Universidade do Minho
Escola de Engenharia

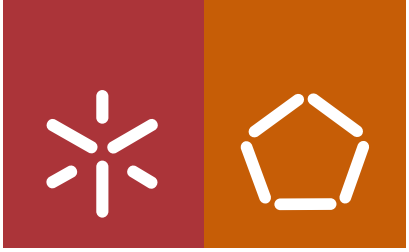
Cristiana Cordeiro de Castro

Monitoring and characterization of yeasts behavior under fermentation processes using technometric approaches

November 2013

Cristiana Cordeiro de Castro **Monitoring and characterization of yeasts behavior under fermentation processes using technometric approaches**

UMinho | 2013



Universidade do Minho
Escola de Engenharia

Cristiana Cordeiro de Castro

Monitoring and characterization of yeasts behavior under fermentation processes using technometric approaches

Doctoral Dissertation for PhD degree in Chemical and Biological Engineering

Supervisors of the thesis:
Doctor José António Couto Teixeira
Doctor Rui Miguel Costa Martins

November 2013

Autor: Cristiana Cordeiro de Castro

E-mail: cristianacastro@deb.uminho.pt/ cris.decastro@gmail.com

Telefone: +351253601966

CC: 12574669

Título da tese:

Monitoring and characterization of yeasts behavior under fermentation processes using technometric approaches

Orientadores:

Doutor José António Couto Teixeira

Doutor Rui Miguel Costa Martins

Ano de conclusão: 2013

Doutoramento em Engenharia Química e Biológica

Universidade do Minho, 21 de Novembro de 2013.

ACKNOWLEDGEMENTS

Undertaking this PhD project has been a truly life-changing experience for me, and it would not have been possible to do without the support and guidance of many people. To those who contributed directly or indirectly to this achievement, I owe my sincere gratitude.

Firstly, I would like to express my sincere thankfulness to my supervisor Doutor José Teixeira for all the support, encouragement, friendship and guidance during my PhD studies. I extend my thanks to my co-supervisor Doutor Rui Martins, for the attendance, scientific knowledge and advices within this work.

I thank Doutor António Vicente for hosting me in the *Laboratório de Indústria e Processos*, to Doutora Lucília Domingues for the kind supply of the industrial PE-2 yeast strain, and Doutor José Maria Oliveira for the technical support in the chromatographic techniques. I also thank Doutor António Silva Ferreira for his indispensable support in the metabolomics field and constant support and motivation.

I acknowledge Fundação para a Ciência e Tecnologia (FCT) for funding me through the scholarship SFRH/BD/46737/2008 and to *Universidade do Minho* for hosting me.

I would also like to thank Engª Madalena Vieira and Sr. Santos, for the technical support and for always being available to help and to provide the necessary technical training.

To Filipe Macieira, João Silva, Ricardo Duarte, Inês Mendes, Caitriona Gunning, Rita Monforte and Carla Oliveira, for their cooperation regarding the experimental work and friendship.

I thank all my colleagues from LIP, for the great lab environment and daily companionship. Also to my friends from DEB and to the "Ladies DEB" team, I thank you all for the amusing moments. A special thank to those who became great friends and helped me in different ways to circumvent most of the less good days during this walk, Ana Cristina, Isa, Kiki, Michele, Joana, Philippe, Miguel, Hélder, Chaves, Artur, Héctor, Xico, thank you for your constant motivation and friendship.

To those who have accompanied me since childhood, Amélia and Arminda, thank you for the continuous companionship, understanding, the always great moments and loyalty over the years. To Marlene, Susana, Nita and Nice, for sticking around along this path, for the constant motivation,

friendship and good moments together. To Bruno, a special friend, who has accompanied me in my professional and personal life and whose dedication and motivation have been a mainstay for me. To Carla, the one with whom I share some professional projects and I have shared the last minute panic while writing this thesis. Thanks for the relaxing moments, the conversations with some sense but mostly nonsense and for being the second *Luna*'s godmother.

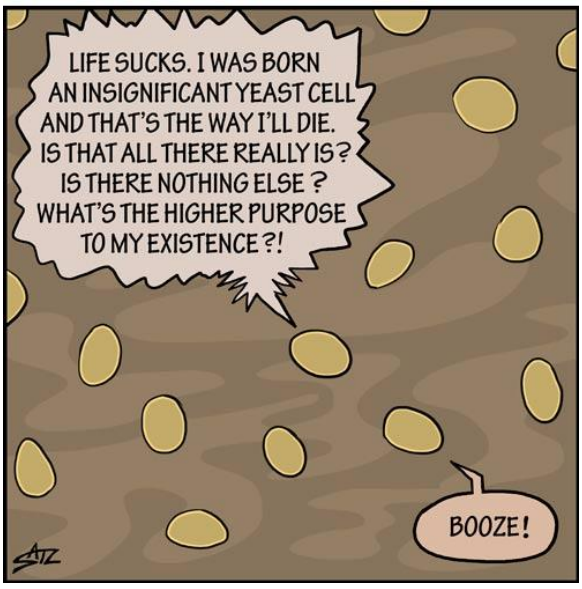
To my grandparents and brother Luís, that although not present among us anymore, have guided and gave me strength every day.

To my parents, Palmira and Manuel, brother Tomás and sister Nanda, the pillars of my life that have followed every step I take, since the moment I could walk. Thanks for the love and the lessons they have given to me, which in just a few words, have taught me to be a better person every day. Thanks for the encouragement, motivation and holding me every moment.

I dedicate this thesis to my parents and nieces Rita and Maria

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

QUALIFICAR É CRESCER.
UNIÃO EUROPEIA
Fundo Social Europeu


LIFE SUCKS. I WAS BORN AN INSIGNIFICANT YEAST CELL AND THAT'S THE WAY I'LL DIE. IS THAT ALL THERE REALLY IS? IS THERE NOTHING ELSE ? WHAT'S THE HIGHER PURPOSE TO MY EXISTENCE ?!
BOOZE!

LIST OF PUBBLICATIONS

According to the 2nd paragraph of the article 8 of the Portuguese Decree-Law no. 388/70, this thesis is based on the following original articles:

- **Castro CC** *et al.*, Phenotypic and physiological characterization of laboratorial and industrial *Saccharomyces cerevisiae* strains as a response to induced stress conditions, *To be submitted.* [Chapter 3]

- **Castro CC** *et al.*, Classification and prediction of *Saccharomyces cerevisiae* strains behavior under induced stress conditions based on metabolic profiles*, To be submitted*. [Chapter 4]

- **Castro CC**, Gunning C, Oliveira CM, Couto JA, Teixeira JA, Martins RC, Silva Ferreira AC (2012) *Saccharomyces cerevisiae* oxidative response evaluation by cyclic voltammetry and gas chromatography–mass spectrometry, *Journal of Agricultural and Food Chemistry,* 60 (29), 7252–7261. [Chapter 5]

- **Castro CC**, Martins RC, Teixeira JA, Silva-Ferreira AC (2014) Application of a high-throughput process analytical technology metabolomics pipeline to *Port* wine forced aging process, *Food Chemistry,* 143, 384-391. [Chapter 6]

SUMMARY

Technometrics concerns on the development and use of statistical methods in different fields, such as biotechnological processes, in order to understand their multivariate and multidimensional complexity. Chemical changes occurring within these processes can be monitored using chemometric tools that combined with bioinformatic methodologies, can provide an enlarged overview of the process, enabling the unbiased study of metabolites and dynamic changes in response to the environmental conditions. For this purpose, different chemometric tools were used, namely relevant principal component analysis (RPCA), multi-way principal component analysis (MPCA), partial least squares logistic regression (PLS-LOG) and unfolded partial least squares (U-PLS).

Phenotypic and physiological behaviors of three different *Saccharomyces cerevisiae* strains, a laboratorial S288c, and two industrials CA11 and PE-2, were evaluated under different stress conditions. Toxic and inhibitory conditions were induced by introducing 1.0% (v/v) ethanol, 1-butanol, isopropanol, *tert*-Amyl alcohol, 0.2% (v/v) furfural and 0.5% (v/v) 5-hydroxymethylfurfural (5-HMF) in batch fermentations with YPD as culture medium. MPCA and PLS-LOG allowed to evidence the different behavior of S288c comparing to PE-2 and CA11, and a higher impact caused by 1-butanol, furfural and 5-HMF in phenotypic and physiological profiles. PE-2 revealed to be the most robust strain, quickly adapting to the environmental conditions, even under the highest stress conditions. It was also observed a correlation between the flocculation profile inhibition under those conditions, with an increased production of intracellular glycerol. This relationship was confirmed by PLS-LOG where intracellular glycerol and trehalose, as well as extracellular acetic acid production showed to be linked to the inhibition of CA11 cells flocculation.

Metabolic changes occurring within CA11 and PE-2 fermentations in the presence of 1-butanol, furfural and 5-HMF were also evaluated, using RPCA. CA11 fermentations enhanced the production of ethanol, isovaleric acid and isoamyl acetate, whereas PE-2 favored the production of more aromatic compounds, such as esters - phenylethyl acetate, ethyl hexanoate, ethyl octanoate and ethyl dodecanoate. These results suggested that PE-2 is less susceptible to the stress effect of the three tested molecules. PLS-LOG models allowed the prediction (R^2=0.90) of the metabolic behavior of both strains during the fermentations: the presence of 1-butanol induced the production of esters ethyl acetate and isoamyl acetate (and its precursor, 3-methyl-1-butanol), as well as butyric acid (which encourages the use of both strains in bio-butanol production systems); CA11 and PE-2 synthesized furfuryl alcohol from furfural; the presence of furfural and 5-HMF induced the production and accumulation of fatty acids in the medium, to counterbalance the inhibitory effects.

The impact of metabolic profile of *S. cerevisiae* PYCC 4653 on its antioxidant capacity, in synthetic grape juice supplemented with phenolics acids was assessed. A bioanalytical pipeline, combining electrochemical features with biochemical background was proposed, for biological systems fingerprinting and sample classification. The electrochemical profile, phenolic acids and the volatile fermentation fraction, were evaluated for 11 days, using cyclic voltammetry, target and non-target metabolic approaches, respectively. It was found that acetic acid, 2-phenylethanol and isoamyl acetate have a significative contribution for samples metabolic variability and the electrochemical features demonstrated redox-potential changes throughout the alcoholic fermentations, showing at the end, a similar pattern to normal wines. *S. cerevisiae* also showed the capacity of producing chlorogenic acid in the supplemented medium fermentation from simple precursors present in the minimal medium. The proposed bioanalytical pipeline proved to be a very efficient strategy for fingerprinting biological systems, by integration of the information from different chemical detectors.

Finally, a non-targeted high-throughput metabolomics pipeline combining GC-MS data preprocessing with multivariate analysis, was developed and integrated in new "*in-house*" software, called *X-Metabolomics* (developed during this thesis). The pipeline was built to enhance the identification of key metabolites involved in the process, through the exploration of the temporal relationships between interesting metabolites related to a chemical phenomenon. It was applied to a *Port* wine "forced aging" process under different oxygen saturation regimes. RPCA showed that the use of extreme oxygen saturation and high temperatures during *Port* wine aging induced the occurrence of chemical reactions undesirable for the aromatic profile, affecting the quality of the final product. Under those conditions an increased production of dioxane and dioxolane isomers and furfural was observed, leading to excessive degradation of the wine aromatic profile, color and taste. The production of dioxane isomer was highly correlated with the production of dioxolane isomer, benzaldehyde, sotolon, and many other metabolites whose identification could be of great interest for their contribution for the final aromatic profile of the *Port* wine.

In sum, during this thesis, the potential of the use of chemometrics and bioinformatics approaches was explored in the characterization (by RPCA and MPCA), classification and prediction (by PLS-LOG and U-PLS, respectively) of physiological, phenotypic and metabolic changes in bioprocesses as an adaptation response to environmental conditions. The joint effect of distinct variables (measured using HPLC, GC-FID, GC-MS and cyclic voltammetry) in multivariate data analysis allowed enhancing the knowledge about chemical and biochemical dynamics in biotechnological processes.

RESUMO

A tecnometria consiste no desenvolvimento e uso de métodos estatísticos em diferentes áreas, tais como processos biotecnológicos, de modo a compreender a sua complexidade multivariada e multidimensional. As alterações químicas que ocorrem nestes processos podem ser monitorizadas utilizando ferramentas de quimiometria que, associadas a métodos de bioinformática, podem proporcionar uma visão alargada do processo e logo, o estudo equitativo dos metabolitos e as alterações dinâmicas em resposta às condições ambientais. Ao longo deste trabalho, diferentes ferramentas de quimiometria foram utilizadas, nomeadamente, *relevant principal component analysis* (RPCA), *multi-way principal component analysis* (MPCA), *partial least squares logistic regression* (PLS-LOG) e *unfolded partial least squares* (U-PLS).

Foi efetuado o estudo de comportamentos fenotípicos e fisiológicos de três estirpes diferentes de *Saccharomyces cerevisiae*, uma laboratorial, S288c, e duas industriais, CA11 e PE -2, sob diferentes condições de stress. Foram adicionadas moléculas tóxicas e inibitórias no meio YPD, nomeadamente, 1,0% (v/v) de etanol, 1-butanol, isopropanol e 2-metil-2-butanol, 0,2 % (v/v) de furfural e 0,5 % (v/v) de 5-hidroxi-metil-furfural (5-HMF). O MPCA e o PLS-LOG evidenciaram o diferente comportamento da estirpe S288c em relação à CA11 e PE-2, e um maior impacto causado pelo 1-butanol, furfural e 5-HMF nos perfis fenotípicos e fisiológicos. A PE-2 revelou ser a estirpe mais robusta e a que melhor se adaptou às condições ambientais impostas, mesmo sob as mais severas. Observou-se uma correlação entre a inibição do perfil de floculação nestas condições, com um aumento da produção de glicerol intracelular. Esta relação foi confirmada utilizando o PLS-LOG onde a produção de glicerol e trealose intracelulares, bem como de ácido acético extracelular mostraram estar associadas ao fenómeno de inibição da floculação das células da CA11.

As alterações metabólicas que ocorrem nas fermentações utilizando a CA11 e PE- 2 na presença de 1-butanol, furfural e 5- HMF também foram avaliadas por RPCA. Enquanto a estirpe CA11 favoreceu a produção de etanol, ácido isovalérico e acetato de isoamilo, a PE-2 levou à produção de outros compostos aromáticos, tais como o acetato de feniletilo, etil hexanoato, octanoato e dodecanoato ao longo das fermentações. Estes resultados reforçam que a PE-2 é menos suscetível ao efeito stressante dessas moléculas. Os modelos PLS-LOG permitiram prever (R^2 = 0,90) o comportamento metabólico de ambas as estirpes, durante as fermentações: a presença de 1-butanol induziu a produção de ésteres de acetato de etilo e acetato de isoamilo (e o seu precursor, 3-metil -1- butanol), bem como o ácido butírico (encorajando a utilização de ambas as estirpes em sistemas de produção de bio-butanol); as estirpes CA11 e PE-2 sintetizaram álcool furfurílico a partir de furfural; a presença de furfural e 5- HMF induziu a produção e acumulação de ácidos gordos, de forma a contrabalançar os efeitos inibitórios na obtenção de energia para as células, metabolizando ácidos gordos no meio.

O impacto do perfil metabólico da *S. cerevisiae* PYCC 4653 sobre a capacidade antioxidante foi avaliado, em fermentações utilizando sumo de uva sintético suplementadas com ácidos fenólicos. Foi apresentada uma metodologia bio-analítica (combinando os perfis eletroquímico e bioquímico) para a caracterização do comportamento da levedura em resposta às perturbações impostas. O perfil eletroquímico, os ácidos fenólicos e a fração volátil das fermentações, foram avaliados durante 11 dias, utilizando a voltametria cíclica, e abordagens metabólicas supervisionadas e não supervisionadas. Verificou-se que o ácido acético, 2- feniletanol e o acetato de isoamilo têm uma contribuição significativa na variabilidade metabólica e as características electroquímicas revelaram as alterações do potencial redox durante as fermentações. O perfil eletroquímico da fermentação alcoólica mostrou, no final, um padrão semelhante ao dos vinhos reais. A *S. cerevisiae* também mostrou a capacidade de produzir ácido clorogénico, no meio de fermentação suplementado a partir de precursores simples, presentes no meio mínimo. A metodologia proposta provou ser uma estratégia eficiente na caracterização de fenómenos biológicos e químicos, através da integração da informação de vários detetores químicos.

Por fim, uma metodologia de processamento metabólico não-direcionado e de alto-débito, combinando o pré-processamento dos dados de GC-MS com a análise multivariada, foi desenvolvida e integrada num novo software, denominado *X-Metabolomics* também desenvolvido no decorrer desta tese. A metodologia foi construída para melhorar a identificação dos metabolitos-chave envolvidos no processo biotecnológico, através da exploração das relações temporais entre os metabólitos interessantes relacionados ao mesmo fenómeno químico. Esta foi aplicada a um processo de "envelhecimento forçado" de vinho do Porto, sob diferentes regimes de saturação de oxigénio. O RPCA mostrou que a utilização da saturação extrema de oxigénio e de temperaturas elevadas durante o envelhecimento do vinho do Porto induziu a ocorrência de reações químicas indesejáveis para o perfil aromático, que afetam a qualidade do produto final. Nestas condições, foi observado um aumento da produção de isómeros de dioxano e dioxolano e furfural, que levaram a uma degradação excessiva do perfil aromático, cor e sabor do vinho. A produção do isómero de dioxano está altamente correlacionada com a produção de um isómero dioxolano, benzaldeído, sotolon, e muitos outros metabolitos, cuja identificação poderia ser de grande interesse pela sua contribuição para o perfil aromático final do vinho do Porto.

Em suma, durante esta tese, foi explorado o potencial da utilização de abordagens de tecnometria, incluindo métodos de quimiometria e bioinformática, na caracterização (por RPCA e MPCA), classificação e previsão (por PLS-LOG e U-PLS respetivamente) das alterações fisiológicas, fenotípicas e metabólicas em bioprocessos, em resposta às condições ambientais. O efeito conjunto de distintas variáveis na análise multivariada, permitiu ampliar o conhecimento acerca das dinâmicas químicas e bioquímicas em processos biotecnológicos.

INDICE

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

5-HMF	5-hydroxymethyl-furfural
AAT	alcohol acetyltransferase (AAT)
ABE	Acetone/butanol/ethanol pathway
Acetyl-CoA	Acetyl coenzyme A
APCI	Atmospheric pressure chemical ionization
APPI	Atmospheric pressure photo ionization
ATP	Adenosine 5'-triphosphate
ATPase	Adenylpyrophosphatase
CA	Cluster analysis
CE	Capillary electrophoresis
CE-MS	Capillary electrophoresis - mass spectrometry
CFU	Colony-forming unit
CI	Chemical ionization
COW	Correlation optimized warping
CV	Cyclic voltammetry
DIMS	Direct injection mass spectrometry
DNA	Deoxyribonucleic acid
DVB/CAR/PDMS	Divinylbenzene/carboxen/polydimethylsiloxane
EI	Electron impact ionization
ESI	Electrospray ionization
FID	Flame ionization detector
FT-MS	Fourier transform – mass spectrometry
GC	Gas chromatography
GC-FID	Gas chromatography – flame ionization detector
GC-MS	Gas chromatography – mass spectrometry
GUI	Graphical user interface
HILIC	Hydrophilic interaction chromatography
HILIC-MS/MS	Hydrophilic interaction chromatography - mass spectrometry/mass spectrometry
HMBD	Human metabolome database
HPLC	High performance liquid chromatography
HPLC-DAD-MS	High-performance liquid-chromatography – diode array detector – mass spectrometry
HPLC-MS	High performance liquid chromatography - mass spectrometry
HPLC-MS/MS	High performance liquid chromatography - mass spectrometry/mass spectrometry

HS/SPME	Head-space solid phase micro extraction
HS-SPME/GC-MS	Head-space/solid-phase-micro-extraction/gas-chromatography-mass-spectrometry
ICA	Independent component analysis
KEGG	Kyoto Encyclopedia of Genes and Genomes
LC	Liquid chromatography
LC-MS	Liquid chromatography – mass spectrometry
LC-MS/MS	Liquid chromatography – mass spectrometry / mass spectrometry
LC-NMR-MS	Liquid chromatography – nuclear magnetic resonance – mass spectrometry
LDA	Linear discriminat analysis
m/z	Mass-to-charge ratio
MALDI	Matrix assisted laser desorption / ionization
MCFA	Medium chain fatty acids
MCP	Micro-channel plate
MPCA	Multi-way principal component analysis
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry ionization
MVDA	Multivariate data analysis
NAD^+	Nicotinamide adenine dinucleotide
NADH	Reduced nicotinamide adenine dinucleotide
$NADP^+$	Nicotinamide adenine dinucleotide phosphate
NADPH	Reduced nicotinamide adenine dinucleotide phosphate
NIPALS	Nonlinear iterative partial least squares
NIST	National Institute of Standards and Technology
NMR	Nuclear magnetic resonance
N-PLS	N-way partial least squares
OD_{600}	Optical density at 600 nm
O-PLS	Orthogolnal partial least squares
PARAFAC	Multi-way decomposition method
PC	Principal component (first - PC1; second - PC2)
PCA	Principal component analysis
PCR	Principal component regression
PLS	Partial least squares
PLS-1	Partial least squares 1
PLS-DA	Partial least squares for discriminant analysis
PLS-LOG	Partial least squares logistic regression

PLSR	Partial least squares regression
PPP	Pentose phosphate pathway
Q-statistic	Square prediction error
Q_α	Q statistic confidence interval
RNA	Ribonucleic acid
ROS	Reactive oxygen species
RPCA	Relevant principal component analysis
RT	Retention time
S/N	Signal-to-noise ratio
SGD	Saccharomyces genome database
SGJ	Synthetic grape juice
SPE	Solid phase extraction
SPME	Solid phase micro-extraction
SVD	Singular value decomposition
T^2_h *Hotelling*	Distance to center
T^2_α	T2h Hotelling confidence interval
TBAP	Tetrabutylammonium
TCA	Tricarboxylic acid cycle
TDC	Time-to-digital converter
TOF	Time-of-flight
U-HPLC	Ultra high performance liquid chromatography
U-HPLC-MS	Ultra high liquid chromatography – mass spectrometry
U-PLS	Unfolded partial least squares
VHG	Very high gravity systems
YM	Yeast malt extract
YPD	Yeast peptone dextrose
YPD_b	Yeast peptone dextrose broth

Notes:

In general, the International System of Units (SI) was used in this work. However, sometimes multiples and sub-multiples of the fundamental units (which are not SI) were used since their use is so common that is allowed by that system. Units not recognized by the SI were also used to express some variables, such as the volume percent (% v/v) and the mass percent (% w/v) to denote the composition of some solutions and the revolutions per minute (rpm) to indicate the agitation rates.

CHAPTER 1

MOTIVATION AND OUTLINE

This chapter introduces the background information about the theme of the work, as well as its objectives. The outline of the thesis is also presented.

1.1 CONTEXT AND MOTIVATION

The understanding of living systems by exploring the dynamic patterns of the relationships of organisms with their environment is the basis of the theory of biological sciences. Yeasts growth, fermentation, maturation and storage processes are some examples of dynamic systems that involve chemical and biochemical changes that confer the character and quality of the final product [1-3].

Fermented food and beverages production is accompanied by the production of several compounds, namely alcohols, aldehydes, organic acids, esters, organic sulfides, carbonyl compounds and fusel alcohols, which contribute to flavors and aromas quality of the final product [4, 5]. The industrial handling of yeasts for food and beverage and also for bio-fuel production, can introduce distinct disturbances, mainly osmotic, oxidative, temperature, nutrient starvation, ethanol toxicity, by-products inhibition, among others [6, 7].

Saccharomyces cerevisiae is one of the most used microorganisms which can undergo different phenotypic, morphological and metabolic or physiological changes [8]. The way how different yeast strains respond to the external conditions can be significantly different, according to yeasts genomic information. Therefore the understanding of how cells behave is of great importance for fermentations monitoring and to provide an external control of the process, inducing cells to grow in a particular conformation or to produce a specific end-product of fermentation [9]. Although the yeast is the most studied eukaryote microorganisms, the increased diversity and complexity of the cellular processes dynamic, sets forth the need of the existence of high-throughput methods of analysis as well as different approaches for handling the massive amounts of information.

Biochemical processes can be monitored by measuring different metabolites and fermentation parameters throughout the process [10]. Liquid or gas chromatography coupled with mass spectrometer [11, 12] and cyclic voltammetry [13, 14] are some of the analytical techniques that have been used in this context. Gathering the maximum metabolic information about the biological process is one challenge of metabolomics field [15, 16], and different high-throughput metabolomics tools have emerged in the last years [17-19] for this purpose. However, the high content of information resulting from these untargeted and high-throughput methods, creates the need for statistical tools capable of extracting the crucial information about '*in vivo*' process,

putting a biological meaning of the preprocessed output and increasing the knowledge about the overall process.

The main goal of this thesis is the application of technometric tools in order to monitor, understand and predict the biochemical changes occurring within dynamic biological systems, as an adaptation to the induced environmental conditions. Different multivariate statistical tools were applied to target physiological and metabolic profiles matrixes from batch fermentations using different *S. cerevisiae* yeast strains. Therefore, in order to enlarge the knowledge of undergoing metabolic changes, a high-throughput metabolomics pipeline was used and applied to a *Port* wine forced aging process. It was also purposed to demonstrate the potential of combining metabolomic preprocessing tools (bioinformatics) with mathematical and statistical methods (chemometrics) for gathering the information about metabolites profiling, relations with other metabolites in the process, contextualization in biochemical pathways and new metabolites discovery.

For this purpose, the following main topics were focused:

- Characterization of physiological behavior of *S. cerevisiae* S288c, CA11 and PE-2 yeast strains under different stress conditions during batch fermentations.
- Early detection of stress molecules throughout the fermentation process using multivariate tools.
- Characterization and prediction of extracellular metabolic profile of CA11 and PE-2 yeast strains under induced stress conditions.
- Evaluation of the impact of a *S. cerevisiae* strain metabolism in the profile of compounds with antioxidant capacity in a synthetic wine during fermentation using cyclic voltammetry and gas chromatography – mass spectrometry.
- Application of a metabolomic pipeline for high-throughput data obtaining and characterization of forced aging process of *Port* wine.

1.2 OUTLINE OF THE THESIS

The main goal of this work was to answer the question "**how is it possible to monitor and maximize the information and knowledge about yeast strains behaviors under different fermentation conditions?**" An attempt to answer it is presented in this thesis, which is divided in eight chapters:

- In **Chapter 1**, the context and motivation of the thesis, as well as the studies aims and the global structure of the thesis are presented.

- The general literature overview concerning to yeasts capabilities and usage, metabolomics potential and tools for fermentations understanding and chemometric methodologies applied to high-throughput metabolomics within chemical and biochemical processes is presented in **Chapter 2**.

- The Experimental Results are presented from **Chapter 3** to **Chapter 6**. Each Chapter includes *Introduction*, *Materials and Methods*, *Results and Discussion* and *Conclusions* sections according to the objective of the experimental work.

- In **Chapter 3** it was performed a physiological characterization of S288c, CA11 and PE-2 yeast strains under different stress conditions in batch fermentations. Multi-way principal component analysis and partial least squares logistic regression were the statistical tools used for the characterization of physiological changes of yeasts, as well as to classify the flocculation phenotype changes in CA11, as a response to the induced stressful molecules within each process.

- In **Chapter 4**, principal component analysis and partial least squares for discriminant analysis were applied to a metabolic dataset resulting from liquid- and gas-chromatography analytical techniques (high performance liquid chromatography (HPLC), gas chromatography–flame ionization detector (GC-FID) and gas chromatography–mass spectrometry (GC-MS)) in order to characterize and predict the metabolic state of industrial yeasts strains (CA11 and PE-2) under induced stress conditions.

- In **Chapter 5** it is proposed a bioanalytical pipeline for biological systems fingerprinting and sample classification by combining electrochemical features with biochemical background. The

methodology was applied to a synthetic wine fermentation where the impact of a *Saccharomyces cerevisiae* strain in the profile of compounds with antioxidant capacity was evaluated.

- A high-throughput metabolomics pipeline, applied to *Port* wine maturation process is presented in **Chapter 6**, where it is shown how it is possible to gather the maximum information of the biochemical changes occurring during the forced aging process, based on the target supervision of interesting metabolites involved in the process.

- **Chapter 7** presents the overall conclusions and suggestions for future work.

- Finally, **Chapter 8** gathers all the references used in the elaboration of this thesis.

CHAPTER 2

LITERATURE REVIEW

In several industrial fermentations, yeasts are subjected to harsh conditions to which they are capable to resist by adapting their entire machinery that is, their genome, metabolome, transcriptome, proteome and/or flux dynamics. Different yeast strains respond in different ways to those conditions, according to their properties, biogeographical distribution or applications.

The knowledge of exometabolome provides improved information about different biochemical changes taking place in the fermentation media. However, the increased complexity of the biological processes leading to cellular structure and function, as well as to the emergence and evolution of organisms and species, emphasizes the need of technometric tools, in order to unravel and explore the enlarged convoluted information of the living system.

The application of technometric tools to metabolomics involves the application of bioinformatic methodologies for high-throughput analysis of the analytical signals, combined with chemometric approaches, *i.e.,* statistical and mathematical methodologies, used to extract the relevant biological information. These tools facilitate the characterization and prediction of yeasts metabolic behavior inside bioreactors, enhancing the understanding of the complex interactions of, and within bionetworks, and fundamental relations essential to microorganisms' life.

This Chapter is focused on the contextualization of the application of technometric tools to metabolomics data sets, for crucial data extraction and metabolic profile characterization and prediction within any biological or biochemical system.

2.1 INTRODUCTION

Industrial microbiology or microbial biotechnology concerns the study and development of technologies to control and manipulate microorganisms' growth and activities in order to produce desirable substances or changes in products promoting economic gain or preventing economic loss [20].

Yeasts are the most extensively used microorganisms in industry. *Saccharomyces cerevisiae* yeast strains are highly specialized organisms, which have evolved to apply their full potential through different environments or ecological niches provided by manufacturers [21]. These strains have been used for a long time as a model for identifying genes and pathways involved in basic cellular processes, including cell cycle, aging, and stress response [22].

Alcoholic fermentations and maturation processes, are two distinct forms of biotechnological applications of yeasts [23], which undergo continuous enzymatic and non-enzymatic changes according to external physical factors and the biological activity of the fermenting organisms [3, 24]. Both processes highly influence the organoleptic character and quality of the fermented product [25], although they introduce chemical or physical parameters, unfavourable to yeasts performance. Chemical changes can lead to nutrients availability or concentration (*e.g.* fermentable sugars, assimilable nitrogen, oxygen, vitamins, minerals) and the presence of inhibitory conditions (*e.g.* ethanol, acetic acid, furfural and 5-hydroxymethyl-furfural), while physical parameters include pH, temperature, agitation and osmotic pressure [21] In far-from optimal fermentation or maturation conditions, yeasts are able to survive, as these are equipped with a molecular machinery capable to maintain their integrity and metabolic activity [26].

The sensing of the environmental signals is carried out by specific receptor proteins, generally located in cells surface that transmit the information by interconnected signal transduction pathways to the different cellular compartments which implement an adaptive response, a process referred to as "stress response" [21]. As a consequence, different chemical and biochemical reactions take place, and yeasts attempt to adapt efficiently to the changing environment and/or unfavorable growth conditions [27].

To understand the overall complexity linked to the different defense mechanisms triggered by yeasts cells is one of the major challenges in biotechnology. In this context, distinct technometric

approaches have been developed, in order to extract the maximum meaningful chemical and biochemical information about the bioprocesses complexity [28]. Technometrics approaches concern to bioinformatic tools, capable of extracting an increased information from any analytical experiment and any type of instrumental signal, combined with chemometric algorithms, that is, mathematical and statistical multivariate methods for extracting the relevant information [29]. These can be applied to the different 'omic' fields (*e.g.* genomics, transcriptomics, proteomics or metabolomics), in order to provide a comprehensive overview of the response of biological systems to disease, genetic and environmental perturbations [30].

Metabolomics was the 'omics' field explored in the present work and consists in the analysis of all the small molecular weight metabolites within the process [31-33]. In metabolomics, the acquisition of a meaningful metabolic information resulting from the activities of metabolic pathways can be enhanced by following: *1)* the application of practical approaches, such as an efficient sample preparation; *2)* the selection of the adequate analytical instrument; *3)* the use of chemometric and bioinformatic tools for selective extraction of metabolites within the different analytical signals; *4)* the use of statistical methodologies, such as principal component analysis and partial least squares regression for data classification and prediction [11, 17, 18, 34-38]. Therefore, it is possible to extract an accurate and relevant metabolic information related to yeasts metabolism changes under specific environmental conditions [32].

So, only a holistic and multi-scale approach comprising analytical chemistry, signal processing, bioinformatics and statistical algorithms produces the necessary amount of information to take advantage of the natural evolution from individuals to colonies into new production strategies.

2.2 YEASTS METABOLISM

Metabolism concerns to all biochemical reactions and transformations of living cells, mediated by enzymatic reactions, that allow microorganisms to grow, reproduce, maintain their structures and respond to the involved environments [39]. German *et al.* [40] define metabolism as the key feature to microorganisms phenotype, and metabolites distribution results from functional genomics.

The main carbon and energy source for yeasts is glucose that is converted to pyruvate throughout the glycolytic pathway and to anabolites and energy in the form of adenosine triphosphate (ATP) throughout the tricarboxylic acid (TCA) cycle [41]. Yeast cells can gain energy from two processes, namely respiration and fermentation [39, 42], both regulated by glucose and oxygen concentrations [41], as it is possible to observe in Figure 2.1. In cellular respiration pyruvate is decarboxilated in the mitochondrion to acetyl coenzyme A (acetyl-CoA), being completely oxidized in the TCA cycle to carbon dioxide (CO_2), energy and intermediates to promote yeast cells growth. In the alcoholic fermentation process, while glucose is gradually used to produce the required energy to keep cells alive, it is not completely oxidized to ethanol and CO_2 [39, 41, 42].

Beyond ethanol and CO_2, during alcoholic fermentation other quantitatively important metabolites are produced, including polyols, specially glycerol and 2,3-butanediol, and organic acids, such as succinic, keto and acetic acids [3, 43], as it is presented in Figure 2.2. Other minor metabolites, such as higher alcohols, esters, volatile fatty acids and carbonyls can be produced, representing < 1% of sugar carbon, which contribute to final flavor-active properties. Because of these metabolic losses occurring in the process, the complete fermentation of hexose sugars only yields 94 - 96% of the theoretical maximum ethanol yield [43]. Yeasts species, fermentation conditions and nutrient content of the wort must modulate the production of these compounds [3, 44].

The growth of microorganisms tends to follow a specific order [43] during alcoholic fermentation. Distinct phases - lag, accelerating growth, exponential, decelerating growth and stationary phases - can be identified, all of which are part of the yeast cells life cycle pitched into fresh wort. Each phase can be described separately and its relative time depends on different factors namely the wort's composition, the environment conditions and the amount of yeast used. Lag phase corresponds to the first phase of the cycle, where yeast cells adapt to the environment while use their internal reserves, namely carbohydrate glycogen, to produce enzymes in order to grow and

ferment the wort. During this lag phase, yeast cells are biochemically active but still do not divide [6].

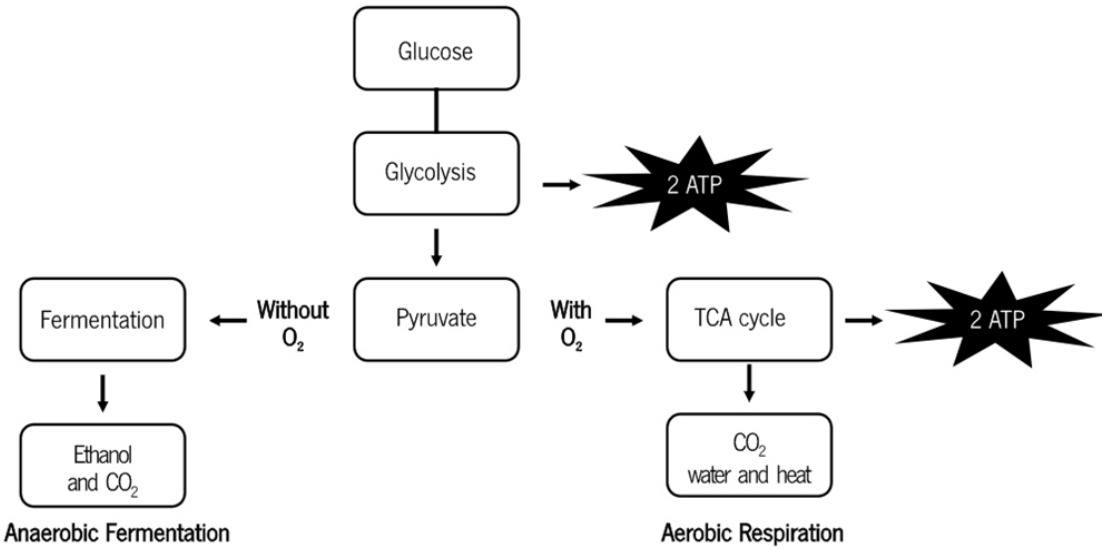


Figure 2.1 The influence of oxygen and ATP throughout yeasts metabolism.

While adapting, yeasts assess the dissolved oxygen level, the overall and relative amounts of amino acids and the overall and relative amounts of sugars present in the wort, which will be further used for cells division [45, 46]. So at the end of lag phase, yeasts will move to the next phase of the life cycle, the growth phase during which yeast cells start to grow and divide, beginning also to store sugar in the form of glycogen for later use. In the third phase, the exponential phase, yeast reproduction is increased as it is now completely adapted to the conditions of the wort and transport of both amino acids and sugars into the cells for metabolism will be very active. During this period fusel alcohols can be produced.

After the exponential phase, in decelerating phase, yeasts begin anaerobic metabolism, with the production of CO_2, since all of the oxygen has now been depleted. Finally, the stationary phase of microbial growth in a batch fermentation process occurs when the number of cells dividing and dying is in equilibrium and can be the result of the following: depletion of one or more essential growth nutrients and/or accumulation of toxic by-products [47].

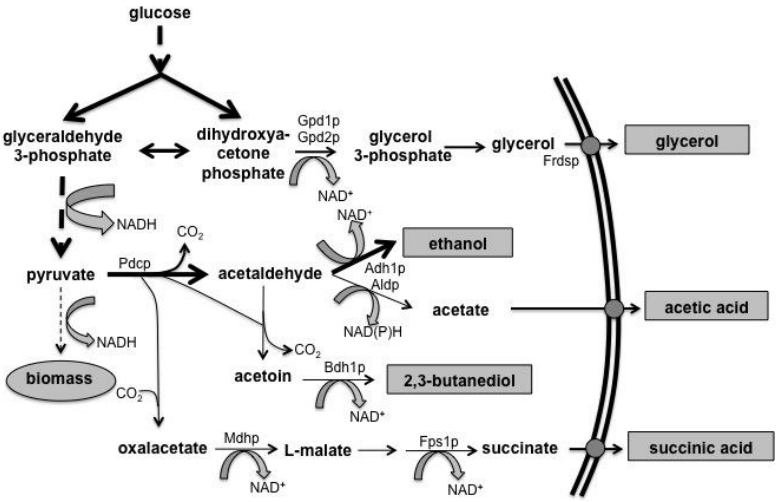


Figure 2.2 Main metabolites and biomass production during alcoholic fermentation by yeasts (Adapted from Ugliano and Henschke (2009) [43]).

2.3 YEASTS IN INDUSTRIAL FERMENTATION PROCESSES

Industrial biotechnology, also known as white biotechnology, is providing some solutions for facing many hurdles introduced by the chemical synthesis of products, namely the generation of large amounts of waste for manufacturing a limited range of products. The implemented solutions result in the depletion of fossil fuel reserves and the increasing global environment problems, filling the higher demand of energy and consumer products. The use of renewable feedstock for chemical production, as well as the use of enzymes and microorganisms, as *S. cerevisiae* yeasts, are some of the strategies applied to make useful products in different industries, such as food, chemicals and feed, paper and pulp, textile and energy [48].

Yeasts have long been used to ferment the sugars of rice, wheat, barley, and corn to produce alcoholic beverages, fuel ethanol and in the baking industry to expand, or raise, dough. The combination of the alcoholic fermentation process resulting metabolites is responsible for the chemical and organoleptic profile of the fermented product. Alcoholic fermentation is a combination of complex interactions involving must or wort variety, microbiota and fermentation technology [49]. In this context, different factors can affect yeasts and fermentations behaviors and thus the quality of the final product, namely fermentation and/or must temperatures [49, 50], inoculum size [51], type of yeasts used [52], fermentable and non-fermentable sugars availability [8, 53], and pH [8].

S. cerevisiae is clearly the ideal eukaryotic microorganism for biological studies as the "awesome power of yeast genetics" has become legendary and the complete sequence of its genome has proved to be extremely useful as a reference towards the sequences of human and other higher eukaryotic genes [54]. In addition, *S. cerevisiae* cells have several prominent useful features, namely the cheap and easy cultivation, short generation times (rapid growth), a highly versatile deoxyribonucleic acid (DNA) transformation system, the detailed genetic and biochemical knowledge accumulated in many years of research. These particularities convert yeast, a simple unicellular eukaryote, into an unique powerful model system for biological research and application in industrial processes [55].

The understanding of *S. cerevisiae* cell cycle, growth, communication and morphology differentiation inside colonies is of great importance for industrial applications of the yeast [56]. Recent studies have shown that cell cycles inside *S. cerevisiae* colonies are highly dynamic, not

only alternating between haploid and diploid, but being capable of changing to filamented morphologies at specific points of the colony due to long-range communication molecules ("*quorum sensing*"), such as fusel alcohols, as a foraging response and subsequent robust adaptation of the colony to more rigorous metabolomic conditions [57].

So, the use of specialized chemometric tools, capable of extracting robust and enlarged information about yeast cells metabolic behavior inside bioreactors, as well as, combined with classification and prediction tools, able to manage and predict the metabolic response of yeasts as response to the environmental conditions, can have profound positive implications in industrial processes [29, 58, 59]: *i)* increase the knowledge about the biological process and how to manage yeast colonies inside biofilms or flocculating fermentations; *ii)* understand the interactions between intracellular and/or extracellular metabolites and metabolic pathways involved in the response mechanism; *iii)* recognize physiological and phenotypic changes occurring in yeasts cells, associated to specific changes in the metabolic behavior; *iv)* manage the colony dynamics to produce desired combinations of metabolites (such as the management of the *Ehrlich* pathway for flavor chemistry) nutraceuticals and anti-oxidant agents production (isoprenoids, flavonoids) [42]; and finally *v)* reduce time, energetic and manufacturing costs in industrial biotechnological processes.

2.3.1 Beverages and flavor production mechanisms

In beverages production, different microorganisms can be used to produce aroma-active compounds, including esters, higher alcohols, carbonyls, volatile fatty acids and sulfur compounds. These compounds are the result of sugar and amino acids metabolism, that provide final chemical quality and character to alcoholic beverages that include wine, cider, sake, beer and distilled spirits (*e.g.* whisky) [3, 8].

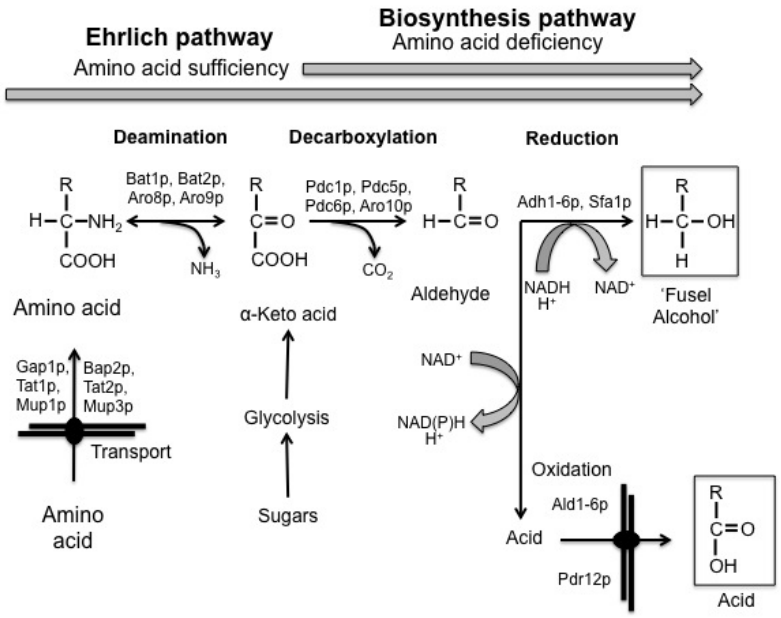


Figure 2.3 Formation of higher alcohols and esters from sugars and amino acids (Adapted from Ugliano and Henschke (2009) [43]).

The complex aroma of beverages can result directly from chemical components of must/wort or from the different mechanisms involved in the production process. Must composition depends on the raw substrates variety, characteristics of the soil, climatic conditions and techniques used in the production process [43]. Production processes include chemical molecules released and/or modified by the action of flavor-active yeasts and bacteria or arises from molecular activities of these microorganisms [4, 5, 60] during fermentation and chemical transformations occurring generally during maturation and aging processes in both bottle or barrel vessels. During fermentation and subsequent maturation and aging, a variety of yeast biochemical mechanisms are involved, including hydrolysis, transformation reactions such as reduction, esterification and decarboxylation, oxidation and metabolite-induced condensation reactions [43], as it is shown in Figure 2.3.

Ethanol, higher alcohols and esters characterize the aroma of the fermented product, being alcohols formed by yeasts and bacteria within the fermentation process [3]. Ethanol is the major product resulting from sugars conversion by yeasts during fermentation, and it is capable of masking the "fruitiness" of beverages, that is the esters contribution [61]. Glycerol is also a major product of alcoholic fermentation [62], chemically it is a polyol with a colorless, odorless and highly

viscous character, and it has also a slightly sweet taste and a oily and heavy mouth-feel [3]. Both ethanol and glycerol production mechanisms are presented in Figure 2.2. The secondary yeasts metabolites include higher alcohols, also known as fusel alcohols, which can have positive or negative effects on the aroma and flavor of the final product. At optimal levels, these contribute with "fruity" character, although, excessive concentrations can result in a strong, pungent smell and taste [44]. Higher alcohols are quite stable throughout the aging process [63]. Branched-chain higher alcohols are produced throughout the *Ehrlich* pathway [4, 5], which involves the degradation of branched-chain amino acids (synthesized via the catabolic or *Ehrlich* pathway or an anabolic pathway though their biosynthetic pathway from glucose [3]). The amino acids uptake by *S. cerevisiae* is mediated by transport proteins, called permeases [64]. The first step of the *Ehrlich* pathway involves the transamination to form the respective α-keto acids, which are therefore converted to the corresponding branched-chain aldehydes by a pyruvate decarboxylase. Finally, an alcohol dehydrogenase catalyses the reduced nicotinamide adenine dinucleotide (NADH)-dependent reduction of this aldehydes to the corresponding fusel alcohol [3]. Instead, the aldehydes might be oxidized to an acid [65].

Volatile acidity leads to a group of volatile organic acids of short carbon chain-length. Free or saturated volatile fatty acids represent 10 – 15% of the total acid content, and of this, acetic acid represents about 90% of the volatile acids produced by yeasts and bacteria [66]. Volatile acids concentrations increase and decrease within beverages aging, mainly due to chemical hydrolysis of some fatty acid ethyl esters occurring during the formation of acid compounds [63].

The production of esters can have a significant influence on the fruity flavors of the final product. Two esters groups can be found, namely ethyl acetate esters and fatty acid ethyl esters, which production can occur by both chemical reactions or via intracellular enzymatic reactions during fermentation [67]. Esters concentrations can differ during the aging process, leading to the occurrence of "fruity" flavors loss. Thus, while acetate esters usually decrease during the aging process, fatty acid ethyl esters behavior depends on the structure of the fatty acid carbon chain. Straight-chain fatty acids ethyl esters concentration decrease over time and branched-chain fatty acid ethyl esters are stable and concentrations can increase during aging [68].

In industrial beverages fermentations, yeasts must respond to environmental fluctuations, leading to dissolved oxygen concentrations, pH, osmotic potential, ethanol concentration, nutrient supply and temperature [8, 69].

After pitching yeast cells into the aerated wort in the fermentor, these start to adapt to the external conditions, and after a brief lag phase, yeasts grow exponentially, rapidly depleting the available oxygen and creating an anaerobic environment [8]. Sugars and assimilable nutrients in the wort are also used rapidly, resulting in carbon and nutrient limitation that lead cells to a resting state [70]. This limitation of carbon and nutrients is coincident with an increase in ethanol concentration, which represents a stress condition to yeasts [3]. Also, the production of other minor metabolites as higher alcohols during the fermentation process, at high concentrations can introduce oxidative stress to cells [71].

The maturation and subsequent aging process can also introduce a series of complex transformations, namely formation or degradation of varietal compounds, oxidation, reduction, esterification or hydrolysis mechanisms, which can result in significant changes to its aroma composition and quality especially when it occurs in oak barrels or bottles [43], as the Port [72, 73] and *Madeira* wine aging [74].

Port wine

Port wine is a fortified wine, with specific quality and character produced in *Douro Region* in the North of Portugal. A complete and fast extraction of both color and flavor from tannins of the grape skins must be performed before adding the fortifying spirit (about 77% alcohol) after two or three days to stop the must fermentation. The unfermented sugars attribute an increased sweetness to the *Port* wine, which at the end of the process reaches 19 - 21% (v/v) alcohol. Ruby *Port* wines are aged in bottles while Tawny *Port* wines are stored in barrel oaks. The different aging conditions of *Port* wines lead to different final aromatic profiles.

The aging process knowledge has an increased commercial interest as this is a time-consuming process during which the wine acquires important and specific organoleptic properties, which is reflected in the final product prices [75]. Usually, during the aging process aromas linked to the fermentation and must properties are lost and new aromas characteristic of older wine or uncommon aromas associated with wines deterioration (or oxidation) are produced [76]. Also, the storage of wine on oak barrels introduces some "woody" characteristics to the wine, such as cis- ("vanilla") and trans- ("coconut like") oak lactones [77], aldehydes - furfural and 4-methylfurfural – that provide "sweet", "butterscotch" and "woody" aromas [78], 2-furanmethanethiol ("roasted

coffee") [79], guaiacol and 4-methylguaiacol, which impart "smoky" aromas to wine and indicate the level of the toasting or charring of oak barrels [77]. An important volatile metabolite which is characteristic of *Port* and *Madeira* wines is called sotolon (3-hydroxy-4,5-dimethyl-2(5H)-furanone) and it is known to have an intense odor of curry [72].

After a long period of maturation, the non-enzymatic reactions, namely the *Maillard* reaction, a form of non-enzymatic browning resulting from the reaction of a reducing sugar and an amino acid, and the related *Strecker* degradation of amino acids induce aroma changes. Browning is a desired characteristic in *Port* wines, although it is unpleasant in white wines [72].

Hydroxymethyl-furfural and furfural can be formed during the *Maillard* reaction or from the dehydratation of sugars in acidic medium and caramelization [80], and their concentrations are higher in wines stored in oak barrels which are mainly formed during the toasting of the oak and can be release to the wine during aging [81].

Also, during the second part of the *Maillard* reaction, the *Strecker* degradation other compounds, such as aldehydes, are formed, which attribute also aromatic characteristics to wines.

2.3.2 Bio-fuels production

The increased awareness worldwide related to energy costs, security and environmental problems concerning to petroleum-derived transportation fuels are drawing an enlarged interest in the alternative bio-fuels. The term bio-fuel leads to any liquid fuel produced from renewable resources as an alternative for petroleum-derived fuel. Bio-fuels have the potential to reduce either the dependency of countries on oil imports and the greenhouse emissions causing the global warming [82]. Two of the most common bio-fuels are ethanol and butanol.

Bio-ethanol production

Bio-ethanol is a renewable and environment friendly alternative to fossil fuels which has been produced using sugars from different sources, namely, plants (*e.g.* sugar cane, sugar beet, sweet sorghum) and plants starch (*e.g.* maize, wheat, rye), both included in the first generation of bio-ethanol processes, and plants cellulose (herbaceous plants, agriculture residues), known as

second generation of bio-ethanol processes [83]. A third generation of bio-ethanol production processes concern to the use of algae, while the fourth generation status is claimed by every new technology, such as genetic manipulation or nanotechnologies [84, 85].

Depending on the biomass feedstock used for bio-ethanol production, different steps are needed in order to produce fermentable single sugar molecules. In first generation processes, pretreatments of sugar containing plants is mainly focused on removal of impurities and disinfection, whereas in second generation starch and cellulose containing plants need to be preprocessed and distinct technologies have been applied for this purpose [86]. Although, the application of preprocessing methodologies, based on enzymatic or chemical hydrolysis processes applied to lignocellulosic biomass, leads to the generation of monomeric sugars, capable to be assimilated by yeasts, stressful conditions are introduced. Oxidative stress is generally caused by the resulting inhibitory material, such as furans derivates – furfural, 5-hydroxymethyl-furfural (5-HMF), phenolics and organic acids [86-88]. 5-HMF concentrations can vary from 2.0 to 5.9 g/L, and furfural is usually found in lower levels than 5-HMF, around 1 g/L, however high enough to be an inhibitor [87]. The use of very high gravity systems (VHG) introduces also osmotic and oxidative stress to yeasts cells leading to the higher concentrations of sugar substrates (> 250 g/L), typical in these systems [89] and thus to the final increased ethanol concentrations in the medium (85 – 100 g/L) [88, 90].

Increased sugar concentrations causes an increased accumulation of intracellular ethanol in yeast cells which promotes an unfavorable effect on intracellular enzymes necessary to ethanol production [91]. The high concentrations of ethanol at the end of VHG process can affect cell membranes, cellular pH and nutrient transport processes [92-94]. Several studies stated that furans are linked to enzymes inhibition, namely alcohol, pyruvate and aldehyde dehydrogenases, and thus the glycolytic pathway, which cause a reduced ATP synthesis and in some cases DNA damages [87, 95-98]. Aliphatic acids also interferes in ATP availability as this is forwarded to maintain the cellular pH and thus inhibits cells growth [87, 99], whereas phenolics mostly affect the cell membrane integrity by disruption of sugar transport and cells growth [100, 101]. The inhibition mechanisms of furan-derivates, weak acids and phenolics are schematized in Figure 2.4. All of these effects concern to a decrease of fermentations performance under VHG conditions, characterized by loss of yeast viability, reduced fermentation rates and incomplete fermentations [102, 103].

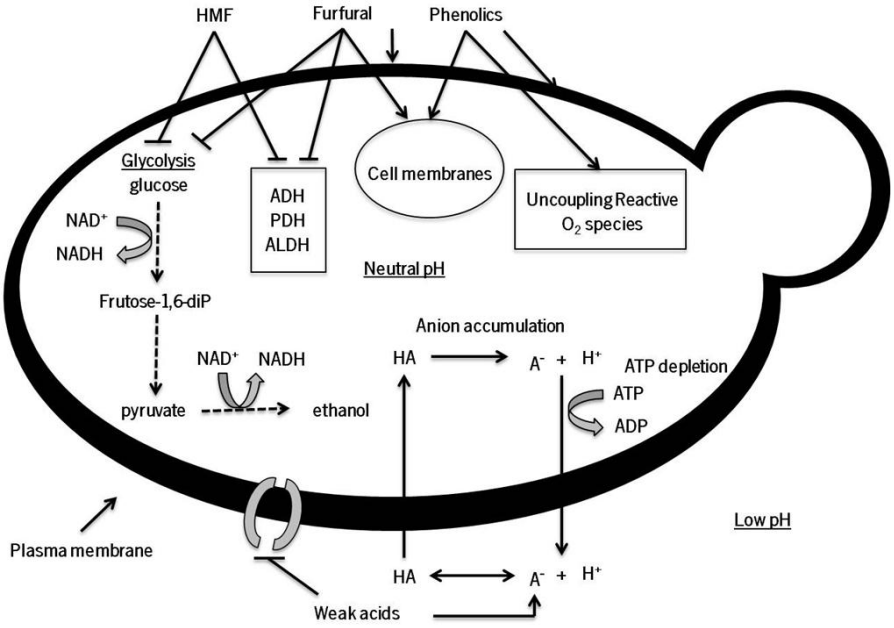


Figure 2.4 Schematic representation of inhibition mechanisms by furan-derivates, weak acids and phenolics (Adapted from Almeida et al (2007) [87]).

Bio-butanol production

Bio-butanol produced from renewable biomass is a promising alternative to ethanol as bio-fuel [104]. It is less volatile, less hygroscopic and less corrosive than ethanol, presents higher energy content, and it is can be used as an additive to gasoline or for replacing it completely without any modification of cars engines [105, 106]. It is also produced by substrates from the first order substrates as starches (corn, wheat, potato) or sugars (sugar beets, sugar cane) and from the second order, as lignocellulosic biomass, such as crop residues, woody crops or energy grasses.

Both native and engineered microorganisms have been used for butanol production [48, 105, 107-111]. 1-Butanol is naturally produced by some *Clostridium* species (*e.g. C. acetobutylicum* and *C. beijerinckii)* mixed with acetone and ethanol [112]. Although other microorganisms with easier genetic manipulation are becoming more competitive than *Escherichia coli* and *S. cerevisiae* [111], *S. cerevisiae* was already engineered either by redirecting amino-acid biosynthetic pathways or by introducing the 1-butanol pathway of *C. acetobutylicum* for 1-butanol and isobutanol production [113, 114].

The production of bio-butanol is generally performed by using the acetone/butanol/ethanol (ABE) fermentation pathway, which consists in two distinct phases: the acidogenic and solventogenic phases (Figure 2.5). The first phase, the acidogenesis, is coupled with cells growth and production of acetic acid, butyric acid and CO_2. The second one, the solventogenesis, starts with the medium acidification and is characterized by the initiation of sporulation and metabolic switch, when usually part of formed acids together with sugar carbon source are metabolized to acetone, ethanol and 1-butanol [115].

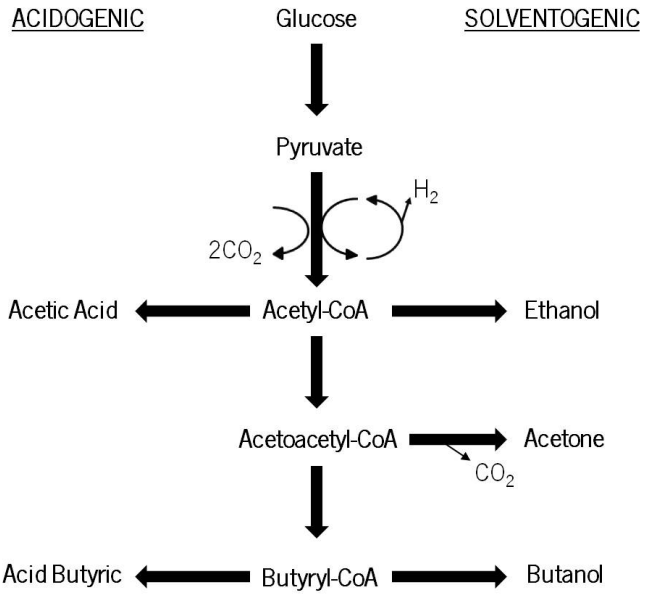


Figure 2.5 Acetone/butanol/ethanol (ABE) mechanism for bio-butanol production.

In addition to pathway engineering challenge, the end product toxicity represents the major limiting factor in developing an effective butanol production process [106, 116, 117]. The toxicity of 1-butanol isomer to cells is higher - inhibitory concentrations above 1.5 – 2% (v/v) - comparing to iso-butanol or 2-butanol, probably because of its higher hydrophobicity which leads to the strongest ability to permeate and/or interact with the cellular membrane [118, 119]. Cells are widely vulnerable to the presence of n-butanol, and different changes can be triggered by its presence, namely, the loss of the cells ability to maintain the internal pH due to the increased proton permeability of the cytoplasmatic membrane, the inhibition of the membrane adenylpyrophosphatase (ATPase) [120, 121], the loss of intracellular molecules, as proteins,

ribonucleic acid (RNA) and ATP (as the fluidity of the membrane increases) [110] and finally the obstruction of glucose uptake [120].

In bio-butanol production industry, product recovery systems are implemented in order to reduce the butanol levels in the medium, and thus to reduce the toxicity to the cells, although, these systems require a significant capital and operating costs to the process [122, 123]. The possibility of operating under increased concentrations of n-butanol [116, 117], by using more tolerant strains would greatly improve process economics and is still a challenge in bio-butanol production.

2.3.3 Flocculation in biotechnological processes

The main goal of biotechnological processes is the increase of productivity. Several strategies have been developed for this purpose, namely, the modification of yeast strains [124, 125], the development and optimization of operational strategies and bioreactors, the improvement of separation processes efficiencies, the utilization of efficient control systems and the use of techniques of cells immobilization, adhesion or flocculation [126, 127]. In industrial processes, the use of flocculating yeasts or strategies that induce flocculation of cells provide several advantages, including, the increased fermentation rates as a result of high cells densities per unit bioreactor volume, the possibility of reusing cells for extended periods of time, the simple separation of cells from the liquid phase, the decreased risk of contamination and the smaller bioreactor volumes that results in reduced capital costs for the process [128].

Yeasts flocculation can occur by three different mechanisms, including, sexual aggregation, chain formation [129] and non-sexual cells aggregation. According to sexual aggregation, complementary α and *a* haploid strains exchange small peptide mating pheromones, α and *a* – factors, responsible for physiological changes that promotes protein-protein bonding of complementary cell walls, enhancing cells aggregation before nucleus fusion to form diploids [126]. The formation of cells chain occurs during yeast cells growth when bud cell fails to breakup from the mother cell, and as both mother and daughter cells continue to form new buds, it results in a chain formation [130]. In non-sexual aggregation, it is known that interactions in flocculation may be mediated by specific recognition mechanisms of cells surface, relating cell surface proteins, called “adhesins” or “flocculins” binding specific amino acids or sugar residues on the surface of adjacent cells [127].

2.3.4 Intracellular metabolites

Some oxidative stresses, such as high ethanol concentrations may interact with the yeast cells membranes changing the membrane polarity and weakening the hydrophobic barrier to the free exchange [131]. It has been reported that trehalose is considered one of the main reserve carbohydrates in yeast and act as yeast protector, maintaining structural integrity of the cytoplasm under stress conditions [132]. The intracellular concentration of trehalose, non-reducing disaccharide consisting of two glucose molecules, has been suggested to play an important role in the ability of many organisms to tolerate adverse environmental conditions [133].

On the other hand, glycerol is known to serve at least two important functions in yeast: (i) as a sink for the excess NADH which is produced by anabolic reactions during anaerobic conditions [134], and (ii) as an osmolyte balancing a high external osmotic pressure during salt or sugar stress due to low water activity [135]. *S. cerevisiae* intracellular glycerol increases with the decrease of the water potential in the medium, playing an important role in the osmoregulation of yeast cells. Recent advances on molecular biology and genetics, showed that glycerol is essential to balance the osmotic stress of the yeast membranes [136].

2.4 SENSORS AND INSTRUMENTAL METHODS

Metabolic changes occurring in "*in vivo*" biological processes can be perceived by metabolites measurements using different analytical sensors and instrumental methods. Some examples include cyclic voltammetry (CV), high performance – liquid chromatography (HPLC), gas chromatography – mass spectrometry (GC-MS), and gas chromatography – flame ionization detector (GC-FID) which have been used in combination with multivariate tools in metabolomics studies [11, 12, 36, 137].

The selective and non-destructive detection of biologically relevant metabolites can be performed using potential control in CV. The ability to selectively detect and electrochemically resolve analytes in a dynamic system by CV results in the large ability for its application in metabolomics studies, because of the higher chemical complexity of samples involved [138]. Also, the combination of CV spectral information with other analytical techniques, such as HPLC and gas chromatography (GC), makes CV a useful tool in metabolomics as it can conduct to rapid metabolic profiling analyses [137].

Liquid and gas chromatography systems are two of the mostly used analytical methods for metabolomics [138], allowing the quantification of important metabolites involved in the dynamic system. These systems are characterized for having a stationary phase chemically bound to the surface and fixed in a column (behaving as a liquid), and a mobile phase which can be a gas (gas-liquid chromatography) or a liquid (liquid-liquid chromatography) [139].

2.4.1 Samples preparation for metabolomics analysis

Practical metabolomics approaches applies efficient sample preparation methodologies, combined with selective extractions coupled with different analytical techniques, in order to gather the maximum information about the metabolome [37, 140]. The preparation of samples for metabolome analysis is of extreme importance and follows a sequence of practical steps, as represented in Figure 2.6, in order to obtain the accurate response of the biological system to genetic or environmental changes [139].

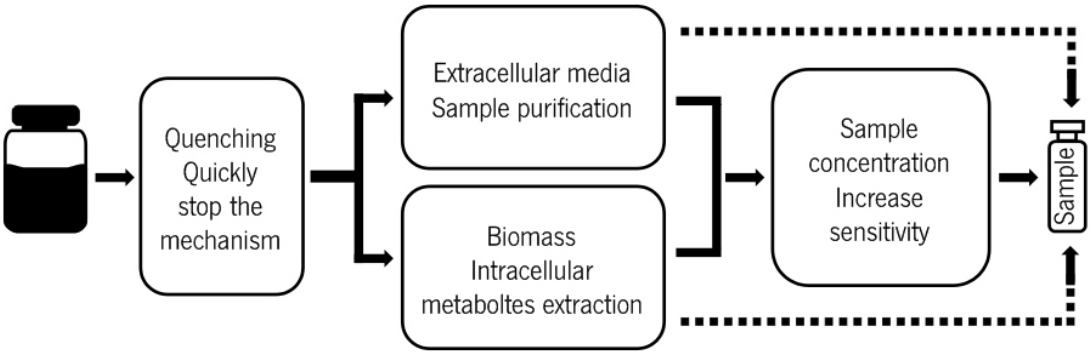


Figure 2.6 General methodologies used for samples preparation for metabolomics approaches (Adapted from Villas-Bôas et al. (2006) [141]).

The initial step which must be taken in consideration to preserve the integrity of metabolites, is the sample collection [142]. As metabolites concentrations inside the biological systems verify quick changes according to the response to the environmental conditions [143]– the typical half-time for an intracellular metabolite is on the order of a second or less, being longer for an extracellular metabolite [144] - a rapid quenching of all biochemical processes is firstly necessary [12, 145]. Villas-Bôas *et al.* [139] summarized some of the quenching methods, including the use of cold methanol [144], liquid nitrogen [146], perchloric acid [147], acid/alkali [148], and their applications, advantages and disadvantages. After the quenching, biomass and extracellular medium can be separated for intra and extracellular metabolome analysis, usually by centrifugation at low temperatures or filtration [149]. For intracellular analysis, metabolites must be extracted from cells fraction, with the minimal losses and preventing any further physical and chemical alterations of the molecules, making them accessible to many different analytical techniques [143, 150].

Intra- and extracellular metabolites are divided into three main classes, namely, polar or water soluble compounds, non-polar or water insoluble compounds and volatile compounds. The extraction of each of these groups can be performed using different methods, according to the nature of the sample [37]. These include solvent extractions - exhaustive extraction with organic solvents [151], the use of a mixture cold methanol: chloroform: buffer [144], boiling ethanol [152], cold methanol [153] and dichloromethane [154] - and mechanical extraction - pressurized liquid extraction [155], microwave and sonic wave [156] and supercritical fluid extraction [157, 158].

After extraction, intracellular and extracellular samples can be directly used for analytical measurement or in some cases, they must be concentrated (extraction methods generally produce high diluted samples) before using them [139].

Samples concentration consists in both removing the excess of solvents used in metabolites extraction from intra- or extracellular fraction or removing water from aqueous samples to avoid any thermal degradation and to enable samples injection in analytical techniques [143]. Different strategies have been also developed for samples concentration for metabolomics analysis, namely, solvent evaporation under vaccum [159], evaporation under anhydrous sodium sulphate stream [154], freeze-drying or liophilization [160] and solid-phase extraction (SPE) and/or solid-phase micro-extraction (SPME) [161].

Complementary to the importance of the selection of efficient sample preparation and separation techniques, gathering the complete information about metabolome within the biological system entails other relevant challenges that must be considered in metabolomics: the complexity and diversity of biological samples, the chemical diversity of small molecules, the large concentration dynamic range (as high as 10^{14}), the incomplete chemical information for identification of all metabolites, the need of reliable bioinformatics tools for non-targeted strategies, and the use of statistical tools capable of extracting the crucial information [11, 31, 143, 162].

2.4.2 Cyclic voltammetry

Voltammetry is one of the techniques which electrochemists employ to investigate electrolysis mechanisms and different voltammetry forms can be found, namely potential step, linear sweep and cyclic voltammetry [163]. For each of these cases a voltage or series of voltages are applied to the electrode and the corresponding current that flows is monitored. The essential elements for measuring the electrochemical signals are: the *electrode*, which is normally made of an inert metal (gold or platinum); the *solvent*, that has high dielectric constant (*e.g.* water or acetonitrile), which allows the electrolyte to dissolve, aiding the passage of current; a *background electrolyte*, a high concentrated inert salt (*e.g.* sodium chloride (NaCl) or tetrabutylammonium (TBAP)) – 0.1 M – which allows the current to pass; and finally the *reactant*, typically in low concentration 10^{-3} M [164].

Cyclic voltammetry (CV) is a very versatile electrochemical technique which allows exploring the mechanics of redox and transport properties of a system in solution [13]. Its versatility combined with ease of measurement has resulted in extensive use of CV in different fields, including electrochemistry, inorganic chemistry, organic chemistry, and biochemistry [165].

In CV, the voltage is swept between two values (as it is possible to observe in Figure 2.7 A) at a fixed rate, and when the voltage reaches V_2, the scan is reversed and the voltage is swept back to V_1. A typical cyclic voltammogram is recorded for a reversible single electrode transfer reaction, and followed by the appearance of the current at a certain potential, as it is shown in Figure 2.7 B. This voltammogram leads to the electrochemical oxidation of a given compound within an inert carbon glassy electrode [166].

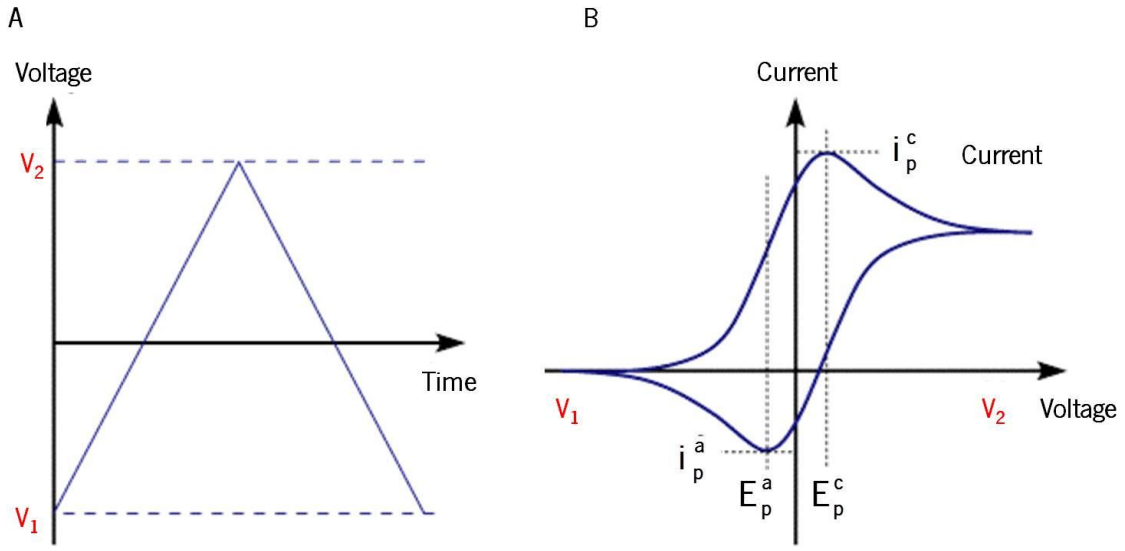


Figure 2.7 Voltage as function of time (A) and current as function of voltage (B) for CV (Adapted from Kilmartin et al (2001) [166]).

The potential of the CV peak reflects the redox properties of the compound, while the value of the current shows the quantity of the compound [164]. For a reversible electrochemical reaction, the CV recorded follows well defined characteristics: the voltage separation between the current peaks is given by Equation 2.1, corresponds to the electric potential in the minimum and maximum current intensity, respectively; peaks voltage positions do not change as function of the voltage scan rate; the ratio of peaks currents intensities is equal to one (Equation 2.2); peak currents are proportional to the square root of the scan rate.

$$\Delta \mathbf{E} = \mathbf{E_p^a} - \mathbf{E_p^c}$$

Equation 2.1

$$\left| \frac{i_p^a}{i_p^c} \right| = \mathbf{1}$$

Equation 2.2

CV can be used for evaluating the antioxidant capacity of a certain sample. The resulting voltammogram provides the information describing the integrated antoxidant capacity of the mixture or phenolics groups present there (*e.g.* catechins, gallates, monophenolics), although it does not allow the identification of antioxidants in complex mixtures [13, 166].

Principal component analysis (PCA) has been applied to CV studies, related to wines oxidation resistance, in order to characterize qualitatively the oxidation status of those wines [14, 167, 168]. Wines and other beverages are chemically complex mixtures where changes in redox potential take place, and their resistance to oxidation depends on the antioxidants quantity and their redox potential [167].

The chemical structure of phenolics acids comprises a hydroxyl group combined with a benzene ring, although the presence of other group and their position in the ring largely affect their electrochemical activity [166, 169], which allows their distinction throughout the voltammogram. Phenolic compounds with the *ortho* and *para* diphenol group show inferior oxidation potentials than those containing the *meta* diphenol group [166].

2.4.3 Chromatography techniques for metabolites separation

In metabolomics field, the separation of small molecules is nearly always based on high-performance chromatography using either a gas or a liquid as mobile phase [162]. The general operating system of the chromatography consists in different modules, namely, the injector, the stationary phase (column) and the detector. The separation process starts by laying a small sample in the mobile phase (using the injection port) at the beginning of the stationary phase (column). Then according to the distribution coefficient of each metabolite, these are separated (with small

differences) through the two-phase system (liquid-liquid or gas-liquid). A plot of intensity *versus* time is the resulting chromatogram as represented in Figure 2.8, where the peaks presented corresponds to the eluted compounds. Each peak is characterized by a retention time (time from the injection to its elution), peak width, peak height (maximal signal) and peak area, which is determined as the area under the curve in Figure 2.8 from the beginning until the end of the peak [170]. The accurate determination of the area under the curve requires a stabilized baseline, which in some cases needs to be adjusted.

Different factors can cause peaks *dispersion* during the chromatographic separation, mainly dependent on the column geometry (which determines the band broadening during metabolites elution), on the flow rate (lower flow rate enhances the dispersion effect), on the stationary and mobile phases used, and on the equilibrium between both phases [170, 171].

Liquid and gas chromatography are equivalent analytical systems and present similar chromatographic profiles, where separation efficiency can be affected by dispersion effects. They can differ in their components, namely, in the mobile phase used, the stationary phase and the detectors used for metabolites identification and quantification [162].

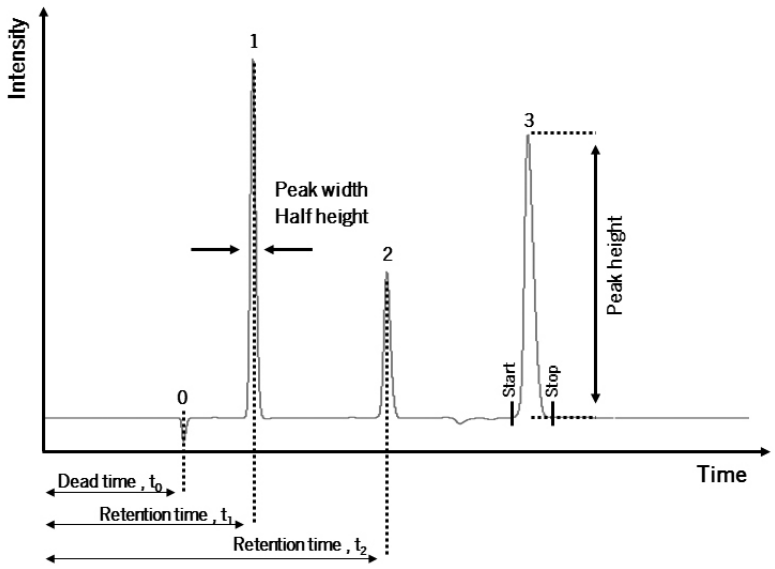


Figure 2.8 General chromatogram profile (Adapted from Villas-Bôas *et al.* (2007) [162]).

Liquid chromatography

Liquid chromatography (LC) is a quite simple methodology to operate and uses a liquid mobile phase to separate metabolites within a column by a pumping system. LC is considered a powerful technique for determining highly polar compounds in several matrices [172], when compared to gas chromatography. Different mobile phases, columns and stationary phases can be used in liquid chromatography, promoting the capacity of dissolving nearly all types of compounds in a mobile phase.

The mobile phase (one solvent or solvent gradients), free from air bubbles, is supplied to the column using high-performance pumps, with a constant and pulse-free flow, using the rate required for metabolites separation. Samples are injected into the solvent stream using the injector to be separated within the column, where the stationary phase is chemically bounded. Therefore, the eluent arising from the columns can pass through a flow cell in a spectrometer for nondestructive detection of compounds with spectrometric features, *e.g.*, a chromophore or a fluorophore. UV and fluorescence spectrometers have a good performance for quantitative analysis and are generally used in high performance liquid chromatography (HPLC).

Other commonly used liquid chromatographic technique is ultra high performance liquid chromatography (U-HPLC), which uses high pressure level instruments (until 15,000 psi) and provides higher chromatographic resolution [173] and peak capacity [174] than HPLC. Also, it improves the separation performance in terms of velocity and high-throughput analysis [175]. Furthermore, the use of hydrophilic interaction chromatography (HILIC) technique or ion-exchange chromatography also improves the detection and separation of many polar and ionic compounds, increasing the metabolic information [36, 176].

Gas-Chromatography

Gas-Chromatography (GC) is a simple and proficient analytical tool, where up to thousand of metabolites can be separated within an hour [170]. The main elements that constitute a gas-chromatography system include a gas supply, an injector, a column inside the oven and finally a detector (or mass spectrometer). GC analysis can be done using constant flow, constant pressure, or a flow program.

The mobile phase of a gas-chromatographer is a gas, generally highly pure helium, delivered from a compressed gas supply. The injection of samples, containing more or less volatile compounds, is the most critical practical step in gas chromatograph. This is due mainly because of the possible occurrence of a slow and incomplete evaporation of the volatile metabolites, and the time needed for transferring the samples through the mobile phase into the column which must be insignificant for the peak width [177].

The stationary phase of these systems is bound to the inner surface of the column - a long opened squeeze bore fused tube - placed in an oven with controlled temperature. The selection of stationary phases depends on the metabolic information aimed to extract: volatile compounds with low retention time typically need a thick stationary phase, while less-volatile compounds require a thin-film column eluting at high temperature. Still, the higher the column length, the higher the retention time is, and thus, the metabolite spends much time in the stationary phase. Although, the longer is this time, the wider the peaks get, because of band broadening effects. The temperature of the oven where the column is placed also influences the distribution coefficient and thus, metabolites separation in GC analysis and temperature variations are defined within the analysis [177].

Gas chromatography is used for the analysis of several compounds, namely trace amounts of organically extractable, non-polar, volatile and highly volatile compounds [178]. Non-volatile metabolites as organic acids, amino acids and sugars, can be also analyzed by gas chromatography however this analysis requires a previous chemical derivatization to induce volatility and enhance thermal stability for further analysis [162].

After metabolites separation within gas chromatography, different detectors can be associated, namely flame ionization detector (FID) in GC-FID and mass spectrometer (MS) in GC-MS. From the industrial point of view, GC-FID is much more attractive, as this is a rapid, high sensitivity and cost-effective methodology for a large number of routine samples analysis [179], and provides the report of carbon-containing compounds. It can be more useful than GC-MS in some cases, namely classification or prediction based on the quantification of specific metabolites, although, peaks identification is not possible using GC-FID [180]. For these reasons, GC-FID hasbeen used in few researches in metabolomics field [181, 182]. On the other hand, GC-MS have been widely used in metabolomic studies, providing a comprehensive quantitative and qualitative analysis of the

metabolome and is characterized by having high sensitivity, reproducibility and robustness [32, 141].

2.4.4 Mass spectrometry

The analysis of the chemical complexity of the metabolome requires the use of multiple technologies to gather a comprehensive visualization of the system [30, 145, 183-185]. The selection of the most suitable methodology is usually a balance between speed, chemical selectivity, and instrumental sensitivity [186].

The initial works in metabolomics were based on nuclear magnetic resonance (NMR) studies, namely on biofluids, cells and tissues [187, 188]. NMR are rapid, highly selective, and non-destructive, however, have a relatively lower sensitivity [16]. Other techniques including capillary electrophoresis (CE) combined with laser – induced fluorescence detection, are highly sensitive, however the chemical selectivity of these systems is limited [189, 190]. Because of their comparatively poor sensitivity cannot be used in lower concentrations metabolites analysis as MS [143]. Subsequently the general concept of metabolomics was extended on the usage of high-resolution separation systems, coupled with mass spectrometer (MS) detector: GC-MS [30, 139, 191], Liquid chromatography – mass spectrometry (LC-MS) [12, 192-194], hydrophilic interaction chromatography - mass spectrometry/mass spectrometry (HILIC-MS/MS) [194], high performance liquid chromatography – mass spectrometry (HPLC-MS) [195], ultra high performance liquid chromatography – mass spectrometry (U-HPLC-MS) [196, 197] and CE [198], which provide an increase of the metabolic information.

MS measurements coupled with chromatographic separation provides an increased sensibility and sensitivity for a rapid qualitative and quantitative analysis, allowing the identification of metabolites [145]. MS has also some potential for determining proteins sequence quickly and efficiently, using low samples quantities in proteomics [143], although it is most used for metabolomics, as it is still more suitable for reproducing phenotype changes of the living system [12, 36, 162, 193].

MS requires the ability of electron impact ionization and fragmentation for detecting any compounds in a specific mass range, typically m/z 10 – 1,000. As a result, compounds that do not form ions within this range will not be detected [179]. So, mass-selective detection of mass

spectrometers operate by ion formation, separation of ions according to their mass-to-charge (m/z) ratio and detection of separated ions [145], providing highly specific chemical information leading to the chemical structure of the metabolites [186]. The information about the molecular mass and/or characteristic fragment-ion information is then used for compounds identification through the spectral matching with data contained in libraries or for *de novo* structural revelation [36]. The limits of detection for many primary and secondary metabolites identification using MS chemical selective systems are very low (µmole to pmole) and small quantities of metabolites are needed to extract the chemical information [186].

Mass spectrometric analyses are performed in the mass spectrometers, which are made up of three main components: an ion source, a mass analyser, and the detector, as it is presented in

Figure 2.9 [162]. Samples contained in gas or liquid mobile phases are supplied to the ionization source, where these are transferred to the gas phase, ionized as positively or negatively charged ions and then transferred to vacuum. These ions are carried to the mass analyser where the separation based on the mass-to-charge ratio (m/z) occurs based on a combination of electric and/or magnetic fields. Once separated, ions enter into the detector where the mass-to-charge ratio and the relative abundance of each ion is determined, and the information is compiled within a spectrum graph [199].

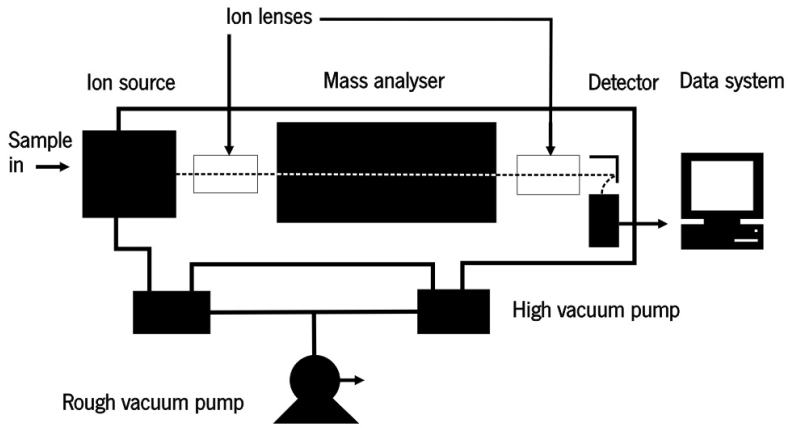


Figure 2.9 Mass spectrometer components for metabolites detection (adapted from Villas-Bôas *et al.* (2007) [162].

Data systems of mass spectrometers are nowadays designed for instruments controlling and also for data processing. In non target approaches, chromatogram processing is performed separately [11, 12].

Ion source

The classical procedure of ionization involves shooting energetic electrons into a gaseous neutral. The ionization sources are typically classified according to the resulting degree of molecules fragmentation. Hard ionization as electron impact ionization (EI) provides considerable and highly reproducible fragmentation patterns for small molecules [200]. Soft ionization techniques mostly used as interfaces to HPLC and MS include chemical ionization (CI), matrix assisted laser desorption / ionization (MALDI), atmospheric pressure chemical ionization (APCI) and atmospheric pressure photo ionization (APPI), which produce larger fragments [200, 201]. The main advantages and disadvantages of these detectors are presented in Table 2.1

The most frequent ionization techniques are EI and electrospray ionization (ESI) commonly used in gas and liquid chromatography, respectively. As ionization in GC-MS occurs under vacuum conditions and in LC-MS at the atmospheric pressure, they require completely different interfaces to mass spectrometer [201]. In GC-columns the EI occurs by impact of a beam of energetic electrons emitted from a heated filament and accelerated (70 eV), under high vacuum ($< 5\times10^{-5}$ hPa). Ions formed by electron impact are dragged out by an electrical acceleration potential [139]. In LC-MS, during the ESI, the eluent is pumped from the LC-column through a narrow steel capillary tube into an open source chamber at atmospheric pressure. Therefore a high voltage is applied to the capillary tube, inducing the emission of a spray (electrospray) of fine highly charged droplet. This electrospray ionizes the molecules in positive or negative ions, depending on the potentials applied on the sprayer and the related counter electrode. Finally ions are driven through a small orifice of heated capillary to the vacuum [139].

Table 2.1 General comparison of ion sources (Adapted from Villas-Bôas et al. (2007) [162])

Ion source	Advantages	Disadvantages
EI	- High performance; - Ease to operate; - High reproducibility;	- Not suitable for non-volatile analytes analysis; - Influence of pyrolisis of analytes in EI spectrum;
ESI	- Detection of non-selective and most ionisable ions; - Well-suitable for polar metabolites detection; - Fast; - High accuracy;	- Matrix effects: loss of sensitivity and discrimination of minor metabolites; - Fine tuning work: flow rate, solvent/sample ratios, etc to get the analytes to ionize
MALDI	- Fast; - Good reproducibility; - Sensitive to small amounts of sample; - Easy spectra; - Accurate; - Not affected by salts; - Soft ionization;	- Fine tuning: spotting plate, getting good crystals, adjusting intensity of laser, finding crystals on plate with sample; - Low shot to shot reproducibility; - Possible occurrence of photo degradation by laser desorption/ionization; - Acidic matrix used in the ionization can cause compounds degradation;
CI	- Less energy transferred causes less fragmentation; - Well suited to the negative ionization mode;	- Spectrum analysis can be influenced by interactions between the reagent carrier gas and the sample; - Higher source pressures required ca difficult high resolution tuning of the ion source;
APPI	- Efficient ionization of non-polar compounds; - Relatively low ionization energy leads to minimal fragmentation;	- Higher signal-to-noise ratios leasing to lower background ionization;
APCI	- Handles high flow rates; - Insensitive to salts;	- Generation of background ions from solvents; - Requires high vaporization temperatures that cause thermal degradation;

Mass analyzers

The determination of the mass-to charge ration of the ions is performed through mass analyser, under vacuum, in order to guarantee that ions do not collide with uncharged molecules, as air or with each other [139]. Mass analysers are divided into two main groups, according to their performance: *nominal mass analysers* (mass resolution: around 1:1000 – 2000; mass accuracy: integer) and *high resolution mass analysers* (mass resolution: 1:7000 - 100,000; mass accuracy: below 1 ppm) [202], which are summarized in Table 2.2.

Ions separated in mass analyzer can be analysed by a detection system that measures the ion current (flux) constantly as a function of the scan in progress and quantifies ions arriving in small

time segments, called bins. The measurement of each ion current is performed by intermediate detectors which concerns either a conversion dynode and electron multipliers in quadrupole and ion-trap analyzers or a micro-channel plate (MCP) detector coupled to time-to-digital converter (TDC) in time-of-flight (TOF) instruments [162].

The resulting mass spectrum displays masses of the ionized molecule and its fragments, which correspond to the masses of the component atoms. This mass spectrum provides enhanced analytical and structural information of the system and there are several available software applications that make the interpretation of MS data relatively easy [139].

Table 2.2 General comparison of mass analyzers (Adapted from Siuzdak (1996) [202] and Villas-Bôas et al. (2007) [162])

Mass analyzer	Typical mass range and resolution	Advantages	Disadvantages
Ion trap	Range m/z: 2000 Resolution: 1500	- Small size; - Medium resolution; - Simple design, low cost; - Well-suited for tandem mass spectrometry; - Easy for positive/negative ions;	- Limited mass range; - Subject to space charge effects and ion molecule reactions; - Artifacts such as harmonics and sidebands are present in the mass spectra;
Quadrupole	Range m/z: 3000 Resolution: 2000	-Tolerant of high pressures; -Well-suited for electrospray; -Ease of switching between positive/negative ions; - Small size; - Relatively low cost; - High sensitivity and reproducibility;	- Mass range limited to about 3000 m/z; - Poor adaptability to MALDI; - Peak heights variable as a function of mass;
TOF	Range m/z: ∞ Resolution: 350	- Highest mass range; - Very fast scan speed; - Simple design, low cost; - Ease adaptation to MALDI;	- Low resolution; - Difficulty of adaptation to electrospray; - Fast digitizers used in TOF can have limited dynamic range; - Limited precursor-ion selectivity for most MS/MS experiments;
Fourier transform – mass spectrometry (FT-MS)	Range m/z: 10,000 Resolution: 30,000	- High resolution; - Well-suited for tandem mass spectrometry;	- High vacuum required; - Expensive; - Instrument massive;

Direct injection mass spectrometry (DIMS)

Separation of complex mixtures is usually required prior to MS analysis, however, direct injection of samples is another approach [200]. Direct injection mass spectrometry (DIMS) consists in injecting directly samples metabolites in front of the ion source of a mass spectrometer without prior chromatographic or electrophoretic separation [203]. This is a high-throughput methodology which allows to process samples within a few minutes. The short analysis time increases inter-sample reproducibility and improves the accuracy of subsequent cluster analysis [200]. DIMS analyses have been performed using electrospray ionization – mass spectrometry (ESI-MS) [204, 205].

The major DIMS disadvantage is the purported 'matrix effect' [206], that can commit the sensitivity and accuracy of the quantification in different matrix composition samples. This approach can be used in non-target metabolomics when the matrix composition variation is low or does not affect the comparison between samples [203].

2.5 METABOLOMICS

Metabolome is defined as the full collection of low-molecular-weight metabolites present within a cell or biological system under a given set of physiological conditions, and is considered the endpoint of "*omics" analysis* [11, 31, 32, 207]. Changes in cells physiology, as a result of gene deletion or overexpression and/or environmental changes, can be measured throughout the metabolome, complementarily to transcriptome and proteome [11, 207]. In Figure 2.10 is presented the relationship between genomics, transcriptomics, proteomics, and metabolomics, called the "*omics"* cascade [30].

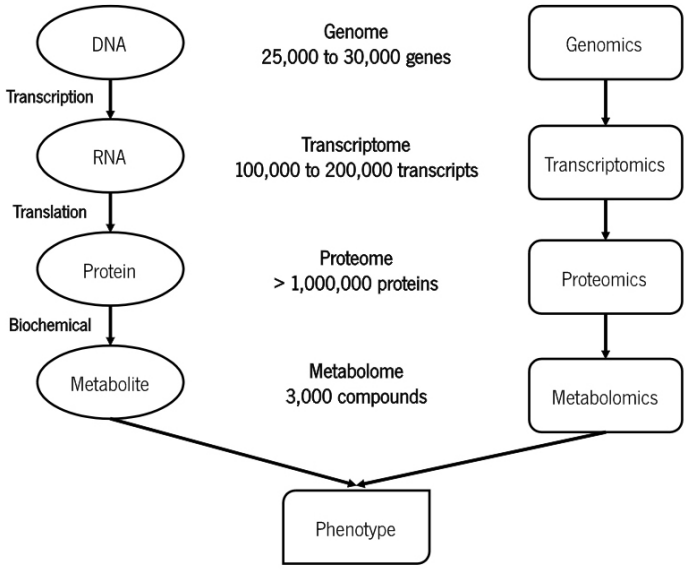


Figure 2.10 The 'omics' cascade summarizing the relationship between metabolomics to genomics, transcriptomics, proteomics and phenotyping in "*systems biology*".

In Figure 2.10, it is possible to observe that being complementary fields, genomics looks for the entire collection of genes which leads to the genome of an organism ("*What can happen...*"), transcriptomics looks at the entire transcriptome, that is the complete set of messenger RNA (mRNA) transcripts ("*What appears to be happening...*"), and finally proteomics looks at the proteome ("*What makes it happen...*") concerning to the expressed proteins that are encoded by the genome. In the same manner, metabolomics can be defined by the complete set of small

molecules involved in general metabolic reactions, synthesized by a cell, tissue or organism ("*What has happened and is happening...*") [143]. The combination of the 'omics' fields can provide a comprehensive overview of the response of biological systems to disease, genetic and environmental perturbations [30].

In spite of this complementarity, metabolome measurement has some advantages [36]. It is just one part of the system biology through it may be the most useful as it is the most direct observation of the status of the cellular physiology [143]. In a biological system, at a given time, the rate of an enzymatic reaction is a function of the available substrates, products and modifiers, as well as gene expression. Although changes occurring in the expression of proteins might have little influence on fluxes [208], the influence on the concentration of intermediary metabolites can be large [209]. The effect of the metabolic pathways activities is reflected more accurately in the concentration of metabolites (or pool of metabolites) comparing to the concentration of the specific enzymes (or mRNAS encoding them) [11]. Indeed, while the genome is what might be expressed and the proteome is what is actually expressed, the metabolome is what is done and represents the current status of a biological system [143].

Metabolomics is the comprehensive analysis of metabolome in which all the metabolites of an organism are identified and quantified [31, 32], and represents the evolution from large-scale analysis of RNA and proteins at the systems level [210]. It also studies the dynamic changes in the metabolome [143]. Different studies on metabolomics field revealed different analysis approaches, presented in Figure 2.11, which can be distinguished between two different methodologies, namely targeted approaches (target analysis and quantitative metabolite profiling) and non-targeted approaches (metabolic fingerprinting and metabolic footprinting) [30, 149, 211].

Target approaches concern to quantitative measurements of metabolites and, while in metabolic profiling sets of analytes of a common biochemical pathway or chemical group are quantified, in target analysis, specific metabolites (substrates or products) are quantified [30, 145, 150, 212-214]. The non-targeted approaches lead to non-quantitative analysis, where statistical tools are used for raw data processing in order to extract distinctive features in different samples types [11, 140, 195, 203, 215].

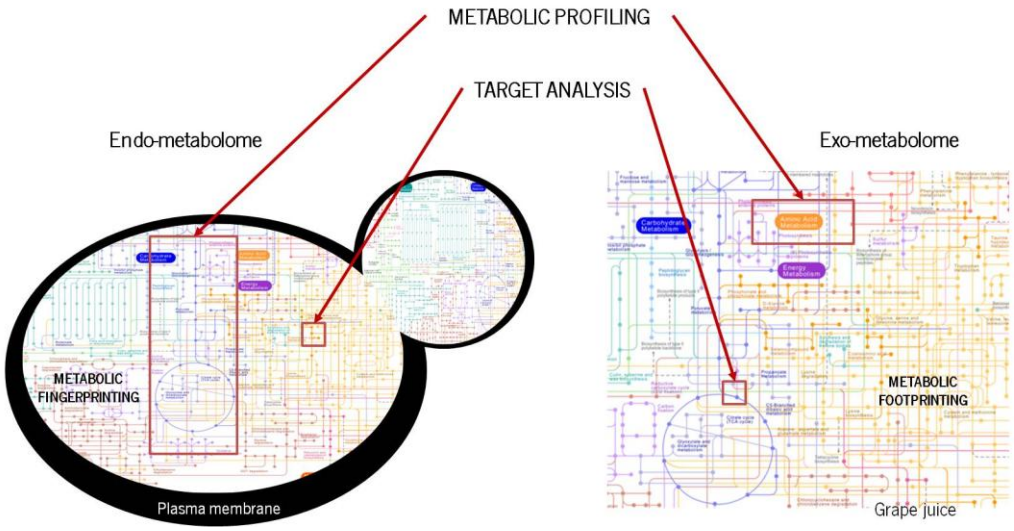


Figure 2.11 Contextualization of different investigations of metabolomics approaches (Adapted from Oldiges et al. 2007 [150]).

2.5.1 Target analysis

Target metabolomics includes target intracellular and/or extracellular metabolites analysis and metabolite profiling and concerns to the quantification of one or a reduced group of metabolites of interest in the dynamic metabolic systems, respectively [194, 203]. In target metabolomics, previous to metabolites measurement it is important to understand the context of their biochemical pathways [216]. Using the internal standards of the defined metabolites it is possible to quantify them in samples, by metabolites concentrations normalization across samples, being possible to have a comprehensive understanding of a wide range of metabolic enzymes, their kinetics and end products, as well as the biochemical pathways where are included [217]. The study of the predefined metabolites or group of metabolites can also highlight novel associations between them leading to specific physiological states [31].

MS - based metabolomics platform after the separation methodology (liquid or gas chromatography) for target metabolomics analysis, allows the exact identification of each molecule which can be fully defined and incorporated into the analysis. Also, prior to the full analysis, it is crucial to define metabolites specific transitions, retention times and dynamic concentrations ranges [217]. In target metabolic analysis and metabolite profiling, where specific metabolites or group of metabolites are measured, the most used platforms include GC-MS [218-220], HPLC-

MS/MS [195, 221, 222], or capillary electrophoresis – mass spectrometry (CE-MS) [190, 198, 223].

2.5.2 Non-target metabolomics

Metabolomics analysis using analytical methodologies including chromatographic systems followed by MS or NMR can produce large amounts of analytical data [211, 224], and once optimized can provide highly quantitative data of many metabolites at the same time [40]. This is called high-throughput or non-target metabolomics analysis and concerns to the comprehensive analysis of all the measurable analytes in each sample, known or unknown. An unbiased and comprehensive high-throughput analysis of a wide range of chemical compounds [31] can provide a critical context necessary to obtain information about cells physiology and metabolism from metabolite data, allowing also the integration of transcriptomics, proteomics, and metabolomics data to achieve a complete systems biology approach [217].

Methodologies for samples preparation, the sensitivity and specificity of the analytical technique, protocols and time required to manage the large amounts of raw data generated, problems in the identification and characterization of unknown small molecules, the reliance on the analytical report of the platform in use and the trend of detection of highly abundant molecules, are the main constraints of non-target approaches [217].

Rapid improvements in MS – based methodologies and computer hardware and software have emerged for a high-throughput metabolomics assessment and large datasets handling [225]. In this context, it is of great importance to establish a robust data handling pipeline, combining different data processing technologies in order to provide an interpretation of the hundreds of chromatographic peaks and mass spectra produced as well as to make meaningful associations with different instruments [31].

Metabolic fingerprinting or endo-metabolome

Metabolic fingerprint (or endo-metabolome) concerns to the quantification of intracellular metabolites involved in the living system, and can be applied to a specific metabolite or to a small,

isolated group of intracellular metabolites [226]. The intracellular metabolism of yeasts is extremely fast - the turnover of an intracellular metabolite can be under 1 second even for metabolites at mM concentrations [11] - and thus the quantification of intracellular metabolites requires an efficient quenching of the cell metabolism followed by an effective separation of intra- and extra- cellular metabolites and subsequent extraction of intracellular compounds [139]. Since the intracellular metabolism is more dynamic than the extracellular, such might be a weakness of this analysis, as it is still not possible to fulfill the need for and accurate, simple and rapid timescale for the turnover of an intracellular metabolite [11]. Different methodologies have been developed for intracellular metabolites determination [11]. These include efficient samples preparation methods (referred in Section 2.4.1), sensitive and accurate methods for detection of metabolites, for improving data mining and data analysis [160].

Metabolic footprinting or exo-metabolome

In dynamics systems such as fermentations using microorganisms, extracellular metabolites include those excreted and secreted by cells into their living environment and also byproducts of the natural environment resources which are left and discharged into the medium [3]. Changes in extracellular metabolites lead to modifications of the fermentation medium as a result of the microorganisms' activities. Extracellular metabolites released by yeasts strains represent the metabolic footprinting (or exo-metabolome) of these microorganisms [11, 215]. Metabolic footprinting resulting from the metabolic activity of yeasts in the process, manipulate the cellular metabolic status, concerning to intracellular metabolism, leading to the production of other by-products in the system [32, 208]. According to Villas-Bôas *et al.* (2005), the metabolic footprinting analysis provides crucial information for fundamental and applied research, namely: different strains distinction, cells "*quorum sensing"* elucidation with metabolites identification and integrated with physiological and phenotypical observations, and metabolic charts of intracellular metabolic pathways optimization. In addition, it can be used for monitoring the physiological status of cells under conditions inducing cellular stresses or in the presence of biomarkers [160].

Metabolic footprinting strategy for high-throughput data analysis follows an ordered methodology which includes extracellular samples preparation, injection in analytical equipments (*e.g.* GC-MS or LC-MS), utilization of chromatographic signals processing methodologies to extract relevant

metabolomic information, and finally use of multivariate statistical tools in order to structure the information obtained, for further understanding of the system [11, 227].

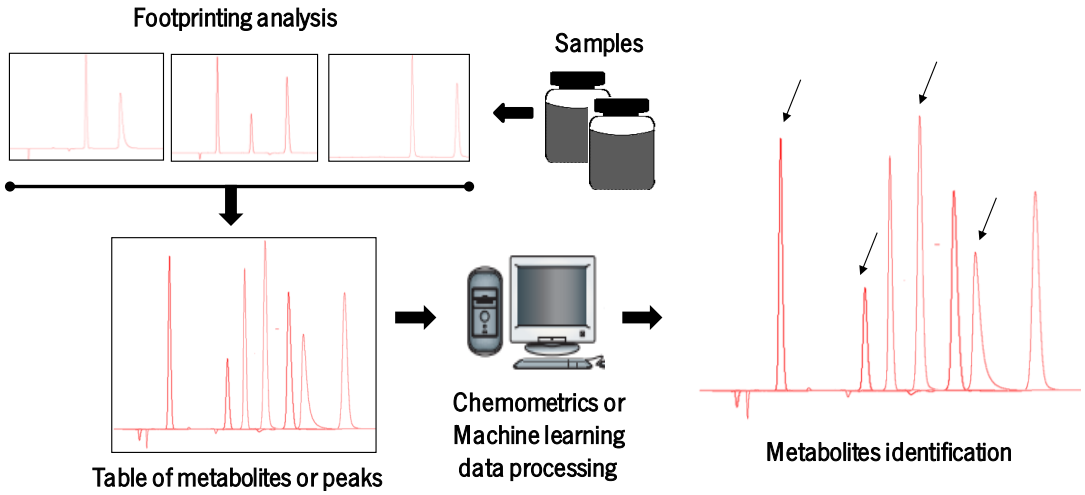


Figure 2.12 Non-target pipeline used in metabolites footprinting analysis (adapted from Kell et al. (2005) [11]).

2.6 TECHNOMETRICS TECHINQUES

Technometrics concerns to the development and use of statistical methods in physical, chemical and engineering science and also information sciences and technology. In all these fields, technometric allows building comparative metrics of product quality and competitiveness for a given product, process or service [228]. Complex processes are characterized by rich and often not recognized information in multivariate and multidimensional data. Technometrics provides the utilization and understanding of the richness of this information by using simple principle models, multidimensional spaces and projections. So, it leads to the use of statistics for the physical, chemical, and engineering sciences [229].

High-throughput non-target metabolomics leads to the study of the chemical processes involving metabolites [32, 145]. Technometric tools applied to metabolomics include bioinformatic and chemometric approaches that can be used for understanding the overall complexity, namely metabolic pathways changes and interactions, interesting metabolites measurements and/or "*de-novo"* metabolites discovery [29, 58, 225].

Bioinformatic tools can be applied to chromatographic signals, in order to deal with analytical experiments and the instrumental signals [29]. In this context, these algorithms provide the use of the maximum potential of those techniques in metabolites extraction and interpretation of the biological information [229, 230]. Also, bioinformatic tools development follows the continuous evolution of the instrumental techniques available, in order to obtain increased number of metabolites and higher peaks efficiencies, to correctly characterize the overall process.

On the other hand, data-driven chemometrics methods enable the analysis of complex multivariate data, whereby the extraction of the relevant information is facilitated. These are generally used to reduce the complexity or number of parameters from metabolomics analysis. Chemometricians use dimensional reduction to identify the key components that seem to contain the maximum amount of information, or that yield the greatest differences [231]

In the following Sections, bioinformatic and chemometric tools that can be used for high-throughput metabolomics analysis are explored.

2.6.1 Bioinformatic tools applied to chromatograms processing

In metabolomics pipeline, data preprocessing is a critical step and demands the reduction of analytical data complexity and the correct extraction of peaks or metabolites present in the raw-data for further data analysis using other algorithms [232-234]. Preprocessing includes the use of the most suitable algorithms for noise filtering and baseline correction (noise or signal discrimination), peaks detection within all samples, deisotoping (clustering the isotopic peaks corresponding to the same compound), alignment, identification and normalization [232].

Several commercial or free available software based on metabolomic raw data preprocessing, peak detection and/or quantification have been developed in the last years [232]. Free software tools for metabolomics data processing are summarized in Table 2.3. Commercial MS software, either provided by the manufacturers of the MS equipment or by independent vendors, generally puts emphasis on the graphical user interface (GUI) and the usability [235]. Some examples of commercial MS software include *MarkerLynk* (Waters Corporation, Milford, MA), *ChromaTOF* (Leco Corporation, St, Joseph, MI), *ChemStation* (Agilent Technologies, Santa Clara, CA), *AnalyzerPro* (SpectralWorks, Runcorn, UK), *ClearView* (Markes International, Rhondda Cynon Taff, UK) and *IonSignature* (Ion Signature Technology, N. Smithfield, RI) [12]. In the recent years, several freely available software packages have been developed for metabolomics data processing by the academic community. In Table 2.3 are presented some examples of freely available software characteristics and tools referred by each one. In this research work only *XCMS*, *MetAlign* ™ and the commercial software application *MarkerLynkx* were explored, and will be explained in more detail.

Raw data files

Raw data resulting from GC-MS or LC-MS analytical tools is the compilation of data points vectors gathered during small successive time slots. Each data point is characterized by m/z and intensity value, and is obtained in different data formats, leading to the analytical instrument. In order to get open each file format, different tools can be used, namely *mzXML, NetCDF* or *mzML* [236-238]. Files in *mzXML* format are generally larger than those in the compressed form, *NetCDF* (binary compact format) and *mzML* (inline zlib compressed format).

Table 2.3 Data preprocessing steps covered by free available tools

Software	Licence and language	Experimental Data	Filtering	Peak Detection	Deisoto-ping	Alignment	Gap filling	Visuali-zation	Identifi-cation	Normali-zation	Multivarite analysis	Metabolic Database connection	Reference
MZmine	*GNU GPL (Java)*	*GC-MS*	■	■	■	■	■	■	■	■	-	-	[239]
MZMine2	*GNU GPL (Java)*	*LC-MS; LC-MS/MS*	■	■	■	■	■	■	■	■	-	■ PubChem; KEGG; METLIN; HMDB	[240]
MetAlign™	*Free*	*LC/GC-MS*	■	■	-	■	-	-	-	-	-	-	[38]
OpenMS	*LGPL (C++ library)*	*LC-MS*	■	■	■	■	-	■	■	-	-	-	[241]
XCMS	*GNU GPL (R)*	*LC-MS*	■	■	-	■	■	■	■	-	-	-	[18]
XCMS2	*GNU GPL (R)*	*LC-MS*	■	■	-	■	■	■	■	-	■	■ (METLIN)	[242]
XCMS Online	*Free (Web based)*	*LC-MS*	■	■	-	■	■	■	■	-	■	■ (METLIN)	[243]
MathDAMP / TriDAMP	*Free (Mathematica)*	*- MS based*	■	■	-	■	-	■	-	-	-	-	[244]
Metabolite Detector	*GNU GPL (C++)*	*GC-MS*	■	■	■	■	■	■	■	-	-	■ NIST library	[245]
TargetSearch	*GPL (R, C)*	*GC-MS*	■	■	-	-	-	-	-	-	-	-	[246]
PyMS	*GNU GPL2 (Python)*	*GC-MS*	■	■	-	■	-	■	■	-	-	-	[235]
MetSign		*LC-MS / DI-MS*	■	■	■	■	-	■	■	■	■	■ MetSign; LIPIDMAPS; KEGG; HMDB	[247]
MetaboAnalyst 2.0	*Free (R)*	*GC-MS / LC-MS*	■	■	-	■	-	■	■	■	■	■ SMPDB; HMDB	[248]

Raw data filtering

Chromatographic signals contain, as any other instrumental signal, three main components, namely, signal, noise and background, which differ in their frequency. The frequency of signal is typically intermediate between noise (highest) and background (lowest) frequencies. Chromatogram enhancement can be obtained by eliminating the noise and background components [29]. Noise filtering or baseline correction provides peaks detection and greatly reduces the detection of false positive features [241]. Different smoothing filters for noise reduction can be applied, namely a peak-area-preserving *Gaussian* low-pass filter, *Savitzky-Golay* filter based on least squares smoothing and non-linear filtering [241, 249, 250].

Noise filtering and peaks detection algorithms are combined within the same algorithm, as *XCMS*, *XCMS*2 or *XCMS* Online [18] combine noise filtering and peaks detection algorithms and *MetAlign* [38] is capable of performing baseline correction, noise filtering, and saturation and mass-peak artifact filtering, using thresholds and time intervals to reduce noise.

Peaks detection and deconvolution

Peaks detection and deconvolution allows to identify and quantify peaks signals corresponding to molecules in the sample, and reduce the data complexity for further analysis [251]. Peaks detection algorithms should be capable of correctly identify true signals and circumvent the false positives. Deconvolution, on the other hand is used to enhance selectivity of a specific chromatographic technique, mainly when the separation conditions cannot be improved to improve peaks separation [29].

Matched filter [18], *centWave* and *centroidPicker* [233] are three different peak detection algorithms, that can be applied in *XCMS*. *Matched Filter* bins the data into slices of 0.1 m/z width and determines the signal in each slice by taking the maximum intensity at each retention time (RT). Then it applies the second-derivate *Gaussian* to filter each slice. Therefore, it uses a signal-to-noise ratio (S/N) cutoff of 10 where the noise is the mean of the unfiltered data. However, this algorithm does not separate pairs of co-eluting peaks that fall within half of m/z bin. In this context, *centWave* can be applied, which finds a region of interest based on the mass accuracy and expected chromatographic peak width. Chromatographic analysis is performed using

continuous wavelet transform, in order to detect peaks with different widths, being the peaks intensity determined by the maximum value of the *centroid* peak in the estimated peaks boundaries. Local baseline and noise are detected by the truncated media and standard deviation in an extended region of interest. Finally *centroidPicker* finds the local maximum of intensity in each spectrum, and for each maximum. If the next scan has the corresponding maximum within the specified m/z window, then connects it with the local maximum in the next scan. On the other hand, if the next scan does not have the corresponding local maximum, it checks the length of the currently connected maxima, and discards the chromatographic peak if the length is not more than the minimum length threshold. When the spectrum has a local maximum with no corresponding maximum in the previous scan, starts a new chromatographic peak.

Deisotoping

During chromatographic separation, a particular molecule may produce a pattern of peaks whose relative heights and m/z lead to the isotopic distribution of the elements composing corresponding ions [252]. In these cases, selecting each observed peak as an unique chemical specie, will result in many false positives identification. Clustering the isotopic peaks corresponding to the same compound (deisotoping) can help on the identification of samples metabolites, as it reduces the complexity of the data by removing the redundant information [232].

Alignment or warping

Chromatographic peaks alignment or warping concerns to the synchronization of the time axis, which is not a trivial problem [29]. This is performed to compare metabolites match features between samples and it should be performed prior to the statistical analysis. Different issues must be taken into consideration within finding matching peaks, namely: *i)* differences in the retention time across samples may be nonlinear; *ii)* a feature in a sample may have multiple possible matching features based on the m/z and retention time values (alignment algorithms should identify the correct match); *iii)* some peaks cannot be present in all samples [232].

In *XCMS*, the alignment algorithms perform the correction of retention times of all samples in a single step [18]. Samples are divided using overlapping bins of 0.25 m/z wide for matching peaks in the mass domain, and then it applies a kernel density estimation method [253] for estimating the RTs distribution of the matching peaks based on the m/z bind, and determines the RT interval based on the estimated distribution. Furthermore, *XCMS* can also use a group of "well-behaved" peaks as temporary standards (that can be also used in the iterative alignment) to determine the non-linear deviation in the RT for each sample and correct it.

MetAlign ™ alignment can be performed using two distinct algorithms, namely rough alignment and iterative alignment [38]. While in the first, peaks are aligned based on the amplitude throughout a dynamic time window defined by the user, in the iterative mode, peaks contained in all data sets are used as reference points and the difference in RT is calculated using a minimum number of reference masses with a certain amplitude in the time window. The difference in RT is calculated in relation to the first data set (the reference), and this difference is then used to correct the shift in the RT for the next alignment cycle. These iterations continue, by reducing the time range dimension, the smallest amount of reference masses and their amplitude, until the dynamic time window ranges the mass peak width.

Comparing the recall / sensitivity and the precision of some algorithms generally used for peaks alignment across samples, Lange *et al.* (2008) [254] concluded that *XCMS* perform better in metabolomic data sets (while *OpenMS* is better for proteomics data alignment).

Signal shifts and correlation optimized warping approaches have been also used for chromatograms warping and their increased efficiency improved their popularity and interest to use in chemometrics [255-257].

Gap filling

After chromatograms alignments some peaks can disappear, as during peaks detection, some of these cannot be detected. The non-detection can occur if peaks have low intensity, or have destroyed quality shape or correspond to an incorrect peak detection. The gap filling recovers the missing peaks from raw-data by statistics prediction of their position [29].

Identification

The identification task is usually more difficult in LC-MS metabolomics analysis comparing to GC-MS. For both analytical techniques, the identification of the peaks is generally based on a combination of the computation efficiency and the analytical data. Different databases have been developed to cooperate with the computational challenge. According to Forcisi *et al.* (2013) [193], the total number of metabolites reported in databases ranges from 1,000 up to 200,000.

In target metabolite profiling methodologies, databases include metabolites and metabolic pathway resources. Some examples are Kyoto Encyclopedia of Genes and Genomes (*KEGG)* [258-260], *MetaCyc* [261], *Lipidmaps* [262] and *BRENDA* [263] which are designed to improve the study of the metabolism and different metabolites transversely to other organisms. These also facilitate several practical applications in biology, such as comparative genomics and target genomics application [231].

In non-target metabolomics, as thousands of metabolites are rapidly characterized at the same time, which are therefore used to identify disease biomarkers or model large-scale metabolic processes [11, 12], databases should assemble more than metabolites and pathways names, the NMR and MS spectra, GC-MS retention indices, chemical structures and chemical concentrations [231]. Some emerging metabolomics databases, which fulfil these requirements, include the human metabolome database (*HMDB*) [264], *METLIN* [265], *BiGG* [266], National Institute of Standards and Technology (*NIST*) [267] and Spectral Data Base (SDBS) [268].

As different GC-MS or LC-MS instrumental setting can be used for metabolomics analysis, the establishment of an universal spectral library becomes very challenging [231]. In this context, efforts have been made to create a MS/MS library from different instruments [269], and the field of mass spectrometry is nowadays driven by high accuracy and mass resolution analytical equipment for annotating metabolites with an increased degree of confidence [33]

Normalization

Normalization is usually performed for data correction, refinement and reducing the overall variability for further mathematical analysis [270, 271].

Different factors can support the use of normalization algorithms during metabolomics analysis, including biological factors - different concentrations magnitudes of the measured metabolites,

distinct fold changes in concentrations of metabolites leading to the imposed variation, and metabolites concentrations variations under similar experimental conditions, - and others – technical deviations concerning to sampling, samples handling and analytical issues; and variance introduced by biological or technical variations [270].

Two different normalization methods can be applied in order to eliminate the undesirable systematic deviation of the measurements. One concerns to the use of internal standards, that is, metabolites representing a group of metabolites and chromatographic behavior. The other concerns to the use of scaling factors, as median and/or average of peaks intensities [232] and dispersion measuring [270], for each sample based on complete dataset. Some of the most used algorithms are auto-scaling [272], Pareto scaling [273], range scaling [274] or vast scaling [275].

Data visualization

Data visualization during metabolomics preprocessing is of great importance as it provides the visual pattern of each data handling algorithm, which allows a faster supervision of metabolites patterns and chromatograms quality [232].

2.6.2 Chemometric analysis

Chemometrics has been applied to organic and analytical chemistry [276], food research [58, 277] and environmental studies [278].

Similarly to other "*metrics*", chemometrics depends on the use of different mathematical models which demands knowledge of statistics, numerical analysis, operation analysis and applied mathematics [229]. The most common statistical tools used in chemometrics for dimensional reduction include the analysis of principal components (PCA), cluster analysis and computer graphics [279]. Predictive modeling (regression and classification) using partial least squares regression (PLSR) is also widely used in chemometrics. Once the chemical properties of data are understood, it is possible to modify the chemometric tools to better explore the chemical data [229].

Exploratory data analysis

Principal component regression (PCR) is usually a way of reducing the data redundancy and dimension. This issue usually occurs in situations when one is measuring m independent variables, which produces an m-dimensional description of the state of the system, with interrelated or exactly the same information [280].

One PCR methodology is PCA which is not a classification algorithm but a non-supervised clustering for important effects detection in data, by reducing the dimensionality of a dataset [231]. These effects can be detected and explored by samples position on the samples space (scores analysis), by variables correspondences inside each principal component (loadings analysis) and also by variance contribution (eigen values) [34]. PCA allows the identification of statistically significant loadings on each principal component providing a better interpretation on how the different variables affect the metabolomics data variability [281].

PCA is a very useful tool as it has the ability of handling collinearity among many variables, and its ability to compress information about many variables into few independent (uncorrelated) principal components [272]. Furthermore, it is particularly useful as it allows one to easily detect, visually or graphically, sample patterns or grouping [231]. Also, detailed descriptions of theoretical and computational aspects of PCA algorithms have been documented, which allows its implementation in different programming languages. In practice, this tool is already available in several user friendly statistical software packages that provide the interpretation of the resulting graphical interfaces.

PCA can be done by eigen values decomposition of a data covariance (or correlation) matrix, singular value decomposition (SVD) of a data matrix or nonlinear iterative partial least squares (NIPALS), usually after mean centering (an normalization or using Z-scores) the data matrix of each attribute [34]. Singular value decomposition algorithm in metabolomics data sets concerns to the decomposition of the initial matrix $\boldsymbol{x}(n,m)$, a scaled matrix where n corresponds to samples and m to variables, to $x_{corr.}$, which is done by (Equation 2.3), where $\hat{x}$ is the signal and $\epsilon(x)$ the estimated noise, respectively.

$$\boldsymbol{X_{corr} = \hat{x} + \epsilon(X)}$$

Equation 2.3

The decomposition of the initial matrix (X), presented in (Equation 2.4), is performed using SVD, which algorithm is presented in (Equation 2.5), where U are the scores, W^T the loadings, and S the singular values [34, 282-284], considering the conditions (Equation 2.6) and (Equation 2.7), where I is the identity matrix (some variations of the technique do not require T to have unit norms) [34].

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x_1^{(1)}} & \cdots & \boldsymbol{x_m^{(1)}} \\ \vdots & \ddots & \vdots \\ \boldsymbol{x_1^n} & \cdots & \boldsymbol{x_m^n} \end{bmatrix}, \boldsymbol{n}\ \text{samples}, \boldsymbol{m}\ \text{variables}$$

Equation 2.4

$$\boldsymbol{X = USW^T}$$

Equation 2.5

$$\boldsymbol{T = US}$$

Equation 2.6

$$\boldsymbol{T^T T = I}$$

Equation 2.7

In order to distinguish between the number of relevant decompositions, a randomization test is performed to the original data matrix (x), to determine the number of relevant singular values [285]. The number of independent singular values and decompositions that discriminate metabolic differences in chromatographic spectra is determined based on the singular values of randomized chromatograms spectrum with the original using equation (Equation 2.8), where $(US)_{relv}$ or T_{relv} and W_{relv}^T are the statistically relevant scores and loadings of x, respectively [281].

$$\boldsymbol{X = T_{relv} W_{relv}^T + \epsilon}$$

Equation 2.8

After signal decomposition into principal components, not all features in the chromatographic spectra fingerprint preserve the same quality. So, the reconstruction of these samples is statistically not possible, which means that features that are not compressed, cannot be analyzed in the scores plot [286]. The quality of features extraction can be assessed by the *Q-statistic* (square prediction error) of the relevant decompositions [287], which is given by Equation 2.9, where E is given by Equation 2.10:

$$\boldsymbol{Q} = \boldsymbol{E}\boldsymbol{E}^{\boldsymbol{T}}$$

Equation 2.9

$$\boldsymbol{E} = \boldsymbol{x} - \boldsymbol{b}\boldsymbol{x}$$

Equation 2.10

The Q statistic confidence interval (Q_α) is determined using the average and standard deviation of Q [288-290]. Samples above Q_α do not present robust feature extractions [291]. In these situations, contribution plot can be estimated in order to understand which variables are affecting the Q-statistics [292, 293], indicating why features are not captured.

The other well known statistic parameter in SVD analysis is the Hotelling (T_h^2) that can be used to measure the distance to the center of data. This parameter is determined by (2.12), where A is given by (2.13) and T by Equation 2.6 for the number of variables (n):

$$\boldsymbol{T_h^2} = \boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{W}\boldsymbol{A}^{-\mathbf{1}}\boldsymbol{W}^{\boldsymbol{T}}\boldsymbol{x}$$

Equation 2.11

$$\boldsymbol{A} = \frac{\mathbf{1}}{\boldsymbol{n}-\mathbf{1}}\boldsymbol{T}\boldsymbol{T}^{\boldsymbol{T}}$$

Equation 2.12

The upper confidence interval for the Hotelling T^2 parameter (T_α^2) is determined using the number of relevant singular values, the number of variables and de distribution value with 1 and n-1 degrees of freedom at α level of significance ($\alpha = 0.05$). Samples above the T_α^2 are considered to present significantly different chromatographic features [294]. So, the Q and T_h^2 statistics plot describe the lack of model fit and the samples variability in the PCA subspace, respectively [292].

PCA methodology features several interpretation facilities, including scores and loadings values, biplots (overlap of scores and a loading plot), diagnostic plot (given by Q and T_h^2 statistics) and

contribution plots, which allows an enlarged analysis and overview of the preprocessed data set. Although, several critical documentations have been reported regarding the application of this statistical methodology to chemometrics measurements [29, 295-298]. The model precision for the given purpose, model diagnostics and interpretational issues are the most common issues discussed. Kozak and Scaman (2008) [299] discussed the capacity of PCs with higher percentage of variation for discriminating the information for samples classification, using some already published papers where it is not possible to perform this discrimination.

Kjeldahl and Bro (2010) [298] discuss some practical issues for visual model interpretations, such as: high variance values of two components cannot confirm, *per se,* the conclusion taken from their visualization; correlations or variance between variables cannot be inferred by the proximity to each other or to the origin axis (it is necessary to check these correlations); model validation reinforces the conclusions from the visualization tools.

One of the most used methodologies for data exploring is cluster analysis (CA). Clustering is the non-supervised, semi-supervised and supervised classification of patterns into groups [300]. Grouping of samples is performed based on their similarities, given by the distance, correlation or some combination of both. Two distinct methods of CA can be found, namely hierarchical algorithms or non-hierarchical algorithms. While in hierarchical algorithms the groups are found and elements are allocated to the groups, and this assignment cannot be altered, in non-hierarchical clustering, the assignment of objects into groups can change throughout the application of the algorithm, being thus more flexible for the optimal solution discovery [301].

The increasing of the complexity of experimental design and the continuous improvement of the performance of the analytical instruments created the need to generalize the multivariate statistic tools [302]. Multi-way principal component analysis (MPCA) is one example of a generalized methodology [303, 304], derived from PCA. This is an efficient tool for reducing the high dimensionality of data arrays.

In MPCA, the initial three-dimensional array $T(n, m, t)$, where n corresponds to samples, m to variables and t to time intervals throughout the experiments. This initial tensor is then unfolded to $X(n, tm)$ to reduce the dimensionality of data in order to apply PCA [10]. The output graphics of MPCA is similar to the PCA statistics algorithm application.

Calibration and prediction methodologies

Samples classification is a crucial step in a pattern recognition problem, which dedicates to identify the class or membership of a given object. On the other hand, the purpose of multivariate prediction in metabolomics is to set up for a mathematical relationship for predicting the values of one or more output variables including wine properties or yeasts behavior, based on the observations taken from samples. Partial least squares regression (PLSR) is a classification and prediction model, a prognostic two-block regression method based on estimated latent variables and applies to the synchronized analysis of two data sets of the same objects [35].

PLSR is particularly used for predicting a set of dependent variables – matrix $\boldsymbol{Y}$ - from large sets of independent variables (called predictors) – matrix $\boldsymbol{X}$ [305], as presented in Equation 2.13. The X matrix is similar to that one used in PCA, where the number of rows (n) is the number of samples and the number of columns (m) is the number of variables. In matrix Y, the number of columns (m) represent the number of responses concerning to n samples.

The principles behind the PLSR are similar to that of PCA, although in PLSR, a second piece of information is used, namely, the labelled set of class identities. It maximizes the covariance between the 'test' or predictor variables and the training variable(s) [231], and describes their common structure [305].

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x_1^{(1)}} & \cdots & \boldsymbol{x_m^{(1)}} \\ \vdots & \ddots & \vdots \\ \boldsymbol{x_1^n} & \cdots & \boldsymbol{x_m^n} \end{bmatrix} \quad \textbf{and} \quad \boldsymbol{Y} = \begin{bmatrix} \boldsymbol{y^1} \\ \vdots \\ \boldsymbol{y^n} \end{bmatrix}, \boldsymbol{n}\ \textbf{samples}, \boldsymbol{m}\ \textbf{variables}$$

Equation 2.13

The linear model resulting from the partial least squares (PLS) technique, concerns in the response of a variable matrix $(\boldsymbol{Y})$, a descriptive or predictor variable matrix (TB), a regression coefficient (C^T) and a noise or error term (ϵ), as it is presented in Equation 2.14. Substituting the predictor variable matrix by U as it is presented in Equation 2.15, the y matrix for PLS model application is given by Equation 2.16.

$$\boldsymbol{Y = TBC^T + \epsilon}$$

Equation 2.14

$$\boldsymbol{U = TB}$$

Equation 2.15

$$\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{C}^{\boldsymbol{T}} + \boldsymbol{\epsilon}$$

Equation 2.16

So, given Equation 2.6 and Equation 2.14, PLS regression finds components from $\boldsymbol{X}$ that are also relevant for $\boldsymbol{Y}$. Distinctively, PLSR searches for a set of orthogonal components (called latent vectors) which performs simultaneous decomposition of $\boldsymbol{X}$ and $\boldsymbol{Y}$ that explain as much as possible the covariance between both. So, the latent variables are obtained by Equation 2.17 and Equation 2.18, with the constraints given by Equation 2.19 and Equation 2.20, and $t^T u$ be maximal.

$$x = tw^T \leftrightarrow \boldsymbol{t} = \boldsymbol{x}\boldsymbol{w}$$

Equation 2.17

$$\boldsymbol{y} = \boldsymbol{u}\boldsymbol{c}^{\boldsymbol{T}} \leftrightarrow \boldsymbol{u} = \boldsymbol{y}\boldsymbol{c}$$

Equation 2.18

$$\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{w} = \mathbf{1}$$

Equation 2.19

$$\boldsymbol{t}^{\boldsymbol{T}}\boldsymbol{t} = \mathbf{1}$$

Equation 2.20

A regression step is followed where the decomposition of $\boldsymbol{x}$ is used to predict $\boldsymbol{y}$**.** When the first latent vector is found, it is subtracted from both $\boldsymbol{x}$ and $\boldsymbol{y}$ and the procedure is re-iterated until $\boldsymbol{x}$ becomes a null matrix [305].

So, PLSR models are used for data calibration and prediction. N-way partial least squares (N-PLS) creates a single model where all $\boldsymbol{y}$-columns contribute to the loadings of the model. A partial least squares–1 (PLS-1) is a model built on a *single* $\boldsymbol{y}$-column and the model reflects only the covariance between the $\boldsymbol{x}$ block and that single $\boldsymbol{y}$-column.

Unfolded-partial least squares

The unfolded partial least squares (U-PLS) works similarly to PLS-1 [306], but firstly the second-order data are first vectorized or unfolded along one of the data dimensions [307, 308]. Therefore

a PLS model is built using the unfolded data and the nominal y-column. For example, applying the U-PLS to chromatographic data matrices - $\mathbf{x}(\mathrm{s},\mathrm{m})$ - where s corresponds to retention times or scans, and m to mass-charge ratio, these are firstly vectorized in $v(sm \times 1)$ vectors, and then a usual PLS model is built using these data together with the vector of calibration concentrations - $\boldsymbol{y}(\,I \times 1)$. This provides a set of loadings P and weight loadings W (bothe size $sm \times A$, where A is the number of latent factors or samples) as well as regression coefficiens $v(a \times 1)$. So, the estimation of the analyte concentration is performed according to Equation 2.21, where t_u is the test sample score, obtained by projecting the vectorized data for the test sample $vec(X_u)$ onto the space of the A latent factors (Equation 2.22), where $vec(.)$ implies the vectorization of the operator.

$$\boldsymbol{y_u = t_u^T v}$$

Equation 2.21

$$\boldsymbol{t_u = (W^T P)^{-1} W^T vec(X_u)}$$

Equation 2.22

Classification and prediction methodologies

PLS-R was not initially designed for classification. However, based on the algorithmic structure of PLS-R, different algorithms were developed, namely partial least squares for discriminant analysis (PLS-DA) and partial least squares logistic regression (PLS-LOG), which were easily adapted for the classification and discrimination problems (*i.e.* supervised classification) [309, 310].

PLS-DA is an extension of the PLSR and enables the classification of the multivariate space directions by fitting a regular PLS model between the x matrix and an artificial y matrix providing discriminant directions with well separated observations, according to class membership [231]. In PLS-DA, the response variable is categorical, *i.e.,* indicates the classes/categories of the samples. [310]. An optimal number of latent variables can be obtained by using cross-validation or external data sets. Cross-validation consists in partitioning many times the dataset into a calibration or training set, from which a model will be built, and a validation or test set and there it will be used for assessing the model's performance [59].

PLS-LOG works as a probabilistic classificatoin model, used for predicting the outcome of a categorical dependent variable, based on the predictor variables. The '*logistic'* or '*logit*' regression provides a qualitative response model on estimating the values of the parameters. In PLS-LOG, only two categories of dependent variables are used. At the end, if a variable exhibits a high estimate, it indicates that the variable had a high information in the classification, given the correlation that the variable has with other variables in the data [311].

PLS-DA and PLS-LOG are PLS-R methodologies designed to operate with high-dimensional and highly correlated data. The capacity of these methods to do inference in high-dimensional space makes them ideal candidates for a nonlinear analysis of data based on an increase of their dimension [312]. These are the main reasons for the ability to be applied to high-throughput non-target metabolomics approaches as classification tools.

CHAPTER 3

YEASTS PHENOTYPIC AND PHYSIOLOGICAL RESPONSES TO INDUCED STRESS CONDITIONS

Yeasts response to stress is a complex phenomena and its understanding is of great importance from both scientific and industrial points of view. In this work, toxic and inhibitory conditions were added to fermentations with three different *Saccharomyces cerevisiae* strains: S288c, CA11 and PE-2. The impact of the inhibitory molecules introduced was monitored using the CO_2 production kinetics while physiological and phenotypic changes of yeasts cells were evaluated by measuring biomass, intra- and extracellular metabolites, as well as, the flocculation profile of CA11 strain. Chemometric tools, as multi-way principal component analysis (MPCA) and partial least squares logistic regression (PLS-LOG) were used for characterizing the yeasts physiological behavior under the induced conditions and to predict the flocculation phenotype according to the measured parameters.

According to CO_2 production kinetics and MPCA analysis, it was found that the presence of 1-butanol, furfural and 5-HMF induced physiological and phenotypic changes of the three yeast strains, as a response adaptation. Changes occurred mainly by the interference of these molecules on the enzymatic activity of the glycolytic pathway and biomass production, at different levels. Under these conditions, it was found a statistically significant decrease on CO_2 production rates, a higher glucose accumulation in the extracellular medium, as well as, higher intracellular and extracellular glycerol levels, compared to the control.

The three yeast strains presented distinct fermentations performances, and the physiological response of the strain S288c showed to be completely different from CA11 and PE-2. CA11 strain produced increased levels of intracellular glycerol and its flocculation capacity markedly changed under 1-butanol, furfural and 5-HMF. This result evidenced the occurrence of the inhibition of the flocculation profile in response to the toxic and inhibitory stresses, which showed to be correlated with the overproduction of intracellular glycerol. This relationship was confirmed by PLS-LOG, where intracellular glycerol and trehalose, as well as extracellular acetic acid production showed to be linked to the inhibition of yeast cells flocculation.

The results presented in this Chapter were adapted from:

Castro CC *et al*. Characterization and prediction of physiological and phenotypic behavior of different Saccharomyces cerevisiae strains under induced stress conditions by multivariate analysis (To be submitted).

3.1 INTRODUCTION

The industrial handling of yeasts in bio-fuel, food and beverage production introduces distinct disturbances mainly osmotic, oxidative, temperature, nutrient starvation, ethanol toxicity, by-products inhibition, among others [7, 8]. The yeasts exposure to these harsh environments usually results in transformations in their genome, cytoplasm and cell membrane, being related with morphological and physiological differentiations, which largely interfere with the population dynamics and fermentation process [313].

Food and beverage fermentation using *Saccharomyces cerevisiae* is accompanied by the production of several compounds, namely alcohols, aldehydes, organic acids, esters, organic sulphides, carbonyl compounds and fusel alcohols [3]. At high concentrations, these fusel alcohols, as isopropanol and *tert*-Amyl alcohol (2-methyl-2-butanol) impart off-flavors and toxic conditions to yeasts cells whereas at low concentrations, together with esters, make an important contribution to flavors and aromas quality of the final product [4, 5]. The presence of some fusel alcohols within the fermentation processes may induce oxidative stress, being associated to morphology abnormalities, including cells filamentation in both haploid and diploid cells [71].

In very high gravity (VHG) fermentation processes, characterized by high saccharides content in the fermentation media, for bio-ethanol production, the range of pre-treatments that can be applied to the lignocellulosic raw materials for obtaining a more accessible feedstock for subsequent fermentation, introduces some inhibitory compounds for yeasts, including furan derivatives (5-Hydroxymethylfurfural – 5-HMF - and furfural) [314]. The composition of sugars and inhibitors in these media vary with raw material and the chosen pre-treatment and hydrolysis methods [315]. Bio-butanol production with engineered *S. cerevisiae* strains has also been explored as a gasoline substitute [111, 113, 114]. Compared to bio-ethanol, bio-butanol presents higher energy content and lower water absorption and volatility, making this an interesting bio-fuel, although it has an increased toxic effect on microorganisms cells [104, 116, 117].

High ethanol and other toxic compounds concentrations introduce some oxidative stresses in the medium, that may interact with yeast cells membranes by changing their polarity and declining its function as a hydrophobic barrier [131].

Trehalose is one of the main reserve carbohydrates in yeasts, which provides cells protection through stress conditions by keeping their structure [132, 133]. This disaccharide is also related

with the protection from lipid peroxidation during oxidative stresses, as it is an important ROS scavenger [316].

Also, glycerol has two essential functions in yeasts cells, namely as a sink for the excess NADH which is produced by anabolic reactions during anaerobic conditions [317] and as an osmolyte, balancing a high external osmotic pressure during salt or sugar stress due to low water activity [318]. *S. cerevisiae* intracellular glycerol increases with the decrease of the water potential in the medium which leads to its role in osmoregulation of yeast cells by balancing the osmotic stress of the membranes [136].

Nowadays, both industry and research groups are focused in developing new strategies to obtain customized microorganisms to enhance a specific bioprocess. These strategies can be performed by genome modifications or by screening existing organisms showing better adaptive responses under specific stresses [88]. Genomic characterization of yeast strains implies inherent costs and labour limitations, hindering the compilation of a complete database of the thousands of different yeasts strains with different properties, bio-geographical distribution and applications [319]. So, one easier and common way to select yeasts strains, according to the bioprocess requirements and purpose, is their phenotypic characterization, allowing to select the most adapted for improved performances under specific fermentations conditions, based on an "*eliminatory*" approach that reduces the initial high number of the candidate strains [320].

Yeast phenotypic switching is the reversible occurrence of the spontaneous emergence of colonies with different morphological and physiological characteristics, providing population variability and contributing to a rapid adaptation to a changing harmful environment. During phenotypic switching of strains, genes representing diverse functions, including metabolism, adhesion, cell surface composition, stress response, signalling, mating type, and virulence are differently expressed, and one third of the differences among cell types are known to be related to metabolic pathways [321]. The study of protective and stress tolerance mechanisms of some indigenous strains during the industrial fermentation processes has raised great interest over the years [8, 88, 322, 323], and several strategies have been suggested to improve yeasts efficiency, including single gene manipulation, evolutionary engineering studies or genome shuffling, as well as optimization of fermentative medium [7, 324].

Fermentations progress is usually monitored by measuring some variables, such as sugars, temperature, pH and dissolved oxygen. However, these are still not enough to detect and diagnose problems in alcoholic fermentations [325]. Chemometrics methods, concerning the use of multivariate "*empirical"* modelling in chemical or biochemical experiments, have been developed for describing fermentation processes by extracting its crucial information and thus increasing the knowledge of the bioprocess [29]. Principal component analysis (PCA) and partial least squares regression (PLSR) are some of the statistical tools developed in this context [29, 229]. The use of both methodologies provides a more realistic framework of chemical and biochemical data analysis, as the joint effect of variables is taken into account, instead of using only one or a very few variables at the same time [29, 229]. These were already used to discriminate different wine strains varieties according to genetic and phenotypic characteristics [319], yeasts metabolic state through fermentations based on spectroscopy [286] and abnormal wine fermentations detection [326].

The present study aims at monitoring and characterizing the physiological effect of the presence of different stress molecules through the fermentations responses of three different strains of *S. cerevisiae* - S288c (reference strain), CA11 (industrial and flocculent) and PE-2 (industrial). Toxic and inhibitory conditions added were 1.0% (v/v) ethanol, 1.0% (v/v) 1-butanol, 1.0% (v/v) isopropanol, 1.0% (v/v) *tert*-Amyl alcohol, 0.2% (v/v) furfural and 0.5% (v/v) 5-HMF. Control fermentations were also performed using each strain, in the absence of stress. The characterization was performed based either on the measurement of physiological parameters, including biomass, CO_2, ethanol, glucose, glycerol and acetic acid concentrations, or on the determination of intracellular concentrations of trehalose and glycerol. With respect to the flocculent strain, CA11, the impact of the induced fermentation conditions on the flocculent phenotype was explored by evaluating its sedimentation profile. Multi-way principal component analysis (MPCA) [326] and partial least squares logistic regression (PLS-LOG) [309, 310], were the statistical methodologies applied to parameters measured during batch fermentations.

3.2 MATERIALS AND METHODS

3.2.1 Fermentation process

Three *S. cerevisiae* strains were used, including: laboratory reference – haploid S288c [327] – an industrial flocculent strain - CA11 - belonging to Federal University of Lavras (Brazil) collection, isolated from sugar cane fermentation process to produce "*cachaça*" [328] – and the industrial strain, isolated from Brazilian industrial bio-ethanol production also from sugar cane fermentation - PE-2 [329]. The incubation was performed in YPD broth (YPD_b) (Sigma Aldrich - ref. Y1375, USA) – 1.0% (w/v) yeast extract, 2.0% (w/v) bacto-peptone and 2.0% (w/v) glucose - during 12 hours at 30 °C under constant agitation (150 rpm) for 12 - 14 hours.

Liquid fermentations were performed in YPD_b, where stress conditions were added, namely 1.0% (v/v) ethanol (≥ 99.8% Riedel-de-Haën, Germany), 1.0% (v/v) 1-butanol (≥ 99.0%, Sigma Aldrich, USA), 1.0% (v/v) isopropanol (≥ 99.5%, Sigma Aldrich, USA), 1.0% (v/v) *tert*-Amyl alcohol (≥ 99.0%, Sigma Aldrich, USA), 0.2% (v/v) furfural (98.0%, Sigma Aldrich, USA), and 0.5% (v/v) 5-HMF (99.0%, Sigma Aldrich, USA). Control fermentations without adding any toxic substance were also performed. These concentrations were considered based on different researches and according to their toxic and inhibitory effects in cells physiology, such as the impact on cells morphology and conformation [71, 107, 111, 115], and/or their influence in fermentation rates reduction [330, 331]. A control fermentation was also done with each strain.

Fermentations were performed using 50 mL of YPD broth in *Erlenmeyer* flasks (100 mL) fitted with perforated rubber stoppers enclosing glycerol-locks for maintenance of anaerobic conditions at 150 rpm orbital agitation, as it is presented in Figure 3.1 [324].

Cell suspension was aseptically collected from incubation medium by centrifugation (10 min at 9000 × *g*, 4 °C) and re-suspended in 0.9% (w/v) NaCl to a concentration of 200 mg fresh yeast/mL. The yeasts cells were pitched at about 1.0×10^6 cells/mL into the culture medium to start the fermentation. Stress conditions were added to the medium except in the control, which was performed in the absence of stress conditions. Liquid fermentations lasted for approximately 24 h at 30 °C, and were monitored by mass loss (related with CO_2 production [324]). Samples were taken within each fermentation condition at times corresponding to different levels of CO_2 production: 0, 10, 40, 85, 95 and 100% of the total concentration of CO_2. For each sample, determinations of yeast cells biomass (dried weight), extracellular glucose, ethanol, glycerol and

acetic acid and intracellular glycerol and trehalose concentrations were performed. For the flocculent strain, CA11, sedimentation assays were also performed in order to understand the effect of each induced stress condition in the flocculation capacity.

Figure 3.1 Experimental setup for liquid fermentations.

3.2.2 Analytical procedures for determination of dried yeast biomass and extracellular metabolites

Samples for yeast dry mass were prepared by centrifuging 5 mL of the fermentation broth sample (10 min, 4800×g, 4 °C) and resuspended in the same volume of 0.9% (w/v) of NaCl to wash the yeast cells. From these, 1.5 mL at each level was centrifuged (10 min, 14000×g, 4 °C) in a pre-weighted dried tube (24 hours at 105 °C). The supernatant was completely removed and the pellet was washed with the same volume of distilled water, mixed in the vortex and centrifuged again and dried overnight at 105 °C. Dried yeast mass was obtained by the difference between the mass of the tube plus yeast pellet and the mass of the dried tube [332].

Batch fermentations samples were centrifuged (10 min, 4800×g, 4 °C) and the supernatant was decarbonated to remove the dissolved CO_2 in order to correctly measure the pH, then the supernatant was stored at -20 °C. Glucose, ethanol, glycerol and acetic acid were measured using High Performance Liquid Chromatography (HPLC) analytical technique with a refractive index (RI) detector and a MetaCarb 87H (300 × 7.8 mm) column at 60 °C, using 0.005 mol/L of H_2SO_4 as eluent at 0.7 mL / min flow rate. All standards employed were of analytical grade.

3.2.3 Intracellular metabolites quantification - trehalose and glycerol – and flocculation assays

Yeasts cells were harvested from the liquid fermentation media at the stationary growth phase and centrifuged (5 min, 4800×g, 4 °C). The supernatant was rejected and cells were weighted and re-suspended in an ice-cold 0.9% (w/v) NaCl solution, to achieve a final concentration of 200 mg of fresh yeast per mL. The cell suspension was mixed in the vortex until the solution was homogeneous and then 0.5 mL of these samples was boiled for 5 min to disintegrate yeast membranes. Boiled samples were then centrifuged (5 min, 13400×g, 4 °C), and the supernatant was collected to a previously weighted tube. Trehalose and glycerol were quantified by HPLC, using the same conditions as for extracellular metabolites quantified (referred in Sub-section 3.2.2). The final concentrations of intracellular trehalose and glycerol were normalized to yeast dry mass, by considering that dry yeast mass corresponded to 30% (w/w) of the fresh yeast mass [324, 333].

3.2.4 Flocculation capacity analysis

The flocculation capacity of *S. cerevisiae* CA11 strain under each fermentation condition was evaluated, according to Soares *et al.* (1992) [334], with slight modifications from Gomes *et al.* (2012) [333]. Samples were firstly deflocculated using a 15 g / L sodium chloride (NaCl), pH 3.0, solution and then 1 mL of 100 mM calcium chloride ($CaCl_2$), pH 3.0, was added to aggregate cells again. 200 µL samples were periodically (every 2 min) taken at a fixed distance of 20 mL in the graduated cylinder during 10 min and resuspended in NaCl for optical density at 600 nm (OD_{600}) measurements. Control assays were also performed without $CaCl_2$ for all the fermentation conditions tested. The sedimentation profiles were obtained by plotting the percentage of yeast cells in suspension for each sample.

3.2.5 Chemometrics applied to fermentations

Six samples were collected from each of the 21 different batch fermentations, concerning to the three tested yeast strains through seven different fermentation conditions. For each sample, physiological parameters were measured. The same number of samples per fermentation provided a homogeneous data matrix, enabling to compare variables discrimination between fermentations.

Multi-way principal component analysis (MPCA), a principal component analysis (PCA) variant method published by Urtubia *et al.* (2012) [326], was used to classify samples according to the measured physiological information. Samples variability and variables contribution for their differences were explored by scores and loadings of the MPCA methodology. Furthermore, partial least squares logistic regression (PLS-LOG) was used to predict the CA strain flocculation phenotype according the induced stress conditions for fermentation.

The 3D data tensor $\boldsymbol{X}$ *(S,V,T),* where S corresponds to samples (strains under each condition), *V* corresponds to variables (parameters measured for each sample), and *T* corresponds to batch fermentation time-course, was organized according to the methodologies purposes. In MPCA the tensor **X** is unfolded to a bi-dimensional matrix for further applying PCA [326]. In MPCA the bi-dimensional $\boldsymbol{M}$ matrix corresponds to $S \times TV$ dimensions, where each row concerns to batch fermentation.

PLS-LOG is a probabilistic classification model [335] which was used for predicting the outcome of the flocculation phenotype of CA11 strain, based on the predictor features, that is, the physiological behavior of the strain under each fermentation condition. So, the **X** tensor, unfolded to the **M'** matrix as $TS \times V$ dimensions, and a *Y* matrix are used for PLS-LOG prediction. The logistical or logit regression determines the impact of the independent variable (the flocculation phenotype) to predict the membership of the dependent variables (physiological parameters), determining the probability of success over the probability of failure of the flocculation phenotype, according to the physiological behavior under the used conditions. The partial least squares (PLS) methodology maximizes the co-variance between both $\boldsymbol{M'}$ and $\boldsymbol{Y}$ matrices, while determines the eigen vectors and eigen values of the co-variance matrix between the two matrices [35].

3.3 RESULTS AND DISCUSSION

3.3.1 Fermentations synchronization

The introduction of stress or inhibitory molecules to fermentations leads to different yields and distinct metabolites production or consumption rates [21]. In this context, the synchronization of fermentations performed under different environmental conditions becomes a crucial step for

comparing the physiological measurements and to evaluate the performance of each *S. cerevisiae* according to the external conditions.

In this study, the way to synchronize fermentations was selecting sampling data points concerning to percentages of CO_2 production (0, 10, 40, 85, 95 and 100% of the total concentration of CO_2), for all fermentation conditions. Several preliminary fermentations were previously performed using the same conditions, in order to accurately define these sampling points concerning to different growing phases of yeast cells: lag, exponential and stationary phases. Thus, it was possible to compare the physiological information of the three yeasts strains used, under different stress conditions.

In Figure 3.2, CO_2 production kinetics within different fermentations for the three strains are presented. According to this figure, similar fermentation profiles were obtained for S288c, CA11 and PE-2 (Figure 3.2 A, B and C, respectively), however, the adaptation of each strain seems to be dependent on the fermentation condition. For the three strains studied, the presence of 1.0% (v/v) ethanol (■), isopropanol (×) and *tert*-Amyl alcohol (+), did not cause a statistically evident change in the CO_2 production profiles (at a level of significance of $\alpha = 0.05$), when compared to the control fermentation (♦). This result suggests that the introduction of such concentrations - 1.0% (v/v) - does not represent a harmful environment for these strains, as these metabolites can be produced during the fermentation process and yeasts can easily adapt to their presence in the medium.

The adaptation to environmental conditions can be observed during the lag phase for the three strains, where it seems to be lower in the presence of unfavourable conditions compared to the control (♦), although less evident for PE-2 (Figure 3.2 C) and CA11 strains (Figure 3.2 B). The stress response mechanism of industrial strains, commonly used in bio-ethanol or "*cachaça*" production, within VHG processes, should easily be able to adapt to ethanol rich environments, as final ethanol concentrations up to 10% (v/v) can be reached [88, 336]. CA11 and PE-2 can tolerate up to 17% (v/v) ethanol [88], and the effects on yeasts physiology induced by the presence of ethanol at high concentrations have been reported for many years [337]. Lorenz *et al.* (2000) [71] have previously demonstrated that the presence of only 1% (v/v) ethanol stimulated hyperfilamentation of diploid cells in nitrogen limiting conditions, however in YPD medium the fermentation rate was not affected nor the morphology of cells (results not shown).

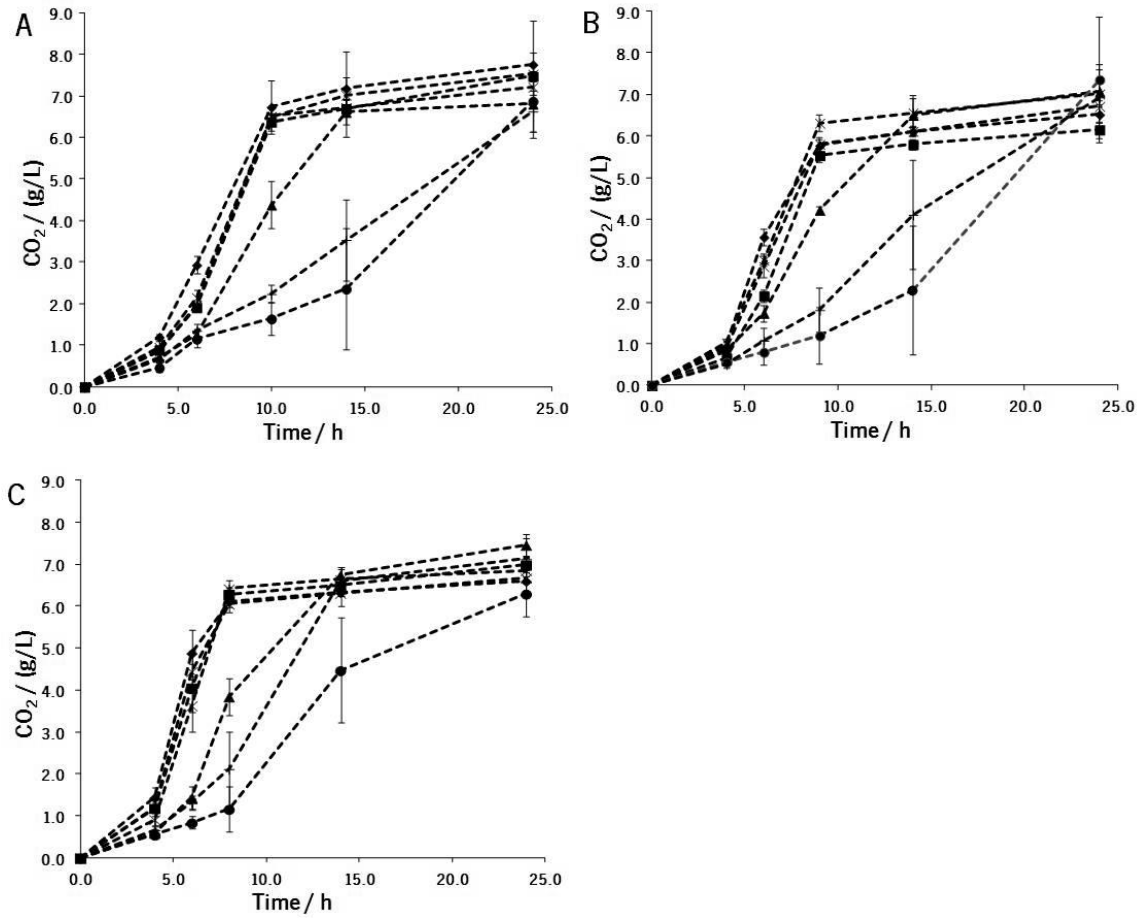


Figure 3.2 CO_2 profiles within YPD_b fermentations using: A – S288c , B – CA11 and C – PE-2, under different fermentation conditions: (♦) control, (■) 1.0% (v/v) ethanol, (▲) 1.0% (v/v) 1-butanol, (×) 1.0% (v/v) isopropanol, (+) 1.0% (v/v) *tert*-Amyl alcohol, (●) 0.2% (v/v) furfural, (-) 0.5% (v/v) 5-HMF.

On the other hand, it is possible to observe that in the presence of 1.0, 0.2 and 0.5% (v/v) of 1-butanol (▲), furfural (-) and 5-HMF (●), respectively, the CO_2 production rate undergoes a statistically significant decrease for the three strains (at a level of significance of α= 0.05), compared to the other fermentation conditions. Figure 3.3 presents the fold-changes, compared to the control fermentation, occurring under these significant conditions.

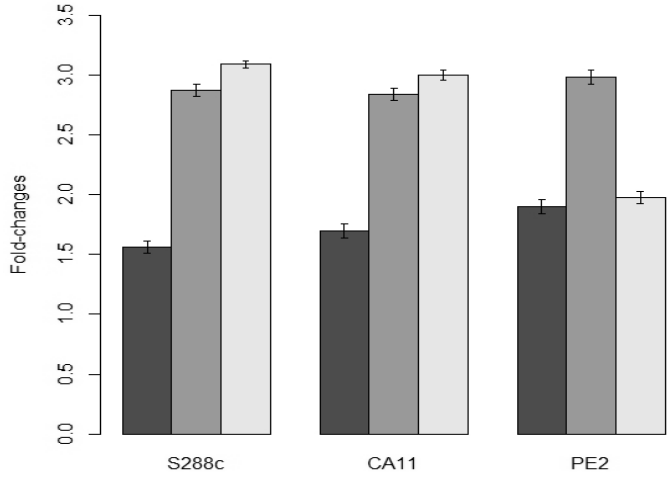


Figure 3.3 Fold-changes in CO_2 production rates, introduced by the presence of the conditions that present a statistically significant effect on growth: ■ 1% (v/v) 1-Butanol; ■ 0.2% (v/v) Furfural; and ■ 0.5% (v/v) 5-HMF.

As it is possible to observe in Figure 3.3, S288c and CA11 showed a similar effect on CO_2 production rates decrease in the presence of 1-butanol, furfural and 5-HMF. The higher decline was verified under 0.5% (v/v) 5-HMF – about 3-fold reduction change -, followed by 0.2% (v/v) furfural and finally 1% (v/v) 1-butanol - 2.9- and 1.6-fold reduction changes. PE-2 verified the higher reduction under 0.2% (v/v) furfural, which was similar to S288c and CA11, followed by the reduction in the presence of 0.5% (v/v) 5-HMF and finally by 1% (v/v) 1-butanol - about 2-fold reduction changes.

Some studies have been performed regarding the furan derivatives effect in fermentation process, especially in bio-ethanol production using lignocellulosic raw materials. Furfural and 5-HMF are known to compete for NADH, interfering with cell glycolysis, during regeneration of NAD^+ [96]. Taherzadeh *et al.* (2000) found that furfural has a more severe inhibitory effect in yeasts cells than 5-HMF [338], and Almeida *et al.* (2007) proved that it is also present in lower levels in the fermentation medium (*e.g.* in VHG medium) [87]. These authors also observed that the tolerance mechanism of furans leads to the increased availability of the NADPH and NADH levels in the cell, mainly produced by the pentose phosphate pathway (PPP). These are the cofactors used by

NADPH-dependent alcohol dehydrogenase and NADH-dependent alcohol dehydrogenase to reduce HMF and furfural, respectively. So, the tolerance mechanism of these molecules depends on the availability of both free NADPH and NADH [339]. An increased activity of PPP might also mediate the protection and repairing of furans induced damage in cells and many NADPH- and NADH-dependent enzymes are responsible for the cellular defence mechanisms against various stresses [340].

In this study, the lower negative effect in CO_2 production rates, which leads to the higher tolerance capacity, of PE-2 under 5-HMF suggests that this strain might have increased PPP activity and thus, higher levels of NADPH which can be used for furans reduction in the medium. This improved tolerance highlights its higher ability to be used in bio-ethanol production through lignocellulosic raw materials. Although furfural and 5-HMF are some of the inhibitory compounds most commonly found in lignocellulosic hydrolysates, a variety of other inhibitors may also be present in the industrial medium, according to acid concentrations, temperature and other conditions used for hydrolysis, which might affect yeasts metabolism [341].

Several studies were already performed in this context, where PE-2 showed a high-quality performance under VHG fermentation conditions, which introduce increased concentrations of saccharides and inhibitory compounds, as well as to higher concentrations of the final ethanol produced [88, 336]. Also different approaches have been used in order to improve its fermenting capacity under harsh conditions [333].

The effect of 1-butanol in the CO_2 production rates of the three strains is very similar, and the significant effect can be mainly explained by the strongest ability of this molecule to permeate and/or interact with the cellular membrane [118, 119]. It can induce the loss of the cells ability to maintain the internal pH (due to the increased proton permeability of the cytoplasmatic membrane), the inhibition of the membrane ATPase [120, 121], the loss of intracellular molecules, as proteins, RNA and ATP (as the fluidity of the membrane increases) [110] and finally the obstruction of glucose uptake [120].

3.3.2 Multiway principal component analysis (MPCA)

MPCA methodology was applied to the unfolded $\underline{X}$ matrix, leading to the physiological parameters representing the stress adaptation behavior of the three strains to each induced condition, and the resulting scores and loadings are presented in Figure 3.4. In this figure, scores concern to samples (represented as symbols with three replicates) distributed through the PC1 *versus* PC2 space while loadings are represented by the arrows. Two relevant orthogonal decompositions were found to explain 77.5% of the physiological data variability (55.5% PC1, 22.0% PC2). According to the relevant decomposition PC1 presented in Figure 3.4, the physiological response to the induced fermentation conditions of the laboratorial strain – S288c (black symbols) – is distinct from the industrial strains – CA11 (blue symbols) and PE-2 (green symbols). On the other hand, the distribution of samples through PC2 emphasizes the physiological differences between fermentations performed in the presence of 1.0% (v/v) 1-butanol (+), 0.2% (v/v) furfural (∇) and 0.5% (v/v) 5-HMF (□) compared to control and in the presence of 1% (v/v) ethanol (Δ), isopropanol (×) and *tert*-Amyl alcohol (◊), suggesting that both groups trigger different stress response mechanisms to yeasts.

The distribution of loadings (variables) through the PC1, in Figure 3.4, evidences the increased contribution of glucose, ethanol and acetic acid concentrations for differences between S288c and the industrial strains CA11 and PE-2, which in turn are most influenced by biomass, CO_2 and glycerol production kinetics. In PC1 it is also evident that the presence of 1-butanol induced the production of higher concentrations of acetic acid, mainly by S288c. In PC2, the presence of furfural and 5-HMF and 1-butanol lead to an increased production of glycerol, higher accumulation of glucose in the fermentation medium and lower concentrations of ethanol and CO_2 produced by the three strains tested. Under these fermentation conditions, the higher accumulation of glucose in the medium occurs when using the S288c.

Acetic acid is an important end-product of energy metabolism [342], and its enhanced production entails the occurrence of an increased production of its precursor acetyl-CoA and energy to survive to unfavourable conditions [343]. In the presence of 1-butanol, an overproduction of acetic acid and a higher accumulation of glucose are evident, mainly during the lag phase (T0 to T2). This fact confirms the enlarged levels of toxicity for S288c, suggesting that it is affecting the glycolytic pathway, which begins with glucose consumption. In *S. cerevisiae,* two main routes are known to produce acetyl-CoA, namely, pyruvate breakdown and β-oxidation of fatty acids [342]. So, in this

context, as pyruvate production by glycolysis is being limited by the presence of such stress, it is suggested that the acetyl-CoA factor can be produced by the fatty acids metabolism [343].

The increased accumulation of glucose in the fermentation medium using S288c in relation to CA11 and PE-2, under the same conditions, highlights the higher susceptibility of S288c to the external toxic and inhibitory conditions. Effectively, CA11 and PE-2 have been used in some research works on VHG fermentation conditions, where they are exposed to high concentrations of saccharides and inhibitory compounds, as well as to higher concentrations of the final ethanol produced, both revealing increased capacity to ferment under such harsh conditions [88, 336].

On the other hand, S288c is a laboratory strain obtained by genetic crosses, completely sequenced and is used as a reference strain being therefore more susceptible to stressful conditions [324].

The increased concentrations of extracellular glycerol produced in response to the presence of 1-butanol, furfural and 5-HMF, mainly by CA11 and PE-2, suggests the increment of the glycerol-3-phosphate dehydrogenase activity [344]. For these two strains, the production of extracellular glycerol seems to occur in order to equilibrate the redox balance and regenerate reduced nicotinamide adenine dinucleotide (NADH) associated with the production of biomass [344, 345], which is also increasing.

The presence of 1% (v/v) ethanol, isopropanol and *tert*-Amyl alcohol molecules in fermentations does not introduce drastic changes in the fermentation medium of the three strains tested, as the three yeasts were able to produce the main end-products, ethanol (E) and CO_2 (C), within fermentations. This suggests that the glycolytic pathway, through which these products are produced, was not affected by the presence of 1% (v/v) ethanol, isopropanol and *tert*-Amyl alcohol molecules at the concentrations tested.

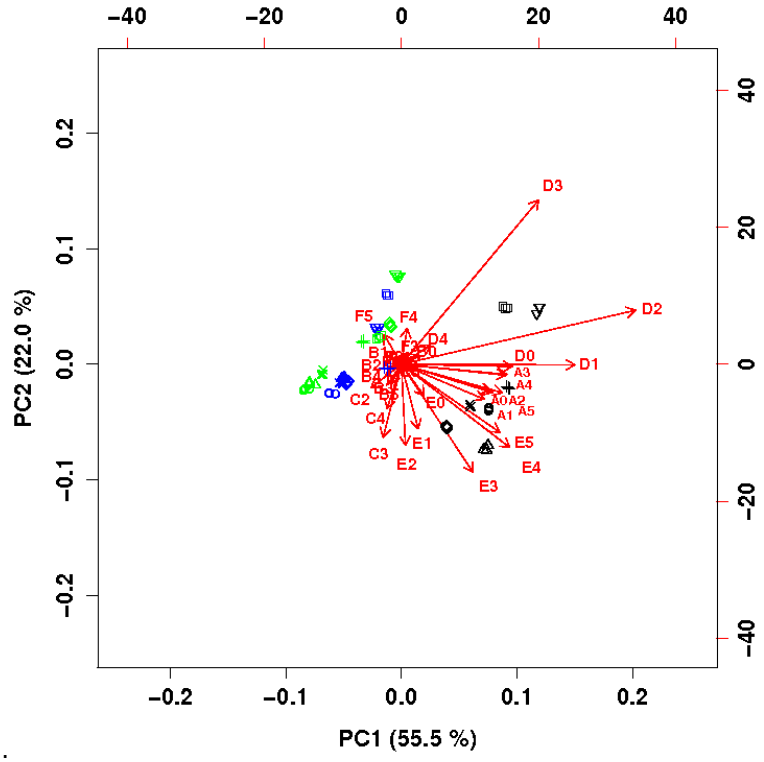


Figure 3.4 Biplot of MPCA scores (symbols) and loadings (arrows), using the three replicates of physiological measurements of S288c (black), CA11 (blue) and PE-2 (green) strains under: o - Control; Δ 1% (v/v) ethanol; + 1% (v/v) 1-butanol; × - 1% (v/v) isopropanol; ◊ - 1% (v/v) tert-Amyl alcohol; ∇ 0.2% (v/v) furfural; and □ 0.5% (v/v) 5-HMF. Loadings correspond to variables measured: acetic acid (A), biomass (B), CO_2 (C), glucose (D), ethanol (E) and glycerol (F) in time 0 to 5.

3.3.3 Intracellular metabolites analysis

Intracellular metabolites were measured at the end of the fermentation processes under all the environmental conditions tested. In Figure 3.5, the intracellular trehalose concentrations of each strain under the different fermentation conditions, normalized to the dried yeast cell mass, are presented. According to this figure it is possible to observe that the industrial flocculent strain CA11 revealed higher levels of trehalose (6.53 to 34.65 mg/g_{DY}) when compared with PE-2 and S288c, for all the induced conditions. The higher concentration was obtained in the presence of 1.0% (v/v) ethanol. The presence of 1.0% (v/v) ethanol, 1-butanol and isopropanol induced an increased production of intracellular trehalose by CA11, when compared to the control. Different authors have shown that contents of intracellular trehalose in flocculent yeasts are directly related to its tolerance to toxic conditions, such as the presence of ethanol [7, 346].

Trehalose content in S288c and PE-2 cells was lower than in CA11, and ranged between 0.14 and 9.65 mg/g_{DY} and 0.02 to 2.48 mg/g_{DY}, respectively. Still, the levels of trehalose in PE-2 were much lower than in S288c.

The higher levels of trehalose for the S288c strain were obtained in the presence of 1.0% (v/v) ethanol, 0.2% (v/v) furfural, and 0.5% (v/v) 5-HMF, whereas for the PE-2 were in the presence 1.0% (v/v) 1-butanol (2.48 mg/g_{DY}) and for 0.5% (v/v) 5-HMF (1.82 mg/g_{DY}).

It was already reported that the presence of alcohols in the fermentation medium induces the production of reactive oxygen species (ROS), which can react with and damage complex cellular molecules, including lipids, proteins, and nucleic acids [316, 347, 348]. Mitochondria electron chain, is the major intracellular source of ROS, namely superoxide anion and hydrogen peroxide [348]. When exposed to oxidative conditions, one of the most important targets for oxygen-derived free radicals, known as reactive oxygen species (ROS), is the membrane phospholipids. This is mainly because of the increased levels of unsaturated fatty acids and the high solubility of molecular oxygen in hydrophobic membranes relative to aqueous environments [316]. So, the sensitivity to both heat and oxidative stress was dependent on membrane lipid composition [347].

According to some authors [349, 350], trehalose is involved in the equilibrium of yeasts membrane and may also be accumulated during exponential growth and used as a reserve sugar under starvation. The production of trehalose within the fermentation process was found to be also related with the protection from lipid peroxidation, during the oxidative stress, being very important in scavenging ROS, similarly to ascorbate, flavonoids and glutathione [316].

As it was referred in Section 3.3.1, pentose phosphate pathway (PPP) seems to have a relevant function in furfural and HMF tolerance, as it regenerates the co-factors needed to reduce the furan-derivates to less-inhibitory molecules [340]. Gorsich *et al.* (2006) showed that an over-expression of the first PPP enzyme, namely *ZWF1*, can commit glucose 6-phosphate to the PPP (to produce D-6-phospho-glucono-δ-lactone), as opposed to other pathways, such as glycolysis or trehalose synthesis [339].

Taking into consideration these studies and the results obtained, in the presence of ethanol, 1-butanol and isopropanol (hydroxyl functional group-containing molecules), CA11 produces higher levels of trehalose for the survival of yeast cells and to reduce lipid peroxidation by free ROS. In S288c and PE-2 fermentations, the higher levels (but much lower than in CA11) of trehalose

produced under furfural and 5-HMF suggested that the activity of *ZWF1* protein is decreased under these conditions.

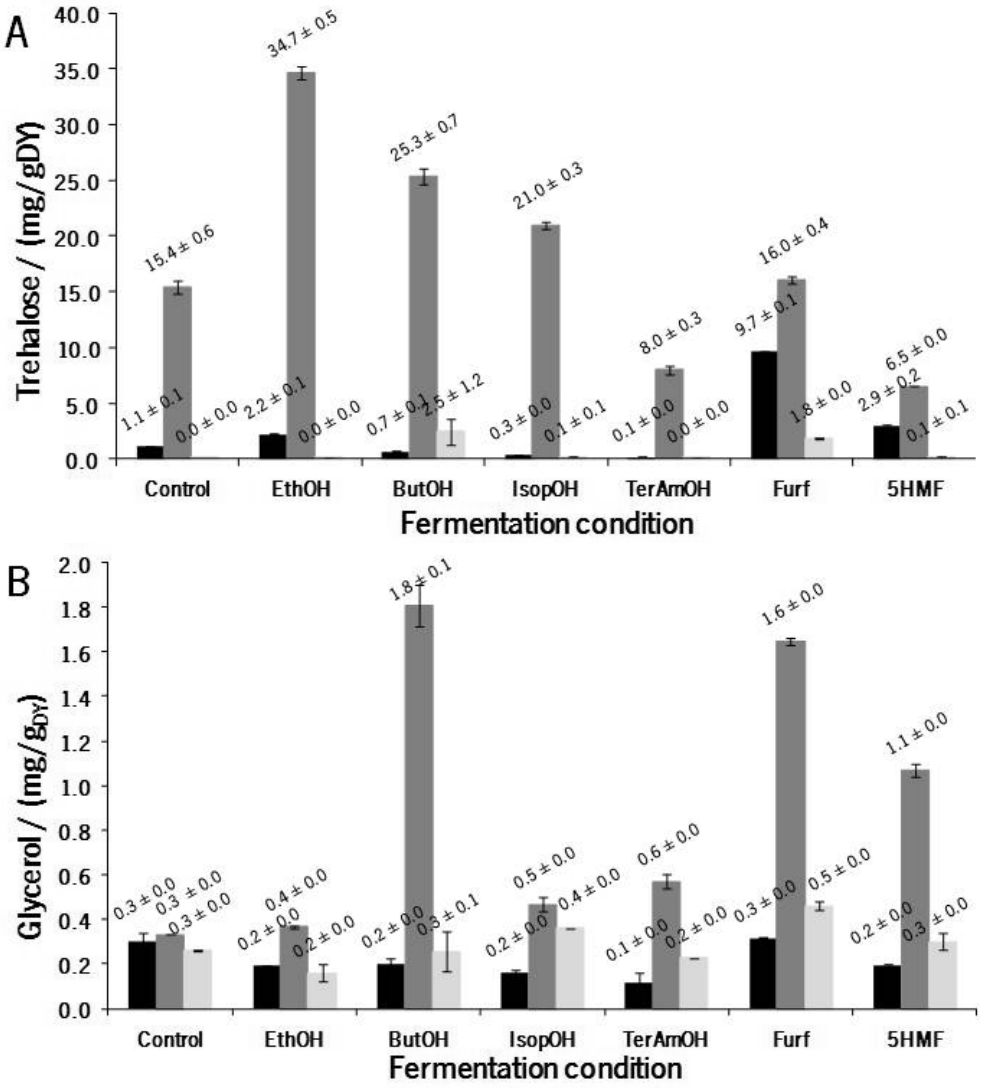


Figure 3.5 Intracellular metabolites at the end of fermentation processes: A – Trehalose; B – Glycerol using ■ S288c, ■ CA11 and ■ PE-2.

Based on the intracellular glycerol measurements presented in Figure 3.5B, it is possible to observe an increase of internal glycerol content for the strain CA11 when compared to S288c and PE-2. Also, for the CA11, 1.0% (v/v) 1-butanol, 0.5% (v/v) 5-HMF and 0.2% (v/v) furfural, revealed higher internal glycerol content values. These results confirm also that the presence of these three molecules induces cell membrane composition changes [92-94, 118, 119], influencing the activity

of the transporters proteins responsible for the regulation of the intra and extracellular osmotic pressure, thus controlling the by-product production and accumulation in the intracellular and extracellular media [351].

3.3.4 Stability of yeast flocculation under stress conditions

The flocculation phenotype of yeast cells entails some advantages when compared with non-flocculating yeasts used in biotechnological processes, mainly the possibility of reusing cells for extended periods of time, the easiness of separating cells from the liquid phase and/or the minimization of the contamination risk [126]. In this study, flocs formed by CA11 in control fermentation, appear to be bigger than in the presence of stress, suggesting that the oxidative conditions induced weakened the forces involved in the flocculation phenomena, resulting, in some cases, in the loss of the ability to form flocs [2]. In order to quantify the effect of stress in the flocculation properties of CA11 strains, the sedimentation capacity was determined under the different oxidative stresses, as displayed in Figure 3.6.

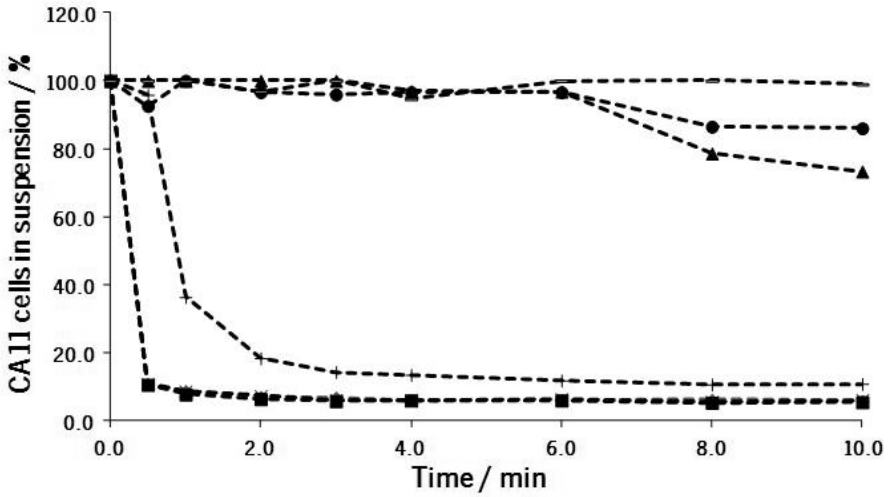


Figure 3.6 Sedimentation capacity of CA11 cells in YPDb under: (♦) control, (■) 1.0% (v/v) ethanol, (▲) 1.0% (v/v) 1-butanol, (×) 1.0% (v/v) isopropanol, (+) 1.0% (v/v) *tert*-Amyl alcohol, (●) 0.2% (v/v) furfural, (-) 0.5% (v/v) 5-HMF.

As it is possible to observe in Figure 3.6, the different induced stresses originated distinct sedimentation behaviors. Enhanced sedimentation profiles were found in the control and under

1.0% (v/v) ethanol and isopropanol, suggesting that these conditions favour the flocculation of yeasts cells. CA11 has a natural flocculent phenotype [352] that is preserved under favourable environmental conditions; the addition of ethanol confirms the results that indicate its positive effect on flocculation leading to the reduction on cell-cell electrostatic repulsion, improving the cell interaction and consequently enhancing flocculation [353-356]. Also, the presence of isopropanol was identified as influencing the phylamentous growth of yeast cells [71] and its addition to the fermentation media did not change the flocculation phenotype [356].

Furthermore, the presence of 1.0% (v/v) *tert*-Amyl alcohol somehow affects the cell wall composition and interferes in the yeast flocculation profile as an adaptive response [2]. Finally, a distinct behavior was observed for CA11 strain under 1, 0.2 and 0.5% (v/v) 1-butanol, furfural and 5-HMF, situations where a high percentage of cells in suspension - 70, 80 and 100% respectively – is observed. This observation was more representative under 0.5% (v/v) 5-HMF, where cells remained suspended in the fermentation medium. Changes in sedimentation profiles of CA11 in the presence of 1-butanol, furfural and 5-HMF suggest that these interfere with the recognition mechanisms of cell surface proteins (called “adhesins” or “flocculins”), and thus affect the flocculation phenotype [127].

So, as it was observed an overproduction of glycerol under 1.0, 0.2 and 0.5% (v/v) of 1-butanol, furfural and 5-HMF, it is suggested that a direct relation between glycerol production and flocculation inhibition for the CA11 yeast strain might occur and thus with the CO_2 production rates. Under toxic conditions, yeasts enable their defence mechanism producing more glycerol and inhibiting flocculation [357, 358]. These results indicate that mechanisms used by CA11 as a response to the stress conditions leads to the production of both trehalose and glycerol within the fermentation process, increasing its chances of survival.

3.3.5 Flocculation classification using partial least squares logistic regression

In order to better understand changes on the flocculation phenotype of CA11, according to the physiological measurements, partial least squares logistic regression (PLS-LOG) was used. The $\boldsymbol{M}'$ matrix, unfolded from the $\boldsymbol{X}$ tensor as $TS \times V$ (where TS correspond to samples took at the different fermentation times and V to the physiological parameters measured at each time), was used for predicting the flocculation phenotype defined in the Y matrix. In Figure 3.7 the PLS-LOG model

results are presented, leading to the model scores (Figure 3.7A) and coefficients (Figure 3.7B). In Figure 3.7A, it is possible to confirm the different behavior of the CA11 yeast strain under different stress conditions. Differences found in the sedimentation profiles of this strain under each induced condition, presented in Figure 3.7 and discussed above in Section 3.3.4, can be predicted by the PLS-LOG model with a high correlation value (R^2= 0.994, p-value < 0.001). This result indicates that changes occurring in the physiological response of CA11 under the different conditions can predict the flocculation profile of this strain under each condition. This can be seen in Figure 3.7A, where the capacity to flocculate is discriminated through the PC1, where fermentations conditions that induced an inhibition of the fermentation profile (presence of 1-butanol, furfural and 5-HMF) and those where this phenotype was not affected, are grouped in two different groups.

In Figure 3.7B, it is possible to observe that the variables that are positively correlated with the flocculation profile are the biomass production and the extracellular ethanol. On the other hand, the extracellular* CO_2 (b), glucose (c), glycerol (e) and acetic acid (f), and the intracellular** glycerol (g) and trehalose (h) appeared correlated with the inhibition of cells flocculation. Between these, intracellular glycerol and extracellular acetic acid showed to have higher weight in the phenotype changes.

The negative effect of glucose, concerning to its higher accumulation in the fermentation medium, under 1-butanol, furfural and 5-HMF indicates that yeast cells are not able to metabolize glucose to pyruvate through the glycolytic pathway, suggesting that this via is being inhibited by the external conditions. So, under these conditions, a decrease of the energy supply (needed for yeast cells growth, adaptation and survival), can be occurring. The depletion of the energy supply can therefore be offset by the use of acetyl-CoA to produced acetic acid or to be used in the TCA cycle in aerobic respiration to produce both energy and electron carriers. The first mechanism underlies the higher production of acetic acid by CA11 (and S288c, mainly in the presence of 1% (v/v) 1-butanol), while the second will provide the generation of energy (ATP) and nicotinamide adenine dinucleotide (NAD^+), which triggers the glycolytic pathway [3].

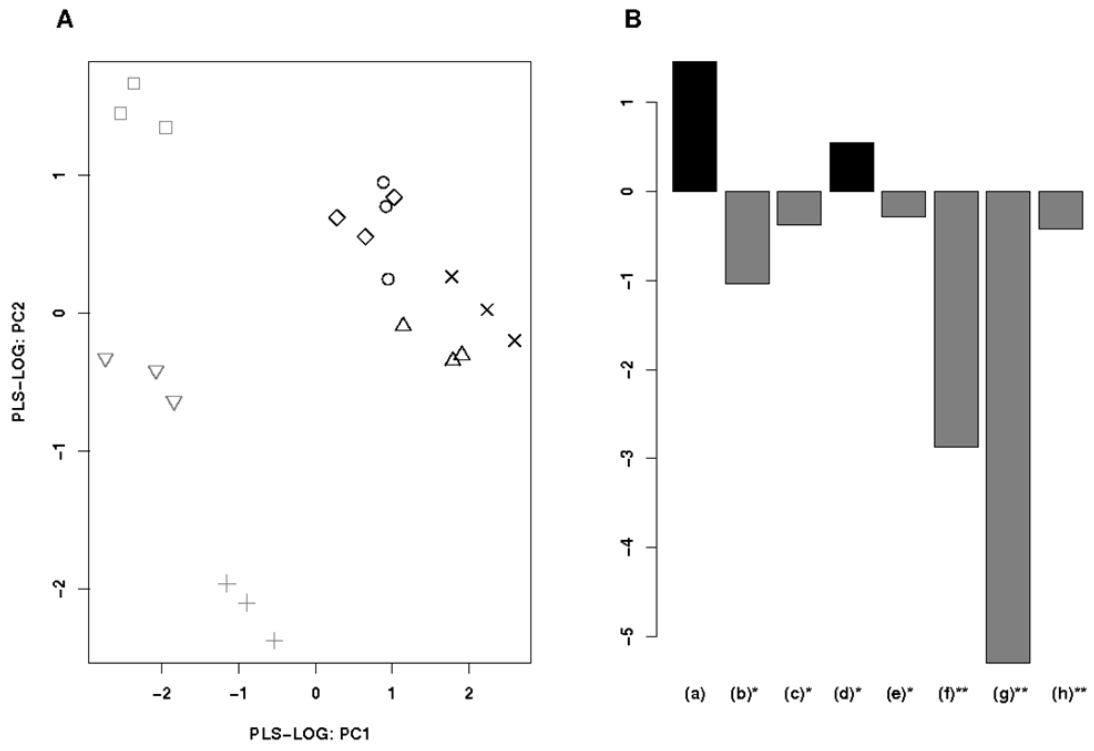


Figure 3.7 Partial least squares logistic regression model results: A – Scores of CA11 (A) strain under: (o) – control, (Δ) - 1% (v/v) ethanol, (+) - 1% (v/v) 1-butanol, (×) - 1% (v/v) isopropanol, (◊) - 1% (v/v) tert-Amyl alcohol, (∇) 0.2% (v/v) furfural, (□) 0.5% (v/v) 5-HMF; and B – Coefficients (variables) weight in flocculation phenotype: (a) – biomass, (b) extracellular CO_2; (c) extracellular glucose, (d) extracellular ethanol, (e) extracellular glycerol, (f) extracellular acetic acid, (g) intracellular glycerol; (h) intracellular trehalose.

So, the higher toxic and inhibitory levels introduced by the presence of 1-butanol, furfural and 5-HMF, lead to decreased energy availability and to increased production of acetic acid, which causes pH changes in the medium. According to these conditions, energy supply and pH modifications seem to directly affect the flocculation expression. On one hand, as flocculation is an energetic-dependent process that requires the presence of a residual external energy source [88], the maintenance of the flocculation profile depends on the energetic availability. On the other hand, unfavorable pH can lead to a reversible denaturation of the flocculins [359], the specific cell surface proteins, capable of binding directly to manose residues present on the wall of adjacent yeast cells, and mediate the flocculation mechanism [360].

The flocculation phenotype is also influenced by the production of trehalose under 1-butanol, furfural and 5-HMF. This metabolite is known to play an important function in cells protection, maintaining the structural integrity of the cytoplasm under stress conditions [132, 133], such as freeze, heat, dehydratation, ethanol, osmotic and oxidative stress [27, 361-363]. In this context, the higher production of trehalose seems to be an adaptation of yeasts cells against the external conditions induced.

Finally, the higher production of intracellular glycerol can be triggered by the excess of NADH in the medium [317], as well as by the existence of a high external osmotic pressure, playing an important role in the osmoregulation of yeast cells, balancing the osmotic stress of the yeast membranes [88]. The increased production of glycerol under these conditions impairs the flocculation expression, which is highly influenced by the carbon source metabolism [355, 364].

3.4 CONCLUSIONS

The understanding of the dynamics of fermentation processes and mechanisms underlying the comprehension of cells adaptation and responses to different environmental conditions is an issue that needs to be clarified. In the present work, three different *S. cerevisiae* strains, a laboratorial strain (S288c) and two industrial strains (CA11 and PE-2), were exposed to different oxidative and inhibitory conditions in batch fermentations. Physiological and phenotypic parameters were measured and analysed for each condition. Therefore, different chemometric tools, namely MPCA and PLS-LOG, were used to both characterize the physiological behavior of three different strains in batch fermentations under different fermentation conditions, and predict the flocculation feature of CA11, based on the physiological changes response of the yeast under those conditions.

The presence of stress molecules through the fermentation processes triggers different stress responses in yeasts strains, interfering with the enzymatic activity of the glycolytic pathway and biomass production.

CA11 and PE-2 were found to be the most robust strains, adapted to resist and survive to the harsh environmental conditions induced. Higher fermentation rates were found when using the two industrial strains, compared to S288c, even in the presence of the most stressful substances, as 1-butanol, furfural and 5-HMF. In addition to the decrease of the CO_2 production rates, the individual measurements of the physiological changes showed that these three molecules enhanced the

production of intracellular trehalose in PE-2 and S288c and intracellular glycerol by the three strains, also inducing the inhibition of the flocculation of CA11.

MPCA allowed characterizing the different behavior of S288c, which showed a decreased ability to resist to the unfavourable environmental conditions, comparing to CA11 and PE-2. In addition, the use of PLS-LOG allowed to classify and predict the flocculation profile changes in CA11, according to the fermentations conditions. It was found a correlation between the inhibition of the flocculation capacity and an overproduction of extracellular acetic acid and intracellular glycerol.

The use of both chemometric tools provided an increased information about *S. cerevisiae* strains physiological and phenotypic responses. These results encourage their application to better explore the metabolic changes occurring on yeast fermentation as a result of changes occurring/being introduced in the fermentation conditions, namely the presence of inhibitory compounds.

CHAPTER 4

CLASSIFICATION AND PREDICTION OF METABOLIC BEHAVIOR OF YEASTS UNDER INDUCED STRESS CONDITIONS

Saccharomyces cerevisiae can undergo different phenotypic, morphological and metabolic or physiological changes, according to environmental conditions, as a way of adapting to harsh growth conditions. The way that different yeast strains respond to the external conditions is generally different, depending on yeasts genomic information. Therefore, the understanding of cells behavior is of great importance for fermentations monitoring and to provide an external control of the process, inducing cells to grow in a particular state or to produce a specific end-product.

In the present work, multivariate methodologies including relevant principal component analysis (RPCA) and partial least squares logistic regression (PLS-LOG) were used for exploratory data analysis and classification. Using these methodologies, it was possible to classify and predict the metabolic behavior of two industrial *S. cerevisiae* strains, CA11 and PE-2, under different induced toxic and inhibitory stresses, namely 1-butanol, furfural and 5-hydroxymethyl-furfural (5-HMF).
The results showed that the cells adaptation response is dependent on the stress molecule used. According to the RPCA results, while CA11 fermentations were characterized by the production of ethanol, isovaleric acid and isoamyl acetate, PE-2 fermentations led to the production of more aromatic compounds, such as esters - phenylethyl acetate, ethyl hexanoate, ethyl octanoate and ethyl dodecanoate. The higher levels of aromatic compounds in PE-2 fermentations indicate that this strain is less susceptible to the stress effect of induced toxic and inhibitory conditions.

PLS-LOG models allowed the prediction of the metabolic behavior of both strains during the fermentations, showing a prediction capacity (R^2) higher than 0.90 ($p < 0.0001$) for all tested conditions. The presence of 1-butanol induced the production of esters ethyl acetate and isoamyl acetate (and its precursor, 3-methyl-1-butanol), as well as butyric acid. The production of the last one seems to indicate that the tested *S. cerevisiae* strains are capable of reducing 1-butanol to butyric acid, which suggests the feasibility of using these both strains in bio-butanol production systems.

Finally, it was found that these yeasts can metabolize furfural to produce furfuryl alcohol, and both furfural and 5-HMF induced the production and accumulation of fatty acids (such as hexanoic and octanoic acid), increasing the medium toxicity and inducing the inhibition of the fermentation process.

The information presented in this Chapter was adapted from:

Castro CC *et al.*, Classification and prediction of *Saccharomyces cerevisiae* strains behavior under induced stress conditions based on target extracellular metabolites profile (To be submitted).

4.1 INTRODUCTION

Fermentation process involves the production of important compounds, that can result from chemical components of musts or from the different mechanisms involved in the production process [365], and are responsible for the quality and character of the final product [4, 5]. During this process, yeasts cells can be subjected to different stress environments, namely osmotic pressure, oxidative and/or inhibitory stresses, which can be introduced either by the fermentation conditions used or by the high concentration of fermentation end-products [71]. Such harsh conditions represent significant obstacles to yeasts performance, leading to metabolic, molecular and transcriptomic changes as a response to the environmental conditions that influence yeasts growth [366]. In bio-ethanol and/or bio-butanol production processes, yeasts can be exposed to inhibitory molecules, such as furfural and 5-HMF, that can be introduced by the substrates pre-treatments [336], and have been associated to the inhibition of important enzymes, namely pyruvate and aldehyde dehydrogenases i in the glycolytic pathway, causing a reduction of the ATP synthesis and DNA damages in some cases [87, 95-98]. Also, the end product toxicity represents a limiting factor for the development of effective production processes, once it might affect cell membranes, cellular pH and nutrient transport processes [92-94], as well as, the inhibition of the membrane ATPase [120, 121], the loss of intracellular molecules, as proteins, RNA and ATP [110] and finally the obstruction of glucose uptake [120].

Thus, highly efficient industrial fermentation processes depend not only on the operational conditions and medium composition, but also require a suitable selection of yeast strains, able to withstand systems conditions [367].

Metabolome analysis involves the identification and quantification of metabolites from a single organism [160]. Compared to genomics, transcriptomics and proteomics, metabolomics is the most straightforward representation of the physiological status of a biological system, as metabolites are more closely linked to the phenotype of an organism [368, 369]. Different approaches have been published for monitoring metabolome changes within the fermentation systems comprising approaches as metabolite target analysis, metabolite profiling, metabolite fingerprinting, metabolite footprinting and flux analysis [208]. Metabolite footprinting or exo-metabolome is related to a completely non-invasive approach for extracellular metabolites measuring [185]. Several advantages have been identified on the measurement of extracellular instead of intracellular metabolites. These include the lower time-consuming for metabolites

extraction; the enhanced reproducibility on the quantification of metabolites; the higher concentration of metabolites that avoid some limitations on the simultaneous detection of large numbers of metabolites, and the improved information about different biochemical processes taking place in the fermentation media, such as the degradation of complex substrates [141]. Gas chromatography (GC) is one of the most used analytical separation techniques for extracellular metabolites identification and quantification [30, 32, 370], that can be coupled with different detectors, such as flame ionization detector (FID) or mass spectrometry (MS) for an accurate measurement of the involved metabolites.

Multivariate data analysis (MVDA) is a strategy that can be used to assign biological or chemical meaning to dynamic systems based on the metabolomics data exploration. This includes data exploration methodologies, such as relevant principal component analysis (RPCA) [281, 371], and classification and prediction algorithms, such as partial least squares for discriminant analysis (PLS-DA) [305, 372]. In metabolomics RPCA is used to maximize the variance between samples by detecting important variables or metabolites statistically relevant on each principal component, that most contribute to data-structure [281]. PLS-DA is used to maximize the co-variance between to independent data-sets, allowing to identify the relevant metabolites for a specific phenotype, as well as to set up for a mathematical relationship for predicting the values of one or more output metabolites [373, 374].

In the present work, both MVDA strategies were applied to exo-metabolomic data matrix, assembling extracellular metabolites acquired using HPLC, GC-FID and GC-MS analytical techniques during batch fermentations. These fermentations were performed using two industrial *Saccharomyces cerevisiae* strains, CA11 and PE-2, isolated from "*cachaça"* and bio-ethanol production in Brazil, respectively. Both strains can tolerate up to 17% (v/v) ethanol [88], and the effects on yeasts physiology induced by the presence of ethanol at high concentrations have been reported for many years [375]. Here, the two strains were exposed to three different stress conditions, induced by the presence of 1.0 % (v/v) butanol, 0.2 % (v/v) furfural and 0.5 % (v/v) 5-HMF. Control fermentations, in the absence of stress conditions were also performed.

The use of statistical tools to understand and predict the biochemical behavior of yeasts cells, under different fermentation conditions, based on the extracellular metabolites measurement was the main objective of this study. This methodology can be later used as a tool for selecting more robust strains for industrial processes application as well as to understand metabolic pathways

preferentially activated by yeasts, which culminates in a different metabolic profile through each fermentation conditions.

4.1.1 MATERIAL AND METHODS

4.1.2 Yeasts and fermentation process

Two industrial *S. cerevisiae* yeasts strains were used, CA11, a flocculent strain isolated from "*cachaça*" fermentation process [328], and PE-2, isolated from Brazilian bio-ethanol production [329]. Both strains were obtained from the microbiological collection of the IBB – Institute for Biotechnology and Bioengineering at the University of Minho.

Yeasts were incubated in YPD broth (YPD_b) medium (1.0 % (w/v) yeast extract, 2 % (w/v) bacto-peptone and 2 % (w/v) glucose - Sigma Aldrich - ref. Y1375) and after 12 h at 30 °C were aseptically collected and pitched at about 1.0×10^6 cells/mL (as described previously in Sub-section 3.2.1) to 50 mL of YPD_b in Erlenmeyer flasks (100 mL) fitted with perforated rubber stoppers enclosing glycerol-locks for anaerobic conditions preserving [324]. In this study, in order to evaluate the metabolic response of yeasts exposed to toxic molecules, the stress substances were added to the YPD_b. These include 1.0 % (v/v) 1-butanol (≥ 99.0%, Sigma Aldrich, USA), 0.2 % (v/v) furfural (98.0%, Sigma Aldrich, USA) and 0.5 % (v/v) 5-HMF (99.0%, Sigma Aldrich, USA). Control fermentations were also performed, using the two strains in the absence of stress.

Fermentations lasted for 24 h, in the presence and absence of stress conditions at 30 °C, and were monitored by weight loss (proportional to CO_2 production) [88]. Samples were taken for fermentation metabolites measurement using HPLC, Gas Chromatography - Flame Ionization Detector (GC-FID) and Gas Chromatography - Mass Spectrometry (GC-MS) analytical techniques.

4.1.3 Analytical methods

Different analytical technologies were used to characterize extracellular metabolic changes occurring within batch fermentations. Glucose, ethanol, glycerol and acetic acid concentrations were obtained by HPLC as described previously in Subsection 3.2.2. The major volatile fraction, including acetaldehyde, ethyl acetate, 1-butanol, 1-propanol, 2-methyl-1-butanol, 3-methyl-1-butanol

and 2-phenylethanol, was measured by GC-FID and the minor volatile fraction was measured by GC-MS and includes isoamyl acetate, ethyl hexanoate, ethyl octanoate, isobutyric acid, ethyl dodecanoate, butyric acid, furfuryl alcohol, isovaleric acid, 2-phenylethyl acetate, benzene ethanol, hexanoic acid, guaiacol, octanoic acid and decanoic acid.

4.1.4 Gas Chromatography – Flame Ionization Detector (GC-FID)

Samples taken during fermentations were filtered (2 μm) and 50 μL of the internal standard (4-nonanol) were added to 5 mL of sample. The volatile fraction semi-quantitative analysis was performed using a flame ionization detector (FID) associated with gas chromatography (GC). A Chrompack CP-9000 gas chromatograph equipped with a split/splitless injector and a flame ionization detector (FID) with a capillary column, coated with CP-Wax 52 CB (50 m × 0.25 mm i.d., 0.2 μm film thickness, Chrompack), was used. Injector and detector temperatures were both set to 250 °C. The oven temperature was held at 40 °C, for 5 min, then programmed to rise from 40 °C to 235 °C, at 3 °C/min, and then finally programmed from 235 °C to 255 °C, at 5 °C/min. The carrier gas was Helium 55 (Praxair) at 103 kPa and the split vent was set to 13 mL /min. Each 3 μL extract was injected in splitless mode (for 15 s). Quantification of volatiles, as 4-nonanol equivalents, was performed by comparing retention indexes with those of pure standard compounds using Varian MS Workstation version 6.6 [376].

4.1.5 Gas Chromatography – Mass Spectrometry (GC-MS)

Volatile fraction of samples taken during fermentation was firstly extracted with dichloromethane and subsequently analyzed by GC-MS, using a Varian 3400 chromatograph and an ion-trap mass spectrometer (Varian Saturn II). 1 μl sample was injected in a capillary column coated with CP-Wax 52 CB (50 m × 0.25 mm i.d., 0.2 μm film thickness, Chrompack). The temperature of the injector ranged from 20 °C to 250 °C, at 180 °C/min. The oven temperature was held at 60 °C, for 5 min, then programmed to rise from 60 °C to 250 °C, at 3°C/min, then held for 20 min at 250 °C and finally programmed to go from 250 °C to 255 °C at 1 °C/min. The carrier gas was Helium at 103 kPa. The detector was set to electronic impact mode (70 eV), with an acquisition range from m/z 29 to m/z 360, and an acquisition rate of 610 ms per scan. The identification of volatiles was

performed using the software Saturn version 5.2 (Varian), by comparing each mass spectra with those of pure standard compounds at each retention time. All the metabolites were quantified as 4-nonanol equivalents [376].

4.1.6 Exploratory metabolomics data analysis using relevant principal component analysis

Relevant principal component analysis (RPCA) is a very common methodology for important effects detection in data, by reducing the dimensionality of a dataset. These effects can be detected and explored by samples position on the samples space (scores analysis), by variables correspondences inside each principal component (loadings analysis) and also by variance contribution (eigen values) [377]. RPCA allows the identification of statistically significant loadings on each relevant principal component providing a better interpretation on how the different variables affect the metabolomics data variability [281].

Metabolites concentrations of samples corresponding to different fermentation conditions were organized in a data matrix $\boldsymbol{X}$ ($n \times m$), where n corresponds to samples and m to chemical compounds (variables), and RPCA was applied. A large part of the structure and variability present in this metabolic original data set can therefore be explained by the resulting reduced and decorrelated principal components (PC), according to the chemical information contained in fermentation samples, allowing to explore its variability and different chemical behaviors during fermentations.

4.1.7 Fermentations classification using partial least squares regression

Partial least squares regression (PLSR) is a prognostic two-block regression method based on estimated latent variables and applies to the synchronized analysis of two data sets of the same objects [373]. Partial least squares logistic regression (PLS-LOG) is an extension of the PLSR and enables the classification of the multivariate space directions by fitting a regular PLS model between the $\boldsymbol{X}$ matrix and an artificial $\boldsymbol{Y}$ matrix (that encodes class memberships by a set of variables) providing discriminant directions with well separated observations, according to class membership. In PLS-LOG, the logistical or logit regression determines the impact of each fermentation condition to predict the membership of the extracellular metabolic changes,

determining the probability of the presence of each stress conditions in the fermentation medium, according to the footprinting metabolome. In this work, PLS-LOG was used to predict metabolic changes occurring within fermentations, according to the external conditions using CA11 and PE-2 yeast strains.

4.2 RESULTS AND DISCUSSION

4.2.1 Classification of yeasts fermentation metabolism under different stress conditions

Major and minor extracellular metabolites were measured within YPD_b batch fermentations, using CA11 and PE-2, both industrial *S. cerevisiae* strains, under control fermentations and in the presence of 1% (v/v) 1-butanol, 0.2% (v/v) furfural and 0.5% (v/v) 5-HMF. The non-volatile fraction of the extracellular medium was quantified by HPLC and includes glucose, ethanol, glycerol and acetic acid, whereas the volatile fraction was measured by GC-FID and GC-MS and includes higher alcohols, sort- and medium-chain fatty acids (MCFA), ethyl and acetate esters, and guaiacol.

Relevant principal component analysis (RPCA) was applied to a $X(n,m)$ matrix, including the m variables - extracellular metabolites measured - within the n samples - taken throughout the different fermentations time-course. In metabolomics analysis, 5000-fold differences in concentrations for different metabolites can be found [270]. As these differences are not proportional to the biological relevance of these metabolites, it was important to scale the data set, when performing the RPCA analysis. Scaling is an approach that divides each variable by a factor, adjusting the differences in fold differences between the metabolites by converting the data into concentrations relative to the scaling factor. In this work, the measure factor used lead to the variance of each variable within all samples, as it is showed in Equation 4.1, where x_i corresponds to each variable value, x_m is the mean value of each variable and sd_i, the internal stardard for each variable. An important aspect of scaling is that it results in the inflation of small values, that can have an undesirable side effect as the influence of the measurement error, that is usually large for small values, is increased as well [270].

$$x = \frac{x_i - \overline{x_m}}{sd_i}$$

Equation 4.1

Figure 4.1 presents the RPCA scores and loadings of CA11 and PE-2 extracellular metabolic profiles. Scores distribution through the two principal components (PC), PC1 and PC2, provided an enlarged view of the relevant metabolic variability, as these decompositions showed the higher discriminant capacity. The scores plot from RPCA analysis of CA11 and PE-2 totalized 47.7 % of the metabolic information variance with discriminant power of 32.3 % PC1 and 13.4% PC2.

According to Figure 4.1, samples were distributed within PC1 as a function of fermentations time-course (from T0 to T5) using both yeast strains, in control and in the presence of each induced condition, underlining the metabolic changes occurring during the fermentation progress. Also, it is possible to observe that the running of the fermentation processes of CA11 or PE-2, based on the extracellular metabolites measurements differed from each other. The RPCA loadings (grey arrows and numbers in Figure 4.1) discriminating samples throughout the first principal component (PC1), showed that in the beginning of the fermentation processes there was a high concentration of glucose (1) in the medium which was being consumed to produce distinct metabolites along the fermentation progress. In the absence of stress (black symbols in the figure), it was observed the production of 2-phenylethanol (15) and acetaldehyde (3), mainly by PE-2. From the exponential phase ahead (T2 to T5), CA11 fermentations lead to the production of ethanol (11), isovaleric acid (7) and isoamyl acetate (19) while PE-2 produced mainly aromatic esters as phenylethyl acetate (20) and the ethyl esters – ethyl hexanoate (C6:0) (21), ethyl octanoate (C8:0) (22) and ethyl dodecanoate (C12:0) (23) – predominantly in the control fermentations (black symbols), and also glycerol (2), acetic acid (4), 2-methyl-1-butanol (13), 3-methyl-1-butanol (14), 1-propanol (12), decanoic acid (10), furfuryl alcohol (16) and guaiacol (24) in the presence of furfural (green symbols) and 5-HMF (red symbols).

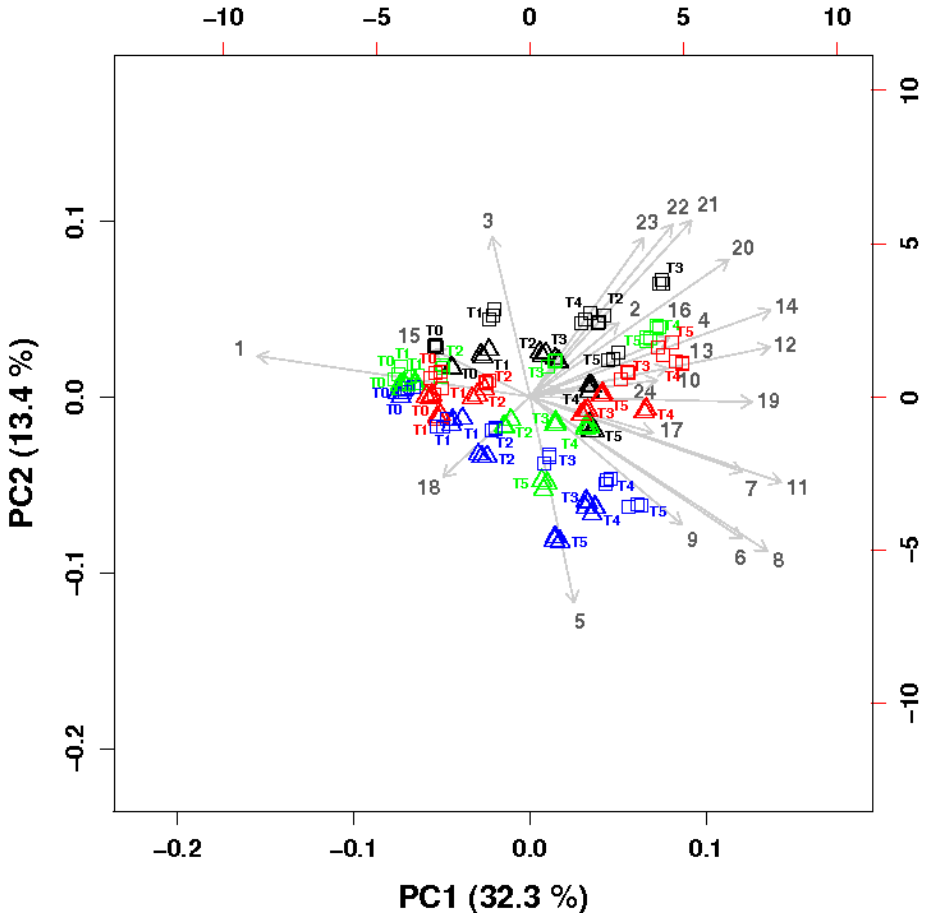


Figure 4.1 Scores and loadings resulting from relevant principal component analysis applied to extracellular matrix using CA11 ▲ and PE-2 ■ yeast strains. Scores correspond to samples took within the fermentations time-course (T0 to T5) under: control (black symbols), butanol (blue symbols), furfural (green symbols), 5-HMF (red symbols); and loadings represented by the arrows correspond to the measure variables: 1 – glucose; 2 – glycerol; 3 – acetaldehyde; 4 – acetic acid; 5 – butyric acid; 6 – isobutyric acid; 7 – isovaleric acid; 8 – hexanoic acid; 9 - octanoic acid; 10 - decanoic acid; 11 – ethanol; 12 – 1-propanol; 13 - 2-methyl-1-butanol; 14 - 3-methyl-1-butanol; 15 - 2-phenylethanol; 16 – furfuryl alcohol; 17 - benzyl alcohol; 18 – ethyl acetate; 19 – 2- phenylethyl acetate; 20 - isoamyl acetate; 21 - ethyl hexanoate; 22 - ethyl octanoate; 23 - ethyl dodecanoate; 24 – guaiacol.

Samples distribution throughout the PC2 discriminates fermentations performed under the different induced conditions, namely the control (black symbols), in the presence of 1% (v/v) 1-butanol (blue symbols), 0.2% (v/v) furfural (green symbols) and 0.5% (v/v) 5-HMF (red symbols), emphasizing the metabolic differences found under each fermentation condition using both strains. Samples dispersion along the PC2 also highlighted the higher deviation of the metabolic behavior of both yeasts in the presence of 1% (v/v) 1-butanol (blue), comparing to the control (black

symbols). In the presence of 1-butanol, the metabolic changes undergone by both strains were similar, and a higher predominance of ethyl acetate (18) was observed in the beginning of the exponential phase, and until the end of the fermentation processes, the production of acids, such as butyric acid (5), isobutyric acid (6), hexanoic acid (9) and isovaleric acid (8) was enhanced.

PE-2 is an isolate from bio-ethanol distilleries, and is one of the most successful strains used by several industries in Brazil, generating about 10% of the world bio-ethanol supply [378]. In VHG fermentations, this strain exhibited an increased fermentation performance, able to produce an ethanol titre of 19.2% (v/v), whereas with CA11 the slowest fermentation amongst different industrial strains was observed together with an incomplete fermentation [324]. CA11 is an isolate from "*cachaça"* fermentation processes and comparing to PE-2, it was found to be more adapted to produce "*cachaça"* as it presented higher amounts of 1,3-butanediol and lower amounts of acetaldehyde and it does not generate propanol [379]. RPCA loadings presented in Figure 4.1 are in agreement with these results, as CA11 seems to be characterized by the production of ethyl acetate (18), mainly, the presence of 1-butanol, and PE-2, in general, by the production of acetaldehyde (3), 1-propanol (12) and acetic acid (4).

Other studies investigated the use of both strains for wines production, namely raspberry wines [380], where both showed a decreased ability to produce volatile aromatic compounds, as ethyl esters, comparing to other strains typically used in wine, "*cachaça"*, and bio-ethanol production. The production of these metabolites by yeasts during fermentations, significantly contributes to the "fruity" flavors of wines [3]. Although, according to the fermentation conditions used in this study, PE-2 showed to be more adapted to produce phenylethyl acetate (20) and other ethyl esters – C6:0 (21), C8:0 (22) and C12:0 (23), while CA11 is characterized by higher production of ethyl acetate, mainly in the presence of 1-butanol. According to Verstrepen *et al.* (2003) [381], esters are synthesized from a fusel alcohol and an active fatty acid (acyl-CoA or acetyl-CoA). Ester synthase is the enzyme that catalyses this reaction and the ester production rate is determined by the concentration of available substrates and the total enzymatic activity. So, any factor, such as the introduction of toxic and inhibitory conditions in the medium as in this study, affecting yeasts metabolism and/or ester synthase gene activity (*e.g. ATF1*), affects the ester synthesis rate.

Under the fermentation conditions induced, PE-2 was also more able to produce higher alcohols, such as 2-phenylethanol (15), 1-propanol (12), amyl alcohols – 2-methyl-1butanol (13) and 3-methyl-1-butanol (14) - and furfuryl alcohol (16) in the presence of furfural. The higher alcohols

production during fermentation could result from keto acid produced either catabolically, involving degradation of an amino acid via the so-called *Ehrlich pathway* [4], or anabolically via the biosynthetic route from the carbon source [382].

CA11 also enhanced the production of acids, especially in the presence of 1-butanol. In wines, the presence of high concentrations of acids may negatively influence their qualities because of the aroma descriptors that include “cheese” and “sweaty” from hexanoic acid (C6:0) (8) and “rancid” and “harsh” from octanoic acid (C8:0) (9) [3]. In the presence of 1-butanol, the production of butyric and isobutyric acids was also enhanced. The presence of these acids could lead to the inhibition and arrest of fermentation, which ultimately blocked the complete transformation of sugars present in the must [383].

In this context, differences found on the higher alcohols and aromatic esters production between PE-2 and CA11 suggest that CA11 is more susceptible to the induced stresses, as it showed decreased prevalence of those metabolites, and the introduction of 1-butanol induced a wide different metabolic behavior for both strains.

One way of understanding the effect of the presence of each stress molecule on the metabolic profile of each *S. cerevisiae* strain is the use of partial least squares logistic regression (PLS-LOG).

4.2.2 Metabolic footprinting prediction using partial least squares logistic regression

Partial least squares logistic regression (PLS-LOG) methodology was used to understand the influence of each yeast strain or fermentation condition in the weight of each variable linked to fermentations behavior. Generally, this methodology imposes a correlation between two independent data matrices, which in this case concern to the $\boldsymbol{X}$ matrix, the metabolic footprinting information (samples *vs* metabolites), already used in RPCA in Section 4.2.1, and a $\boldsymbol{Y}$ matrix, with the quality information about the presence or absence of each stress in the fermentation medium (samples *vs* variables) [305]. *Samples* correspond to the sampling time of each fermentation; *metabolites* consist in the extracellular metabolites measured using the different analytical methodologies; and *variables* correspond to the different stress conditions used in fermentations: control, 1 % (v/v) 1-butanol, 0.2 % (v/v) furfural and 0.5 % (v/v) 5-HMF. In Table 4.1 it is presented the summary of the PLS-LOG models prediction, which is linked to the capacity of predicting the metabolic changes, according to the external conditions imposed using the two different *S.*

cerevisiae strains. According to Table 4.1, high correlation indices were found for the different fermentation conditions tested. This suggests that the metabolic behavior of each strain was directly correlated to the impact of the induced fermentation conditions.

Table 4.1 Prediction capacity of PLS-LOG models of the metabolic behavior based on the footprinting characterization (*** p-value < 0.0001)

Fermentation condition	*Prediction capacity (R^2)*
Control	*0.921 ****
1% (v/v) 1-butanol	*0.952 ****
0.2% (v/v) furfural	*0.957 ****
0.5% (v/v) 5-HMF	*0.901 ****

The way how the toxic and inhibitory environment influences the metabolic behavior can be understood by the scores and coefficients of the PLS-LOG models. In Figure 4.2, the scores (Figure 4.2 A) and coefficients (Figure 4.2 B) allow the understanding of the effect of each induced stress condition in the production (positive correlation) or the consumption (negative correlation) of the extracellular metabolites measured.

The scores plots resulting from PLS-LOG models, allow to observe that samples are clustered in the same distint group, in Figure 4.2A, which can explain the higher prediction capacity observed in Table 4.1. The impact of the presence of each toxic and inhibitory condition in the medium, namely, 1-butanol, furfural, and 5-HMF (Figure 4.2), in the metabolic behavior of CA11 and PE-2, was explored by the PLS-LOG models coefficients, presented in Figure 4.2B.

According to PLS-LOG coefficients of the model predicting the metabolic behavior of both strains in the presence of 1-butanol, it is possible to observe that the production of ethyl acetate (18), butyric acid (5), 3-methyl-1-butanol (14) and isoamyls acetate (19) was enhanced (positive weight) while the opposite occurred in the production of phenylethyl acetate (20), furfuryl alcohol (16) (negative weight). Acetate esters production by *S. cerevisiae* yeasts during fermentation is dependent of three enzymes, namely acetyltransferase, ethanol acetyltransferase and isoamyls alcohol acetyltransferase [3] or an ester synthase that produces acetate esters from ethanol and the respective acids [384]. Nordström (1961) [317] demonstrated that ethyl acetate production can be produced as an energy-requirement metabolic process. The formation or consumption of acetyl-

CoA is known to affect ethyl acetate synthesis [385]. So, according to these possibilities, it was suggested that in response to the increased toxicity introduced by 1-butanol, both strains use the acetyl-CoA factor, which can be produced by the fatty acids metabolism [343], for the production of these acetate esters. This can also explain the increased effect on the production of 3-methyl-1-butanol (isoamyls acohols), as it is the isoamyl acetate precursor [386]. Finally, it has been previously shown that butyric acid could be produced either by *Clostridium saccharoperbutylacetonicum N1-4* fermentation [387] or by the biotransformation of butanol by *Acetobacter aceti [388],* which suggests that *S. cerevisiae* CA11 and PE-2 might have this ability which reveals that it could be interesting to explore both strains in the production of bio-butanol production.

According to the coefficients obtained from the PLS-LOG model coupled with the metabolic information under the presence of 0.2% (v/v) furfural (Figure 4.2 B2), it is possible to observe that this condition affects positively the production of furfuryl alcohol (16) and hexanoic acid (8) and negatively the production of guaiacol (24) and octanoic acid (9). Furfuryl alcohol is the result of the furfural reduction, which was already correlated with the vital need for overcoming the toxic effects of furfural, a reaction that strongly demands for NADH and so, it can result in insufficient ATP generation to sustain cells growth [389]. The presence of hexanoic acid indicated that this is not being converted to the respective ester [314, 365].

In 5-HMF fermentations using both industrial strains (Figure 4.2 B3), it can be observed that the production of 2-phenylethyl acetate (20) and fatty acids – hexanoic (8) and octanoic (9) acids – was enhanced. 5-HMF is known to inhibit some specific enzymes (alcohol dehydrogenase, pyruvate dehydrogenase and aldehydes dehydrogenase) as well as the glycolysis (either enzyme and/or cofactors), and so, it interferes with the energetic balance for yeasts growth. Under these circumstances, as a stress response adaptation, yeasts synthesized acetate esters, such as 2-phenylethyl acetate and accumulated fatty acids. The increased accumulation of fatty acids is known to induces fermentation inhibition [390], by providing an enhanced toxicity in the medium, linked to a decreased membrane integrity [391, 392].

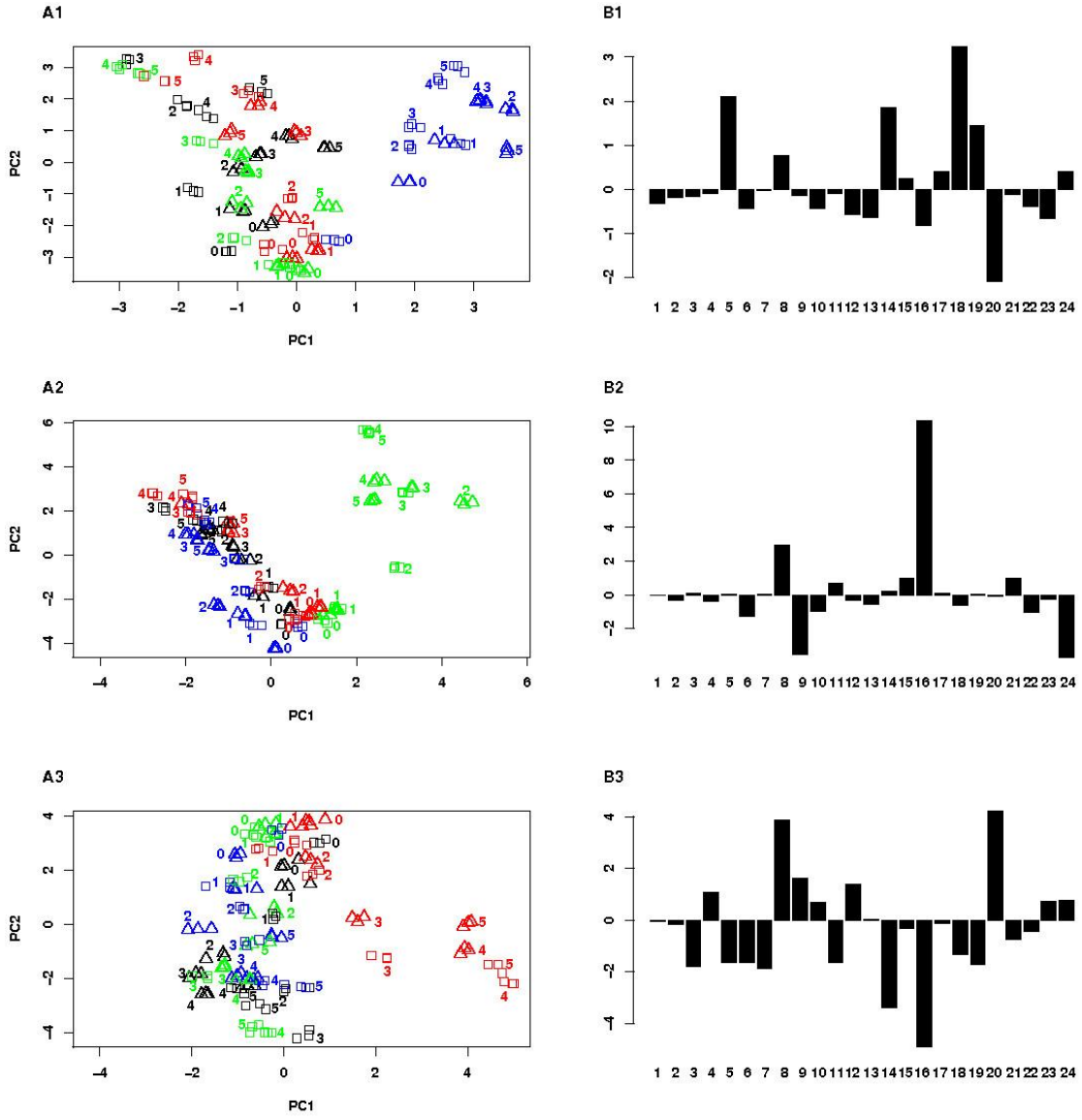


Figure 4.2 Partial least squares logistic regression models: A – scores; B – coefficients; for predicting the metabolic behavior of yeasts under: 1 – 1% (v/v) 1-butanol; 2 – 0.2% (v/v) furfural; and 3 – 0.5% (v/v) 5-HMF. Coefficients bars in B correspond to: 1 – glucose; 2 – glycerol; 3 – acetaldehyde; 4 – acetic acid; 5 – butyric acid; 6 – isobutyric acid; 7 – isovaleric acid; 8 – hexanoic acid; 9 - octanoic acid; 10 - decanoic acid; 11 – ethanol; 12 – 1-propanol; 13 - 2-methyl-1-butanol; 14 - 3-methyl-1-butanol; 15 - 2-phenylethanol; 16 – furfuryl alcohol; 17 - benzyl alcohol; 18 – ethyl acetate; 19 – 2- phenylethyl acetate; 20 - isoamyl acetate; 21 - ethyl hexanoate; 22 - ethyl octanoate; 23 - ethyl dodecanoate; 24 – guaiacol.

4.3 CONCLUSIONS

The way of how cells adapt and respond to different environmental conditions in the fermentation process is a crucial step for the fermentation process manipulation and monitoring. In the present work, the footprinting metabolic behavior of two industrial strains, CA11 and PE-2 in the presence of 1-butanol, furfural and 5-HMF, was evaluated and classified using multivariate tools, namely, relevant principal component analysis and partial least squares logistic regression.

Yeast strains presented different metabolic information, which can be accurately predicted by applying the logistic regression, and while CA11 guided the fermentation process for the production of carboxylic acids in the medium, PE-2 promoted mainly the production of aromatic compounds, including higher alcohols and esters. The increased capacity of PE-2 for producing more aromatic compounds under the same conditions of CA11 suggested a higher robustness of PE-2.

In the presence of 1-butanol, the metabolic adaptation showed to be very distinct when compared to the response to the furan derivates. However, the capacity of both strains to produce butyric acid from 1-butanol, encourages both strains to be used and possibly engineered for bio-butanol production.

The understanding of how cells behave under specific conditions in terms of phenotypic, morphological and metabolic changes or physiological parameters is crucial for fermentation process monitoring and to a possible external control of the process by inducing cells to grow in a particular rate, as well as to produce a specific end-product of fermentation.

CHAPTER 5

EVALUATION OF *SACCHAROMYCES CEREVISIAE* OXIDATIVE RESPONSE USING ANALYTICAL TOOLS

This Chapter is focused on the evaluation of the impact of *Saccharomyces cerevisiae* metabolism in the profile of compounds with antioxidant capacity in a synthetic wine during fermentation. Chemometrics tools, namely relevant principal component analysis (RPCA) and unfolded partial least squares (U-PLS) were applied to this data matrix, in order to increase the knowledge about metabolites synthesized as response to the perturbation induced. So, a bioanalytical pipeline, which allows for biological systems fingerprinting and sample classification by combining electrochemical features with biochemical background is proposed. To achieve this, alcoholic fermentations of a minimal medium supplemented with phenolic acids, were evaluated daily during 11 days, for electrochemical profile, phenolic acids and the volatile fermentation fraction, using cyclic voltammetry, HPLC-DAD and HS-SPME/GC-MS (target and non-target approaches), respectively. It was found that acetic acid, 2-phenylethanol and isoamyl acetate are compounds with a significative contribution for samples metabolic variability and the electrochemical features demonstrated redox-potential changes throughout the alcoholic fermentations, showing at the end, a similar pattern to normal wines. Moreover, *Saccharomyces cerevisiae* had the capacity of producing chlorogenic acid in the supplemented medium fermentation from simple precursors present in the minimal medium.

The information presented in this Chapter was published as:

Castro CC, Gunning C, Oliveira CM, Couto JA, Teixeira JA, Martins RC, Silva Ferreira AC (2012) *Saccharomyces cerevisiae* Oxidative Response Evaluation by Cyclic Voltammetry and Gas Chromatography–Mass Spectrometry, *Journal of Agricultural and Food Chemistry,* 60 (29), 7252–7261.

5.1 INTRODUCTION

Fermentation is an often used method for preservation. Alcoholic fermentation concerns to the conversion of grape juice sugars, namely glucose and fructose, to ethanol and carbon dioxide by wine yeasts [3]. Thus, a product with a high ethanol content immediately restricts the possibility of deterioration by other microorganisms, as well as lower pH and changes in redox potential [393]. In this context, alcoholic fermentation contributes to the resistance of the beverage to microbial spoilage [394]. Wine oxidation mechanism on the other hand, represents different challenges for wine-makers. The resistance of wines to oxidation is influenced by their composition, their exposure to oxygen and the antioxidants-containing concentrations [168]. Phenolic compounds are antioxidant compounds, known to be related to the wine oxidation capacity [168]. These are currently of great interest in the wine industry, as a result of their health benefits and radical scavenging properties [395].

The understanding of yeasts behavior within fermentations under oxidative conditions can be assessed using different analytical detectors, such as, electrochemistry and mass spectrometry methods. Combining the information obtained from both detectors seems to be an advantage as it will be possible to correlate metabolic, chemical and electrochemical responses providing tools to better understand the complexity of the overall system.

Cyclic voltammetry can be used for electroactive activity scanning. A voltammogram provides information about the type of antioxidants present, as well as quantitative information about the likelihood of oxidation of particular substances [166] enabling the understanding of how these compounds are metabolized and changed during fermentation.

Key volatile metabolites can be identified and quantified using a Head-Space/Solid-Phase-Micro-Extraction/Gas-Chromatography-Mass-Spectrometry (HS-SPME/GC-MS) technique by directly integrating chromatogram peaks area (namely target mode), or by a non-targeted methodology, an unbiased approach, towards understanding the overall biological system [32, 396]. The non-targeted approach allows a faster metabolic overview, avoiding the time-consuming need for any prior assignment of chemical classification of the molecular structure for hundreds of datasets [17].

Studying *S. cerevisiae* metabolism during fermentation is effectively a process of taking a series of snapshots of the metabolism at different stages of a very dynamic process [210], where some fermentation pathways are being activated, while other are being down-regulated.

In this study, the question of how yeast metabolism affects the concentration of phenolic compounds was addressed and monitored indirectly and directly, using cyclic voltammetry and HPLC/DAD/MS detection, respectively. In addition, the effect of the phenolic acids on the aromatic profile of the beverage was explored using HS-SPME/GC-MS analysis (Figure 5.1), conducted using both non-target and target approaches. The information gathered by the different detectors used, was then categorized using statistical methods and structured in terms of similarities and differences between samples, allowing the biological information interpretation and the oxidation resistance interpretation.

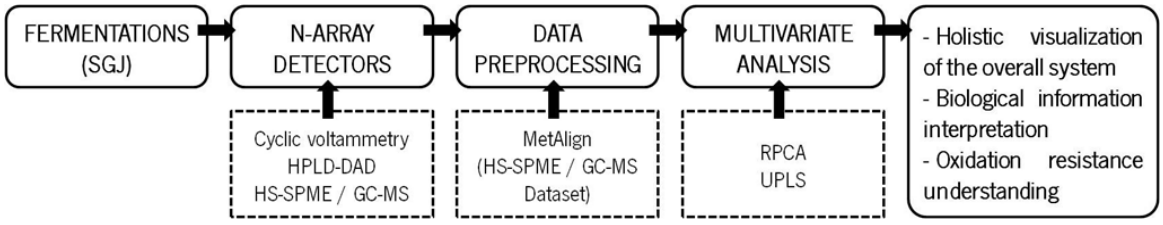


Figure 5.1 Schematic representation of the pipeline presented in this study.

5.2 MATERIAL AND METHODS

5.2.1 Minimal medium preparation protocol

Figure 5.2 presents the minimal medium preparation protocol for all fermentations studied. For this purpose, 3 liters of a synthetic grape juice (SGJ) was prepared and adjusted as follows. *(i)* addition of yeast (Control) and *(ii)* and *(iii)* addition of yeast and phenolic acids (Replica 1 and Replica 2).

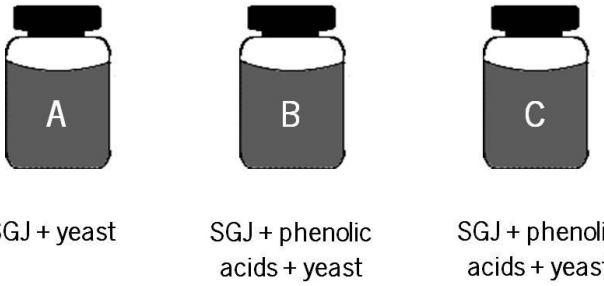


Figure 5.2 Methodology used for fermentation media preparation: SGJ - Synthetic grape juice; A – Control: fermentation in the absence of phenolic acids (900 mL); B - Replica 1 (900 mL); and C - Replica 2 (900 mL): fermentations with phenolic acids supplemented medium.

5.2.2 Synthetic grape juice

A synthetic grape juice (SGJ), a less complex and more reproducible matrix, was made according to the procedure described by Ciani and Ferraro [397]. SGJ was produced by combination of three aqueous solutions, prepared and sterilized separately. *Solution 1*: 110 g/L D-glucose (99.5%, Sigma), 100 g/L D-fructose (99.0%, Sigma), 10 mg/L ergosterol (95%, Sigma-Aldrich), 0.1% (v/v) Tween 80 (for synthesis, Merck); *Solution 2*: 6.0 g/L L-(+) tartaric acid (99.5%, Merck), 3.0 g/L L-(-) malic acid (99.5%, Fluka), 0.5 g/L citric acid (99.5%, Sigma-Aldrich); *Solution 3*: 1.7 g/L Yeast Nitrogen Base (Bacto Difco), 2.0 g/L Casamino acids (Bacto Difco), 0.2 g/L anhydrous calcium chloride (99.5%, Merck), 0.8 g/L L-arginine hydrochloride (98%, Sigma), 1.0 g/L L-(-) proline (99%, Sigma), 0.1 g/L L-(-) tryptophan (98%, Sigma), 0.1 g/L phenylalanine (98%, Sigma) and 0.1 g/L L-tyrosine (98%, Sigma).

Solutions 2 and 3 were adjusted to pH 3.5 with NaOH (2M) and HCl (1M), before sterilizing. SGJ was then added to Control (Figure 5.2A), Replica 1 (Figure 5.2B) and Replica 2 (Figure 5.2C).

5.2.3 Phenolic acids addition

The following phenolic acids were added to Replica 1 and Replica 2 (Figure 5.2B and C) before addition of the yeast: hydroxybenzoic acids - gallic acid monohydrate (99%, Sigma-Aldrich), protocatechuic acid (99%, Sigma- Aldrich), vanillic acid (97%, Sigma-Aldrich) - and hydroxycinnamic

acids - caffeic acid (99%, Sigma-Aldrich), *para*-coumaric acid (98%, Sigma-Aldrich) and ferulic acid (99%, Sigma-Aldrich). The final concentration of each phenolic acid was 15 mg/L each.

5.2.4 Yeast addition

Saccharomyces cerevisiae PYCC 4653 strain (Portuguese Yeast Culture Collection) was used for alcoholic fermentation in the Control, Replica 1 and Replica 2 (Figure 5.2A, B and C). Yeast cultures were previously grown in Yeast extract-Malt extract (YM) medium for a minimum of two days at 30 °C in an orbital incubator, then collected after centrifugation (9000 rpm, 15 minutes, 25 °C), and re-suspended in Ringer solution before addition to the fermentation medium.

The yeasts cells were pitched at about 1.0×10^{6} colony-forming unit (CFU)/mL into the culture medium, adjusted by microscopic enumeration with a cell-counting hematocytometer (Neubauer chamber; Merck) to start the fermentation. Fermentations of control and supplemented medium were carried out at 18 °C in 1 litre sterile Schott flasks equipped with cotton-plugs after filling with 900 mL of fermentation medium (Figure 5.2). The low temperature was chosen to simulate the white wine fermentation [398]. Alcoholic fermentations were monitored for 11 days, until the viable counts of yeast cells had fallen below 1.0×10^{6} CFU/mL.

Samples were taken and analyzed daily, for 11 days, by HS-SPME / GC-MS for metabolic changes, by cyclic voltammetry for electrochemical changes, and by HPLC - DAD for quantification of specific antioxidants. Residual sugars and ethanol were analyzed by HPLC - RI and used as fermentation control parameters.

5.2.5 Cyclic voltammetry analysis

Experiments were performed using a potentiostat (microAutolab Type III with an Autolab Faraday Cage) and voltammograms were obtained with a scan rate of 100 mV with a set potential of 2.4 mV, between -0.2 V to 1.2 V. The working electrode was a 3 mm Glassy Carbon disk in combination with a Metrohm tipholder, cleaned by polishing with 3 μm alumina powder between acquisitions. A saturated calomel electrode was used as a reference electrode in conjunction with a

platinum counter electrode. Each acquisition required 12 mL of undiluted sample. Voltammograms were treated by General Purpose Electrochemical System (GPES) 4.9 software.

5.2.6 HPLC-DAD analysis – Quantification of phenolic acids and *para*-chlorogenic acid

A Beckman model 126 quaternary solvent pump system, equipped with System 32 Karat 5.0 software and a 168 rapid scanning, UV-visible photodiode array detector, was used. The absorption spectra were recorded between 270 and 550 nm. Stationary Phase: Chromolith Performance RP-18e (100 x 4.6 mm) (Merck, Germany). Mobile Phase: Solvent A: acetonitrile/water (5:95 v/v) (Merck pure grade and pure water) with 0.2% TFA (Sigma-Aldrich, Germany). Solvent B: acetonitrile (100%) (Merck pure grade) with 0.2% of TFA; flow rate = 3 mL/minute. The following gradient was employed: 0-2 minutes (0% B); 2-6 minutes (10% B); 6-10 minutes (20% B); 10-12 minutes (0% B); post time of 3 minutes. Each run took 15 minutes to complete, and all relevant compounds had eluted by 8 minutes. Hydroxybenzoic and hydroxycinnamic acids were detected at 280 and 320 nm, respectively. Along with these six phenolic acids (added to the medium preparation protocol) para-chlorogenic acid was also detected. Retention times were as follows: gallic acid (0.8 min); protocatechuic acid (1.2 min); vanillic acid (3.0 min); caffeic acid (3.5 min); *para*-coumaric acid (5.3 min); ferulic acid (6.2 min). Identification: Phenolics were identified by comparison with pure, authentic, commercially available standards' retention times and UV-visible photodiode array spectra.

5.2.7 HPLC - RI analysis - Quantification of residual sugars and ethanol

A Beckman Model 126 quaternary solvent pump system equipped with an autosampler and a RI detector was employed. The acquisition was done using 32 Karat 5.0 software. Stationary-phase: Aminex hpx-87H (300 x 7.8mm) from Bio-Rad. Mobile-phase: sulphuric acid (2.5mM); flow rate = 0.6 mL/minute.

5.2.8 LC-MS analysis – Detection of chlorogenic acid

The chromatographic system was consisted of a Prostar 210 LC pump (Varian, CA, USA) coupled with a Varian 1200 triple quadrupole mass spectrometer (Varian, CA, USA) with electrospray ionization. A 5 μm C18 column (4.6mmx100mm, Merck) was used for the separation at a flow rate of 0.4 mL/min. The separation was performed by gradient elution (eluent A, water with 0.1% formic acid; eluent B, 100% methanol) in 33 minutes. For MS/MS fragmentation, argon atoms were used (pressure 1.20 mtorr; collision energy of 15 V). Data were acquired by Varian LC-MS 1200L Workstation. LC and MS-MS were established by ESI-LC/MS under negative ion mode. Structural identification of chlorogenic acid (MW = 354; [M-H] = 353(191); RT = 15.82 min] was performed by comparison of the retention time and mass spectra of the phenolic standard with the supplemented medium [399].

5.2.9 Gas-Chromatography analysis

Head Space - Solid Phase Micro Extraction (HS - SPME)

Volatile compound analyses using the analytical SPME technique were performed according to Ferreira and Guedes [400]. The used fiber was coated with a divinylbenzene / carboxen / polydimethylsiloxane (DVB / CAR / PDMS), 50/30 μm (Supelco, Bellefonte, Pa., USA). For each SPME analysis, 5 mL of sample was placed in a vial with 20 mL capacity together with a small stirring magnet, and stirred at 1300 rpm, spiked with an internal standard (IS) (20 μL of methanolic solution of 3-octanol, 46.3 mg/L) while immersed in a water-bath at 36 °C. The SPME needle then pierced the septum and the fiber was extended through the needle to expose the stationary phase with the head-space of the sample for 5 minutes. Afterwards, it was removed from the vial and inserted into the injection port of the gas chromatograph for 5 minutes. The extracted chemicals were thermally desorbed, at 220 °C, and transferred directly to the analytical column. Fibers were cleaned before each micro-extraction process to prevent contamination by inserting the fiber in the auxiliary injection port at 220 °C for 30 minutes.

Mass spectrometry analysis

Samples were analyzed using a Varian CP-3800 gas chromatograph (Walnut Creek, CA, USA) equipped with a Varian Saturn 2000 mass selective detector and Saturn GC/MS workstation software version 5.51.14. The column used was a STABILWAX-DA (60 m x 0.25 mm, 0.25 μm) fused silica (Restek, Bellefonte, PA, USA). The injector port was heated to 220 °C. The oven temperature was 40 °C (for 1 minute), then increased at 2 °C / minute to 220 °C. The carrier gas was Helium (Gasin, Portugal), at 1 mL/minute at constant flow. All mass spectra were acquired in the electron impact (EI) mode with the Ion Trap detector set as follows: transfer line, manifold and trap temperatures 170, 60 and 150 °C, respectively. The mass range was 33 m/z to 350 m/z, with a scan rate of 6 scan/s and without solvent delay. The emission current was 50 μA, and the electron multiplier was adjusted according to the auto-tune procedure. The maximum ionization time was 25.000 μs with an ionization storage level of 35 m/z with a pre-scan time of 100 μs. The analysis was performed in Full Scan mode.

Two different approaches were used to analyse mass chromatograms: target and non-target approaches.

5.2.10 Multivariate analysis

In the non-target approach, the raw chromatograms were imported to *MetAlign*™ [38] for spectral alignment and differentiation. The optimization of the software parameters was performed taking into account the need to preserve the original compounds features after preprocessing. The settings used in the preprocessing software are presented in Table 5.1. At the end of the alignment of dataset, the Excel compatible output matrix was obtained and subjected to multivariate analysis techniques, namely, relevant principle component analysis (RPCA) [281] and partial least squares regression (PLSR) [305]. All programming and statistical analyses were performed using R (R-Project R, http://www.r-project.org/).

The RPCA algorithm is a blind and non-supervised method by which samples are grouped together, and relevant features, discriminating between the samples in the time course, can be captured [401]. Time-course fermentation metabolic direction can be observed in relevant PC's scores and interpreted by using the relevant PC's loadings. These indicate the fragments and compounds that

are formed and consumed during the fermentation process, as well as, measuring the contribution of them in the different fermentations. Chromatograms were normalized by the internal standard (3-octanol) by mean scaling and division by standard deviation for the assignment of the same degree of importance to the preprocessing resulting metabolites. It is important to note that, the normalization can be considered an artefact for data interpretation as it is attributed the same degree of importance to the true signal and to noise, and thus, metabolites which are considered to be important to explain samples metabolic profile in each fermentation time can correspond not to valid compounds but to noisy compounds.

Table 5.1 *MetAlign™* preprocessing optimized parameters

MetAlign™ Part	Parameter	Value
A - Baseline and noise elimination parameters	*retention begin (Scan.nr)*	262
	retention end (Scan.nr)	4347
	maximum amplitude	10000
	peak slope factor (xNoise)	1
	peak threshold factor (xNoise)	2
	peak threshold (Abs.Value)	10
	average peak width	10
B - Scaling and aligning data sets	*begin 1st region (Scan.Nr / Max.Shift)*	261 / 20
	end 1st region (Scan.Nr / Max.Shift)	4347 / 20
	pre-align processing	iterative
	maximum shift per 100 scans	35
	min factor (xNoise) (1st iteration / last iteration):	3 / 2
	min number of masses (1st iteration / last iteration)	8 / 3
	group	1

The mass spectra features of each volatile compound were validated using two methods: *1)* target features by retention time correspondence; and *2)* correlation between features documented in the NIST 98 MS library.

PLSR was mainly used to determine metabolic co-expression between known and/or unknown compounds present in GC-MS chromatograms. PLSR decomposes the co-variance matrix between the sample chromatogram and supervised metabolites matrix to develop a linear relationship between them, using only the projections that maximize the correlation between the two datasets.

Thus, co-expression between supervised metabolites and all metabolites present in a GC-MS chromatogram can be studied by analyzing the PLS coefficients. Therefore, the amount of metabolites obtained by direct peak width integration in the supervised target methodology can be plotted against the overall folded matrix, obtained from the non-target approach, for classification and validation purposes. In this case, we used the unfold PLSR technique (U-PLS) [307, 308]. The aligned chromatograms form a tri-linear tensor that can be unfolded into a single matrix. U-PLS follows the same algorithms of PLSR [401], maximizing the unfold chromatogram co-variance matrix, to obtain the U-PLS. After U-PLS algorithm application, peaks and fragments highly correlated to the supervised metabolites can be visualized as a chromatogram, where positively correlated peaks are co-expressed with the reference compounds and the negatively correlated peaks are inversely correlated. This allows for the analysis and validation of compounds corresponding to the holistic approach that are being consumed or produced in the fermentation through time.

A set of compounds, including the identified as the most contributory (presented in the loadings plot) for samples differentiation and dispersion through the RPCA scores plot, as well as other known compounds that participate with those in certain pathways, were identified in a target mode. This was done by comparison with mass spectra, obtained from the samples, with the retention times of pure commercially available standards injected using the same conditions, and by comparing the Kovats indices and the mass spectra present in the NIST 98 MS library. These selected compounds were then normalized to the internal standard, in a selected ion current mode. Ions selected were respectively, for (IS) 3-octanol (99%, Sigma) m/z = 83; 2-phenylethanol (98%, Sigma-Aldrich) m/z = 91; phenylacetaldehyde (90%, Sigma-Aldrich) m/z = 91; isoamyl acetate (95%, Sigma-Aldrich) m/z = 43; isoamyl alcohol (98%, Sigma-Aldrich) m/z = 55; acetic acid (99.5%, Sigma) m/z = 43; 2-methoxy-4-vinylphenol (98%, Sigma) m/z = 150. Results from the target approach were used to compare and validate other key compounds obtained in the non-target approach, and the combination of both approaches allows obtaining a valid interpretation of the overall process.

5.3 RESULTS AND DISCUSSION

5.3.1 Fermentation performance

The effect of phenolic acids supplementation in yeast growth and the fermentation performance were studied during alcoholic fermentations using a control (in the absence of phenolic acids) against a supplemented (with phenolic acids) medium, as it is presented in Figure 5.3. Each fermentation was monitored for 11 days, and the kinetics of its control parameters (fructose, glucose and viable plate counts) were recorded. Comparing the kinetics of monitoring parameters of both control and supplemented fermentations, no statistically differences were found between them ($p < 0.05$). The maximum viable cell density of *S. cerevisiae* obtained reached around 10^7 CFU/mL from 10^6 CFU/mL inocula, as expected in this SGJ fermentations [402]. Similar viable plate count behavior in the control and supplemented fermentations indicates that the addition of phenolic acids does not inhibit yeast growth and fermentation performance, not affecting fructose or glucose consumption, also.

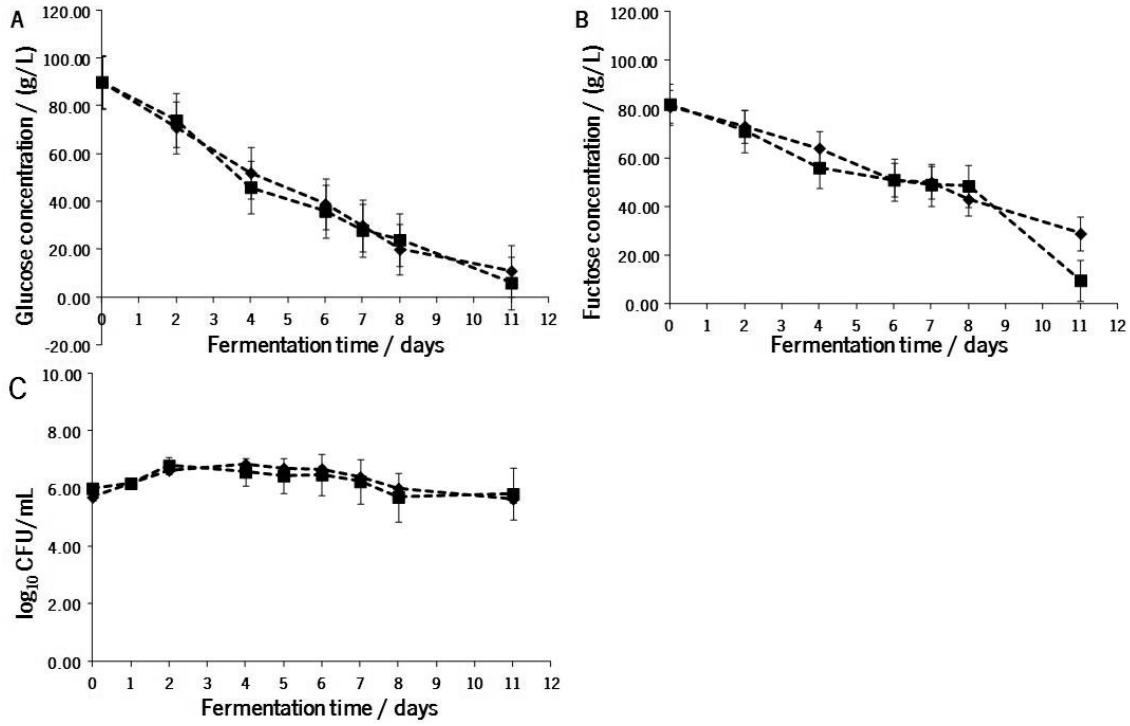


Figure 5.3 Fermentation monitoring kinetic parameters in both control (■) and supplemented (♦) medium: A - glucose concentrations / (g/L); B - fructose concentrations / (g/L); and C - cells concentrations / (log_{10} CFU/mL).

5.3.2 Electrochemical analysis

Cyclic voltammetry was used for electroactive activity scanning providing a holistic vision of the entire fermentation process, which later should be tentatively understood and validated with selected molecules by the HPLC - DAD analysis. The oxidation curves from both control (dashed line) and supplemented medium (full line) are represented in Figure 5.4, at the beginning - T0 (Figure 5.4A), and at the end - T11 (Figure 5.4B) of the fermentation process.

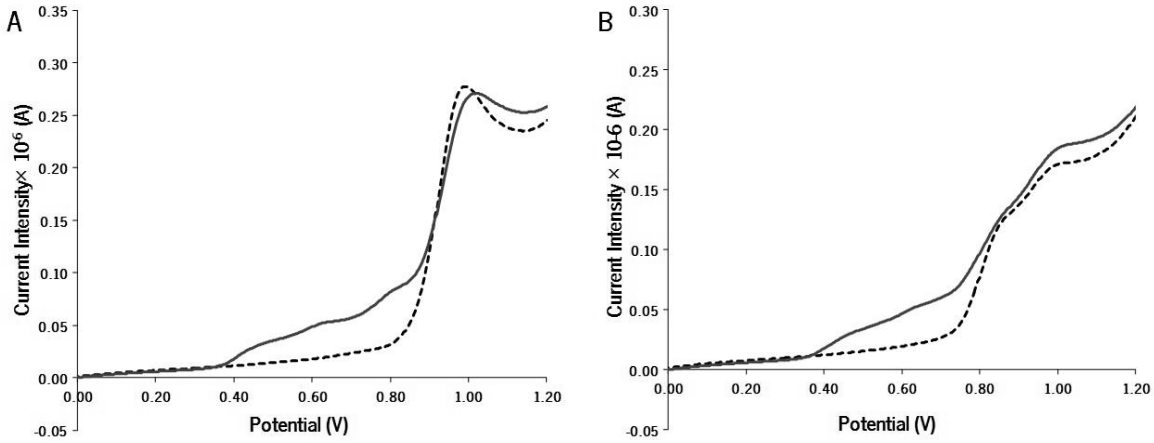


Figure 5.4 Electrochemical signals of control (dashed line) and supplemented (full line) fermentations in: A - the beginning of the fermentation process (T0); and B - the end of the fermentation process (T11).

In Figure 5.4 it is possible to observe different electrochemical signals between the control and the supplemented media (dashed and full lines, respectively), as well as different profiles in the beginning and end of the fermentation processes (Figure 5.4A and Figure 5.4B, respectively) as a result of a higher concentration of electro-active compounds in the beginning of the fermentation process. Supplemented media voltammograms (full line) display peaks at lower potentials (0.4 to 0.8 V) which can be reasonably assumed to be due to the added phenolic compounds (as the control voltammogram does not show this deformation in this potential range). Precise identification of these compounds, related to each peak's position, cannot be ascertained, nevertheless, their structure influences the anodic peak position, as it has been previously reported in the literature [32] .Using the compound's structure and the evidence from published studies, it can be tentatively assumed that, in potential range (0.4 to 0.8 V), gallic acid has the lowest formal

potential, followed by caffeic and protocatechuic acids, ferulic and vanillic acids, and *para*-coumaric acid [32]. The higher number of hydroxyl substituents present on the benzene ring decreases the formal potential of the phenolic compound [403]. Gallic acid, a benzoic acid derivative with 3 available hydroxy groups, is therefore easily oxidized [166, 404]. In the same way, caffeic and protocatechuic acids have an easily oxidizable *ortho*-diphenol group which makes them more easily oxidizable than a phenolic acid with an isolated phenol group like *para*-coumaric acid. The other phenolics which have significantly higher formal potentials (ferulic and vanillic acids) lacked an *ortho*-diphenol which is dependent on oxidation of an isolated phenol group adjacent to a methoxy group. These compounds are therefore expected to be less active as antioxidants where reducing ability is the key requirement. The oxidation of these phenolics, which could be due to processes involving one or two electrons [405] produced broad peaks and is largely irreversible [166]. The type of electro-active compounds changed during fermentation (T0 to T11), as indicated by the voltammograms (Figure 5.4A, and Figure 5.4B). Voltammograms clearly show a decrease in the amplitude of the broad band at approximately 1 V. Conversely, one broad band at 0.8 V is brought about by fermentation, indicating that a change in the overall resistance to oxidation does indeed occur during alcoholic fermentation and shows a pattern similar to that of normal wines [168]. It must be remembered that our medium is of a considerably minimal nature, and that it is interesting that such a medium produces the same type of electro-active substances as natural must.

5.3.3 HPLC - DAD analysis

Concurrent with the electrochemical study, the kinetics of the added phenolic acids to the supplemented media were quantified and monitored using HPLC – DAD, and are presented in Figure 5.5. By observing phenolic acids kinetics (Figure 5.5) it was possible to detect that phenolic acids most affected by fermentation were *para*-coumaric (■) and ferulic acids (♦) (Figure 5.5A). In fact, it is known that during the fermentation process, *para*-coumaric and ferulic acids can be used for 4-vinylphenol and to 2-methoxy-4-vinylphenol (4-vinylguaiacol) or vanillic acid formation, respectively [3, 406, 407]. In fact, the final concentration of the ferulic acid is lower than the vanillic acid. The phenolic acids presented in Figure 5.5B and C were not largely changed during the fermentation time-course. Protocatechuic acid (♦) concentration (Figure 5.5B) suffers a small decrease at the beginning of alcoholic fermentation, but thereafter remains constant until the end

of the time-course sampling time. Caffeic (■) (Figure 5.5B), gallic (♦) and vanillic (■) (Figure 5.5C) acids did not show any significant change in concentration throughout the fermentation. Besides the study of the added phenolic compounds, one compound of particular interest was found increasing in both control and supplemented media, after 4 days of alcoholic fermentation, which growth pattern, in control medium, is shown in Figure 5.5D.

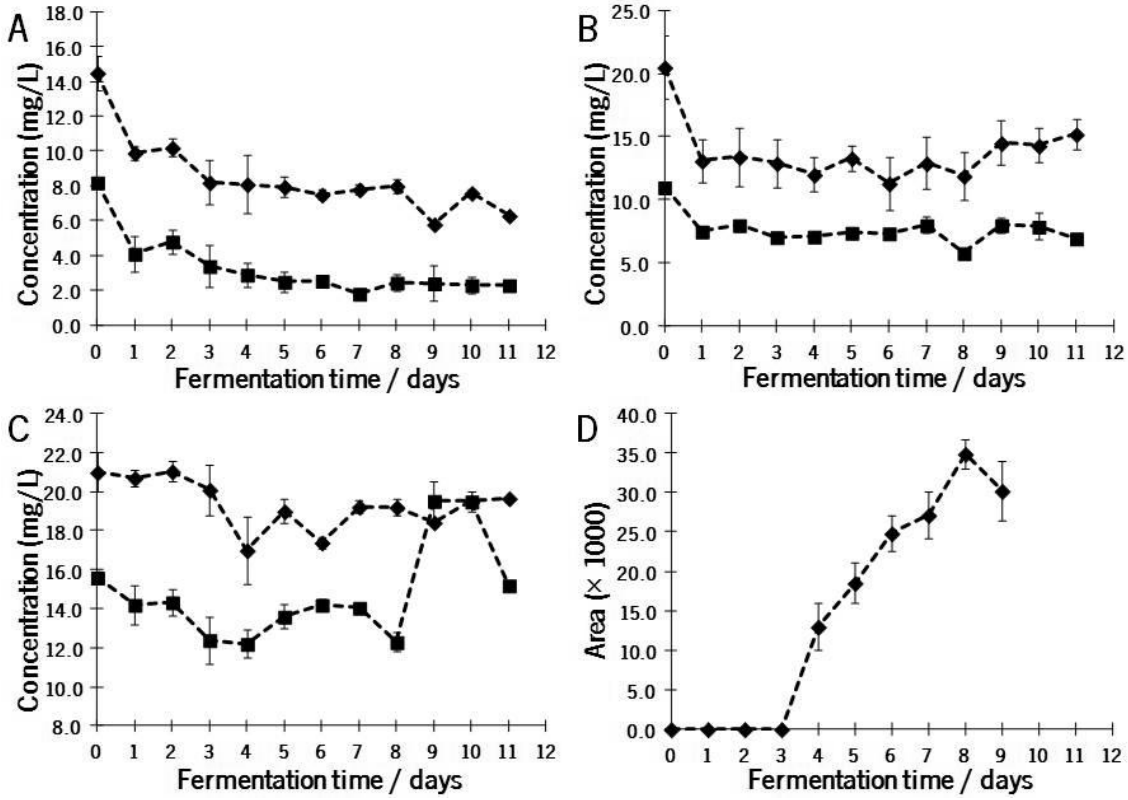


Figure 5.5 Phenolic acids kinetics during the supplemented fermentation process: A - *para*-coumaric (■) and ferulic (♦); B - caffeic (■) and protocatechuic (♦); C - vanillic (■) and gallic (♦); D - unknown compound (♦).

Attempts were made to further identify this unknown compound, and it was found to have the same retention time as the standard for chlorogenic acid. These results were confirmed by LC-MS analyses (results not shown). The discovery of chlorogenic acid in the control medium was unexpected as *S. cerevisiae* is not a known producer of any phenolic acids. This is of particular interest, because its presence in the control medium suggests that *S. cerevisiae* has the ability to produce antioxidants "*de novo*" from simple precursors such as sugars and amino acids. Chlorogenic acid is the condensation product of quinic and caffeic acids and furthermore, it is a

product of an intermediate in aromatic acid biosynthesis and a late intermediate of phenol biosynthesis [408]. While it is not documented that yeast produces this acid, *Saccharomyces cerevisiae* does contain some of the enzymes and certain pathways used by plants for its production. *Saccharomyces cerevisiae* perform aromatic acid *de novo* biosynthesis via the Shikimate pathway [406], creating a possible pathway for production of quinic acid. Furthermore, the exploration of the *Saccharomyces* Genome Database (SGD) shows that the yeast contains the PAL gene (phenylalanine ammonia lyase, ENZYME: 4.3.1.5), which encodes the enzyme needed to catalyze the deamination of phenylalanine to trans- cinnamate and ammonia. These results clearly indicate that more research and more exhaustive study can be necessary to validate and understand the reasons for production of chlorogenic acid during alcoholic fermentation.

5.3.4 Metabolomic analysis

In addition to the electrochemical changes, it is known that during the fermentation of wine, different biological processes occur which lead to different organoleptic characteristics of the final product, depending on the conditions of the fermentation medium [3]. In this context, it seems noticeable that through the interpretation of the holistic view that can be provided by the non-target analysis of the overall process as well as the kinetics of individual compounds of interest in target mode, it may be possible to understand the metabolic phenomena concomitant with assimilation or processing of phenolic compounds added the fermentation medium.

After the chromatograms preprocessing using the *MetAlign*™ software [38], the resulting dataset was subjected to a non-supervised approach for classifying the metabolic information. This included the RPCA analysis, which was used to impose a statistical structure on the pre-processed HS-SPME/GC-MS obtained dataset. Figure 5.6 presents the samples scores Gabriel plot, where Principal Component 1 (PC1) and Principal Component 2 (PC2) have a discriminated power of 67.14% and 10.82%, respectively, totalizing approx. 78% of the chemical information explanation.

Samples distribution in the PC1 *vs* PC2 space presented in Figure 5.6, shows similar features of the control fermentation samples (Grcsc) compared to the supplemented media - Replica 1 (Grsc1) and 2 (Grsc2). PC1 is the most relevant component and is the result of the variable time during the fermentation process, as samples distributed in the PC1 axis space have a time-dependence. This

demonstrates that the metabolic profile is changing over time, during the fermentation process in all media.

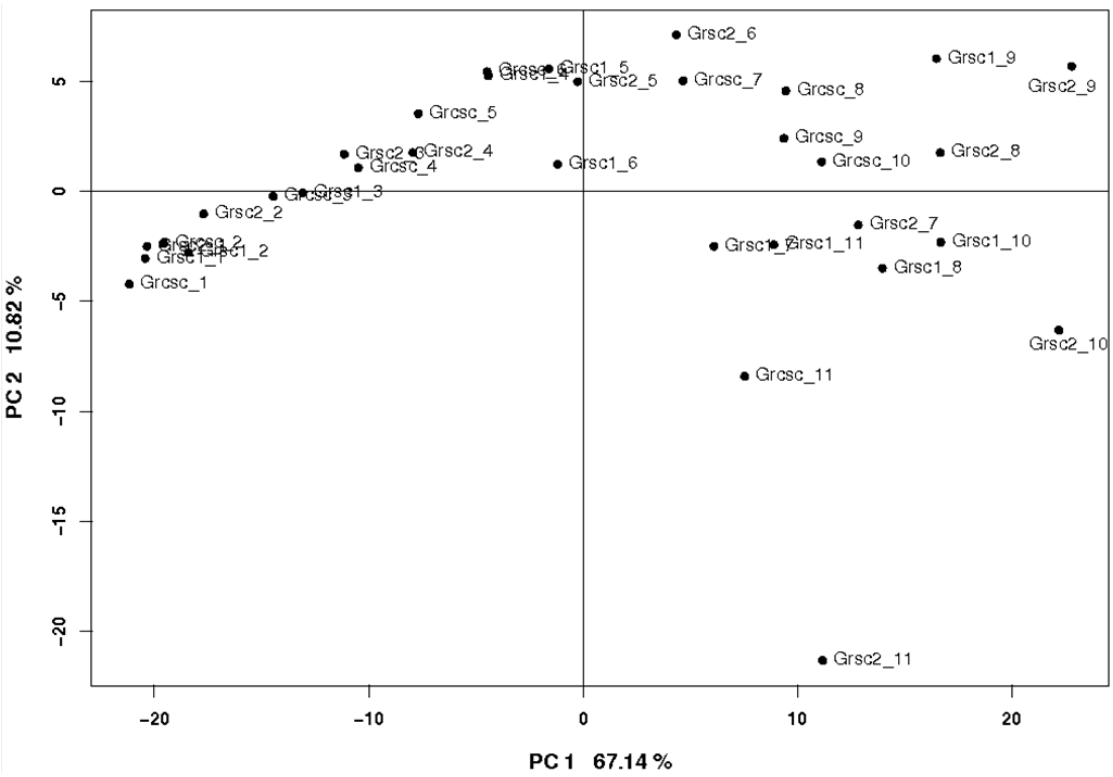


Figure 5.6 Scores plot resulting from Principal Component Analysis method: PC1 and PC2.

The interpretation of the metabolic information that differentiates samples throughout the fermentation can be done analysing the loadings of the singular value decomposition analysis, represented in Figure 5.6 and described in Table 5.2. Peaks reported in Figure 5.7 correspond to those who must contribute to samples distribution throughout the PC1 axis (*xx* axis), in this case, were identified as isoamyl acetate, acetic acid and 2-phenylethanol (Figure 5.7b, Figure 5.7e and Figure 5.7g, respectively), which kinetics can be found in the target approach (Figure 5.9). Peaks labelled in Figure 5.7a, Figure 5.7c, Figure 5.7d and Figure 5.7f, were tentatively identified as unknown compounds, and more efforts should be done in order to interpret the results. In fact, the most laborious task linked to the non-target approach is the identification step, usually by searching the NIST library and by direct comparison after injection of standards [400]. In this work, as the main objective is to present a tool to provide an enlarged view of the overall system, we did not carry out a detailed identification and quantification of all the metabolites.

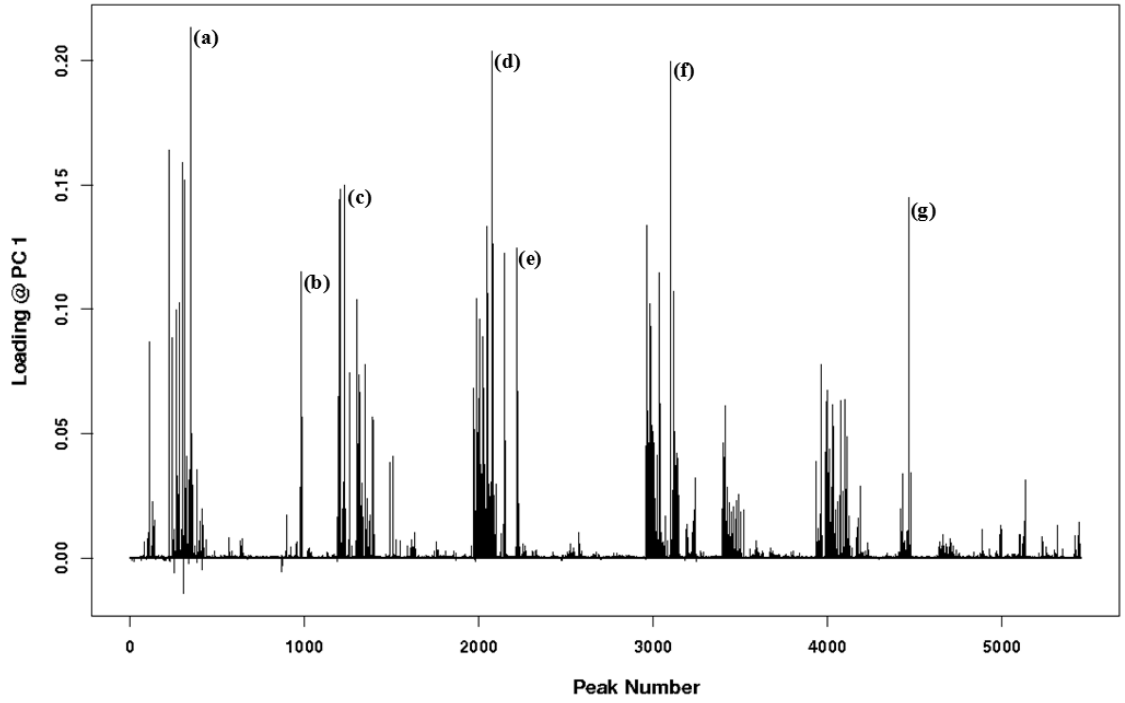


Figure 5.7 Loadings plot corresponding to the PC1, resulting from Principal Component Analysis method (the annotation of the highlighted fragments is presented in Table 5.2).

Table 5.2 Tentative identification of the loadings plot fragments corresponding to the first principal component resulting from principal component analysis method

Label	Peak No	Scan No	m/z	Tentative identification
(a)	349	389	46 (1); 73 (0.19); 92 (0. 09); 47 (0.05)	**Unknown**
(b)	980	724	43(1); 55(0.54); 70(0.45); 41(0.24)	**Isoamyl acetate**
(c)	1203	975	39(1); 41(0.94); 70(0.88)	**Unknown**
(d)	2076	1661	41(1); 173(0.82); 57(0.68); 56(0.52)	**Unknown**
(e)	2220	1712	43(1); 45(0.78); 60(0.32); 61(0.09)	**Acetic acid**
(f)	3101	2256	41(1); 201(0.98); 60(0.35); 44(0.25)	**Unknown**
(g)	4468	2983	91(1); 92(0.63); 122(0.22)	**2-Phenylethanol**

Partial least squares regression (PLSR) was also used to determine metabolic correlations between 2-phenylethanol and all peaks obtained in the GC-MS chromatograms. 2-phenylethanol fermentation kinetic obtained by direct peak-width integration in the target approach was plotted against the data matrix resulting from *MetAlign*™ preprocessing in the non-target approach, for classification purposes. This metabolite was arbitrarily selected, as the main purpose is to show that besides being possible the characterization of the system holistically, the statistical tools used in this pipeline may also be crucial in understanding the correlations, then co-expressions, of interesting metabolites, which can be correlated with the phenolic acids kinetics, during the fermentation process.

Figure 5.8 shows the resulting display of U-PLS coefficients of the model performed using 2-phenylethanol amounts, and higher coefficients correspond to the most correlated scans with 2-phenylethanol. The image map colour (blue to red scale) of the point determines the magnitude of correlation, being higher for red. In Figure 5.8, it is possible to observe that 2-phenylethanol is strongly correlated with 2-phenylethanol (scan 2983) (Figure 5.8a), acetic acid (scan 1712) (Figure 5.8b), 2,3-dihydro-3,5,dihydroxy-6-methyl-4(H)-pyran-4-one (scan 3802) (Figure 5.8c) and 5-hydroxymethyl-furfural (scan 4292) (Figure 5.8d), which means that the kinetics of production of each compound is similar during the fermentation process. The assignment of scan numbers to their corresponding compounds was facilitated by the target approach, which identified a set of metabolites associated with alcoholic fermentation, and validated their mass spectral profiles. Using this classification method, and after the identification process, new information about metabolic pathways may be revealed.

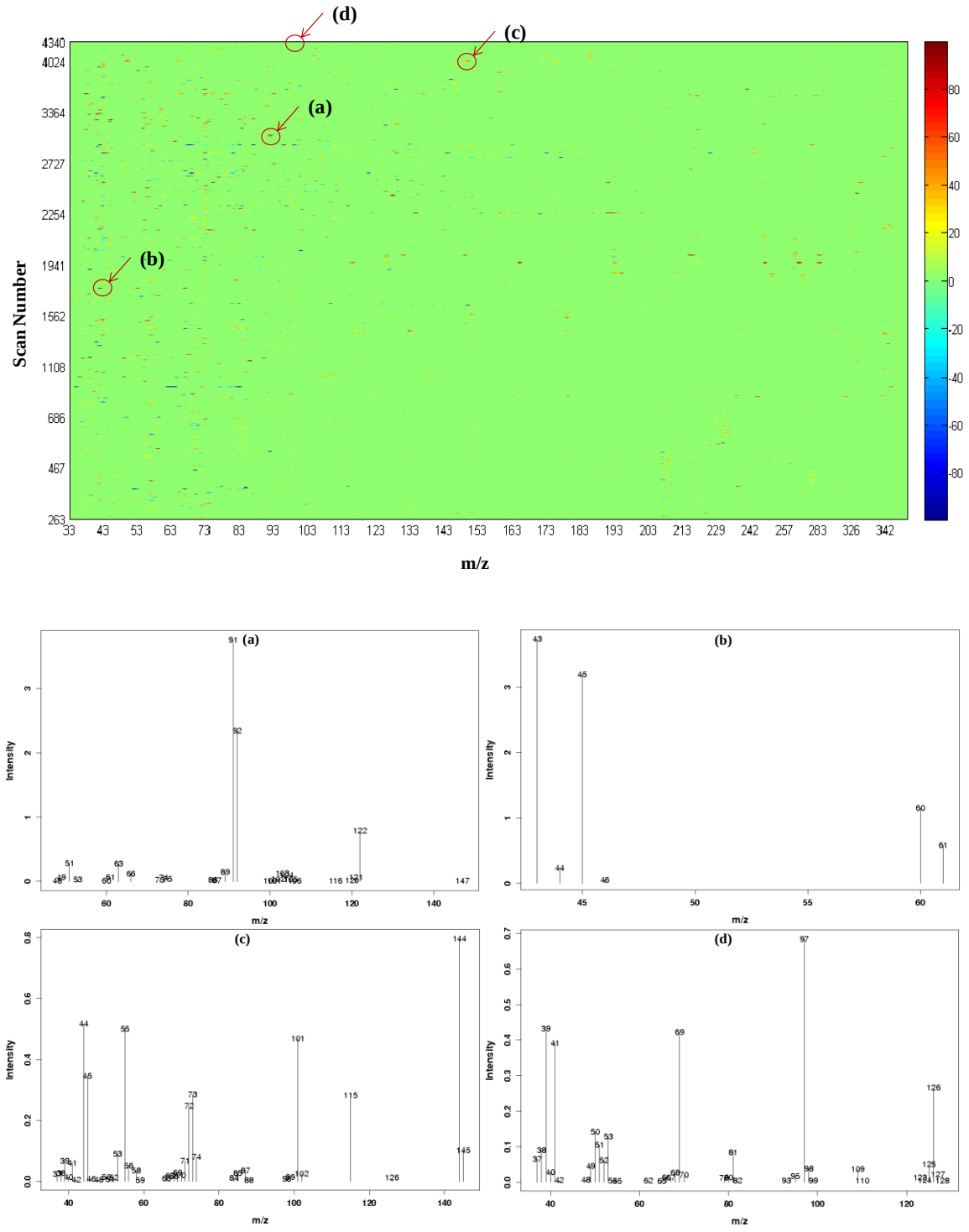


Figure 5.8 Partial least squares regression resulting coefficients correlations with 2-phenylethanol throughout the fermentation process: (a) 2-phenylethanol (scan 2983); (b) acetic acid (scan 1712); (c) 2,3-dihydro-3,5,dihydroxy-6-methyl-4(H)-pyran-4-one (scan 3802); and (d) 5-hydroxymethyl-furfural (scan 4292).

The evolution throughout fermentation of some secondary metabolites, including those identified as the most contributory for the explanation of the fermentation progress, is presented in Figure 5.9. Higher alcohols and esters, including 2-phenylethanol and isoamyl acetate, and those known to be correlated with them (phenylacetaldehyde and isoamyl alcohol), as well as the acetic acid were identified and quantified during the fermentation process. The fermentation kinetics of each metabolite in both the supplemented and control media showed to be statistically similar ($p<0.05$).

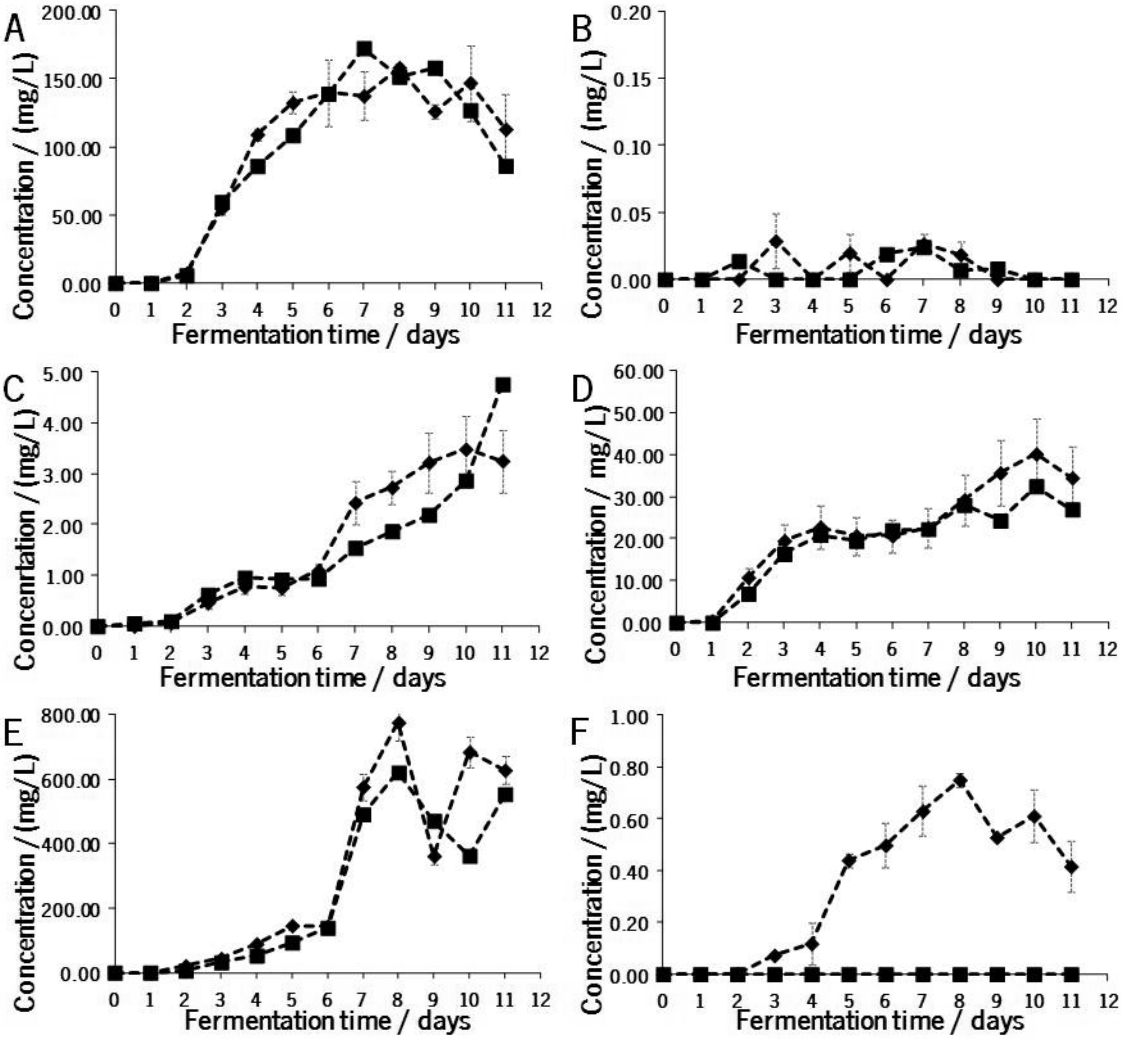


Figure 5.9 Target analysis of some volatile metabolites kinetics found to be important for samples dispersion in the scores plan and some known to be correlated with those in both control (■) and supplemented (♦) medium: A - 2-phenylethanol; B - phenylacetaldehyde; C - isoamyl acetate; D - isoamyl alcohol; E - acetic acid; F - 2-metoxy-4-vinylphenol.

So, in Figure 5.9 A, it is possible to observe a greater increase of the 2-phenylethanol until the 8th day of the process. The production of this fusel alcohol is mainly associated with the phenylalanine degradation by the *Ehrlich* pathway [4, 5], where phenylacetaldehyde is an intermediary compound (being consumed as is produced, as it is presented in Figure 5.9B).

Comparing both 2-phenylethanol and phenylacetaldehyde kinetics, it is possible to confirm their relationship as when the concentration of phenylacetaldehyde decreases to zero the concentration of 2-phenylethanol no longer increases. This also suggests that phenylalanine, initially added to the SGJ, was totally converted into 2-phenylethanol during the fermentation processes by *S. cerevisiae* after 8 days of fermentation.

The synthesis of isoamyl acetate (Figure 5.9C) by *S. cerevisiae* is catalyzed by a group of enzymes called alcohol acetyltransferase (AAT) by utilizing higher alcohols, in this case the isoamyl alcohol (Figure 5.9D), and acetyl-CoA (resulting from pyruvate metabolism) as substrates [3]. The amount of isoamyl acetate and isoamyl alcohol appears to increase throughout the control and supplemented fermentation processes. This means that as isoamyl alcohol is being produced, *S. cerevisiae* provides the conductive and continuous condition for the production of isoamyl acetate. As isoamyl alcohol is the result of L-leucine degradation by the *Ehrlich* pathway [4, 5], this is also suggested that this amino acid is present in both fermentation media throughout the fermentation processes.

Acetic acid is also a normal end product of the alcoholic fermentation [3, 409], and its concentration seems to increase until the 8th day of both fermentation processes, decreasing from then onwards until the end of fermentation (Figure 5.9E). In this case, as glucose is constantly consumed (Figure 5.3A) for formation of the end products (i.e. ethanol, glycerol, and acetic acid), the final decreasing of the acetic acid amount could be due to a residual substrate concentration available. Comparing 2-methoxy-4-vinylphenol kinetics in both the supplemented and control fermentations (Figure 5.9F), it is possible to verify different behaviors between them. In the supplemented media, the amount of this volatile phenol also increases until the 8th day; however, in the control fermentation there was no evidence of 2-methoxy-4-vinylphenol production. These kinetic differences in control and supplemented fermentations, lead us to suggest that the addition of phenolic acids are significantly affected the primary metabolism or rate of production of secondary metabolites. As it was discussed, 2-methoxy-4-vinylphenol is the result of ferulic acid metabolism, one of the added phenolic acids [3, 406].

5.4 CONCLUSIONS

In this work, the influence of the *S. cerevisiae* strain (PYCC 4653) metabolism on the antioxidant capacity of a synthetic wine was evaluated. The variation of the phenolic acids content within the fermentation process was determined by HPLC-DAD and the redox-potential was measured by cyclic voltammetry. A higher concentration of electro-active compounds was found in the beginning of the fermentation process. According to the voltammograms information, the fermentation process introduces a variation of the electro-active compounds leading to an overall resistance to oxidation, within time-course. The synthetic medium shows a pattern similar antioxidant capacity to that of normal wines.

The non-target bioanalytical pipeline combining the electrochemical features with the metabolic information allowed to understand that the electrochemical features representing the redox-potential changes throughout the alcoholic fermentation process is somehow accompanied with the production of some metabolites, such as 2-phenythanol, acetic acid and isoamyl acetate. Also, the use of the classification methodology called U-PLS allowed to find out metabolites, which kinetics are well correlated with 2-phenylethanol, which include the acetic acid, 2,3-dihydro-3,5,dihydroxy-6-methyl-4(H)-pyran-4-one and 5-hydroxymethyl-furfural. Moreover, *S. cerevisiae* had the capacity of producing chlorogenic acid in the supplemented medium fermentation from simple precursors present in the minimal medium

In this work, the combination of information from different detectors, including cyclic voltammetry, liquid and gas chromatography in this work, revealed exciting prospects to explore and gather the maximum information regarding complex systems, such as real must fermentation.

CHAPTER 6

X-METABOLOMICS: HIGH-THROUGHPUT METABOLOMICS PIPELINE APPLIED TO *PORT* WINE FORCED AGING PROCESS

Metabolomics aims at gathering the maximum amount of metabolic information for a total interpretation of biological systems. In this Chapter, a process analytical technology pipeline, combining gas chromatography – mass spectrometry data preprocessing with multivariate analysis, is presented. This pipeline is integrated in a metabolomics in-house platform called *X-Metabolomics*, which was demonstrated by application to a *Port* wine "forced aging" process under different oxygen saturation regimes at 60 °C, in order to characterize the overall biological process. It was found that extreme "forced aging" conditions promote the occurrence of undesirable chemical reactions by production of dioxane and dioxolane isomers, furfural and 5-hydroxymethylfurfural, which affect the quality of the final product through the degradation of the wine aromatic profile, color and taste. Also, were found high kinetical correlations between these key metabolites with benzaldehyde, sotolon, and many other metabolites that contribute for the final aromatic profile of the *Port* wine. The use of the kinetical correlations in time-dependent processes as wine aging can further contribute to biological or chemical systems monitoring, new biomarkers discovery and metabolic network investigations.

The information presented in this Chapter was published as:

Castro CC, Martins RC, Teixeira JA, Silva-Ferreira AC (2013) Application of a high-throughput process analytical technology metabolomics pipeline to *Port* wine forced aging process, *Food Chemistry*, 143, 384-391.

6.1 INTRODUCTION

Metabolomics is a recent "*omics"* field and is based on biology science devoted to identify phenotypes, the connectivity that relates their constituents and the dynamical response to perturbations, as well as the relationships with the rest of the cellular molecular machinery, at the regulatory levels such as genomics, transcriptomics and proteomics [28].

The study of such complex matrix on a chemical composition perspective can be described as being *multi-scale*, *i.e.* several orders of magnitude (ppm-ppt) and *multivariate*, *i.e.* large diversity of chemical substances, and constitutes an extremely challenging problem being the main goal to obtain a comprehensive profile of the molecular biology machinery and biochemistry. Ideally it would require a true "*omic sensor"*, *i.e.*, with an unlimited linear zone of response and sensitive to all substances resulting in a global picture of the bioprocess [410].

A high-throughput metabolomic chromatography systems consists in hardware (analytical equipment) and software (signal processing, data storage and multivariate analysis), which are becoming a trend in modern biotechnology [212]. The most common chromatographic platforms for metabolomics include gas chromatography – mass spectrometry (GC-MS), liquid chromatography – mass spectrometry (LC-MS), capillary electrophoresis – mass spectrometry (CE-MS), high performance liquid chromatography – diode array detection (HPLC-DAD) and nuclear magnetic resonance (NMR), all of which can be brought together for a wide screening of natural metabolites in complex biological samples. These methods become more appealing to systems biology and molecular biology when the screening is subjected to signal processing, turning chromatography into a high-throughput system capable of extracting peaks by an automated technique, which start to reveal a holistic view of the metabolism [139].

GC-MS is specially suited for the study of yeast volatile metabolites [12, 370]. These important in fundamental biological functions, such as signalling and precursors of biochemical pathways, as well as in the formation of aroma from wines that result from fermentation and aging, either in barrels or bottles. Several hundreds of compounds can be captured by GC-MS, and each compound produces a unique mass spectral fingerprint, which is afterwards used for metabolites recognition and quantification [12]. Classical mass spectroscopy is highly laborious and a significant amount of time is necessary for peak analysis, fingerprint recognition, identification and quantification by an analyst, that makes scientific discovery or holistic characterization impractical

[256]. In this context nontargeted methodologies are extremely important, with an unbiased approach, toward understanding the biological system [32, 137, 396]. Chromatography enables physico-chemical deconvolution of the complex matrix components during the analytical process originating a pattern, that is, a chromatogram. Although it allows samples classification based on the different patterns, it is important to relate that there are inherent variations in the chromatographic process, mainly related to the retention time (peak shifts) and baseline drifts [232]. Therefore, in order to use GC-MS data on a high-throughput base, those issues must be focused on the preprocessing data handling prior to multivariate analysis [256]. In that regard, attempts at the automatic processing of chromatograms have been developed, including preprocessing tools, such as: noise filtering, baseline correction, peak detection, alignment, identification and normalization algorithms [18, 38, 232].

Several pieces of commercial or free software based on metabolomic raw data preprocessing, peak detection and/or quantification have been developed in the last years - msInspect [411], *MZmine* [239], *MetAlign*™ [38], *OpenMS* [241], *XCMS* [18], *SpecArray* [412], *XAlign* [413], *MassUntangler* [414], *MathDAMP* [244], *MetaboliteDetector* [245] - and some from instrument companies - *Waters MarkerLynx, ThermoFisher SIEVE, Agilent MassHunter, Applied BiosystemsMarkerView, Shimadzu Profiler AM +* and *LECO ChromaTOF* [12]. More recently, bioinformatics web applications namely *MeltDB* [415], *TICL* [416], *MetaboAnalyst* [248] and *MetabolomeExpress Project* [417] have also emerged, focused on the biological interpretation via multivariate explorative and statistical analysis of preprocessed datasets. Furthermore, some of these software programs also include metabolomic data integration with transcriptomics, proteomics and/or genomics databases [415] - *KEGG* [258-260], *BRENDA* [263], MetaCyc [261] - or metabolomic databases - *NIST* [267] and *SDBS* [268].

In this Chapter, *X-Metabolomics*, a new software metabolomic pipeline is presented and applied to a *Port* wine data set. *X-Metabolomics* concerns in a stepwise approach for data handling and data processing, which combines different sources of algorithms consolidated on a tool, providing metabolites screening, quantification, and biological interpretation for a better understanding of biological systems It can be used as a diagnostic tool for high-throughput data supervision and validation, allowing the biological interpretation, providing the analysis of compounds expression and co-expression in the overall network. This pipeline can be used as a multidisciplinary tool, as it can be useful for: *i)* bioanalytics - to obtain the metabolic matrix of compounds and possibly

explore *de-novo* interesting compounds and increase the number of known molecules or metabolic pathways; *ii)* bioengineers - to understand and monitor the biological system, as well as to reconstruct the metabolic network of the system; and *iii)* bioinformatics - to create and improve preprocessing tools for preserving the integrity of various compounds features.

The main goal of the application of the *X-Metabolomics* pipeline to a *Port* wine forced aging case study, is to evaluate the impact of the presence of oxygen and higher temperature during the "forced aging" of a *Port* wine matrix, using the bioanalytical pipeline developed. *Port* wine aging process is an extremely complex chemical process that has been extensively studied by our research group [1, 72, 73], where several chemical mechanisms take place and are responsible for differences in sensory perception affecting final product quality. The study of such matrix involves the use of both chromatographic signal preprocessing and MVA methodologies, together with the exploration of candidate metabolites expression and co-expression within the overall aging process (see Figure 6.1). Candidate metabolites correspond to MVA model variables, candidate to explain specific pathways of the metabolism that have chemical meaning and can be used for understanding the overall process. The co-expression of a candidate metabolite constitutes a powerful feature, enabling the study of potential interconnections between the candidates formation and consequently allowing further research for overlapping/connections between chemical and/or biochemical mechanisms. In this context, the global high-throughput pipeline methodology provides increased rate of metabolites identification involved in the chemical process contributing to further build the metabolic network of the overall process.

6.2 *X-METABOLOMICS* WORKFLOW

X-Metabolomics was designed for aiding metabolomics research, providing a useful laboratory pipeline for real-time GC-MS diagnosis. The pipeline is based on R statistical programming environment and Tcl/Tk X-window graphical user interface. The *X-Metabolomics* native processing pipeline works as follows: Preprocessing: *i)* chromatograms import (directly or selected samples), peaks extraction and alignment; or *ii)* direct importation of resulting data set from other preprocessing software; Post-processing: *i)* supervised filtering; *ii)* fragments classification; *iii)* multivariate statistics for data interpretation and classification; *iv)* building the identification and

composition tables; and *v)* compounds co-expression and expression in time-course (see Figure 6.2). The data-handling software configurations must be specified on a configurations text file.

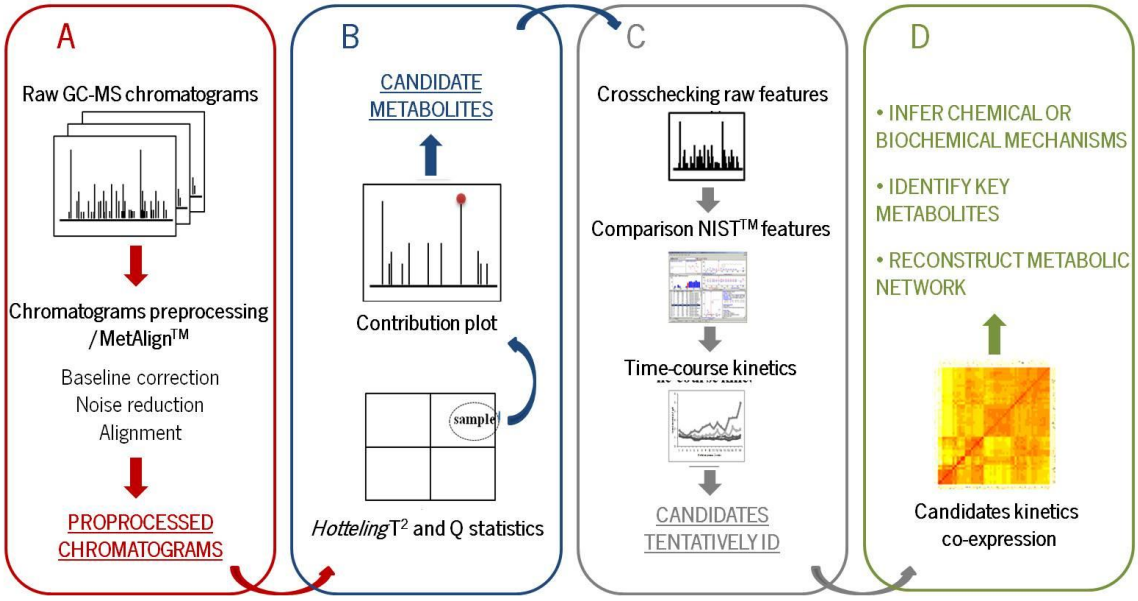


Figure 6.1 High-throughput pipeline methodology for GC-MS data processing: A - GC-MS data preprocessing; B - Multivariate Analysis; C - Candidate metabolites identification; D - Temporal relationships of candidate metabolites for process contextualization.

6.2.1 Data import

The native preprocessing of *X-Metabolomics* is based on the *XCMS* methodology [18] and includes feature extraction, peaks grouping and scan alignment. However, it is also possible to import preprocessed data from other sources, such as *MZmine* [239] or *MetAlign*™ [38], as it supports other open file formats (*netCDF, ASCII, mzXML*) (see Figure 6.2).

6.2.2 Preprocessing

GC-MS chromatograms preprocessing can be performed using *XCMS* methodology [18], which is originally implemented, *MZmine* [239] or *MetAlign*™ [38], and then proceed with the *X-Metabolomics* post-processing. Whatever the preprocessing software in use, the input parameters should be optimized as it is important to preserve the original compounds features for further analysis [137]. The optimization settings for each preprocessing software can be either instrument

dependent and so, constant for all the chromatograms of the data set, or less objective, distinguishing the real chromatographic peaks versus noise or window sizes in which peaks in two chromatograms are considered the same [418]. The optimization of the preprocessing parameters must be performed taking into account both the preservation of the integrity of a particular set of volatiles features (used for preprocessing supervision) and the maximum of total extracted scans. For these reasons, the optimization of the settings is, usually, performed in a data-driven mode, being re-set until both assumptions are satisfied. In this context this is the starting point from which biological and biochemical processes understanding is made.

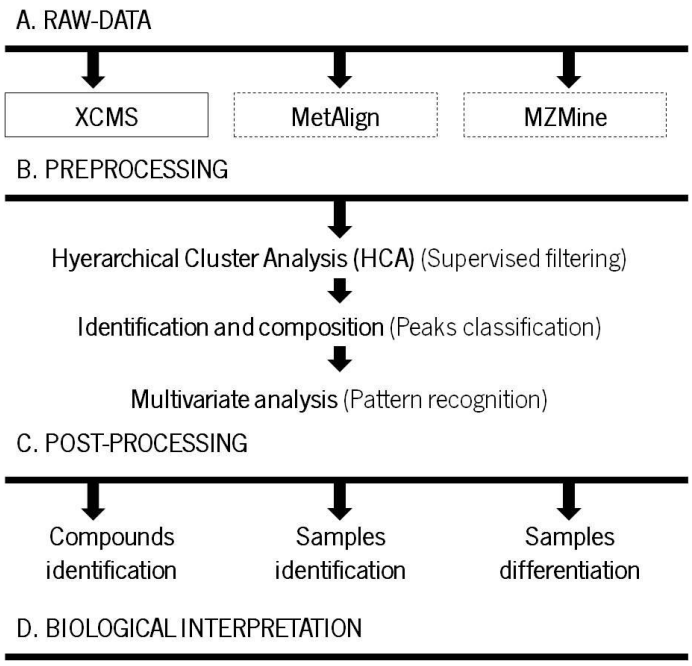


Figure 6.2 Implementation steps of *X-Metabolomics*, the in-house developed pipeline for high-throughput metabolomics.

There is no consensus about the benchmarking criteria which should be used for comparing the quality of algorithms applicable at different steps of preprocessing [232]. For peak detection step, F-factor, which is a combination of recall and precision, was used by Tautenhahn *et al.* (2008) [233] and Lange *et al* (2008) [38] used the same methodology for the alignment step, while Koh *et al.* (2010) [39] used the comparison of R^2 and Q^2 values and prediction accuracy of ordinary partial least squares for discriminant analysis (OPLS-DA) models.

During the preprocessing, peaks found to represent the same scan across samples, based on similar retention times, are placed into groups, *i.e.*, by performing retention time correction by interpolation with a fitting function (linear or β-spline), the final matrix ***M*** (m,n) is obtained, where n is the fragment number and m the sample (Figure 6.3). The nonlinear retention time alignment process is performed by identifying groups of peaks (a master peak list) to use as standard, and then the retention time deviation of each sample according to the standard is calculated in an iterative mode and peaks are aligned using the non-linear warping. The ***M*** matrix is further used in post-processing and therefore for exploring the time-course kinetics and co-expression of compounds providing the system interpretation.

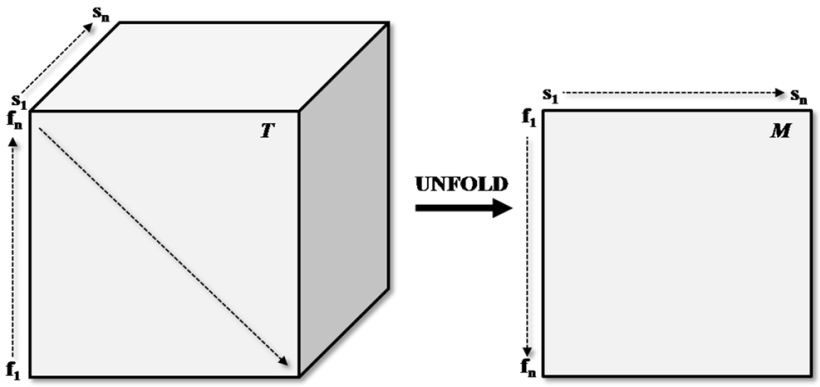


Figure 6.3 Retention time correction between all samples chromatograms.

6.2.3 Post-processing

The matrix resulting from the preprocessing methodologies, which includes the extracted peaks, must be filtered and the summary of the resulting metabolites features and samples information can be reached in the identification and composition tables, respectively. The final composition table is therefore explored using multivariate analysis (MVA), where samples and metabolites features can be classified using relevant principal component analysis (RPCA). Also, it is possible to determine metabolites expression and co-expression throughout the process, as well as, to predict compounds kinetics using Partial Least Squares Regression (PLSR) algorithm.

Candidate peaks filtering

Two different methodologies are implemented for peaks filtering after preprocessing. These include the threshold filtering and the supervised filtering using hierarchical cluster analysis (HCA) [17]. The threshold filtering applies a maximum threshold value leading to peaks intensity present in chromatograms and so, peaks which intensity is above the defined threshold are excluded. This methodology allows eliminating peaks that might be saturated in the matrix [419], ensuring that only peaks with a linear behavior are therefore used. On the other hand, the HCA allows the selection / deletion of groups of candidate metabolites, which features are grouped in each branch of the dendrogram, according to features correlations. HCA is a supervised filtering, which allows to explore candidate metabolites clustered in the same branch and select the interesting candidates, eliminating those known as contaminants, coming from organic solvent, glass ware or detectors artifacts, and ion complexes formed [420]. The filtering step can reduce the detection or false positive features [232], in order to undertake an accurate identification of candidate metabolites involved in the process. In *X-Metabolomics* pipeline, each cluster can be individually explored to further understand compounds expression and co-expression over time, as well as, metabolic pathways contextualization.

Relevant principal component analysis

MVA involves the use of mathematical and statistical tools to extract the information from complex datasets (*e.g.* GC–MS datasets) [421], considering the system as a multi-dimensional comparison of the different chemical information present in the sample [396]. Relevant principal component analysis (RPCA) [422] is a non-supervised methodology that determines relevant orthogonal decompositions of the information able to discriminate between samples (scores), and variables (loadings), that provide relevant interpretation of significant variables (or metabolites) on each principal component [422].

The Diagnostic plot (Q statistics *vs Hotelling* T^2) is performed because not all features in the metabolic matrix (resulting from the preprocessing algorithm) preserve the same quality after data decomposition into relevant principal components [286]. In these cases, their reconstruction is statistically impossible, and their metabolic information is widely different from the average. Q statistics (square prediction error) is determined to assess the feature extraction quality, and the

line crossing the Q axis correspond to the statistic confidence interval (Qα) determined using the average and standard deviation of Q at a level of significance of α = 0.05. Samples above Qα do not represent robust feature extractions [423]. On the other hand, *Hotelling* T^2 measures the distance to the center of the data. The line crossing the Hotelling T^2 axis corresponds to the upper confidence interval of the distribution, determined at a level of significance of α = 0.05. Samples above the $T\alpha^2$ are considered to present significantly different features [286, 424]. The determination of the Q statistic confidence interval (Qα) limits the Q values above which features are not robust.

The use of both scores, *Hotteling* T^2 and *Q* statistics – namely diagnostic plot - and the contribution plot for the identification of variables responsible for a specific behavior/observation being outside the normal operation conditions is quite common [291, 425, 426].

In *X-Metabolomics* it is also possible to explore peaks present in the contribution or the loadings plots, and when selecting each peak, the fingerprint/feature of the selected molecule is reconstructed, as in the example presented in Figure 6.4. As such it allows the molecules identification, by comparing their mass spectra with databases [267] and the Kovats index.

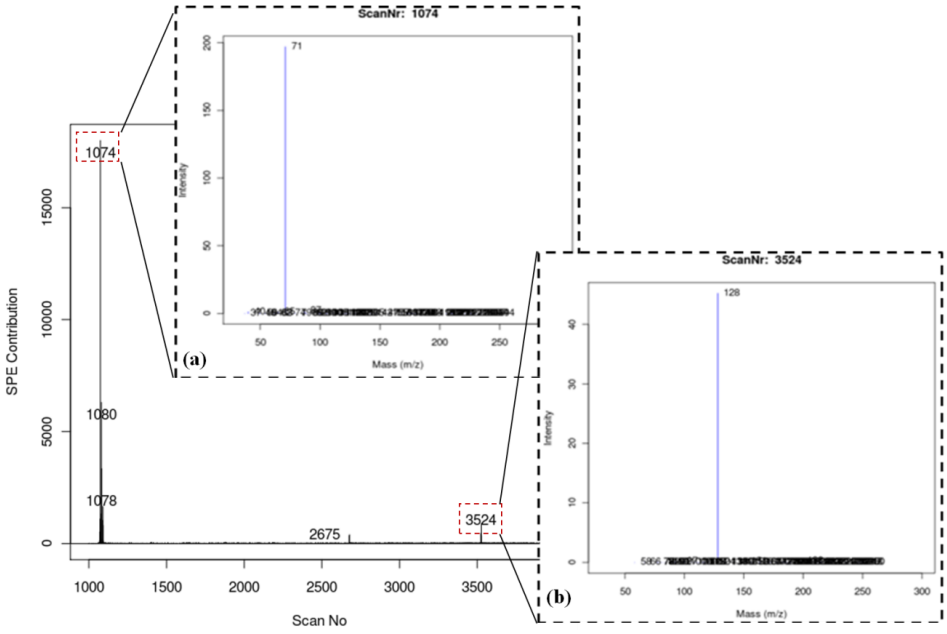


Figure 6.4 Exploration of peaks included in Contribution plot of a given sample.

Metabolites expression through time-course and co-expression

X-Metabolomics produces both the identification and quantification tables. The identification table is produced by imposing the minimum number of fragments, from the original feature, necessary for compounds identification (3 or more) and performing the *Sinkhorn* factorization for checking peak consistency among samples [257, 427]. Only identifiable and coherent peaks which hold the same mass fingerprint among all samples build up the identification table. Data in this table is presented for each candidate compound with corresponding normalized fragments intensity for identification. The composition table includes compounds concentrations obtained directly by linear relation to the internal standard for each sample, taking into account the maximum intensity of one m/z channel.

Using the composition table, the time-expression of each molecule can be explored throughout the biological process or through the samples that we are interested in. The way that the composition table is determined is that by the maximum intensity of one m/z channel, which makes the interpretation less sensitive, and obviously leads to variations of the preprocessed expression when compared to the raw data expression of each molecule. However, the interest in the time-expression is to explore not the exact concentration value in each sample but the tendency of the molecule concentration through samples, which makes this methodology acceptable for our purposes.

The use of temporal correlations between metabolites is essential to understand the overall metabolism in terms of degradation, production or control cycles within the process. These correlations can be used for selecting candidate metabolites, interesting to be studied and identified. So, the co-expression of compounds is obtained by Pearson's correlation coefficient [428] of the composition table. Correlations can further be analyzed in time-course expression, for the correct diagnosis and interpretation of temporal relationships.

Partial least squares regression (PLSR)

PLSR is a statistical tool useful to predict a set of dependent variables from a very large set of independent variables [305]. This methodology is also included in *X-Metabolomics* pipeline, and can be used for predicting the metabolites statistically associated to a given molecule kinetic in the overall process.

6.3 CASE STUDY

6.3.1 Wine material and treatments

Young *Port* wine characterized by a pH = 3.4, 2.5 mg/L of dissolved oxygen, 17 mg/L of free SO_2, 150 g/L of reducing sugars and 20.5% (v/v) alcohol made on the year of the experiment (without any oak contact) was used in this experiment.

Different oxygen treatments were introduced: 0 (P#_NoO2), 1 (P#_1inj), 2 (P#_2inj) and 5 (P#_5inj) saturations in glass vessels filled with 500 mL of *Port* wine. Oxygen saturation was obtained by stirring each sample vigorously for about 1 hour until an oxygen concentration of about 8 to 9 mg/L was reached. This was performed in a laminar flow chamber under UV light to prevent microbial contamination. Oxygen injections were measured with a WTW 340 Oxygen Probe [400], during 18 weeks ('P1' to 'P18') of storage at 60 °C in a temperature controlled incubator, and discrete samples were obtained heuristically for each oxygen regime and further analyzed by GC-MS (see section 2.4 GC-MS analysis, for specifications).

The forced aging experiment [167, 400] was implemented to simulate the typical oxidation aroma of *Port* wine by promoting chemical changes on wine composition. Samples were supplemented with different oxygen regimes and kept at high temperatures (60 °C). Although those extreme conditions are not representative of real aging process, they were selected in order to be able to reproduce in the laboratory on a reasonable time the aging process, in spite of the risk of promoting other chemical reactions which would not occur in the normal process. The length of the duration for the forced aged protocol was sensory-driven. In fact, at each sampling point, samples were submitted to sensory analysis in order to validate that the product was still perceived as *Port* wine and it was observed that after 18 weeks it not accepted as such.

6.3.2 Chemicals

All chemicals employed were of analytical grade: anhydrous sodium sulphate (HPLC grade) (Merck, Darmstadt, Germany), dichloromethane (Lab Scan, Sowinskiego, Gliwice), 3-octanol (97%) (Sigma-Aldrich, USA), *cis*- and *trans*-5-hydroxy-2-methyl-1,3-dioxane (> 99.0%, Sigma-Aldrich, USA), cis-

and trans-4-hydroxymethyl-2-methyl-1,3-dioxolane (> 99.0%, Sigma-Aldrich, USA), benzaldehyde (> 98.0%, Sigma-Aldrich, USA), furfural (> 97.0%, Sigma-Aldrich, USA), 5-hydroxymethylfurfural (5-HMF) (> 98.0%, Sigma-Aldrich, USA); and 3-hydroxy-4,5-dimethyl-2(5H)-furanone (> 99.0%, Sigma-Aldrich, USA).

6.3.3 Volatiles extraction

A liquid-liquid extraction was performed to extract the volatile fraction from each sample. The procedure used was as follows: 5 g of anhydrous sodium sulphate and 50 µL of internal standard - 3-octanol - were added to 50 mL of sample and were extracted twice with 5 mL of dichloromethane using a magnetic stir bar for 5 minutes per extraction, 2 mL of the resulting organic phase were concentrated under a nitrogen stream 4 times [400].

6.3.4 GC-MS analysis

The forced aging experimental protocol was performed in duplicate for practical reason and some samples were analyzed by GC-MS on the replicate trial. These GC-MS targeted analysis were used as a crosscheck procedure on the evaluation and validation of both metabolites features preservation after the preprocessing methodology and the kinetics of each metabolite.

Discrete sample extracts were analyzed using a Varian CP-3800 gas chromatograph (USA) equipped with a Varian Saturn 2000 mass selective detector (USA) and a Saturn GC/MS workstation software version 5.51. The column used was STABILWAX-DA (60 m x 0.25 mm, 0.25 µm) fused silica (Restek, USA). The injector port was heated to 220 °C. The split vent was opened after 30 sec. The carrier gas was Helium C-60 (Gasin, Portugal) at 1 mL/min, constant flow. The oven temperature was 40 °C (for 1 min), then increased at 2 °C/min to 220 °C and held for 30 min. All mass spectra were acquired in the electron impact (EI) mode. The ion trap detector was set as follows: The transfer line, manifold and trap temperatures were respectively 230, 45 and 170 °C. The mass range was 33 to 350 m/z, with a scan rate of 6 scan/sec. The emission current was 50 µA, and the electron multiplier was set in relative mode to auto-tune procedure. The

maximum ionization time was 25,000 µsec, with an ionization storage level of 35 m/z. The injection volume was 1 µL and the analysis was performed in Full Scan mode.

6.3.5 GC-MS data preprocessing

GC-MS ion-trap raw chromatograms were converted to text format (*.txt) using MASSTransit™ (Version 3.0, Palisade Corporation) and imported to *MetAlign*™ software. This preprocessing software was used for baseline correction, accurate mass calculation, data smoothing and noise reduction, followed by a spectral alignment of the GC-MS data [38]. Aligned chromatograms matrix was imported to the *X-Metabolomics* pipeline, and was subjected to multivariate analysis and the identification and quantification tables were built for further obtain the expression and co-expression of candidate molecules within the "forced aging" *Port* wine process.

6.3.6 Multivariate analysis (MVA)

After RPCA performing, the diagnostic and contribution plots were used together for the selection of samples standing out of the latent model according to their metabolic characteristics, in response to the external conditions imposed, and to analyze variables/candidate metabolites, to be responsible for these differences during the *Port* wine "forced aging" process.

6.3.7 Metabolites identification and quantification

Both the identification and composition tables are constructed in *X-Metabolomics*, using the aligned matrix (*.csv) derived from *MetAlign*™ preprocessing software [38]. These tables allow identifying variables, that is, candidate metabolites which are responsible for the deviations of samples in the diagnostic plot.

The identification table was built using the discrimination of the m/z channels and the corresponding normalized intensity in each scan. This was executed by imposing the minimum number of fragments, from the original feature, necessary for compounds identification (3 or more) and performing the *Sinkhorn* factorization for checking peak consistency among samples [257,

427]. Only identifiable and coherent peaks which hold the same mass fingerprint, among all samples build up the identification table, are used for a correct identification of molecules by comparing their fingerprint with databases [267].

The composition table was built using compounds concentrations obtained directly by linear relation to the internal standard for each sample (in this case, 3-octanol), taking into account the maximum intensity of one m/z channel.

6.3.8 Metabolites kinetics and co-expression in time-course

The temporal variations and/or interactions between metabolites reproduce the cellular dynamics about the chemical or biochemical systems in response to the environmental conditions. The kinetics or time-expression of each candidate molecule was directly obtained from the composition table, which was determined by the maximum intensity of one m/z channel. This fact can make the interpretation less sensitive, leading to variations of the preprocessed expression when compared to the raw data expression of each molecule. However, the main interest in the time-expression is to explore the tendency of the molecule concentration through samples.

The co-expression of compounds is obtained by Pearson's correlation coefficient [429] of the composition table. Correlations between the kinetics of candidates and a key metabolite, known to be involved in a specific chemical phenomenon in the process, are displayed in a heatmap (performed in R-Project 2.15.0), in a color range from white-to-red, which correspond to correlations between $0.8 \leq R^2 \leq 1$. Thus, clusters from the heatmap can be used for interpretation of temporal relationships of candidate metabolites allowing process contextualization, identification of the involved metabolites and metabolic network reconstruction [28].

6.4 RESULTS AND DISCUSSION

6.4.1 Robustness of GC-MS preprocessing using *MetAlign*™

Due to *multi-scale* nature of *Port* wine chromatograms matrix, operational parameters of *MetAlign*™ [38] preprocessing software tool, need to be optimized in order to provide simultaneous extraction

of high and low concentration compounds. The optimization was performed according to the analytical equipment characteristics, and taking into account the preservation of the original features of a sub-set of volatiles. Volatiles were selected to monitor the impact of the different algebraic treatments in each metabolite spectra fingerprint within data processing. The criteria for metabolites selection were: *(i)* volatiles recognized as products of oxidation, *Maillard* reaction products and combination of both; *(ii)* substances present at different scales (*i.e. ppm-ppt*) and simultaneously acquired on the same analytical run.

The selected volatiles were: the heterocyclic acetals of glycerol – dioxolane 1 and dioxolane 2 (isomers *cis*- and *trans*-4-hydroxymethyl-2-methyl-1,3-dioxolane) - the furanic aldehydes: furfural, 5-HMF and phenylacetaldehyde, a *Strecker* aldehyde [430], dependent of oxygen and temperature during aging.

The original features of each metabolite were compared with the preprocessed resulting features in order to ensure the validity of the preprocessing methodology. Table 6.1 presents the correlation coefficients between features of both raw data and *MetAlign*™ [38]. These features can be used as a validation index for the preprocessing software tool preserving the original chromatographic features. The high correlation coefficients (R^2) are presented in Table 6.1, indicating the preprocessing parameters used to maintain the integrity of the overall matrix in the forced aging of *Port* wine. As consequence, the features of the selected compounds with different magnitudes (*ppm-ppt*) were preserved, showing the strength of the preprocessed matrix which can be used for understanding chemical compound interaction in the *Port* wine "forced aging" process.

Table 6.1 Correlation coefficients between raw and *MetAlign*™ data features using quantifier ion of the selected metabolites: (a) Quantifier ion; (b) Normalized value

Metabolite [b]	m/z [a]	R^2
Dioxolane 1	103	0.9242
Dioxolane 2	103	0.9200
Furfural	95	0.9700
5-HMF	97	0.7120
Phenylacetaldehyde	91	0.9564

6.4.2 Multivariate analysis (MVA)

Performing the RPCA to the *MetAlign*™ [38] final matrix, four relevant orthogonal decompositions were found to explain a total of 63.2% of the chromatographic data variability (22.3% PC1, 21.3% PC2, 11.7% PC3 and 7.9% PC4). The most relevant components for capturing samples variability within the oxidation under each oxygen saturation regime are PC1 and PC2. According to both relevant decompositions, samples are distributed through the components as a function of the oxidation process occurring in the *Port* wine, which induces chemical composition changes in samples (Figure 6.5). Chemical differences occurring in the beginning of the "forced aging" process are less representative than those taking place in the final weeks (samples closest to week 18 – "P18") of the oxidation process as samples corresponding to these times are spread through the PC2 and PC1, respectively and the variability of the chemical information is higher in PC1. So, it can be stated that the complexity of the chemical profile of the resulting *Port* wine made under the different oxygen saturation regimes at high temperature increases during the maturation process leading to the production of compounds of during the storage time.

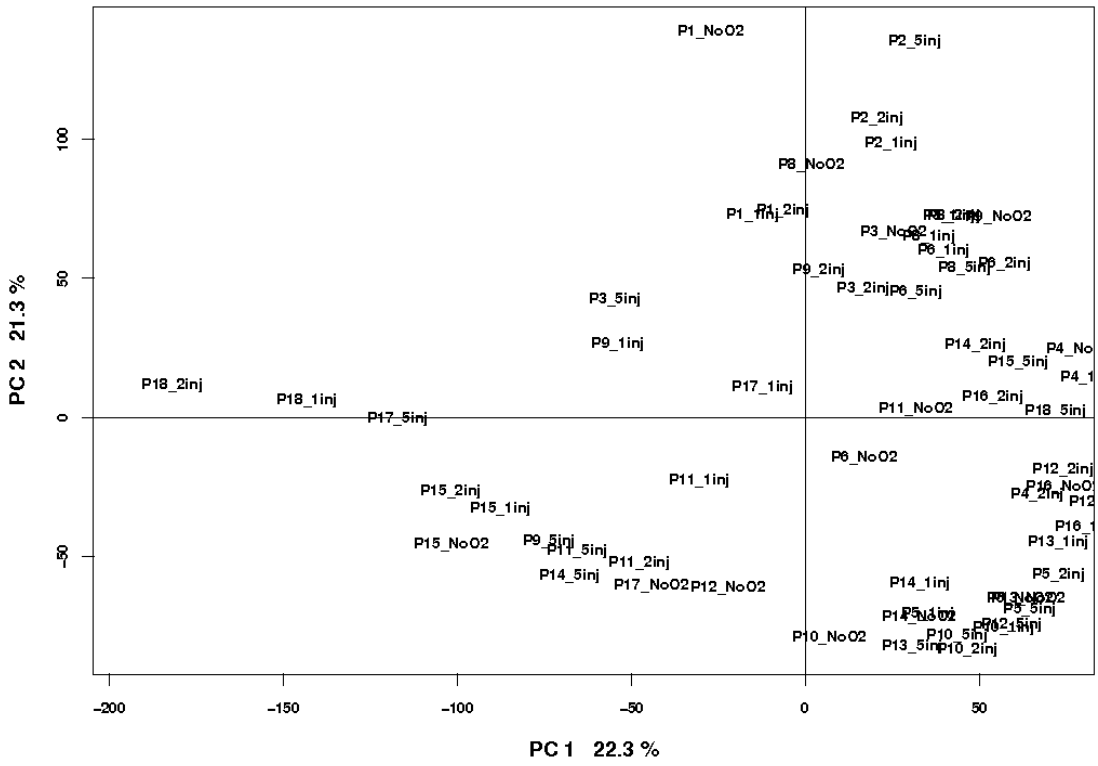


Figure 6.5 Scores resulting from relevant principal component analysis.

6.4.3 Diagnostic and Contribution Plots – Selection of candidate molecules

To understand the role of dissolved-oxygen regime and high temperatures during *Port* wine "forced aging" process, the selection of candidate metabolites was performed by employing diagnostic (*Q* statistics and *Hotelling T²*) and contribution plots, which are presented in Figure 6.6 and Figure 6.7, respectively [425] [431, 432]. According to the Diagnostic plot in Figure 6.6, sample P18_5inj is above the Qα, which means that the reconstruction of its metabolic information is statistically impossible.

In practice, P18_5inj was exposed to extreme conditions (high temperature, 60°C, under a saturation regime of 5 injections of O_2) for 18 weeks, and its sensorial analysis highlights an unconformity of its chemical profile comparing to other samples recognized as *Port* wine. So, as the chemical information of the sample reflects the impact of the extreme conditions during storage, and it cannot be accepted as a normal *Port* wine, this was used as a control for monitoring the aging process of these wines.

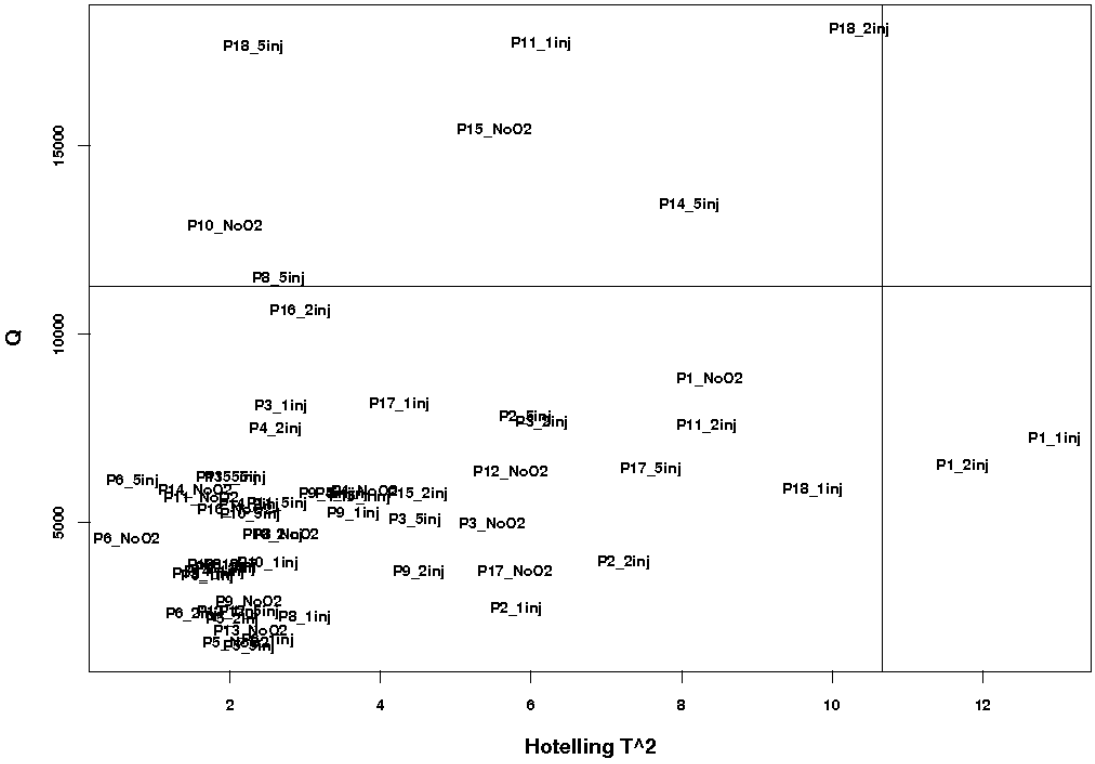


Figure 6.6 Diagnostic plot (Hotteling T^2-Q statistics) of Multivariate Analysis model.

In contribution plot of Figure 6.7 are visible the candidate variables/scans (1 to 14) representing metabolites that mostly explain chemical differences from the latent model. These candidates addresses the chemical quality of the final *Port* wine in the presence of the higher levels of oxygen after 18 weeks of "forced aging" at high temperature (60°C). The contextualization of these candidate molecules through the different chemical pathways involved in the "forced aging" of *Port* wine entails their correct identification according to each mass spectra fingerprint, being essential for further comprehension of the temporal relationships with the entire matrix hich constitutes the aromatic profile of the *Port* wine samples.

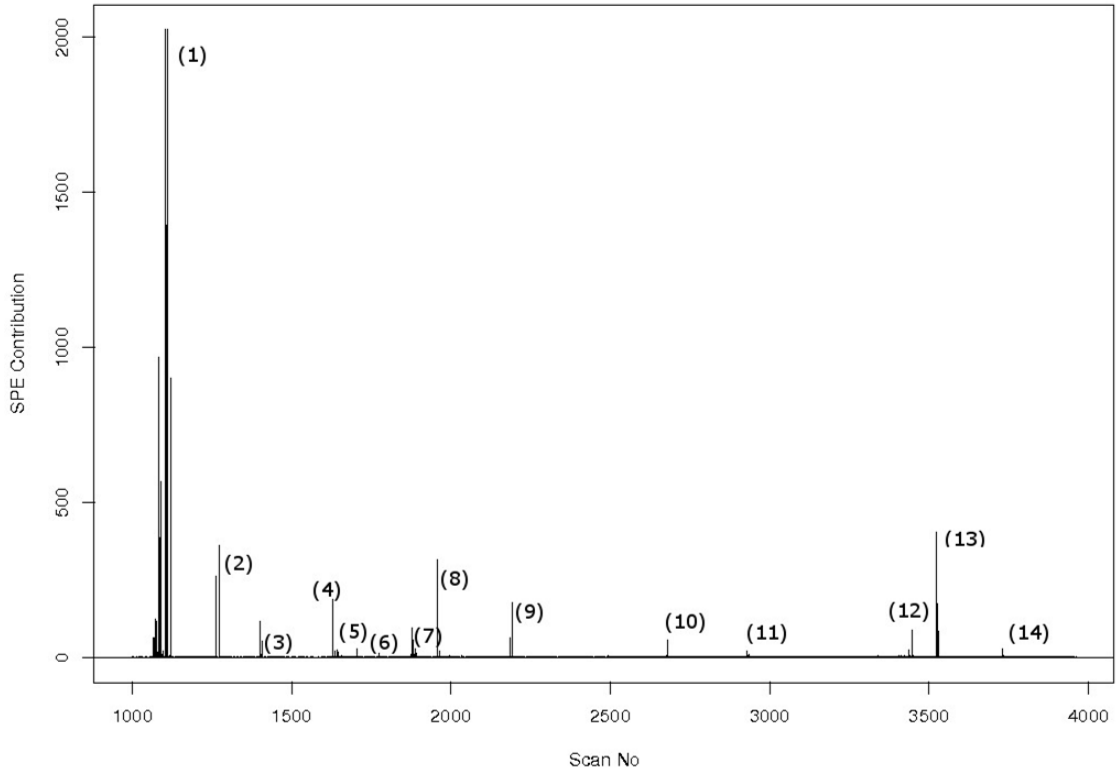


Figure 6.7 Contribution plot of sample "P18_5inj" with relevant candidate scans highlighted.

6.4.4 Candidate metabolites identification and time-expression

In the present pipeline methodology, mass spectra features from the candidate metabolites selected in the contribution plot are validated by *i)* crosschecking their presence in the raw data; *ii)* comparing their features with those present in NIST 98 MS library [267] and; *iii)* time-expression of each metabolite throughout the maturation process under each condition.

The mass spectra (MS) fingerprints of the highlighted metabolites of Figure 6.7 were built using the identification table from the data matrix resulting from *MetAlign*™ preprocessing and were used for tentatively identifying the key metabolites [m/z (%)] that mostly influence the *Port* wine aging process. MS features of metabolites tentatively identified as unknown are presented: 1 – 3-methyl-1-butanol; 2 – unknown: unkn_1 [(100); 43 (67); 55(15); 88 (7); 59 (7)] – Scan 1274 ; 3 – unknown: unkn_2 [(100); 43 (21); 91 (13); 119 (12); 75 (5); 59 (5)] – Scan 1403; 4 – unknown: unkn_3 [43 (100); 115 (46); 55 (41); 45 (30); 59 (29); 73 (21); 39 (21); 67 (19)] – Scan 1629; 5 – furfural; 6 – dioxane isomer; 7 – dioxolane isomer; 8 – unknown: unkn_4 [73 (100); 45 (33); 91 (15); 43 (14); 55 (7); 74 (6); 57 (4)] – Scan 1959; 9 - unknown: unkn_5 [(100); 93 (91); 59 (90); 45 (83); 121 (80); 136 (70)] – Scan 2216; 10 – 2-phenylethanol; 11 – diethyl malate; 12 – diethyl tartrate; 13 – monoethyl succinate; 14 – 5-HMF.

The presence of furfural, dioxane and dioxolane isomers within the "*forced aging*" of *Port* wine is indicative of the occurrence of wine oxidation during the process, as it was reported by [73] and [167]. According to time-course kinetics of furfural and dioxane isomer, presented in Figure 6.8A and Figure 6.8 B, respectively, concentrations of both metabolites increase during "forced aging". Furfural and 5-hydroxymethylfurfural are derived from carbohydrate dehydration followed by cyclation in *Maillard*-type systems and are generally correlated to wine browning during aging [433]. It is also evident that higher concentrations of oxygen during the maturation process induce higher production of this heterocyclic acetal (Figure 6.8B) and so, the formation of these off-flavors is due to the oxidative degradation of the wine [24].

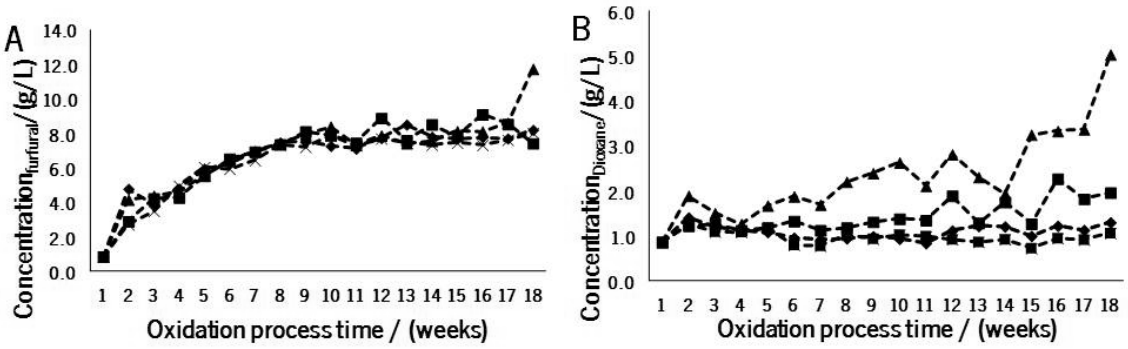


Figure 6.8 Time-course kinetics of the highlighted candidate metabolites: A - furfural; B - *cis*- and *trans*-5-hydroxy-2-methyl-1,3-dioxane; (×) No O2; (♦) 1 Injection O2; (■) 2 Injection O2; (▲) 5 Injection O2.

The presence of 2-phenylethanol and 3-methyl-1-butanol at high concentrations is mainly due to the occurrence of the ethanol fermentation and widely contribute to the aromatic profile of *Port* wine [4, 5]. Another noteworthy result is the high impact of diethyl esters (diethyl tartrate and diethyl malate) as well as the monoethyl succinate in the *Port* wine 'forced' aging, which concerns to the extended esterification reactions occurring through the process [434].

So, according to the chemical background of sample "P18_5inj", translated with the variables or candidate metabolites identified in the contribution plot, it is possible to analyze that the presence of such extreme conditions (18 weeks of storage under the highest oxygen saturation regime studied) induces the occurrence of redox mechanisms which are closely associated to the "oxidative spoilage" of the wine during the maturation storage [167]. This mechanism, and thus the presence of the candidate molecules identified, is in most cases associated with the loss of the original *Port* wine aromas, the development of unpleasant aromas as well as with changes in the *Port* wine color or taste [73]. For these reasons, it is possible to consider that "P18_5inj" sample is not representative of a typical aged *Port* wine by the conventional aging process, since the chemical composition evidences the occurrence of deterioration mechanisms inherent to the aging process occurring under drastic conditions.

Furthermore, regarding the correlation between candidate molecules kinetics over time is an innovative approach that allows increasing the knowledge of the overall process, providing the co-expression, enhancing time-course relationships of the highlighted metabolites and other relevant molecules. The co-expression between compounds allows a kinetical comparison enabling correlations discovery between them, facilitating "mechanistic contextualization".

6.4.5 Candidate metabolites co-expression

Dioxane isomer, was reported to be one of the indicators of the *Port* wine aging [73], it was used to understand the co-expression of other compounds present in the overall metabolites matrix.

The exploration of dioxane kinetical correlations, allows understanding its temporal relationships with other metabolites for further contextualize them within a chemical phenomenon that can be being activated through the process, which is essential for further metabolic network reconstruction.

Scans corresponding to metabolites with correlation coefficients similar or higher than 0.8 ($R^2 \geq 0.8$) according to dioxane isomer are present in the heatmap of Figure 6.9. Features from scans included in the heatmap of Figure 6.9 were also tentatively identified, by supervising with the raw-data chromatogram features and also by comparing the mass feature with NIST 98 MS library [267], and using the time-expression of scans discriminated in the heatmap.

So, according to Figure 6.8, it was found that the production of dioxane isomer, is highly correlated with the production of dioxolane isomers – dioxolane 1 (R^2 = 0.98), dioxolane 2 (R^2 = 0.99), benzaldehyde (R^2 = 0.89) and sotolon (R^2 = 0.81). The high correlations between these compounds is in agreement with the reported bibliography, as all these metabolites are known to be related with the *Port* wine aging process [1, 72, 73]. Dioxane isomer production is a result of the condensation reaction between glycerol and acetaldehyde under the low pH during wine aging [435].

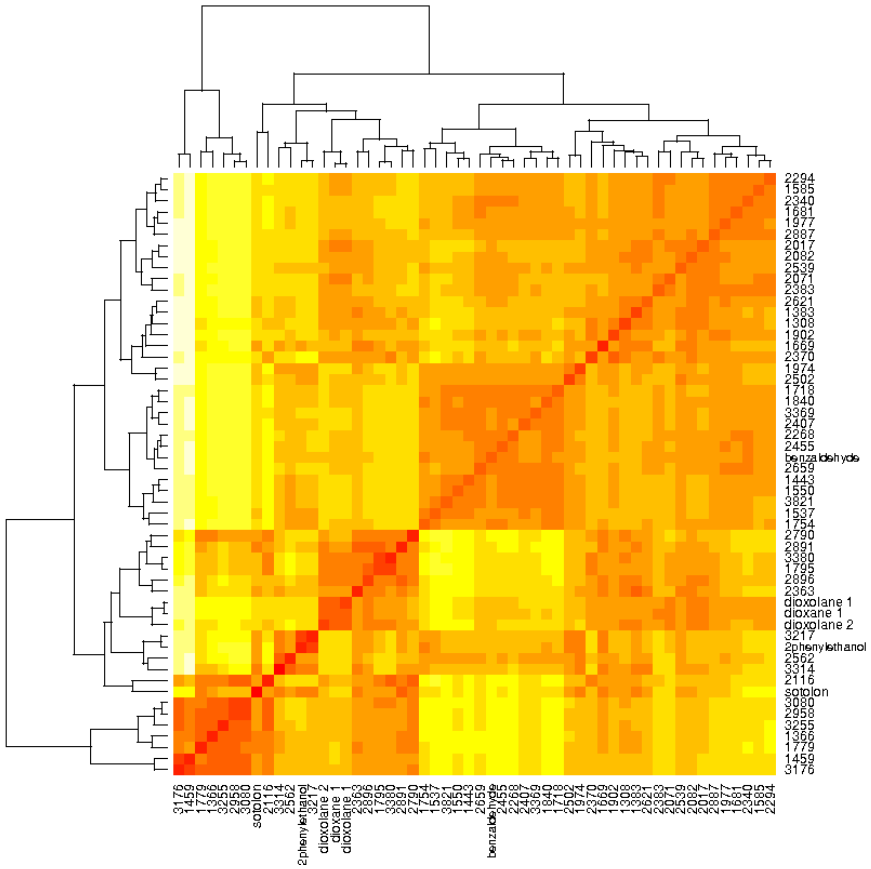


Figure 6.9 Heatmap matrix correlations (0.8 ≤ R2 ≤ 1 corresponding to white to red range colors) with dioxane isomer.

6.5 CONCLUSIONS

This research presented a metabolic pipeline methodology, implemented in the in-house platform *X-Metabolomics*, which allows identifying metabolites involved in dynamic processes in a high-throughput mode. This pipeline enables the elucidation about the network reaction mechanisms and pathways relationships, starting with a set of known compounds clustered together according to their kinetical correlations and representing a specific chemical pathway.

Applying the present pipeline to a *Port* wine "forced aging" process, furfural and dioxane isomer were found to be key molecules involved in the process, which production is relevant for the quality of the *Port* wine. Moreover, dioxane isomer production is highly correlated with dioxolane isomers, dioxane isomer, benzaldehyde and sotolon production, which have been already reported in other studies, as being result of physical and chemical changes occurring over the wine aging process.

Temporal relationships between well known compounds and the matrix chromatograms, it was possible to understand potential interconnections between the candidates metabolites formation. Future researches of connections between chemical and/or biochemical mechanisms would provide further information about pathways and their reconstruction.

This study revealed several advantages leading to the use of the metabolomics pipeline methodology, encouraging also its application in different research fields, as it provides a greater, faster and reliable understanding of biological or chemical processes occurring through time-course.

CHAPTER 7

GENERAL CONCLUSIONS AND FINAL REMARKS

This chapter presents the concluding remarks and the main outcomes of this thesis.

7.1 GENERAL CONCLUSIONS

In this thesis, the potential of technometric approaches, combining chemometric and bioinformatic methodologies, was explored in the characterization, classification and prediction of physiological, phenotypic and metabolic changes as an adaptation response of different strains of *Saccharomyces cerevisiae* to environmental conditions. Methodologies used for data characterization include relevant principal component analysis (RPCA) and multi-way principal component analysis (MPCA), whereas for prediction and classification the methodologies used were unfolded partial least squares (U-PLS) and partial least squares logistic regression (PLS-LOG). The combined effect of distinct variables (measured using HPLC, GC-FID, GC-MS and cyclic voltammetry) explored by multivariate data analysis, allowed enhancing the knowledge about chemical and biochemical dynamics in biotechnological processes.

It was found that the physiological, phenotypic and metabolic responses of three different strains of *S. cerevisiae* (S288c, CA11 and PE-2) triggered by the presence of toxic molecules, such as 1-butanol, are different from those obtained in the presence of inhibitory molecules, as furfural and 5-HMF. PE-2 was found to be the most robust strain, able to resist under toxic and inhibitory conditions in YPD_b batch fermentations. The flocculation profile of CA11 was also found to be correlated with the production of glycerol and trehalose, as well as with acetic acid, as a response to the induced stress conditions. The use of chemometric tools, such as RPCA and PLS-LOG, proved to be extremely suitable in the characterization, classification, and prediction of physiological, phenotypic and metabolic responses of *S. cerevisiae.*

In synthetic wine fermentations, it was found that changes occurring in the electro-active compounds variations in the medium are linked to antioxidant capacity response of *S. cerevisiae* PYCC 4653, which induced also metabolic changes. The use of the bioanalytical pipeline combining the electrochemical signal with the target and non-target metabolomics, using multivariate analysis strategies, allowed to understand that the antioxidant capacity variations were accompanied by the production of important metabolites, such as 2-phenylethanol, acetic acid and isoamyl acetate. The use of the prediction methodology called U-PLSR allowed to find out metabolites, which kinetics were well correlated with the production of 2-phenylethanol within the fermentation, such as, acetic acid, 2,3-dihydro-3,5,dihydroxy-6-methyl-4(H)-pyran-4-one and 5-

HMF. It was also found that *S. cerevisiae* was able to produce chlorogenic acid in synthetic medium.

In order to increase the capacity of gathering the maximum metabolic information about a biological system, by exploring the complexity of a given biochemical process, a metabolomics pipeline for high-throughput data analysis was developed. Bioinformatic and chemometric tools were integrated in a new software, *X-Metabolomics*, developed during this thesis, which was built for increasing the information of metabolic changes occurring in response to environmental conditions.

The pipeline was applied to *Port* wine, and kinetics of known metabolites and candidate metabolites were found. Temporal relationships between well-known metabolites and the entire matrix, allowed to understand potential interconnections between the candidates metabolites formation. Future researches of connections between chemical and/or biochemical mechanisms would provide chemical responses according to the interest on a specific process.

The increment provided by the application of this pipeline in metabolites identification, encourages its application in different research fields, as it provides a greater, faster and reliable understanding of biological or chemical processes along time.

In sum, this thesis revealed the suitability of technometrics approaches, including different chemometrics and bioinformatics tools, for the monitorization and characterization of yeasts behavior during fermentation. The increment of the pipeline in the candidate metabolites identification encourages the application of this approach in different research fields, as they provide a greater, faster and reliable understanding of biological or chemical changes occurring on biotechnological processes.

7.2 SUGGESTIONS FOR FUTURE WORK

This thesis is inserted in the pioneer area of developing methods, sensors and software for the holistic monitoring of biological systems. Although this work allowed to validate the suitability of technometric approaches, combining different methodologies, for monitoring yeasts behavior as an adaptation to the environmental conditions within fermentations, there are still some hurdles that are necessary to solve, which include:

- The use of control systems for monitoring fermentations "*at-line",* that provide rapid and accurate responses about the metabolic and physiological behavior of cells inside bioreactors. These systems could provide sampling representativeness and miniaturization of fermentative reactors for high-throughput screening;
- The use of different smaller size sensors, with better separation columns and ionization control, as well as, faster scan methods;
- Development of improved analytical methods for compounds extraction, contaminants detection and minimization, providing enhanced time/space resolution and sample size;
- The development simplified, generalized and robust methods for chromatograms processing. None of state-of-the-art approaches are fully optimized for automatic processing of chromatograms without analytical chemist supervision. Significant errors occur in peak extraction, correct alignment, deconvolution and the accurate identification of compounds. Furthermore, as mass spectroscopy technologies and methods are very diverse, software has been dedicated directly to just some standard applications, and even so, these are complicated to operate and to optimize processing variables;
- Development of new methodologies, especially in the integration of time-course data with pattern recognition and pathway networks reconstruction, so that, the underlying biochemistry present in high-throughput data can be correctly put into the biological context and not as a pure classification study.

CHAPTER 8

REFERENCES

This Chapter lists all the references that contributed to the elaboration of this written work.

1 Silva Ferreira AC, J Monteiro, C Oliveira and P Guedes de Pinho (2008) Study of major aromatic compounds in port wines from carotenoid degradation, *Food chemistry*, 110 (1): p. 83-87.

2 Verstrepen K, G Derdelinckx, H Verachtert and F Delvaux (2003) Yeast flocculation: what brewers should know, *Applied microbiology and biotechnology*, 61 (3): p. 197-205.

3 Swiegers JH and IS Pretorius (2005) Yeast modulation of wine flavor, *Advances in applied microbiology*, 57: p. 131-175.

4 Ehrlich F (1907) Über die Bedingungen der Fuselölbildung und über ihren Zusammenhang mit dem Eiweissaufbau der Hefe, *Berichte der deutschen chemischen Gesellschaft*, 40 (1): p. 1027-1047.

5 Hazelwood LA, JM Daran, AJ van Maris, JT Pronk and JR Dickinson (2008) The Ehrlich pathway for fusel alcohol production: a century of research on *Saccharomyces cerevisiae* metabolism, *Appl Environ Microbiol*, 74 (8): p. 2259-66.

6 Gibson BS, SJ Lawrence, JP Leclaire, CD Powell and KA Smart (2007) Yeast responses to stresses associated with industrial brewery handling, *FEMS Microbiology Review*, 31: p. 535-69.

7 Zhao XQ and FW Bai (2009) Yeast flocculation: new story in fuel ethanol production, *Biotechnology Advances*, 27 (6): p. 849-856.

8 Gibson BR, SJ Lawrence, JPR Leclaire, CD Powell and KA Smart (2007) Yeast response to stresses associated with industrial brewery handling, *FEMS Microbiol*, 31: p. 535-569.

9 Hensing MCM, RJ Rouwenhorst, JJ Heijnen, JP van Dijken and JT Pronk (1995) Physiological and technological aspects of large-scale heterologous-protein production with yeasts, *Antonie van Leeuwenhoek*, 67 (261-279).

10 Urtubia A, G Hernández and JM Roger (2012) Detection of abnormal fermentations in wine process by multivariate statistics and pattern recognition techniques, *Journal of Biotechnology*, 159: p. 336-341.

11 Kell DB, M Brown, HM Davey, WB Dunn, I Spasic and SG Oliver (2005) Metabolic footprinting and systems biology: the medium is the message, *Nature Reviews Microbiology*, 3 (7): p. 557-565.

12 Dunn WB, D Broadhurst, P Begley, E Zelena, S Francis-McIntyre, N Anderson, M Brown, JD Knowles, A Halsall, JN Haselden, AW Nicholls, ID Wilson, DB Kell and R Goodacre (2011) Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry, *Nat Protoc*, 6 (7): p. 1060-83.

13 Chevion S, MA Roberts and M Chevion (2000) The use of cyclic voltammetry for the evaluation of antioxidant capacity, *Free Radical Biology & Medicine*, 8 (6): p. 860-870.

14 Martins RC, R Oliveira, F Bento, D Geraldo, VV Lopes, P Guedes-de-Pinho, CM Oliveira and AC Silva-Ferreira (2008) Oxidation management of white wines using cyclic voltammetry and

multivariate process monitoring, *Journal of Agricultural and Food Chemistry*, 56 (24): p. 12092-12098.

15 Steuer R (2006) On the analysis and interpretaton of correlations in metabolomic data, *Briefings in Bioinformatics*, 7 (3): p. 151-158.

16 Wishart DS (2008) Quantitative metabolomics using NMR, *TrAC Trends in Analytical Chemistry*, 27 (3): p. 228-237.

17 Tikunov Y, A Lommen, CH de Vos, HA Verhoeven, RJ Bino, RD Hall and AG Bovy (2005) A novel approach for nontargeted data analysis for metabolomics. Large-scale profiling of tomato fruit volatiles, *Plant Physiol*, 139 (3): p. 1125-37.

18 Smith CA, EJ Want, G O'Maille, R Abagyan and G Siuzdak (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification, *Anal Chem*, 78 (3): p. 779-87.

19 Luedemann A, K Strassburg, A Erban and J Kopka (2008) TagFinder for the quantitative analysis of gas chromatography–mass spectrometry (GC-MS)-based metabolite profiling experiments, *Bioinformatics*, 24 (5): p. 732-7.

20 McGraw-Hill (2002) *in Encyclopedia of Science and Technology*, McGraw-Hill, Editor: New York.

21 Bauer FF and IS Pretorius (2000) Yeast stress response and fermentation eficiency: how to survive the making of wine - a review, *S. Afr. J. Enol. Vitic*, 21.

22 Váchová L, M Cáp and Z Palková (2012) Yeast colonies: a model for studies of aging, environmental adaptation, and longevity, *Hindawi Publishing Corporation: Oxidative Medicine and Cellular Longevity*, 2012: p. 1-8.

23 Samuel D (1996) Investigation of ancient Egyptian baking and brewing methods by correlative microscopy, *Science*, 273: p. 488-490.

24 Silva-Ferreira A, P Guedes-de-Pinho, P Rodrigues and T Hogg (2002) Kinetics of oxidative degradation of white wines and how they are affected by selected technological parameters, *Journal of Agricultural and Food Chemistry*, 50 (21): p. 5919–5924.

25 du Toit WJ, J Marais, IS Pretorius and M du Toit (2006) Oxygen in must and wine: a review, *South African Journal of Enoloy and Viticulture*, 27 (1): p. 76-94.

26 Palková Z and L Váchová (2006) Life within a community: benefit to yeast long-term survival, *FEMS microbiology reviews*, 30 (5): p. 806-824.

27 Attfield PV, A Raman and CJ Northcott (1992) Construction of Saccharomyces cerevisiae strains that accumulate relatively low concentrations of trehalose, and their application in testing the contribution of the disaccharide to stress tolerance, *FEMS microbiology letters*, 94 (3): p. 271-276.

28 Feist AM, MJ Herrgård, I Thiele, JL Reed and BØ Palsson (2009) Reconstruction of biochemical networks in microorganisms, *Nature Reviews Microbiology*, 7 (2): p. 129-143.

29 Daszykowski M and B Walczak (2006) Use and abuse of chemometrics in chromatography, *Trends Analyt Chem*, 25: p. 1081-1096.

30 Dettmer K, PA Aronov and BD Hammock (2007) Mass spectrometry-based metabolomics, *Mass Spectrometry Reviews*, 26 (1): p. 51-78.

31 Bino RJ, RD Hall, O Fiehn, J Kopka, K Saito, J Draper, BJ Nikolau, P Mendes, U Roessner-Tunali and MH Beale (2004) Potential of metabolomics as a functional genomics tool, *Trends in Plant Science*, 9: p. 418-425.

32 Fiehn O (2002) Metabolomics - the link between genotypes and phenotypes, *Plant molecular biology*, 48 (1): p. 155-171.

33 Matsuda F, Y Shinbo, A Oikawa, MY Hirai, O Fiehn, S Kanaya and K Saito (2009) Assessment of metabolome annotation quality: a method for evaluating the false discovery rate of elemental composition searches, *PLoS ONE*, 4 (7490).

34 Abdi H and LJ Williams (2010) Principal component analysis, *Wiley Interdisciplinary Reviews: Computational Statistics*, 2: p. 433-459.

35 Abdi H (2003) Partial Least Squares (PLS) Regression, *in Encyclopedia of Social Sciences Research Methods.*, M. Lewis-Beck, A. BrymanandT. Futing, Editors., Sage: Thousand Oaks (CA).

36 Goodacre R, S Vaidyanathan, WB Dunn, GG Harrigan and DB Kell (2004) Metabolomics by numbers: acquiring and understanding global metabolite data, *Trends in Biotechnology*, 22 (5): p. 245-252.

37 Villas-Bôas SG, S Rasmussen and GA Lane (2005) Metabolomics or metabolite profile?, *Trends Biotechnol*, 23 (8): p. 385-386.

38 Lommen A (2009) MetAlign: interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing, *Anal Chem*, 81 (8): p. 3079-86.

39 Feldmann H (2012) Yeast Metabolism, *in Yeast: Molecular and Cell Biology*, H. Feldmann, Editor., Wiley-VCH Verlag GmbH & Co. KGaA.

40 German JB, BD Hammock and SM Watkins (2005) Metabolomics: building on a century of biochemistry to guide human health, *Metabolomics*, 1 (1): p. 3-9.

41 Danø S (1999) *Glycolitic oscillations in yeast cells: The nonlinear dynamics of metabolism studied in living yeast cells*, *in Department of Chemistry*, H.C.Ørsted Institute of University of Copenhagen: Copenhagen.

42 Bekatorou A, C Psarianos and AA Koutinas (2006) Production of food grade yeasts, *Food Technol Biotechnol*, 44 (3): p. 407-415.

43 Ugliano M and PA Henschke (2009) Yeasts and wine flavour, *in Wine Chemistry and Biochemistry*, M. V. a. P. Moreno-Arribas, M.C., Editor., Springer: Spain.

44 Lambrechts MG and IS Pretorius (2000) Yeast and its importance to wine aroma*, South African Journal of Enoloy and Viticulture*, 21 (97-129).

45 Guinard J-X, M Miranda and MJ Lewis (1989) *Yeast biology and beer fermentation*, *in Zymurgy*. p. 14.

46 Fix GJ and LA Fix (1997) An analysis of brewing techniques: Brewers Publications.

47 Shuler ML and F Kargi (2002) Bioprocess engineering basic concepts. 2nd ed, Upper Saddle River, NJ: Prentice Hall.

48 Chen X (2011) *Metabolic engineering of Saccharomyces cerevisiae amino acid metabolism for production of chemicals of industrial interest*, Technical University of Denmark.

49 Torija MJ, N Rozès, M Poblet, JM Guillamón and A Mas (2002) Effects of fermentation temperature on the strain population of *Saccharomyces cerevisiae, International Journal of Food Microbiology*, 80: p. 47-53.

50 Fleet GH and GM Heard (1993) Yeasts: growth during fermentation, *in Wine Microbiology and Biotechnology*, G. M. Fleet, Editor., Harwood Academic Publishers: Chur, Switzerland. p. 22-54.

51 Carrau F, K Medina, L Fariña, E Boido and E Dellacassa (2010) Effect of *Saccharomyces cerevisiae* inoculum size on wine fermentation aroma compounds and its relation with assimilable nitrogen content*, International Journal of Food Microbiology*, 143 (1-20): p. 81-85.

52 Mingorance-Cazorla L, JM Clemente-Jiménez, S Martínez-Rodríguez and FJ Las Heras-Vázquez (2003) Contribution of different natural yeasts to the aroma of two alcoholic beverages*, World Journal of Microbiology & Biotechnology*, 19: p. 297-304.

53 Barnett JA (1976) The utilization of sugars by yeasts*, Advances in Carbohidrate Chemistry and Biochemistry*, 32: p. 125-234.

54 da Silva Filho EA, HF de Melo, DF Antunes, SKB dos Santos, AM Resende, DA Simões and MA de Morais Jr (2005) Isolation by genetic and physiologycal characteristics of a fuel-ethanol fermentative Saccharomyces cerevisiae strain with potential for genetic manipulation*, Journal of Industrial Microbiology and Biotechnology*, 32 (10): p. 481-486.

55 Simon JA and A Bedalov (2004) Yeast as a model system for anticancer drug discovery*, Nature Reviews Cancer*, 4: p. 481-187.

56 Hogan D (2006) Quorum sensing: alcohols in a social situation*, Current Biology*, 16 (2): p. 447-458.

57 Dickinson JR (2008) Filament formation in *Saccharomyces cerevisiae* - a review*, Folia Microbiology*, 53 (1): p. 3.

58 Guidetti R, R Beghi and V Giovenzana (2012) Chemometrics in food technology, *in Chemometrics in practical applications*, K. Varmuza, Editor., InTech.

59 Liang Y and L Yi (2010) The Basis of Chemometrics, Shanghai, China: East China University of Science and Technology Press.

60 Schreier P (1979) Flavour composition of wines*, CRC Critical Reviews in Food Science and Nutrition*: p. 59-111.

61 Escudero A, E Campo, L Fariña, J Cacho and V Ferreira (2007) Analytical characterization of the aroma of five premium red wines: insights into the role of odor families and the concept of frutiness of wines*, Journal of Agricultural and Food Chemistry*, 55: p. 4501-4510.

62 Gancedo C, JM Gancedo and A Sols (1968) Glycerol metabolism in yeast. Pathways of utilization and production*, European Journal of Biochemistry*, 5: p. 165-172.

63 Blake A, Y Kotseridis, ID Brindle, D Inglis, M Sears and GJ Pickering (2009) Effect of closure and packaging type on 3-alkyl-2-methoxypyrazines and other impact odorants of Riesling and Cabernet Franc wines*, Journal of Agricultural and Food Chemistry*, 57: p. 4680-4690.

64 Didion T, M Grauslund, MC Kielland Brandt and HA Anderson (1996) Amino acids induce expression of *BAP2*, a branched-chain amino acid permease gene in *Saccharomyces cerevisiae, FEBS Letters*, 326: p. 29-32.

65 Derrick S and PJ Large (1993) Activities of the enzymes of the Ehrlich pathway and formation of branched-chain alcohols in *Saccharomyces cerevisiae* and *Candia utilis* grown in continuous-culture on valine or ammonium as sole nitrogen-source*, Journal of General Microbiology*, 139: p. 2783-2792.

66 Fowles GWA (1992) Acids in grapes and wines: a review*, Journal of Wine Research*, 3: p. 25-41.

67 Mason AB and J-P Dufour (2000) Alcohol acetyltransferases and the significance of ester synthesis in yeast*, Yeast*, 16: p. 1287-1298.

68 Díaz-Maroto C, R Scheider and R Baumes (2005) Formation pathways of ethyl esters of branched short-chain fatty acids during wine aing*, Journal of Agricultural and Food Chemistry*, 53: p. 3503-3509.

69 Briggs DE, CA Boulton, PA Brookes and R Stevens (2004) Brewing: science and practice, *ed. Woodhead*, Cambridge, UK.

70 Nicolaou SA, SM Gaida and ET Papoutsakis (2010) A comparative view of metabolite and substrate stress and tolerance in microbial bioprocessing: from biofuels and chemicals, to biocatalysis and bioremediation*, Metabolic Engineering*, 12: p. 307-331.

71 Lorenz M, NS Cutler and J Heitman (2000) Characterization of alcohol-induced filamentous growth in Saccharomyces cerevisiae*, Molecular Biology of the Cell*, 11: p. 183-199.

72 Silva Ferreira AC, JC Barbe and A Bertrand (2003) 3-Hydroxy-4,5-dimethyl-2(5H)-furanone: a key odorant of the typical aroma of oxidative aged Port wine*, J Agric Food Chem*, 51 (15): p. 4356-63.

73 Silva Ferreira AC, JC Barbe and A Bertrand (2002) Heterocyclic acetals from glycerol and acetaldehyde in Port wines: evolution with aging*, J Agric Food Chem*, 50 (9): p. 2560-4.

74 Câmara JS, JC Marques, MA Alves and AC Silva-Ferreira (2004) 3-Hydroxy-4,5-dimethyl-2(5H)-furanone levels in fortified Madeira wines: relatioship to sugar content*, Journal of Agricultural and Food Chemistry*, 52: p. 6765-6769.

75 Vernon LS and JLC Johannes (1995) Phenolic browning: a perspective from grape and wine research, *in Enzymatic browning and its prevention*, A. C. Society, Editor: New York.

76 Lambropoulos I and IG Roussis (2007) Inhibition of the decrease of volatile esters and terpenes during storage of a white wine and a model wine medium by caffeic acid and gallic acid*, Food Research International*, 40: p. 176-181.

77 Jarauta I, J Cacho and V Ferreira (2005) Concurrent phenomena contributing to the formation of the aroma of wine during aging in oak wood: an analytical study*, Journal of Agricultural and Food Chemistry*, 53: p. 4166-4177.

78 Campo E, J Cacho and V Ferreira (2008) The chemical characterization of the aroma of dessert and sparkling white wines by gas chromatography - olfactometry and chemical quantitative analysis*, Journal of Agricultural and Food Chemistry*, 56: p. 2477-2484.

79 Blanchard L, T Tominaga and D Dubourdieu (2001) Formation of furfurylthiol echibiting a strong coffee aroma during oak barrel fermentatin from furfural released by toasted staves*, Journal of Agricultural and Food Chemistry*, 49: p. 4833-4835.

80 Antonelli A, F Chinnici and F Masino (2004) Heat-induced chemical modiciation of grape must as related to its concentration during the production of tradicional balsamic vinegar: a preliminary approach*, Food Chemistry*, 88: p. 63-68.

81 Monforte AR (2013) *Changes in the volatile fraction composition of Port wines during aging: a mechanistic approach*, *in Escola Superior de Biotecnologia*, Universidade Católica Portuguesa: Porto.

82 Karimi M (2012) *Optimization of butanol pathway in metabolically engineered Saccharomyces cerevisiae*, *in Department of Chemical and Biological Engineering*, Chalmers University of Technology: Göteborg.

83 Mussato SI, G Dragone, PMR Guimarães, JP Silva, LM Carniero, IC Roberti, AA Vicente, L Domingues and JA Teixeira (2010) Technological trends, global market, and challenges of bioethanol production*, Biotechnology Advances*, 28: p. 817-830.

84 Dragone G, B Fernandes, AA Vicente and JA Teixeira (2010) Current Research Technology and Education Topics, *in Applied Microbiology and Microbial Biotechnology*, A. Méndez-Vilas, Editor.

85 Tsutsumi A and Y Kansha (2011) Co-production of bioethanol and power, *in Biofuel's engineering process technology*, D. M. A. d. S. Bernardes, Editor., InTech.

86 Ruiz HA, DP Silva, DS Ruzene, LF Lima, AA Vicente and JA Teixeira (2012) Bioethanol production from hydrothermal pretreated wheat straw by a flocculating Saccharomyces cerevisiae strain - effect of process conditions*, Fuel*, 95: p. 528-536.

87 Almeida J, T Modig, A Petersson, B Hahn-Hägerdal, G Liden and MF Gorwa-Grauslund (2007) Increased tolerance and conversion of inhibitors in lignocellulosic hydrolysates by Saccharomyces cerevisiae*, J Chem Technol Biotechnol*, 82: p. 340–349.

88 Pereira FB, PMR Guimarães, JA Teixeira and L Domingues (2011) Robust industrial Saccharomyces cerevisiae strains for very high gravity bio-ethanol fermentations.*, Journal of Bioscience and Bioengineering*, 112 (2): p. 130-136.

89 Bai W, WA Anderson and M Moo-Young (2008) Ethanol fermentation technologies from sugar and starch feedstocks*, Biotechnol Adv*, 26: p. 89-105.

90 Puligundla P, D Smogrovicova, VS Obulam and S Ko (2011) Very High Gravity (VHG) ethanolic brewing and fermentation: a research update*, Journal of Industrial Microbiology and Biotechnology*, 38 (9): p. 1133-1144.

91 Panchal CJ and GG Stewart (1980) The effect of osmotic pressure on the production and excretion of ethanol and glycerol by a brewing yeast strain*, J Inst Brew*, 86: p. 207-210.

92 Ingledew WM (2009) The alcohol textbook, Nottingham, UK: Nottingham University Press

93 Ma M and ZL Liu (2010) Comparative transcriptome profiling analyses during the lag phase uncover YAP1, PDR1, PDR3, RPN4, and HSF1 as key regulatory genes in genomic adaptation to the lignocellulose derived inhibitor HMF for Saccharomyces cerevisiae.*, BMC Genomics*, 24: p. 11-660.

94 Walker G (2011) 125th Aniversary review: fuel alcohol: current production and future challenges*, J Inst Brew*, 117: p. 3-22.

95 Banerjee N, R Bhatnagar and L Viswanathan (1981) Inhibitioin of glycolysis by furfural in *Saccharomyces cerevisiae, Eur J Appl Microb Biotechnol*, 11: p. 226-228.

96 Modig T, G Liden and MJ Taherzadeh (2002) Inhibition effects of furfural on alcohol dehydrogenase, aldehyde dehydrogenase and pyruvate dehydrogenase.*, Biochemical Journal*, 363 (3): p. 769-776.

97 Almeida J, M Bertilsson, MF Gorwa-Grauslund, S Gorsich and G Liden (2009) Metabolic effects of furaldehyde and impacts on biotechnological processes*, Appl Microbiol Biotechnol*, 82: p. 625-638.

98 Liu ZL (2011) Molecular mechanisms of yeast tolerance and in situ detoxification of lignocellulose hydrolysates*, Appl Microbiol Biotechnol*, 90: p. 809-825.

99 Zaldivar J, A Martinez and LO Ingram (1999) Effect of selected aldehydes on the growth and fermentation of ethanologenic Escherichia coli, *Biotechnol Bioeng*, 65: p. 24-33.

100 Heipieper HJ, FJ Weber, J Sikkema, H Keweloh and J Debont (1994) Mechanisms of resistance of whole cells to toxic organic solvents, *Trends Biotechnol*, 12: p. 409–415.

101 Pienkos PT and M Zhang (2009) Role of pretreatment and conditioning processes on toxicity of lignocellulosic biomass hydrolysates, *Cellulose*, 16: p. 743–762.

102 Piddocke MP, S Kreisz, HP Heldt-Hansen, KF Nielsen and L Olsson (2009) Physiological characterization of brewer's yeast in high-gravity beer fermentations with glucose or maltose syrups as adjunts, *Appl Microbiol Biotechnol*, 84: p. 453-464.

103 Kapu NS, M Piddocke and JJN Saddler (2013) High gravity and high cell density mitigate some of the fermentation inhibitory effects of softwood hydrolysates, *AMB Express*, 3: p. 15.

104 González-Ramos D, M van den Broek, AJA van Maris, JT Pronk and J-MG Daran (2013) Genome-scale analyses if butanol tolerance in *Saccharomyces cerevisiae* reveal an essential role of protein degradation, *Biotechnology for Biofuels*, 6 (48).

105 Dürre P (2007) Biobutanol: an attractive biofuel, *Biotechnology Journal*, 2: p. 1525-1534.

106 Fischer CR, D Klein-Marcuschamer and G Stephanopoulos (2008) Selection and optimization of microbial hosts for biofuels production, *Metabolic Engineering*, 10: p. 295-304.

107 Dürre P (2011) Fermentative production of butanol-the academic perspective, *Current Opinion in Biotechnology*, 22: p. 331-336.

108 Inui M, M Suda, S Kimura, K Yasuda, H Suzuki, H Toda, S Yamamoto, S Okino, N Suzuki and H Yukawa (2008) Expression of Clostridium acetobutylicum butanol synthetic genes in *Escherichia coli Appl Microbiol Biotechnol* 77: p. 1305–1316.

109 Atsumi S, AF Cann, MR Connor, CR Shen, KM Smith, MP Brynildsen, KJ Chou, T Hanai and JC Liao (2008) Metabolic engineering of Escherichia coli for 1-butanol production, *Metabolic Engineering*, 10: p. 305–311.

110 Nielsen DR, E Leonard, SH Yoon, HC Tseng, C Yuan and KL Prather (2009) Engineering alternative butanol production platforms in heterologous bacteria, *Metabolic Engineering*, 11: p. 262–273.

111 Ghiaci P, J Norbeck and C Larsson (2013) Physiological adaptations of *Saccharomyces cervisiae* evolved for improved butanol tolerance, *Biotechnology for Biofuels*, 6 (101).

112 Dürre P (1998) New insights and novel developments in clostridial acetone/butanol/isopropanol fermentation, *Applied Microbiology and Biotechnology* 49: p. 639–648.

113 Chen X, KF Nielsen, I Borodina, MC Kielland-Brandt and K Karhumaa (2011) Increased isobutanol production in *Saccharomyces cerevisiae* by overexpression of genes in valine metabolism. , *Biotechnology Biofuels*, 4: p. 1754–6834.

114 Steen EJ, R Chan, N Prasad, S Myers, CJ Petzold, A Redding, M Ouellet and JD Keasling (2008) Metabolic engineering of *Saccharomyces cerevisiae* for the production of n-butanol, *Microbiology Cell Factory*, 7 (36).

115 Patakova P, M Daniel, M Rychtera, M Linhova, P Fribert, Z Muzikova, J Lipovsky, L Paulova, M Pospisil, G Sebor and K Melzoch (2011) Perspectives of biobutanol production and use *in Biofuel's Engineering Process Technology*, M. A. Bernardes, Editor., InTech.

116 Knoshaug EP and M Zhang (2009) Butanol tolerance in a selection of microorganisms, *Applied Biochemitry and Biotechnology*, 153: p. 13–20.

117 Liu S and N Qureshi (2009) How microbes tolerate ethanol and butanol, *N Biotechnol*, 26: p. 117–121.

118 Moreira AR, DC Ulmer and JC Linden (1981) Butanol toxicity in the butylic fermentation, *Biotechnology Bioengineering Symposium*, 11: p. 567–579.

119 Vollherbst-Schneck K, JA Sands and BS Montenecourt (1984) Effect of butanol on lipid composition and fluidity of *Clostridium acetobutylicum* ATCC 824, *Applied Environmental Microbiology*, 47: p. 193-194.

120 Bowles L, K. and W Ellefson, L. (1985) Effects of butanol on *Clostridium acetobutylicum, Appl Environ Microbiol* 50: p. 1165–1170.

121 Ingram LO (1986) Microbial tolerance to alcohols: role of the cell membrane, *Trends Biotechnology*, 4: p. 40–44.

122 Zheng YN, LZ Li, M Xian, YJ Ma, JM Yang, X Xu and DZ He (2009) Problems with the microbial production of butanol, *Journal of Industrial Microbiology and Biotechnology*, 36: p. 1127–1138.

123 Ennis BM, CT Marshall, IS Maddox and AHJ Paterson (1986) Continuous product recovery by in-situ gas stripping/condensation during solvent production from whey permeate using *Clostridium acetobutylicum, Biotechnology Letters*, 8: p. 725–730.

124 Pagliardini J, G Hubmann, C Bideaux, S Alfenore, E Nevoigt and SE Guillouet (2010) Quantitative evaluation of yeast's requirement for glycerol formation in very high ethanol performance fed-batch process, *Microbial Cell Factories*, 9 (36).

125 Białecka-Florjańczyk E and AU Kapturowska (2012) Genetically modified baker's yeast Saccharomyces cerevisiae in chemical synthesis and biotransformations, *in Chemical Biology*, P. D. Ekinci, Editor., InTech.

126 Domingues L, AA Vicente, N Lima and JA Teixeira (2000) Applications of yeast flocculation in biotechnological processes, *Biotechnology and Bioprocess Engineering*, 5: p. 288-305.

127 Verstrepen KJ and FM Klis (2006) Flocculation, adhesion and biofilm formation in yeasts, *Molecular Microbiology*, 60 (1): p. 5-15.

128 Domingues L, JA Teixeira and N Lima (1999) Construction of a flocculent Saccharomyces cerevisiae fermenting lactose, *Applied Microbiology Biotechnology*, 51: p. 621-626.

129 Eddy AA and MAD Phil (1958) Composite nature of the flocculation process of top and bottom strains of Saccharomyces, *J Inst Brew*, 64: p. 143-151.

130 Miki BLA, NH Poon, AP James and VL Seligy (1982) Possible mechanism for flocculation interations governed by gene *FLO1* in Saccharomyces cerevisiae, *J Bacteriol*, 150: p. 878-889.

131 Sharma SC (2006) A possible role of trehalose in osmotolerance and ethanol tolerance in Saccharomyces cerevisiae, *FEMS microbiology letters*, 152 (1): p. 11-15.

132 Wiemken A (1990) Trehalose in yeast, stress protectant rather than reserve carbohydrate, *Antonie van Leeuwenhoek*, 58 (3): p. 209-217.

133 Thevelein JM (1984) Regulation of trehalose mobilization in *fungi, Microbiological Reviews*, 48 (1): p. 42.

134 Nordström K (1966) Yeast growth and glycerol formation, *Acta Chem Scand*, 20 (4): p. 1016-1025.

135 Blomberg A and L Adler (1992) Physiology of osmotolerance in fungi 1, *Advances in microbial physiology*, 33: p. 145.

136 Nevoigt E and U Stahl (2006) Osmoregulation and glycerol metabolism in the yeast *Saccharomyces cerevisiae, FEMS Microbiology Reviews*, 21 (3): p. 231-241.

137 Castro CC, C Gunning, CM Oliveira, JA Couto, JA Teixeira, RC Martins and A Silva-Ferreira (2012) Saccharomyces cerevisiae Oxidative Response Evaluation by Cyclic Voltammetry and Gas Chromatography–Mass Spectrometry, *Journal of Agricultural and Food Chemistry*, 60: p. 7252 - 7261.

138 Mensack MM, RE Holcomb and CS Henry (2013) Potential of microfluidics and single cell analysis in metabolomics (micrometabolomics), *in Metabolomics in practice: successful strategies to generate and analyse metabolic data*, M. LämmerhoferandW. Weckwerth, Editors., Wiley-VCH: Singapore.

139 Villas-Bôas SG, S Mas, M Akesson, J Smedsgaard and J Nielsen (2005) Mass spectrometry in metabolome analysis, *Mass Spectrom Rev*, 24 (5): p. 613-46.

140 Carrau FM, K Medina, L Farina, E Boido, PA Henschke and E Dellacassa (2008) Production of fermentation aroma compounds by *Saccharomyces cerevisiae* wine yeasts: effects of yeast assimilable nitrogen on two model strains, *FEMS Yeast Res*, 8: p. 1196-1207.

141 Villas-Bôas SG, S Noel, GA Lane, G Attwood and A Cookson (2006) Extracellular metabolomics: a metabolic footprinting approach to assess fiber degradation in complex media, *Analytical Biochemistry*, 349: p. 297-305.

142 Tuck MK, DW Chan, D Chia, AK Godwin, WE Grizzle, KE Krueger, W Rom, M Sanda, S L., S Stass, W Wang and DE Brenner (2009) Standard operating procedures for serum and plasma

collection: early detection research network consensus statement standard operating procedure integration working group, *J Proteome Res*, 8 (1): p. 113-117.

143 de Hoffman E and V Stroobant (2007) Mass Spectrometry: principles and applications, England: John Wiley & Sons Ltd.

144 Koning W and K van Dam (1992) A method for determination of changes of glycolytic metabolites in yeast on a subsecond time scale using extraction at neutral pH, *Analytical Biochemistry*, 204 (118-123).

145 Dunn WB and DI Ellis (2005) Metabolomics: current analytical platforms and methodologies, *TrAC Trends in Analytical Chemistry*, 24 (4): p. 285-294.

146 Hajjaj H, PJ Blanc, G Goma and J François (1998) Sampling techniques and comparative extraction procedures for quantitative determination of intra- and extracellular metabolites in filamentous fungi, *FEMS Microbiol Let*, 164 (195-200).

147 Cole HA, JW Wimpenny and DE Hughes (1967) Atp pool in *Escherichia Coli*. I. Measurement of pool using a modified luciferase assay, *Biochimica et Biophysica Acta* 143: p. 445–453.

148 Shryock JC, R Rubio and RM Berne (1986) Extraction of adenine nucleotides from cultured endothelial cells, *Anal Biochem*, 159: p. 73–81.

149 Mashego MR, K Rumbold, M De Mey, E Vandamme, W Soetaert and JJ Heijnen (2007) Microbial metabolomics: past, present and future methodologies.

150 Oldiges M, S Lütz, S Pflug, K Schroer, N Stein and C Wiendahl (2007) Metabolomics: current state and evolving methodologies and tools, *Appl Microbiol Biotechnol*, 76 (3): p. 495-511.

151 Starmans DAJ and HH Nijhuis (1996) Extraction of secondary metabolites from plant material: A review *Trends Food Sci Technol* 7:p. 191-197.

152 Entian KD, FK Zimmermann and I Scheel (1977) A partial defect in carbon catabolite repression mutants of *Saccharomyces cerevisiae* with reduced hexose phosphorylation, *Mol Gen Genet*, 156: p. 99–105.

153 Maharjan RP and T Ferenci (2003) Global metabolite analysis: The influence of extraction methodology on metabolome profiles of *Escherichia coli, Anal Biochem* 313: p. 145–154.

154 Oliveira JM, M Faria, F Sá, F Barros and IM Araújo (2006) C6-alcohols as varietal markers for assessment of wine origin, *Anal. Chim. Acta*, 563: p. 300–309.

155 Bethin B, H Danz and M Hamburger (1999) Pressurized liquid extraction of medicinal plants, *J Chromatogr A* 837: p. 211–219.

156 Young JC (1995) Microwave-assisted extraction of the fungal metabolite ergosterol and total fatty acids, *J Agric Food Chem*, 43: p. 2904–2910.

157 Young JC and DE Games (1993) Supercritical fluid extraction and supercritical fluid Chromatography of the fungal metabolite ergosterol, *J Agric Food Chem* 41: p. 577–581.

158 Sasidharan K, T Soga, M Tomita and DB Murray (2012) A yeast metabolite extraction protocol optimised for time-series analysis, *PLoS ONE*, 7 (8): p. e44283.

159 Ferreira V, P Fernández, J Meléndez and J Cacho (1995) Analytical characteristics of sample evaporation with the micro-Kuderna-Danish concentrator, *J Chromatogr A* 695: p. 41–55.

160 Villas-Bôas SG, J Hojer-Pedersen, M Akesson, J Smedsgaard and J Nielsen (2005) Global metabolite analysis of yeast: evaluation of sample preparation methods, *Yeast*, 22: p. 1155-1169.

161 Pawliszyn J (1997) Solid Phase Microextraction: theory and practice: Wiley-VCH.

162 Villas-Bôas SG, U Roessner, MAE Hansen, J Smedsgaard and J Nielsen (2007) Metabolome analysis: an introduction: John Wiley & Sons, Inc.

163 Kounaves SP (1997) Voltammetric techniques, *in Handbook of instrumental techniques for analytical chemistry*, F. A. Settle, Editor., Prentice Hall PTR: Arlington, Virginia.

164 Roginsky V and EA Lissi (2005) Review of methods to determine chain-breaking antioxidant activity in food, *Food Chemistry*, 92: p. 235-254.

165 Kissinger PT and W Heineman (1983) Cyclic voltammetry, *Journal of Chemical Education*, 60 (9).

166 Kilmartin PA, H Zou and AL Waterhouse (2001) A cyclic voltammetry method suitable for characterizing antioxidant properties of wine and wine phenolics, *Journal of agricultural and food chemistry*, 49 (4): p. 1957-1965.

167 Silva Ferreira AC, C Oliveira, T Hogg and P Guedes de Pinho (2003) Relationship between potentiometric measurements, sensorial analysis, and some substances responsible for aroma degradation of white wines, *J Agric Food Chem*, 51 (16): p. 4668-72.

168 Rodrigues A, AC Silva Ferreira, P Guedes de Pinho, F Bento and D Geraldo (2007) Resistance to oxidation of white wines assessed by voltammetric means, *Journal of agricultural and food chemistry*, 55 (26): p. 10557-10562.

169 Simic A, D Manojlovic, D Segan and M Todorovic (2007) Electrochemical behavior and antioxidant and prooxidant activity of natural phenolics, *Molecules*, 12: p. 2327-2340.

170 McNair HM and JM Miller (2009) Basic Gas Chromatography: John Wiley & Sons, Inc.

171 Giddings JC (2002) Dynamics of Chromatography: principles and theory, Danvers, MA, USA: CRC.

172 Iwasaki Y, Y Nakano, K Mochizuki, M Nomoto, Y Takahashi, R Ito, K Saito and H Nakazawa (2011) A new strategy for ionization enhancement by derivatization for mass spectrometry., *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* , 879 p. 1159–1165.

173 Guillarme D, J Schappler, S Rudaz and J-L Veuthey (2010) Coupling ultra-high-pressure liquid chromatography with mass spectrometry., *TrAC, Trends in Analytical Chemistry*, 29: p. 15-27.

174 Wilson ID, JK Nicholson, J Castro-Perez, JH Granger, KA Johnson, BW Smith and RS Plumb (2005) High resolution "Ultra Performance"Liquid Chromatography coupled to oa-TOF Mass Spectrometry as a tool for differential metabolic pathway profiling in functional genomic studies *J. Proteome Res.*, 4 (2): p. 591–598.

175 Plumb RS, JH Granger, CL Stumpf, KA Johnson, BW Smith, S Gaulitz, ID Wilson and J Castro-Perez (2005) A rapid screening approach to metabonomics using UPLC and oa-TOF mass spectrometry: application to age, gender and diurnal variation in normal/Zucker obese rats and black, white and nude mice., *Analyst.*, 130 (6): p. 844-849.

176 Idborg H, L Zamani, PO Edlund, I Schuppe-Koistinen and SP Jacobsson (2005) Metabolic fingerprinting of rat urine by LC/MS Part 1. Analysis by hydrophilic interaction liquid chromatography-electrospray ionization mass spectrometry, *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* 828: p. 9-13.

177 Grob KJ (2001) Split and splitless injection for quantitative gas chromatography: concepts, processes, practical guidelines, sources of errors, Weinheim, Germany: Wiley-VCH.

178 Iwasaki Y, T Sawada, K Hatayama, A Ohyagi, O Tsukuda, K Namekawa, R Ito, K Saito and H Nakazawa (2012) Separation technique for the determination of highy polar metabolites in biological samples, *Metabolites*, 2: p. 496-515.

179 Jumtee K, T Bamba and E Fukusaki (2009) Fast GC-FID based metabolic fingerprinting of Japanese green tea leaf for its quality ranking prediction, *J Sep Sci*, 32: p. 2296-2304.

180 Ochi H, T Bamba, H Naito, K Iwatsuki and E Fukusaki (2012) Metabolic fingerprinting if hard and semi-hard natural cheeses using gas chromatography with flame ionization detector for practical sensory prediction modeling, *Journal of Bioscience and Bioengineering*, 114 (5): p. 506-511.

181 Ma C, H Wang, X Lu, G Xu and B Liu (2008) Metabolic fingerprinting investigation ofArtemisia annua L. in different stages of development by gas chromatography and gas chromatography-mass spectrometry, *J Chromatogr A* 1186: p. 412–419.

182 Yeh TF, CR Morris, B Goldfarb, HM Chang and JF Kadla (2006) Utilization of polar metabolite profiling in the comparison of juvenile wood and compression wood in loblolly pine (Pinus taeda), *Tree Physiology*, 26 (11): p. 1497-1503.

183 Seger C and S Sturm (2007) Analytical aspects of plant metabolite profiling platforms - current standings and future aims. (Review), *J. Proteome Res.* , 6: p. 480-497.

184 Lenz EM and ID Wilson (2007) Analytical strategies in metabonomics, *J Proteome Res*, 6 (2): p. 443-458.

185 Hollywood K, DR Brison and R Goodacre (2006) Metabolomics: current technologies and future trends, *Proteomics*, 6: p. 4716-4723.

186 Bedair M and LW Summer (2008) Current and emergif mass-spectrometry technologies for metabolomics, *Trends in Analytical Chemistry*, 27 (3): p. 238-250.

187 Nicholson JK, MJ Buckingham and PJ Sadler (1983) High resolution 1H n.m.r. studies of vertebrate blood and plasma, *Biochem J*, 211: p. 605-615.

188 Nicholson JK, DP Higham, JA Timbrell and PJ Sadler (1989) Quantitative high resolution 1H NMR urinalysis studies on the biochemical effects of cadmium in the rat, *Mol Pharmacol*, 36 (3): p. 398-404.

189 Ramautar R, A Demirci and GJ De Jong (2006) Capillary electrophoresis in metabolomics, *Trends in Analytical Chemistry*, 25: p. 455-466.

190 Ramautar R, GW Somsen and GJ de Jong (2009) CE-MS in metabolomics, *Electrophoresis*, 30 (1): p. 276-291.

191 Kopka J (2006) Current challenges and developments in GC-MS based metabolite profiling technology, *Journal of Biotechnology*, 124: p. 312-322.

192 Gika HG, GA Theodoridis, JE Wingate and ID Wilson (2007) Within-day reproducibility of an HPLC-MS-based method for metabonomic analysis: an application to human urine, *J Proteome Res*, 6: p. 3291-3303.

193 Forcisi S, F Moritz, B Kanawati, D Tziotis, R Lehnmann and P Schmitt-Kopplin (2013) Liquid chromatography - mass spectrometry in metabolomics research: mass analysers in ultra high pressure liquid chromatography coupling, *Journal of Chromatography A*, 1292: p. 51-65.

194 Sun T, SJ Wetzel, ME Johnson, AS Beth and J Patton-Vogt (2012) Development and validation of a hydrophilic interaction liquid chromatography-tandem mass spectrometry method for the quantification of liquid-related extracellular metabolites in *Saccharomyces cerevisiae, Jounal of Chromatography B*, 897: p. 1-9.

195 Lu W, E Kimball and JD Rabinowitz (2006) A high-performance liquid chromatography-tandem mass spectrometry method for quantitation of nitrogen-containing intracellular metabolites, *J Am Soc Mass Spectrom*, 17 (37–50).

196 Wang J, D Leung and W Chow (2010) Applications of LC/ESI-MS/MS and UHPLC QqTOF MS for the Determination of 148 Pesticides in Berries, *J. Agric. Food Chem.*, 58 (10): p. 5904–5925.

197 Wang J, W Chow and D Leung (2011) Applications of LC/ESI-MS/MS and UHPLC/Qq-TOF-MS for the determination of 141 pesticides in tea, *J AOAC Int.* , 94 (6): p. 1685-1714.

198 Soga T, Y Ohashi, Y Ueno, H Naraoka, M Tomita and T Nishioka (2003) Quantitative metabolome analysis using capillary electrophoresis mass spectrometry, *J Proteome Res*, 2 (5): p. 488-494.

199 Zhang NR, S Yu, P Tiller, S Yeh, E Mahan and WB Emary (2009) Quantitation of small molecules using high-resolution accurate mass spectrometers - a different approach for analysis of biological samples, *Rapid Comunications in Mass Spectrometry*, 23 (7): p. 1085-1094.

200 Carrasco-Pancorbo A, N Navas-Iglesias and Cuadros-Rodríguez (2009) From lipid analyss towards lipidomics, a new challenge for the analytical chemistry of the 21st century. Part I: modern lipid analysis, *Trends in Analytical Chemistry*, 28 (263-268).

201 McEwen CN and RG McKay (2005) A combination atmospheric pressure LC/MS:GC/MS ion source: advantages of dual AP-LC/MS:GC/MS instrumentation, *American Society for Mass Spectrometry*, 16: p. 1730-1738.

202 Siuzdak G (1996) Mass spectrometry for biotechnology, *ed. A. Press*.

203 Mapelli V, L Olsson and J Nielsen (2008) Metabolic footprinting in microbiology: methods and applications in functional genomics and biotechnology, *Trends Biotechnol*, 26 (9).

204 Matyash V, G Liebisch, TV Kurzchalia, A Shevchenko and D Schwudke (2008) Lipid extraction by methyl-ter-butyl ether for high-throughput lipidomics, *The Journal of Lipid Research*, 49: p. 1137-1146.

205 Liebisch G, M Binder, R Schifferer, T Langmann, B Schulz and G Schmitz (2007) Glycerophospholipid identification and quantitation by electrospray ionization mass spectrometry *in Methods in Enzymology*, R. Khosravi-Far, Z. Zakeri, R. A. LockshinandM. Piacentini, Editors., Elsevier: Amsterdam, The Netherlands.

206 King R, R Bonfiglio, C Fernandez-Metzler, C Miller-Stein and T Olah (2000) Mechanistic investigation of ionization suppression in electrospray ionization, *Journal of the American Society of Mass Spectrometry*, 11 (11): p. 942-950.

207 Oliver SG, MK Winson, DB Kell and F Baganz (1998) Systematic functional analysis of the yeast genome, *Trends in Biotechnology*, 16 (9): p. 373-378.

208 Fiehn O (2001) Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks, *Comparative and Functional Genomics*, 2: p. 155-168.

209 Cristoni S and LR Bernardi (2003) Development of new methodologies for the mass spectrometry study of bioorganic macromolecules, *Mass Spectrometry Review*, 22: p. 369-406.

210 Weckwerth W and K Morgenthal (2005) Metabolomics: from pattern recognition to biological interpretation, *Drug discovery today*, 10 (22): p. 1551-1558.

211 Schiesel S, M Lämmerhofer and W Lindner (2010) Multitarget quantitative metabolic profiling of hydrophilic metabolites in fermentation broths of β-lactam antibiotics production by HILIC-ESI-MS/MS., *Analytical Bioanalytical Chemistry*, 396 (5): p. 1655-1679.

212 Kaddurah-Daouk R and KRR Krishnan (2009) Metabolomics: a global biochemical approach to the study of central nervous system diseases, *Neuropsychopharmacology Reviews*, 34 (173-186).

213 Gika HG, E Macpherson, GA Theodoridis and ID Wilson (2008) Evaluation of the repeatability of ultra-performance liquid chromatography-TOF-MS for global metabolic profiling of human urine samples, *J Chromatogr B Analyt Technol Biomed Life Sci*, 871 (2): p. 299-305.

214 Gika HG, GA Theodoridis and ID Wilson (2008) Hydrophilic interaction and reversed-phase ultra-performance liquid chromatography TOF-MS for metabonomic analysis of Zucker rat urine, *J. Sep. Sci.*, 31: p. 1598-1608.

215 Kaderbhai NN, DI Broadhurst, DI Ellis, R Goodacre and DB Kell (2003) Functional genomics via metabolic footprinting: monitoring metabolite secretion by *Escherichia coli* tryptophan metabolism mutants using FT-IR and direct injection electrospray mass spectrometry, *Comparative and Functional Genomics*, 4 (4): p. 376-391.

216 Watkins SM, PR Reifsnyder, HJ Pan, JB German and EH Leiter (2002) Lipid metabolome-wide effects of the PPARgamma agoinst rosiglitazone, *L Lipid Res*, 43 (11): p. 1809-1817.

217 Roberts LD, AL Souza, RE Gerszten and CB Clish (2012) Targeted metabolomics, *in Metabolomics*.

218 Morgenthal K, S Wienkoop, M Scholz, J Selbig and W Weckwerth (2005) Correlative GC-TOF-MS based metabolite profiling and LC-MS based protein profiling reveal time-related systemic regulation of metabolite-protein networks and improve pattern recognition for multiple biomarker selection., *Metabolomics* 1:p. 109-121.

219 Bicalho B, F David, K Rumplel, E Kindt and P Sandra (2008) Creating a fatty acid methyl ester database for lipid profiling in a single drop of human blood using high resolution capillary gas chromatography and mass spectrometry., *J Chromatogr A.* , 1211 (1-2): p. 120-128.

220 Tian J, P Sang, P Gao, R Fu, D Yang, L Zhang, J Zhou, S Wu, X Lu, Y Li and G Xu (2009) Optimization of a GC–MS metabolic fingerprint method and its application in characterizing engineered bacterial metabolic shift, *J. Sep. Sci.*, 32: p. 2281 – 2288.

221 Bajad SU, W Lu, EH Kimball, J Yuan, C Peterson and JD Rabinowitz (2006) Separation and quantitation of water soluble cellular metabolites by hydrophilic interaction chromatography-tandem mass spectrometry, *J Chromatogr A* 1125: p. 76-88.

222 Luo B, K Groenke, R Takors, C Wandrey and M Oldiges (2007) Simultaneous determination of multiple intracellular metabolites in glycolysis, pentose phosphate pathway and tricarboxylic acid cycle by liquid chromatography-mass spectrometry, *J. Chromatogr. A*, 1147: p. 153–164.

223 Timischl B, K Dettmer, H Kaspar, M Thieme and PJ Oefner (2008) Development of a quantitative, validated CE-TOF-MS method and its integration with high-confidence feature detection for metabolomics, *Electrophoresis* 29: p. 2203-2214.

224 Nicholson JK, E Holmes and ID Wilson (2005) Gut microorganisms, mammalian metabolism and personalized health care, *Nature Reviews Microbiology*, 3: p. 431-438.

225 Katajamaa M and M Oresic (2005) Processing methods for differential analysis of LC/MS profile data, *BMC Bioinformatics*, 6: p. 179.

226 Buchholz A, J Hurlebaus, C Wandrey and R Takors (2002) Metabolomics: quantification of intracellula metabolite dynamics, *Biomolecular Engineering*, 19: p. 5-15.

227 Hilario M, A Kalousis, C Pellegrini and M Müller (2006) Processing and classification of protein mass spectra, *Mass Spectrometry Reviews*, 25: p. 409-449.

228 Gruup H and S Maital (2001) Improving existing products: optimal incremental innovation, *in Managing new product development and innovation: a microeconomic toolbox*. Edward Elgar Publishing.

229 Wold S (1995) Chemometrics; what do we mean with it, and what do we want from it?, *Chemometrics and Intelligent Laboratory Systems*, 30: p. 109-115.

230 Boccard J, J-L Veuthey and S Rudaz (2010) Knowledge discovery in metabolomics: an overview of MS data handling, *J. Sep. Sci.*, 33: p. 290-304.

231 Wishart DS (2007) Current progress in computational metabolomics, *Briefings in Bioinformatics*, 8 (5): p. 279-293.

232 Castillo S, P Gopalacharyulu, L Yetukuri and M Oresic (2009) Algorithms and tools for the preprocessing of LC-MS metabolomics data, *Chemometrics and intelligent laboratory systems*, 108 (1): p. 23-32.

233 Tautenhahn R, C Böttcher and S Neumann (2008) Highly sensitive feature detection for high resolution LC/MS, *BMC Bioinformatics*, 9 (504).

234 Koh Y, KK Pasikanti, CW Yap and ECY Chan (2010) Comparative evaluation of software for retention time alignment of gas chromatography / time-of-flight mass spectrometry - based metabonomic data, *Journal of Chromatography A*, 1217: p. 8308-8316.

235 O'Callaghan S, DP De Souza, A Isaac, Q Wang, L Hodkinson, M Olshansky, T Erwin, B Appelbe, DL Tull, U Roessner, A Bacic, MJ McConville and VA Likić (2012) PyMS: a Python toolkit for processing of gas chromatography-mass spectrometry (GC-MS) data. Application and comparative study of selected tools, *BMC Bioinformatics*, 13 (115).

236 Martens L, M Chambers, M Sturm, D Kessner, F Levander, J Shofstahl, WH Tang, A Römpp, S Neumann, AD Pizarro, L Montecchi-Palazzi, N Tasman, M Coleman, F Reisinger, P Souda, H Hermjakob, PA Binz and EW Deutsch (2011) mzML–a community standard for mass spectrometry data, *Mol Cell Proteomics*, 10 (1).

237 Deutsch EW (2010) Mass spectrometer output file format mzML, *Methods Mol Biol*, 604: p. 319-331.

238 Deutsch EW (2008) mzML: a single, unifying data format for mass spectrometer output, *Proteomics*, 8: p. 2776-2777.

239 Katajamaa M, J Miettinen and M Oresic (2006) MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data, *Bioinformatics*, 22 (5): p. 634-6.

240 Pluskal T, S Castillo, A Villar-Briones and M Oresic (2010) MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data, *BMC Bioinformatics* 11 (395).

241 Sturm M, A Bertsch, C Gropl, A Hildebrandt, R Hussong, E Lange, N Pfeifer, O Schulz-Trieglaff, A Zerck, K Reinert and O Kohlbacher (2008) OpenMS - an open-source software framework for mass spectrometry, *BMC Bioinformatics*, 9: p. 163.

242 Benton HP, DM Wong, SA Trauger and G Siuzdak (2008) XCMS2: Processing tandem mass spectrometry data for metabolite identification and structural characterization., *Anal Chem*, 80 (16): p. 6382–6389.

243 Tautenhahn R, GJ Patti, D Rinehart and G Siuzdak (2012) XCMS Online: A web-based platform to process untargeted metabolomic datas.,
Anal Chem, 84 (11): p. 5035–5039.

244 Baran R, H Kochi, N Saito, M Suematsu, T Soga, T Nishioka, M Robert and M Tomita (2006) MathDAMP: a package for differential analysis of metabolite profiles, *BMC Bioinformatics*, 7: p. 530.

245 Hiller K, J Hangebrauk, C Jager, J Spura, K Schreiber and D Schomburg (2009) MetaboliteDetector: comprehensive analysis tool for targeted and nontargeted GC/MS based metabolome analysis, *Anal Chem*, 81 (9): p. 3429-39.

246 Cuadros-Inostroza A, C Caldana, H Redestig, M Kusano, J Lisec, H Peña-Cortés, L Willmitzer and MA Hannah (2009) TargetSearch - a Bioconductor package for the efficient preprocessing of GC-MS metabolite profiling data, *BMC Bioinformatics*, 10 (428).

247 Wei X, W Sun, X Shi, I Koo, B Wang, J Zhang, X Yin, Y Tang, B Bogdanov, S Kim, Z Zhou, C McClain and X Zhang (2011) MetSign: a computational platform for high-resolution mass spectrometry-based metabolomics, *Anal Chem.*, 83 (20): p. 7668-7675.

248 Xia J, N Psychogios, N Young and DS Wishart (2009) MetaboAnalyst: a web server for metabolomic data analysis and interpretation, *Nucleic Acids Res*, 37 (Web Server issue): p. W652-60.

249 Press W, S Teykolsky, W Vetterling and B Flannery (2002) Numerical recipes in C++: the art of scientific computing: Cambridge University Press.

250 Savitzky A and M Golay (1964) Smoothing and differentiantion of data by simplified least squares procedures, *Analytical Chemistry*, 36: p. 1627-1639.

251 Katajamaa M and M Oresic (2007) Data processing for mass spectrometry-based metabolomics, *J Chromatogr A*, 1158: p. 318-328.

252 Jaitly N, A Mayampurath, K Littlefield, JN Adkins, GA Anderson and RD Smith (2009) Decon2LS: An open-source software package for automated processing and visualization of high resolution mass spectrometry data, *BMC Bioinformatics*, 10: p. 87.

253 Silverman BW (1982) Kernel estimation with cross-validation using the fast Fourier transform, *Appl Stat*, 31: p. 93-99.

254 Lange E, R Tautenhahn, S Neumann and C Gröpl (2008) Critical assessment of alignment procedures for LC-MS proteomics and metabolomics measurements, *BMC Bioinformatics*, 9: p. 375.

255 Nielsen NPV, JM Carstensen and J Smedsgaard (1998) Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping, *Journal of Chromatography A*, 805 (1-2): p. 17-35.

256 Christensen JH, G Tomasi and AB Hansen (2005) Chemical fingerprinting of petroleum biomarkers using time warping and PCA, *Environmental Science & Technology*, 39 (1): p. 255-60.

257 Walczak B and W Wu (2005) Fuzzy warping of chromatograms, *Chemometrics and Intelligent Laboratory Systems*, 77 (1-2): p. 173-180.

258 Kanehisa M and S Goto (2000) KEGG: kyoto encyclopedia of genes and genomes, *Nucleic Acids Res*, 28 (1): p. 27-30.

259 Kanehisa M, S Goto, M Furumichi, M Tanabe and M Hirakawa (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs, *Nucleic Acids Res*, 38 (Database issue): p. D355-60.

260 Kanehisa M, S Goto, M Hattori, KF Aoki-Kinoshita, M Itoh, S Kawashima, T Katayama, M Araki and M Hirakawa (2006) From genomics to chemical genomics: new developments in KEGG, *Nucleic Acids Res*, 34 (Database issue): p. D354-7.

261 Caspi R, H Foerster, CA Fulcher, P Kaipa, M Krummenacker, M Latendresse, S Paley, SY Rhee, AG Shearer, C Tissier, TC Walk, P Zhang and PD Karp (2008) The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases, *Nucleic Acids Res*, 36 (Database issue): p. D623-31.

262 Fahy E, M Sud, D Cotter and S Subramaniam (2007) LIPID MAPS online tools for lipid research, *Nucleic Acids Research* 35: p. 606-612.

263 Schomburg I, A Chang, O Hofmann, C Ebeling, F Ehrentreich and D Schomburg (2002) BRENDA: a resource for enzyme data and metabolic information, *Trends Biochem Sci*, 27 (1): p. 54-6.

264 Wishart DS, T Jewison, AC Guo, M Wilson and C Knox (2013) HMDB 3.0—The Human Metabolome Database *Nucleic Acids Research*, 41: p. 801-807.

265 Smith CA, G O'Maille, EJ Want, C Qin, SA Trauger, TR Brandon, DE Custodio, R Abagyan and G Siuzdak (2005) METLIN: A Metabolite Mass Spectral Database *Therapeutic Drug Monitoring*.

266 Schellenberger J, JO Park, TC Conrad and BØ Palsson (2010) BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions, *BMC Bioinformatics* 11 (213).

267 Linstrom PJ, WG Mallard, S National Institute of and Technology (1997) *NIST Chemistry WebBook*, National Institute of Standards and Technology Gaithersburg, MD.

268 Saito T, K Hayamizu, M Yanagisawa, O Yamamoto, N Wasada, K Someno, S Kinugasa, K Tanabe, T Tamura and J Hiraishi (2006) Spectral Database for Organic Compounds (SDBS). National Institute of Advanced Industrial Science and Technology (AIST), Japan.

269 Hopley C, T Bristow, A Lubben, A Simpson, E Bull, K Klagkou, J Herniman and J Langley (2008) Towards a universal product ion mass spectral library-reproducibility of product ion spectra across eleven different mass spectrometers*, Rapid Commun.Mass Spectrom*, 22: p. 1779-1786.

270 Van den Berg R, H Hoefsloot, J Westerhuis, A Smilde and M van der Werf (2006) Centering, scaling, and transformations: improving the biological information content of metabolomics data*, BMC genomics*, 7 (1): p. 142.

271 Shurubor YI, U Paolucci, BF Krasnikov, WR Matson and BS Kristal (2005) Analytical precision, biological variation, and mathematical normalization in high data density metabolomics*, Metabolomics*, 1 (1).

272 Jackson JE (1991) A user's guide to principal components John Wiley & Sons Inc.

273 Eriksson L, E Johansson, N Kettaneh-Wold and S Wold (1999) Scaling. In Introduction to multi- and megavariate data analysis using projection methods
(PCA & PLS) *Umetrics*: p. 213-225.

274 Smilde AK, MJ van der Werf, S Bijlsma, B van der Werff-van der Vat and RH Jellema (2005) Fusion of mass-spectrometry-based metabolomics data*, Anal Chem* 77: p. 6729-6736.

275 Keun HC, TMD Ebbels, H Antti, ME Bollard, O Beckonert, E Holmes, JC Lindon and JK Nicholson (2003) Improved analysis of multivariate data by variable stability scaling: application to NMR-based metabolic profiling. *, Anal Chim Acta*, 490: p. 265-276.

276 Maeder M and P King (2012) Analysis of Chemical Processes, Determination of the Reaction Mechanism and Fitting of Equilibrium and Rate Constants, *in Chemometrics in Practical Applications*, K. Varmuza, Editor., InTech.

277 Munck L, L Nørgaard, SB Engelsen, R Bro and CA Andersson (1998) Chemometrics in food science - a demonstrations of the feasibility of a highly exploratory, inductive evaluation strategy of fundamental scientific significance*, Chemometrics and Intelligent Laboratory Systems*, 44 (1-2): p. 31-60.

278 Einax JW, HW Zwanziger and S Geiß (2004) Chemometrics in environmental analysis.

279 Frank IE and JH Friedman (2012) A statistical view of some chemometric regression tools*, Technometrics*, 35 (2): p. 109-135.

280 Antoniewicz MR, G Stephanopoulos and JK Kelleher (2006) Evaluation of regression models in metabolic physiology: predicting fluxes from isotopic data without knowledge of the pathway*, Metabolomics*, 2 (1): p. 41-52.

281 Martins RC, VV Lopes, P Valentão, J Carvalho, P Isabel, MT Amaral, MT Batista, PB Andrade and BM Silva (2008) Relevant principal component analysis applied to the

characterisation of Portuguese heather honey, *Natural Product Research: Formerly Natural Product Letters*, 22 (17): p. 1560-1582.

282 Jolliffe I (1986) Principal Component Analysis, New York, USA.: Springer.

283 Krzanowski WJ (2000) Principles of multivariate analysis: Oxford University Press Oxford:.

284 Baig S and F Rehman (2006) Signal modeling using singular value decomposition, *in Advances in Computer, Information, and Systems Sciences, and Engineering*. Springer: Netherlands.

285 Manly BF (1998) Randomization, Bootstrap and Monte Carlo Methods in Biology, London, UK.: Chapman and Hall.

286 Castro CC, JS Silva, VV Lopes and RC Martins. (2009!) *Yeast metabolic state identification by fiber optics spectroscopy*. *in Applications of optics and photonics*,Portugal.

287 Jackson J and G Mudholkar (1979) Control procedures for residuals associated with principal component analysis, *Technometrics*, 21: p. 341–349.

288 Nomikos P and J MacGregor (1994) Monitoring batch process using multiway principal component analysis, *AIChE Journal*, 40 (8): p. 1361-1375.

289 Nomikos P and J MacGregor (1995) Multivariate spc charts for monitoring batch processes, *Technometrics*, 37 (1): p. 41-59.

290 Westerhuis J, S Gurden and A Smilde (2000) Generalized contribution plots in multivariate statistical process monitoring, *Chemometrics and intelligent laboratory systems* 51 (1): p. 95–114.

291 Conlin AK, EB Martin and AJ Morris (2002) Confidence limits for contribution plots, *Chemometrics and Intelligent Laboratory Systems*, 14: p. 725-736.

292 Miller P, RE Swanson and CE Heckler (1998) Contribution plots: a missing link in multivariate quality control, *Applied mathematics and computer science*, 8: p. 775-792.

293 Dunia R, SJ Qin, TF Edgar and TJ McAvoy (1996) Identification of faulty sensors using principal component analysis, *AIChE Journal*, 42 (10): p. 2797-2812.

294 Qin S (2003.) Statistical process monitoring: basics and beyond, *Journal of Chemometrics*, 17: p. 480-502.

295 DiFoggio R (1995) Examination of some misconceptions about near-infrared analysis, *Applied Spectroscopy*, 49: p. 67-75.

296 Brereton RG (2006) Conseuqneces of sample size, variable selection, and model validation and optimisation, for predicting classification ability from analytical data, *Trends Analyt Chem*, 25: p. 1103-1111.

297 Kourti T (2009) Multivariate statistics process control and process control using latent variables, *in Comprehensive Chemometrics*, S. Brown, Tauler, R. and Walczak, Editor., Elsevier: Oxford. p. 21-54.

298 Kjeldahl K and R Bro (2010) Some common misunderstandings in chemometrics, *Journal of Chemometrics*, 24: p. 558-564.

299 Kozak M and S C.H (2008) Unsupervised classification methods in food sciences: discussion and outlook, *Journal of Food Science*, 88: p. 1115-1127.

300 Frades I and R Matthiesen (2010) Overview on techniques in cluster analysis, *Methods Molec Biol*, 593: p. 81-107.

301 Johnson RA and DW Wichern (2007) Applied multivariate statistical analysis, *ed. P. Education*.

302 Henrion R, G Henrion and GC Onuoha (1992) Multi-way principal components analysis of a complex daa array resulting from physicochemical characterization of natural waters, *Chamometrics and Intelligent Laboratory Systems*, 16: p. 87-94.

303 Tucker LR (1964) The extension of factor analysis to three-di-mensional matrices. Contributions to Mathematical Psychology, *ed. H. Gulliksen, Frederiksen, N.*, New York.

304 Kroonenberg PM and J de Leeuw (1980) Principal component analysis of three-mode data by means of alternating least squares algorithms, *Psychometriha*, 45: p. 69-97.

305 Abdi H (2003) *Partial least squares regression (PLS-regression)*, Thousand Oaks, CA: Sage. p. 792-795.

306 Damiani PC (2011) Determination of atenolol in human urine by emission–excitation fluorescence matrices and unfolded partial least-squares with residual bilinearization, *Talanta*, 85: p. 1526-1534.

307 Kiers HAL (2000) Towards a standardized notation and terminology in multiway analysis, *Journal of Chemometrics*, 14 (3): p. 105-122.

308 Chevion S, MA Roberts and M Chevion (2001) The use of cyclic voltammetry for the evaluation of antioxidant capacity, *Bio-assays for oxidative stress status (BOSS)*: p. 120.

309 Tenenhaus A, A Giron, E Viennet, M Béra, G Saporta and B Fertil (2007) Kernel logistic PLS: A tool for supervised nonlinear dimensionality reduction and binary classification, *Computational Statistics & Data Analysis*, 51: p. 4083-4100.

310 Alvarez-Guerra, D Ballabio, JM Amigo, R Broc and JR Viguri (2010) Development of models for predicting toxicity from sediment chemistry by partial least squares-discriminant analysis and counter-propagation artificial neural networks, *Environmental Pollution*, 158: p. 607-614.

311 Gusnantoa A, A Ploner, F Shuweihdi and Y Pawitanb (2013) Partial least squares and logistic regression random-effects estimates for gene selection in supervised classification of gene expression data, *Journal of Biomedical Informatics*, 46: p. 697-709.

312 Tenenhausa A, A Giron, E Viennet, M Béra, G Saporta and B Fertil (2007) Kernel logistic PLS: A tool for supervised nonlinear dimensionality reduction and binary classification, *Computational Statistics & Data Analysis*, 51: p. 4083-4100.

313 Ding M-Z, B-Z Li, J-S Cheng and Y-J Yuan (2009) Metabolome analysis of differential responses of diploid and haplid yeast to ethanol stress*, OMICS: A Journal of Integrative Biology*. p. 553-561.

314 Palmqvist E, JS Almeida and Hahn-Hägerdal (1999) Influence of furfural on anaerobic glycolytic kinetics of Saccharomyces cerevisiae in batch culture*., Biotechnology and Bioengineering*, 62 (4): p. 447-454.

315 Westman JO, M Taherzadeh and CJ Franzén (2012) Inhibitor tolerance and flocculation of a yeast strain suitable for second generation bioethanol production*, Electronic Journal of Biotechnology*, 15 (3).

316 Herdeiro RS, MD Pereira, AD Panek and EC Eleutherio (2006) Trehalose protects *Saccharomyces cerevisiae* from lipid peroxidation during oxidative stress*, Biochemica et Biophysica Acta*, 1760: p. 340-346.

317 Nordström K (1961) Formation of ethyl acetate in fermentation with brewer's yeast*, Journal Institute of Brewing*, 67.

318 Blomberg A and L Adler (1992) Physiology of osmotolerance in *fungi Advances in Microbial Physiology*, 33: p. 145.

319 Franco-Duarte R, L Umek, B Zupan and D Schuller (2009) Computational approaches for the genetic and phenotypic characterization of a Saccharomyces cerevisiae wine yeast collection*, Yeast*, 26: p. 675-692.

320 Caridi A, JA Cufari and D Ramondino (2002) Isolation and clonal pre-selection of enological *Sacharomyces, The Journal of General and Applied Microbiology*, 48: p. 261-267.

321 Jain N, F Hasan and BC Fries (2008) Phenotypic switching in fungi*, Current fungal infection reports*, 2 (3): p. 180-188.

322 Hansen J and MC Kiekkand-Brandt (1996) Modification of biochemical pathways in industrial yeasts*, Journal of Biotechnology*, 49: p. 1-12.

323 Dijken JP, J Bauer, L Brambilla, P Duboc, JM Francois, C Gancedo, MLF Giuseppin, JJ Heijnen, M Hoare, HC Lange, EA Madden, P Niederberger, J Nielsen, JL Parrou, T Petit, D Porro, M Reuss, N van Riel, M Rizzi, HY Steensma, CT Verrips, J Vindelov and JT Pronk (2000) An interlaboratory comparison of physiological and genetic properties of four Saccharomyces cerevisiae strains*, Enzyme and Microbial Technology*, 26: p. 706-714.

324 Pereira FB, PMR Guimarães, JA Teixeira and L Domingues (2010) Selection of Saccharomyces cerevisiae strains for efficient very high gravity bio-ethanol fermentation processes*, Biotechnology Letters*, 32: p. 1655-1661.

325 Urtubia A and JM Roger (2011) Predictive power of LDA to discriminate abnormal wine fermentations*, Journal of Chemometrics*, 25: p. 382-388.

326 Urtubia A, G Hernández and JM Roger (2012) Detection of abnormal fermentations in wine process by ultivariate statistics and pattern recognition techniques*, Journal of Biotechnology*, 159: p. 336-341.

327 Mortimer R and J Johnston (1986) Genealogy of principal strains of the yeast genetic stock center*, Genetics*, 113 (1): p. 35-43.

328 Freitas Schwan R, AT Mendonça, J da Silva, V Rodrigues and AE Wheals (2001) Microbiology and physiology of cachaça (aguardente) fermentations*, Antonie van Leeuwenhoek*, 79 (1): p. 89-96.

329 Basso LC, HV De Amorim, AJ De Oliveira and ML Lopes (2008) Yeast selection for fuel ethanol production in Brazil*, FEMS yeast research*, 8 (7): p. 1155-1163.

330 Palmqvist E, H Grage, NQ Meinander and B Hahn-Hägerdal (1999) Main and interaction effects of acetic acid, furfural, and p-hydroxybenzoic acid on growth and ethanol productivity of yeasts*, Biotechnology and Bioengineering*, 63 (1): p. 46-55.

331 Palmqvist E and B Hahn-Hägerdal (2000) Fermentation of lignocellulosic hydrolysates. II: Inhibitors and mechanisms*, Bioresource Technology*, 74: p. 25-33.

332 Cunnif P (1999) Official methods of analysis of AOAC international, *ed. M. GaithersburgandA. International*.

333 Gomes DG, PMR Guimarães, FB Pereira, JA Teixeira and L Domingues (2012) Plasmid-mediate transfer of FLO1 into industrial Saccharomyces cerevisiae PE-2 strain creates a strain useful for repeat-batch fermentations involving flocculation-sedimentation.*, Bioresource Technology*, 108: p. 162-168.

334 Soares EV, JA Teixeira and M Mota (1992) Interaction between flocculent and nonflocculent cells of Saccharomyces cerevisiae*, Canadian Journal of Microbiology*, 38 (9): p. 969-974.

335 Bastien P, V Esposito and M Tenenhaus (2005) PLS generalised linear regression*, Computational Statistics & Data Analysis*, 48: p. 17-46.

336 Ruiz HA, DP Silva, DS Ruzene, LF Lima, AA Vicente and JA Teixeira (2012) Bioethanol produciton from hydrothermal pretreated wheat straw by a flocculating *Saccharomyces cerevisiae* strain - Effect of process conditions*, Fuel*, 95: p. 528-536.

337 Stanley D, A Bandara, S Fraser, PJ Chambers and GA Stanley (2010) The ethanol stress response and ethanol tolerance of *Saccharomyces cerevisiae, Journal of Applied Microbiology*, 109 (1): p. 13-24.

338 Taherzadeh M, L Gustafsson, C Niklasson and G Lidén (2000) Physiological effects of 5-hydroxymethylfurfural on *Saccharomyces cerevisiae, Applied Microbiology and Biotechnology*, 53: p. 701-708.

339 Gorsich SW, BS Dien, NN Nichols, PJ Slininger, ZL Liu and CD ASkory (2006) Tolerance to furfural-induced stress is associated with pentose phosphate pathway genes *ZWF1, GND1, RPE1, and TKL1* in *Saccharomyces cerevisiae, Applied Microbiology and Biotechnology*, 71: p. 339-349.

340 Carmel-Harel O and G Storz (2000) Roles of the glutathione- and thioredoxin-dependent reduction systems in the *Escherichia coli* and *Saccharomyces cerevisiae* responses to oxidative stress*, Annual Reviews in Microbiology* 54: p. 439–461.

341 Mussatto SI and IC Roberto (2004) Alternatives for detoxification of diluted-acid lignocellulosic hydrolyzates for use in fermentative processes – a review*, Bioresource Technology*, 93: p. 1–10.

342 Tielens AGM, KWA van Grinsven, K Henze, JJ van Hellemond and W Martin (2010) Acetate formation in the energy metabolism of parasitic helminths and protists*, International Journal of Parasitology*, 40: p. 387-397.

343 Tielens AGM, C Rotte, JJ van Hellemond and M Martin (2002) Mitochondria as we don't know them*, Trends in Biochemical Sciences*, 27 (11).

344 Remize F, JL Roustan, JM Sablayrolles, P Barre and S Dequin (1999) Glycerol overproduction by engineered Saccharomyces cerevisiae wine yeast strains leads to substancial changes in by-product formation and to a stimulation of fermentation rate in stationary phase*, Applied and Environmental Microbiology*, 65 (1): p. 143-149.

345 Erasmus DJ, M Cliff and JJ van Vuuren (2004) Impact of yeast strains on the production of acetic acid, glycerol and the sensory attributes of icewine*, Journal of Enology and Viticulture*, 55: p. 371-378.

346 Li Q, X-Q Zhao, AK Chang, QM Zhang and FW Bai (2012) Ethanol-induced yeast flocculation directed by the promoter of TPS1 encoding trehalose-6-phosphate synthase 1 for efficient ethanol production*, Metabolic Engineering*, 14: p. 1-8.

347 Alvarez-Peral FJ, O Zaragoza, Y Pedreño and J-C Argüelles (2002) Protective role of trehalose during severe oxidative stress caused by hydrogen peroxide and the adaptive oxidative stress response in *Candida albicans, Microbiology*, 148: p. 2599-2606.

348 Moradas-Ferreira P, V Costa, P Piper and W Mager (1996) The molecular defences against reactive oxygen species in yeast*, Molecular Microbiology*, 19 (4): p. 651-658.

349 Francois JM and JL Parrou (2001) Reserve carbohydrates metabolism in the yeast Saccharomyces cerevisiae*, FEMS Microbiology Review*, 25: p. 125-145.

350 Schulze U, ME Larsen and J Villadsen (1995) Determination of intracellular trehalose and glycogen in Saccharomyces cerevisiae*, Analytical Biochemistry*, 228: p. 143-149.

351 Tamás MJ, S Karlgren, RM Bill, K Hedfalk, L Allegris, M Ferreira, JM Thevelein, J Rydström, JGL Mullins and S Hohmann (2003) A short regulatory domain restricts glycerol transport through yeast Fps1p*, The Journal of Biological Chemistry*, 278 (8): p. 6337-6345.

352 Schwan RF, AT Mendonça, JJ da Silva, V Rodrigues and AE Wheals (2001) Microbiology and physiology of cachaça (aguardente) fermentations*, Antonie van Leeuwenhoek*, 79 (1): p. 89-96.

353 Dengis PB, L Nélissen and PG Rouxhet (1995) Mechanisms of yeast flocculation: comparison of top-and bottom-fermenting strains*, Applied and environmental microbiology*, 61 (2): p. 718-728.

354 Soares EV and A Vroman (2003) Effect of different starvation conditions on the flocculation of Saccharomyces cerevisiae.*, Jounal of Applied Microbiology*, 95: p. 325-330.

355 Sampermans S, J Mortier and EV Soares (2005) Flocculation onset in Saccharomyces cerevisiae: the role of nutrients.*, Journal of Applied Microbiology*, 98: p. 525-531.

356 Claro F, K Rijsbrack and EV Soares (2006) Flocculation onset in Saccharomyces cerevisiae: effect of ethanol, heat and osmotic stress*, Journal of applied microbiology*, 102 (3): p. 693-700.

357 Påhlman A-K, K Granath, R Ansell, S Hohmann and L Adler (2001) The yeast glycerol 3-phosphatases Gpp1p and Gpp2p are required for glycerol biosynthesis and differentially involved in the cellular responses to osmotic, anaerobic, and oxidative stress*, Journal of Biological Chemistry*, 276 (5): p. 3555-3563.

358 Guo B, CA Styles, Q Feng and GR Fink (2000) A Saccharomyces gene family involved in invasive growth, cell–cell adhesion, and mating*, Proceedings of the Nacional Academy of Sciences*, 97 (22): p. 12158-12163.

359 Li E, F Yue, Q Chang, X Guo, X He and B Zhang (2013) Deletion of intragenic tandem repeats in unit C of FLO1 of *Saccharomyces cerevisiae* increases the conformational stability of flocculin under acidic and alkaline conditions*, PLoS ONE*, 8 (1).

360 Stratford M (1989) Evidence for two mechanisms of flocculation in *Saccharomyces cerevisiae, Yeast*, 5: p. 441-445.

361 Eleutherio EC, PS Araujo and AD Panek (1993) Protective role of trehalose during heat stress in *Saccharomyces cerevisiae Cryobiology*, 30 (6): p. 591-596.

362 Kim J, P Alizadeh, T Harding, A Hefner-Gravink and DJ Lionsky (1996) Disruption of the yeast ATH1 gene confers better survival after dehydration, freezing, and ethanol shock: potential commercial applications*, 62*, 5 (1563-1569).

363 Hounsa CG, EV Brandt, JM Thevelein, S Hohmann and BA Prior (1998) Role of trehalose in survival of *Saccharomyces cerevisiae* under osmotic stress*, Microbiology*, 144 (Pt3): p. 671-680.

364 Soares EV (2010) Flocculation in *Saccharomyces cerevisiae:* a review*, Journal of Applied Microbiology*, 110: p. 1-18.

365 Ugliano M and PA Henschke (2009) Yeasts and wine flavour, *in Wine Chemistry and Biochemistry*, M. C. Moreno-Arribas, Editor., Springer: Spain.

366 Bafrncová P, D šmogrovičová, I Slavikova, J Patkova and Z Dömény (1999) Improvement of very high gravity ethanol fermentation by media supplementation using Saccharomyces cerevisiae*, Biotechnology Letters*, 21 (4): p. 337-341.

367 Silva-Filho EA, SK dos Santos, AM Resende, MA de Morais Jr and DA Simões (2005) Yeast population dynamics of industrial fuel-ethanol fermentation process assessed by PCR-fingerprinting*, Antonie van Leeuwenhoek*, 88: p. 13-23.

368 Roessner U and J Bowne (2009) What is metabolomics all about?*, Biotechnoiques*, 46: p. 363-365.

369 Van der Werf MJ, KM Overkamp, B Muilwijk, L Coulier and T Hankemeier (2007) Microbial metabolomics: toward a platform wih full metabolome coverage*, Analytical Biochemistry*, 370: p. 17-25.

370 Kell DB, M Brown, HM Davey, WB Dunn, I Spasic and SG Oliver (2005) Metabolic footprinting and systems biology: the medium is the message*, Nature Reviews Microbiobiology*, 3 (7): p. 557-565.

371 Daffertshofer A, CJC Lamoth, OG Meijer and PJ Beek (2004) PCA in studying coordination and variability: a tutorial*, Clinical Biomechanics*, 19: p. 415-428.

372 Krishnan A, LJ Williams, AR McIntosh and H Abdi (2011) Partial Least Squares (PLS) methods for neuroimaging: a tutorial and review.*, Neuroimage*, 56: p. 455-475.

373 Sun Q-S, Z Jin, P-A Heng and D-S Xia. (2005!) *A novel feature fusion method based on partial least squares regression*. *in ICAPR'05 Proceedings of the Third international conference on Advances in Pattern Recognition - Volume Part I* Heidelberg: Springer-Verlag Berlin.

374 Zha Y and PJ Punt (2013) Exometabolomics approaches for studying the application of lignocellulosic biomass as fermentation feedstock*, Metabolites*, 3: p. 119-143.

375 Stanley D, A Bandara, S Fraser, PJ Chambers and GA Stanley (2010) The ethanol stress response and ethanol tolerance of *Saccharomyces cerevisiae, Journal of Applied Microbiology*, 109: p. 13-24.

376 Oliveira JM, Z Genisheva and L Lima. (2009!) *Easy and accurate methodology to quantify volatile compounds in fermented beverages*. *in VIth Symposium in Vino Analytica Scientia*,Angers.

377 Krazanowski WJ (1998) Principles of multivariate analysis: a users perspective, Oxford: Clarendon Press.

378 Argueso JL, MF Carazzolle, PA Mieczkowski, FM Duarte, OVC Netto, SK Missawa, F Galzerani, GGL Costa, RO Vidal, MF Noronha, M Dominska, MG Andrietta, SR Andrietta, AF Cunha, LH Gomes, FC Tavares, AR Alcarde, FS Dietrich, JH McCusker, TD Petes and GA Pereira (2009) Genome structure of a Saccharomyces cerevisiae strain widely used in bioethanol production*, Genome Research*, 19 (12): p. 2258-2270.

379 Soares TL, CF Silva and RF Schwan (2011) Monitoring the fermentation process for cachaça production usng microbiological and physico-chemical methods with different *Saccharomyces cerevisiae* isolates*, Food Science and Technology*, 31 (1).

380 Duarte WF, G Dragone, DR Dias, JM Oliveira, JA Teixeira, JB Almeida e Silva and RF Schawan (2010) Fermentative behavior of *Saccharomyces* strains during microvinification of raspberry
juice (*Rubus idaeus L.*)*, International Journal of Food Microbiology*, 143: p. 173-182.

381 Verstrepen KJ, G Derdelinckx, J-P Dufour, J Winderickx, JM Thevelein, IS Pretorius and FR Delvaux (2003) Flavor-active esters: adding frutiness to beer*, Journal of Bioscience and Bioengineering*, 96 (2): p. 110-118.

382 Díaz-Montaño DM, ML Délia, M Estarrón-Espinosa and P Strehaiano (2008) Fermentative capability and aroma compoudns production by yeast strains isolated from *Agave tequiliana* Weber juice*, Enzyme and Microbial Technology*, 42 (7): p. 608-616.

383 Lafon-Lafourcade S, C Geneix and P Ribereau-Gayon (1984) Inhibition of alcoholic fermentation of grape must by fatty acids produced by yeasts and their elimination by yeast ghosts*, Applied and Environmental Microbiology*, 47: p. 1246–1249.

384 Hilary A and B Peddie (1990) Ester formation in brewery fermentations*, Journal Institute of Brewing*, 96: p. 327-331.

385 Howard D and RG Anderson (1976) Cell-free synthesis of ethyl acetate by extracts from *Saccharomyces cerevisiae, Journal Institute of Brewing*, 82: p. 70-71.

386 Myers J, P Issenberg and EL Wick (1970) L-leucine as a precursor of isoamyl alcohol and isoamyl acetate, volatile aroma constituents of banana fruit discs*, Phytochemistry*, 9 (8): p. 1693-1700.

387 Al-Shorgani NKN, MS Kalil, E Ali, WMW Yusoff and AA Hamid (2012) Enhancement of biobutanol production by butyric acid addiction using *Clostridium saccharoperbutylacetonicum N1-4* (ATCC 13564)*, Biotechnology*, 11 (6): p. 326-332.

388 Druaux D, G Mangeot, A Endrizzi and J-M Belin (1999) Bacterial bioconversion of primary aliphatic and aromatic alcohols into acids: effects of molecular structure and physico-chemical conditions*, Journal of Chemical Technology and Biotechnology*, 68 (2): p. 214-218.

389 Taherzadeh M, L Gustafsson, C Niklasson and G Lidén (2000) Inhibition effects of furfural and aerobic batch cultivation of *Saccharomyces cerevisiae* growing on ethanol and/or acetic acid*, Journal of Bioscience and Bioengineering*, 90 (4): p. 374-380.

390 Legras JL, C Erny, C Le Jeune, M Lollier, Y Aldolphe, C Demuyter, P Delobel, B Blondin and F Karst (2010) Activation of two different resistance mechanisms in *Saccharomyces cerevisiae* upon exposure to octanoic and decanoic acids*, Applied and Environmental Microbiology*, 76 (22): p. 7526-7535.

391 Liu P, A Chernyshov, T Nadji, Y Fu, J Dickerson, S Sandmeyer and L Jarboe (2013) Membrane stress caused by octanoic acid in *Saccharomyces cerevisiae, Applied Microbiology and Biotechnology*, 97 (7): p. 3239-3251.

392 Bardi L, C Crivelli and M Marzona (1998) Esterase activity and release of ethyl esters of medium-chain fatty acids by *Saccharomyces cerevisiae* during anaerobic growth*, Canadian Journal of Microbiology*, 44 (12): p. 1171-1176.

393 Nicoli MC, R Toniolo and M Anese (2004) Relationship between redox potential and chain-breaking activity of model systems and foods*, Food Chemistry*, 88 (1): p. 79-83.

394 Leistner L and LGM Gorris (1995) Food preservation by hurdle technology*, Trends in Food Science & Technology*, 6 (2): p. 41-46.

395 Waterhouse AL and VF Laurie (2006) Oxidation of wine phenolics: A critical evaluation and hypotheses*, American Journal of Enology and Viticulture*, 57 (3): p. 306-313.

396 Rodrigues JA, AS Barros, B Carvalho, T Brandão, AM Gil and ACS Ferreira (2011) Evaluation of beer deterioration by gas chromatography-mass spectrometry/multivariate analysis: A rapid tool for assessing beer composition*, Journal of Chromatography A*, 1218 (7): p. 990-996.

397 Ciani M and L Ferraro (1996) Enhanced Glycerol Content in Wines Made with Immobilized *Candida stellata* Cells*, Applied and Environmental Microbiology*, 62 (1): p. 128-132.

398 Jackson R (2008) Wine science: principles and applications. 3rd ed, San Diego: Academic Press.

399 Demiray S, ME Pintado and PML Castro (2009) Evaluation of phenolic profiles and antioxidant activities of Turkish medicinal plants: Tilia argentea, Crataegi folium leaves and Polygonum bistorta roots*, World Academy of Science, Engineering and Technology*, 54: p. 312-317.

400 Silva Ferreira AC, T Hogg and P Guedes de Pinho (2003) Identification of key odorants related to the typical aroma of oxidation-spoiled white wines*, Journal of Agricultural and Food Chemistry*, 51 (5): p. 1377-81.

401 Huang J, H Wium, KB Qvist and KH Esbensen (2003) Multi-way methods in image analysis-relationships and applications*, Chemometrics and Intelligent Laboratory Systems*, 66 (2): p. 141-158.

402 Péres-Nevado F, H Albergaria, T Hogg and F Gírio (2006) Cellular death of two non-Saccharomyces wine-related yeasts during mixed fermentations with Saccharomyces cerevisiae*, International Journal of Food Microbiology*, 108 (3): p. 336-345.

403 Rice-Evans CA, NJ Miller and G Paganga (1996) Structure-antioxidant activity relationships of flavonoids and phenolic acids*, Free Radical Biology and Medicine*, 20 (7): p. 933-956.

404 Hendrickson HP, M Sahafayen, MA Bell, AD Kaufman, ME Hadwiger and CE Lunte (1994) Relationship of flavonoid oxidation potential and effect on rat hepatic microsomal metabolism of benzene and phenol*, Journal of Pharmaceutical and Biomedical Analysis*, 12 (3): p. 335-341.

405 Shinohara T, S Kubodera and F Yanagida (2000) Distribution of phenolic yeasts and production of phenolic off-flavors in wine fermentation, *Journal of Bioscience and Bioengineering*, 90 (1): p. 90-97.

406 Herrmann KM and LM Weaver (1999) The shikimate pathway, *Annual Review of Plant Biology*, 50 (1): p. 473-503.

407 Mathew S, TE Abraham and S Sudheesh (2007) Rapid conversion of ferulic acid to 4-vinyl guaiacol and vanillin metabolites by *Debaryomyces hansenii, Journal of Molecular Catalysis B: Enzymatic*, 44 (2): p. 48-52.

408 Braus GH (1991) Aromatic amino acid biosynthesis in the yeast *Saccharomyces cerevisiae*: a model system for the regulation of a eukaryotic biosynthetic pathway, *Microbiology and Molecular Biology Reviews*, 55 (3): p. 349.

409 Ludovico P, MJ Sousa, MT Silva, C Leão and M Corte-Real (2001) *Saccharomyces cerevisiae* commits to a programmed cell death process in response to acetic acid, *Microbiology*, 147 (9): p. 2409.

410 Jerez CA (2008) The use of genomics, proteomics and other OMICS technologies for the global understanding of biomining microorganisms, *Hydrometallurgy*, 94: p. 162-169.

411 Bellew M, M Coram, M Fitzgibbon, M Igra, T Randolph, P Wang, D May, J Eng, R Fang, C Lin, J Chen, D Goodlett, J Whiteaker, A Paulovich and M McIntosh (2006) A suite of algorithms for the comprehensive analysis of complex protein mixtures using high-resolution LC-MS, *Bioinformatics*, 22 (15): p. 1902-9.

412 Li XJ, EC Yi, CJ Kemp, H Zhang and R Aebersold (2005) A software suite for the generation and comparison of peptide arrays from sets of data collected by liquid chromatography-mass spectrometry, *Mol Cell Proteomics*, 4 (9): p. 1328-40.

413 Zhang D, J Suen, Y Zhang, Y Song, Z Radic, P Taylor, MJ Holst, C Bajaj, NA Baker and JA McCammon (2005) Tetrameric mouse acetylcholinesterase: continuum diffusion rate calculations by solving the steady-state Smoluchowski equation using finite element methods, *Biophysical journal*, 88 (3): p. 1659-1665.

414 Ballardini R, M Benevento, G Arrigoni, L Pattini and A Roda (2011) MassUntangler: a novel alignment tool for label-free liquid chromatography-mass spectrometry proteomic data, *J Chromatogr A*, 1218 (49): p. 8859-68.

415 Neuweger H, SP Albaum, M Dondrup, M Persicke, T Watt, K Niehaus, J Stoye and A Goesmann (2008) MeltDB: a software platform for the analysis and integration of metabolomics experiment data, *Bioinformatics*, 24 (23): p. 2726-32.

416 Antonov AV, S Dietmann, P Wong and HW Mewes (2009) TICL–a web tool for network-based interpretation of compound lists inferred by high-throughput metabolomics, *FEBS J*, 276 (7): p. 2084-94.

417 Carroll AJ, MR Badger and A Harvey Millar (2010) The MetabolomeExpress Project: enabling web-based processing, analysis and transparent dissemination of GC/MS metabolomics datasets, *BMC Bioinformatics*, 11: p. 376.

418 Peters S, E van Velzen and HG Janssen (2009) Parameter selection for peak alignment in chromatographic sample profiling: objective quality indicators and use of control samples, *Anal Bioanal Chem*, 394 (5): p. 1273-81.

419 Eichelberger JW, WL Budde and LE Slivon (1987) Existence of self chemical ionization in the ion trap detector, *Analytical Chemistry*, 59 (22): p. 2730-2732.

420 March RE (1997) An introduction to quadrupole ion trap mass spectrometry, *Journal of Mass Spectrometry*, 32 (4): p. 351-369.

421 Cozzolino D, WU Cynkar, N Shah, RG Dambergs and PA Smith (2009) A brief introduction to multivariate methods in grape and wine analysis, *International Journal of Wine Research*, 1: p. 123-130.

422 Martins RC, VV Lopes, AA Vicente and JA Teixeira (2008) Computational shelf-life dating: Complex systems approaches to food quality and safety, *Food and Bioprocess Technology*, 1 (3): p. 207-222.

423 Conlin A, E Martin and A Morris (2000) Confidence limits for contribution plots, *Journal of Chemometrics*, 14: p. 725-736.

424 Qin S (2003) Statistical process monitoring: basics and beyond, *Journal of Chemometrics*, 17: p. 480-502.

425 Hotteling H (1947) *Multivariate Quality Control Illustrated by the Air Testing of Sample Bombsites*, New York: McGraw-Hill. p. 111.

426 Kourti T and JF MacGregor (1996) Multivariate SPC methods for process and product monitoring, *Journal of Quality Technology*, 28: p. 409-427.

427 Sinkhorn R (1964) A relationship between arbitrary positive matrices and doubly stochastic matrices, *The Annals of Mathematical Statistics*, 35 (2): p. 876-879.

428 Rodgers JL and WA Nicewander (1988) Thirteen ways to look at the correlation coefficient, *The American Statistician*, 42 (1): p. 59-66.

429 Rodgers JL and WA Nicewander (1988) Thirteen ways to look at the correlation coefficient, *American Statistical Association*, 42 (1): p. 59-66.

430 Hofmann T and P Schieberle (2000) Formation of aroma-active Strecker-aldehydes by a direct oxidative degradation of Amadori compounds, *Journal of Agricultural and Food Chemistry*, 48 (4301-4305).

431 Mason RL, ND Tracy and JC Young (1995) Decomposition of T2 for multivariate control chart interpretation, *Journal of quality technology*, 27 (2): p. 99-1108.

432 Mason RL and JC Young (1999) Improving the sensitivity of the T2 statistic in multivariate process control*, Journal of quality technology*, 31 (2): p. 155-165.

433 Maillard B (1971) *Additions radicalaires de diols et de leurs dérivés: diesters et acétals cycliques*, University Bordeaux 1 - France: Bordeaux. p. 325.

434 Pereira AC, MS Reis, PM Saraiva and JC Marques (2010) Aroma trends in GC7MS profiles of liqueur wines*, Analytica Chimica Acta*, 659: p. 93-101.

435 Müller C, R Kepner and A Webb (1978) 1,3-dioxanes and 1,3-dioxolanes as constituents of the acetal fraction of Spanish fino sherry*, American Journal of Enology and Viticulture*, 29: p. 207-212.

SCIENTIFIC OUTPUT

Papers in International Journals

Castro CC *et al.*, Classification and prediction of *Saccharomyces cerevisiae* industrial strains behavior under induced stress conditions based on metabolic profiles, *To be submitted.*

Castro CC *et al.*, Phenotypic and physiological characterization of laboratorial and industrial *Saccharomyces cerevisiae* strains as a response to induced stress conditions, *To be submitted.*

Castro CC, Martins RC, Teixeira JA, Silva-Ferreira AC (2014) Application of a high-throughput process analytical technology metabolomics pipeline to *Port* wine forced aging process, *Food Chemistry*, 143, 384-391.

Castro CC, Gunning C, Oliveira CM, Couto AJ, Teixeira JA, Martins RC, Silva-Ferreira AC (2012) *Saccharomyces cerevisiae* oxidative response evaluation by cyclic voltammetry and gas chromatography–mass spectrometry, *Journal of Agricultural and Food Chem*istry, 60 (29), 7252–7261.

Oral Communications

Oliveira CM, **Castro CC**, Gunning C, Couto J, Teixeira JA, Martins RC, Silva Ferreira AC (2011) *Saccharomyces cerevisiae* oxidative response by cyclic voltammetry and GC-MS, *9ème Symposium International d'OEnologie de Bordeaux*, Bordeaux, France.

Martins RC, **Castro CC**, Teixeira JA, Silva-Ferreira AC (2009) The X-windows high-throughput mass spectroscopy pipeline, In: *Encontro Nacional de Cromatografia 6*, 51p., Funchal, Portugal.

Papers in Conference Proceedings

Silva J, **Castro CC**, Vicente AA, Tafulo P, Jorge P, Martins RC (2011) Yeast metabolic state identification using micro-fiber optics spectroscopy, *Proceedings of International Conference on Applications of Optics and Photonics,* Portugal, pp 8001-202.

Abstracts in Conference Proceedings

Castro CC, Silva JS, Silva SS, Macieira da Silva F, Oliveira JM, Martins RC, Teixeira JÁ (2011) Fermentative metabolism of yeast strains under induced stress conditions. *Semana da Escola de Engenharia*, Guimarães, Portugal.

Castro CC, Teixeira JA, Silva-Ferreira AC, Martins RC, (2010) *X-Metabolomics*: A Software Tool for Dynamic Mass Spectroscopy Automated Signal Processing*, ICSB - 11th International Conference on Systems Biology*, Edinburgh, UK.

Martins RC, **Castro CC**, Silva-Ferreira AC, Calapez A, Lopes VV (2010) OpenPAT: Process analytical technometrics framework for unscrambling complex biological information, *ICSB - 11th International Conference on Systems Biology*, Edinburgh, UK.

Castro CC, Lopes VV, Silva JS, Teixeira JA, Martins RC (2008) Feasibility of Monitoring S. cerevisiae morphological state by difusive reflectance UV-VIS-SWNIR spectroscopy, *26th SMYTE (Small Meeting on Yeast Transport and Energetics)*, Braga, Portugal.