



HAL
open science

Identification et cartographie des métabolites secondaires de nouvelles souches de levures tropicales.

Olivier Perruchon

► **To cite this version:**

Olivier Perruchon. Identification et cartographie des métabolites secondaires de nouvelles souches de levures tropicales.. Chimie organique. Normandie Université, 2021. Français. NNT : 2021NORMR087 . tel-03655135

HAL Id: tel-03655135

<https://theses.hal.science/tel-03655135v1>

Submitted on 29 Apr 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Normandie Université

THÈSE

Pour obtenir le diplôme de doctorat

Spécialité CHIMIE

Préparée au sein de l'Université de Rouen Normandie

Identification et cartographie des métabolites secondaires de nouvelles souches de levures tropicales.

Présentée et soutenue par
OLIVIER PERRUCHON

Thèse soutenue le 10/12/2021
devant le jury composé de

MME SOIZIC PRADO	PROFESSEUR DES UNIVERSITES, MUSEUM HISTOIRE NATURELLE	Rapporteur du jury
M. DAVID TOUBOUL	DIRECTEUR DE RECHERCHE, Université Paris-Saclay	Rapporteur du jury
M. CARLOS AFONSO	PROFESSEUR DES UNIVERSITES, Université de Rouen Normandie	Membre du jury
M. PIERRE CHAMPY	PROFESSEUR DES UNIVERSITES, Université Paris-Saclay	Membre du jury
MME ISABELLE SCHMITZ-AFONSO	INGENIEUR DE RECHERCHE (RECHERCHE CNRS), Université de Rouen Normandie	Membre du jury
M. ABDELHAKIM ELOMRI	PROFESSEUR DES UNIVERSITES, Université de Rouen Normandie	Directeur de thèse

Thèse dirigée par **ABDELHAKIM ELOMRI, CHIMIE ORGANIQUE, BIOORGANIQUE, REACTIVITE, ANALYSE**

“So, come outside, it’s time to see the tide

It’s out of sight, but never out of mind”

Oliver Sykes

Remerciements

Les personnes citées ici sauront à quel point je peux leur être reconnaissant, donc je ferai relativement simple. Alors c'est parti.

Je tiens en premier lieu à remercier vivement Pr. Soizic Prado et Dr. David Touboul d'avoir accepté d'être rapporteurs de mes travaux de thèse et de participer à mon jury. Je remercie également Pr. Pierre Champy d'avoir accepté de faire partie de mon jury de thèse en tant qu'examineur.

Je remercie l'équipe du CIRM-Levures : Cécile Grondin, Serge Casaregola, plus récemment Jean-Luc Legras, ainsi que toutes les personnes que j'ai pu rencontrer lors de ma visite à Jouy-en-Josas. Merci pour votre aide et pour m'avoir permis de replonger le temps d'une journée dans l'univers de la microbiologie. C'était un plaisir de participer à ce projet avec vous !

Un grand merci aux membres de l'équipe 1 qui ont fait de ces trois ans une expérience humaine aussi intéressante (et fun) qu'elle peut l'être professionnellement. Merci à Corinne, Hélène et Marie ; merci à Pr. Carlos Afonso pour ses conseils à propos des différentes réalisations et d'avoir accepté de participer à mon jury. A ceux qui pour moi font partie des meubles : Clément, Estelle, Julien, Matthieu, Valentina et Sébastien. A ceux qui, partis vers de nouveaux horizons, ont le don de nous manquer : Johann et Valentin. A ceux en cours de thèse à qui l'avenir ne peut que sourire : Charlotte, Nathaniel, Anaïs, Ael, Aurélien et Hend ! Enfin, Maxime, notre petit nouveau, bienvenue chez nous ; et bien sûr Oscar, qui ne veut certainement pas de dédicace personnelle : en voilà une.

Un grand merci à tous mes amis proches, ma bande du collège et du lycée (Chloé, Manon, Ophélie, Narindra, Matthieu, Caroline), l'équipée de la fac (Fabien, Florian, Matthieu, Charlotte, Julie), les copains enseignants de SVT (Léna, Nicolas, Camille, Pierre, et surtout James pour m'avoir prêté ton bureau !). Merci pour toutes ces années, pourvu que ce jeu dure encore un peu. Maxime et Idir, cela fait tellement longtemps que je pense ne rien avoir à dire de plus.

Un immense merci à ma co-encadrante Mme Isabelle Schmitz-Afonso pour son aide, pour m'avoir formé sur les instruments, pour le temps passé à relire mes productions ; également une éternelle reconnaissance à mon directeur de thèse Pr. Abdelhakim Elomri pour le soutien, l'aide durant le projet, et pour m'avoir permis de rentrer dans le monde de l'enseignement.

Je ne peux exprimer en ces lignes tout ce que je dois à mes parents, Evelyne et Philippe. Alors, de manière sobre, merci pour tout ce que vous avez pu faire, c'était top. Profitez bien de votre retraite !

Et un mot final pour Caroline, l'autre face de la pièce : merci d'être là, tout simplement.

Résumé

Cette thèse s'intitule « Identification et cartographie des métabolites secondaires de nouvelles souches de levures tropicales ». L'objectif principal est le développement d'une méthodologie de déréplication pour l'exploration des métabolites secondaires extraits de cultures de champignons unicellulaires. La méthode développée fait intervenir la chromatographie liquide à ultra-haute performance couplée à la spectrométrie de masse avec l'outil des réseaux moléculaires en complémentarité de la spectrométrie de masse à ultra-haute résolution. La méthodologie permet d'annoter les métabolites détectés par le procédé analytique, c'est-à-dire formuler des hypothèses sur l'identité et la structure de nombreuses molécules produites par les nouvelles espèces de levures sur lesquelles elle est appliquée. De plus, des cartographies et profils métaboliques sont dressés grâce aux diagrammes et réseaux moléculaires construits. L'application de la méthodologie à de nouvelles souches de levures a permis l'annotation de plusieurs centaines de métabolites, parmi lesquels une large proportion d'alcaloïdes mais aussi des acides aminés, des terpénoïdes et des stérols.

Mots-clés : déréplication ; métabolomique non ciblée ; UHPLC-MS/MS ; réseaux moléculaires ; FTICR-MS.

Summary

This PhD thesis project is named "Identification and cartography of secondary metabolites from new tropical yeast strains". The main objective is the development of a dereplication methodology and the application of it to the exploration of the secondary metabolites from unicellular fungal extracts. The developed method implies ultra-high-performance liquid chromatography hyphenated with mass spectrometry with the use of the molecular networking tool in combination with the ultra-high-resolution mass spectrometry. The methodology allows the annotation of the detected metabolites from acquired data, that is to say the formulation of hypotheses about the structure and identity of the numerous molecules produced by the analyzed yeast. Moreover, cartography and metabolic profiles are provided thanks to the built diagrams and molecular networks. The application of the methodology to new yeast strains led to the annotation of hundreds of metabolites, with a rich content in alkaloids, but also in amino-acids, terpenoids and sterols.

Keywords: dereplication; untargeted metabolomics; UHPLC-MS/MS; molecular networking; FTICR-MS.

Sommaire / Table des matières

Sommaire / Table des matières	9
Abréviations	12
Introduction	15
Préambule : Les produits naturels et les moyens analytiques applicables à leur exploration	19
I. La nature, source essentielle pour l'humanité.....	20
1. Les substances naturelles	20
2. La biodiversité structurale	21
II. Les levures	26
1. Définitions.....	26
2. <i>Saccharomyces cerevisiae</i> , un modèle omniprésent.....	26
3. Autres genres de levures	27
III. La métabolomique	29
1. Définition et historique.....	29
2. Le profilage métabolique.....	30
3. Différentes approches	31
4. La déréplication	32
IV. Les méthodes d'extraction de métabolites appliquées aux microorganismes	33
1. Arrêt de la prolifération cellulaire	34
2. Disruption mécanique des cellules	35
3. Élimination des débris cellulaires	36
4. Influence de la température sur l'extraction.....	36
5. Concentration et reprise des extraits	37
V. Les techniques analytiques utilisées en métabolomique	37
1. Aperçu des méthodes employées	38
2. La séparation chromatographique des composés d'un mélange	39
3. Les méthodes spectroscopiques.....	46
4. L'application de la résonance magnétique nucléaire à la métabolomique	52
5. La spectrométrie de masse	56
6. Les possibilités d'injection en spectrométrie de masse	65
7. Bilan du profilage métabolique de <i>S. cerevisiae</i>	74
8. Traitement de données en métabolomique.....	79
Chapitre I : Développement de la méthodologie de déréplication de nouvelles levures.	87
I. Sélection des souches.....	88

II.	Développement du protocole d'extraction des métabolites secondaires.....	92
1.	Culture de levures.....	92
2.	Extrait apolaire de levures.....	93
3.	Extrait polaire de levures.....	93
4.	Extrait du témoin, le milieu de culture sans levures.....	94
5.	Préparation des échantillons pour analyse.....	94
III.	Analyse d'extraits de levures pour évaluation de la méthode d'extraction.....	94
1.	Etude préliminaire d'extraits de <i>S. cerevisiae</i>	95
2.	Vérification de l'efficacité de l'extraction par chromatographie sur couche mince.....	96
IV.	Optimisation de la méthode d'extraction.....	98
	Chapitre II : Optimisation de la méthode analytique et application du traitement de données pour l'exploration d'extraits de levures.....	101
I.	Les réseaux moléculaires pour la métabolomique.....	102
1.	Conversion des données spectrales.....	102
2.	Construction des réseaux.....	105
II.	Application et optimisation des outils de traitement de données.....	106
1.	MZMine2 pour le prétraitement des données d'analyse UHPLC-MS/MS.....	106
2.	Comparaison de l'outil UNIFI avec MZMine2 pour la détection des pics.....	107
3.	Optimisation du prétraitement de données spectrales.....	108
4.	Traitement de données : <i>feature-based molecular networking</i>	112
III.	Finalisation de la méthode.....	114
1.	Comparaison des différents extraits.....	114
2.	Paramètres de chromatographie liquide.....	116
3.	Modification du protocole d'extraction pour regrouper les extraits.....	117
4.	Analyse par GC-MS de l'extrait apolaire.....	117
5.	Spectrométrie de masse et problème de fragmentation en source.....	120
6.	Vérification des triplicats d'analyse de <i>S. cerevisiae</i>	123
IV.	Résultats sur <i>Saccharomyces cerevisiae</i>	124
1.	Annotation des composés.....	126
2.	Investigation manuelle des résultats.....	128
	Chapitre III : Ajout de la FTICR-MS et approfondissement du traitement de données pour application sur <i>S. cerevisiae</i>.	135
I.	La FTICR-MS appliquée à la déréplication.....	136
1.	Optimisation pratique de la méthode.....	137
2.	Corrélation entre résultats UHPLC-MS/MS et FTICR-MS.....	138
3.	Traitement de données FTICR-MS.....	139
II.	Résultats de la déréplication de <i>S. cerevisiae</i> : publication des résultats.....	141

Chapitre IV : Application de la méthode de déréplication à de nouvelles souches de levures.	169
I. Ajout de nouveaux outils de traitement de données	170
1. Utilisation de modules supplémentaires de MZMine2	170
2. Application d'outils intégrés à la plateforme GNPS	170
II. Résultats de l'application de la méthode finalisée à deux souches récemment découvertes : <i>Starmerella reginensis</i> et <i>Starmerella kourouensis</i>	171
III. Perspectives : <i>Kurtzmaniella hittingeri</i> comme support de l'amélioration de la gamme de composés détectés	197
1. Confirmation de l'efficacité de conservation d'extraits secs	197
2. Perspectives : améliorations analytiques par emploi de différentes sources d'ionisation	198
Conclusion	201
Matériel et méthodes	205
Références	212
Annexe 1	224
Annexe 2	239

Abréviations

Bien que les termes employés soient en français, les abréviations utilisées dans cet ouvrage correspondent principalement aux termes anglais, afin d'uniformiser le langage utilisé avec celui tiré de la littérature scientifique.

ACN : acétonitrile.

ANP-LC-MS : chromatographie liquide en phase normale aqueuse couplée à la spectrométrie de masse.

APCI : ionisation chimique à pression atmosphérique (*Atmospheric Pressure Chemical Ionization*).

APPI : photoionisation à pression atmosphérique (*Atmospheric Pressure Photoionization*)

BPC ou BPI : chromatogramme du pic de base (*Base Peak Chromatogram* ou *Base Peak Ion chromatogram*).

CE : électrophorèse capillaire (ou *capillary electrophoresis*).

CE-MS : électrophorèse capillaire couplée à la spectrométrie de masse.

DI-MS : spectrométrie de masse en introduction ou injection directe (ou *direct injection-mass spectrometry*).

DNP : *Dictionary of Natural Products*, base de données.

EI : ionisation électronique.

EIC ou XIC : chromatogramme d'ion extrait (*Extracted Ion Chromatogram*).

ESI : ionisation par électronébulisation (*electrospray ionisation*).

Extrait DCM/AE : extrait dichlorométhane / acétate d'éthyle de levures.

Extrait M : extrait méthanolique de levures.

FBMN : *feature-based molecular networking* ou réseaux moléculaires basé sur les variables.

Feature : traductible par le terme « variable », il s'agit d'une entité spectrale décrite au cours du traitement de données de spectrométrie de masse par sa mesure de masse (MS1), sa fragmentation (MS2) et son temps de rétention.

FTICR : analyseur à résonance cyclotronique ionique par transformée de Fourier (*Fourier Transform Ion Cyclotron Resonance*).

FTIR : spectroscopie infrarouge à transformée de Fourier.

GC-MS : chromatographie en phase gazeuse couplée à la spectrométrie de masse.

GNPS : *Global Natural Products Social molecular networking*, plateforme en ligne de création de réseaux moléculaires disposant d'un grand nombre d'outils pour l'annotation et l'interprétation de données spectrales.

H/C : rapport hydrogène / carbone.

HILIC : chromatographie liquide d'interactions hydrophiles.

HPLC : chromatographie liquide à haute performance.

HS-SPME : *Headspace Solid Phase Micro Extraction*, ou micro-extraction par sorption sur phase solide dans l'espace de tête.

LC-MS : chromatographie en phase liquide couplée à la spectrométrie de masse.

MRM : *Multiple Reaction Monitoring*, mode d'acquisition ciblée de spectres MS/MS.

MS : spectrométrie de masse.

MS1 : le spectre de niveau MS1 représente une acquisition d'une gamme de masse donnée, sans fragmentation.

MS/MS ou MS2 : spectrométrie de masse en tandem ; spectre résultant de la fragmentation d'un ion.

m/z : rapport masse sur charge, utilisé pour indiquer un ion en spectrométrie de masse.

MSI : *Metabolomics Standard Initiative*, guide de niveaux d'annotations en métabolomique établie en 2007.

NAP : *Network Annotation Propagation*, outil d'analyse de réseaux moléculaires sur GNPS.

NIST : *National Institute of Standards and Technology*, nom donné à la base de données spectrales couramment utilisée pour l'annotation de composés analysés en GC-MS.

O/C : rapport oxygène / carbone.

PCA : analyse par composante principale.

PLS-DA : régression des moindres carrés partiels.

Analyseur Q : analyseur quadripolaire, ou quadrupôle.

QTOF : analyseur quadrupôle couplé à un analyseur temps de vol.

RID : détecteur à indice de réfraction, ou réfractomètre différentiel.

RMN : résonance magnétique nucléaire.

RP-LC-MS : chromatographie liquide en phase inverse couplée à la spectrométrie de masse.

SIM : *Single Ion Monitoring*, mode de sélection utilisé pour la quantification de composés ciblés d'un échantillon en spectrométrie de masse, qui permet de suivre un rapport m/z donné.

S/N : rapport signal sur bruit.

TIC : chromatogramme en courant ionique total (*Total Ion Current*).

TOF : analyseur à temps de vol (Time Of Flight).

t-SNE : *t-distributed Stochastic Neighbor Embedding*, algorithme utilisé par le logiciel français de réseaux moléculaires MetGem.

UHPLC (ou UPLC) : chromatographie liquide à ultra-haute performance.

UV-Vis : ultraviolet-visible.

YPD : *Yeast Peptone Dextrose*, milieu de culture.

Introduction

D'une époque antérieure à l'Antiquité jusqu'à aujourd'hui, l'humanité a su tirer profit de son environnement. La nature est pour l'être humain tantôt un modèle, tantôt un puits de ressources. Celle-ci sont nécessaires à la survie, à l'alimentation, à la construction ou encore aux soins des maux. Les effets recherchés au travers de l'utilisation de sources naturelles pour la nutrition ou la fabrication de remèdes sont bien souvent liés à des entités chimiques présentes à l'intérieur des ressources citées ou sécrétées par celles-ci : les substances naturelles.

Les substances naturelles peuvent avoir de multiples rôles. Dans l'alimentation, elles peuvent être à l'origine de certaines propriétés organoleptiques, tel que les arômes. Elles peuvent également être employées pour le traitement de maladies ou des symptômes de celles-ci. Enfin, si elles ne sont pas utilisées telles quelles, elles sont potentiellement des modèles pour des procédés d'hémisynthèse : une substance naturelle est utilisée comme base puis modifiée artificiellement à l'aide de réactions chimiques pour obtenir une structure bien précise. Les métabolites secondaires, souvent retrouvés au sein des substances naturelles, sont des produits ou intermédiaires de voies métaboliques et ont des rôles divers pour les êtres vivants, comme la communication ou la défense. Bien qu'ils ne participent pas directement à la nutrition, le maintien ou la reproduction, ils sont essentiels et la variété d'effets de ces composés n'a d'égal que la diversité structurale qu'ils sont en mesure de présenter.

On retrouve parmi les sources naturelles les plus exploitées un grand nombre d'espèces appartenant au règne végétal. De nombreuses cultures utilisent des plantes dans des procédés de médecine traditionnelle, et une multitude de molécules utilisées dans l'industrie pharmaceutique en proviennent également. Cependant, les microorganismes ne sont pas en reste et parmi eux, les champignons unicellulaires que l'on nomme « levures » sont impliqués dans de multiples activités humaines. Les levures sont utilisées pour la fermentation de boissons alcoolisées, ou encore en tant qu'organismes transformés génétiquement pour servir d'usines à molécules pharmaceutiques. De nombreuses espèces de levures continuent à être découvertes et caractérisées au fil des ans. Cette flore levure représente un vivier de molécules potentiellement bioactives et intéressantes à plusieurs niveaux, dans l'univers de l'industrie comme du côté de la recherche scientifique.

C'est dans ce contexte que s'inscrit le projet de ma thèse présentée dans cet ouvrage. Celle-ci s'intitule « Identification et cartographie des métabolites secondaires de nouvelles souches de levures tropicales ». L'objectif est le développement d'une méthodologie de dérégulation combinant des technologies de spectrométrie de masse à haute résolution avec des outils bio-informatiques afin d'explorer les métabolites secondaires de souches de levures exotiques récemment découvertes.

La première phase de la thèse se focalisera sur le développement de la méthodologie qui sera appliquée à l'exploration des extraits de cultures de microorganismes. L'extraction des métabolites est une première étape cruciale du procédé et n'est pas à négliger, les métabolites détectés par les techniques analytiques utilisées dépendent grandement du protocole appliqué. La principale méthode analytique envisagée implique l'utilisation de la chromatographie en phase liquide à haute performance couplée à la spectrométrie de masse (UHPLC-MS/MS). Cette technique est utilisée couramment en métabolomique, la discipline qui se concentre sur l'étude du métabolome d'un organisme. Elle permet la déréplication des échantillons analysés, c'est-à-dire l'analyse et l'identification des composants d'un mélange complexe sans la nécessité de les isoler au préalable les uns des autres. Les petites molécules sont séparées selon leurs propriétés physico-chimiques puis ionisées. La spectrométrie de masse permet de détecter les molécules après la séparation chromatographique et d'obtenir des informations sur leurs masses. Cette caractéristique intrinsèque à chaque composé et les informations des fragments formés pendant l'expérience permettent de déterminer la formule moléculaire des analytes et de formuler des hypothèses quant à leurs structures.

La méthodologie développée tirera profit des dernières avancées en termes d'outils bioinformatiques pour le traitement de données. Plus particulièrement, les données spectrales serviront à la construction de réseaux moléculaires via l'interface en ligne de *Global Natural Products Social molecular networking* (GNPS). Ce système utilise l'information de fragmentation des molécules fournie par la spectrométrie de masse pour regrouper les composés dont la structure est similaire dans une représentation en réseaux connectés. En effet, des substances faisant partie d'une famille chimique commune auront des groupements spécifiques et se fragmenteront de façon semblable dans l'instrument. Lesdites substances sont alors représentées sous forme de nœuds et connectées entre elles. Les réseaux moléculaires permettent donc la cartographie des métabolites des échantillons analysés.

Plusieurs méthodes analytiques seront étudiées afin de réaliser les explorations les plus exhaustives possibles des nouvelles souches de levures. Le laboratoire COBRA dispose d'équipements analytiques de pointe aux principes variés. Les instruments de spectrométrie de masse peuvent être couplés avec les systèmes de chromatographie en phase liquide ou en phase gazeuse. La spectrométrie de masse à haute résolution est disponible via des analyseurs à temps de vol, tandis que la spectrométrie de masse à résonance cyclotronique ionique rend accessible les mesures à ultra-haute résolution. Les différentes techniques seront appréciées afin de jauger les combinaisons les plus adéquates à nos objectifs de déréplication d'extraits naturels.

La méthodologie développée sera appliquée à de nouvelles souches de levures récoltées récemment lors d'une exploration de la flore levure en Guyane française et caractérisées par notre partenaire scientifique, le CIRM-Levures. Nous avons donc accès à de nombreuses nouvelles espèces, parmi lesquelles peuvent être citées *Kurtzmaniella hittingeri*, *Starmerella reginensis* ou *Starmerella kourouensis*.

Le présent ouvrage sera organisé en plusieurs chapitres. Le préambule détaillera le contexte scientifique ainsi que les principes analytiques applicables dans les études métabolomiques menées durant cette thèse. Les chapitres suivants seront le moyen d'expliquer la manière dont l'extraction des métabolites a été mise au point, puis comment la méthodologie a été développée. Les techniques analytiques et les différentes étapes de l'optimisation de celles-ci seront également présentées. Enfin, les résultats de l'exploration des souches de levures seront communiqués.

Préambule : Les produits naturels et les moyens
analytiques applicables à leur exploration

I. La nature, source essentielle pour l'humanité

Depuis des temps immémoriaux, l'humanité, qu'elle soit de culture nomade où implantée en civilisation sédentaire, a su tirer profit et exploiter les ressources naturelles de son environnement. Elle a commencé par subvenir à ses besoins primaires en nourriture par la chasse et la cueillette, ou en abri et habitation par l'exploitation de matériaux tels que le bois et les roches. L'espèce humaine a ensuite diversifié ses activités et son utilisation des ressources naturelles, par exemple dans la fermentation de breuvages par des microorganismes ou l'utilisation de plantes médicinales pour soigner les blessures et les maux [1].

1. Les substances naturelles

Les substances naturelles sont des molécules qui jouent un rôle majeur dans l'histoire et l'évolution de l'humanité, et continueront dans cette voie si l'exploitation de celles-ci se fait de manière raisonnée et durable. Ces molécules correspondent à des produits ou des intermédiaires faisant partie des voies métaboliques des organismes vivants et utiles voire essentiels à leur fonctionnement : ce sont donc des métabolites.

Le métabolisme primaire correspond aux voies essentielles à la survie des individus et des espèces. Il intervient dans le développement d'un organisme, son maintien dans l'environnement, sa reproduction et le transfert de l'information génétique à sa descendance. Un grand nombre de métabolites primaires est identique dans tous les êtres vivants, car ce sont les briques universelles qui permettent le maintien des espèces.

Les métabolites secondaires sont des composés élaborés à partir de ces briques de construction provenant du métabolisme primaire. Bien qu'ils n'interviennent pas dans les voies métaboliques essentielles, ils permettent aux organismes une meilleure adaptation à leur environnement [2]. Ils leur servent dans la communication, dans la défense, ou dans leurs relations avec d'autres individus de la même espèce voire d'espèces différentes. Par exemple, certaines molécules de plantes permettent la répulsion de prédateurs, la défense contre certains pathogènes ou encore l'attraction d'insectes pollinisateurs et d'animaux disséminant les graines au gré de leurs activités [3–6]. Ces molécules sont spécifiques à chaque espèce et ont une distribution restreinte dans la nature ; ainsi, la plupart des espèces de plantes, de bactéries ou de champignons, et dans quelques cas des espèces animales,

produiront leurs propres métabolites secondaires. Cela conduit à une colossale diversité structurale de composés potentiellement produits par les sources naturelles [7].

2. La biodiversité structurale

Les métabolites secondaires diffèrent des métabolites primaires par leur fonction et par leur distribution dans la nature. Certains métabolites secondaires sont produits par des voies de biosynthèse qui produisent également des métabolites primaires qui sont donc structurellement très proches. Ils peuvent alors appartenir aux mêmes familles chimiques de molécules [8]. On retrouvera néanmoins, parmi une grande diversité de classes chimiques, quelques grandes familles de produits naturels ayant fait l'objet de nombreux travaux. La majeure partie d'entre eux est originaire d'espèces végétales, car ce sont les principales sources naturelles ayant été étudiées à ce jour [9]. Un exemple de la diversité des origines des produits naturels répertoriés dans la base de données « *Dictionary of Natural Products* » est présenté en Figure 1.

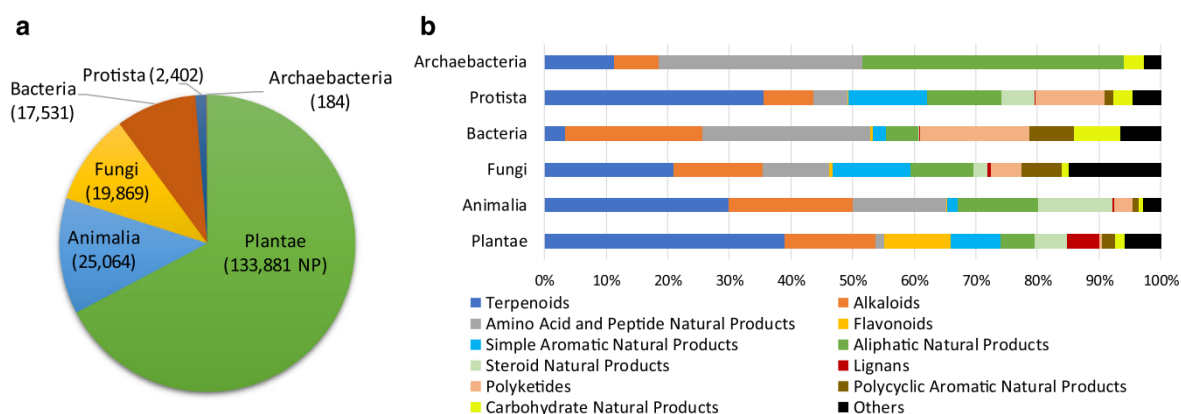


Figure 1 : répartition des produits naturels répertoriés dans le *Dictionary of Natural Products*. A : distribution des produits naturels en fonction du « royaume » du vivant d'origine. B : distribution des classes chimiques répertoriées de ces produits naturels, en fonction du « royaume » [9].

a. Quelques exemples de produits naturels

Trois familles majeures de métabolites secondaires représentent souvent les branches principales de la classification des produits naturels : les composés phénoliques, les terpénoïdes et les alcaloïdes [8,10]. Cependant, de nombreuses classes chimiques sont représentées parmi les produits naturels, comme indiqué en Figure 1.

Les composés phénoliques et associés

De nombreux composés phénoliques font partie des métabolites secondaires. Les lignines par exemple sont omniprésentes chez les plantes vasculaires et permettant la structure et la défense des végétaux. Les flavonoïdes sont des molécules très variées qui comprennent par exemple des pigments qui assurent aux plantes une photoprotection, mais aussi l'attraction de pollinisateurs [11–13]. Certaines molécules peuvent aussi être utilisées pour leur propriétés antioxydantes, anticancéreuses ou bactéricides [14].

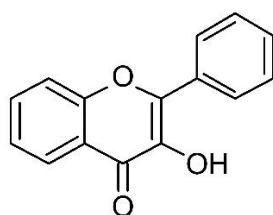


Figure 2 : exemple d'un flavonoïde, le 3-hydroxy-2-phénylchromén-4-one

Les alcaloïdes

On retrouve également une grande diversité d'alcaloïdes parmi les produits naturels. Ces molécules à bases azotées et souvent hétérocycliques, aux structures variées, proviennent pour beaucoup d'entre elles de voies de synthèse d'acides aminés [10,15]. Les alcaloïdes ont été grandement exploités par l'humanité pour leurs propriétés, allant de l'effet psychoactif, stimulant ou dopant, comme avec la caféine, à des vertus analgésiques ou anticancéreuses [8,16]. La morphine est par exemple un alcaloïde extrait de l'opium, c'est-à-dire le latex du pavot *Papaver somniferum* [17]. La quinine est un alcaloïde aux multiples effets, dont antipaludique, provenant du quinquina [18]. Les taxines sont un autre exemple de molécules de type alcaloïde. On y retrouve le paclitaxel, produit par certains champignons et présent dans l'écorce d'if, doté d'une activité anticancéreuse [19].

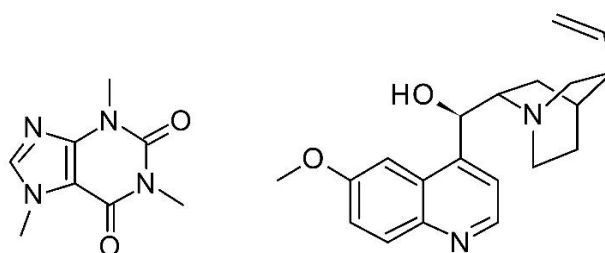


Figure 3 : exemple d'alcaloïdes, la caféine à gauche et la quinine à droite.

Les terpénoïdes

La famille des terpénoïdes correspond à des molécules dérivées de l'unité isoprène et comprend parmi d'autres des stéroïdes, des caroténoïdes et des acides gibbérelliques [8]. Les monoterpénoïdes retrouvés dans les huiles essentielles (un type d'échantillon naturel extrait de manière spécifique) dégagent en général de fortes odeurs. On trouve également des molécules servant à la défense de certaines espèces végétales [20]. Les terpénoïdes sont en partie responsables des arômes du vin, ou peuvent être utilisés pour certaines propriétés antimicrobiennes et fongicides, ce qui les rend exploitables par l'espèce humaine dans les industries agro-alimentaire, pharmaceutique et cosmétique, notamment la parfumerie [21].

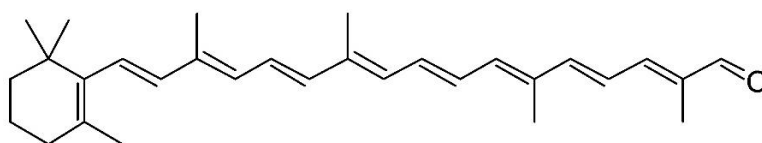


Figure 4 : exemple d'un tétraterpénoïde, l'apocaroténal.

D'autres produits naturels

La recherche autour des substances naturelles peut aussi aider dans le monde de l'agriculture, en remplaçant des pesticides par des alternatives dérivées de molécules naturelles, car certains composés synthétisés par les plantes leur permettent de se défendre contre des pathogènes dont la présence est grandement délétère pour les cultures [4].

Bien que ces exemples traitent particulièrement de molécules originaires de végétaux, Les champignons et les bactéries ne sont pas en reste. Par exemple, la famille des pénicillines est représentée par plusieurs antibiotiques de type bêta-lactamine produites par des champignons du genre *Penicillium*. La cyclosporine, elle aussi isolée de champignons microscopiques, est utilisée depuis les années 1980 comme agent immunosuppresseur.

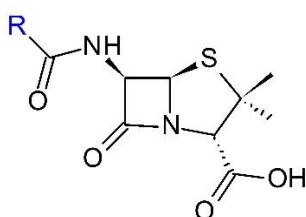


Figure 5 : exemple de produit naturel issu de champignons, la pénicilline.

Un exemple de quelques autres molécules d'origine naturelle est présenté en Figure 6.

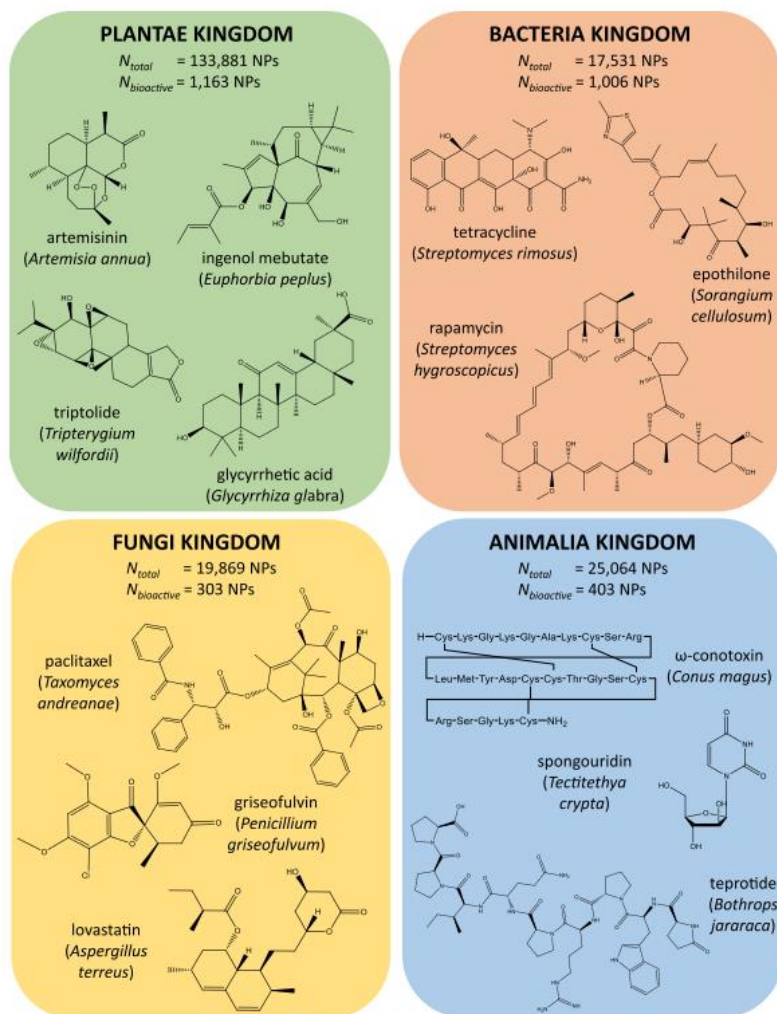


Figure 6 : exemples de structures de composés bioactifs retrouvés chez des plantes, des bactéries, des champignons ainsi que des animaux [9].

b. Les sources naturelles de métabolites secondaires

La diversité structurale des produits naturels, ou métabolites secondaires, n'a d'égal que la diversité des sources naturelles qui se présentent à nous. Historiquement, les êtres humains ont exploité un grand nombre de plantes, amenant à une prépondérance de nos jours du règne végétal dans la chimie des substances naturelles et dans la recherche de composés aux propriétés potentiellement intéressantes [9,22]. Cependant, de nombreuses espèces de bactéries et de champignons ont également été employées par l'humanité au cours des âges, ne serait-ce que par exemple dans la

production de vinaigre par la fermentation de bactéries acétiques ou la production de breuvages alcoolisés avec l'aide des voies métaboliques de champignons unicellulaires tels que les levures.

Les bactéries et les champignons représentent aujourd'hui une partie de la recherche de produits naturels. Néanmoins, la diversité des espèces existantes est si vaste que seule une infime partie des espèces connues à ce jour a été explorée par rapport à ce qu'elles peuvent produire. Parmi les 200 000 composés enregistrés dans le *Dictionary of Natural Products* (DNP), seulement 10 % d'entre eux sont d'origine fongique [9].

Les substances naturelles ont des origines variées, et les classes chimiques auxquelles elles appartiennent sont également très diverses. La répartition des différentes activités biologiques répertoriées dans le DNP en fonction de la classe chimique est présentée en Figure 7 et témoigne de la richesse d'effets possibles des produits naturels [9]. Il est donc potentiellement très utile de se pencher sur des sources naturelles peu explorées et sur les métabolites secondaires des bactéries et des champignons. Les champignons unicellulaires notamment sont très simples à cultiver et peuvent fournir des composés potentiellement intéressants dans les domaines pharmaceutique, cosmétique ou agro-alimentaire.

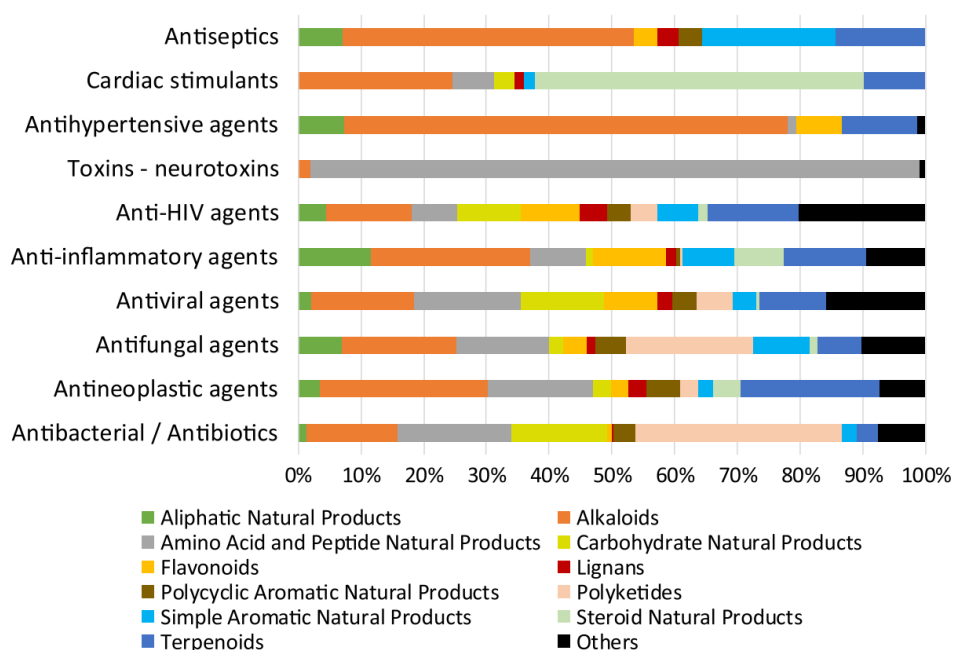


Figure 7 : distribution des principales classes chimiques des produits naturels répertoriés dans le DNP comme ayant une activité biologique importante [9].

II. Les levures

1. Définitions

Les levures sont des champignons unicellulaires capables de fermentation de matière organique pour subvenir à leurs besoins et se reproduire. Ces champignons ubiquitaires sont retrouvés sur une très large variété de substrats, allant des sols jusqu'aux insectes en passant par de nombreuses espèces de plantes [23].

Les levures sont pour la plupart apparentées soit aux ascomycètes, soit aux basidiomycètes, et les genres les plus connus sont les *Saccharomyces* (auquel appartient la levure emblématique *Saccharomyces cerevisiae*) et les *Candida*. Elles sont exploitées depuis l'Antiquité pour leurs capacités métaboliques dans la fabrication de boissons alcoolisées, telles que des bières ou des vins, ou dans la fabrication de pain et autres pâtes levées. Le terme « levure » vient d'ailleurs de l'observation de ces fermentations, la présence de ces microorganismes conduisant à la libération de CO₂ et faisant « lever » le pain. Après que Louis Pasteur ait prouvé qu'il s'agit d'organismes vivants, les levures sont également employées à l'époque contemporaine dans la fabrication de produits pharmaceutiques, d'exhausteurs de goût ou de compléments alimentaires.

2. *Saccharomyces cerevisiae*, un modèle omniprésent

L'espèce de levure la plus connue, étudiée et utilisée est *Saccharomyces cerevisiae*. Des souches de cette espèce ont été isolées sur de nombreux types de substrats naturels tels que des fleurs, des fruits, des sols ou des insectes [23–26]. De nombreuses souches appartiennent à cette espèce, comme par exemple la souche communément appelée « levure de bière » utilisée pour sa capacité de fermentation alcoolique, ou encore la souche appelée « levure de pain » utilisée dans la panification. *S. cerevisiae* est très adaptable et peut se développer sur de nombreux milieux différents.

La souche S288c de *S. cerevisiae* est très utilisée pour la recherche en laboratoire en tant que modèle d'organisme eucaryote, et a été le premier organisme eucaryote dont le génome fut déterminé par séquençage systématique. Ce génome est conservé dans la *Saccharomyces Genome Database* (SGD) [27]. Il existe ainsi une très grande diversité de souches de *S. cerevisiae*, amenant ainsi à une grande diversité génétique et à une grande diversité de produits de fermentations selon ses multiples utilisations dans les activités humaines, pour la production de pain, de vin, de bière, de thé kombucha,

de kéfir ou la fermentation du cacao [28,29]. Certaines souches sont également employées dans la production de bioéthanol utilisable comme biocarburant [30,31].

Les procédés antiques de fabrication d'aliments et de boissons à partir de l'action des levures ont servi de base à l'utilisation actuelle des levures comme outils biotechnologiques dans l'industrie. Les levures sont aujourd'hui des organismes aptes à être transformés facilement par des techniques de bio-ingénierie génétique. En effet, de nombreux travaux et de multiples procédés industriels utilisent la biologie moléculaire pour faire exprimer à *S. cerevisiae* des gènes permettant la production de produits biopharmaceutiques. La fermentation contrôlée dans des bioréacteurs est utilisée pour la production de vaccins contre l'hépatite B [32], pour la production de molécules utilisées comme vaccin contre le papillomavirus humain [33], ou encore pour la production industrielle d'albumine, d'hirudine, d'insuline et de glucagon [34].

3. Autres genres de levures

Bien que *S. cerevisiae* soit une espèce des plus représentées dans la recherche et dans l'industrie, d'autres levures appartenant à des genres différents font également l'objet d'études ou sont impliquées dans des procédés de production faisant appel à leurs capacités de fermentation. Certains genres sont très fréquents dans les travaux menés sur les microorganismes, tel que le genre *Candida*. D'autres sont cependant plus restreints, comportent moins d'espèces répertoriées et chaque nouvelle espèce peut représenter un vivier inexploré de molécules potentiellement bioactives.

Le genre *Candida* comprend un grand nombre d'espèces appartenant également à l'ordre des *Saccharomycetales*. Ces levures sont très répandues et sont commensales de l'être humain ; on les retrouve naturellement sur la peau et les muqueuses. Certaines espèces, notamment la plus connue *Candida albicans*, provoquent cependant des mycoses chez des humains et d'autres animaux quand les défenses immunitaires sont affaiblies. Des souches appartenant à ce genre sont reportées dans la littérature pour leurs activités fermentatives. Par exemple, l'espèce *Candida shehatae* est capable de fermenter des sucres, le glucose, l'hexose et le pentose, en produisant une faible quantité d'éthanol [35]. Une autre étude traite d'inoculations de *Saccharomyces cerevisiae* conjointement avec la levure *Candida stellata* pour la fermentation de vin, produisant ainsi des arômes intenses de miel et d'abricot [36].

Le genre *Starmerella* fut introduit en 1998 avec l'espèce *Starmerella bombicola*, un téléomorphe de *Candida bombicola* [37]. Un téléomorphe est le synonyme d'une autre espèce de champignon, à la

différence près qu'elle est dans un stade de reproduction sexuée ; c'est un terme utilisé en biologie, quand les champignons concernés peuvent également avoir une forme se reproduisant de manière asexuée dite « anamorphe ». Les levures du genre *Starmerella* sont décrites comme osmotolérantes, capables de fermentation, et utilisant peu de composés carbonés pour survivre et croître. Elles sont souvent, d'après la littérature, isolées à partir de fleurs et d'insectes associés à ces fleurs. Une souche de *S. bombicola* a également été isolée dans la pâte d'un gâteau italien, le *Panettone* [38]. Cette espèce a été le sujet de plusieurs études, et plus particulièrement sa capacité à produire des sophorolipides, des composés glycolipidiques tensio-actifs intéressants d'un point de vue industriel, car ils peuvent servir de base pour des synthèses moléculaires, peuvent servir de détergents et peuvent présenter des activités antimicrobiennes et anticancéreuses [39,40]. *Starmerella bacillaris*, une autre espèce du genre, a été isolée lors de la fermentation de grains de café [41]. D'autres espèces de *Starmerella* ont été caractérisées à partir du thé kombucha, une boisson fermentée asiatique ayant de multiples bienfaits rapportés, telles que des activités antioxydante, antimicrobienne, antistress et antidiabétique parmi d'autres [42,43]. Enfin, l'activité de fermentation de certaines *Starmerella* a été observée, et ces levures sont considérées comme utiles lors de la fermentation alcoolique de vins, car elles permettent d'améliorer la complexité sensorielle du vin tout en réduisant la proportion d'éthanol [44]. L'impact de *S. bacillaris* sur les profils aromatiques de vins a également été étudié par des techniques analytiques [45,46].

Tout comme le genre *Starmerella*, le genre *Kurtzmaniella* fut proposé en 2008 pour s'accommoder au stade reproductif d'une souche de *Candida cleridarum* [47]. A partir de la création du genre *Kurtzmaniella*, plusieurs formes de levures *Candida* ont été déplacées en tant que synonymes dans le genre *Kurtzmaniella* [48,49]. Bien que le genre soit nouveau, des souches de certaines espèces ont été caractérisées à partir de charcuteries fermentées européennes, de type salami, témoignant ainsi de la possibilité de production par ces levures de composés intéressants pour l'agro-alimentaire [50,51].

En tant que dernier exemple de nombreux genres de champignons considérés comme des levures, le genre *Metschnikowia* comprend des espèces principalement retrouvées sur des plantes. Elles sont isolées au niveau de fleurs, de fruits ou d'autres tissus végétaux ainsi que dans le tube digestif d'insectes associés aux plantes en question. Certaines espèces sont même aquatiques et proviennent de prélèvements réalisés sur des crustacés ou des poissons. Certaines souches sont utilisées pour la fermentation du vin, en co-culture avec d'autres levures mais aussi seules, remplaçant ainsi les autres levures dont *S. cerevisiae*. La synthèse de métabolites secondaires spécifiques aux *Metschnikowia* génère de nouveaux profils sensoriels pour la boisson fermentée [52]. Des souches de *Metschnikowia*

pulcherrima ont également été étudiées en tant qu'agents de contrôle pour limiter la prolifération de pathogènes sur des fruits après la récolte [53].

III. La métabolomique

1. Définition et historique

La métabolomique est l'un des noms donnés à l'étude du métabolome d'un organisme. Né en 1998, ce terme est construit par analogie aux termes « génomique », qui est l'étude du génome d'un organisme dans sa globalité, « transcriptomique », l'étude des ARN produits de transcription, et « protéomique », l'étude des protéines d'un organisme. Les sciences « omiques » se sont développées à la fin des années 1990. L'appellation « profilage métabolique » est cependant utilisée depuis les années 1970, mais les méthodes se développent et la discipline ne prolifère qu'à la fin du 20^{ème} siècle [54]. On retrouvera également les termes de « métabonomique » et de « profilage des métabolites ».

La métabolomique correspond aux études et aux moyens déployés pour l'analyse (qu'elle soit qualitative, quantitative ou une combinaison des deux) de l'ensemble des métabolites en général de poids moléculaire inférieur à 1500 Da, d'un organisme ou d'une partie d'un organisme et ce dans une situation donnée [55,56]. Le métabolome représente le produit final de l'expression des gènes, la définition du phénotype biochimique du sujet d'étude. Il découle donc du génome, du transcriptome et du protéome des cellules en question, d'où l'exemple de cascade des sciences omiques présentée en Figure 8 [57].

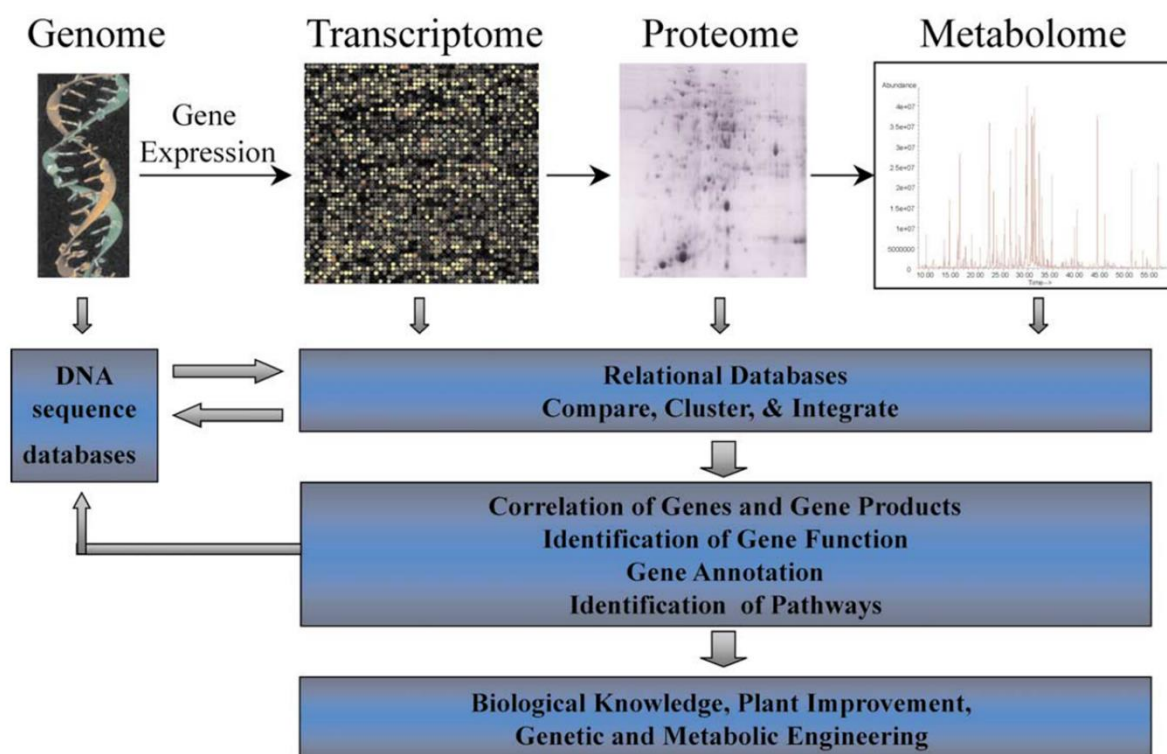


Figure 8 : Exemple de génomique fonctionnelle. Le métabolome est le témoin des conséquences des modifications du génome, et son analyse représente un outil très utile pour déduire les fonctions des gènes [57].

2. Le profilage métabolique

L'apparition du terme « profilage métabolique » date du début des années 1970 en tant qu'approche analytique servant à décrire des voies métaboliques, des mécanismes ou des captures d'un état métabolique et physiologique à un instant donné. Ce terme est donc antérieur à l'utilisation du mot « métabolomique », mais décrit des objectifs d'étude similaires.

Le profilage métabolique a longtemps été utilisé dans des études de fluides biologiques humains, principalement l'urine mais aussi du sérum sanguin, en biomédecine pour la recherche et le diagnostic de maladies humaines [54,58–60]. Il est donc appliqué depuis la fin du 20^{ème} siècle à des études cliniques et à la caractérisation de marqueurs de pathologies.

Le profilage métabolique a ensuite fait lentement son chemin jusqu'aux années 1990, avec l'élargissement des sujets d'étude vers des matrices autres qu'humaines et l'étude de produits naturels. Par exemple, la première application du profilage métabolique aux plantes est réalisée en 1991, dans le but de décrire des réponses physiologiques à des stress environnementaux [61]. Le profilage

métabolique est aussi appliqué en 1998 sur des lipides d’huitres pour étudier la pollution marine [62]. C’est ensuite au début des années 2000 que les études de métabolomique sur les plantes fleurissent. Des méthodes de profilage ont été appliquées à la comparaison de phénotypes biochimiques de plantes dont le génome varie afin d’approfondir la compréhension de certaines voies complexes de régulation [63,64].

3. Différentes approches

L’approche choisie pour mener une étude métabolomique dépend évidemment de la question biologique posée concernant le projet de l’étude ; le workflow général en métabolomique est résumé en Figure 9. Les différentes étapes seront détaillées dans les parties suivantes. En réponse à une question biologique donnée, l’échantillon est prélevé, préparé et analysé. S’ensuit alors un traitement de données qui aboutit à une interprétation des résultats obtenus pour répondre à la problématique de départ.

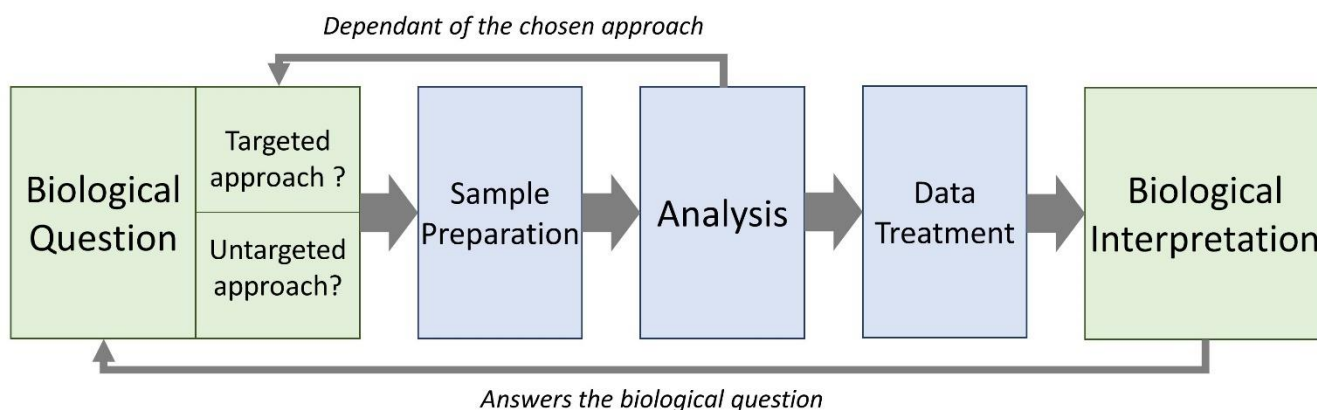


Figure 9 : schéma du workflow général des études de métabolomique.

a. Études ciblées et non ciblées

Les études de métabolomique peuvent être ciblées, c’est-à-dire dirigées vers la détection et la quantification de molécules prédéfinies au sein d’un échantillon. Les analytes recherchés sont déterminés en amont du processus analytique, et ce dernier est adapté à l’analyse des composés visés. La plupart des études de métabolomique ciblées se focalisent sur quelques métabolites ou des voies spécifiques, par exemple sur les voies centrales du métabolisme du carbone, pour répondre à des questions biologiques précises telles que les réactions intracellulaires d’être vivants suite à des

modifications environnementales. Les études ciblées peuvent également être quantitatives, et analyser les proportions des différentes molécules d'un échantillon biologique par rapport aux autres.

Il existe également des procédés d'analyse non-ciblées, permettant de caractériser tous les composés d'un extrait naturel ou des marqueurs spécifiques mais dont la structure n'est pas attendue ou connue d'avance ; ce sont donc des études plus exploratoires, parfois qualifiées de « balayage » ou « screening ». La méthode analytique appliquée se doit d'être plus exhaustive, et d'analyser une gamme large de composés.

b. Études du métabolisme primaire ou secondaire

Les approches de métabolomique peuvent être appliquées à l'étude du métabolisme primaire d'un organisme ou d'un tissu. Les métabolites primaires sont indispensables au fonctionnement physiologique de l'organisme et commun à tous les êtres vivants. Le métabolisme primaire est ciblé dans des travaux qui visent à répondre à une question biologique précise, observer des réponses métaboliques à des changements environnementaux, ou encore déchiffrer certaines voies de biosynthèse.

Les recherches sur le métabolisme secondaire représentent une vaste diversité des sujets d'application, allant des mêmes objectifs que pour le métabolisme primaire jusqu'à la recherche de nouvelles substances naturelles potentiellement intéressantes. Comme expliqué précédemment, les métabolites secondaires font partie de voies métaboliques impliquées dans la communication, la défense, la structure ou d'autres fonctions qui ne sont ni la reproduction ni la survie des organismes qui les produisent. Ils représentent ainsi une diversité structurale colossale, et sont souvent spécifiques des êtres vivants dont ils sont originaires : chaque espèce produit sa propre diversité de biomolécules.

4. La déréplication

En chimie des substances naturelles, les approches métabolomiques sont très intéressantes pour la découverte de nouvelles molécules bioactives potentiellement intéressantes dans le secteur pharmaceutique. En effet, grâce aux moyens analytiques mis en œuvre dans ce domaine, il est possible de réaliser la déréplication des extraits naturels étudiés. Ce terme date de 1999 et décrit le fait de procéder à l'identification des constituants d'un mélange sans même les isoler du reste du mélange par des techniques de séparation physiques classiques, ou d'extraction [65].

L'avantage de la déréplication est de permettre un gain de temps considérable afin de se concentrer sur les substances présentant un intérêt vis-à-vis des objectifs de travaux. En effet, isoler chaque constituant d'un mélange puis analyser ceux-ci, pour qu'au final les molécules soient déjà connues, décrites et à l'activité biologique déjà renseignée, fait l'effet d'une grande perte de temps et d'énergie pour qui mène la recherche. Ainsi, les méthodes de profilage métabolique présentent un grand intérêt dans la déréplication d'extraits naturels et la découverte de nouveaux composés à intérêt thérapeutique.

IV. Les méthodes d'extraction de métabolites appliquées aux microorganismes

Les méthodes d'extraction de molécules intracellulaires comme extracellulaires à partir de microorganismes sont légèrement différentes des méthodes appliquées à l'extraction de composés de plantes pluricellulaires. L'extraction des métabolites des microorganismes est une étape extrêmement importante qui nécessite des précautions. La nature et la quantité de molécules extraites dépendent de multiples facteurs. Appliquées à des microalgues ou des champignons microscopiques, ces méthodes sont très variables et sont destinées aussi bien à des cultures en milieu liquide qu'à des cultures sur gélose solide. Le choix de la méthode dépend énormément des métabolites visés.

Les métabolites primaires, acides aminés, oses, lipides et acides nucléiques sont retrouvés dans la majorité des organismes vivants et à toutes les périodes de leur existence. Les études visant le métabolisme primaire se focalisent sur les abondances relatives des différentes molécules pour caractériser, par exemple, des réponses métaboliques à des stress environnementaux. Elles concernent des métabolites souvent fragiles, susceptibles d'être dénaturés par les élévations de température ; les composés sont conservés grâce à un froid important. L'étude du métabolisme secondaire est plus permissive vis-à-vis de la chaîne du froid, car les métabolites secondaires sont principalement des composés plus résistants à la température.

Dans la majeure partie des cas, les procédés d'extraction à partir d'une culture font partie de la préparation des échantillons, et vont se dérouler en trois grandes étapes résumées en Figure 10 [66]. Cela commence par l'interruption du métabolisme cellulaire, se poursuit par l'extraction des métabolites dont le principe va dépendre de l'échantillon et des choix des utilisateurs, puis par la concentration de l'extrait pour donner un échantillon final.

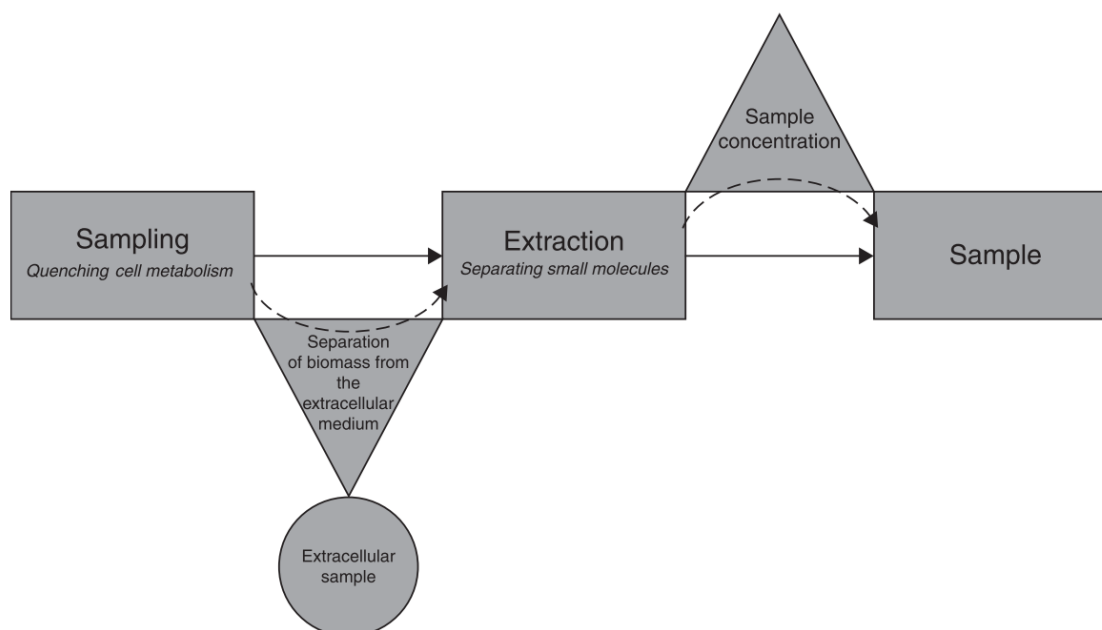


Figure 10 : étapes générales incluant l'extraction de petites molécules de matrices naturelles. Les flèches pleines représentent les événements obligatoires, tandis que les flèches pointillées représentent des alternatives pour améliorer l'analyse [66].

1. Arrêt de la prolifération cellulaire

a. Définition et objectif

La constante dans le procédé d'extraction de métabolites d'une culture de microorganismes est l'utilisation de solvant organique pour arrêter la prolifération des cellules cultivées. L'arrêt de la prolifération est également appelé « *quenching* ». Le méthanol est utilisé comme solvant dans de nombreux travaux [67–69]. Eteindre le métabolisme d'une culture cellulaire a pour but d'arrêter les différentes activités enzymatiques et de figer les concentrations intracellulaires des métabolites. Cela permet donc d'avoir une capture plus précise du contenu en métabolites des organismes à étudier, facilitant donc les interprétations des taux des molécules ciblées dans le cas de comparaison de larges cohortes d'échantillons.

Au-delà de l'aspect *quenching* du métabolisme, l'utilisation du méthanol, tout comme celle d'autres solvants organiques, a également un intérêt direct dans l'extraction des métabolites. L'ajout d'un solvant organique permet de déstabiliser les membranes plasmiques des organismes, ainsi que les parois glycoprotéiques des champignons unicellulaires. Cela résulte dans l'éclatement partiel ou

complet des cellules et donc à la libération dans le solvant des métabolites intracellulaires, amenant ainsi à un extrait de culture.

b. Choix du solvant

Outre le méthanol, l'extraction par solvant organique peut aussi se faire par l'ajout d'éthanol bouillant [68,70], par un mélange de chloroforme et de méthanol amenant à l'obtention d'une fraction polaire et apolaire [71,72], par l'ajout d'acétonitrile [73], par l'ajout d'un mélange de dichlorométhane et d'acétate d'éthyle [74], ou bien par un mélange des différents solvants énumérés précédemment [75,76].

Le choix du solvant revêt une importance particulière car il conditionne les métabolites extraits. L'utilisation d'un mélange de solvant contenant de l'eau et du méthanol aura pour effet d'extraire les métabolites les plus polaires de la culture cellulaire concernée. A l'inverse, l'utilisation de dichlorométhane ou de chloroforme dans un mélange de solvant permettra l'extraction des métabolites de polarité moyenne à faible. Enfin, il est à noter que plusieurs étapes d'extractions par solvants sont possibles, par exemple avec une première étape avec un mélange moyennement polaire voire apolaire, puis une seconde macération des milieux de culture avec des solvants plus polaires. Une plus large gamme de polarité des métabolites est ainsi couverte.

2. Disruption mécanique des cellules

L'ajout de solvant organique dans la culture de microorganismes a pour effet, comme énoncé précédemment, de déstabiliser les membranes et parois. L'éclatement des cellules et la libération des métabolites peuvent être finalisés avec des méthodes de disruption mécaniques sur la culture. Le traitement aux ultrasons permet de casser les membranes de façon uniforme et homogène [74,77,78]. Cela a cependant le défaut de générer de la chaleur lors d'une utilisation prolongée, ce qui peut avoir un effet délétère sur certaines molécules. La culture peut également être mélangée vigoureusement pendant un temps long (de plusieurs dizaines de minutes à plusieurs heures) avec une assistance mécanique, tel qu'un agitateur automatique [79,80]. Les cellules peuvent aussi être brisées à l'aide d'un broyage mécanique par mélange vigoureux accompagné de billes en verre ou en céramique [67,71,81]. Enfin, le gel-dégel rapide des cultures, souvent à l'aide de glace ou d'azote liquide, entraîne une modification très rapide du volume d'eau du cytoplasme et amène ainsi à l'éclatement des membranes [69,72].

3. Élimination des débris cellulaires

Une fois les membranes et parois déstabilisées et les cellules éclatées, les métabolites sont libérés dans le solvant d'extraction. Cependant, les débris de membrane, les macromolécules telles que les protéines et parfois des morceaux d'un milieu de culture solide peuvent être présents dans l'échantillon. La filtration est une étape nécessaire afin de se débarrasser du matériel biologique.

La plupart des protocoles d'extraction présentent une ou plusieurs étapes de filtration afin d'éliminer un maximum de débris cellulaires ou de protéines, souvent avec des filtres en polytétrafluoroéthylène (PTFE) car ils résistent aux solvants les plus agressifs, comme par exemple le dichlorométhane [74,81,82]. Les débris cellulaires peuvent également être éliminés via centrifugation puis récupération de l'extrait par prélèvement du surnageant [73].

4. Influence de la température sur l'extraction

Selon l'objectif de l'étude et la catégorie de métabolites visés, la température de l'environnement direct de l'extraction peut être prise en considération. Comme indiqué précédemment, lorsque le métabolisme primaire est ciblé (pour rappel, acides aminés, lipides, sucres, acides nucléiques, c'est-à-dire composants de base du vivant), l'objectif est d'observer une capture la plus précise possible des différentes concentrations de métabolites au sein d'un organisme. Ainsi, une très basse température sera nécessaire et ce tout au long du protocole. Les bacs de glace et contenant remplis d'azote liquide sont mis à contribution pour conserver le froid le plus bas durant l'extraction. Le quenching au méthanol peut être réalisé à -40 °C [81,83]. Les méthodes de disruption mécanique générant de la chaleur, telle que l'ultrasonication, sont évitées au profit par exemple du broyage à l'aide de billes en verre, qui peut être réalisé au froid si l'on plonge la cuve contenant l'échantillon dans un bac de glace [67,71]. Enfin, les extraits sont soit analysés directement pour éviter tout temps qui pourrait provoquer une détérioration des métabolites, soit conservés à basse température à -80 °C .

La température d'extraction visée dans les études ciblant le métabolisme secondaire est bien plus permissive, les molécules synthétisées dans ces voies sont souvent plutôt résistantes et ne seront pas rapidement dégradées par les voies métaboliques des cellules. La plupart des protocoles d'extraction de tels travaux permettent de maintenir l'environnement direct d'extraction à température ambiante [74,76,79]. Dans le cas spécifique de l'analyse de composés volatils par chromatographie en phase gazeuse couplée à la spectrométrie de masse (GC-MS), l'extraction de ces métabolites peut être réalisée

par micro-extraction sur phase solide à partir de l'espace de tête (HS-SPME). Ici, la température d'extraction atteint les 250 à 270 °C afin de faire passer les analytes à l'état de vapeur et de les piéger dans la phase solide.

En conclusion, le choix de la température du protocole d'extraction et son application doivent se faire en connaissance des métabolites visés et en considération avec la méthode analytique envisagée pour l'analyse des composés des extraits.

5. Concentration et reprise des extraits

L'échantillon peut être concentré par évaporation sous vide. Cette étape permet d'éliminer le solvant organique utilisé pour l'extraction et donc de maîtriser la concentration finale des échantillons. L'extrait, qu'il soit sec après évaporation ou de nature liquide, est souvent conservé pour analyse ultérieure. Dans certains cas, quand l'analyse des extraits est proche, les extraits sont conservés au réfrigérateur ; autrement, la congélation est utilisée pour garder les extraits sur le long terme. La température de conservation peut aller jusqu'à plusieurs dizaines de degrés en dessous de zéro, jusqu'à - 80 °C [67,71].

Après les étapes d'extraction des métabolites vient la préparation des échantillons pour analyse à partir des extraits obtenus. Lorsque les extraits sont évaporés à sec, il est nécessaire de maîtriser la nature du solvant de reprise et de vérifier la solubilité de l'extrait sec dans le solvant sélectionné [75,79,82]. Le solvant de reprise d'un extrait sec doit être sélectionné selon les mêmes critères qu'un solvant utilisé dans la première étape d'extraction de métabolites : il doit être choisi en fonction de la gamme de polarité des métabolites visés. Le solvant doit également être adapté en fonction du système analytique utilisé après l'extraction afin de ne pas endommager ce dernier et de rendre possible l'analyse.

V. Les techniques analytiques utilisées en métabolomique

Le profilage métabolique et les approches métabolomiques permettent la définition des fonctions biochimiques de certaines voies métaboliques des organismes, et permettent de déchiffrer certaines relations moléculaires existantes au sein de ces voies. Cependant, pour pouvoir analyser et comparer un large panel d'échantillons, la méthode analytique appliquée se doit d'être rapide, fiable, répétable et sensible.

1. Aperçu des méthodes employées

Les techniques analytiques utilisées en métabolomique sont diverses et permettent d'acquérir une multitude d'informations sur les échantillons analysés. La métabolomique fait appel à des techniques analytiques spectroscopiques, des méthodes de couplage de chromatographie avec la spectrométrie de masse, la plus commune étant la chromatographie en phase gazeuse couplée à la spectrométrie de masse, utilisée depuis les débuts du profilage métabolique (GC-MS) et suivie par la chromatographie liquide couplée à la spectrométrie de masse (LC-MS), ainsi que l'analyse par résonance magnétique nucléaire (RMN) [55]. C'est un domaine complexe, car contrairement à la génomique, qui se concentre sur l'analyse d'assemblages de 4 acides nucléiques, et à la protéomique et ses 22 acides aminés protéinogènes, la diversité des molécules observables en métabolomique est colossale. Les limitations des approches métabolomiques au début des années 2000 étaient liées au manque d'exhaustivité des analyses.

Bien que l'analyse d'absolument toutes les petites molécules d'un organisme ne soit pas réalisable, les améliorations des instrumentations au cours des deux dernières décennies permettent de compenser ces limitations et d'améliorer la gamme de composés détectés. Les données spectrales générées par les approches métabolomiques sont lourdes, c'est pour cette raison que les analyses statistiques multivariées (telles que l'analyse par composante principale ; PCA) sont très utiles à la validation et à l'interprétation des résultats, et qu'il existe une grande nécessité d'outils bio-informatiques pour le stockage, le traitement informatisé et l'analyse de telles données [57].

Le profilage métabolique est appliqué aux levures dès le début des années 2000, notamment sur la levure modèle *S. cerevisiae* et son métabolome. On parle également d'exométabolome lorsque l'analyse se focalise sur le produit de la fermentation de ladite levure, par exemple dans le cas des nombreux travaux concernant les analyses de vins [84]. Les rubriques suivantes détailleront les différents principes inhérents à chaque méthode d'analyse utilisée en métabolomique. Après l'introduction des méthodes séparatives appliquées à l'étude du métabolome des organismes et plus particulièrement des levures, les exemples d'application à *S. cerevisiae* seront fournis pour chacune des méthodes analytiques traitées.

2. La séparation chromatographique des composés d'un mélange

Les techniques chromatographiques sont des méthodes séparatives utilisées en chimie pour séparer les composants d'un mélange et ainsi faciliter leur identification et leur quantification.

Les composants du mélange injecté sont entraînés par une phase mobile (un solvant, un gaz, ...) tout en passant au contact d'une phase stationnaire solide. Les interactions entre les composés et cette phase stationnaire vont ralentir leur avancée, les espèces migrant le plus vite seront séparées des espèces les plus lentes et toutes seront détectées à un point précis dans l'espace expérimental. On parle ainsi de temps de rétention des composés, et celui-ci sera d'autant plus élevé que le composé sera retenu par la phase stationnaire.

L'appareil servant à cette séparation est un chromatographe, et le résultat graphique pouvant être tiré d'une telle expérience est un chromatogramme. La détection des composés en fin de séparation peut se faire par des techniques spectroscopiques, ou grâce à des techniques d'analyse chimique comme l'ionisation de flamme ou la spectrométrie de masse.

Bien qu'il existe de multiples méthodes de chromatographie, et ce en raison de nombreuses phases mobiles, d'interactions possibles avec la phase stationnaire et de combinaisons de techniques, nous ne traiterons que celles ayant eu un impact durant ces travaux de thèse, c'est-à-dire la chromatographie sur couche mince, la chromatographie en phase gazeuse et la chromatographie en phase liquide. Elles sont présentées en premier car elles peuvent être utilisées dans de multiples configurations de couplages utilisées en analyse métabolomique.

a. La chromatographie sur couche mince

Les premières utilisations de la chromatographie sur couche mince (ou TLC pour *thin layer chromatography*) datent de la fin du 19^{ème} siècle pour la séparation d'acides organiques ou de deux enzymes [85]. Cette méthode est donc utilisée pour séparer ou purifier les composants d'un mélange par le biais d'une phase stationnaire en couche mince en matériau adsorbant, le plus souvent de la silice ou de la cellulose. Les composés à séparer sont entraînés par une phase mobile composée d'un solvant ou d'un mélange de solvant. C'est la polarité de cet éluant qui va influencer la migration des composés le long de la plaque de phase stationnaire : par exemple, plus un composé est polaire et plus son affinité avec une phase stationnaire polaire sera grande, donc moins il migrera vite.

A la fin de l'expérience, l'élution des composés peut être constatée grâce à l'observation de tâches sur la plaque. Ces tâches seront visibles grâce à une lampe à ultra-violet, ou grâce à une révélation par des réactifs chimiques. La position des tâches permet de calculer le rapport frontal de chaque composé (c'est-à-dire le rapport entre la distance ligne de dépôt-composé migré sur celle de la ligne de dépôt-front de solvant) et de comparer la migration des différentes substances séparées. D'une autre manière, la séparation des composés peut être réalisée à des fins préparatoires pour isoler physiquement les différentes molécules avant l'application d'autres traitements. Le principe de la technique de TLC est illustré en Figure 11 [86].

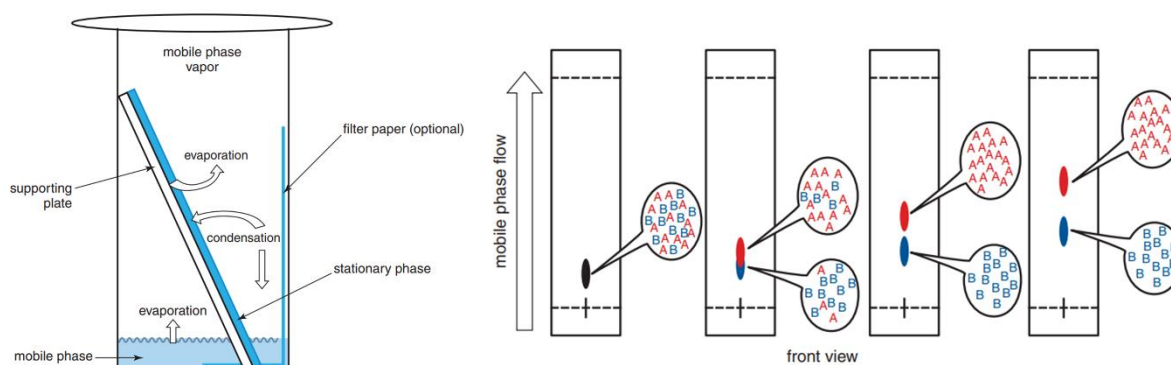


Figure 11 : schéma du montage de chromatographie sur couche mince et de l'élution des composés d'un mélange sur la phase stationnaire [86].

b. La chromatographie en phase gazeuse

La chromatographie en phase gazeuse (ou GC pour *gas chromatography*) a été inventée dans les années 1950 [87]. Elle permet de séparer les espèces chimiques présentes dans un mélange, celui-ci souvent plus complexe que ceux habituellement séparés par TLC. La GC cible des molécules volatiles ou apolaires, qui sont vaporisées avant d'être séparées dans une colonne chauffée. Cette technique est donc destructrice de l'échantillon analysé, c'est pourquoi elle est utilisée dans un but analytique et non préparatoire.

Injection de l'échantillon

L'échantillon peut être sous forme liquide, donc un mélange complexe dans un solvant apolaire ou moyennement polaire. L'échantillon est alors soumis à une forte température, supérieure à la température d'ébullition du solvant et des composés à analyser, afin que ceux-ci soient vaporisés. L'échantillon peut aussi être sous forme gazeuse, c'est-à-dire l'espace de tête d'un flacon contenant

l'échantillon sous forme liquide qui a été chauffé ; les composés à analyser se sont donc évaporés dans cet espace de tête.

Bien que la GC-MS soit utilisée classiquement pour l'analyse de molécules apolaires et volatiles, il est possible d'analyser des molécules polaires. Pour ce faire, les composés doivent être dérivatisés durant l'étape de préparation d'échantillon, c'est-à-dire qu'un groupe fonctionnel spécifique est ajouté aux molécules afin d'en réduire la polarité ou de les rendre volatiles [71,88,89]. Le plus souvent, la dérivatisation avec des groupes silyles est faite pour améliorer la séparation chromatographique, la détection des métabolites polaires ainsi que la reproductibilité de l'analyse. L'injecteur permet ensuite d'introduire l'échantillon vaporisé dans le flux de gaz vecteur.

Séparation des composés

Après injection, les analytes sont emmenés par un gaz vecteur inerte vis-à-vis des composés du mélange et de la phase stationnaire, comme l'hélium. La phase stationnaire est greffée sur le support interne de la colonne, et peut par exemple être en polyéthylène glycol ou en polydiméthylsiloxane, toutes deux polaires. La sélectivité et polarité de la phase stationnaire induiront des interactions physico-chimiques différentes en fonction des analytes et permettront de les séparer. La colonne est chauffée, le plus souvent définie selon un gradient de température croissant, maintenant les analytes en phase gazeuse. Ce gradient de température permet de progressivement vaporiser les composés et les séparer. Il est également possible de coupler deux colonnes de GC l'une à la suite de l'autre, les deux ayant des phases stationnaires différentes, afin de créer une seconde dimension de séparation dans le processus d'analyse.

Afin de caractériser les molécules du mélange, le système de chromatographie en phase gazeuse est équipé d'un détecteur. Il en existe de différentes natures et aux différents principes. Par exemple, un détecteur à ionisation de flamme (FID) emploiera une tension importante et une flamme afin d'ioniser les molécules sortant de la colonne et une électrode détectera le courant électrique ainsi créé. Il est également possible de coupler le système chromatographique avec un spectromètre de masse pour un profilage précis des molécules contenues dans le mélange complexe (détails sur la spectrométrie de masse dans les parties suivantes). Pour terminer, une illustration des composants de base d'un chromatographe en phase gazeuse est fournie en Figure 12 [87].

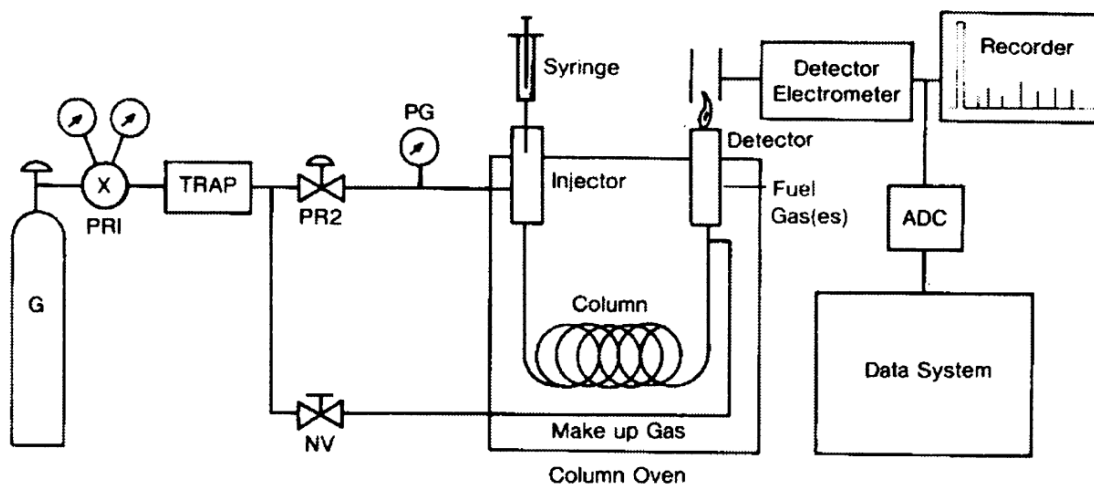


Figure 12 : schéma des composants de base d'un appareillage pour chromatographie en phase gazeuse. A gauche du four, les systèmes permettant l'arrivée des différents gaz. A droite du four, les systèmes de détection et d'acquisition de données [87].

c. La chromatographie en phase liquide

La chromatographie en phase liquide (ou LC pour *liquid chromatography*) fit ses débuts en 1931 avec un procédé de séparation de pigments, mais les termes de « chromatographie de partition liquide-liquide » n'apparurent que durant les années 1940, et la technique s'affina durant les décennies suivantes [85]. La chromatographie en phase liquide, tout comme la chromatographie sur couche mince, peut être séparative ou analytique. La chromatographie séparative isole les composés les uns des autres et permet le fractionnement d'un échantillon. Analytiquement, elle peut être appliquée à des études quantitatives comme qualitatives. Techniquement parlant, elle comprend les méthodes utilisant une phase mobile liquide, dont la TLC détaillée précédemment, la chromatographie sur papier, ainsi que la chromatographie sur colonne qui sera détaillée davantage dans cette partie. Dans tous les cas et comme dans les méthodes déjà décrites, la séparation des composés entraînés par la phase mobile s'opère en fonction des interactions chimiques entre lesdites substances et les molécules composant à la fois la phase mobile et la phase stationnaire.

De l'injection à la détection

Le procédé de chromatographie en phase liquide se rapproche grandement de celui de la chromatographie en phase gazeuse. L'échantillon est préparé au préalable dans un solvant ou un mélange de solvant similaire à ceux utilisés en phase mobile ; les deux entités doivent être miscibles l'une dans l'autre, afin de ne pas créer de problème au moment de l'injection. Une petite quantité du

mélange est injectée grâce à une boucle d'injection dans le flux de solvant, et les composés de ce mélange traversant le système vont atteindre la colonne où la séparation s'opérera. De multiples principes de détection des molécules séparées existent. Les détecteurs ont été optimisés depuis l'époque de la création de la LC, le détecteur par absorption dans l'ultra-violet a été rejoint par des réfractomètres, des fluoromètres ou des détecteurs électrochimiques. Enfin, la spectrométrie de masse est également utilisée en couplage avec la chromatographie en phase liquide pour la caractérisation des composés.

Types de chromatographie en phase liquide

Les différents types de chromatographie liquide sont en lien avec la grande variété d'interactions qui entrent en jeu lors d'une expérience.

La chromatographie d'adsorption se base sur l'adsorption des composés sur la phase stationnaire en fonction de leur polarité. La chromatographie d'affinité se base sur des interactions ayant également lieu dans le domaine de la biologie, comme la liaison d'un anticorps avec son ligand, tandis que la chromatographie par échange d'ions sépare les molécules en fonction de leurs groupes chargés. La chromatographie d'exclusion stérique utilise une phase stationnaire poreuse, et les composés sont séparés en fonction de leur taille (les composés plus grands n'entreront pas dans les pores et élueront avant les composés de plus faible taille). La chromatographie de paires d'ions permet de séparer des ions spécifiques à l'aide de contre-ions dotés de chaînes alkyles. La chromatographie liquide à interaction hydrophile (HILIC) permet la rétention des petites molécules polaires et leur séparation selon un gradient de phase aqueuse. Enfin, la chromatographie de partage représente le principe le plus utilisé dans les études de métabolomique dans lequel les composants du mélange injecté sont séparés en fonction de leur affinité avec la phase mobile et avec la phase stationnaire, c'est-à-dire en fonction de la polarité de ces différents éléments.

Phase normale et phase inverse

La chromatographie de partage peut être appliquée en phase dite « normale » : un solvant moyennement polaire ou apolaire est utilisé pour faire éluer les composés de l'échantillon en fonction de leurs interactions avec la phase stationnaire polaire, comme un gel de silice. Les composés apolaires élueront plus vite et auront un faible temps de rétention, tandis que les composés polaires seront retenus sur la colonne et élueront à un temps de rétention plus élevé.

La chromatographie liquide en phase « inverse » est réalisée avec un solvant polaire en guise de phase mobile (tel que l'eau ou le méthanol), avec un gradient de solvants et un flux passant au travers d'une

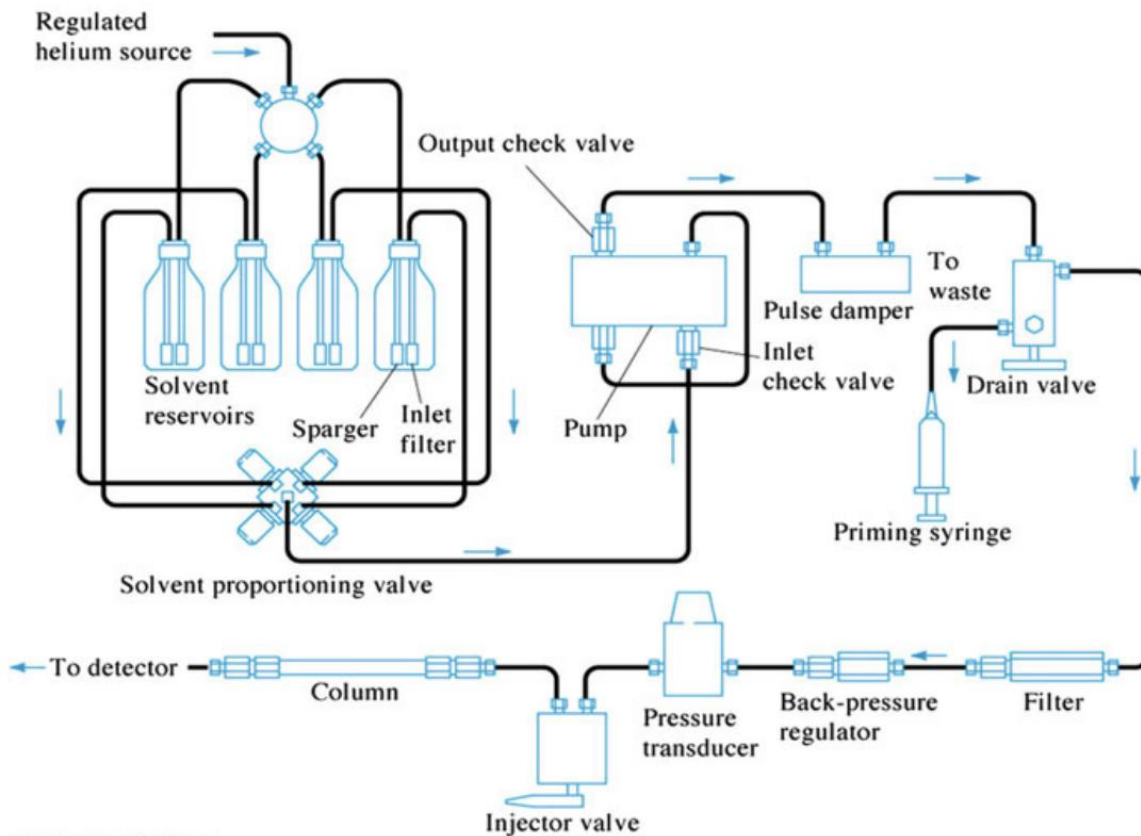
colonne chromatographique contenant le plus souvent de la silice greffée de C18 (des molécules comprenant une chaîne de 18 atomes de carbone) en guise de phase stationnaire apolaire. Les composés polaires élueront vite tandis que les molécules apolaires auront un temps de rétention plus élevé.

La chromatographie HILIC est basée en général sur une phase stationnaire de type polaire, selon plusieurs types de support différents (aminopropyl, diol, zwitterionique...) et la phase mobile est majoritairement organique contenant de l'eau comme éluant fort. L'eau forme une couche polaire dans laquelle les analytes polaires vont interagir plus efficacement avec la surface de la phase stationnaire. Ainsi, la chromatographie de type HILIC permet une bonne rétention des composés polaires.

La chromatographie liquide à haute performance

Grâce à l'utilisation de particules de petite taille et à l'application de pression sur la phase mobile, l'appellation « chromatographie liquide à haute vitesse » apparut à la fin des années 60. Rapidement s'ensuivit l'évolution de la technique en « chromatographie liquide à haute performance » (HPLC), et les surfaces adsorbantes des colonnes chromatographiques continuèrent à s'améliorer depuis cette époque [85]. Datant des années 1970, l'HPLC part du principe de la chromatographie sur colonne. Ici, la phase mobile est pompée au travers de la chaîne chromatographique, et donc de la colonne, sous une haute pression (en moyenne approximativement entre 170 et 350 bars) [90]. Par rapport à la LC sur colonne classique, le diamètre de la colonne est plus faible tout comme le diamètre des particules à l'intérieur de celle-ci (entre 3 à 5 μm de diamètre habituellement). La pression interne plus importante est due à ce faible diamètre et au débit important fourni par la pompe.

Afin d'illustrer à la fois l'instrumentation de chromatographie liquide ainsi que les différentes parties d'une chaîne chromatographique d'HPLC, un schéma d'un système HPLC est fourni en Figure 13 [90]. Les avantages de l'utilisation de l'HPLC sont liés à ses attributs décrits précédemment, et sont une plus grande vitesse de séparation (et donc d'expérience), ainsi qu'une meilleure sensibilité et résolution des pics chromatographiques par rapport à la LC.



© 2007 Thomson Higher Education

Figure 13 : schéma de l'instrumentation HPLC. On y observe les composants classiques de chromatographie liquide (injecteur, colonne, détecteur) ainsi que ceux spécialisés pour l'HPLC (réglages de pression) [90].

La chromatographie liquide à ultra-haute performance

Les méthodes de chromatographie en phase liquide furent encore améliorées dès l'arrivée des techniques de chromatographie à ultra-haute performance (UPLC, ou UHPLC). Initialement lancée en 2004 par Waters Corporation UPLC, les deux acronymes sont aujourd'hui utilisés depuis que les autres constructeurs de chaînes de chromatographie ont également produits des instruments au principe similaire sous l'appellation UHPLC. Dans tous les cas, c'est le principe de l'HPLC qui est ici raffiné et la principale amélioration développée concerne le diamètre des particules à l'intérieur des colonnes chromatographiques.

Les colonnes d'UHPLC contiennent des particules d'un diamètre inférieur à $2\ \mu\text{m}$; l'appareillage nécessite alors une capacité de pression supérieure pour délivrer le débit adéquat, pouvant aller de 500 à 1000 bars. La vitesse et la résolution des séparations chromatographiques par UHPLC rendent cette méthode idéale pour un couplage avec une détection par spectrométrie de masse [91].

Types de colonnes chromatographiques

Il existe différents types de colonnes selon l'usage que l'on fait de la chromatographie ; par exemple, un usage analytique avec détection par spectrométrie de masse nécessite une colonne capillaire (très faible diamètre interne).

Les colonnes utilisées en UHPLC pour le mode de séparation en phase inverse sont variées. Par exemple, les colonnes CSH (*Charged Surface Hybrid particle technology*) permettent une amélioration de la sélectivité et de l'acuité des pics via l'utilisation de particules chargées greffées à bas niveau sur la surface interne de la colonne. Les colonnes BEH (*Ethylene Bridged Hybrid particle technology*) sont greffées de chaîne alkyles droites et permettent une meilleure rétention, sélectivité ainsi qu'une forme de pics améliorée. Les colonnes HSS (*High Strength Silica particle technology*) permettent une meilleure rétention et un bon équilibre de séparation entre molécules polaires et hydrophobes grâce à une stabilité du gel de silice améliorée par rapport à la fabrication classique des colonnes d'HPLC. Enfin, les colonnes PST (*Peptide Separation Technology*) repose sur le principe des colonnes BEH mais au diamètre adapté à la séparation de peptides.

3. Les méthodes spectroscopiques

Les méthodes spectroscopiques, notamment couplées à la séparation chromatographique, sont utilisées pour dresser des profils biochimiques d'extraits biologiques ou de cultures cellulaires. Les méthodes spectroscopiques, comme expliqué précédemment, se basent sur l'analyse de molécules qui interagissent avec une radiation électromagnétique d'une longueur d'onde donnée, souvent en l'absorbant partiellement. Les études de profilage métabolique appliquant des méthodes spectroscopiques à l'espèce modèle ont fait appel à la détection par indice de réfraction (RID), la détection dans le spectre ultraviolet-visible (UV-Vis), la spectroscopie Raman et la spectroscopie infrarouge à transformée de Fourier (FTIR).

Ces méthodes sont moins informatives que les approches par résonance magnétique nucléaire et par spectrométrie de masse mentionnées précédemment et détaillées dans les prochaines parties, et ce à cause d'un manque de résolution des molécules entre elles et du manque d'informations structurales fournies par de telles analyses. Elles permettent cependant une analyse globale rapide et robuste d'un mélange biologique ou d'une cellule entière.

Bien que le profilage métabolique de *S. cerevisiae* soit principalement représenté par les méthodes de spectrométrie de masse ou de résonance magnétique nucléaire, certains travaux traitent de méthodes

spectroscopiques. Les objectifs sont variés et spécifiquement dirigés sur la caractérisation des empreintes métaboliques de certaines souches, pour répondre à des questions biologiques précises, comparer différentes souches ou espèces entre elles, ou encore profiler une certaine classe de molécules.

a. Détection par indice de réfraction

Principe

La détection par indice de réfraction, souvent mentionnée par le nom du détecteur à indice de réfraction (RID) ou du réfractomètre différentiel (DRI), se base sur l'indice de réfraction relatif de l'analyte par rapport au solvant le contenant. Dans un détecteur à indice de réfraction, une radiation lumineuse est émise par l'appareil. Or, la trajectoire de l'onde lumineuse peut dévier au moment de quitter un matériau et d'entrer dans un nouveau. L'indice de réfraction d'un milieu est donc mesuré en fonction de l'amplitude de cette déviation. Il y aura besoin d'un témoin, c'est-à-dire d'un contenant avec le même solvant que l'échantillon à analyser. Celui-ci sert de référence et la différence est mesurée entre le témoin et le contenant de l'échantillon au moment où celui-ci passe au niveau du détecteur. Un schéma résumant le principe de fonctionnement du détecteur est présenté en Figure 14 [92].

Le RID est principalement utilisé comme détecteur après une séparation par chromatographie liquide à haute performance (HPLC-RID). La méthode de détection étant non destructrice, ce détecteur peut être utilisé en parallèle d'autres systèmes de détection.

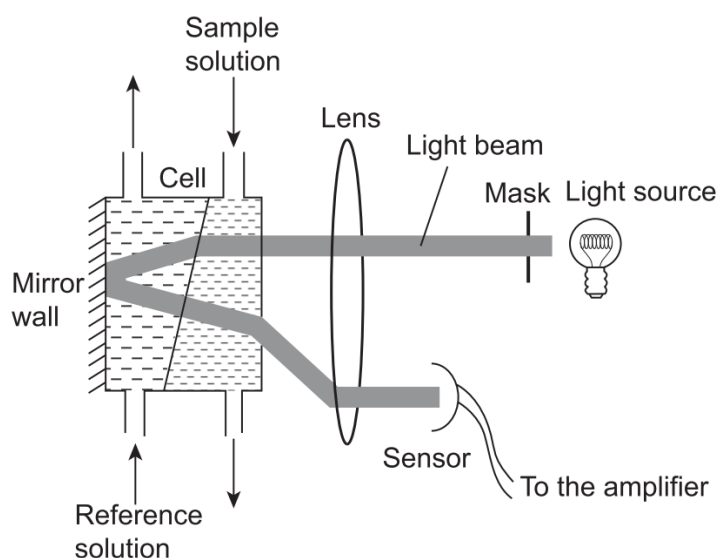


Figure 14 : schéma d'un détecteur à indice de réfraction [92].

Profilage de S. cerevisiae par RID

Le profilage et la quantification d'alcools, de sucres et d'acides organiques ont été réalisés de manière ciblée sur des échantillons provenant de cultures de *S. cerevisiae*. L'objectif de telles études fut de caractériser la réponse métabolique de la levure à un stress environnemental, où la détection RID a été utilisée en complément d'une analyse par LC-MS.

L'analyse de sucres et d'acides organiques a permis de caractériser la réponse métabolique de la levure au préalable stressée avec de l'acide acétique face à l'ajout de sulfate de zinc [93]. L'addition de zinc diminue la production de dérivés réactifs de l'oxygène (ROS) par les levures, et modifie la concentration de certains métabolites du métabolisme central du carbone. L'éthanol, l'acétate et le glucose ont été quantifiés via un RID dans une étude visant à caractériser les effets du milieu de culture sur le profil métabolique de la levure [94]. Enfin, les intermédiaires de la glycolyse ont été profilés par LC-RID en combinaison avec la détection UV-Vis chez une souche génétiquement transformée, afin de caractériser les états thermodynamiques du métabolisme de ces souches de *S. cerevisiae* [95].

Ces travaux démontrent l'intérêt de l'utilisation de la détection RID, notamment lorsque celle-ci est couplée à d'autres méthodes.

b. Détection par absorption des rayons ultraviolet-visibles

Principe

Les détecteurs à rayons ultraviolet-visibles (UV-Vis) partent du même principe que la détection par RID mais en utilisant les propriétés d'absorbance des rayons ultraviolets et visibles des analytes. Souvent couplée à la chromatographie en phase liquide, cette détection est utilisée sur des composés absorbant dans les gammes ultraviolets et visibles du spectre électromagnétique, de 190 à 380 nm de longueur d'onde en général pour l'UV mais pouvant aller plus loin dans le cas des détecteurs UV-Vis. Le rayon UV est émis par une lampe à deutérium, le rayon émis dans le spectre visible provient quant à lui d'une lampe à tungstène, et le schéma de l'analyseur est très proche de celui d'un RID (voir Figure 14). Il s'agit d'une méthode utilisée pour la détection et la quantification de métabolites.

La détection UV est plus performante lorsque le détecteur est équipé de barrettes de diodes (DAD). Cela signifie que les récepteurs sensibles aux rayonnements sont multiples, c'est-à-dire au moins 1000 fois plus nombreux qu'un détecteur UV-Vis classique. Cela permet l'acquisition de données spectrales d'absorbance dans l'UV d'une large gamme de longueurs d'ondes, faisant ainsi du DAD le haut niveau de la détection UV-Vis.

Profilage de S. cerevisiae par UV-Vis

Concernant le profilage métabolique de *S. cerevisiae*, la détection UV-Vis a été appliquée conjointement à la détection RID dans l'étude citée précédemment, visant à caractériser les états thermodynamiques du métabolisme de souches modifiées de *S. cerevisiae* [95]. La détection UV, couplée à l'HPLC, a également été appliquée à l'exométabolome de *S. cerevisiae* par l'analyse de vin fermenté par la levure, afin de caractériser le profil moléculaire et métabolique de la production d'arômes [96]. Les résultats ont montré des différences selon la provenance de la souche, qu'elle provienne de la fermentation de la cachaça, du saké ou du vin, au niveau de la production d'esters de type acétate, d'esters d'éthyle, d'acides et d'alcools, avec la souche standard de laboratoire S288c.

c. Spectroscopie Raman

Principe

La spectroscopie Raman se base sur les interactions entre une lumière monochromatique, typiquement un laser dans le visible, l'infrarouge ou l'ultraviolet, émise par le détecteur et les vibrations moléculaires des molécules d'intérêt. En interagissant avec les vibrations ou autres excitations de l'échantillon, l'énergie du laser est légèrement décalée, donnant ainsi des informations sur les modes vibrationnels des molécules analysées [97].

Profilage de S. cerevisiae par Raman

La spectroscopie Raman a été appliquée à *S. cerevisiae* pour contrôler la production de β -carotène de souches modifiées par ingénierie génétique, ainsi que pour estimer la composition lipidique et les quantités de protéines durant le procédé [98]. Un exemple de spectre Raman des souches de *S. cerevisiae*, provenant de l'étude citée, est fourni en Figure 15. Les résultats permirent d'optimiser le procédé de production du β -carotène par les souches hyper-productrices utilisées.

Malgré le peu d'informations structurales que procure la spectroscopie Raman, elle a l'avantage d'être non-destructrice et de pouvoir être réalisée sur des cellules entières, ce qui limite le temps investi dans l'extraction de métabolites.

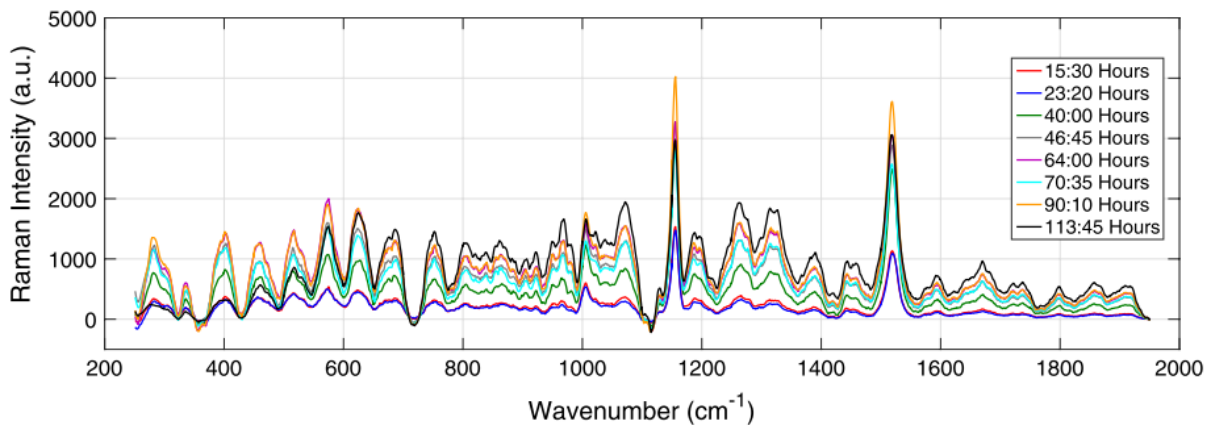


Figure 15 : spectres Raman acquis durant la croissance d'une culture de *S. cerevisiae* de souche SM14 productrice de β -carotène [98].

d. Spectroscopie infrarouge à transformée de Fourier : infrarouge proche et infrarouge moyen

Principe

La spectroscopie infrarouge à transformée de Fourier (FTIR) est basée sur la mesure des interactions de la matière avec des radiations dont la longueur d'onde appartient au domaine des infrarouges. Ces interactions peuvent être l'absorption, l'émission et la réflexion, mais les études de matrices biologiques se concentrent principalement sur l'absorption et la réflexion.

Au sein de l'instrument FTIR, une lumière infrarouge contenant de nombreuses longueurs d'ondes différentes passe au travers d'un interféromètre et de l'échantillon à analyser, puis l'absorption de ce rayonnement est détectée. Des modifications des combinaisons de fréquences émises durant l'analyse permettent d'obtenir de multiples points de données. Le recouplement de ceux-ci crée un jeu d'interférences constructrices et destructrices nommé spécifique de l'échantillon analysé.

La représentation brute des données d'absorption, nommée interférogramme, est convertie en spectre à l'aide de l'outil mathématique de traitement de données nommé transformée de Fourier. Les différents éléments de l'instrumentation FTIR sont présentés en Figure 16 [99].

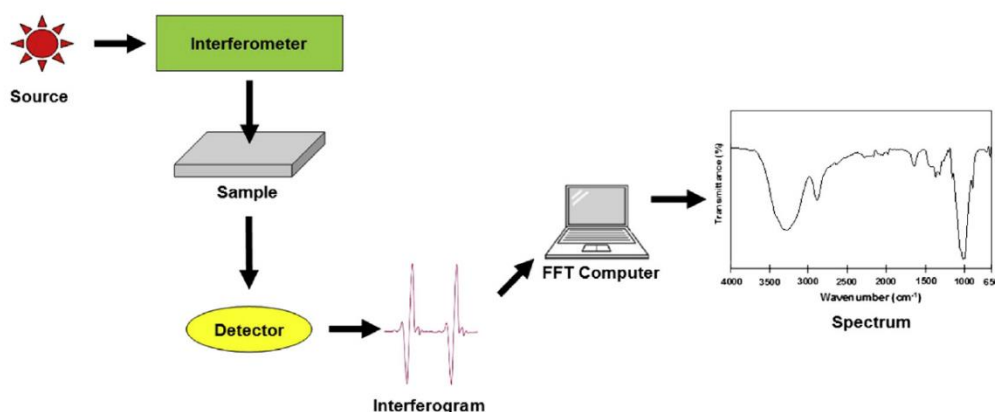


Figure 16 : composants de base d'un instrument de spectroscopie infrarouge à transformée de Fourier [100].

Souvent combinée à des méthodes d'analyses multivariées (telles que la PCA), c'est une méthode rapide pour le balayage global d'un échantillon ou d'une cellule entière. La mesure de l'absorption en infrarouge permet la quantification de certains composés et la comparaison de plusieurs échantillons entre eux.

Malgré sa vitesse d'application et sa simplicité d'utilisation par rapport à d'autres techniques, elle a le désavantage de ne pas fournir d'informations individuelles détaillées sur les composés analysés et de nécessiter un étalonnage par rapport aux composés ciblés.

*Profilage de *S. cerevisiae* par FTIR*

La nature d'un spectre FTIR est complexe car elle additionne le signal de toutes les liaisons chimiques interagissant avec le rayonnement au sein de l'échantillon. Ainsi, c'est une méthode rapide et efficace pour balayer une large collection de mutants afin de les différencier grâce à la caractérisation de leurs empreintes biochimiques spécifiques, notamment au niveau des voies de biosynthèse des lipides [101]. Les méthodes faisant appel à la spectroscopie dans l'infrarouge moyen sont également une alternative rapide et non-invasive pour la description des phénotypes de souches de levures et pour la distinction de leurs différents états métaboliques [102]. Elle est aussi appliquée en combinaison à la spectroscopie dans le domaine visible pour étudier des souches mutantes de *S. cerevisiae* [103]. Ces méthodes représentent ainsi une technique de balayage pour la discrimination de souches lorsque le phénotype des microorganismes n'est pas évident.

La spectroscopie FTIR a également permis d'analyser des cellules entières de *S. cerevisiae* sans besoin d'extraction préalable de métabolites afin de comparer les profils de levures fraîches avec ceux de levures lyophilisées [104]. Après comparaisons appuyées par des analyses chémométriques (PCA et

régression des moindres carrés partiels ; PLS-DA), des profils similaires entre les deux situations ont été reportés, avec une variation moindre pour les souches lyophilisées.

Enfin, la spectroscopie infrarouge a été appliquée à l'analyse de l'exométabolome de *S. cerevisiae* pour profiler les produits majeurs de la fermentation du vin par des souches hyper-productrices de glycérol [105]. La méthode a aussi été employée pour discriminer différents vins rouges fermentés soit par *S. cerevisiae* seule, soit par un cocktail de levures comprenant l'organisme modèle [106]. Cependant, la méthode par infrarouge proche (*near infrared*, NIR) ne semble pas assez sensible pour discriminer certaines co-fermentations de vins dont les propriétés organoleptiques sont très proches.

4. L'application de la résonance magnétique nucléaire à la métabolomique

La spectroscopie par résonance magnétique nucléaire (RMN) est une technique analytique non destructrice développée dans les années 1940. Elle est basée sur la visualisation des interactions magnétiques entre noyaux atomiques de molécules d'un mélange liquide ou d'un échantillon solide [107]. La RMN fit son chemin au cours de la deuxième moitié du 20^{ème} siècle, en s'améliorant dans les années 1970 grâce au développement des aimants supraconducteurs, puis ensuite avec le développement des bases de la transformation de Fourier. Elle devint au 21^{ème} siècle un outil phare de la chimie analytique pour l'élucidation structurale (grâce au concept de « déplacement chimique ») et de l'imagerie médicale (imagerie par résonance magnétique nucléaire, ou IRM).

Principe de la RMN

Les protons des noyaux atomiques étant des particules chargées, ils peuvent être considérés comme des aimants élémentaires. Ainsi, ils possèdent une propriété quantique nommée spin, qui correspond à leur moment magnétique. L'application d'un puissant champ magnétique artificiel uniforme sur l'échantillon à analyser va aligner les spins des atomes d'une certaine manière, créant ainsi des niveaux d'énergie internes différents de l'état normal. Pour atteindre des niveaux d'énergie plus élevés, le champ magnétique doit être plus puissant et à plus haute fréquence ; la sensibilité de la méthode dépend de ce paramètre. Ainsi, les spins du niveau d'énergie le plus bas pourront être excités vers le niveau plus élevé. C'est la magnétisation provoquée par la relaxation de ces spins qui sera mesurée par l'instrument de RMN. La RMN est applicable aux noyaux possédant des spins non nuls, et est utilisée le plus souvent sur le proton (¹H) et sur le carbone 13 (¹³C).

En pratique, les scans sont accumulés pour améliorer la sensibilité tout en conservant un bon rapport signal sur bruit (S/N) ; les signaux de l'échantillon sont augmentés et le bruit de fond est diminué. La

donnée brute d'une expérience RMN est une FID (*free induction decay*, ou signal de précession libre). De nombreux traitements mathématiques vont servir à traiter cette donnée brute pour donner un spectre exploitable, y compris la transformée de Fourier. La Figure 17 montre un exemple de FID, puis son spectre traité après les multiples opérations mathématiques [107].

La RMN bidimensionnelle

La spectroscopie RMN peut également être bidimensionnelle, afin de détecter des relations de proximité spatiale ou de proximité au travers des liaisons covalentes entre les atomes. Différents cycles de phase sont employés pour mesurer les déplacements chimiques de deux noyaux. Parmi les techniques de RMN en deux dimensions, la 2D-COSY montre les couplages scalaires entre deux atomes d'hydrogène ; la 2D-HSQC se concentre sur les atomes d'hydrogène et de carbone adjacents ; la 2D-HMBC sur les hydrogènes et carbones proches de deux ou trois liaisons ; enfin la 2D-TOCSY montre les corrélations entre les hydrogènes d'un même système de spin. Toutes ces techniques permettent d'obtenir des informations précises sur la configuration spatiale des atomes au sein d'une molécule analysée, et donc une meilleure élucidation structurale.

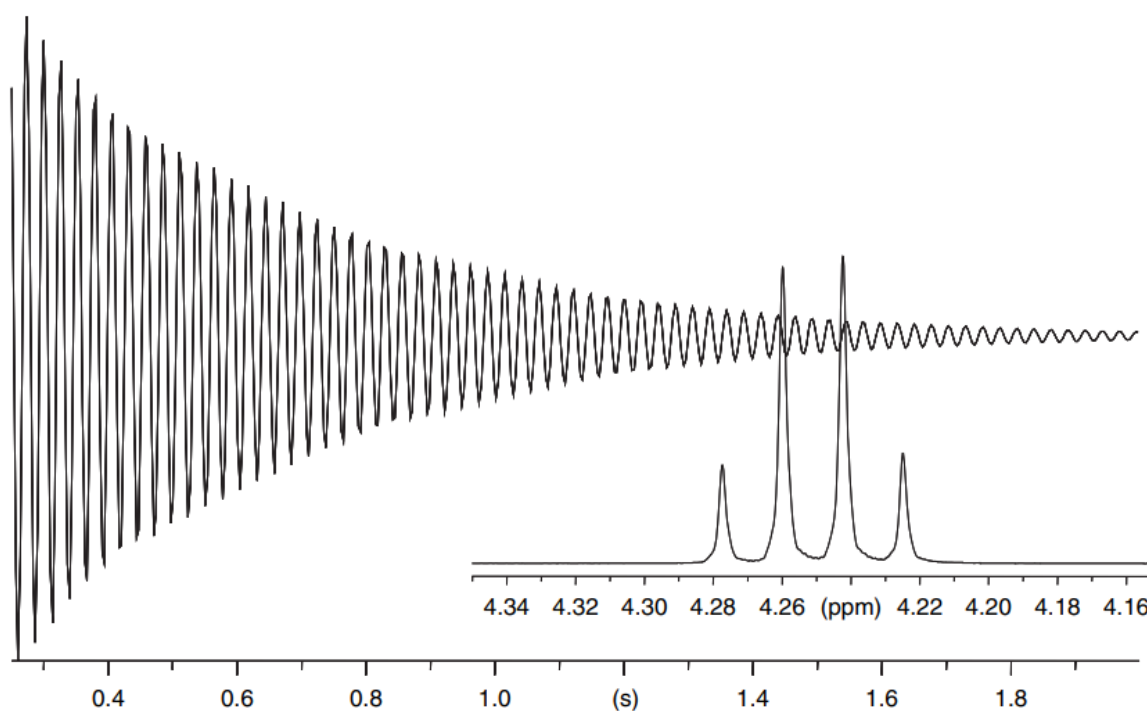


Figure 17 : exemple de résultat d'une analyse par RMN ; signal de précession libre, et spectre traité (notamment par transformée de Fourier) [107].

La RMN en métabolomique

La spectroscopie RMN est utilisée en métabolomique pour obtenir des empreintes biochimiques ou faire de l'élucidation structurale. L'intensité des pics présents dans les spectres acquis par RMN montre une corrélation linéaire avec la concentration des molécules analysées, rendant ainsi la RMN utile pour les études quantitatives. Le traitement des données a une importance particulière pour la métabolomique par RMN, notamment les analyses statistiques multivariées [108]. Cependant, dans le cas d'étude de plantes par exemple, l'identification structurale reste difficile, car la variété de molécules est grande et le nombre d'espèces différentes au sein d'un échantillon est important. Elle reste possible, grâce par exemple à l'analyse en parallèle de standards physiques purs des composés à identifier, ou encore de comparaison avec des bases de données spectrales disponibles en ligne.

La spectroscopie RMN a déjà été employée pour le profilage métabolique de fluides biologiques (urine ou sang), de cellules et de tissus, de plantes, de produits naturels, d'interactions environnementales ou encore d'aliments [109]. Il est à noter que c'est une méthodologie relativement récente et que de nombreux développements sont encore en cours dans ce domaine.

La spectroscopie RMN est une technique robuste et reproductible permettant la quantification et l'élucidation structurale des composés abondants au sein d'un mélange. Ceci peut être réalisé sans fractionner les extraits, c'est-à-dire que la RMN est efficace pour la déréplication d'extraits naturels. Cependant, le traitement de données sur des mélanges complexes est difficile, et un fractionnement de l'échantillon peut être conseillé. La RMN peut en effet rencontrer des difficultés pour détecter et identifier certaines molécules difficilement accessibles, par exemple celles qui se décomposent facilement ou celles qui ne peuvent être isolées d'un mélange [110]. Le temps de préparation est moins important que dans le cas d'analyses par spectrométrie de masse, mais le temps d'analyse peut de son côté être bien plus long (jusqu'à plusieurs heures, notamment pour les expériences de RMN bidimensionnelle). De plus, la spectroscopie RMN est moins sensible que des méthodes basées sur la spectrométrie de masse. L'intensité du signal étant linéairement corrélée avec la concentration de composés, la RMN semble très efficace pour des études quantitatives des métabolites majoritaires d'un échantillon. L'autre principale force de cette méthodologie est l'information structurale que fournit la RMN, ce qui est très intéressant pour des travaux sur les produits naturels inconnus.

*Profilage de *S. cerevisiae* par RMN*

Concernant les études sur la levure *S. cerevisiae*, les petites molécules intracellulaires et extracellulaires ont été profilées par des techniques de spectroscopie RMN et analysées via des méthodes statistiques

multivariées dans le but d'identifier des mutants en prenant en compte leurs voies métaboliques [111]. Le même type de procédé a été appliqué pour caractériser les altérations du métabolisme causées par une mutation du gène de la succinate déshydrogénase (SDH) [112]. 36 métabolites de l'exométabolome ont ainsi pu être identifiés et quantifiés, parmi lesquels des acides aminés et des molécules du métabolisme central du carbone ; les mutations de la SDH ont engendré des altérations significatives de ces voies métaboliques.

La levure a été profilée par RMN du proton ($^1\text{H-NMR}$) dans le but d'optimiser une méthode d'extraction de métabolites [113]. Dans la même étude, une méthode de RMN en phase solide à haute résolution en rotation à l'angle magique (HRMAS, fait intervenir une rotation de l'échantillon) a été optimisée pour l'analyse de levures en tant que méthode complémentaire. L'avantage de l'HRMAS est la rapidité de l'analyse et le fait qu'elle peut être réalisée sur des cellules intactes, limitant le temps de préparation en esquivant l'étape d'extraction. Des profils similaires ont été obtenus entre la $^1\text{H-NMR}$ et l'HRMAS, menant à l'identification et la quantification de 50 composés différents.

Toujours par RMN du proton, les effets des stress de température sur le métabolome de la levure ont aussi été étudiés [72]. Bien que l'analyse des données acquises après des expériences de RMN puissent être chronophage, des outils de traitement de données automatisés existent, tel que le package BATMAN du logiciel de statistique R utilisé dans cette dernière étude. Ce procédé a permis l'identification de 38 métabolites parmi lesquels des acides aminés, des sucres, des nucléotides et des acides organiques. 28 métabolites ont été significativement affectés par les modifications de température. Les mêmes auteurs ont étudié l'acclimatation de la levure aux températures basses comme hautes et ont révélé des modifications de la composition en lipides, des voies énergétiques et des métabolites secondaires du microorganisme [114]. L'utilisation de MCR-ALS (*Multivariate Curve Resolution - Alternating Least Squares*) pour traiter les données de $^1\text{H-NMR}$ a permis l'identification de 42 métabolites. Enfin, une forme d'exométabolome de *S. cerevisiae* a été profilée par l'analyse des petites molécules de cidre fermenté par différentes souches [115]. Une dominance de glycérol et d'acides organiques a été rapportée, aux côtés de molécules aromatiques, de sucres et de molécules aliphatiques.

La spectroscopie RMN bidimensionnelle a également été appliquée au profilage métabolique de *S. cerevisiae*. Une expérience $^1\text{H-}^{13}\text{C-HSQC}$ a permis d'étudier la réponse métabolique de la levure à des stress (ajout de sel ou d'éthanol, stress osmotique ou oxydatif et variations de pH) via des analyses

multivariées, et des variations métaboliques furent observées pour 31 métabolites primaires et acide aminés [116].

En guise d'exemple de résultat d'analyse d'extrait de levure, un spectre original de RMN du proton réalisé sur un extrait de *S. cerevisiae* est fourni en Figure 18 [117].

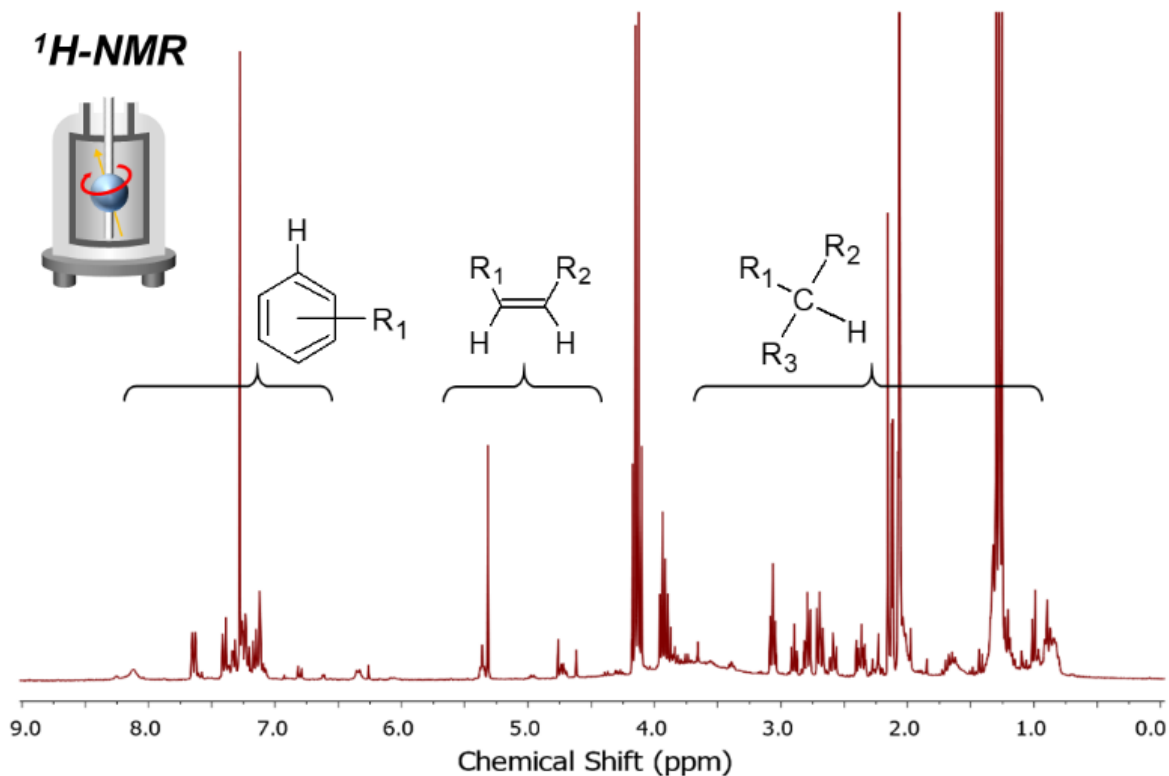


Figure 18 : spectre $^1\text{H-NMR}$ d'un extrait dichlorométhane / acétate d'éthyle (50/50, v/v) acquis sur un instrument équipé d'un aimant de 300,1 MHz. L'extrait sec a été resolubilisé dans du chloroforme deutéré [117].

5. La spectrométrie de masse

La spectrométrie de masse est une technique analytique qui permet la mesure du rapport masse sur charge m/z d'un ion. L'intérêt de cette technologie réside dans le fait que ce rapport m/z est une caractéristique intrinsèque de l'ion. Par déduction, la spectrométrie de masse apporte des informations concernant la formule moléculaire de l'espèce analysée et des formules structurales grâce à la réalisation d'expériences de spectrométrie de masse en tandem (MS/MS). Les principaux atouts de la spectrométrie de masse sont sa sélectivité, sa sensibilité importante avec des limites de détection très basses ainsi que sa vitesse d'analyse élevée [118]. Il faut souligner que la spectrométrie de masse implique une grande diversité d'instruments et d'applications.

Principe de la spectrométrie de masse

La première étape de l'analyse par spectrométrie de masse est l'ionisation en phase gazeuse ou en phase condensée des composés d'un échantillon par une source d'ionisation après son injection dans le système. L'ionisation, selon la technologie employée, peut créer des fragments des composés de l'échantillon. Les ions, et éventuels fragments, générés sont séparés ensuite dans un analyseur, en fonction de leur rapport masse sur charge (noté m/z), puis sont détectés à l'aide d'un détecteur convertissant les ions en signal électrique. Un diagramme simplifié des différents éléments constituant un spectromètre de masse est fourni en Figure 19 (inspiré de la littérature [118]).

Le résultat brut d'une expérience de spectrométrie de masse est un spectre de masse représenté avec en abscisse le rapport m/z des ions détectés et en ordonnée l'intensité de ces ions. En effet, les ions sont détectés en fonction de leur abondance, rendant la spectrométrie de masse intéressante pour les études quantitatives en plus de son intérêt pour des études qualitatives. De plus, les molécules organiques possèdent naturellement un certain pourcentage d'isotopes comme le ^{13}C . Il est possible d'observer sur un spectre de masse, en fonction de la résolution de l'analyseur, les différents isotopes des ions analysés. Cela peut apporter une information sur la formule moléculaire, car l'allure des massifs isotopiques est liée aux abondances relatives d'isotopes différents selon les éléments et donc est spécifique de la composition élémentaire. Enfin, la spectrométrie de masse en tandem (MS/MS) se focalise sur l'analyse de fragments d'un ion précurseur sélectionné. L'interprétation des spectres MS/MS nous apporte des informations structurales sur la molécule analysée.

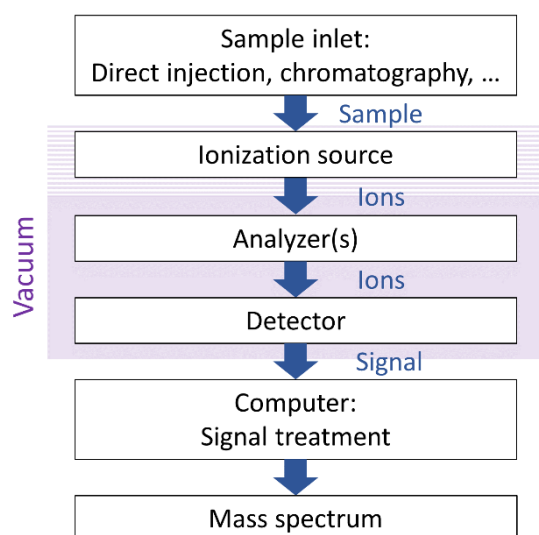


Figure 19 : diagramme de base des composants formant un spectromètre de masse.

La spectrométrie de masse est née à la fin du 19^{ème} siècle, d'après les expériences de J.J. Thomson après sa découverte de l'électron et de la détermination de son rapport m/z . Le premier véritable spectromètre de masse fut construit en 1912 et employé à l'analyse de gaz tel que du dioxygène ou du diazote. Les évolutions de la technique s'ensuivirent tout au long du 20^{ème} siècle, avec le développement de nouvelles sources, de nouveaux analyseurs, ou encore le premier couplage avec la chromatographie gazeuse en 1956 [118]. L'instrumentation en spectrométrie de masse connut cependant sa plus rapide évolution depuis les trois dernières décennies, marquant des améliorations spectaculaires dans la sensibilité des analyses, l'accès à l'ionisation des molécules de grande masse, la précision de mesure ainsi que la résolution des analyseurs de masse (permettant de séparer des composés ayant une masse très proche).

a. Les sources d'ionisation

Depuis les débuts de la spectrométrie de masse, de nombreux principes d'ionisation ont été développés puis améliorés. Chaque source aura des caractéristiques spécifiques qui permettront l'ionisation de différents types de molécules (polaires, apolaires...). Certaines techniques seront dites « dures » car elles conduisent à une fragmentation importante, tandis que d'autres seront dites « douces » et formeront des espèces intactes. Seules certaines de ces méthodes d'ionisation, emblématiques ou ayant un intérêt significatif dans ces travaux de thèse, seront détaillés ci-après.

Ionisation électronique

L'ionisation électronique (EI), autrefois également nommée impact électronique, est une méthode d'ionisation ayant lieu en phase gazeuse et nécessite que la source soit sous vide. Son utilisation est limitée aux composés volatiles et aux molécules stables face à la température. Au sein de la source, un filament émet des électrons qui sont accélérés par une différence de potentiel et qui vont interagir avec les molécules du mélange injecté. Si l'énergie cinétique des électrons est suffisante, ceux-ci vont arracher un électron de la molécule et la transformer en cation radical, voire même la fragmenter en fonction de son énergie interne. Un schéma d'une source EI est fourni en Figure 20 [118]. Les molécules ionisées par EI ont tendance à se fragmenter de manière importante ; ainsi, l'EI est une source relativement simple, équipée sur de nombreux spectromètres de masses, et les analyses qui l'utilisent fournissent de nombreuses informations structurales. Elle permet le couplage avec la chromatographie en phase gazeuse ; on parle alors de GC-MS.

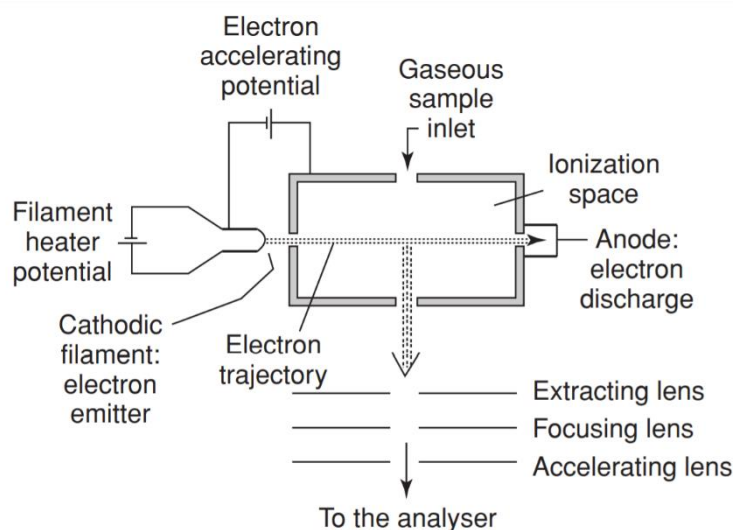


Figure 20 : schéma d'une source d'ionisation électronique [118].

Les sources d'ionisation à pression atmosphérique

Les sources d'ionisation à pression atmosphérique sont des sources ne présentant pas de passage sous vide entre l'injection de l'échantillon et l'ionisation de ses composants. Les systèmes d'ionisation à pression atmosphérique sont considérés comme des sources d'ionisation douce. L'ionisation à pression atmosphérique permet les couplages de la spectrométrie de masse, avec la chromatographie en phase liquide ou l'électrophorèse capillaire (LC-MS et CE-MS respectivement).

Source d'ionisation par électrospray

La source d'ionisation par *electrospray* (ESI), ou électronébulisation, est une source à pression atmosphérique qui génère majoritairement des espèces moléculaires intactes avec peu ou pas de fragmentations. Elle permet l'ionisation de molécules polaires en solution par la formation d'ions qui peuvent être multichargés. Cette méthode d'ionisation permet l'analyse de petits composés mais aussi de très grandes molécules comme les protéines. L'ESI fonctionne de la manière suivante : un capillaire amenant à la source les composants du mélange injecté est porté à un potentiel élevé, et des gouttelettes de solvant contenant les molécules se forment. Un champ électrique intense est appliqué entre le capillaire et une contre-électrode. Les gouttelettes hautement chargées sont formées et sont pulvérisées ; les gouttelettes du nuage deviennent de plus en plus fines. Le solvant des gouttelettes s'évapore progressivement jusqu'à la limite dite « de Rayleigh ». A ce niveau, elles explosent pour libérer les molécules à analyser, et ces dernières seront ainsi déprotonées, en mode d'ionisation négatif, ou protonées, en mode d'ionisation positif, et notées $[M - zH]^{z-}$ ou $[M + zH]^{z+}$; voir Figure 21 pour une illustration du principe de nébulisation [118]. Le nombre de charges portées par les molécules

est variable, et est potentiellement bien plus important plus les molécules sont de haut poids moléculaire (à partir de 1000 Dalton approximativement). Il est important de noter qu'il existe des cas où ce n'est pas un proton qui se fixe mais un autre élément, présent dans l'échantillon analysé ou dans le solvant servant à l'injection : par exemple, l'addition d'un sodium sur le composé amènera à la formation d'un adduit sodium et sera noté $[M + Na]^+$.

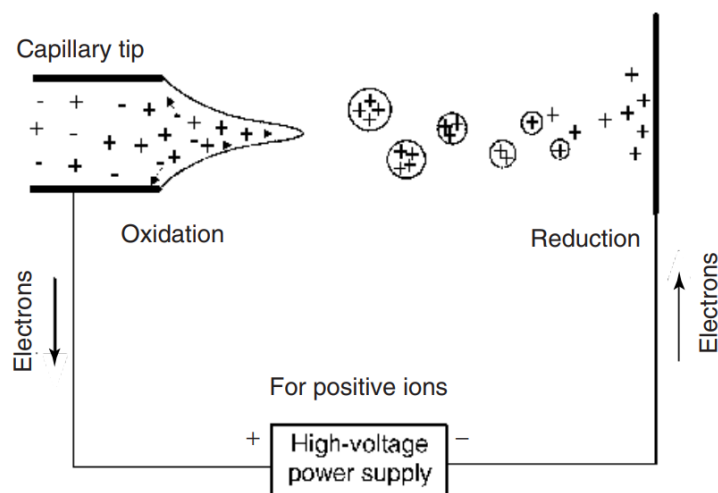


Figure 21 : schéma de la nébulisation des molécules d'un mélange injecté par la source ESI [118].

Ionisation chimique à pression atmosphérique

La source d'ionisation chimique à pression atmosphérique (ou APCI pour *Atmospheric Pressure Chemical Ionization*) est une source principalement appliquée à l'ionisation de petites molécules en phase gazeuse. Elle génère majoritairement des ions monochargés et contrairement à l'électronébulisation permet l'ionisation de molécules peu polaires. Le principe de la source APCI est décrit en Figure 22 [118]. L'échantillon introduit est pulvérisé en un fin brouillard à l'aide du gaz nébuliseur, généralement de l'azote à haut débit. La haute température de la source permet la vaporisation et la désolvatation des composés du mélange. A la sortie du nébuliseur, une électrode émet des électrons sous la forme d'une décharge corona conduisant à l'ionisation du gaz (formation d'ions primaires tels que N_2^{*+} ou O_2^{*+}) puis indirectement du solvant et de l'analyte. Les processus de transfert de charge ou de transfert de protons entraînent la formation d'ions $[M + H]^+$ en mode d'ionisation positif et $[M - H]^-$ en mode d'ionisation négatif ; la fréquence élevée de collisions à pression atmosphérique favorise l'ionisation. Les ions sont ensuite conduits dans l'analyseur.

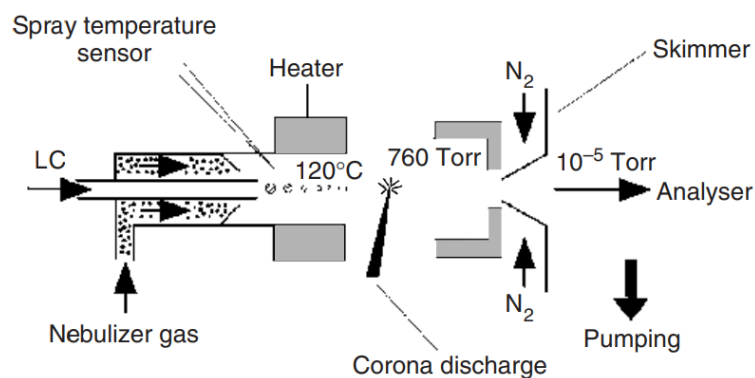


Figure 22 : schéma des composants et du principe d'une source APCI [118].

Photoionisation à pression atmosphérique

La source à photoionisation à pression atmosphérique (ou APPI pour *Atmospheric Pressure Photoionization*) est le système d'ionisation douce le plus récent. La Figure 23 illustre le principe et les différentes parties qui composent cette source [118]. L'échantillon est vaporisé à l'aide d'un gaz nébuliseur et d'un compartiment chauffé similaire à celui de la source APCI. Les photons de haute énergie émis par une lampe UV interagissent avec les molécules du mélange injecté et induisent des réactions en phase gazeuse qui amènent à la formation d'ions. Avec des photons de 10 eV émis par une lampe au Krypton, l'énergie est suffisamment élevée pour ioniser les composés d'intérêt et suffisamment basse pour ne pas ioniser les molécules de solvant (eau, méthanol, ...). Le principe présente des analogies à celui de l'APCI, sauf que c'est un flux de photons et non d'électrons qui ionise les analytes. En pratique l'ionisation directe des analytes est souvent peu efficace. L'efficacité de l'ionisation peut être améliorée de 10 à 100 fois par l'ajout d'un dopant photoionisable, tel que l'acétone ou le toluène, en flux complémentaire dans la source. Le dopant va s'ioniser en premier puis il y aura un transfert de charge avec les molécules de l'échantillon pour les ioniser. L'APPI peut générer des cations radicaux $M^{\bullet+}$ (en absence de solvants protiques) tout comme des molécules protonées $[M + H]^+$ (en présence de solvants protiques) en mode d'ionisation positif, ainsi que des $M^{\bullet-}$ et des $[M - H]^-$ en mode d'ionisation négatif. Le principal avantage de l'APPI est sa capacité à ioniser des molécules peu polaires qui ne sont pas ionisées facilement en ESI.

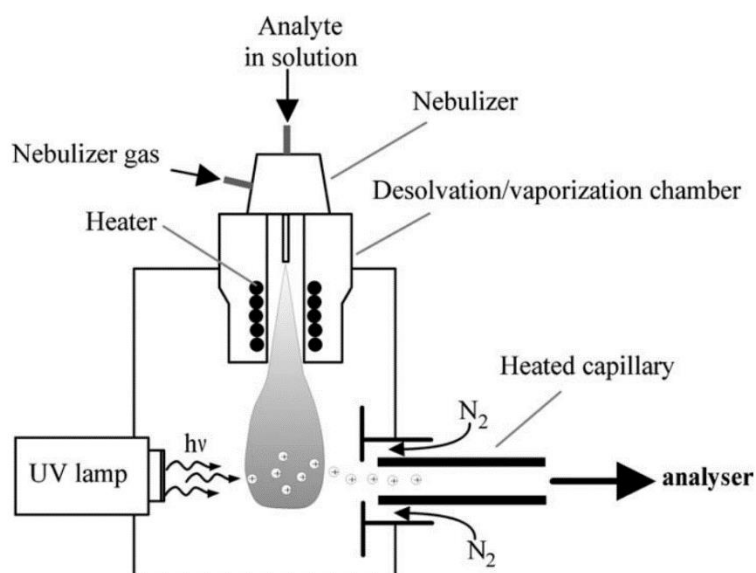


Figure 23 : schéma du principe et des composants d'une source APPI [118].

b. Les analyseurs

Une interface permet le transfert des ions de la source d'ionisation à pression atmosphérique dans l'analyseur sous vide poussé. Cette interface comprend un jeu de lentilles électrostatiques, séparant le trajet des ions entre plusieurs compartiments chacun disposant d'une pompe à vide (pompes à palettes et turbomoléculaires) et des guides d'ions multipolaires. Les ions sont accélérés le long du trajet grâce à des tensions électriques. Il existe une variété d'analyseurs en spectrométrie de masse, reposant sur différents principes physiques. Au sein de ces analyseurs, les ions sont toujours séparés selon leur rapport masse sur charge (m/z), et ce à l'aide de champs électriques, de champs magnétiques ou de combinaisons des deux. Chaque analyseur a ses forces et faiblesses, et les principales caractéristiques à prendre en compte sont la résolution, la précision de mesure de masse, la vitesse d'analyse et la gamme de masse détectable.

En spectrométrie de masse, la résolution R est généralement calculée par $R = m/\Delta m$, où Δm correspond à la largeur d'un signal à mi-hauteur pour un rapport m/z donné. Ainsi, plus la résolution est élevée et plus il est possible de distinguer deux ions dont le rapport m/z est proche ; on parle en général de basse résolution en dessous de $R = 10000$, et de haute résolution au-dessus de cette valeur. Seuls les principes d'analyseurs ayant un impact dans ces travaux de thèse seront détaillés ci-après.

Analyseur quadripolaire

L'analyseur quadripolaire, ou quadripôle (Q) est un analyseur de basse résolution composé de quatre électrodes parallèles de section circulaire ou hyperbolique [118]. Une électrode est soumise au même

potentiel que son homologue opposé, et à un potentiel inverse aux électrodes adjacentes. Un ion entrant entre ces barres est soumis à des champs électriques oscillants, et la séparation s'effectue en fonction de la stabilité de sa trajectoire. Cette stabilité dépend de la masse et de la charge des ions analysés ainsi que des tensions radiofréquences appliquées. Les quadropôles montrent néanmoins des faiblesses dans l'analyse d'ions de haute masse, car leurs trajectoires peuvent montrer une stabilité médiocre et provoquer l'écrasement des ions sur les barres. L'analyseur quadripolaire est très performant dans le cadre d'une analyse ciblée (*selected ion monitoring*) mais peu performant en mode non ciblé (mode balayage ou *scan*) du fait dans ce cas d'un faible rapport cyclique (ou *duty cycle*) et de sa faible résolution. C'est pourquoi le quadropôle est souvent couplé à un analyseur de plus haute résolution du type Orbitrap ou temps de vol dans des instrument hybrides.

Il est possible de combiner plusieurs quadropôles entre eux et de mettre en place un système de triple-quadropôle très utile pour la spectrométrie de masse en tandem : le premier quadropôle peut sélectionner un ion, le deuxième faire office de cellule de collision pour le fragmenter et le troisième scanner les fragments de cet ion. Le mode MRM (*Multiple Reaction Monitoring*), permis par les triple quadropôles, sélectionne des ions précurseurs, fragmente ceux-ci et détecte des fragments spécifiques, ce qui se révèle utile pour les études quantitatives ciblées.

Analyseur par temps de vol

L'analyseur par temps de vol (TOF, time of flight) est un concept apparu dans les années 1950 qui a subi de grandes améliorations depuis les années 1980. Les ions entrant dans un TOF sont accélérés par une tension, puis traversent ensuite un tube de vol sous vide, sans champ appliqué, à une vitesse constante [118]. Leur temps de traversée du tube de vol permet la détermination du rapport m/z : plus un ion le traverse rapidement et plus sa masse est faible. Le TOF nécessite un flux d'ions pulsés, ce qui est obtenu directement dans le cas d'une source de desorption/ionisation laser. Dans le cas d'une source d'ionisation fonctionnant en continu, une géométrie orthogonale entre la source et l'analyseur permet d'obtenir un flux d'ion pulsé.

Le TOF a l'avantage d'avoir une gamme de mesure de rapport m/z très grande, et peut détecter de très grosses molécules, jusqu'à plusieurs centaines de kDa. Ces analyseurs à haute résolution sont également très sensibles et leur vitesse d'analyse est très rapide (jusqu'à 100 Hz) [119].

Pour améliorer leur résolution, les analyseurs TOF sont généralement équipés d'un *reflectron*, ou d'un miroir électrostatique. Le dispositif positionné à la fin du trajet des ions est constitué d'une série d'électrodes circulaires ; ces électrodes appliquent un champ électrique de direction opposée à la

tension accélératrice, entraînant le renvoi des ions [120]. Le réflectron sert à corriger les décalages d'énergie cinétique qui peuvent se créer lors de l'accélération d'ions de même masse à l'entrée du tube, améliorant ainsi la résolution. Les ions les plus rapides vont pénétrer plus profondément dans les lentilles avant d'être renvoyés. Le détecteur est alors situé en bout de course du côté de la source.

Le TOF peut être couplé à un quadropôle et une cellule de collision (Qq-TOF) constituant ainsi un instrument hybride permettant de faire des expériences de spectrométrie de masse avec une analyse en haute résolution.

Analyseur à résonance cyclotronique ionique par transformée de Fourier

L'analyseur à résonance cyclotronique ionique par transformée de Fourier (ou FTICR, *Fourier Transform Ion Cyclotron Resonance*), développé au cours des années 1970, est considéré comme un analyseur à ultra-haute résolution (aux côtés de la technologie Orbitrap). En effet, la FTICR-MS a montré des performances importantes en matière de résolution ; une résolution de plusieurs millions peut être atteinte dans certaines conditions. Une telle résolution implique un temps de détection relativement long pouvant atteindre plusieurs secondes. La FTICR-MS est principalement utilisée en introduction directe mais peut aussi être couplée à la GC et la LC. Elle présente également l'une des meilleures précisions de mesure de masse parmi les technologies actuelles avec des erreurs de mesures pouvant atteindre 0,1 ppm [119].

L'analyseur à résonance cyclotronique ionique est composé d'une cellule ICR (ou piège de Penning) situé au centre d'un aimant supraconducteur horizontal. Les ions sont piégés par l'association d'un champ magnétique et d'un champ électrique. La cellule ICR est constituée de plaques sur lesquelles sont appliquées des tensions continues (piégeage axial) ou radiofréquences (étape de détection). Le champ magnétique soumet les ions à la force de Lorentz qui va conduire à piéger les ions de façon radiale au sein de la cellule avec une trajectoire circulaire, dans un mouvement dit « cyclotronique » (circulaire et uniforme selon une certaine fréquence). La Figure 24 illustre la structure d'une cellule ICR ainsi que les mouvements des ions [118]. Chaque m/z aura une fréquence cyclotronique caractéristique mais celle-ci est non détectable car les ions évoluent avec un faible rayon cyclotronique et un mouvement non cohérent. Pour rendre ce signal détectable, un courant alternatif sera appliqué aux deux électrodes d'excitation de la cellule. Les ions seront accélérés et le mouvement des ions sera en phase ce qui générera un courant induit qui sera détecté par les plaques de détection.

Sur un FTICR moderne, les ions sont produits à l'extérieur de la cellule ICR. Ils sont généralement accumulés sur un piège linéaire externe puis sont injectés simultanément, à basse vitesse, dans la

cellule ICR. Les performances d'un FTICR sont liées en particulier au champ magnétique. Plus celui-ci est intense et plus la résolution et la vitesse d'analyse seront élevées. Les instruments commerciaux actuels vont de 7 T à 15 T. Il existe toutefois deux FTICR 21 T installés au NIMFL (Floride) et au PNNL (Etat de Washington).

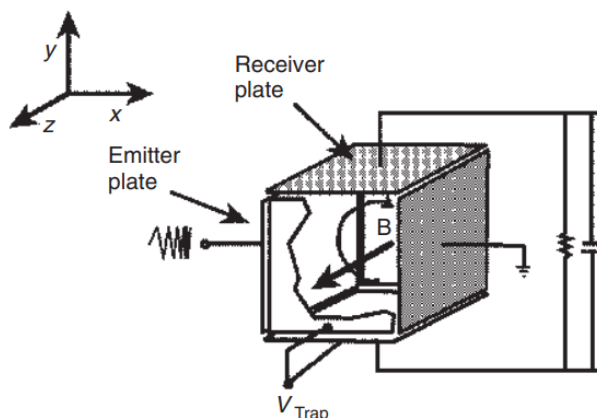


Figure 24 : schéma de la cellule ICR. Le champ magnétique est appliqué selon l'axe z. Les ions sont injectés le long de l'axe z, puis suivent un mouvement cyclotronique autour de l'axe z dans le plan xy [118].

6. Les possibilités d'injection en spectrométrie de masse

Les extraits naturels sont qualifiés de mélanges complexes car ils contiennent une grande variété de biomolécules avec une importante diversité structurale. L'analyse de ces mélanges complexes est souvent réalisée par couplage des techniques chromatographiques avec la spectrométrie de masse. Les possibilités d'analyses ciblées et non ciblées ainsi que les outils de traitement automatisés pour l'assignation de descripteurs aux différents pics facilitent l'interprétation des résultats de telles études.

Le résultat d'une analyse en couplage GC-MS ou LC-MS sera une série de spectres de masse enregistrés au cours du temps. Ils seront représentés sous la forme d'un chromatogramme reconstitué à partir des données de spectrométrie de masse. Ainsi, chaque point du chromatogramme correspond à un spectre de masse. Le chromatogramme peut être un « courant ionique total » (ou TIC pour *Total Ion Current Chromatogram*) et prendre en compte pour chaque point la somme des intensités de tous les ions du spectre de masse correspondant. Il peut être un « chromatogramme du pic de base » (ou BPC pour *Base Peak Chromatogram*, ou BPI pour *Base Peak Ion chromatogram*), similaire au TIC mais où l'intensité d'un point du chromatogramme est celle du pic le plus intense dans le spectre de masse associé. Il peut

également être un « chromatogramme d'ion extrait » (ou EIC, XIC pour *Extracted Ion Chromatogram*) à partir d'un rapport m/z ou associé à un intervalle plus ou moins large suivant la résolution du spectromètre de masse. Les EIC permettent d'obtenir une très grande sélectivité.

a. Couplage de la GC avec la spectrométrie de masse

Principe du couplage

Le couplage de la chromatographie en phase gazeuse avec la spectrométrie de masse (GC-MS) permet la séparation puis l'introduction des analytes en phase gazeuse. La GC-MS est utilisée pour l'identification de molécules, mais aussi pour la quantification de composés présents à de très faibles concentrations, jusqu'à l'état de traces ; elle est ainsi particulièrement adaptée aux analyses menées dans la police judiciaire, aux contrôles anti-dopage, à la détection de résidus d'armes chimiques, à l'analyse de parfums et d'autres échantillons. L'utilisation d'un quadripôle ou d'un triple quadripôle la rend très utile pour la détection ciblée de molécules spécifiques.

En métabolomique, la GC-MS est employée pour l'analyse de composés apolaires, volatils ou rendus volatils après dérivation [121]. Elle permet de séparer les isomères et fournit, grâce aux processus de fragmentation (dus aux sources EI utilisées), des informations structurales essentielles à l'identification de molécules.

La base de données la plus connue pour annoter les composés analysés par GC-MS est celle du *National Institute of Standards and Technology*, abrégée en NIST. Plusieurs études font cas d'annotations attribuées aux composés grâce à la comparaison des données de fragmentation et de temps de rétention avec cette base de données [45,122–125].

*Profilage métabolique ciblé de *S. cerevisiae* par GC-MS*

La GC-MS est une méthode classique utilisée pour le profilage métabolique de *S. cerevisiae* grâce à sa sensibilité et aux informations structurales qu'elle peut fournir. Certains travaux se focalisent sur les approches ciblées. Le rapport m/z de l'analyte déterminé par l'utilisateur est utilisé pour le paramétrage du mode SIM (*Single Ion Monitoring*). L'analyte est ensuite fragmenté dans la source EI et ses fragments sont analysés.

Les objectifs des approches ciblées sont variés. Elle peuvent s'appliquer à la quantification de composés tels que des sucres, des acides organiques ou des acides aminés, à l'analyse du métabolisme primaire pour étudier la production de certaines molécules par des souches modifiées par ingénierie génétique,

ou à l'optimisation d'une méthode d'extraction permettant une perte minimale de métabolites [88,126].

Certaines études sont ciblées sur les voies métaboliques susceptibles de subir des altérations dans le cas de stress environnementaux, par exemple lors de l'ajout de sulfate de zinc à des cultures stressées par de l'acide acétique, ou bien lors de l'ajout d'éthanol [88,122]. Le premier cas a révélé une diminution du nombre d'espèces réactives de l'oxygène libérées après ajout de sulfate de zinc, ainsi qu'une accumulation de glutathion en lien avec la protection des cellules contre l'oxydation provoquée par l'acide acétique, tandis que le deuxième cas démontre une inhibition de la glycolyse des cellules en présence d'une forte concentration en éthanol.

Des analyses ciblées ont été conduites pour l'analyse des acides aminés et non-aminés de *S. cerevisiae* afin de mettre en place une méthode d'extraction la plus large, ou encore dans le but de comprendre les fonctions de certains gènes grâce aux différents métabolites produits [80,127]. Les avantages de l'approche ciblée sont la possibilité de quantifier les composés analysés et le fait que les équipements de GC-MS équipés de quadropôles sont parmi les plus accessibles. Les analyses ciblées sont rapides et sensibles (avec une faible limite de détection [88]) mais nécessitent cependant la sélection au préalable par l'utilisateur des métabolites à analyser.

Profilage métabolique non ciblé de S. cerevisiae par GC-MS

Les approches non ciblées par GC-MS sur *S. cerevisiae* se concentrent par exemple sur l'identification de métabolites secondaires et le lien entre celles-ci avec leur potentielle activité antimicrobienne, ou encore sur l'analyse de composés volatils qui pourraient faire office d'agents de contrôle contre des champignons pathogènes [125,128].

Les profils métaboliques ont également été étudiés pour caractériser les réponses de la levure à des changements de milieu, permettant de noter une diminution des niveaux d'acides aminés essentiels à la synthèse de protéines et à la croissance cellulaire, ou encore le comportement des voies métaboliques dans le cas de traitements avec de la patuline [94,123].

Des souches de *S. cerevisiae* transformées génétiquement ont été profilées pour étudier certains gènes et identifier des protéines, utilisées par des virus pour leur réplication, pouvant être des cibles de médicaments antiviraux [129].

Profilage métabolique de l'exométabolome de S. cerevisiae par GC-MS

Le métabolome de différents vins fermentés notamment par *S. cerevisiae*, ainsi que leur volatilome, ont aussi été analysés par GC-MS pour définir les composés qui ont un impact sur les propriétés organoleptiques du vin, c'est-à-dire les arômes, saveurs et les sensations en bouche. Ces études font état de la présence de nombreux sucres, acides aminés, acides gras, acétals, acides organiques, alcools, cétones et aldéhydes, mais surtout de terpènes et d'esters [89,130,131]. Il a été établi que les terpènes et les esters ont une grande influence sur les arômes des breuvages alcoolisés.

Des vins fermentés par *S. cerevisiae* ont été analysés dans le but d'être comparés avec d'autres vins fermentés par des espèces n'appartenant pas au genre *Saccharomyces* [45,84,132]. Ces travaux montrent que les profils de composés sont différents selon la levure utilisée pour la fermentation et changent lors de co-cultures entre différentes espèces, que certaines combinaisons de souches permettent une diminution des taux d'éthanol, d'acide acétique et d'acétate d'éthyle tout en amenant une nouvelle diversité sensorielle aux breuvages.

Le volatilome de *S. cerevisiae* a également été analysé et a permis de définir la présence de nombreux alcools et esters au sein des métabolites émis par la levure [124]. Les analyses concernant le volatilome d'une culture de microorganismes nécessitent une étape de préparation supplémentaire de l'échantillon qui est l'extraction des métabolites volatils. Celle-ci se fait très souvent par HS-SPME (*Headspace Solid Phase Micro Extraction*, ou micro-extraction par sorption sur phase solide dans l'espace de tête) [130]. Enfin, cette même étude montre qu'il est possible d'améliorer la séparation chromatographique des substances lors d'une expérience de GC-MS en couplant deux colonnes de GC l'une après l'autre. On parle alors de GC bidimensionnelle (GCxGC).

Limites de la méthode

Un inconvénient rencontré avec la GC-MS est la nécessité de dériver les molécules polaires pour pouvoir les analyser, car il peut s'agir d'une étape particulièrement chronophage prenant de 2 heures à plus de 20 heures dans certaines situations [122]. Les étapes d'extraction par HS-SPME peuvent également être chronophages. Au-delà de ces désavantages, la GC-MS reste une technique fiable et rapide utilisée couramment pour le profilage métabolique d'extraits de levures, principalement dans les études ciblées, quantitatives, les analyses de métabolites volatils ou les études sur les boissons fermentées. Il est important de noter que le traitement de données obtenues par GC-MS peut être plus complexe notamment en raison du grand nombre de fragmentation obtenu par ionisation électronique.

b. Couplage de la LC avec la spectrométrie de masse

Principe du couplage

Le couplage de la chromatographie en phase liquide avec la spectrométrie de masse (LC-MS) est une technique versatile destinée à la séparation et l'analyse de molécules non volatiles. Une large gamme en polarité des métabolites peut être détectée au sein d'un échantillon biologique, notamment grâce à la variété de configurations possibles [133].

Ce couplage fut plus difficile à développer et mettre en œuvre que la GC-MS, car l'échantillon est introduit en phase liquide et la spectrométrie de masse nécessite des ions en phase gazeuse. Une interface entre les deux est nécessaire pour passer d'un environnement pressurisé vers un milieu sous vide. Le développement des sources à pression atmosphérique ont permis le couplage LC-MS.

La LC-MS en métabolomique

La LC-MS est utilisée dans beaucoup de secteurs, en biotechnologie, en chimie de l'environnement, pour l'analyse agro-alimentaire, dans les secteurs pharmaceutiques et cosmétiques. Le couplage LC-MS est très populaire en métabolomique dans l'analyse de produits naturels car informative, sensible et tirant le meilleur des deux technologies combinées. La LC-MS permet la détection d'une large gamme de métabolites [134]. Ce couplage est réalisé en chromatographie en phase normale, phase inverse ou en chromatographie HILIC. Il est nécessaire, en LC-MS, de choisir la phase la plus adaptée aux métabolites analysés [135,136]. La LC-MS permet d'analyser des molécules polaires sans les dériver. Comparativement à une analyse par introduction directe, l'étape de séparation chromatographique permet une séparation des isomères et de limiter les problèmes d'effets matrices.

*Profilage ciblé de *S. cerevisiae* par LC-MS*

Des analyses ciblées des métabolites primaires d'extraits de cultures de *S. cerevisiae* ont été réalisées pour identifier et quantifier des composés d'un échantillon et développer des méthodes de marquage de métabolites avec des isotopes [137,138]. Ces analyses ciblées sont réalisées le plus souvent avec un instrument de type triple quadropôle. La quantification permet également la caractérisation des changements des concentrations en métabolites intracellulaires dans différentes situations. Les profils métaboliques de plusieurs souches ont pu être comparés, ou encore la réponse de la levure à l'ajout de sulfate de zinc comme expliqué précédemment [88,93]. Les métabolites primaires sont aussi ciblés et quantifiés dans le cas d'études du cycle cellulaire et de la glycolyse chez la levure [139]. Le profilage

métabolique ciblé par LC-MS a aussi pu servir à l'optimisation de méthodes d'extraction avec une perte minimale de métabolites intracellulaires [68,126].

La source ESI opérée en mode d'ionisation négatif permet en particulier de favoriser la détection des molécules ayant un caractère acide sous la forme de molécules déprotonées [88]. La gamme de métabolites détectée est néanmoins améliorée si deux analyses sont conduites en parallèle, selon les deux polarités. Cela a permis à certains travaux de compléter les profils métaboliques provenant d'analyses ciblées de vins fermentés par *S. cerevisiae* et de caractériser des variations considérables de leur composition au cours de la fermentation [89].

Profilage non ciblé de S. cerevisiae par LC-MS

Les approches non ciblées par LC-MS sont supportées par l'utilisation d'analyseurs à haute résolution tels que les analyseurs TOF et Orbitrap. La LC-MS supporte l'annotation des métabolites par la comparaison de temps de rétention, des mesures de masse précise et de la fragmentation avec les bases de données spectrales [89]. La MS/MS est un outil puissant pour l'identification structurale. De plus, la séparation chromatographique permet la séparation des isomères. La gamme de métabolites détectés est étendue par l'utilisation des deux polarités d'ionisation et de différentes sources d'ionisation. Ces avantages et possibilités de la LC-MS sont soulignés par certaines études, notamment dans le cas du profilage de souches de *S. cerevisiae* stressées par exposition à une concentration élevée de calcium [135].

Des expériences de LC-ESI-LTQ-Orbitrap ont été menées sur des extraits de la levure [140]. La comparaison des masses mesurées avec la base de données YMDB a permis d'annoter 65 métabolites de *S. cerevisiae*, comprenant de nombreux alcools, acides aminés et glucides. La chromatographie HILIC a également été combinée à une instrumentation ESI-TOF-MS pour le profilage métabolique de la levure et pour caractériser sa réponse à l'ajout de sulfate de cuivre [136]. Cet ajout entraîne la modification significative de la concentration de 14 métabolites ayant un lien avec la protection contre le stress oxydatif et les dégâts sur l'ADN.

La technique d'UHPLC-MS/MS a été employée pour l'analyse de la composition lipidique d'extraits de la levure afin de caractériser les changements métaboliques permettant l'acclimatation à la température, en combinaison avec la RMN [114]. Les membranes des levures semblent ajuster leur composition lipidique et ainsi leur fluidité pour que la cellule s'adapte aux températures basses comme élevées.

Avantages et limites de la LC-MS pour le profilage

La LC-MS est une technique précise et fournissant une bonne fiabilité des profils métaboliques. Moins rapide que la GC-MS, elle permet l'analyse de molécules polaires sans dérivatisation et la séparation des isomères. De multiples technologies sont accessibles comme la très haute résolution et la fragmentation précise d'ions sélectionnés.

La gamme de métabolites détectés est étendue grâce aux différentes configurations du couplage (en phase inverse ou normale, les gradients et solvants choisis, les sources et analyseurs). Cette gamme se révèle différente de celle accessible par la GC-MS. Il est d'ailleurs à noter que la LC-MS est souvent utilisée pour l'analyse de *S. cerevisiae* en complément d'autres techniques, notamment de la GC-MS [88,89,93].

Cependant, malgré ces avantages, la co-élution est toujours possible. Une meilleure séparation chromatographique nécessite de longues expériences et de longues séquences. D'autres inconvénients peuvent exister, tels que la distorsion de la ligne de base, le décalage des temps de rétention entre les analyses consécutives, et la déformation des pics. Ces inconvénients tendent néanmoins à disparaître grâce aux technologies et instruments modernes. La LC-MS reste un outil très intéressant pour la métabolomique et le profilage métabolique.

c. Couplage de l'électrophorèse capillaire avec la spectrométrie de masse

Principe du couplage

L'électrophorèse capillaire (ou CE, pour *capillary electrophoresis*) est une technique séparative où les analytes sont injectés dans le système, traversent un capillaire de très fin diamètre et migrent à travers une solution d'électrolytes. Un champ électrique est appliqué au milieu expérimental, et les composés sont séparés en fonction de leur mobilité électrophorétique, grandeur liée à la charge et la taille des molécules.

Souvent utilisée avec une source ESI, l'électrophorèse capillaire couplée à la spectrométrie de masse (CE-MS) permet d'analyser des molécules chargées ou très fortement polaires. C'est un couplage résolutif, sensible et nécessitant un petit volume d'échantillon. Il est utilisé de manière qualitative et quantitative dans plusieurs domaines où des biomolécules sont analysées, tels que la protéomique ou la médecine clinique.

Profilage de S. cerevisiae par CE-MS

Plusieurs études traitent de la métabolomique sur *S. cerevisiae* par CE-MS, autant en mode d'ionisation positif que négatif. La plupart impliquent des analyses ciblées sur les molécules du métabolisme central du carbone de la levure. Les voies métaboliques de *S. cerevisiae* ont été étudiées dans l'optique d'améliorer la production d'acide succinique par des souches transformées grâce à l'interruption de deux gènes et l'introduction d'un gène spécifique à une autre levure, *Schizosaccharomyces pombe* [141]. Les métabolites primaires de la levure, notamment ceux du cycle du carbone et des acides aminés, ont aussi été analysés pour caractériser la réponse du microorganisme à un stress induit par l'ajout d'acide lactique [142]. Ces mêmes composés en addition de sucres phosphatés, d'acides organiques et de nucléotides ont été ciblés pour étudier l'effet d'un ajout de L-méthionine sur le métabolome de *S. cerevisiae* dans le but d'optimiser la production de S-adénosyl-L-méthionine par les cultures [143]. Les métabolites des voies centrales du cycle du carbone ont encore été profilées et identifiées grâce aux modèles de migration et à la comparaison des rapport m/z avec des standards afin de caractériser les changements métaboliques de souches transformées [144].

S. cerevisiae a également fait l'objet de profilage métabolique pour modéliser des stratégies chimiométriques de fusion de données provenant de CE-MS et de LC-MS. Des méthodes de statistiques multivariées ont été appliquées (MCR-ALS, ou *multivariate curve resolution alternating least square*) [145]. Grâce aux bases de données, la complémentarité des deux méthodes employées est démontrée (20 métabolites en commun entre les 40 détectés en CE-MS et les 32 en LC-MS).

Avantages et limites du couplage

La CE-MS est moins fréquemment utilisée en métabolomique que les couplages par techniques chromatographiques. La variabilité des temps de migration rend la CE-MS moins reproductible que les autres méthodes analytiques décrites dans ces travaux. La sensibilité vis-à-vis de la concentration des métabolites est également plus faible. Cette dernière caractéristique a cependant été améliorée durant la dernière décennie et la détection peut désormais atteindre l'ordre de la nanomole.

La CE-MS reste une technologie utile en métabolomique ciblée comme non ciblée [146]. Elle complète la LC-MS en phase inverse tout comme la GC-MS, tout en utilisant un faible volume d'échantillon et en montrant une haute sélectivité pour les composés très polaires ou chargés, tels que les acides aminés ou les nucléotides. Ce couplage est communément utilisé avec des analyseurs comme le TOF pour le profilage métabolique d'échantillons cliniques, d'extraits de microorganismes, de plantes ou d'aliments.

d. Hors couplage, la possibilité d'introduction directe

Principe

La spectrométrie de masse en injection directe (DI-MS), ou en infusion, consiste en l'introduction de l'échantillon dans le spectromètre de masse sans passer par une étape de séparation par chromatographie ou par électrophorèse capillaire. L'échantillon est injecté grâce à un flux continu de solvant. Cela présente l'avantage d'aboutir à l'obtention rapide d'une empreinte métabolique de l'échantillon injecté.

Les inconvénients majeurs de la DI-MS sont la perte des bénéfices de la séparation chromatographique face aux autres couplages, et donc la vulnérabilité aux effets de matrice et à la suppression ionique lors de l'ionisation. En retour, la DI-MS permet l'accès au plein potentiel de la spectrométrie de masse à ultra-haute résolution, notamment avec les instruments de type Orbitrap et les analyseurs à résonance cyclotronique ionique (FTICR-MS). La spectrométrie de masse à ultra-haute résolution fournit des mesures de masse très précises, une discrimination efficace des composés isobares et l'attribution des formules moléculaires uniques aux *features* détectées. La FTICR-MS présente le plus grand pouvoir résolutif. Le couplage avec les méthodes séparatives est possible, mais l'utilisateur ne profite alors pas pleinement des capacités de précision de l'instrument, d'où l'injection directe. En effet cette dernière permet de pouvoir effectuer des temps de détection plus long pour avoir une meilleure résolution. Elle permet également d'avoir un nombre d'ions constant dans la cellule ICR pour avoir une précision de mesure de masse optimale. Les instruments de type FTICR ont un cout élevé et des frais de maintenance plus importants que les autres spectromètres de masse à haute résolution.

La FTICR-MS est appliquée dans les études sur des matrices biologiques pour la caractérisation des signatures chimiques d'échantillons complexes, principalement grâce à l'attribution de compositions élémentaires précises et la construction de diagrammes de van Krevelen. La technique a par exemple été appliquée à l'analyse de vins pour décrire la complexité chimique d'un vin rouge et d'un vin blanc, ainsi qu'à celle de 150 échantillons de whisky provenant de distilleries différentes, révélant l'impact du bois utilisé lors du vieillissement en fût [147,148]. Pour donner un exemple de son application en métabolomique, la FTICR-MS a été employée en combinaison avec la LC-MS/MS dans l'exploration des métabolites secondaires de champignons du genre *Alternaria*, démontrant des différences entre les profils métaboliques de deux espèces [149].

La DI-ESI-MS a été utilisée pour profiler les métabolites de *S. cerevisiae* dans le but de discriminer un certain nombre de mutants, notamment avec l'aide d'analyses multivariées pour comparer les résultats des différentes analyses [150,151]. Il existe également une technique en temps réel d'analyse de métabolites volatils injectant des molécules du mélange en phase gazeuse vers une source d'ionisation par électrospray secondaire (SESI) [152]. Ces travaux réalisés avec un analyseur de type Orbitrap ont résulté dans la détection d'approximativement 300 molécules formant un profil métabolique riche de la levure.

Des analyses par FTICR-MS ont été réalisées pour le profilage de métabolites non-volatiles provenant de vins fermentés par *S. cerevisiae* et par d'autres souches de levures n'appartenant pas au genre *Saccharomyces* afin de caractériser l'impact des microorganismes choisis sur la composition du vin [46]. Les analyses FTICR-MS ont été conduites en mode d'ionisation *electrospray* négatif. Les données acquises permettent des analyses multivariées comme la PCA (analyse par composante principale) et les tests ANOVA. Les formules moléculaires ont été attribuées de façon automatisée, permettant la représentation par des diagrammes de van Krevelen [153,154]. Les vins issus de fermentation par une unique espèce sont clairement discriminés grâce au processus analytique, et les co-fermentations de la boisson par plusieurs souches augmentent la diversité moléculaire, démontrant les interactions entre les différentes levures durant le procédé de fermentation. Les analyses de vins fermentés par *S. cerevisiae* ont abouti à des diagrammes de van Krevelen présentant un fort contenu en molécules composées de CHO et de CHOS, dans les zones correspondant aux molécules glucidiques, polyphénoliques et les acides aminés. Les annotations ont pu être confirmées grâce à des comparaisons sur des bases de données ainsi que des analyses ciblées par LC-MS/MS.

7. Bilan du profilage métabolique de *S. cerevisiae*

Durant les deux dernières décennies, *Saccharomyces cerevisiae* a été le sujet de nombreuses études faisant intervenir des méthodes métabolomiques. Les objectifs sont variés, allant de l'exploration des métabolites secondaires de la levure à la caractérisation des réponses métaboliques face à des stress environnementaux en passant par l'analyse des produits de fermentation du vin. Afin d'illustrer ces idées, le Tableau 1 résume les différentes méthodes utilisées sur *S. cerevisiae* ainsi que le contexte desdites études [117].

Tableau 1 : résumé des techniques de chimie analytiques employées pour le profilage métabolique de *S. cerevisiae* et corrélées au contexte de chaque étude [117].

Method	Context & Aim of the studies	References
<i>Spectroscopic approaches</i>		
RID	Metabolic responses against an environmental stress or a treatment characterization Targeted analyses to quantify primary metabolites of <i>S. cerevisiae</i> mutants	[93,94] [95]
UV-Vis	Targeted analyses to quantify primary metabolites of <i>S. cerevisiae</i> mutants Red wine analysis, <i>S. cerevisiae</i> non-volatile exometabolome analysis	[95] [96]
Raman	Measurement of lipid composition and protein amounts on entire cells	[98]
FTIR	Red wine fermentation product analysis and red wine characterization Entire cell analysis to differentiate fresh cell and lyophilized cell metabolic profiles Metabolic profiling to discriminate yeast mutants or engineered strains	[105,106] [104] [101–103]
<i>NMR approaches</i>		
NMR	Metabolic profiling applied to yeast mutant characterization Characterization of metabolism alterations due to mutations Metabolic responses against an environmental stress or a treatment characterization Optimization of extraction protocol Cider analysis, <i>S. cerevisiae</i> non-volatile exometabolome analysis	[111] [112] [72,114,116] [113] [115]
HRMAS	HRMAS analysis on entire cells	[113]
<i>Separative techniques hyphenated with mass spectrometry approaches</i>		
GC-MS	<i>S. cerevisiae</i> secondary metabolites, antimicrobial molecules analyses and drug screening Flavoring compounds, volatilome exploration and analysis of <i>S. cerevisiae</i> fermented red wine (exometabolome) Metabolic responses against an environmental stress or a treatment characterization Central metabolites, carbon metabolism and metabolic pathway analyses Optimization of extraction or quenching techniques / protocols	[125,128,129] [29,46,48,50,57,59,60,84,85] [93,94,122] [88,123,127] [71,80,126]
LC-MS	Targeted analyses to identify and quantify primary metabolites	[88,137–139]

	Metabolic responses against an environmental stress or a treatment characterization	[26,41,42,63,64,68]
	Red wine analysis, <i>S. cerevisiae</i> non-volatile exometabolome analysis	[89]
	Optimization of extraction protocol	[68,126]
	Untargeted metabolic profiling, or metabolome exploration	[140]
CE-MS	Metabolic profiling to enhance bioengineered-yeast production	[141]
	Metabolic responses against an environmental stress or a treatment characterization	[142,143]
	Targeted metabolic profiling to characterize transformed strains or mutants	[144]
	Optimization of a data fusion strategy	[145]
DI-MS	Metabolic profiling applied to yeast mutant characterization	[150,151]
	<i>S. cerevisiae</i> volatilome exploration	[152]
	Analysis of <i>S. cerevisiae</i> fermented red wine (exometabolome)	[46]

La diversité des raisons de pratiquer le profilage métabolique n'a d'égal que la variété des moyens analytiques appliqués.

Les méthodes spectroscopiques telles que l'absorption dans l'UV-Vis, la spectroscopie infrarouge ou la spectroscopie Raman ont été employées sporadiquement pour le profilage métabolique de la levure modèle. Elles représentent des techniques rapides pour un profilage global d'un échantillon ou d'une cellule entière sans besoin d'extraction préliminaire. Ces techniques manquent toutefois d'informations structurales et de compréhension moléculaire des métabolomes analysés.

Afin de répondre au besoin d'informations structurales, les méthodes basées sur la spectrométrie de masse et les couplages avec la chromatographie ont émergé et se sont imposées en métabolomique.

La GC-MS a été largement appliquée dès le début du profilage de *S. cerevisiae*, pour l'analyse d'extraits de culture ou des composés volatils responsables des arômes des vins. Elle permet l'acquisition d'informations structurales de par la possibilité de fragmentation des ions. La GC-MS se restreint cependant à l'analyse de composés apolaires ou volatils, bien que la dérivation soit possible pour l'analyse de composés polaires.

Les expériences de LC-MS sont plus longues que leurs analogues en phase gazeuse, mais les possibilités de phases, de colonnes et de solvants élargissent la gamme de molécules détectées. La LC-MS est de nos jours une référence pour l'analyse non ciblée en métabolomique. La CE-MS est une méthode qui

complémentaire bien les expériences de GC-MS et LC-MS en montrant une grande sélectivité pour les molécules chargées et hautement polaires. Enfin, la DI-MS, bien que perdant l'avantage de la séparation des composants de l'échantillon, autorise l'accès au plein potentiel de la spectrométrie de masse à ultra-haute résolution comme la FTICR-MS.

Une autre technique intéressante en profilage métabolique est la résonance magnétique nucléaire. C'est une méthode non-destructive, robuste et reproductible très utile pour les études quantitatives. De plus, la RMN (et tout particulièrement la RMN bidimensionnelle) montre un grand intérêt pour l'élucidation structurale de composés inconnus, et ce même sans séparation au préalable des composés. C'est donc un outil efficace de dérégulation. La RMN est moins sensible et moins adaptée aux mélanges complexes que la spectrométrie de masse, mais fournit plus d'informations structurales.

Ces avantages et inconvénients sont résumés dans le Tableau 2. La connaissance de ce que chaque technique peut offrir peut amener l'utilisateur sur une conclusion quant à la technique (ou la combinaison de techniques) idéale pour réaliser au mieux les objectifs d'une étude. *S. cerevisiae* a été largement étudiée mais reste un sujet d'étude intéressant car de nombreuses souches existent et tout n'est pas connu. La connaissance du métabolome de la levure avance en parallèle des évolutions technologiques des outils analytiques. Dans le cas du profilage métabolique, la méthode doit être soigneusement sélectionnée en considération des molécules à analyser et de la nature de l'échantillon. La clé d'une analyse plus exhaustive réside dans la combinaison complémentaire de plusieurs de ces techniques analytiques.

Tableau 2 : avantages et inconvénients des techniques analytiques utilisées pour le profilage métabolique de *S. cerevisiae* [117].

Method	Analysed samples & compounds	Advantages	Drawbacks
<i>Spectroscopic approaches</i>			
RID	Alcohols, sugars, acids	• Fast non-destructive techniques	• Less accurate than MS
UV-Vis	Central carbon metabolites, wine volatile compounds	• Overall profiling and quantification	and NMR
Raman	β -carotene, lipids, proteins from intact cells	• Provides real-time information on entire cells without metabolite extraction	• Less structural information
FTIR	Metabolome, exometabolome		

NMR approaches

NMR	Metabolome, exometabolome; amino-acids, lipids, carbohydrates	<ul style="list-style-type: none"> • Accurate, robust, non-destructive technique • Powerful in accurate structural elucidation and compound identification 	<ul style="list-style-type: none"> • Less sensitive than MS • Difficult treatment with very complex mixtures
HRMAS	Analysis of whole cells		

Separative techniques hyphenated with mass spectrometry approaches

GC-MS	Metabolome, exometabolome; volatile, non-polar compounds (terpenes, esters); derivatized polar molecules	<ul style="list-style-type: none"> • Separation and resolution of isomers • Sensitive, mass information and quantification • EI allows fragmentation and structural information 	<ul style="list-style-type: none"> • Derivatization is time-consuming • Destructive
LC-MS	Metabolome, exometabolome; wide range of metabolites (lipids, amino-acids, organic acids, alkaloids, carbohydrates)	<ul style="list-style-type: none"> • Versatile MS method, isomer resolution • Wide range of detected metabolites • Various mobile and stationary phases, ionization sources and high-resolution analyzers • Powerful data treatment tools 	<ul style="list-style-type: none"> • Liquid chromatography experiments may be longer
CE-MS	Metabolome; highly polar and charged molecules (amino-acids, peptides)	<ul style="list-style-type: none"> • Complements well other MS techniques • Charged and highly polar molecules detection • Small volume of sample 	<ul style="list-style-type: none"> • Less reproducible than other separative methods • Slightly less sensitive
DI-MS	Metabolome, exometabolome; wide range of molecules (carbohydrates, polyphenols, amino-acids)	<ul style="list-style-type: none"> • Access to ultra-high-resolution MS • Accurate mass measurement (unique molecular formula attribution) • Fine and powerful cartography tools 	<ul style="list-style-type: none"> • No chromatographic separation • High cost, difficult maintenance

Ces travaux sur le profilage métabolique de S. cerevisiae ont fait l'objet d'une publication : O. Perruchon et al., Microchem. J. 170 (2021) 106704.

8. Traitement de données en métabolomique

La spectrométrie de masse a pu profiter au cours des trois dernières décennies de grandes avancées technologiques, et le domaine de la métabolomique a pu en profiter et évoluer dans le même sens. Plus particulièrement, le couplage HPLC-MS a connu un essor pour la séparation et l'identification d'une large gamme de métabolites au sein d'un échantillon. De plus, la spectrométrie de masse en tandem permet l'obtention d'informations structurales significatives [158,159].

Le revers de la médaille des progrès liés aux méthodes analytiques tient dans le fait que les techniques de spectrométrie de masse à haute résolution en métabolomique génèrent une quantité colossale de données à traiter. Ainsi, de nombreux outils informatiques et automatisés ont pu voir le jour et être développés au cours des dernières années ; ces outils innovants sont constamment en développement et continuent à être améliorés. L'acquisition des données s'ensuit du traitement et de l'interprétation de celles-ci, et les outils bio-informatiques automatisés présentent les avantages d'atténuer le caractère chronophage de ces étapes tout en esquivant un grand nombre d'erreurs possibles lors du traitement manuel des données.

Il est important de noter que dans le cadre d'études métabolomique non ciblée, il convient de substituer le terme « identification » par « annotation ». En effet, une identification est considérée comme certaine et requiert une corrélation de résultats de temps de rétention et de mesure de masse d'un composé avec ceux d'un standard analysé. Or, lors du traitement d'un grand nombre de données d'un extrait naturel, il existe de nombreux composés dont les standards n'auront pas pu être analysés en parallèle. Ainsi, les identifications laissent place à des annotations, qui correspondent dans le sens à des hypothèses sur l'identité d'un composé détecté. Le terme « annotation » sera principalement utilisé dans la suite de ce mémoire.

a. Le pré-traitement des données spectrales

Les études de profilage métabolique, de caractérisation de biomarqueurs ou encore d'exploration de produits naturels se sont épanouies ces dernières années [9]. Afin d'utiliser au mieux les multiples outils disponibles pour interpréter les signaux acquis dans le cadre d'études métabolomiques, il convient de pré-traiter les données spectrales acquises afin d'en permettre l'injection dans l'algorithme desdits outils ou de faire corréler la présence de métabolites avec un processus biologique. Le pré-traitement permet ainsi d'uniformiser les données afin qu'elles soient comparables statistiquement, mais aussi

d'aligner plusieurs échantillons entre eux ou préparer les données spectrales à un traitement informatisé plus approfondi.

La notion de features

Le pré-traitement des données spectrales, et le traitement de données en général, prend en compte différents descripteurs qui permettent de définir un signal détecté : la mesure de rapport m/z (extrait des données de MS1), la fragmentation du composé si elle a lieu (extrait des données de MS2), ainsi que le temps de rétention du composé. Ces descripteurs réunis permettent de définir une *feature* : ce nom féminin (pouvant être traduit par *variable*) désigne alors un composé, et ses caractéristiques peuvent être affinées au cours des étapes de traitement de données. Le terme *feature* sera employé durant les prochaines rubriques de ce manuscrit pour définir les entités spectrales au cours du traitement.

Le logiciel de traitement XCMS

Dans le but de traiter des données acquises en LC-MS, le logiciel libre XCMS (*various forms (X) of Chromatography Mass Spectrometry*) fut développé et écrit à partir du langage de programmation spécialisé dans les statistiques R dans le milieu des années 2000. XCMS utilise des algorithmes de détection des pics et des méthodes d'alignement non-linéaire des temps de rétention pour faire le lien entre les ions aux temps de rétention et m/z similaires entre plusieurs analyses. Ces informations facilitent l'annotation des composés [160]. Le rapport signal/bruit est également pris en compte pour écarter les signaux générés à partir du bruit de fond observable dans les données brutes.

L'alignement des données permet de relier une même molécule présente dans deux échantillons différents. Ces informations permettent de calculer un profil de correction pour les temps de rétention, et les intensités relatives correspondantes aux différents composés sont directement comparées afin d'identifier les changements significatifs entre les différents échantillons. XCMS a été développé dans le but d'être indépendant des fabricants d'instruments, permettant ainsi le traitement des données de plusieurs types de formats, dont les formats libres. Le logiciel a été amélioré en 2012 en s'agréant d'une version en ligne nommée XCMS Online et se parant d'une interface graphique intuitive [161].

Le logiciel de traitement MZMine2

Dans la même optique, le logiciel MZMine, et par extension sa deuxième version MZMine 2, permet le pré-traitement des données d'expériences LC-MS/MS en convertissant les données spectrales en listes de *features* [162,163].

Dans sa première version de 2006, MZMine inclus des méthodes de traitement et des outils de visualisation de données de profilage moléculaire par spectrométrie de masse. Le logiciel supporte alors les données au format libre (par exemple le format mzXML), peut réaliser des traitements de données automatisés en « lot », c'est-à-dire le traitement d'un grand nombre de fichiers en même temps et selon les mêmes paramètres.

Le logiciel MZMine 2 correspond à une mise à jour de 2010. L'interface a été améliorée et les différents outils disponibles pour traiter des chromatogrammes et spectres sont nommés *modules*. Cet outil n'a pas cessé d'être mis à jour depuis sa sortie et de nouveaux modules sont fréquemment ajoutés aux versions en développement.

MZMine2 permet la visualisation des chromatogrammes sous la forme de TIC, de BPC ou encore de représentations 3D : une première dimension correspond au temps de rétention, une deuxième fait apparaître les spectres de masse disponibles pour chaque temps de rétention et la troisième mesure l'intensité des signaux. Les différents modules de traitement de MZMine2 permettent d'extraire l'information de la donnée brute. La première étape correspond à la détection des rapports m/z significatifs au-dessus d'un seuil défini de bruit de fond. Les chromatogrammes d'ions extraits (EIC) de chaque rapport m/z peuvent alors être construits à partir des ions détectés dans plusieurs spectres de masse consécutifs. Ils sont ensuite déconvolués, une étape qui permet la résolution des isomères ainsi que de retirer les données correspondant au bruit de fond.

Il est également possible d'aligner des analyses ou des échantillons différents pour comparer ceux-ci entre eux ou discriminer des composés d'un échantillon de blanc ou de témoin. Les fragments en source (visibles en MS1), certains adduits et les complexes (dimère ou trimères de molécules) peuvent être identifiés afin d'écarter les signaux non informatifs. Des formules moléculaires peuvent être attribuées aux ions détectés, ainsi que des annotations grâce à la comparaison avec des bases de données internes ou en ligne. Chacune des étapes décrites est gérée par un module spécifique. Un exemple de la variété de modules disponibles est présenté en Figure 25 [163].

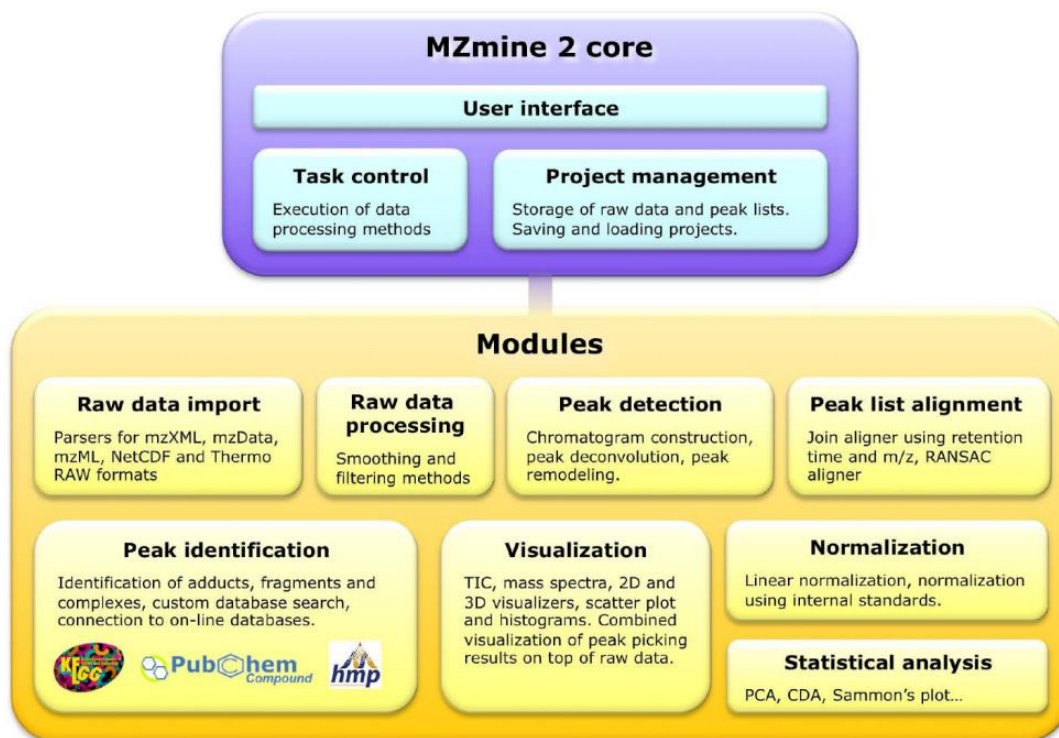


Figure 25 : architecture de MZMine 2 et modules inhérents au logiciel tels qu'ils étaient présentés en 2010 [163].

b. Processus d'annotation de composés

Il est possible d'attribuer des annotations aux composés lors du traitement des données de LC-MS. Les données empiriques telles que le temps de rétention, le rapport m/z et les données de fragmentation MS/MS peuvent être utilisées et comparées avec les données des bases de données spectrales afin d'annoter les *features*.

Comparaison avec les bases de données

Il est ainsi possible de comparer le m/z issu de la mesure de masse précise avec des bases de données telles que la Yeast Metabolome Database (YMDB), la Human Metabolome Database (HMDB) ou MassBank [164–167]. La YMDB est une base de données spécifiquement destinée aux métabolites de levures, principalement *S. cerevisiae*. Elle comprend actuellement plus de 16000 molécules très variées, ainsi que des informations supplémentaires sur les voies métaboliques associées aux métabolites compilés. De multiples rubriques apportent des renseignements sur chaque page dédiée à une molécule, telles que les numéros d'identification et synonymes du composé, la classe moléculaire, les propriétés physico-chimiques ou les propriétés biologiques. La HMDB ressemble beaucoup à la YMDB concernant son interface, et compile quant à elle plus de 110000 métabolites retrouvés dans le corps

humain. Elle est aussi bien utilisée pour des études en métabolomique que dans le domaine biomédical et la découverte de biomarqueurs. HMDB fournit également des spectres MS/MS expérimentaux ou des données de fragmentation théoriques (dite *in silico*), afin de permettre les annotations par comparaison. Enfin, MassBank regroupe plus de 85000 spectres d'environ 15000 composés [167]. Cette base de données permet des recherches à partir du m/z , à partir d'une liste de pics, ou bien à partir des pics obtenus en MS/MS.

c. Les réseaux moléculaires

Principe des réseaux moléculaires

La construction de réseaux moléculaires est une approche de bio-informatique récente et novatrice utilisée dans le traitement, la visualisation et l'interprétation des données acquises lors d'expériences de spectrométrie de masse en tandem (MS/MS) et de LC-MS/MS. Les réseaux moléculaires se basent sur les données MS/MS des *features* détectés. En effet, les composés partageant des similarités de fragmentation sont considérés comme partageant des similarités structurales, car les groupements ou noyaux en commun forment les mêmes fragments pour des composés proches. Les réseaux moléculaires permettent d'accélérer et de préciser l'interprétation des données, une étape habituellement chronophage et complexe dans les projets de métabolomique.

Construction de réseaux par GNPS

La plateforme *Global Natural Products Social molecular networking* (GNPS), développée par un laboratoire de l'Université de Californie San Diego et mise en ligne en 2014, rend accessible aisément la construction de réseaux moléculaires par la communauté scientifique [168]. L'algorithme MS-Cluster fusionne les spectres identiques, puis les spectres MS/MS sont appairés en fonction de leurs rapports m/z et de l'intensité des ions. Le degré de ressemblance entre deux spectres est mesuré à l'aide du « cosine score », variable allant de 0 à 1. Plus les spectres MS/MS sont similaires, plus la grandeur s'approche de 1. Le cosine score minimal pour appairer deux composés entre eux est paramétrable, au côté de nombreux autres paramètres servant à raffermir ou relâcher la tolérance de la construction des réseaux.

Il est à noter que plusieurs analyses peuvent être injectées sur la plateforme pendant le processus. Les réseaux construits vont alors prendre en compte ces échantillons différents et relier les composés similaires entre les différents échantillons, permettant ainsi leur comparaison.

Annotation des composés

La plateforme GNPS a également accès à plusieurs bases de données spectrales participatives, s'agréant constamment de nouveaux spectres MS/MS. Lors de la construction de réseaux, les spectres MS2 contenus dans ces bases sont comparés aux spectres expérimentaux, afin d'annoter les composés correspondant à ces spectres. Le principe de construction de réseaux et d'annotation est résumé en Figure 26 [168]. Des paramètres sont indiqués par l'utilisateur tels que la tolérance de rapport m/z entre l'ion expérimental et celui de la base de données, ou le nombre de fragments en commun entre le spectre soumis et ceux des bibliothèques en ligne.

Visualisation des réseaux

Après construction, les réseaux moléculaires peuvent être visualisés à l'aide du logiciel Cytoscape. Les différents échantillons ou réplicas peuvent être colorés différemment afin de les comparer.

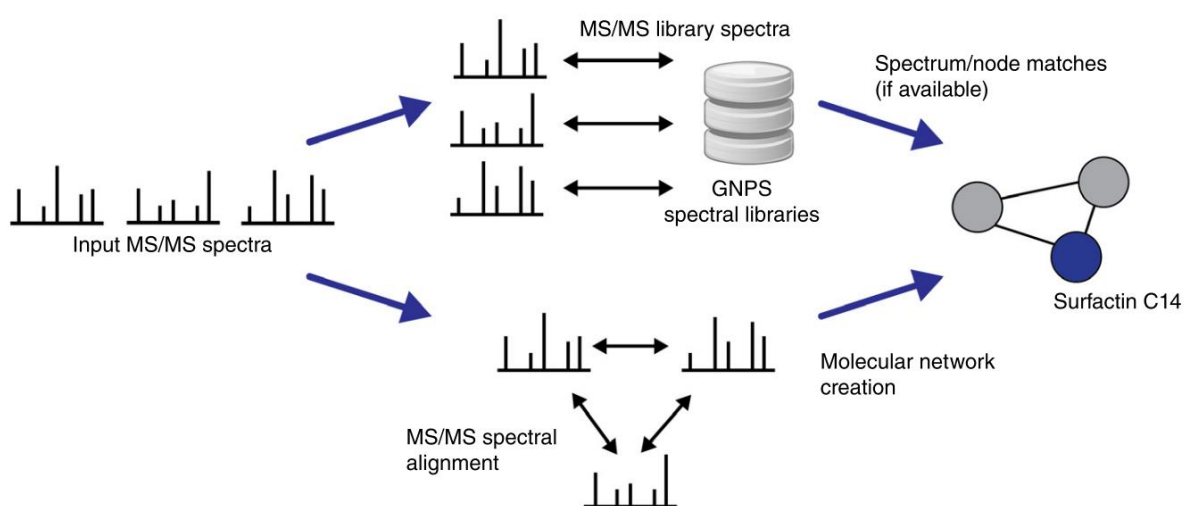


Figure 26 : principe de construction de réseaux moléculaires et d'annotation de composés par GNPS [168].

Construction de réseaux par MetGem

Le logiciel MetGem, développé en France, reprend quant à lui le principe de réseaux moléculaires en l'améliorant par l'ajout d'un algorithme de traitement statistique multivarié, le t-SNE (*t-distributed Stochastic Neighbor Embedding*) [169]. L'utilisation de MetGem requiert le pré-traitement des données de LC-MS/MS par MZMine 2. L'algorithme t-SNE réduit les dimensions de représentation des réseaux et permet la reconnaissance des modèles de fragmentation. La distance entre les différents nœuds au sein du réseau est prise en compte et a une importance dans l'interprétation des données. Le logiciel

MetGem propose une représentation en réseaux classique et la représentation spécifique au t-SNE. L'interface est intuitive et un grand nombre de paramètres réglables en temps réel.

L'intérêt de l'utilisation de réseaux moléculaires est d'améliorer le processus d'annotation en partant d'une base d'hypothèse. Les composés structurellement proches sont rassemblés, et comme ces molécules font probablement partie de la même classe chimique, l'attribution d'annotations est grandement accélérée. Il y a également un intérêt de cartographie, pour visualiser les différents métabolites présents dans les échantillons et de les comparer entre eux. L'outil de réseaux moléculaires est utilisé dans la littérature pour l'exploration de produits naturels à la recherche de molécules bioactives aux propriétés pharmaceutiques potentiellement intéressantes, par exemple pour la déréplication d'extraits de cultures bactériennes ou encore le profilage des composés phénoliques de fruits [155,170,171].

d. Le traitement spécifique aux données de FTICR-MS

Attribution de formules moléculaires uniques

Dans le cas d'analyse de matrices complexes par spectrométrie de masse, des difficultés d'attribution d'une formule moléculaire à un composé sont rencontrées lorsque plusieurs formules sont possibles avec un faible indice d'erreur, et ce même avec des analyseurs à haute résolution (tel que le TOF, aux alentours de 5 ppm d'erreur). Les instruments à ultra-haute résolution, et en particulier la FTICR-MS, possèdent le pouvoir résolutif et la précision nécessaires pour résoudre ce problème (avec une résolution de plus d'un million et une erreur inférieure à 0,1 ppm). De telles capacités permettent l'attribution d'une formule moléculaire unique à un signal détecté.

La cohérence de la formule moléculaire est aidée par le respect des sept règles d'or établies dans la littérature pour l'attribution de formule [172]. Ces règles dites heuristiques permettent d'éviter les formules aberrantes en considérant les éléments observables dans le monde vivant. On y retrouve des restrictions vis-à-vis du nombre d'éléments, la cohérence de l'enveloppe isotopique des signaux, la cohérence du rapport hydrogène / carbone, puis les différents rapports entre les éléments azote, oxygène, phosphore et soufre sur l'élément carbone. On retrouve également le respect des règles concernant la valence des atomes concernés, la prise en compte des rapports élémentaires les plus probables, ainsi que la présence ou non de groupements triméthylsilyles lors d'analyses en GC/MS. La plupart des logiciels permettent l'application des sept règles d'or lors de l'attribution automatisée. La confiance dans ces formules moléculaires uniques est renforcée par la comparaison de la structure

isotopique fine simulée de la formule attribuée avec celle observable sur les données spectrales [173]. Cela permet en métabolomique de confirmer la présence d'éléments spécifiques tels que le soufre.

Représentation des données

Les données d'attribution de formules peuvent ensuite être utilisées pour la création d'une variété de diagrammes servant alors à l'interprétation des résultats et la comparaison de différents échantillons entre eux. Il est possible de construire des diagrammes de van Krevelen, un type de représentation conçu pour l'analyse de produits pétroliers et où les *features* sont répartis selon leurs rapports oxygène / carbone (O/C) et hydrogène / carbone (H/C) [153,154]. Cela permet d'estimer les différentes classes chimiques au sein d'un échantillon et la répartition des composés dans celles-ci. En effet, des molécules composées de nombreux atomes de carbone et de peu d'oxygènes, tels que des lipides, se situeront dans une zone très éloignée d'espèces très oxygénées comme des glucides. Le diagramme de van Krevelen est donc un puissant outil de cartographie des substances présentes dans un échantillon.

Les diagrammes d'équivalents de doubles liaisons contre le nombre de carbone (*DBE vs carbon number*) permettent de caractériser la composition des échantillons complexes en fonction des doubles liaisons et de l'aromaticité des composés, permettant ainsi une interprétation alternative des classes chimiques de l'échantillon [174,175].

Enfin, les diagrammes de Kendrick regroupent dans une représentation graphique les composés d'un échantillon répartis selon leur défaut de masse de Kendrick (base sur la masse d'un groupement CH_2) ainsi que selon le nombre d'hétéroatomes sélectionnés [176].

Chapitre I : Développement de la méthodologie de déréplication de nouvelles levures.

Le projet de ma thèse s'intitule « Identification et cartographie des métabolites secondaires de nouvelles souches de levures tropicales ». Il s'agit d'un sujet transdisciplinaire qui conjugue la chimie des substances naturelles avec les technologies de chimie analytique. L'objectif principal est la sélection de nouvelles souches de levures isolées en Guyane française sur des matrices diverses, puis l'exploration des métabolites secondaires que ces microorganismes peuvent produire grâce à l'instrumentation de pointe dont le laboratoire COBRA est équipé.

Plusieurs volets analytiques furent envisagés, en commençant par les méthodes de chromatographie liquide couplée à la spectrométrie de masse, et plus particulièrement l'UHPLC-MS/MS. La spectrométrie de masse en tandem pour fragmenter les composés des échantillons analysés est alors valorisée par l'application d'outils bio-informatiques innovants, notamment les réseaux moléculaires. La méthode de chromatographie gazeuse couplée à la spectrométrie de masse est également envisagée pour compléter la gamme de métabolites détectés par UHPLC-MS/MS en terme de polarité.

La thèse a ainsi pu débiter sur la sélection des différentes souches fournies par notre partenaire scientifique, la branche « Levures » du Centre International de Ressources Microbiologiques (CIRM-Levures ; <https://www6.inrae.fr/cirm/Levures>), le développement d'une méthode d'extraction rapide, simple et la plus exhaustive possible pour extraire les métabolites secondaires des champignons étudiés, ainsi que l'application de la méthode analytique envisagée par UHPLC-MS/MS à des extraits de levures.

I. Sélection des souches

La nature comporte de nombreuses sources de produits naturels et de biomolécules intéressantes. De nos jours le domaine s'étend et de nombreuses recherches s'organisent pour diversifier les matrices explorées et identifier un plus grand nombre d'espèces végétales mais aussi des bactéries ou des champignons. Dans cette optique, l'exploration de souches de levures nouvelles peut être tout à fait intéressante pour la découverte de composés potentiellement intéressants pour les domaines pharmaceutiques, agro-alimentaires ou cosmétique.

a. Exploration de la flore levure en Guyane française

Les équipes du CIRM-Levures ont récemment réalisé une exploration de la flore levure en Guyane française. Les substrats de prélèvement sont variés et vont de matrices animales comme des insectes,

des serpents ou des grenouilles jusqu'à différents organes de plantes (fruits, fleurs ou gousses). Les souches prélevées ont ensuite été lyophilisées sous vapeur d'azote puis conservées à – 80 °C.

Les différentes levures isolées ont ensuite été les sujets d'études microbiologiques afin de caractériser les espèces récoltées par des techniques de biologie moléculaires. Le CIRM-Levures a ainsi pu nous fournir un tableau comprenant les différentes souches. Ce tableau répertorie plus de 350 prélèvements auxquels sont attribués un genre et une espèce. Il a donc été nécessaire de sélectionner les souches qui allaient être le sujet d'étude de la déréplication par les méthodes analytiques.

b. Critères de sélection

Le premier critère de sélection des souches de levures est le critère de nouveauté de l'espèce sélectionnée. Une souche isolée qui se révèle être une souche différente d'une espèce déjà bien connue n'est pas retenue, car il y a des chances que l'organisme en question ait déjà été exploré pour ses petites molécules produites. Une nouvelle espèce caractérisée aura le critère d'originalité qui, bien que non nécessaire dans l'absolu, assure que la déréplication d'un extrait biologique de cette souche apporte véritablement de nouveaux éléments de connaissance vis-à-vis des métabolites des champignons de ce genre.

Ensuite, un second critère est l'appartenance d'une souche à un genre déjà connu. La littérature scientifique concernant les genres indiqués dans le tableau fourni par notre partenaire a donc été balayée afin d'en isoler les genres les plus intéressants. Ce qui rend un genre intéressant pour l'exploration est la présence dans la littérature d'études démontrant qu'un organisme appartenant audit genre soit capable, selon les conditions et le milieu, de produire des métabolites possédant une ou plusieurs applications qui se révéleront intéressantes pour l'industrie.

c. Exclusion de certains genres

Certains genres furent exclus de la sélection pour certaines raisons. La première est le grand nombre de travaux déjà réalisés sur des espèces du genre, d'autant plus s'ils traitent d'analyses des voies métaboliques des levures. C'est par exemple le cas du genre *Pichia* ayant fait l'objet de dizaines de milliers de publications. Il a été largement étudié pour ses arômes produits et autres molécules responsables des propriétés organoleptiques des aliments dont la fermentation implique des microorganismes de cette famille. D'autres genres tels que les *Cryptococcus*, les *Trichosporon* ou les *Geotrichum* sont dans le même cas et ne sont pas idéaux pour nos travaux exploratoires.

Enfin, le genre *Candida* a également été écarté car de nombreuses levures considérées comme pathogènes en font partie. Nous fûmes néanmoins confrontés à certaines difficultés lors de la compilation d'information sur les genres des souches proposés par le CIRM-Levures, difficultés liées au concept de synonymes chez les microorganismes. En effet, certains champignons classés dans des espèces différentes peuvent être considérés comme synonymes et posséder deux noms d'espèces, dans deux genres différents, selon leur stade sexuel.

d. Kurtzmaniella et Starmerella

Deux espèces du genre *Starmerella* et une espèce du genre *Kurtzmaniella* ont été sélectionnées comme sujets d'étude pour cette thèse.

Kurtzmaniella hittingeri

Kurtzmaniella hittingeri a été sélectionné pour son appartenance au genre *Kurtzmaniella*, un genre plutôt récent et n'ayant pas été le sujet de nombreuses études. La souche isolée possède le numéro d'identification CLIB 1609 et a été récoltée sur des échantillons de baies près de la ville de Kourou, puis caractérisée [177]. C'est le critère de nouveauté qui a primé dans le choix de cette souche. De plus, certaines espèces appartenant au genre *Kurtzmaniella* ont été isolées dans des produits fermentés, notamment des produits de charcuterie, rendant ce genre intéressant selon le second critère [50,51].

A titre d'illustration, une photographie d'une culture de la nouvelle espèce *Kurtzmaniella hittingeri* est présentée en Figure 27 [177].

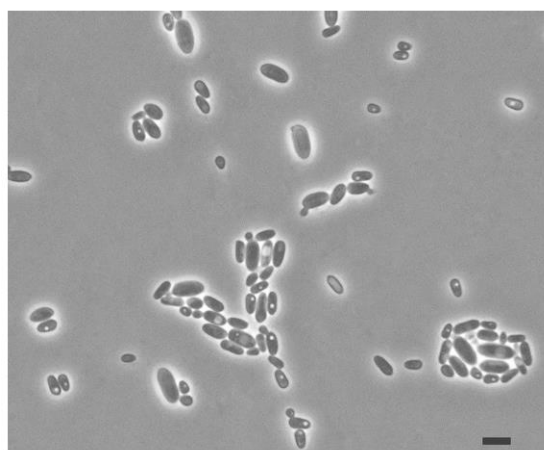


Figure 27 : photographies au microscope de cellules d'une culture de *Kurtzmaniella hittingeri* UFMG-CM-Y272T sur un milieu agar YM incubé pendant 3 jours à 25 °C. Echelle 5µm (barre noire) [177].

Starmerella reginensis et *Starmerella kourouensis*

Deux souches du genre *Starmerella* ont également été sélectionnées pour cette étude. Il s'agit de *Starmerella reginensis* et *Starmerella kourouensis*, respectivement notées CLIB 1634 et CLIB 1707. Elles ont été récoltées sur des fleurs près de la ville de Regina pour *S. reginensis* et près de la ville de Kourou pour *S. kourouensis* [178]. Le genre est intéressant car certaines espèces sont capables de produire des glycolipides tensio-actifs pouvant être utilisés comme détergents ou ayant des propriétés antimicrobiennes ou anticancéreuses [39,40]. Des espèces de *Starmerella* sont également impliquées dans la fermentation de certains aliments tels que le thé kombucha, une boisson d'origine asiatique présentant plusieurs bienfaits parmi lesquels on retrouve des activités antimicrobiennes, antioxydantes, antidiabétique ou antistress [42,43]. Les deux souches sélectionnées sont aussi nouvellement caractérisées et répondent au critère d'originalité.

Des photographies de cultures de *Starmerella reginensis* et *Starmerella kourouensis* sont présentées en Figure 28 [178].

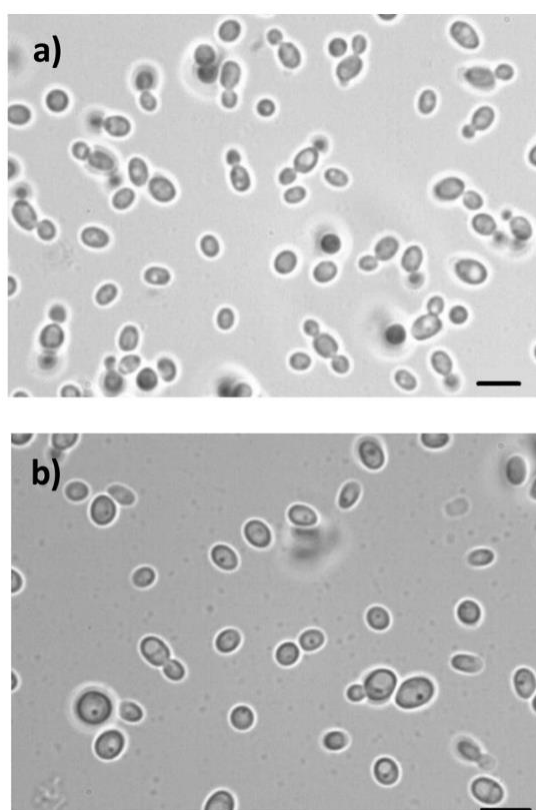


Figure 28 : photographies au microscope de cellules végétatives et bourgeonnantes, levures cultivées sur un milieu Yeast Extract Peptone Dextrose (YPD) et agar pendant 2 jours à 28 °C. Echelle 10 μm (barre noire). (a) *Starmerella reginensis* f.a., sp. nov. CLIB 1634T. (b) *Starmerella kourouensis* f.a., sp. nov. CLIB 1707T. [178]

Ces nouvelles espèces de levures appartenant à des genres intéressants sont un bon exemple des possibilités d'exploration que propose les sources naturelles. Les métabolites secondaires de levures représentent un vivier important de molécules potentiellement bioactives, il est donc d'intérêt de cartographier ces composés via des technologies modernes et des méthodes analytiques les plus précises, rapides et sensibles possibles.

II. Développement du protocole d'extraction des métabolites secondaires

Afin de réaliser la dérégulation des métabolites secondaires de souches de levures, ceux-ci doivent d'abord être extraits des cultures de microorganismes. Pour cela, le protocole d'extraction doit être optimisé de manière à être rapide, reproductible et le plus large possible en termes de polarité des métabolites extraits. Il doit également être adapté à la nature de la culture des microorganismes, qui peut être un milieu liquide ou solide.

1. Culture de levures

Au cours de cette thèse, des cultures de la souche de laboratoire CLIB 338 de la levure modèle *Saccharomyces cerevisiae* ont été réalisées. De nombreuses études existant sur ce microorganisme phare, nous avons choisi celle-ci comme référence pour l'optimisation et la validation des protocoles mis en place. Les souches de levures sont conservées lyophilisées et congelées. Les cultures sont ensuite réalisées en boîte de Petri, sur un milieu solide YPD (*Yeast Peptone Dextrose*) à base d'agar, puis une culture secondaire a été réalisée par ensemencement sur un milieu YNB (*Yeast Nitrogen Base*). Cette culture a ensuite été incubée 3 jours à 26,5 °C.

Nous visons dans cette étude les métabolites secondaires de nouvelles souches de levures. Le protocole a été mis au point grâce à plusieurs sources d'inspiration, dont les références sont indiquées dans le préambule (sur l'extraction de métabolites secondaires d'ascomycètes marins, on en parle encore sur la comparaison de multiples méthodes [74,77]).

2. Extrait apolaire de levures

La première étape intégrée au protocole d'extraction fut l'ajout de solvant organique. Au vu de la littérature et du choix de la plus large gamme de polarités de métabolites possibles, nous avons choisi un mélange de solvants de polarité moyenne : un mélange de dichlorométhane et d'acétate d'éthyle (50/50, v/v). La gélose des cultures est découpée afin de la placer dans un flacon, puis le mélange de solvants est ajouté. Celui-ci a pour effet d'éteindre les voies métaboliques des levures cultivées (le *quenching*) et de déstabiliser les membranes et parois cellulaires. Il est important ici de noter que les métabolites excrétés par les cellules font également partie des métabolites extraits, car tout le milieu de culture baigne dans le mélange.

L'extraction se poursuit par la disruption mécanique des cellules par sonication. Le traitement aux ultrasons de 30 minutes permet de terminer l'éclatement des cellules et donc de déverser dans le mélange de solvant les métabolites intracellulaires. L'extrait contient à ce stade les métabolites intracellulaires tout comme les métabolites extracellulaires.

Le mélange est ensuite filtré sur un fritté à large porosité. Cela permet l'élimination des plus gros débris, comme les résidus de gélose ou les macromolécules telles que les grosses protéines ou les morceaux de paroi et de membrane. Les restes de milieu de culture sont ensuite repris avec un mélange dichlorométhane / acétate d'éthyle (50/50, v/v) puis regroupés pour donner, après évaporation sous vide à l'aide d'un bain d'eau à 41°C, un extrait sec nommé DCM/AE. L'extrait sec, dont la masse est de 31 mg, est conservé au réfrigérateur à 4 °C jusqu'au moment de sa préparation pour analyse.

3. Extrait polaire de levures

Nous avons choisi de pratiquer une seconde extraction à l'aide d'un solvant d'une polarité différente du premier mélange de solvants afin d'améliorer la gamme de métabolites extraits. Après l'extraction de polarité moyenne, les résidus de milieu de culture sont repris avec du méthanol. Un rapide traitement aux ultrasons est appliqué, puis l'extrait est filtré. Il est à noter que le méthanol a un effet bien plus agressif sur la gélose du milieu de culture, détachant bien plus de macromolécules de la structure et rendant la filtration plus longue et compliquée. L'extrait méthanolique lui-même en devient plus dense et plus pâteux que l'extrait DCM/AE.

L'extrait méthanolique, nommé extrait M, est ensuite évaporé selon la même procédure que l'extrait DCM/AE. Cet extrait visqueux a une masse de plus de 800 mg avec cette première version de la méthode

d'extraction, soit une masse très supérieure à celle de l'extrait DCM/AE. Un échantillon de 10 mg a été collecté et conservé au réfrigérateur pour préparation avant analyse.

4. Extrait du témoin, le milieu de culture sans levures

Des échantillons de milieu de culture, c'est-à-dire seulement la gélose d'agar avec le milieu YPD, ont été préparés dans les mêmes conditions que les cultures de levure. Ils ont également subi le protocole d'extraction et abouti à la production d'extraits DCM/AE et M.

5. Préparation des échantillons pour analyse

Les extraits DCM/AE de culture de levures ainsi que les extraits de milieux ont été solubilisés dans du méthanol pour analyse. Le volume ajouté a été adapté à la masse des extraits secs pour obtenir des échantillons à une concentration d'environ 10 mg/mL. Dans les cas où des dépôts se sont formés, un rapide traitement aux ultrasons a permis de solubiliser l'extrait dans le solvant. Nous avons également décidé d'ajouter une étape de filtration par filtre PTFE (Polytétrafluoroéthylène) de 0,45 µm à ce stade afin d'être certains d'éliminer les restes de débris cellulaires, de débris de milieux de culture ou de protéines et d'éviter toute pollution dans l'instrument analytique. Une centrifugation à 6000 rpm pendant 7 minutes a aussi été réalisée sur un autre jeu d'extraits afin de comparer la solution la plus viable pour l'élimination des résidus de milieu et autres débris cellulaires. Les échantillons ont ensuite été dilués au 1/10 dans un mélange d'eau et d'acétonitrile (Eau / ACN ; 50/50, v/v) pour analyse.

Les extraits M ont également été solubilisés dans du méthanol dans le but d'obtenir des échantillons à 10 mg/mL. Ils ont cependant présenté plus de dépôt et d'éléments en suspension que les extraits DCM/AE, et ce même après passage aux ultrasons. Ils ont ensuite été filtrés et dilués au 1/10 dans le mélange eau / acétonitrile (50/50, v/v).

III. Analyse d'extraits de levures pour évaluation de la méthode d'extraction

Après extraction des métabolites des cultures de levures, les échantillons préparés ont été analysés par chromatographie liquide à ultra-haute performance couplée à la spectrométrie de masse en tandem.

De premières analyses ont donc été menées afin de mettre au point le procédé analytique, en partant sur des modèles classiques de séquences et de paramètres analytiques utilisés en UHPLC-MS/MS.

1. Etude préliminaire d'extraits de *S. cerevisiae*

a. Analyses UHPLC-MS/MS

Les extraits ont d'abord été séparés par UHPLC en phase inverse puis détectés par ESI-QTOF-MS. Les instruments utilisés sont indiqués dans la section Matériel et Méthodes en page 205. Les analyses sont réalisées avec une gamme de masse m/z 50-2000. Pour l'obtention de spectres MS/MS, un mode de balayage permettant le choix automatique de l'ion précurseur est sélectionné. Pour chaque scan MS1, les rapports m/z d'intensités les plus élevées sont sélectionnés en tant qu'ions précurseurs. Ces ions sont ensuite fragmentés. Après l'acquisition de ces spectres MS/MS, les ions précurseurs choisis au préalable sont exclus pour quelques secondes, afin de ne pas les fragmenter à nouveau et de fragmenter de nouveaux composés. Tout au long de l'analyse, cette méthode est maintenue afin d'obtenir des informations de fragmentation sur le plus de composés possibles. L'acquisition de données MS/MS est une condition indispensable pour l'application des réseaux moléculaires envisagée par la suite.

b. Paramètres d'acquisition

L'étalonnage en masse de l'instrument est réalisé en externe avant chaque série d'analyses grâce à une solution de formiate de sodium. Un étalonnage en direct, pendant les analyses, est aussi possible sur l'instrument grâce au système de « *lock mass* ». Un flux de leucine-enképhaline est injecté en continu dans la source au travers d'un capillaire d'introduction dédié, différent du capillaire de l'échantillon. Une rotation périodique d'une pièce métallique permet l'acquisition du signal de leucine-enképhaline de manière indépendante. Un étalonnage interne peut ensuite être réalisé en continu tout au long de l'analyse.

Enfin, les analyses sont toutes réalisées de manière successive en mode d'ionisation positif ainsi qu'en mode d'ionisation négatif. Chaque échantillon est ainsi injecté deux fois. Le fait de réaliser ces deux expériences permet encore une fois d'améliorer la gamme des molécules détectées, car toutes les molécules ne s'ionisent pas de la même manière en ESI positif qu'en ESI négatif.

2. Vérification de l'efficacité de l'extraction par chromatographie sur couche mince

Les premières analyses sur les extraits de *S. cerevisiae* et sur les milieux de cultures ont donné des résultats avec de très faibles intensités, ne permettant la détection que d'un faible nombre de métabolites avec peu de spectres MS/MS acquis. En effet, le seuil d'intensité pour la fragmentation d'un ion doit être suffisamment élevé afin d'obtenir des informations structurales.

Des expériences de chromatographie sur couche mince de silice ont été réalisées sur les mêmes échantillons analysés par UHPLC-MS/MS. Celles-ci avaient pour objectif la vérification rapide de l'efficacité des extractions réalisées sur les cultures de levures. La principale hypothèse émise est une extraction efficace mais une concentration trop faible de métabolites au sein des extraits pour une détection suffisamment informative lors des analyses.

a. Séparation par CCM

Une partie des extraits DCM/AE et M a été récupérée et dissoute dans une goutte de méthanol. Ces échantillons sont ensuite séparés sur plaque de silice à l'aide d'un mélange dichlorométhane / acétate d'éthyle (50/50, v/v) en guise de phase mobile. Les extraits DCM/AE et M de cultures de levures ont été comparés entre eux, et l'extrait DCM/AE de levure a été comparé avec l'extrait DCM/AE du milieu de culture sans levures. Les résultats sont présentés en Figure 29.

Solvant de migration :
Dichlorométhane /
acétate d'éthyle
(50/50, v/v)

Révélation à la vanilline
sulfurique, observée
sous lampe UV 254 nm

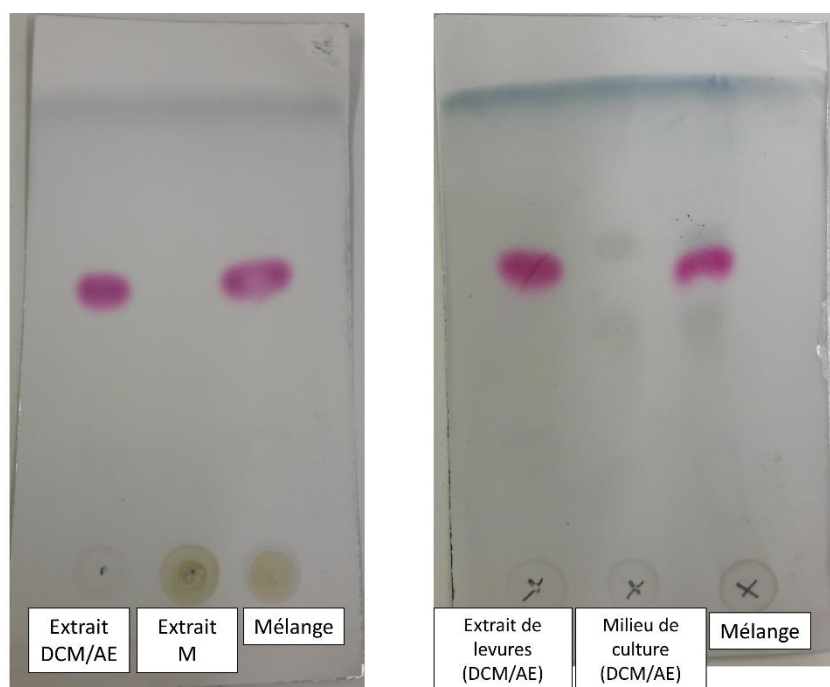


Figure 29 : expériences de séparation sur chromatographie sur couche mince d'extraits de levures ; phase mobile : dichlorométhane / acétate d'éthyle (50/50, v/v).

b. Révélation et résultats

Après révélation à la vanilline sulfurique et observation, des tâches sont observables pour les extraits de levures DCM/AE et M. Des métabolites sont donc bien présents dans ces échantillons. Les molécules dans l'extrait DCM/AE ont bien migré avec la phase mobile elle-même également composée de dichlorométhane et d'acétate d'éthyle. Les substances extraites par le méthanol n'ont quant à elle pas migré car elles sont certainement bien plus polaires que celles présentes dans le premier extrait.

La présence de composés exclusifs aux levures est vérifiée grâce à la deuxième partie de l'expérience (partie de droite de la Figure 29). La migration des composés de l'extrait DCM/AE est toujours visible sous la forme d'une grosse tâche rose, tandis que la migration des molécules de l'extrait DCM/AE du milieu de culture ne donne aucun résultat visible. Les molécules détectées sont donc confirmées comme originaires des levures cultivées.

Cette expérience permet de vérifier que l'étape d'extraction a fonctionné, que les extraits sont bien chargés en métabolites exclusifs des levures, c'est-à-dire ne provenant pas de l'action du solvant organique sur le milieu de culture. Les deux extraits DCM/AE et M ont également leurs propres métabolites extraits, de polarités différentes. Le manque de signal des analyses UHPLC-MS/MS décrit précédemment est donc très probablement lié à la concentration des échantillons et donc aux étapes d'évaporation et de dilution du protocole d'extraction.

IV. Optimisation de la méthode d'extraction

Nous savons que l'extraction a été menée à terme et que les échantillons collectés contiennent des métabolites exclusifs aux levures. Cependant les analyses préliminaires réalisées par UHPLC-MS/MS nous montrent des données spectrales globalement peu intenses. Le manque d'intensité en MS/MS pose un frein à l'utilisation des réseaux moléculaires autour duquel le projet s'articule pour l'interprétation des données et l'exploration de souches de levures.

Une étape d'optimisation de la méthode d'extraction des métabolites de levures ainsi que de la préparation d'échantillon est nécessaire. Les cibles sont principalement l'amélioration de la concentration par évaporation, pour donner des extraits plus concentrés tout en rendant la méthode la plus répétable possible, ainsi que la dilution des extraits lors de la préparation pour l'analyse.

a. Concentration de l'extrait

De nouvelles cultures sur milieu solide de la levure *S. cerevisiae* sont utilisées pour optimiser la méthode d'extraction. Les mêmes étapes d'extraction sont réalisées, et l'évaporation est poussée à son maximum, jusqu'à ce que l'extrait soit le plus sec possible, afin de concentrer au mieux l'extrait.

Cette manière de concentrer les extraits par évaporation permet d'obtenir des extraits M ayant une masse plus faible que la méthode précédente, de l'ordre de 200 mg (contre 800 mg auparavant), témoignant ainsi de la meilleure concentration des échantillons. La masse des extraits DCM/AE ne change quant à elle pas, et va de 10 mg à 30 mg dans la plupart des cas. Ceci est explicable par le fait que le mélange dichlorométhane / acétate d'éthyle était très simple à évaporer, même avec la méthode moins approfondie. Les flacons contenant les extraits secs sont conservés à $-20\text{ }^{\circ}\text{C}$.

b. Préparation des échantillons

La solubilisation de l'extrait est une étape simple mais très importante et ne doit pas être négligée. Les résultats des analyses conduites en aval en dépendent.

Différents essais de solubilité ont été menés pour éviter la création de dépôts ou de suspension de particules qui pourraient polluer ou endommager les instruments analytiques. Les extraits DCM/AE sont ainsi repris dans du méthanol, tandis que les extraits M sont repris dans de l'eau. Le volume ajouté est ajusté en fonction de la masse connue des extraits, dans le but d'obtenir des échantillons finaux dont la concentration est de l'ordre de 10 mg/mL. Les échantillons sont ensuite dilués dans un mélange

eau / ACN (50/50, v/v), un mélange similaire aux solvants utilisés comme phases mobiles dans les expériences d'UHPLC-MS/MS. Le facteur de dilution appliqué est de 1/2 pour garantir une concentration des échantillons supérieure et ainsi obtenir un meilleur signal. Les échantillons sont finalement filtrés sur filtre PTFE de porosité 0,45 μm , pour éliminer les éventuels restes de débris des membranes cellulaires, du milieu de culture ou des protéines.

Le protocole final qui sera conservé tout au long de la thèse pour l'extraction des métabolites des cultures de souches de levure exotiques tout comme pour l'extraction de *S. cerevisiae* est résumé en Figure 30.

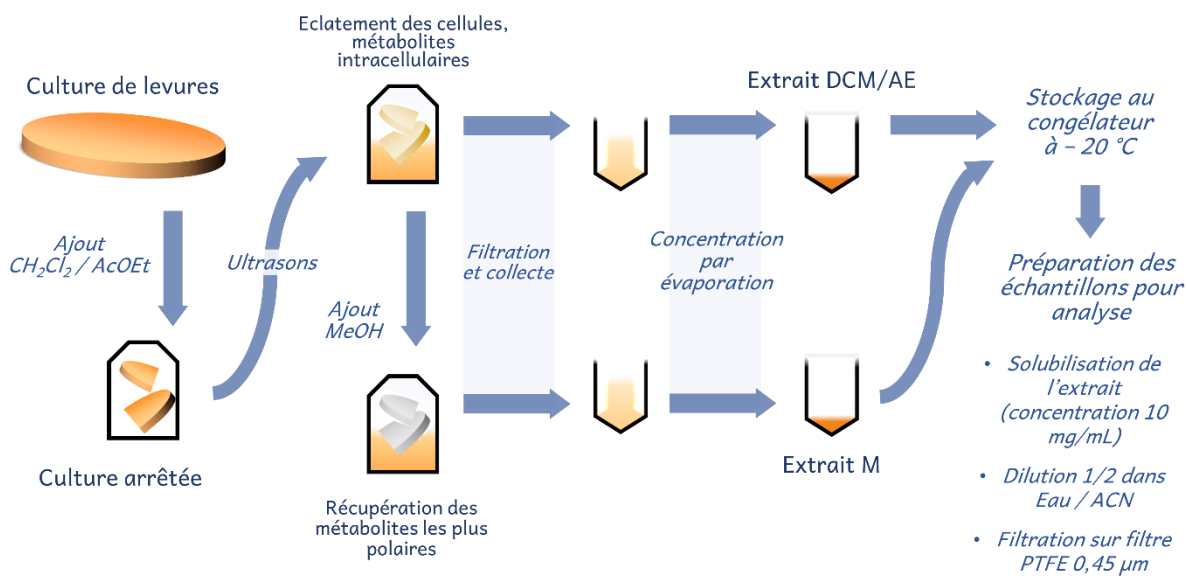


Figure 30 : schéma du protocole final d'extraction de métabolites de cultures de souches de levures.

Chapitre II : Optimisation de la méthode analytique et
application du traitement de données pour
l'exploration d'extraits de levures

I. Les réseaux moléculaires pour la métabolomique

Les expériences UHPLC-MS/MS génèrent de lourds jeux de données. L'utilisation d'outils automatisés de traitement de données est nécessaire pour limiter le caractère chronophage d'une investigation manuelle des résultats analytiques. La plateforme en ligne de *Global Social Natural Products molecular networking* (GNPS [168]) fut sélectionnée pour la construction de réseaux moléculaires. Il s'agit un bon exemple de l'application de programmes informatisés pour analyser des données et les représenter d'une façon aidant l'utilisateur à accélérer le processus d'interprétation de résultats. L'objectif est la création de réseaux et la visualisation de ceux-ci sous le logiciel libre Cytoscape afin d'aider à l'exploration des extraits de levures.

1. Conversion des données spectrales

La première étape nécessaire à la construction de réseaux moléculaires sur l'interface GNPS est la conversion des données dans un format libre. En effet, en spectrométrie de masse, chaque constructeur d'instrument utilise un format de données spécifique pour la création de fichiers après une analyse. Par exemple, les instruments Waters permettent l'acquisition de données sous format « .raw », tandis que ceux de Bruker acquièrent des spectres au format « .d ». Il existe cependant des formats libres, tels que le format « mzML », le format « mzXML » ou le format « MGF ».

a. Conversion par MSConvert

Dans le but de construire des réseaux moléculaires sur la plateforme GNPS à partir des données UHPLC-MS/MS, il est nécessaire de convertir les données du format « .raw » vers un format libre. Notre choix s'est porté sur la conversion en format « mzML » ou en format « mzXML » grâce à l'application MSConvert, une interface utilisateur téléchargeable en local faisant partie de la suite d'outils ProteoSafe. L'interface MSConvert est illustrée dans la Figure 31.

Les données spectrales peuvent être chargées sous différents formats, venant de différents instruments, et le format de sortie est sélectionné (« mzML », « mzXML », ou « MGF » parmi d'autres). Il est possible d'appliquer des opérations sur les spectres traités. Par exemple, la fonction la plus utilisée est celle du « *peak picking* », ou détection du signal. Un algorithme transforme les pics d'un spectre de masse. Ils ont dans un premier temps une allure Gaussienne et l'apex se situe au niveau du rapport m/z .

Ce sont des profils, et plus la résolution d'un analyseur est grande, plus les profils sont fins et les rapport m/z proches sont différenciés.

L'algorithme permet ensuite de convertir les signaux en profils dans un format en « bâtons », à une dimension, caractérisés par une valeur de m/z en abscisse et une intensité sur l'axe des ordonnées. On appelle ce format de pic un « centroïde ». Cette valeur de m/z en abscisse correspond à la valeur m/z mesurée de l'ion considéré. Ces spectres de masse centroïdés permettent l'obtention de données plus légères en terme de mémoire dans le système informatique (un bâton est plus léger à enregistrer qu'une multitude de points formant une enveloppe). Il est également favorable de centroïder les spectres de masse pour la création de réseaux moléculaires par GNPS.

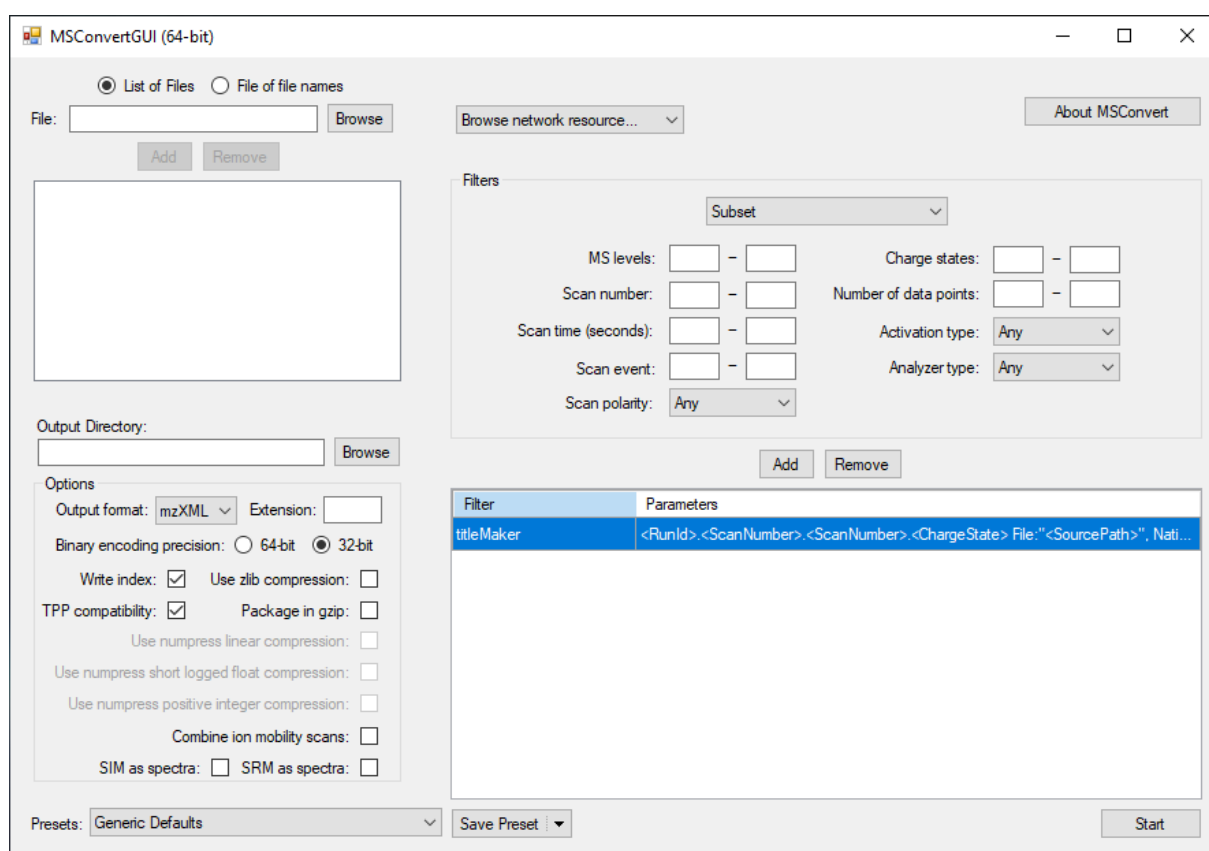


Figure 31 : interface de l'application MSConvert servant à convertir des formats de données spectrales variés en formats libres.

Cette étape de centroïde est donc essentielle, et l'algorithme employé pour ce traitement l'est tout autant. Nous nous sommes cependant rendu compte que le centroïde des données au format Waters (.raw) par l'application MSConvert créait un décalage dans la mesure du rapport m/z , et donc dans la mesure de masse. Un tel décalage en masse est forcément délétère pour la qualité des analyses car il impacte toutes les étapes de traitement de données ainsi que d'interprétation des résultats avec des

erreurs d'attribution de formules moléculaires. Nous avons comparé le rapport m/z d'une molécule connue avant centroïde et après centroïde avec la masse théorique du composé en question. Nous avons déterminé que le décalage en masse était créé quel que soit l'algorithme sélectionné. Il était donc nécessaire de modifier le processus de pré-traitement de données en trouvant un moyen de convertir les données sans créer de décalage ou encore un moyen de centroïder les données spectrales avant une conversion par MSConvert.

b. Centroïde des pics des spectres de masse acquis par MassLynx

Le logiciel utilisé pour piloter le spectromètre de masse de l'instrumentation UHPLC-MS/MS du laboratoire, MassLynx, est également utilisable pour la visualisation et un traitement basique des données. Le module « *Automatic peak detection* » centroïde les données spectrales directement sous MassLynx. Deux spectres, l'un centroïdé et l'autre non, sont fournis en guise d'illustration en Figure 32. Ce module ne crée pas de décalage en masse. Cette étape supplémentaire a permis de centroïder les données en amont de la conversion. Les données sont allégées et les pics sont détectés selon le rapport m/z le plus juste possible. Les spectres de masse sont ensuite convertis via MSConvert du format constructeur vers un format libre avant de les utiliser pour la création de réseaux moléculaires par GNPS.

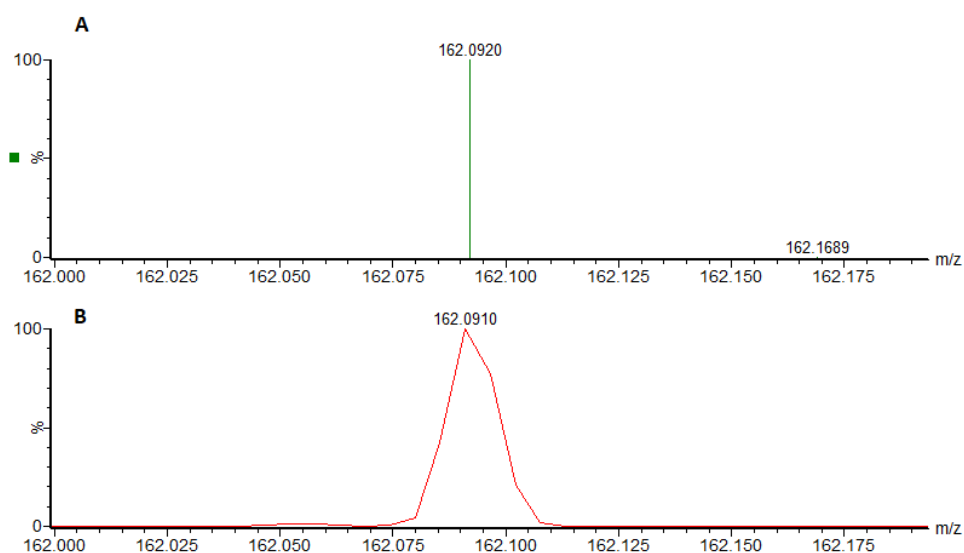


Figure 32 : exemple de spectre centroïdé sous MassLynx. A : spectre centroïdé grâce au module « *Automatic peak detection* ». B : spectre en mode profil. Issu d'une analyse d'un extrait DCM/AE de *S. cerevisiae* en mode d'ionisation positif.

2. Construction des réseaux

Des réseaux moléculaires furent construits à partir des analyses par UHPLC-MS/MS d'extraits de levures ayant subi le protocole d'extraction et de préparation d'échantillon optimisé. Le workflow classique a été sélectionné, avec une sélection des groupes pour séparer et comparer les analyses entre elles. Aucun pré-traitement des données spectrales n'a été intégré au procédé d'analyse de données à ce stade, les réseaux furent construits directement après conversion des données.

La visualisation des réseaux moléculaires créés montre les composés fragmentés sous forme de nœuds, ceux-ci étant connectés entre eux en fonction de leurs similitudes structurales. Plus les spectres MS/MS sont semblables et plus le *cosine score* se rapproche de 1. Les réseaux construits ne présentent qu'assez peu de composés qui se connectent entre eux. On observe néanmoins des annotations fournies par le système de comparaison avec les bases de données spectrales de GNPS. Une capture des différents nœuds résultant de l'une de ces analyses est présentée en Figure 33. Les réseaux sont visualisés sous Cytoscape, un logiciel avec de nombreuses options de paramétrage. Des analyses d'un extrait de culture de levure avec celles d'un extrait de milieu de culture sont comparées.

Le code couleur permet d'identifier l'échantillon de provenance du composé correspondant au nœud considéré. Cette configuration permet d'envisager les avantages d'un système regroupant les données entre elles pour les traiter, les analyser et/ou les comparer entre elles. Elle pointe cependant également les limites du traitement actuel par le workflow classique de GNPS.

En effet, aucun ion de l'extrait de levure n'est en commun avec le milieu de culture. De plus, bien que les composés soient séparés selon leur échantillon d'origine (grâce au code couleur), il reste possible dans cette situation qu'un composé A soit présent dans les deux extraits comparés mais qu'il apparaisse sous la forme de deux nœuds différents, un nœud par extrait d'origine. Il existe aussi le cas inverse où des isomères peuvent être considérés comme un seul et même composé bien qu'ils éluent à des temps de rétention différents, le workflow actuel ne prenant pas ce descripteur en compte à l'heure où ces tests ont été conduits.

Ces écueils peuvent nuire à l'interprétation des résultats, néanmoins les étapes de pré-traitement par MZMine2 sont déjà envisagées à ce stade de la thèse, et pourraient remédier à ces problèmes [179].

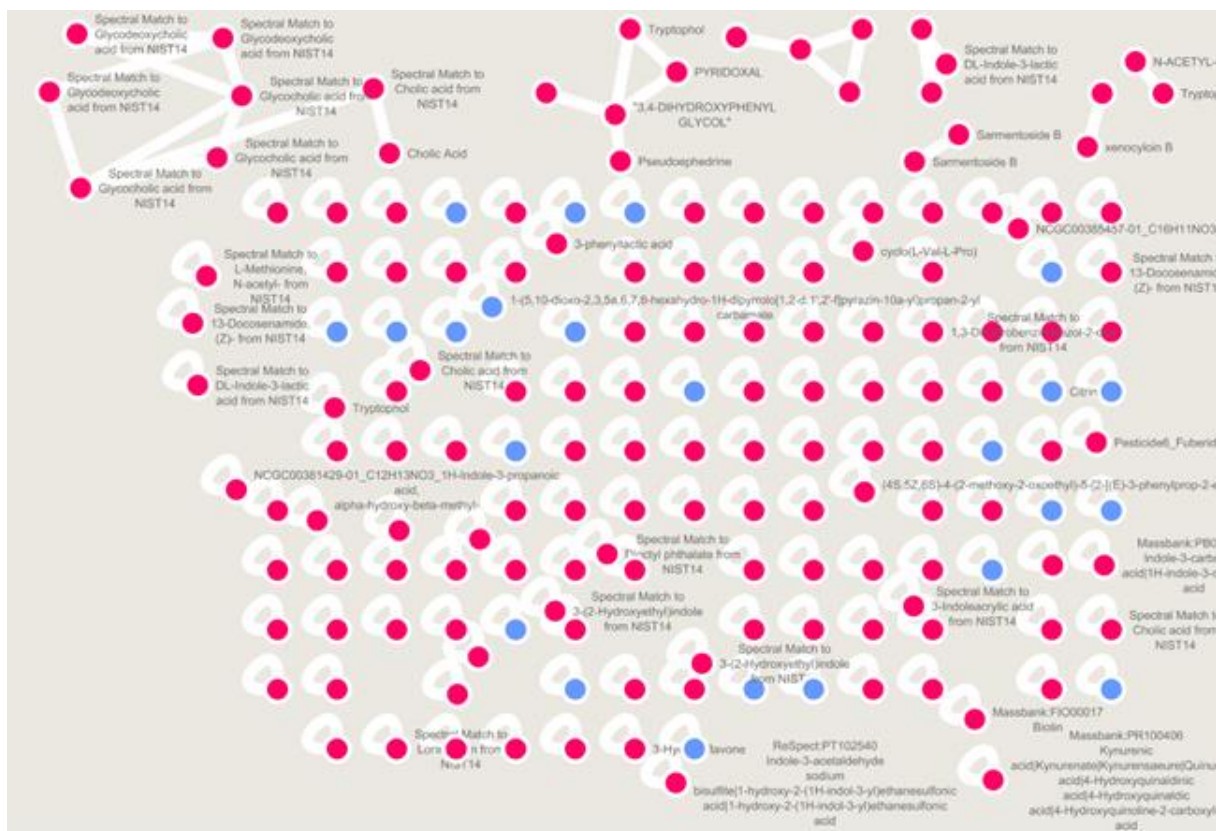


Figure 33 : résultats de traitement par GNPS d'analyse UHPLC-MS/MS d'un extrait de culture de *S. cerevisiae* (nœuds rouges) avec celle d'un extrait de milieu de culture (bleu).

II. Application et optimisation des outils de traitement de données

L'utilisation classique de l'outil de réseaux moléculaires a révélé quelques inconvénients et limites dans sa mise en œuvre. Le pré-traitement des données en amont de la construction de réseaux moléculaires est une solution pour pallier à ces problèmes. Ces outils permettent l'emploi d'algorithmes afin de détecter les différentes masses et de convertir les données en *features* ainsi que de les trier et les aligner. L'utilisateur peut alors limiter les doublons, prendre en compte le temps de rétention des composés pour éviter la confusion entre des isomères mais aussi limiter la création de nœuds correspondant à des signaux du bruit de fond.

1. MZMine2 pour le prétraitement des données d'analyse UHPLC-MS/MS

L'outil MZMine2 est spécialisé dans la visualisation et le traitement de données LC-MS. Il possède l'avantage de permettre l'import de données Waters au format « .raw » et de se passer de l'étape de conversion par MSConvert. Nous conservons néanmoins la fonction de centroïde des données

spectrales par MassLynx car celle-ci s'est révélée plus performante pour la détection de pics et pour garantir la précision de mesure de masse.

Durant le traitement, les données spectrales sont transformées en *features*. Chacune est caractérisée par un rapport m/z , un temps de rétention, un EIC du rapport m/z correspondant ainsi que des spectres MS/MS associés. Des modules spécifiques créent des listes de *features*, et optimisent ces listes en dissociant les isomères ou en supprimant les signaux du bruit de fond. De plus, l'interface permet la prévisualisation des graphiques traités, chromatogrammes comme spectres de masse, pour chaque étape de traitement.

Le logiciel est également défini comme très utile pour le pré-traitement des données spectrales issues d'expériences LC-MS en vue de construire des réseaux moléculaires, en raison de la prise en compte du temps de rétention dans la création des tables de *features* ce qui permet de se passer de l'algorithme MSCluster de GNPS [179].

2. Comparaison de l'outil UNIFI avec MZMine2 pour la détection des pics

L'outil UNIFI est également évoqué pour le pré-traitement des données avant export pour construction de réseaux moléculaires sur GNPS. Il s'agit d'un logiciel du constructeur Waters spécialisé dans la visualisation et le traitement automatisé des données acquises au cours d'expériences LC-MS. L'objectif de l'utilisation du logiciel UNIFI est d'exploiter ses fonctionnalités pour la détection des masses au sein des données spectrales, avant de les exporter en format libre.

Le logiciel Unifi applique la correction liée à la mesure de la *lockmass* (l'étalon semi-interne ionisé en parallèle de l'échantillon), tandis que MZMine2 n'applique pas cette correction. Cela signifie que les données comparées doivent provenir d'analyses différentes, les unes avec correction et les autres sans.

Des analyses supplémentaires ont été réalisées sur les extraits de levures. Un outil Microsoft Office Excel, sous forme de macro programmée à partir de l'interface Visual Basic for Applications (VBA), a été créé afin de comparer deux listes de pics entre elles. La comparaison de listes d'ions détectés par MZMine2 d'un côté et détectés par UNIFI de l'autre montre que les deux procédés donnent des résultats très similaires (82 % des masses retrouvées sont communes). La plupart des composés détectés exclusivement par le logiciel UNIFI et manquant à ceux détectés par MZMine2 correspondent à des ions provenant du bruit de fond.

En conclusion, le logiciel UNIFI pourrait être une alternative très utile pour la détection de masses et une identification grâce à la comparaison avec les bases de données spectrales accessibles par ce logiciel. Cependant, les listes de pics obtenues étant similaires, l'outil MZMine2 est sélectionné pour la suite du traitement des données UHPLC-MS/MS issues d'analyses d'extraits de levure. Ce choix est confirmé par l'existence de nombreux modules inhérents à MZMine2 permettant d'approfondir le traitement de données après l'étape de détection des masses.

3. Optimisation du prétraitement de données spectrales

Une fois le choix fait du logiciel MZMine2 pour le pré-traitement des données de LC-MS, les différents modules destinés à être utilisés ont été sélectionnés et les paramètres optimisés pour être adaptés au mieux au procédé de création de réseaux et à nos attentes.

a. Import des données et détection des masses

Les données UHPLC-MS/MS sont dans un premier temps importées dans MZMine2, toujours en format « .raw » mais déjà centroïdées par l'outil adéquat de MassLynx. Le module de détection de masses permet de détecter les pics des spectres MS1 et des données de fragmentation MS2. L'algorithme sélectionné se base sur l'intensité de signal. Les pics suffisamment intenses sont sélectionnés et les autres sont écartés, considérés comme appartenant au bruit de fond. Un seuil d'intensité très bas est choisi ; c'est très permissif, cependant les ions du bruit de fond conservés à cette étape seront éliminés aux étapes suivantes. Il est aussi possible d'observer le bruit de fond sur les spectres de masse acquis pour choisir un seuil.

b. Construction des chromatogrammes, ou listes de features

Les listes de masses créées à l'étape précédente sont ensuite utilisées par le module de construction de chromatogrammes pour générer des listes de *features*. Le module en question connecte les valeurs de rapport m/z des ions détectés lorsqu'elles sont retrouvées dans des scans consécutifs, puis crée un EIC pour ces ions. Cela permet notamment d'éliminer les artefacts si un pic n'apparaît que sur un scan, chose peu probable avec un couplage LC-MS. Les listes de *features* sont créées et représentent les données expérimentales qui seront manipulées par les prochains modules. Elles correspondent en réalité à un regroupement d'EIC où chaque *feature* est caractérisée par un rapport m/z , un temps de rétention, un chromatogramme d'ion extrait ainsi que des données MS/MS correspondantes. Les isomères ne sont pas encore séparés à cette étape, il est donc possible de retrouver des EIC avec

plusieurs pics correspondant au même m/z mais à différents temps de rétention. Encore une fois, différents algorithmes sont disponibles pour la construction de chromatogrammes et différents paramètres sont à régler, comme le nombre de scans consécutifs minimal pour considérer un pic comme une molécule détectée, le seuil d'intensité de ce pic ou une tolérance de différence de rapport m/z entre les pics des différents scans considérés.

c. Déconvolution des chromatogrammes, et résolution des isomères

L'étape suivante est probablement l'une des étapes les plus importantes et délicates du processus de traitement des données par MZMine2. La déconvolution des chromatogrammes s'effectue sur une liste entière et affecte toutes les *features*, qu'elle contient. Elle permet le raffinement des données, c'est-à-dire que chaque EIC est analysé par l'algorithme sélectionné selon les paramètres indiqués. La forme des pics chromatographiques est évaluée pour éliminer les pics aberrants, comme c'est le cas lors d'un relargage en fin de gradient. Les EIC comprenant plusieurs pics correspondant à des isomères sont séparés, donnant ainsi une *feature* par isomère, avec le même rapport m/z mais des temps de rétention différents. Les pics appartenant au bruit de fond retenus à l'étape précédente sont écartés grâce à ce traitement.

Différents algorithmes existent, mais c'est à l'heure actuelle la déconvolution par l'algorithme *ADAP Wavelet* qui est la plus utile vis-à-vis de nos jeux de données UHPLC-MS [180]. On retrouve parmi les paramètres des seuils de signal sur bruit, des seuils d'intensité, des gammes de temps de rétention pour considérer un signal comme pic chromatographique mais aussi la gamme de temps de rétention qui permet de définir la notion de *wavelet*. Pour faire simple, il s'agit des opérations qui permettent de définir un pic chromatographique selon la forme du début et de la fin d'un signal. Ce paramètre peut permettre par exemple la résolution de pics chromatographiques aux temps de rétention très proches et la séparation de ceux-ci en deux *features* distinctes. Ce paramètre crucial est délicat à régler et nécessite une optimisation minutieuse à chaque jeu de données. La déconvolution est également l'étape qui confirme l'association des différentes *features* avec les spectres MS/MS correspondants. L'ion décrit comme *feature* est le précurseur sélectionné pour la fragmentation qui a donné ces scans MS/MS.

d. Suppression des isotopes et autres doublons superflus

Le module de groupement des isotopes permet de supprimer les *features* superflues qui correspondent à des ions isotopiques contenant un ou plusieurs ^{13}C d'une molécule déjà détectée sous sa forme sans isotopes dans une autre *feature*.

Il est également possible d'utiliser divers modules de filtrage sur les listes de *features*. L'une de ces options permet par exemple de filtrer les doublons, quand deux *features* partagent toutes leurs caractéristiques, rapport m/z , temps de rétention, et correspondent en fait à la même molécule détectée. Cela permet d'alléger les listes en éliminant les variables superflues.

Une autre option disponible est le filtrage des éléments qui ne sont reliés à aucune donnée spectrale de MS/MS. Elle permet de retirer les variables en question afin de ne garder que les *features* disposant de spectres MS/MS. Cela peut se révéler utile pour un traitement par GNPS, car cet outil utilise les données de fragmentation aussi bien pour la construction de réseaux avec les composés de nos analyses mais aussi pour la comparaison des données expérimentales avec les bases de données spectrales. Cependant, nous ferons un usage plus approfondi des listes de *features* et nous ne souhaitons pas négliger des signaux même si les composés en question n'auront pas de spectres MS/MS attribués. Ce module restera donc optionnel dans notre cas.

e. Alignement des échantillons

Une autre étape optionnelle mais extrêmement intéressante pour le pré-traitement des données en vue d'utiliser les réseaux moléculaires pour la suite du projet est l'alignement des échantillons. Plusieurs listes de *features* correspondant à des échantillons différents peuvent être alignées entre elles. Le rapport m/z et le temps de rétention sont utilisés pour fusionner les variables communes et les intensités relatives à chaque échantillon sont affichées.

L'alignement est très utile pour observer l'occurrence des molécules au sein des échantillons et comparer les composés en commun, mais également dans notre cas pour écarter les composés provenant du milieu de culture et retrouvés aussi dans les échantillons d'extraits de levure. Des réseaux moléculaires créés avec des données alignées et le workflow classique de GNPS perdent l'information des groupes et de l'échantillon d'origine des molécules, mais l'étape d'alignement se révèle essentielle dans le workflow spécialisé de GNPS nommé « *feature-based molecular networking* » (ou FBMN ; détaillé ci-après).

Il est possible d'utiliser un module optionnel de *gap-filling* après l'alignement d'échantillons afin de croiser de nouveau les données et de retrouver les composés des différentes analyses qui auraient pu échapper à l'algorithme de l'alignement.

Ces étapes principales du traitement de données spectrales par MZMine2 et sélectionnées pour faire partie de notre workflow analytique sont résumées dans la Figure 34.

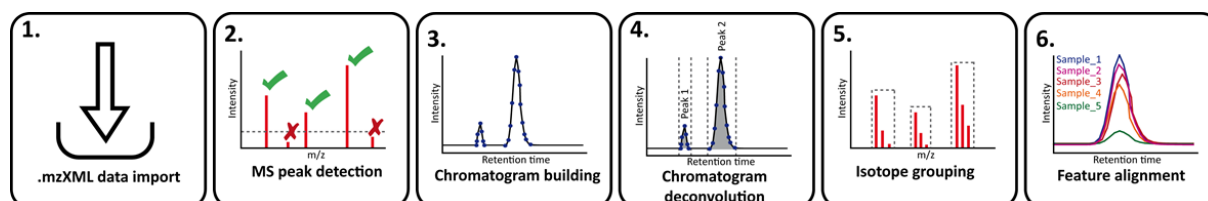


Figure 34 : étapes principales de traitement de données LC-MS par le logiciel MZMine2. Issu de la documentation en ligne de la plateforme GNPS.

f. Attribution de formules moléculaires et détection des fragments en source et adduits

MZMine2 dispose également de modules permettant d'ajouter des informations aux éléments des listes de *features*. Nous avons donc intégré dans le processus de traitement de données l'attribution automatisée de formules moléculaires aux différents signaux des échantillons. Le module en question est paramétré pour respecter les sept règles d'or de l'attribution de formule moléculaire.

Il est aussi possible d'annoter automatiquement certains composés après comparaison des du rapport m/z avec les entrées disponibles sur certaines bases de données. Il fut choisi d'intégrer la recherche automatisée sur la base de données YMDB, bien que les résultats d'une telle attribution d'annotation doivent être soigneusement vérifiés car seul le rapport m/z est utilisé pour celle-ci. Une annotation en faux positif est donc fortement possible car l'attribution ne prend pas en compte le spectre de fragmentation.

Il est aussi possible de rechercher les fragments en source (fragments qui se forment dans la source et apparaissent donc en MS1 plutôt qu'en MS2), les adduits ainsi que les complexes (tels que les dimères ou trimères de molécules qui peuvent aussi se former dans la source d'ionisation). Nous appliquons ces étapes afin de filtrer les signaux non informatifs et les écarter des étapes d'interprétation manuelle des résultats.

À la suite de toutes ces étapes de traitement, les données sont enfin exportées dans un format adéquat à l'utilisation par GNPS.

4. Traitement de données : *feature-based molecular networking*

a. Principe du FBMN

GNPS s'est vu amélioré, après le début de ce projet de thèse, via l'ajout sur la plateforme d'une nouvelle méthode de construction de réseaux. Le *feature-based molecular networking*, abrégé en FBMN ou traduit littéralement par *réseaux moléculaires basés sur les variables*, est une méthode développée récemment pour utilisation sur l'interface GNPS [181].

Ce workflow se focalise sur les différentes variables chromatographiques détectées lors d'étapes de pré-traitement de données d'expériences LC-MS ainsi que sur l'alignement de ces données. Les données doivent être traitées en amont via des logiciels comme MZMine2 ou XCMS.

La prise en compte de l'alignement des données permet la résolution des isomères. Les doublons sont également évités car le temps de rétention est considéré. Elle permet également une analyse semi-quantitative. Bien qu'elle ne puisse pas être qualifiée pleinement de quantitative sans un ajout d'étalon interne dans nos échantillons, la prise en compte des intensités des signaux permet de jauger l'abondance des composés au sein des échantillons alignés. Le workflow garde toutes les fonctionnalités classiques de construction de réseaux moléculaires de GNPS, en reliant les molécules entre elles selon leurs similitudes de fragmentation et en comparant les données MS/MS avec des bibliothèques spectrales pour attribuer des annotations aux composés. Le FBMN s'intègre donc dans le processus analytique de notre projet comme une amélioration de notre étape de création de réseaux moléculaires, en s'imbriquant juste après le pré-traitement des données UHPLC-MS/MS par MZMine2.

b. Application du FBMN pour améliorer notre traitement de données

L'application du FBMN a nécessité des étapes d'optimisation logicielle puis d'optimisation de la représentation sous Cytoscape pour obtenir des réseaux informatifs. Les étapes de pré-traitement sur MZMine2 décrites dans la rubrique précédente ont été optimisées selon les paramètres retrouvés dans la littérature et dans la documentation en ligne de GNPS (*FBMN with MZMine*) [179].

L'étape d'export des données est ici différente d'un export de données pour le workflow classique. Plusieurs fichiers doivent être téléchargés sur la plateforme GNPS. Le premier contient les données spectrales, le plus souvent au format « MGF ». Un deuxième fichier au format « CSV » fournit les données semi-quantitatives provenant des tableaux MZMine2, c'est-à-dire les valeurs d'intensité retrouvées pour chaque *feature* dans les échantillons alignés. Enfin, le fichier de métadonnées permet

d'attribuer des groupes et autres caractéristiques selon le choix de l'utilisateur et relie ces informations aux différents échantillons. Cela peut être le nom de l'échantillon, le nom de l'espèce dont il provient ou le solvant ayant servi à l'extraction. Les paramètres du FBMN ont aussi été optimisés pour s'adapter à nos données spectrales, bien que peu de changements ait été opérés par rapport aux paramètres indiqués par défaut lors de la création d'une séquence.

c. Optimisation de la représentation graphique adaptée au FBMN

Concernant la visualisation sous Cytoscape, l'apparence des nœuds choisie correspond à un cercle coloré contenant un diagramme circulaire dont la couleur des sections correspond à l'échantillon d'origine du composé. Cela permet de regrouper les composés retrouvés dans plusieurs échantillons alignés. La présence et l'abondance des dites molécules dans les différents extraits sont comparées. Le code couleur permet aussi d'écarter rapidement les composés non exclusifs aux levures, en attribuant aux molécules présentes dans le milieu de culture ou dans les blancs analytiques une couleur caractéristique. Cette optimisation graphique de l'interface de visualisation Cytoscape est illustrée par les exemples présentés en Figure 35.

Les exemples de réseaux démontrent l'intérêt de la méthode FBMN pour l'investigation des résultats d'analyses UHPLC-MS/MS de substances naturelles. Le réseau du tryptophol montre un cas simple : les composés ont été annotés (soit par annotation automatique par GNPS, soit par investigation manuelle ; la notion d'investigation manuelle sera détaillée dans une autre rubrique). Le code couleur indique la provenance des molécules, entre extrait DCM/AE, extrait M ou encore milieu de culture pour le *Tetrahydroharman-beta-carboline-3-carboxylic acid*.

Le réseau de la méthionine présente un cas similaire. La méthionine est présente dans le milieu de culture, ce qui semble cohérent pour un acide aminé essentiel dans un milieu riche. Connaître sa position dans les réseaux et les connexions avec d'autres nœuds aide à annoter les dérivés de la méthionine qui sont exclusifs aux extraits de levure. Enfin, l'exemple du réseau de composés liés à l'acide cholique montre également que certains réseaux connectent des composés présents dans le milieu comme dans les extraits. Ces réseaux restent informatifs bien qu'ils seront écartés des processus d'investigation manuelle car nous nous concentrons sur les métabolites exclusifs aux levures étudiées.

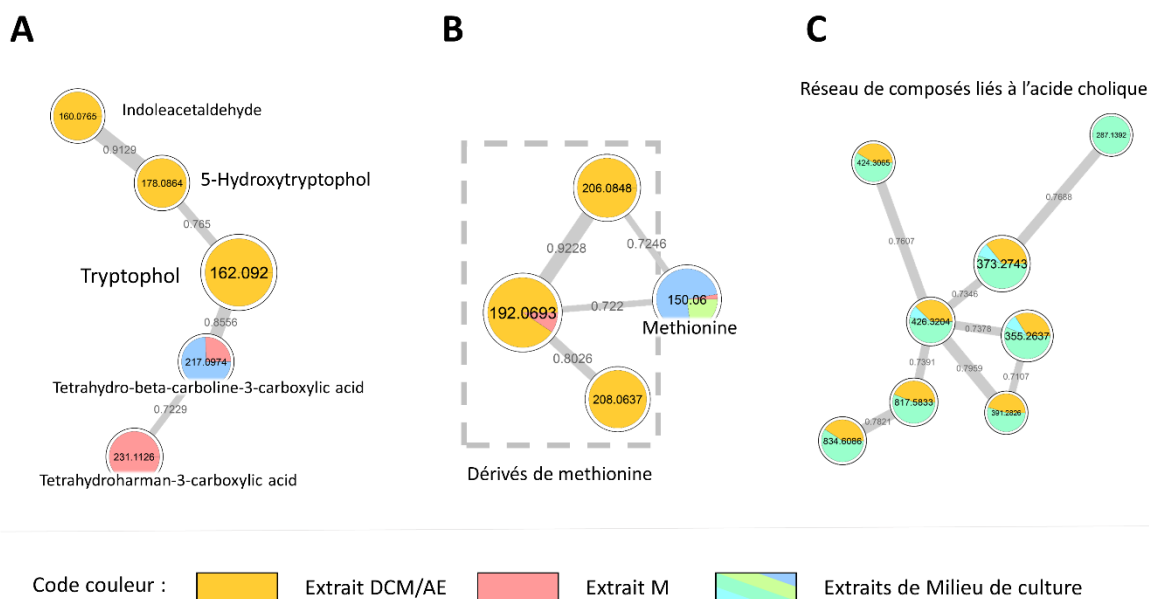


Figure 35 : extraits de réseaux moléculaires créés à partir d'analyses UHPLC-MS/MS d'extraits DCM/AE et M de *S. cerevisiae* (analyses en mode d'ionisation positif). Les réseaux sont construits grâce au workflow FBMN et visualisés sous le logiciel Cytoscape. Le code couleur de l'échantillon d'origine est indiqué. A : réseau de composés connectés au tryptophol. B : réseau de la méthionine et de quelques dérivés. C : réseau de composés en lien avec l'acide cholique.

III. Finalisation de la méthode

A ce stade de la thèse, les principales étapes de protocole d'extraction des métabolites de cultures de levures ainsi que d'analyse des échantillons par UHPLC-MS/MS sont mises en place, tout comme les phases de pré-traitement et de traitement automatisé des données pour l'accélération et l'amélioration de l'interprétation des résultats analytiques. Il reste cependant certains points, du côté de l'extraction et de la préparation d'échantillon, et du côté analytique, qui doivent être vérifiés et certaines pistes méritant d'être explorées avant de qualifier la routine analytique de « définitive ».

1. Comparaison des différents extraits

Le protocole d'extraction est basé sur l'ajout de solvant organique sur les cultures cellulaires. Le mélange de dichlorométhane et d'acétate d'éthyle et le méthanol permettent de couvrir une large gamme de polarité de métabolites extraits. Il était néanmoins nécessaire de vérifier le contenu de

l'extrait méthanolique, car la présence de nombreux doublons et l'absence de composés exclusifs rendrait ce dernier chronophage et inutile.

a. Comparaison des extraits réalisés avec différents solvants d'extraction

Les résultats des analyses des extraits DCM/AE et M de *S. cerevisiae* ont suivi notre procédé de traitement de données. Les listes de *features* de MZMine2 et les réseaux ont été comparés entre eux pour déterminer si l'extrait DCM/AE et l'extrait M présentent des composés en commun et des composés exclusifs. Le premier constat est que l'extrait DCM/AE est plus riche en petites molécules que l'extrait M. On retrouve globalement deux fois plus de *features* pour l'extrait DCM/AE, le nombre de nœuds dans les réseaux moléculaires est plus important tout comme le nombre de composés annotés par le procédé d'annotation de GNPS. Ce constat semble néanmoins très cohérent car l'extrait DCM/AE est la première extraction opérée sur les cultures de levures, donc le mélange de polarité moyenne extrait le plus grand nombre de métabolites disponibles.

L'extrait DCM/AE et l'extrait M présentent certains signaux en commun, ceux-ci partageant le même *m/z* et le même temps de rétention. Cela paraît également cohérent car certaines molécules sont présentes en abondance dans les levures. Elles sont extraites et analysées de la même manière et sont donc définies par des descripteurs similaires. En dehors de ces composés en commun, des composés exclusifs sont retrouvés dans l'extrait M, tels que certains acides organiques et des sucres. Ceci témoigne de l'intérêt de conserver une seconde extraction pour récupérer les métabolites les plus polaires de nos levures.

b. Comparaison des extraits de levure avec les extraits des milieux de culture

Les extraits et les milieux de culture ont aussi été comparés à ce stade. Les résultats ont montré un certain nombre d'ions en commun entre l'extrait DCM/AE de culture de levure et l'extrait DCM/AE de milieu sans levure (41 % en mode positif et 28 % en mode négatif). En parallèle, l'extrait M de culture de levure semble présenter un nombre plus important de composés en commun avec l'extrait M de milieu de culture seul (65 % en mode positif et 76 % en mode négatif). Nous pouvons en conclure que l'extraction par ajout de méthanol reste très utile car elle permet de récupérer des composés polaires exclusifs aux levures, mais qu'elle présentera globalement moins de molécules exclusives que l'extrait DCM/AE. Celles-ci seront à écarter par alignement des échantillons pendant le traitement par MZMine2.

2. Paramètres de chromatographie liquide

a. Gradient spécifique aux extraits méthanoliques

Le gradient de solvants utilisé initialement pour l'analyse UHPLC-MS/MS est un gradient linéaire commençant par une proportion de 99 % d'eau (additivée de 0,1 % d'acide formique) et de 1 % d'acétonitrile (avec la même proportion d'acide formique). Le gradient progresse ensuite de façon linéaire jusqu'à atteindre 100 % d'acétonitrile avec 1 % d'acide formique, puis passe par une étape de stabilisation avant de revenir à la situation initiale. Ce gradient permet une séparation très correcte des métabolites présents dans les extraits DCM/AE.

Les molécules des extraits M, plus polaires, ont cependant tendance avec ce gradient à éluer à des temps de rétention très bas. Un gradient spécifique aux extraits méthanoliques a été mis en place. Avec les mêmes phases mobiles, le gradient commence avec la même situation initiale et progresse bien plus lentement en première partie, avant d'augmenter rapidement vers les 100 % d'acétonitrile dans la deuxième partie du gradient.

Après observation des chromatogrammes, on constate une meilleure séparation des pics chromatographiques. Les métabolites détectés sont mieux répartis sur les BPC. Aucun nouveau composé n'est mis en évidence. Le nouveau gradient ne modifie que les temps de rétention. L'allure des spectres de masse MS1 comme MS2 est inchangée, tout comme les réseaux moléculaires. Par contre, le gradient empêche l'alignement de données entre les extraits DCM/AE séparés par le gradient classique et les extraits M séparés par le gradient spécifique. Cet alignement s'étant révélé très utile pour l'interprétation des données, le gradient classique est finalement conservé dans le procédé analytique UHPLC-MS/MS, et le gradient spécifique aux extraits méthanoliques ne sera pas utilisé.

b. Essai de l'utilisation du méthanol comme phase mobile

Une phase mobile constituée de méthanol (avec ajout d'acide formique à hauteur de 0,1 % du mélange) fut testée pour remplacer la phase mobile composée d'acétonitrile et d'acide formique. Les résultats montrèrent que ce changement de phase mobile n'apporta rien de mieux par rapport à la phase mobile à base d'acétonitrile initialement choisie. La phase mobile au méthanol ne fut donc pas conservée pour le protocole analytique du projet.

3. Modification du protocole d'extraction pour regrouper les extraits

Toujours dans un souci d'optimisation des étapes du procédé analytique, il fut essayé de regrouper les extraits réalisés avec le mélange dichlorométhane / acétate d'éthyle d'un côté et avec le méthanol de l'autre.

En pratique, 30 mL de méthanol sont ajoutés sur la culture de levures dans le flacon contenant déjà le mélange dichlorométhane / acétate d'éthyle. L'échantillon est soumis au traitement par ultrasons, filtré et concentré par évaporation sous vide. L'évaporation s'est révélée très difficile, notamment à cause du méthanol qui arrache une grande partie de la gélose constituant le milieu de culture. Un extrait de 261 mg a été récupéré, puis solubilisé dans du méthanol, filtré et dilué pour analyse.

Les résultats de l'analyse de ce nouvel extrait DCM/AE/M sont comparés avec ceux des analyses des extraits DCM/AE et M analysés séparément. Le chromatogramme du nouvel extrait correspond globalement à l'addition des chromatogrammes des extraits DCM/AE et M. Beaucoup de composés sont retrouvés en commun. Certains pics chromatographiques ne sont cependant pas retrouvés, et restent exclusifs aux extraits analysés séparément. Pour cette raison, cette nouvelle méthode d'extraction ne sera pas conservée pour la suite de la thèse. De plus, il est nécessaire de rappeler que les extraits DCM/AE secs ont une masse comprise entre 10 mg et 30 mg, tandis que les extraits M secs ont une masse proche des 200 mg. Cela confirme une sorte d'incompatibilité entre les deux types d'extraits, car les mélanger crée une énorme différence de concentration au détriment des métabolites extraits par le DCM/AE qui montrent la plus grande richesse.

4. Analyse par GC-MS de l'extrait apolaire

Des analyses par chromatographie en phase gazeuse couplée à la spectrométrie de masse ont été réalisées sur un extrait DCM/AE d'une culture de *S. cerevisiae*. Cet extrait a été choisi pour sa polarité basse à moyenne. Les métabolites présents dans cet échantillon sont plus susceptibles d'être adaptés à la GC-MS que les molécules polaires contenues dans les extraits M. L'extraction des métabolites de la culture sélectionnée a été faite selon le même protocole que pour les analyses UHPLC-MS/MS.

Les analyses ont été conduites sur un instrument de GC-MS du constructeur JEOL équipé d'une source à impact électronique (EI) et d'un analyseur à temps de vol (TOF). Le chromatogramme de pic de base de l'analyse de l'extrait DCM/AE de *S. cerevisiae* est présenté en Figure 36 en guise d'illustration. Les spectres intégrés au chromatogramme acquis par GC-EI-TOF sont analysés manuellement et comparés

avec la base de données spectrale NIST. Il s'agit d'une méthode déjà appliquée massivement dans le traitement de données GC-MS issues d'analyses de levures dans la littérature [45,122–125].

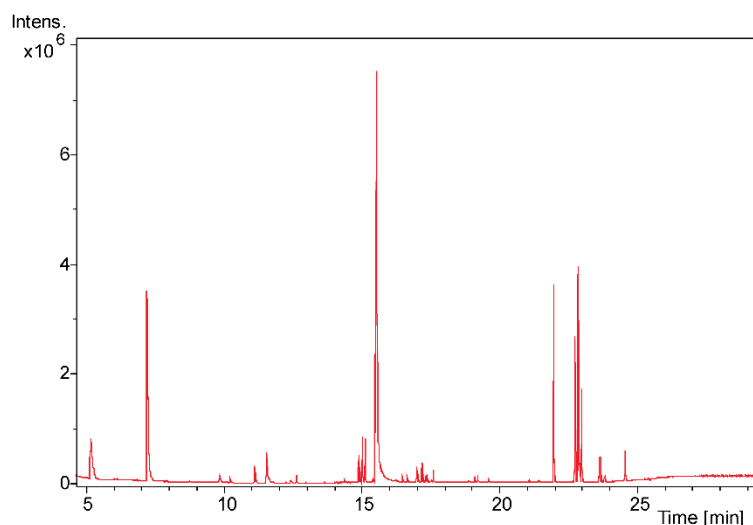
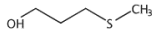
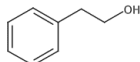
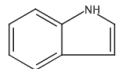
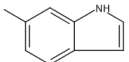
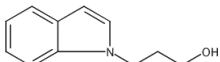
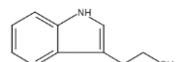
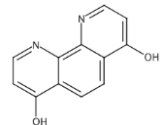
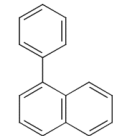
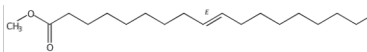


Figure 36 : BPC d'une analyse par GC-EI-TOF d'un extrait DCM/AE d'une culture de *S. cerevisiae*.

Un blanc analytique composé uniquement d'un mélange des solvants d'extraction et de dilution des échantillons a également été analysé par GC-MS. Les résultats de cette analyse montrent que le blanc est peu contaminé. L'analyse de l'échantillon de levure, après dilution, révèle quant à elle un plus grand nombre de composés et l'intensité globale est plus élevée que celle du blanc analytique. Après comparaison des spectres de fragmentation avec la base de données NIST, on retrouve dans notre extrait de levures quelques composés annotés tels que le 2-Phényléthanol ou l'acétate de tryptophol. Le signal n'étant pas optimal, de nouvelles analyses furent conduites à l'aide du même échantillon mais à plus forte concentration (concentration semblable à celle utilisée pour les analyses UHPLC-MS/MS). Un extrait des résultats d'annotation de cette seconde analyse est fourni dans le Tableau 3.

Tableau 3 : extrait des résultats d'analyse par GC-MS d'un extrait DCM/AE d'une culture de *S. cerevisiae*. Acquisition sur un appareil GC-EI-TOF et comparaison des spectres de fragmentation obtenus avec la base de données spectrales NIST.

Temps de rétention	Composé identifié	Formule moléculaire	Masse moléculaire	Pourcentage de corrélation	Structure
5,28	3-(Methylthio)-1-propanol	C ₄ H ₁₀ O S	106,19	95%	
7,3	Benzeneethanol	C ₈ H ₁₀ O	122,16	99%	
9,94	1H-Indole	C ₈ H ₇ N	117,15	97%	
11,2	1H-Indole, 6-methyl-	C ₉ H ₉ N	131,17	97%	
15,09	1H-Indole-1-propanol	C ₁₁ H ₁₃ NO	175,23	89%	
15,62	Tryptophol	C ₁₀ H ₁₁ NO	161,2	72%	
16,55	1,10-Phenanthroline-4,7-diol	C ₁₂ H ₈ N ₂ O ₂	212,2	77%	
16,71	Naphtalene, 1-phenyl	C ₁₆ H ₁₂	204,27	90%	
18,94	9-Octadecenoic acid, methyl ester, (9E)-	C ₁₉ H ₃₆ O ₂	296,49	95%	

Bien que les résultats de l'analyse GC-MS fournissent quelques annotations, une majorité de ces dernières ont déjà été retrouvées grâce aux analyses UHPLC-MS/MS du même extrait. Les autres annotations tirées de la comparaison avec la base de données NIST ont été jugées peu fiables lors de l'interprétation des résultats. En se basant sur ces éléments, ainsi que sur le caractère extrêmement chronophage du traitement non automatisé des données de GC-MS, cette technique ne sera pas conservée dans le workflow analytique final pour la déréplication des souches de levures exotiques.

5. Spectrométrie de masse et problème de fragmentation en source

a. Fragmentation en source, visible dans le signal MS1

Après les premières étapes d'optimisation du procédé d'analyse et les investigations manuelles des résultats obtenus sur des extraits de levures, un problème analytique fut remarqué au niveau des chromatogrammes et spectres de masse. Les listes de *features* générées sur MZMine2 présentent un grand nombre de composés aux rapports m/z différents mais aux temps de rétention identiques. De plus, des composés de rapports m/z plus élevés présentent des ions fragments de plus bas rapport m/z correspondant à ces co-élutions sur leurs spectres MS/MS. Cela signifie que ces molécules de plus grande masse sont également fragmentées dans la source d'ionisation, car certains fragments détectés en MS2 sont aussi détectés en MS1. Ces fragments sont des fragments formés en source. La fragmentation en source peut être due à une tension d'ionisation trop élevée pour les molécules analysées.

Les conséquences de ce problème sont multiples et représentent une perte d'informations au niveau de l'interprétation des données. Premièrement, cela crée de nombreuses *features* qui ne sont pas véritablement des métabolites des levures. De plus, les composés fragmentant en source perdent en intensité, donc le spectre MS2 est moins informatif. Enfin, les fragments en source, détectés en MS1, peuvent eux aussi être sélectionnés pour fragmentation MS/MS et occuper du temps d'analyse au détriment d'autres composés qui pourraient être sélectionnés à la place (pour rappel, les composés sont sélectionnés selon leur intensité, mais sont exclus après une première sélection pour que d'autres composés soient également fragmentés).

b. Visualisation des fragments en source

Avant de trouver des solutions pour limiter le phénomène de fragmentation en source, il convient de trouver une méthode efficace pour jauger la quantité de ces fragments en source dans nos analyses. Premièrement, la comparaison des chromatogrammes permet d'identifier les pics témoins d'une co-élution. Les ions communs entre la MS1 et la MS2 sont et les précurseurs sont déterminés. Un exemple de visualisation manuelle des fragments en source est présenté en Figure 37. Deux ions co-éluent. L'analyse du spectre MS/MS de l'ion au plus grand m/z (ici le 154,05) révèle la détection de l'ion de plus faible m/z (le 136,04). Cet ion de plus grande masse est donc précurseur du second. Il se fragmente dans la source, permettant la détection de l'ion de plus basse masse.

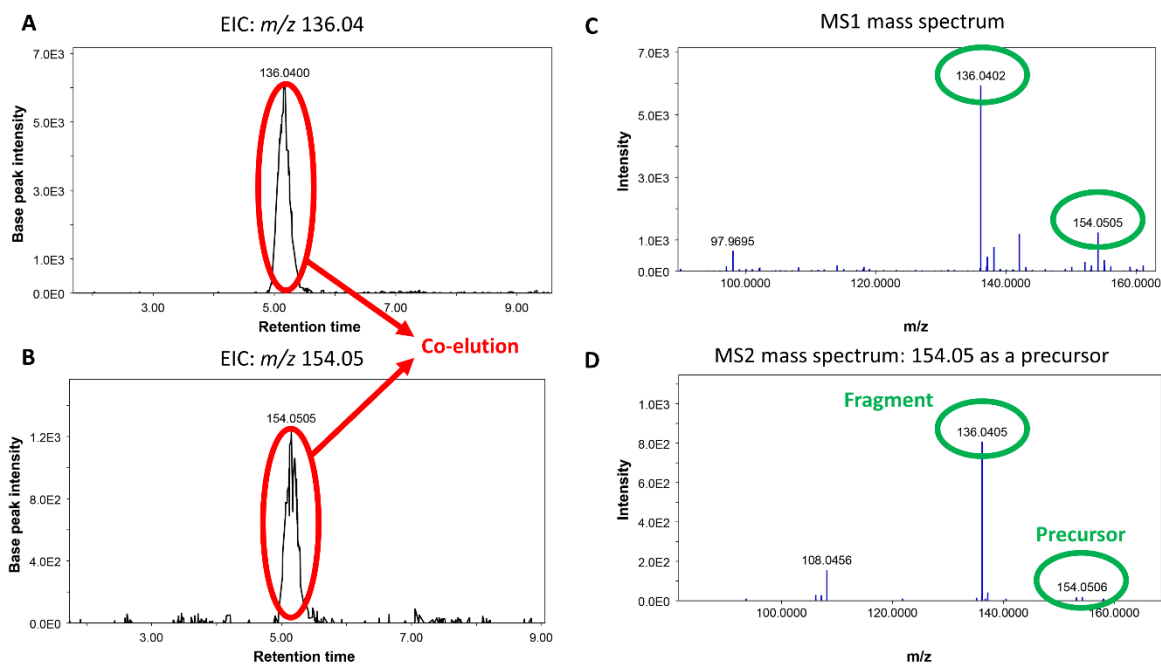


Figure 37 : exemple de visualisation manuelle de fragments en source. A et B : EIC des ions de m/z 136,04 et 154,05 co-éluant aux alentours du temps de rétention 5,20 minutes. C : spectre de masse acquis en MS1 ; les deux ions 136,04 et 154,05 sont visibles. D : spectre de fragmentation de l'ion 154,05 ; l'ion 136,04 est un fragment du 154,05.

L'analyse manuelle des fragments en source étant très chronophage, le module de recherche de fragments de MZMine2 a été utilisé pour déterminer la quantité de signaux en MS1 correspondant à des fragments d'autres molécules de plus grande masse. Pour résumer succinctement le gain de temps, le module de MZMine2 réalise automatiquement en quelques secondes une recherche de fragments ayant pris plusieurs heures manuellement.

c. Paramètres de la source d'ionisation aidant à limiter les fragments en source

Les paramètres sur lesquels nous pouvons jouer pour tenter de limiter la fragmentation en source sont la tension du capillaire de la source ESI, la tension appliquée au cône d'échantillonnage et au cône d'extraction, ainsi que le débit et la température du gaz de désolvation.

Ces cinq paramètres concernant la source d'ionisation de l'équipement de spectrométrie de masse ont été les cibles d'un plan d'expérience réalisé sur le logiciel MODDE. Cet outil permet de créer des variations de paramètres qui sont ensuite testées sur une séquence d'analyses. Chaque paramètre reçoit une valeur minimale et une valeur maximale, et un jeu d'une trentaine d'analyses est créé par le logiciel. Ces analyses présentent toutes une configuration différente de valeurs de paramètres

comprises dans l'intervalle indiqué par l'utilisateur. Un exemple d'intervalles et les unités des paramètres testés sont indiqués dans le Tableau 4.

Tableau 4 : paramètres testés pour la minimisation des fragments en source par plan d'expérience à l'aide de MODDE. Exemple d'intervalles testés pour l'optimisation en mode d'ionisation positif.

Paramètre	Unité	Intervalle	Pas
Tension de capillaire	kV	2 à 3,5	0,5
Tension du cône d'échantillonnage	V	10 à 35	1
Tension du cône d'extraction	V	2 à 5	0,5
Flux de gaz de désolvatation	L/h	400 à 800	20
Température de désolvatation	°C	300 à 500	10

La trentaine d'expériences d'UHPLC-MS/MS a été réalisée sur un même échantillon. Les résultats ont servi à déterminer le pourcentage d'intensité correspondant à des signaux de fragments en source au sein desdites analyses. La proportion de fragments en source dans les analyses réalisées varie de 36 % à 66 % du signal. Ces résultats sont ensuite traités par le logiciel MODDE. La grandeur « intensité des fragments » est indiquée comme paramètre prioritaire à minimiser. Un compromis doit cependant être fait par le logiciel, qui doit minimiser cette variable tout en gardant un bon signal d'intensité. Il détermine ensuite quelle combinaison de paramètres permet de limiter au mieux les fragments en source.

Le facteur identifié comme ayant le plus d'influence sur notre problème de fragments est la tension du cône d'échantillonnage. Loin derrière en termes d'influence, la tension du capillaire, la tension du cône d'extraction et le flux de gaz de désolvatation sont définis comme ayant une influence modérée sur la problématique de fragments en source, tout comme la température de désolvatation. Un résumé de ces résultats est présenté en Figure 38.

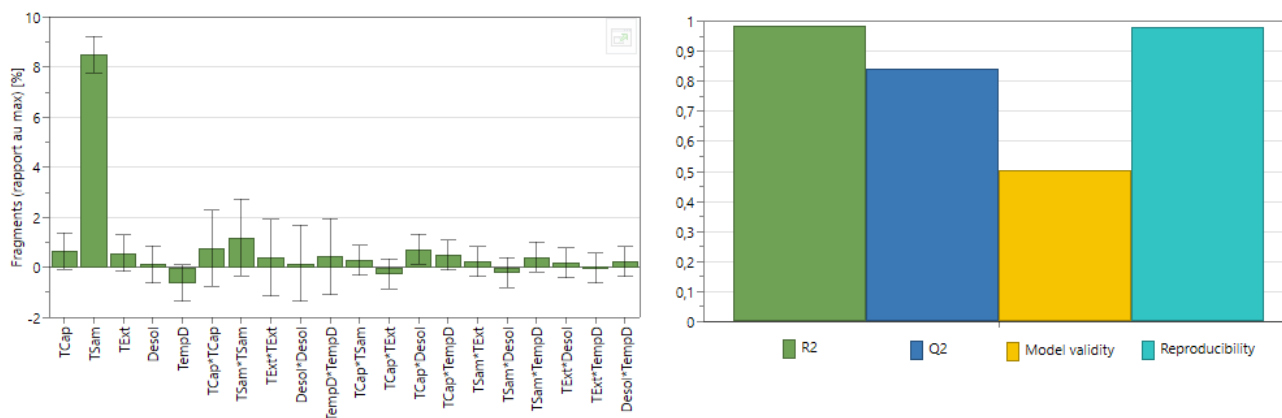


Figure 38 : résumé des résultats d'optimisation par plan d'expérience vis-à-vis du problème de fragments en source. A gauche : impact des paramètres testés, seuls ou combinés entre eux. A droite : résumé des méthodes statistiques d'évaluation appliquées par MODDE. Plus les valeurs sont élevées et meilleure est la confiance dans le modèle.

Après détermination des valeurs optimales, de nouvelles analyses ont été conduites avec lesdits paramètres et ont amené à une diminution des fragments en source : une moyenne de 17 % de fragments est retrouvée en mode d'ionisation positif, et une moyenne de 7 % est retrouvée en mode d'ionisation négatif. Ces paramètres sont ensuite adaptés pour également maximiser le nombre de signaux obtenus, car ils influencent aussi sur le nombre de *features*. Ce sont ces paramètres finaux qui sont conservés dans le protocole analytique du projet.

6. Vérification des triplicats d'analyse de *S. cerevisiae*

La reproductibilité de la méthode analytique mise au point au cours de ce projet de thèse est vérifiée grâce à l'analyse de triplicats biologiques de cultures de la levures *S. cerevisiae*. Plusieurs cultures différentes de la même souche du microorganisme modèle subissent le protocole d'extraction puis les analyses par UHPLC-MS/MS.

Les résultats de ces analyses sont ensuite comparés grâce à un pré-traitement à l'aide du logiciel MZMine2. Cela permet d'homogénéiser les jeux de données en traitant les résultats des expériences de la séquence d'analyses chromatographiques selon les mêmes paramètres. Ces jeux de données spectrales sont ensuite exportés pour la construction de réseaux moléculaires. L'utilisation du FBMN permet de comparer au mieux l'abondance des ions au sein des échantillons alignés.

La variabilité évaluée inclut à la fois la variabilité du processus analytique, liée à l'extraction suivie de l'analyse, et la variabilité biologique, due aux cultures différentes. Une bonne constance des signaux est retrouvée entre les répliques biologiques (respectivement 95 % et 91 % de signaux en commun entre les triplicats des trois cultures en mode d'ionisation positif et en mode d'ionisation négatif). Ainsi, la reproductibilité est satisfaisante et nous a permis de confirmer cette méthode analytique par UHPLC-MS/MS pour la continuité du projet.

IV. Résultats sur *Saccharomyces cerevisiae*

Comme expliqué dans les sections précédentes, le procédé analytique d'exploration d'extraits de souches de levures par UHPLC-MS/MS. Ces étapes mises bout à bout fournissent des résultats bruts sous différentes formes. Les tableaux de *features* générés par le traitement MZMine2 d'échantillons alignés sont exportés sous un format facilement manipulable avec Microsoft Office Excel. Les modules de visualisation des chromatogrammes, des spectres de masse et des données de fragmentation, sous MZMine2 permettent de plonger plus loin dans l'interprétation des données spectrales. Enfin, GNPS fournit des réseaux moléculaires aidant à l'exploration des familles de molécules présentes dans les extraits et présentant les annotations attribuées par l'algorithme de la plateforme californienne.

L'étape suivante dans le cadre de ce projet de thèse est l'investigation manuelle des résultats afin de dresser une cartographie globale des métabolites de levures. En pratique, il s'agit de combiner les informations dont nous disposons et celles que nous pouvons acquérir sur les bases de données en ligne et d'y ajouter notre expertise scientifique afin d'attribuer des annotations aux *features* détectées. Il s'agit d'une étape cruciale du workflow de déréplication qui permet au final l'obtention d'une représentation en réseaux moléculaires et d'un tableau fournissant des informations sur le contenu des microorganismes étudiés. Des exemples de chromatogrammes obtenus pour l'analyse d'extraits de *S. cerevisiae* sont présentés en Figure 39 et en Figure 40.

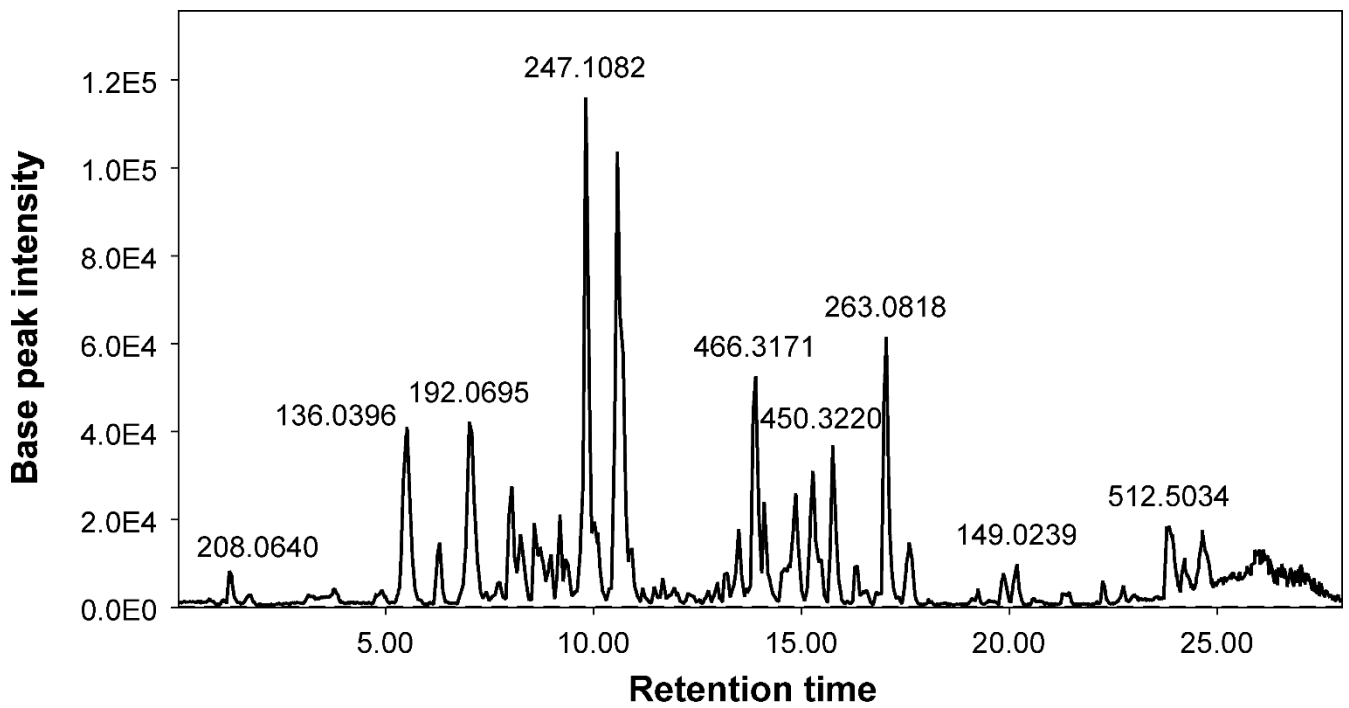


Figure 39 : chromatogramme de pic de base (BPC) d'une analyse UHPLC-MS/MS d'un extrait DCM/AE d'une culture de *S. cerevisiae*.

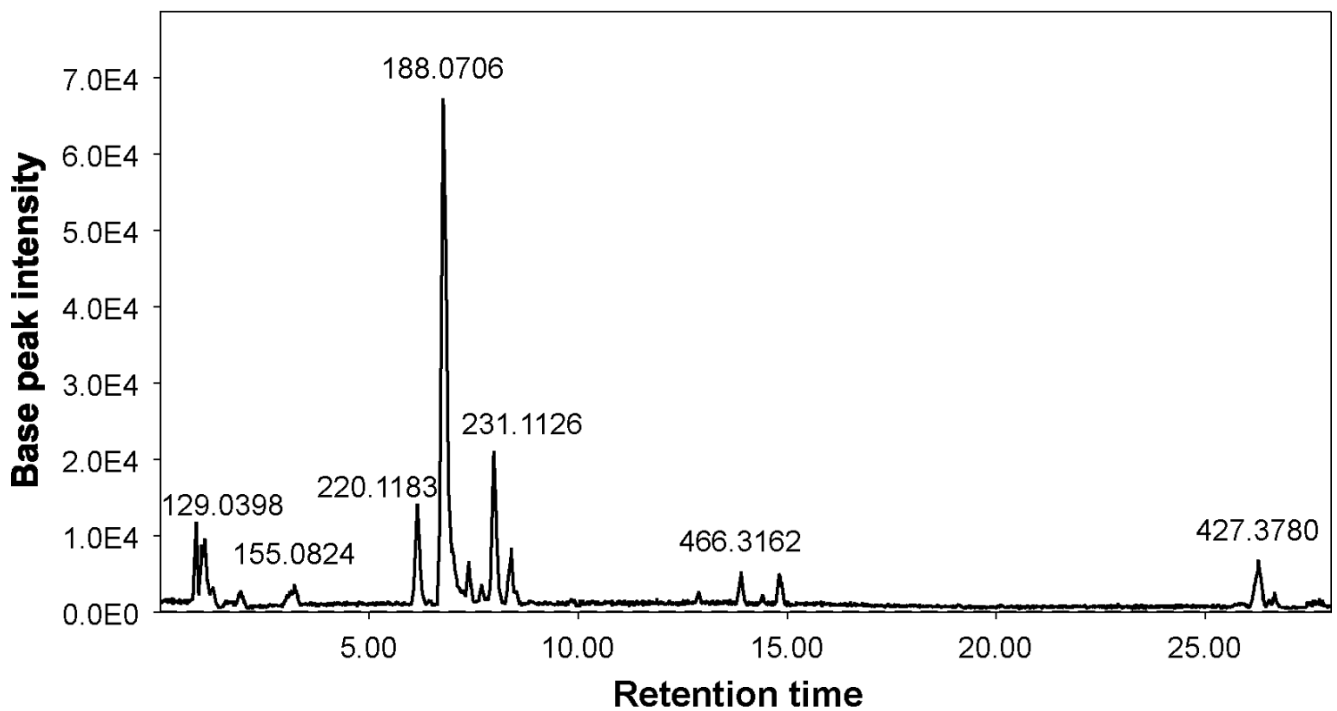


Figure 40 : chromatogramme de pic de base (BPC) d'une analyse UHPLC-MS/MS d'un extrait M d'une culture de *S. cerevisiae*.

1. Annotation des composés

Le processus d'annotation peut sembler abstrait et parfaitement subjectif. L'utilisateur va définir si son annotation est fiable ou non selon le degré de correspondance entre les descripteurs expérimentaux et les informations acquises sur les bases de données [182]. Il existe néanmoins des projets mis en place dans le but de standardiser les résultats de traitement de données de métabolomique. La *Metabolomics Standard Initiative* (MSI), établie en 2007, propose une charte des niveaux d'annotations [183]. De multiples méthodes pour rapporter les résultats d'études de métabolome sont étudiées, en passant par la description précise des méthodes d'extraction et le renseignement des paramètres analytiques des méthodes employées. Concernant le processus d'annotation de composés, ses auteurs proposent un système à niveau mettant en avant la distinction entre identification et annotation explicitée précédemment.

Le *niveau 1* d'annotation est le niveau le plus élevé atteignable avec ce système. Il s'agit de l'identification complète d'un composé présent, c'est-à-dire la certitude du nom et de la structure de la molécule considérée. Il est nécessaire, pour atteindre ce niveau d'identification, de comparer les descripteurs empiriques de nos analyses avec ceux de standards analysés sur le même instrument au laboratoire. Les données de temps de rétention, de rapport m/z et de fragmentation doivent être similaires afin de pouvoir identifier clairement un métabolite.

Le *niveau 2* correspond à une annotation putative d'un composé du mélange analysé, c'est-à-dire une hypothèse sur son identité et sa structure. Bien que cette hypothèse puisse être fiable, elle ne peut devenir identification qu'en cas de comparaison avec un standard analysé en parallèle. Une annotation de niveau 2 peut être obtenue après comparaison des descripteurs ou des propriétés physico-chimiques du composé avec les informations acquises dans des bases de données, comme la masse mesurée d'une molécule et son spectre de fragmentation. C'est précisément ce que fait GNPS pour annoter les composés ; les annotations réalisées ainsi seront donc de facto de niveau 2.

Le *niveau 3* est attribué à un composé lorsque les informations possédées sont suffisantes pour supposer une famille chimique pour la *feature* considérée, mais néanmoins insuffisante pour attribuer un nom et une structure précise à ladite molécule. Il peut s'agir par exemple d'une incertitude face à certains isomères, mais avec des propriétés physico-chimiques ou une fragmentation en MS/MS typiques d'une famille chimique. Concrètement, la création de réseaux moléculaires permet d'attribuer rapidement des annotations de niveau 3 aux nœuds présents dans nos cartographies. En effet, les similitudes de fragmentation témoignent de ressemblances structurales, donc potentiellement de

composés partageant des groupements chimiques caractéristiques de familles moléculaires. Les composés des mêmes classes auront alors tendance à se connecter entre eux au sein des réseaux.

Enfin, le *niveau 4* correspond à un composé inconnu. Le manque d'informations des bases de données comparables aux données empiriques empêche la formulation d'une hypothèse sur l'identité et la classe chimique de la molécule étudiée. Ces annotations restent néanmoins intéressantes : il peut s'agir de composés connus dans la littérature mais que l'utilisateur n'a pas réussi à identifier à partir des résultats expérimentaux tout comme de composés véritablement inconnus, jamais isolés au préalable d'une source naturelle et donc originaux.

Certains projets reprennent ces niveaux d'annotations pour y ajouter des niveaux intermédiaires en fonction de la quantité et de la qualité des informations disponibles pour les différentes *features*. Les 4 niveaux de la MSI restent cependant un standard applicable à toute étude de métabolomique, notamment celles employant les techniques non-ciblées de chromatographie couplée à la spectrométrie de masse. La Figure 41 propose un schéma récapitulatif de ces niveaux d'annotations. Ce système d'annotations se révèle ainsi très pratique dans les études mettant en œuvre la déréplication d'extraits naturels, car il permet d'évaluer rapidement quels sont les composés déjà bien connus et identifiés dans la littérature et les composés inconnus originaux potentiellement intéressants à caractériser.

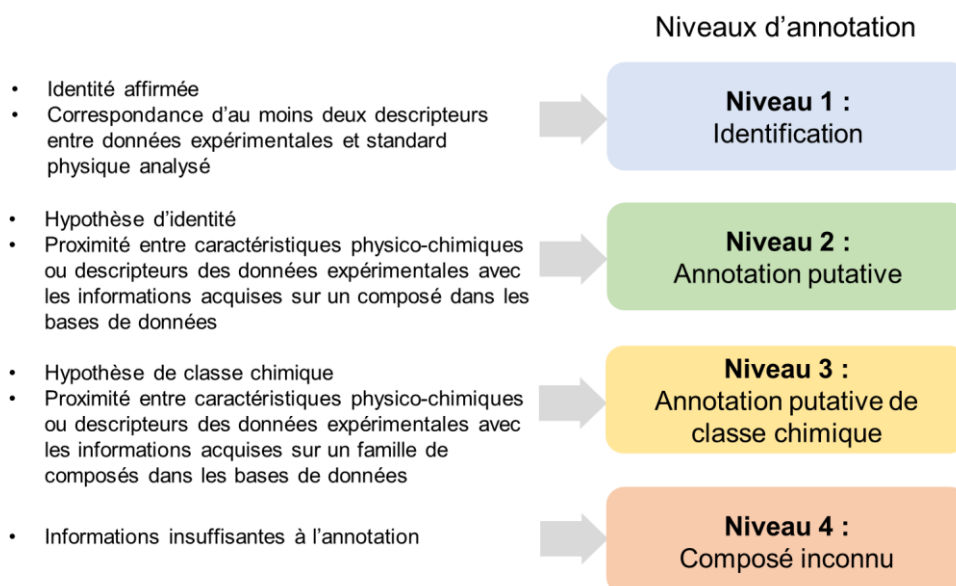


Figure 41 : schéma récapitulatif des différents niveaux d'annotations proposés par la Metabolomics Standard Initiative (MSI).

2. Investigation manuelle des résultats

Bien que toutes les informations obtenues expérimentalement soient considérées simultanément pour tenter l'annotation des métabolites des extraits analysés, il est possible de décrire les étapes de cette investigation dans un ordre logique et chronologiquement cohérent avec le procédé d'interprétation en lui-même.

a. Annotations à partir des réseaux moléculaires

La création de réseaux moléculaires par GNPS s'accompagne d'une recherche sur les bibliothèques spectrales disponibles en ligne et d'une comparaison des données MS/MS injectées sur la plateforme avec les données expérimentales publiques fournies par d'autres utilisateurs. Il s'agit d'une manière simple d'attribuer des annotations de niveau 2 aux *features* détectées dans nos analyses. De plus, de par la comparaison entre les fragmentations observées, ces annotations sont plutôt fiables. Il convient néanmoins de vérifier celles-ci par une visualisation des « spectres miroirs » fournis par GNPS. Un exemple de ce type de spectre est fourni en Figure 42.

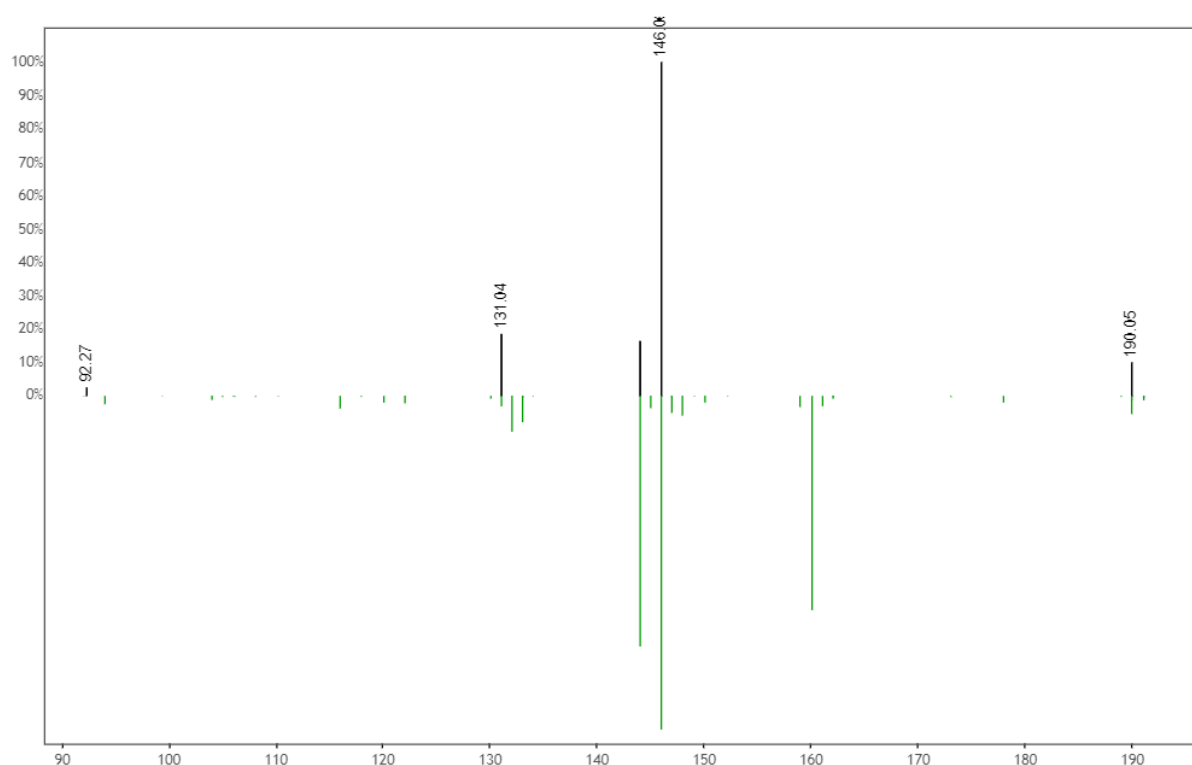


Figure 42 : spectre miroir fourni par GNPS à partir d'une comparaison avec des bases de données spectrales. Ce spectre correspond à l'annotation de l'acide hydroxyindoleacétique.

Les réseaux moléculaires, une fois construits et visualisés sous le logiciel Cytoscape, permettent une analyse visuelle des résultats des expériences UHPLC-MS/MS. Les annotations attribuées par la comparaison avec les bases de données peuvent être affichées en tant qu'étiquettes des nœuds appartenant à ces réseaux. Les connexions établies entre les différents métabolites sont également très informatives, car elles permettent de regrouper un certain nombre de molécules entre elles au sein d'une famille chimique commune. Par exemple, si l'un des composés d'un réseau possède une attribution, on peut supposer que les substances appartenant à ce réseau font partie de la même classe moléculaire. On peut donc attribuer facilement des annotations de niveau 3 à partir des réseaux moléculaires.

b. Annotation à partir des données spectrales

Certains composés détectés ne pourront être annotés via les réseaux moléculaires, car pour une proportion d'entre eux, aucun lien n'est établi avec d'autres composés. La comparaison des spectres MS/MS avec des bases de données spectrales pourra cependant être réalisée afin d'attribuer des annotations de niveau 2 auxdites *features*. Les bases YMDB et HMDB, par exemple, fournissent des spectres MS2 expérimentaux ou encore des spectres de fragmentation *in silico*. Il est donc possible de déterminer des correspondances au niveau de la fragmentation entre nos données (caractérisées alors par un rapport m/z représentatif de la masse de la molécule) et les spectres en ligne des métabolites possédant la même masse.

De la même manière, cette comparaison peut être réalisée à partir de nœuds faisant partie de réseaux caractéristiques de certaines familles chimiques. L'annotation est alors accélérée, car il s'agit dans ce cas d'attribuer une structure et un nom à une molécule dont nous supposons au préalable l'appartenance à une classe de composés.

Les données MS/MS sont donc une source d'informations très importante pour l'attribution d'annotations de deuxième niveau aux *features* détectées. Dans les cas où l'annotation de niveau 2 n'est pas possible (si la correspondance entre les données MS/MS n'est pas assez claire pour différencier des composés de masse très proche mais de formule chimique différente, ou encore pour discriminer deux isomères), les spectres de fragmentation peuvent néanmoins aider à l'attribution d'une annotation de troisième niveau, si les fragments observés dans ces spectres sont caractéristiques de groupements chimiques spécifiques à des familles de molécules (par exemple, si l'on observe un fragment caractéristique d'un groupement indole au sein d'un métabolite).

c. Annotation des métabolites de *S. cerevisiae*

Les extraits de cultures de *S. cerevisiae* ont été analysés par UHPLC-MS/MS et les données acquises ont servi à la construction de réseaux moléculaires ainsi que de tableaux regroupant les *features* détectées. Il est à noter que les données provenant des extraits DCM/AE et M sont alignées et donc traitées ensemble à partir de ce stade. Les données provenant d'analyses en mode d'ionisation positif sont cependant traitées séparément des données acquises en mode d'ionisation négatif.

Après ces étapes, un grand nombre de variables ont pu être écartées des résultats. Ces *features* correspondent par exemple à des artefacts, des molécules provenant de la phase stationnaire de la colonne de chromatographie, à des composés du blanc analytique, des composés exclusifs au milieu de culture ou encore des adduits, dimères ou fragments en source. Une fois ces *features* écartées, il reste alors les exploitables. Ce sont les composés exclusifs aux levures ou les composés du milieu formant des réseaux avec les métabolites d'origine fongique et donc pouvant aider à l'exploration de ces derniers. Les résultats de ce classement des *features* sont présentés en Figure 43.

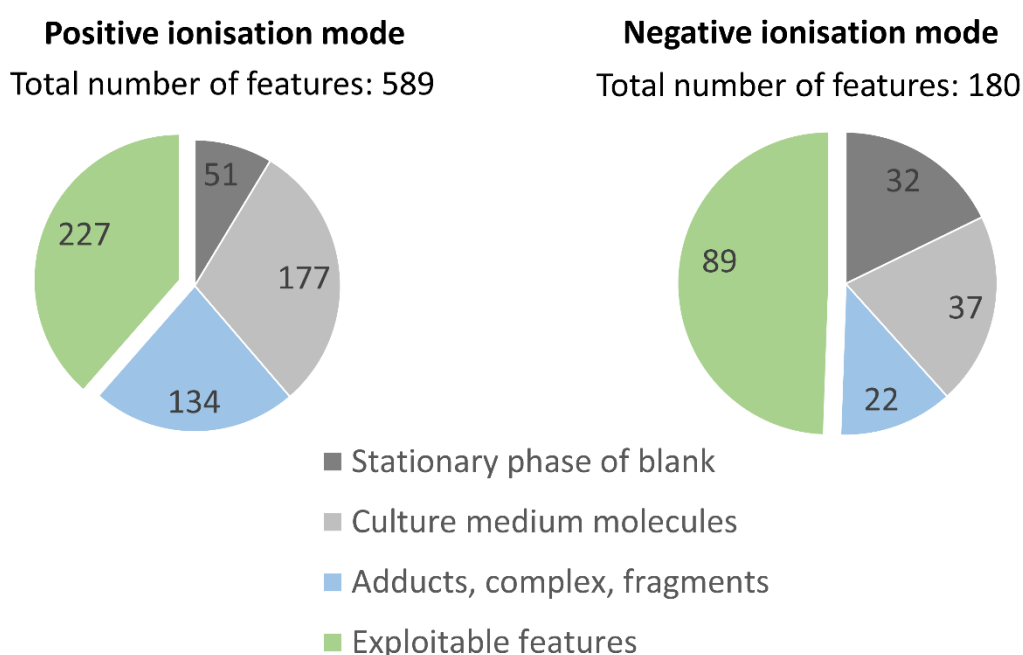


Figure 43 : résultats d'analyse UHPLC-MS/MS d'extraits de *S. cerevisiae* ; classement des *features* exploitables et des *features* non informatives écartées.

Après ce premier tri permis par l'alignement de données, les éléments exploitables ont été investigués dans le but d'explorer au mieux les métabolites de la levure *S. cerevisiae* à l'aide d'un maximum d'informations disponibles. Par exemple, les composés présents dans le milieu de culture et dans les extraits de levures ainsi que les composés du milieu de culture exclusivement mais également présents

dans des réseaux moléculaires ont été conservés car ils peuvent renfermer des informations de fragmentation pouvant apporter une aide dans l'annotation des composés exclusifs aux molécules détectées dans les extraits de microorganismes. Un certain nombre d'annotations a donc pu être réalisé grâce à l'exploration des données, et les statistiques concernant ces résultats sont présentés en Figure 44.

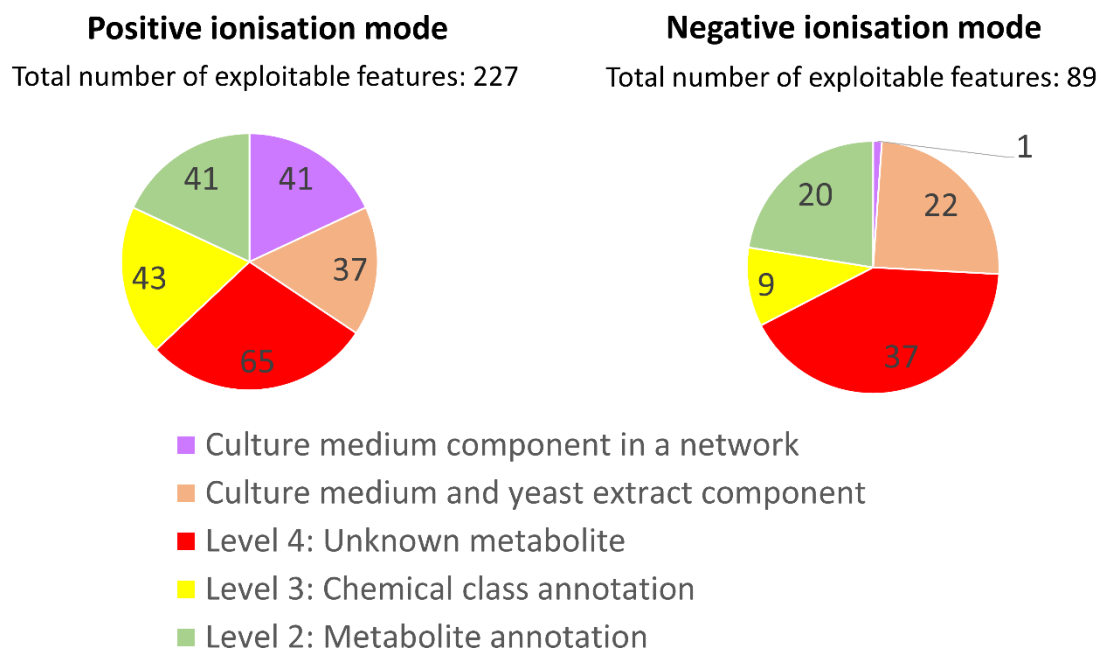


Figure 44 : résultats d'analyse UHPLC-MS/MS d'extraits de *S. cerevisiae* ; features exploitables pour l'investigation des résultats et niveaux d'annotation après exploration manuelle.

d. Réseaux moléculaires

Un extrait des réseaux moléculaires qui ont été construits à partir des analyses d'extraits de *S. cerevisiae* est présenté ci-après en Figure 45. Ce sont ces réseaux qui ont aidé à l'attribution d'annotations de niveau 3 ainsi que certaines annotations de niveau 2 concernant les métabolites de la levure modèle.

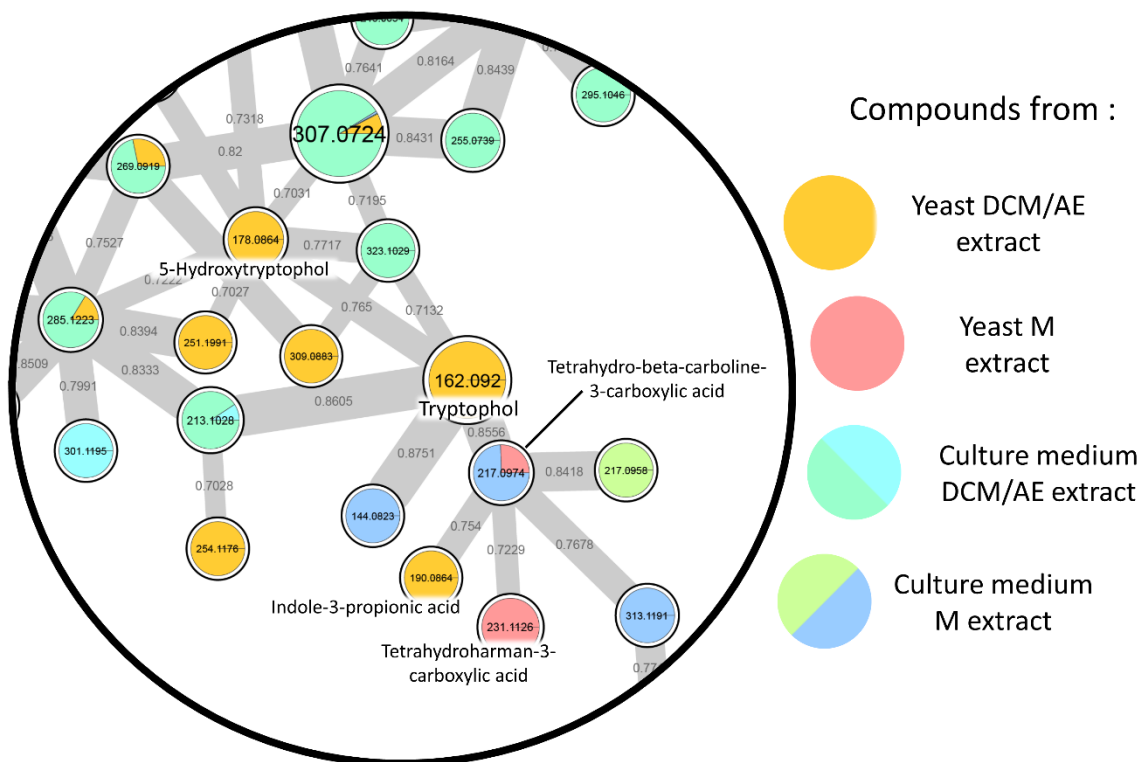


Figure 45 : extrait de réseaux moléculaires construits à partir d'analyses UHPLC-MS/MS d'extraits DCM/AE et M de culture de *S. cerevisiae* conduites en mode d'ionisation positif et négative. La couleur des nœuds correspond à l'échantillon d'origine des composés. Leur taille est réglée proportionnellement à l'intensité des composés dans leurs analyses. Certains composés sont annotés.

e. Conclusion et limites quant aux résultats d'analyse de *S. cerevisiae*

Les résultats d'analyse d'extraits de *S. cerevisiae* sont satisfaisants. La méthode analytique est solide, fournit des informations de valeurs et le procédé d'interprétation des données est très informatif. Les représentations des résultats, que ce soit les tableaux ou les réseaux moléculaires, fournissent une cartographie intéressante des métabolites de levures, avec un relativement grand nombre de molécules annotées ainsi que des composés inconnus repérés. La méthode mise en place semble donc très intéressante pour l'exploration non-ciblée des métabolites secondaires d'autres souches de levures, et est par extension applicable à d'autres sources de produits naturels.

Il existe cependant des points d'amélioration qui ressortent des résultats indiqués. Les niveaux d'annotation sont bien fournis pour un sujet d'étude de métabolomique non ciblée. Néanmoins, le niveau le plus intéressant d'annotation, le niveau 2, correspond à moins d'un tiers des annotations attribuées aux métabolites détectés, aussi bien en mode d'ionisation positif que négatif. La proportion

d'annotations putatives mériterait à être augmentée, tout comme le nombre de composés inconnus gagnerait à être diminué dans le but de rendre les résultats plus informatifs sur ce que la levure est en mesure de produire.

Il existe également un certain nombre de cas où une ambiguïté est repérée dans l'annotation des molécules. La mesure de masse à haute résolution avec une précision moyenne inférieure à 3 ppm ne permet pas toujours l'attribution d'une formule moléculaire unique ; plusieurs formules possibles sont alors proposées par MZMine2. Cela peut donc entraîner une difficulté à discriminer la bonne formule et donc la possibilité d'attribuer un nom et une structure précise à une *feature*.

Ces limites nous ont donc amené à ce stade de la thèse à réfléchir à une nouvelle méthode, une technique analytique dont les résultats, combinés à ceux déjà acquis par UHPLC-MS/MS, nous permettraient d'affiner les hypothèses concernant les molécules détectées dans nos analyses et améliorer les statistiques des niveaux des annotations attribuées.

Chapitre III : Ajout de la FTICR-MS et
approfondissement du traitement de données pour
application sur *S. cerevisiae*.

I. La FTICR-MS appliquée à la déréplication

La spectrométrie de masse à ultra-haute résolution est de plus en plus utilisée en métabolomique, et ce sur de multiples applications [147–149,184]. La FTICR-MS est une technologie puissante qui fournit de précieuses informations. La mesure de masse très précise et l'ultra-haute résolution permettent l'attribution de formules moléculaires uniques à chaque signal détecté, et ces attributions permettent la cartographie des signaux acquis pendant les analyses, ce qui se révèle très utile pour la comparaison inter-échantillon. La différence de résolution entre une analyse FTICR-MS et une analyse UHPLC-MS/MS (réalisée avec un analyseur de type Q-TOF) est illustrée en Figure 46.

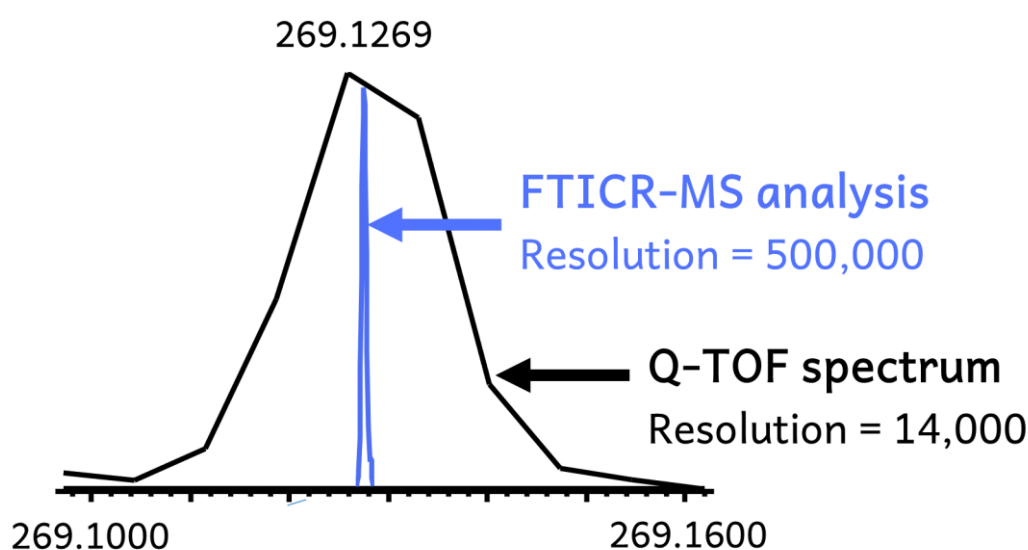


Figure 46 : superposition de deux spectres de masse en mode profil acquis sur le même échantillon ; en bleu, analyse par FTICR-MS ; en rouge, analyse par UHPLC-MS/MS (analyseur Q-TOF).

La détermination d'une formule moléculaire unique est le premier avantage qu'il est possible de tirer d'une expérience FTICR-MS. L'idée, à ce stade de la thèse, serait alors d'exploiter ce point fort pour l'amélioration du processus d'annotation des composés. En effet, certaines *features* présentent dans les premiers résultats sur *S. cerevisiae* plusieurs formules chimiques possibles. Ces ambiguïtés amènent à une incertitude concernant l'attribution d'annotation des certaines molécules ; l'annotation de niveau 2 et parfois même de niveau 3 se complique si la formule moléculaire n'est pas certaine.

Les représentations de données expérimentales après des analyses par FTICR-MS sont multiples : il est possible de dresser des diagrammes de van Krevelen ou des cartographies *DBE vs carbon number* par exemple. Ce type de graphique apporte un degré d'information supplémentaire aux échantillons

naturels analysés en cartographiant les signaux détectés selon une signification chimique. Cette cartographie s'avère intéressante pour dresser des profils métaboliques des extraits à analyser, et pour comparer lesdits extraits entre eux. Il s'agit donc d'un deuxième avantage à l'utilisation de cette technique de spectrométrie de masse en métabolomique.

Des expériences de FTICR-MS pourraient permettre de lever les ambiguïtés sur la détermination des formules moléculaires des métabolites détectés et apporteraient une information très intéressante pour notre projet de cartographie de métabolites secondaires. C'est pour ces raisons qu'il fut décidé d'incorporer des analyses par FTICR-MS à notre workflow de déréplication afin de compléter les résultats obtenus tout en fournissant de nouvelles données.

1. Optimisation pratique de la méthode

Dilution des échantillons

Les échantillons préparés à partir des extraits secs de culture de levures ont été analysés par UHPLC-MS/MS avec une concentration finale estimée approximativement à 5 mg/mL.

Une telle concentration est adéquate pour l'utilisation avec l'instrumentation ESI-TOF-MS/MS de la partie centrale du procédé analytique du projet. Elle est cependant loin d'être recommandée pour des expériences de FTICR-MS. Cette technologie présente une très grande sensibilité et une telle concentration créerait une intense pollution de l'instrument, notamment car les expériences conduites en DI-MS n'ont pas l'avantage d'être en couplage avec la chromatographie liquide, celle-ci pouvant diffuser l'échantillon le long du gradient de solvants appliqués (alors qu'en introduction directe, tous les composés du mélange sont injectés puis ionisés simultanément).

Des analyses sur les extraits de levure dilués selon une cascade de dilutions ont été réalisées afin de déterminer le facteur de dilution le plus adapté à un compromis entre l'amélioration du signal acquis et une potentielle pollution de l'appareil. Les échantillons analysés en UHPLC-MS/MS sont donc dilués au 1/10 000^{ème} avant analyse par FTICR-MS.

Étalonnage interne par ajout d'ions connus

La FTICR-MS permet des mesures de masse très précises ; il est néanmoins nécessaire d'étalonner en interne les spectres après l'acquisition afin de maximiser la précision de ces mesures. Un mélange d'ions de masses connues est donc ajouté à chaque échantillon analysé (*Agilent Technologies tuning mix*, ions à m/z 118,0863 ; 322,0481 ; 622,0290 ; 922,0098). Le volume d'ajout a été optimisé à la suite de

plusieurs analyses à différentes concentrations de ce mélange, afin de maximiser le signal des ions connus sans que ce signal soit plus intense à celui des métabolites détectés, exclusifs aux extraits de levure.

2. Corrélation entre résultats UHPLC-MS/MS et FTICR-MS

Après des analyses préliminaires d'extraits de *S. cerevisiae* par FTICR-MS, les résultats acquis ont été comparés avec ceux obtenus par UHPLC-MS/MS. Une grande partie des signaux ont été retrouvés en commun entre les deux techniques analytiques, que ce soit en mode d'ionisation positif qu'en mode d'ionisation négatif. Cependant, quelques *features* firent une exception à ce constat, et la raison de ces formules manquantes fut investiguée.

Plusieurs hypothèses ont été formulées pour trouver les causes de ce problème. La première concerne les paramètres de transmission des ions de la source vers la cellule ICR, c'est-à-dire les paramètres qui influencent la sensibilité de l'appareillage quant à l'analyse d'ions de plus haut rapport m/z (au-delà de m/z 500). Ces paramètres sont des tensions appliquées à certaines parties de l'instrument : l'amplitude du temps de vol, l'amplitude de l'octopôle, la fréquence du transfert optique, la tension de la plaque déflectrice, la tension des « entonnoirs » traversés par les ions et celle appliquée au capillaire de sortie. Les paramètres en question ont été réglés au cours de plusieurs expérimentations afin d'améliorer la transmission des ions de plus haut rapport m/z . L'objectif était de trouver un compromis pour la transmission aussi bien des plus bas rapports m/z (de l'ordre de 100) que les plus hauts (aux alentours de 950), afin de combler le manque de certains composés déjà détectés en UHPLC-MS/MS. Ces essais ne furent malheureusement pas concluants, car les réglages améliorant la détection des composés de plus haut rapport m/z entraînaient une diminution de l'information acquise pour les molécules à basse masse.

La deuxième hypothèse avancée concerne la potentielle dégradation des composés qui aurait pu se réaliser entre l'analyse UHPLC-MS/MS et l'analyse par FTICR-MS. En effet, six mois se sont écoulés entre les analyses des extraits de levures par les deux méthodes citées. Il est donc possible qu'il y ait eu une perte de métabolites pendant le temps de conservation de l'échantillon liquide. Des analyses UHPLC-MS/MS furent donc conduites selon les mêmes paramètres que les premières afin de vérifier la détection des différents signaux manquants. Cependant, les composés qui faisaient défaut à l'analyse FTICR furent retrouvés durant les expériences par LC-MS sur les échantillons d'origine. L'hypothèse de dégradation des molécules est donc à écarter.

En conclusion, la non-détection de certaines espèces est certainement due aux phénomènes de compétition à l'ionisation et suppression du signal à cause de l'introduction directe de l'échantillon dans la source *electrospray*. Les composés présents dans l'échantillon sont tous ionisés simultanément, et des interactions sont possibles entre eux au moment de l'ionisation. Il en résulte alors une perte au niveau de certains signaux. Ceci n'a pas lieu en UHPLC-MS/MS car les composés sont séparés lors de leur passage au travers de la colonne chromatographique. Ils sont ionisés séparément et progressivement pendant la durée de l'expérience, limitant ainsi les problèmes de suppression de signal. Aucune solution n'eut raison de cet écueil et certaines *features* détectées en UHPLC-MS/MS ne purent être associées à une formule brute unique ; ceci n'est le cas que pour une petite fraction des résultats, et les analyses FTICR-MS restent très informatives pour la suite de la thèse.

3. Traitement de données FTICR-MS

Les analyses conduites par FTICR-MS ont la caractéristique de générer des jeux de données encore plus lourds que les analyses UHPLC-MS/MS. La sensibilité et la grande résolution de cette technique permettent l'obtention de très gros fichiers présentant une importante quantité de signaux à traiter.

Attribution de formules moléculaires

Dans le cadre de cette thèse, l'attribution de formule moléculaire a été réalisée grâce à un outil d'attribution automatisé du logiciel développé par le constructeur de l'appareil sur lequel les analyses furent faites (Bruker). En effet, DataAnalysis propose un module nommé SmartFormula. Les paramètres réglables lors de l'utilisation de cet outil permettent lors de la génération des formules moléculaires de respecter les sept règles d'or de l'attribution d'une composition chimique [172].

Une étape de tri manuel des formules moléculaires fut introduite dans le procédé de traitement de données, afin d'éliminer les potentielles formules aberrantes, conserver une formule moléculaire unique pour chaque signal et ainsi gagner en confiance dans les résultats. Cette étape fut réalisée toujours dans le respect des sept règles d'or.

Soustraction des ions du milieu de culture

L'application de l'algorithme décrit précédemment sur les résultats d'analyse d'extraits de levures ont donné de longues listes de formules moléculaires. Afin d'exclure les formules qui furent attribuées à des composés qui ne sont pas exclusifs aux levures mais qui appartiennent également au milieu de culture, il a été nécessaire de trouver une méthode pour discriminer les ions appartenant aux extraits

de *S. cerevisiae* de ceux provenant du milieu de culture. Un module appartenant au logiciel DataAnalysis fut appliqué aux résultats d'analyse ; celui-ci soustrait des listes de pics sélectionnées les pics présents dans une autre analyse, au choix de l'utilisateur. Il fut donc appliqué aux analyses d'extraits de levures afin d'en soustraire les signaux détectés dans les analyses correspondant aux milieux, afin d'écartier ceux-ci et de mettre en lumière les composés exclusifs aux microorganismes.

Cartographie des métabolites de levures

Le processus de traitement de données FTICR-MS fut complété d'une étape de construction de diagrammes pour la cartographie des extraits analysés. Ce sont les diagrammes de van Krevelen qui furent sélectionnés pour représenter les résultats obtenus. Pour rappel, il s'agit de graphiques où les *features* détectées sont placées selon le rapport O/C et le rapport H/C de la formule moléculaire qui leur est attribuée. La position des points correspondant aux métabolites dans les différentes zones du diagramme amène à l'observation des proportions de la variété de classes chimiques représentée dans l'échantillon analysé. L'attribution d'un code couleur aux composés permet également de figurer les différentes compositions chimiques des molécules de l'échantillon (en atomes de C, H, O, N, P et S par exemple). Enfin, une telle cartographie permet une comparaison entre différents échantillons pour lesquels un diagramme a été réalisé.

II. Résultats de la déréplication de *S. cerevisiae* : publication des résultats

Les résultats précédemment obtenus pour l'exploration des métabolites de *S. cerevisiae* furent grandement améliorés après l'ajout de la technique de FTICR-MS au processus analytique. La FTICR-MS apporta beaucoup à l'annotation des composés. L'ajout des formules moléculaires précises et un approfondissement de l'investigation manuelle des résultats ont grandement augmenté le nombre d'annotations avec des niveaux intéressants (les niveaux 2 et 3, respectivement l'annotation putative d'une molécule et l'annotation de classe chimique). Cela signifie que certains composés inconnus, auparavant annotés au niveau 4, appartiennent désormais au niveau 3 et possèdent une classe chimique. Certaines *features*, auparavant de niveau 3, ont désormais une annotation correspondant à un nom et à une structure moléculaire précise. Les informations concernant les petites molécules de la souche modèle sont donc grandement affinées par les résultats obtenus. De plus, des standards ont été analysés pour permettre des annotations de niveau 1 (liste disponible en Annexe page 207).

Il s'agit d'une première utilisation de la combinaison de l'UHPLC-MS/MS, du traitement par réseaux moléculaires et de la FTICR-MS à *S. cerevisiae*. Au total, 128 *features* exclusives à la levure ont été détectées. Grâce au procédé analytique, les annotations de composés et de classes chimiques atteignent respectivement 89 % en mode d'ionisation positif et 87 % en mode d'ionisation négatif, ce qui est élevé pour une étude métabolomique. Les annotations démontrent un riche contenu en alcaloïdes avec 58 % des *features* détectées faisant partie de cette famille de molécules, ainsi qu'une bonne proportion en acides aminés et dérivés. De plus, la FTICR-MS apporte une nouvelle dimension aux résultats avec la cartographie par diagrammes de van Krevelen, confirmant au passage les familles moléculaires majoritaires.

Ces résultats sur la déréplication de *S. cerevisiae* ont conduit à une publication dans le *Journal of Pharmaceutical and Biomedical Analysis (Elsevier)* [185]. Cette publication est jointe à partir de la page suivante ; le format est adapté au style de ce manuscrit de thèse. Les figures sont insérées dans le texte telles qu'elles le sont dans la publication finale. Les références sont numérotées selon l'ordre propre à l'article, et donc différemment de l'indexation réalisée dans l'ouvrage présent. Les informations supplémentaires sont fournies dans la partie Annexes.

Référence : O. Perruchon et al., *J. Pharm. Biomed. Anal.* 195 (2021) 113857.

<https://doi.org/10.1016/j.jpba.2020.113857>

Combination of UHPLC-MS/MS-molecular networking approach and FTICR-MS for the metabolic profiling of *Saccharomyces cerevisiae*

Olivier Perruchon^a, Isabelle Schmitz-Afonso^a, Cécile Grondin^b, Serge Casaregola^b, Carlos Afonso^a, Abdelhakim Elomri^a

^aNormandie Univ, UNIROUEN, INSA Rouen, CNRS, COBRA (UMR 6014), Rouen 76000, France

^bUniversité Paris-Saclay, INRAE, AgroParisTech, Micalis Institute, CIRM-Levures, 78350, Jouy-en-Josas, France

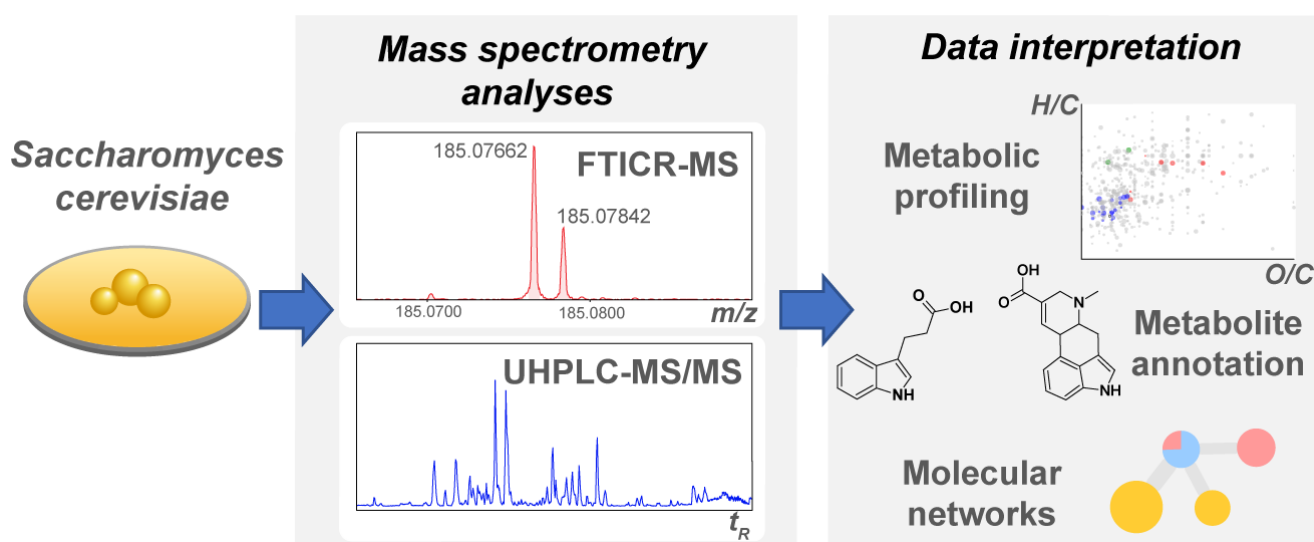
Corresponding author:

E-mail address: hakim.elomri@univ-rouen.fr (A. Elomri).

Keywords:

FTICR-MS; Mass spectrometry; Dereplication; Molecular networking; *Saccharomyces cerevisiae*; Metabolomics

Received: 11 September 2020. / *Received in revised form:* 18 November 2020. / *Accepted* 16 December 2020. / *Available online:* 19 December 2020



Abstract

Natural products are a reliable source of bioactive molecules and represent an industrial and pharmaceutical stake. Indeed, the model yeast species *Saccharomyces cerevisiae* is a well-known eukaryotic organism largely used as a biotechnological tool, but still a topical subject of study. In this work, the exploration of *Saccharomyces cerevisiae* is taken further through an untargeted metabolomics workflow. The aim is to enrich databases and bring new information about the standard *S. cerevisiae* strain in a given medium. Analytical methods and bioinformatics tools were combined in a high-throughput methodology useable to dereplicate many types of biological extracts and cartography secondary metabolites. Ultra-high performance liquid chromatography-tandem mass spectrometry (UHPLC-MS/MS) analyses were carried out and spectral data were pre-processed to build molecular networks. Annotations were attributed to compounds through comparison with databases and manual investigation of networks. Ultra-high resolution Fourier-transform ion cyclotron resonance mass spectrometry (FTICR-MS) brought additional information thanks to a higher dynamic range and enhanced UHPLC-MS/MS results by unveiling ambiguities and bringing accurate molecular formulae. Therefore, accurate and reliable annotated features resulted from the UHPLC-MS/MS data while FTICR-MS provided an overall cartography of metabolites thanks to van Krevelen diagrams. Various small molecules such as amino acids derivatives and indole alkaloids have been determined for the first time in this yeast. The complementarity of FTICR-MS and UHPLC-MS/MS for secondary metabolite annotation brought this new mapping of *S. cerevisiae*.

1. Introduction

Saccharomyces cerevisiae is one of the best-known model systems for eukaryotic cell biology. It belongs to the *Saccharomycotina* yeasts, a subphylum of Ascomycete unicellular fungi found on a wide variety of substrates such as insects, plants or soils [1]. The *S. cerevisiae* strain S288c is a widely used laboratory strain and was the first eukaryotic organism to have its genome fully sequenced through a systematic sequencing project. This reference sequence is stored in the *Saccharomyces* Genome Database (SGD) [2]. Yeasts are involved in numerous agri-food applications like winemaking, brewery, bakery [3] or drugs as probiotics, but *S. cerevisiae* is also an effective biotechnological tool for the production of biopharmaceuticals: hormones like insulin, or vaccines such as hepatitis B vaccine and human papillomavirus vaccine [4–6]. Fermentation and other biological processes make yeasts capable of producing interesting secondary metabolites that represent a large structural biodiversity. As natural products can show biological effects against many diseases, determining what *S. cerevisiae* naturally synthesizes represents an industrial, pharmaceutical and scientific stake. Previous studies have characterized some of these small molecules by gas chromatography-mass spectrometry (GC–MS) analyses [7,8], liquid chromatography-mass spectrometry (LC–MS) analyses [9] or carried out metabolic profiling by nuclear magnetic resonance (NMR) techniques [10].

The last two decades have brought many technological advances in mass spectrometry, concerning ionization sources, analyzers or hyphenation with other separation techniques. Therefore, metabolomics research has evolved and among common techniques, high performance liquid chromatography (HPLC) is an efficient starting point to separate compounds according to their polarity. The coupling of liquid chromatography with high-resolution mass spectrometry (HPLC-HRMS), with atmospheric pressure ionization such as electrospray ionization, allows a screening of a wide range of metabolites. In addition, tandem mass spectrometry (MS/MS), and in particular high resolution MS/MS spectra, affords structural information on metabolites [11,12].

High-resolution mass spectrometry analyses in metabolomics generate an important amount of data that require automated data processing. Indeed, the complete analytical workflow in metabolomics studies implies data acquisition, data processing and finally data interpretation. With regard to data processing, computerized and automated tools are great time savers as manual MS-data deciphering is heavily time-consuming. The use of algorithms also avoids human error. As available applications are in development, there are a lot of parameters to optimize. For example, the open software MZMine2 processes HPLC-MS/MS data to convert spectral data in feature lists [13,14].

With regard to data interpretation the first descriptor used is the mass determination. Accurate mass-to-charge ratios (m/z) measurement allow to attribute molecular formulae that can be compared to databases to annotate the features (e.g. Yeast Metabolome Database, Human Metabolome Database, MassBank [15–18]). High-resolution analyzers such as time-of-flight and orbitrap present a mass accuracy up to 5 ppm. With such accuracy there are however often several possible molecular formulas for a given signal even considering specific criteria such as the seven golden rules [19]. Fourier Transform Ion Cyclotron Resonance systems (FTICR-MS) have the required resolution ($>1,000,000$) and accuracy (<0.5 ppm) to solve this issue [20]. FTICR-MS is used in metabolomics for rapid metabolome fingerprinting or exploratory profiling with direct injection approaches [21,22]. The high quality of high field FTICR spectra allow to attribute unique molecular formulae to each ion with great confidence. The use of isotopic fine structure can be used to confirm the presence of some specific elements such as sulfur [23]. This annotated data can be used to generate molecular mapping using for instance van Krevelen diagrams [24,25].

The other descriptors used in data interpretation are the tandem mass spectra of metabolite ions. Fragmentation data are used on the Global Natural Products Social molecular networking platform to build molecular networks (<https://gnps.ucsd.edu>), as structurally related molecules share similar fragmentation patterns [26]. In addition, GNPS compares uploaded MS/MS spectra to constantly enrich databases and annotate compounds from the feature list. This recent bioinformatics tool is already a valuable instrument in the exploration of microorganisms and in drug discovery [27,28]. The combination of MZMine2 and molecular networking proves to be very worthwhile to accelerate the dereplication of biological samples and subsequently focus on potential unknown compounds.

In metabolomics studies, annotation is carried out based on the Metabolomics Standard Initiative annotation levels thanks to all descriptors obtained previously [29]. Ranging across four different levels, it respectively covers fully identified molecules, putatively annotated compounds, putatively annotated compound classes and unknown metabolites.

Thus, we present here an in-depth exploration of *S. cerevisiae* reference strain S288c extracts via the complementarity of UHPLCMS/MS and Direct Injection FTICR-MS, completed by molecular networking data treatment. This approach accelerated significantly the deciphering of the biological extracts, and could be applied to other complex mixtures.

2. Material and methods

2.1. Statement of human and animal rights

This article does not contain any studies with human participants or animals.

2.2. Yeast culture

Saccharomyces cerevisiae S288c (CLIB 338) standard laboratory strain is preserved at the Biological Resource Centre CIRM-Levures (INRAE, Jouy-en-Josas, France; www.inra.fr/cirm_eng/Yeasts). This strain was routinely grown on YPD (1 % yeast extract, 1 % peptone, 1 % glucose) agar medium at 26.5 °C. For analysis, strains were streaked from an overnight culture on YNB (6.7 % Yeast Nitrogen Base, 1 % peptone, 1 % glucose) agar medium and incubated for 3 days at 26.5 °C. One culture have been extracted, and control samples, also called culture medium extracts were prepared in the same way, except that no yeast strain were added to the culture medium.

2.3. Metabolites extraction

Dichloromethane, ethyl acetate and methanol are HPLC gradient grade solvents. Intracellular and extracellular metabolites were extracted by addition of 15 mL of 1:1 v/v dichloromethane and ethyl acetate on the culture, helped by a 30 min sonication (Branson B5200–54, 450 W). The extract was filtered through a sintered glass funnel (porosity from 16 m to 40 m). 10 mL of dichloromethane and ethyl acetate (1:1, v/v) were added to the medium, sonicated 10 min and filtered. The two gathered dichloromethane / ethyl acetate extracts were concentrated by solvent dry evaporation under vacuum at 40 °C to give a brown residue of 8.5 mg named “DE extract” and stored at – 20 °C. 10 mL of methanol were added to the same growing medium, the mixture was sonicated 5 min and filtered. The extracted metabolites were then concentrated by evaporation. A brown gel residue of 183.5mg was obtained, named “M extract” and stored at – 20 °C. Culture medium without yeasts went through the same process (leading to a CDE extract and a CM extract), prepared and analyzed in the same way.

2.4. Sample preparation for analysis

Methanol, water and acetonitrile are Fisher Chemical Optima LC–MS grade. DE and CDE extracts were resuspended in methanol, M and CM extracts were resuspended in water to a concentration of 50 mg/mL. Every sample was half-diluted with water / acetonitrile (1:1 v/v), which corresponds to the mobile phase, for analysis. All samples were filtered through PTFE 0.45 m filters. Analytical blanks were prepared with the same solvents as the mobile phase and analyzed before and at the end of each analytical sequence to further clean-up the spectral data of background interference signals, like plasticizers or solvent impurities. Pure chemical standards (Table S1) were individually dissolved in water / acetonitrile (1:1 v/v) and mixed, then diluted to 0.01 mg/mL and analyzed by UHPLC-MS/MS.

2.5. UHPLC-ESI-QTOF-MS/MS analysis

Chromatographic separation was carried out using a UHPLC system (Dionex Ultimate 3000 UPLC+, Thermo Scientific, San Jose, CA, USA) equipped with a C18 silica-based Column (Acquity UPLC HSS T3 1.8 m x 1.0 mm x 100 mm, Waters Corporation, Milford, MA, USA) with a prefilter of 0.2 m. The column was kept at 40 °C during the analysis. An autosampler kept the samples at 4 °C. The injection volume was 1 L. The solvents used for gradient separation are the following: water + 0.1 % formic acid (solvent A); acetonitrile + 0.1 % formic acid (solvent B); solvents are Fisher Chemical Optima LC–MS grade; formic acid is LiChropur (Merck) for LC–MS. The flow rate was 0.1 mL/min. Gradient consisted of 1 % B at 0 min; 1 % B at 1 min; going up linearly to 100 % B at 23 min then maintained 100 % B for 2 min; finally, 1 % B as initial conditions at 25.2 min. The run ended after 30 min including the stabilization step.

UHPLC was coupled to a hybrid quadrupole-time of flight analyzer (QTOF, SYNAPT G2 HDMS, Waters MS Technologies, Manchester, UK) equipped with an Electrospray Ionization source (ESI). The instrument was driven by MassLynx 4.1 software. Experiments were realized in positive and in negative ionization mode. m/z range was 50–2000. The ionization parameters and acquisition parameters for the data dependent scan method are indicated in Table S2. The instrument was externally calibrated with a sodium formate mixture (10 % formic acid / 0.1 M NaOH / acetonitrile in 1:1:8, v/v/v). The lock mass used during runs as internal calibrant was a Leucine-Enkephalin solution (2 ng/L in water/isopropanol 1:1, v/v); the lock mass ions were m/z 556.2771 and m/z 554.2615 in positive and negative ionization mode respectively.

2.6. Spectral data processing

UHPLC-MS/MS data were firstly centroided on MassLynx 4.1 with the “Automatic peak detection” module. Raw data (Waters “.raw” format) were then imported within the MZMine2 software. Features are characterized by variables as m/z ratio, retention time (RT in minutes), intensity, and are linked to corresponding MS/MS spectra. Retention time is considered so isomers can be discriminated. Peaks were detected with the “Exact mass” algorithm from the “Mass detection” module then detected masses were built into a chromatogram with “ADAP Chromatogram Builder” [30]. The chromatogram was deconvoluted with the Wavelet (ADAP) algorithm, removing background noise features then MS2 scans were paired to corresponding MS1 features. Isotopes were grouped and duplicate features were removed. Then, feature lists from samples and controls analyzed with the same ionization polarity were aligned. The “Formula prediction” module attributed molecular formulae following the seven golden rules [19]. Fragments and adducts were automatically searched with adequate modules. The detailed parameters for each module are shown in Table S3. Feature lists were then exported in a fitted format for GNPS [14].

2.7. Molecular networking on global natural products social molecular networking

We exported MZMine2 pre-processed feature lists and imported it on the GNPS platform. A molecular network was created with the feature based molecular networking workflow (<https://ccms-ucsd.github.io/GNPSDocumentation/featurebasedmolecularnetworking/>) on the GNPS website (<http://gnps.ucsd.edu>). The data was filtered by removing all MS/MS fragment ions within +/- 17 Da of the precursor m/z. MS/MS spectra were window filtered by choosing only the top 6 fragment ions in the +/- 50 Da window throughout the spectrum. The precursor ion mass tolerance was set to 0.02 Da and a MS/MS fragment ion tolerance of 0.02 Da. A network was then created where edges were filtered to have a cosine score above 0.7 and more than 4 matched peaks. Further, edges between two nodes were kept in the network if and only if each of the nodes appeared in each other’s respective top 10 most similar nodes. Finally, the maximum size of a molecular family was set to 0, and the lowest scoring edges were removed from molecular families until the molecular family size was below this threshold. The spectra in the network were then searched against GNPS’ spectral libraries. The library spectra were filtered in the same manner as the input data. All matches kept between network spectra and library spectra were required to have a score above 0.7 and at least 4 matched peaks. Molecular networks were visualized through Cytoscape.

2.8. FTICR-MS analysis

Analyses were carried out on a FTICR instrument (SolariX XR FTMS, Bruker Daltonics) equipped with a 12 T superconducting magnet and a dynamically harmonized ICR cell. The instrument is equipped with an electrospray ionization source. The same samples analyzed by UHPLC-MS/MS were analyzed by direct injection FTICR-MS after dilution (1/10,000) with water / acetonitrile (1:1, v/v), at a flow of 120 L/h. Each sample was analyzed both in positive and negative ionization mode. Source and analyzer parameters (shown in Table S4) were inspired by previous work and chosen in order to guarantee a robust and sensitive analysis [31]. FTICR-MS data were treated with Data Analysis 5.0 (Bruker). Molecular formulae were attributed considering M + H, Na and K adducts for positive mode, M – H and Cl for negative mode, for C₁–100H₀–200N₀–800–20S₀–1 (then P₀–1) species. The tolerance was set to 0.3 ppm, RDBE from 0 to 80 and H/C ratio from 0 to 3. To exclude media components and artefacts, masses from analytical blanks and culture medium were subtracted from the sample MS scan for each extract in both ionization modes (“Xpose” mode with a ratio of 5).

3. Results and discussion

3.1. Feature determination and molecular networking

Dichloromethane / ethyl acetate (DE) and methanol (M) extracts from the strain *S. cerevisiae* S288c, together with culture medium control extracts, have been analyzed by UHPLC-MS/MS and Direct Injection FTICR-MS. Fig. 1 summarizes the entire analysis and interpretation workflow. Figure S1 shows the base peak ion chromatogram (BPC for DE and M extracts in positive and negative ionization mode). UHPLC-MS/MS data were processed within MZMine2. During the automated data treatment, background noise features, blank features, in-source fragments, adducts and dimers were discarded. Yeast extract data and control sample data were aligned. The feature tables for each ionization mode were exported to build molecular networks on GNPS. Alignment and molecular networking allowed the removal of features exclusively found in culture medium extracts. The medium components present in networks are kept as they can help us to decipher yeast extract component that are linked to molecular

networks, requiring the investigation of these additional networks. It is important to note that primary metabolites (e.g. amino-acids) are in the culture medium, and often cluster with structurally similar yeast metabolites. The feature tables and the molecular networks are the first form of results needed to begin data interpretation. Manual interpretation of result lists is a crucial step to confidently annotate the yeast extract compounds.

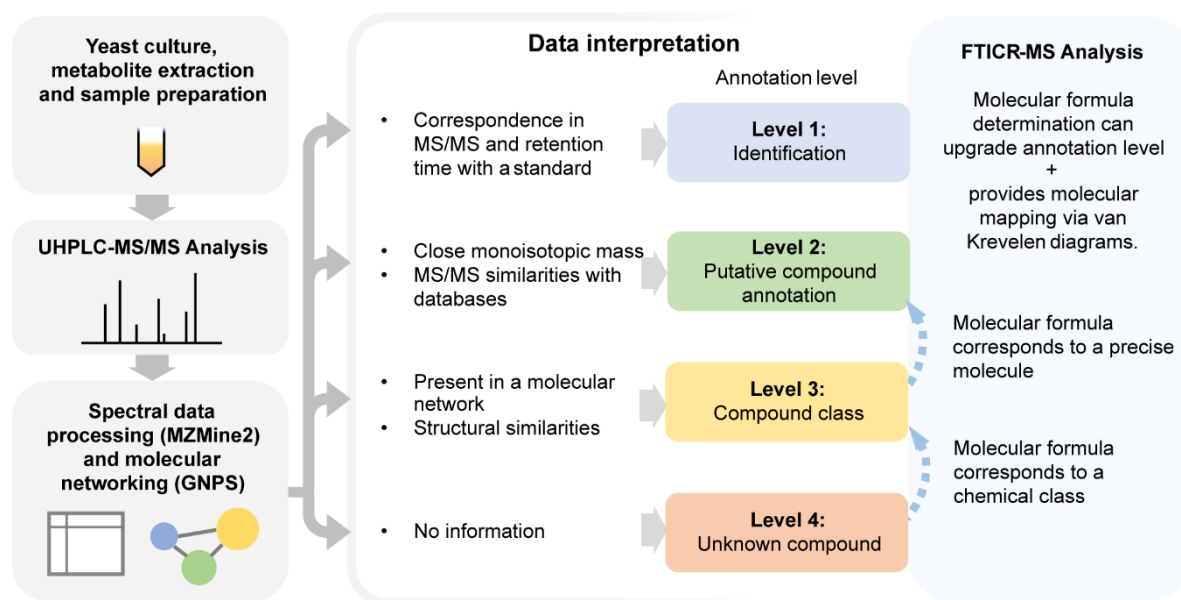


Fig. 1. Summary diagram of the yeast extract analysis workflow with the different annotation levels. Metabolites are extracted from yeast cultures to prepare the samples that are analyzed through UHPLC-MS/MS. Spectral data are processed and used to build molecular networks. Annotations are attributed and can be upgraded thanks to FTICR-MS analysis.

Annotations were attributed via database comparison and were supported by molecular network investigation. For instance, a feature was detected at m/z 231.1126 in positive mode, RT = 7.98 min. This node was included in a large molecular network, where indole alkaloids have already been annotated. Molecular networks built on GNPS displayed structurally related nodes. By comparing MS/MS spectra with tryptophol MS/MS spectrum and with MS/MS databases (e.g. HMDB), we could lead to an annotation that fitted with the molecular formula of the medium feature at m/z 217.0974: 1,2,3,4-Tetrahydro-beta-carboline-3-carboxylic acid (Figure S2a). As this node was a crossroad of two other nodes, we used it to determine hypotheses for m/z 190.0864 and m/z 231.1126, attributed respectively 3-Indolepropionic acid and 1-Methyl-1,2,3,4-tetrahydro--carboline-3-carboxylic acid. These annotations corroborate of an indole alkaloid network (Fig. 2a). As compound hypotheses could

be given, the annotation level is 2 for these compounds. Molecular networking permits the annotation spread and accelerates the whole annotation process. 25 level 2 annotations were provided thanks to progressing through molecular networks (21 in positive ionization mode and 4 in negative). Figure S3 shows all molecular networks for *S. cerevisiae* analysis including annotations after manual investigation. We observed that positive ionization mode led to bigger and more complex molecular networks. This is due to the structural families of the detected compounds, which led to more informative MS/MS spectra, therefore more networks. The largest network is related to alkaloids, as mentioned before, with an indole part and a quinoline part. Another small network corresponds to sulfated amino-acids. The last one clusters only medium and is also found in negative mode, alongside small networks of sulfated amino-acids, tryptophan derivatives and alkaloids.

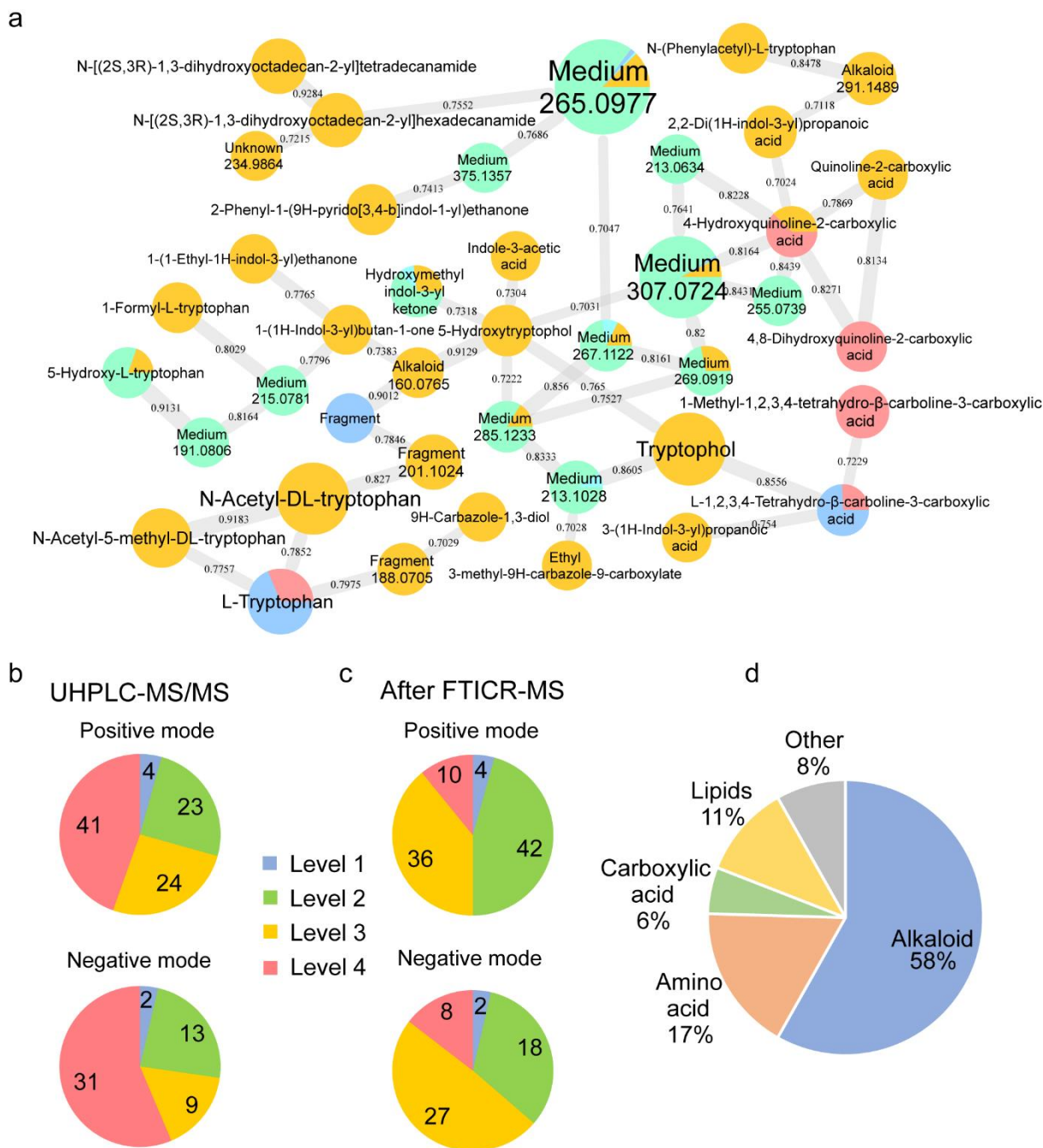


Fig. 2. Results from UHPLC-MS/MS-molecular networking workflow on *Saccharomyces cerevisiae* extracts. *a*: example of alkaloid network built in GNPS (positive mode). Color is the sample origin (Orange: DE; Red: M; Green and Blue: culture medium). Node labels display the proposed metabolite, or the compound class with the m/z ratio. Node size is proportional to intensity, edge size to cosine score. *b*: features from the analysis by UHPLC-MS/MS. The number of features is displayed after a comparison with databases and GNPS annotations. *c*: feature number after FTICR-MS analysis and deeper investigation step. *d*: compound classes pie chart based on UHPLC-MS/MS analysis enhanced by FTICR-MS. Compounds from DE extract and M extract analyzed in positive and negative mode are present.

To assess the confidence level of the annotations, we considered the metabolomics annotation level system [29]. The compounds may be given putative annotations on the basis of molecular formula, database search and molecular networking. Level 1 are identifications allowed by retention time and MS/MS correlation between the studied feature and a pure standard. In our work, six pure standards were compared and six features were identified as level 1 (Table 1). Structural similarities associated to database match led to level 2 annotations. When structural similarities were characterized between an unknown compound and a putatively annotated compound, molecular networking was incredibly useful to get level 3 annotations. Finally, all compounds that cannot be associated to known molecules are considered level 4 or unknown compounds. In parallel, additional data treatment and bibliographical search was done to increase the molecular network information. Fig. 2b summarizes the number of annotated features for each level after the first manual data deciphering that led to annotations for 55 % of the 92 features in positive mode and for 44 % of the 55 features in negative, based on UHPLC-MS/MS analyses.

3.2. Contribution of FTICR-MS

The direct injection ESI-FTICR analyses, yielded an average mass accuracy of 0.08 ppm. After analysis subtraction, 1676 ions have been detected from positive ion spectra of DE extract, and 794 in negative ion spectra, with signal to noise of 9. M extract yielded 1604 ions in positive ion spectra, and 1471 in negative. From which unique molecular formulae have been attributed considering parameters told in Material and methods section (603 for DE POS, 403 for DE NEG, 404 for M POS, 304 for M NEG). Isotopic fine structure was used to confirm the presence of specific elements. In a few cases in which 2 possible molecular formulae are proposed, manual control was used to discard wrong attribution (keeping the formulae with the best accuracy in ppm, or discarding aberrant formulae with O/C ratio > 1 or formulae without enough H or too much N for the number of C).

The data from the two techniques were manually linked by searching corresponding m/z between UHPLC-MS/MS data and FTICR-MS data. Some molecular formulae were not attributed in UHPLC-MS/MS due to the lower time-of-flight precision. The highest FTICR mass measure precision was able to enhance molecular formulae determination. Therefore, more than 91 % of the features already detected by UHPLC-MS/MS could be linked to an accurate molecular formula in positive mode, and 96 % in negative mode. The compounds whose formulae were not retrieved tend to level 4 annotations and remain unknown. These values prove the efficacy of analyzing yeast extracts with UHPLC-MS/MS

on a side and FTICR-MS on the other. It should be pointed out that compared to the UHPLC-MS/MS analysis the DI FTICR analysis yielded a large amount of sodium and potassium adduct due to the lack of chromatographic separation.

It can be shown that the manual linking of FTICR-MS molecular formulae to UHPLC-MS/MS features enhanced significantly the quality of the obtained metabolomics results. The major interest is revealed when the chromatographic workflow did not lead to a precise annotation. The lower precision of TOF analysis, with errors around 2–5 ppm, could lead to several possible molecular formulae, from two to more than seven for high m/z ratio compounds, and then prevent the feature annotation. FTICR helped to discriminate the right formula. Level 2 annotations were confirmed if the given formula corresponds to the compound hypothesis. Some level 3 compounds had a chemical class attributed (e.g. alkaloid) thanks to molecular networking and common fragments, and FTICR-MS provided the missing molecular formula that allowed, combined with MS/MS data, a level 2 annotation. Level 4 unknown compounds at least have a formula. An empiric example of a hypothesis ambiguity was a feature detected at RT 12.76 min (Figure S4). The FTICRMS data allowed us to determine the molecular formula among the three possible ones, confirming the Paspalic acid ($C_{16}H_{16}N_2O_3$) hypothesis. The enhancement of the deeper manual interpretation after FTICR-MS is shown in Fig. 2c. Thus, 89 % of annotations is reached in positive mode and 87 % in negative mode. FTICR-MS enhanced previous results as a fast and complementary technique. In fact, the separation based on LC-MS methods has the advantage on isomer identification and matrix ion suppression effects reduction whereas the direct FTICR injection has the advantage on mass measure accuracy, leading to unique molecular formula determination. The critical annotation step is therefore improved. FTICR-MS also have the advantage of the sensitivity and dynamic range that allows the observation of features that could not be detected in UHPLC-MS/MS.

Based on FTICR-MS data, global metabolic profiling was established for *S. cerevisiae* extracts. Molecular formulae were automatically attributed during data treatment, then O/C ratios were plotted against H/C ratios to create van Krevelen diagrams. These diagrams are bidimensional graphical representations of the global elemental composition of each sample. The van Krevelen diagrams presented in Fig. 3a depicts a qualitative description of the four extracts, that is to say a snapshot of what *S. cerevisiae* can produce when cultivated on the studied medium. The dot color testifies the chemical composition in C, H, O, N and S of the determined formulae, so the compositional differences between extracts may be examined. In Figure S5, blue dots represent the common masses detected both by FTICR-MS and LC-MS/MS. Thus, more features are observed in FTICR-MS analyses, which enhanced the metabolite coverage of the study by its sensitivity and dynamic range. Namely, 603 and

404 features were detected in positive ion mode (respectively in DE extract and in M extract) and 404 and 304 features in negative ion mode. The accuracy of this MS technology reveals the richness of metabolite classes within the yeast extracts. By comparing our data with previously described LC-MS/MS data (Figure S6), alkaloids, lipids and amino-acids were placed. Van Krevelen diagrams from the literature [21,32] allowed to confirm lipid region and amino-acid region, and to locate small acids. Compared to these papers, a very few compounds of the “oxy-aromatics” region were detected.

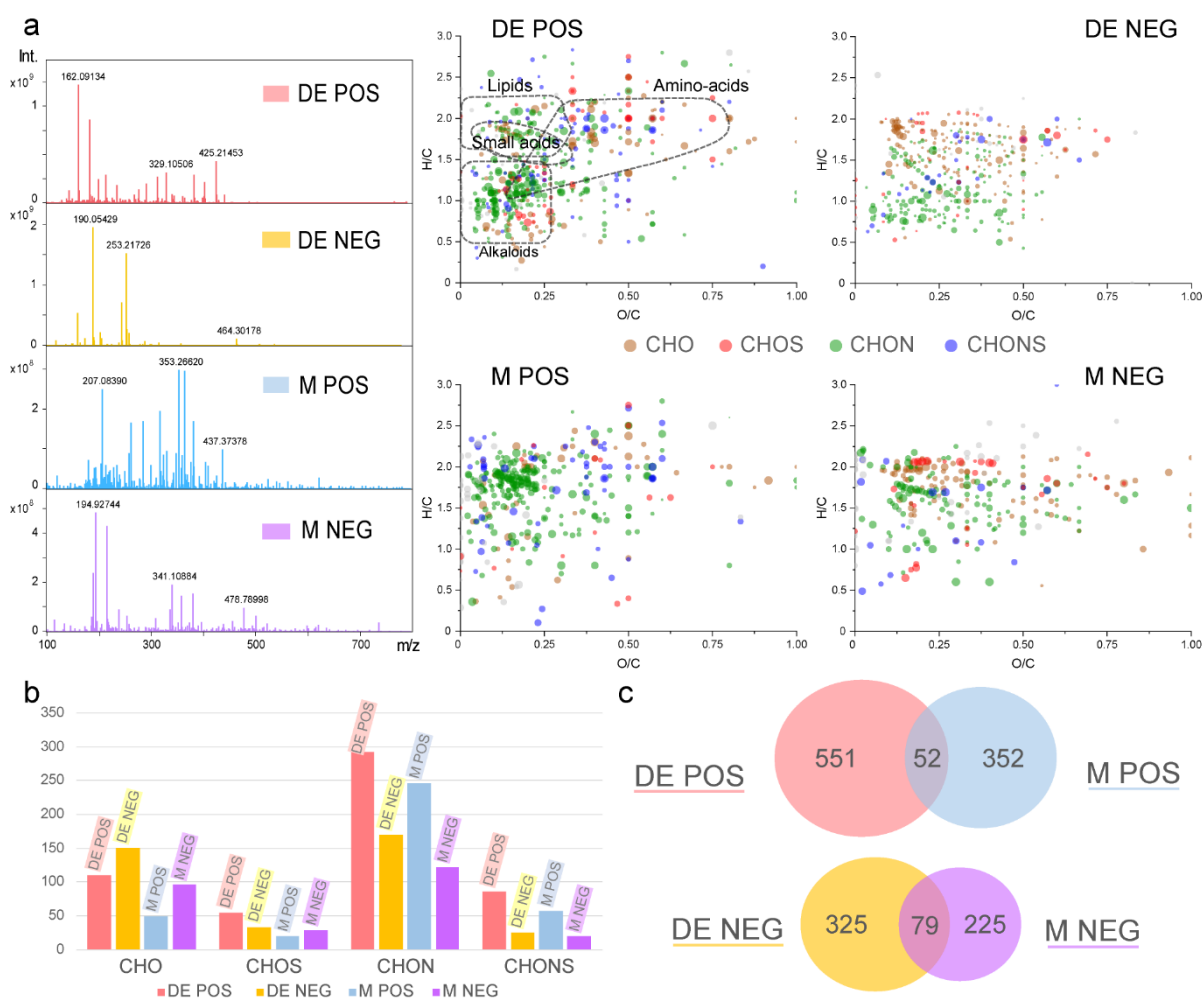


Fig. 3. Representations of FTICR-MS data from *Saccharomyces cerevisiae* extract analysis after chemical formula attributions. *a*: FTICR-MS spectra and corresponding O/C vs. H/C van Krevelen diagrams. Brown dots are CHO features; red dots are CHOS; green dots are CHON; blue dots are CHONS. *b*: histograms of FTICR-MS data distribution among chemical composition types (four most numerous classes) in number of features. Red bars are for DE extract in POS ionization mode; yellow bars are for DE NEG; blue bars are for M POS; purple bars are for M NEG. *c*: Venn diagrams comparing attribution number between DE extracts and M extracts, in positive and negative ionization modes.

The *S. cerevisiae* extracts appear to have a highly diverse content of alkaloids of diverse polarities, as it was seen according to LC MS/MS results. Many were detected in the DE extract. Alkaloids are often found in non-polar extracts, but in this case, there is a lot of CHON and CHONS molecules in the methanol extract too. These alkaloids present in the M extract are more polar than average and may contain one or several hydroxyl groups. A lot of CHON signals are found in the M POS van Krevelen diagram, in an area classically dedicated to lipids and non-polar compounds. These could be saturated alkaloids or amino-acids derivatives with hydroxyls for example. Van Krevelen diagrams were built to discriminate nitrogen-rich features from others (Fig. S7): the bluer the dot, the more nitrogen-rich is the formula. We can see compounds with a lot of nitrogen, confirming an overlay between the alkaloid region and lipid region (indicated on Fig. 3a). The methanol extracted a lot of alkaloids with a high H/C ratio. Van Krevelen diagrams are a useful tool for the comparison and the differentiation of the different yeast extracts, here from the same strain, but it could be used to compare extracts from different strains or species.

The histograms shown in Fig. 3b represent the distribution of masses among chemical compositions (CHO, CHOS, CHON and CHONS). The most represented composition is CHON, which corresponds to the numerous alkaloids and amino-acid derivatives found in the samples, followed by CHO, CHOS and CHONS, and far ahead from CH, CHNS and phosphorus type molecules. The fermentation ability of yeast explains the prominence of CHON and CHONS chemical space. Data were compared to build Venn diagrams (Fig. 3c), comparing for each polarity the presence of detected masses between both DE and M extracts. As in LC–MS/MS data, DE extract analyses provided more features than M extract analyses, which can be attributable to the sequence of extraction (DE comes first and extracts more metabolites). Positive ionization mode also led to more detected masses. Ions that are common between DE and M extract are found in the intersection of the two circles. The amount of common masses is significantly smaller than each extract exclusive species. This is explained by the polarity difference of the used solvent mixtures, and by the order of the extractions (the culture is extracted by dichloromethane / ethyl acetate, then by methanol). The methanol extraction picked some metabolites that dichloromethane / ethyl acetate did not. It confirms the value of using two different extractions to expand the metabolite coverage.

3.3. *Saccharomyces cerevisiae* secondary metabolites

The annotated compounds are listed in Table 1, together with their annotation level, the origin of the extract and the ionization polarity. We also summarized in Fig. 2d the proportions of the different substance types. All the detected lipids may be part of the membrane system remaining after the solvent based extraction. Except for glycerophosphoryl choline, they are all detected in the DE extract, the most non-polar extract. Amino-acids may come from strain primary metabolism, but are also derived compounds such as N-Acetyl-methyltryptophan or Kynurenic acid. "Other" compile diverse families, such as sugars. Alkaloids are a major family part of the small molecules of this strain, in the form of indole alkaloids or quinoline alkaloids. All these compounds are presumably secondary metabolites produced by the yeast, such as Paspalic acid (Peak ID 69), which is an indolo-isoprenic alkaloid identified in *Claviceps purpurea* [33] and used in the hemisynthesis of therapeutic alkaloids. The (2S,5R)-5-(1H-Indol-3-yl)-4-oxo-2,3,5,6-tetrahydro1H-azepino[4,5-b]indole-2-carboxylic acid (Peak ID 75, also called Malassezindole A) is an indole alkaloid involved in the tryptophan metabolism [34]. Some other features have been sparsely described like Ethyl 3-methyl-9H-carbazole-9-carboxylate (Peak ID 62) and 5-(3-Methyl-2-butenyl)-1H-indole-3-carboxylic acid (Peak ID 84). To the best of our knowledge, this is the first time that these compounds have been described in *S. cerevisiae*. It is important to note that focused structural studies, such as NMR experiments or MS experiments with standards, are necessary to bring structural evidences and convert putative annotations into identifications. The possibility of several isomers and isobars still exists. However, we obtained evidence of the presence of many unexpected compounds so far, such as the indole and quinoline alkaloids described here for the first time in this species.

Previous studies have implemented gas chromatography to analyze *S. cerevisiae* secondary metabolites [7,8]. Compared to this work, the *S. cerevisiae* strain was cultivated in a solid medium, the extraction method started with a dichloromethane / ethyl acetate mixture to obtain a moderately polar fraction followed by a second extraction with methanol. It was expected to find more diversified chemical classes, the method was rapid and presented a good level of annotation confidence. The GC-MS analysis by Al-Jassani et al. [7] led to non-polar and volatile compound determination such as terpenoids, whereas we characterized more polar or averagely polar compounds. The two studies appear to be complementary, the GC-MS workflow determined the volatile molecules synthesized during fermentation, while we analyzed intracellular and excreted secondary metabolites. The differences may also be explained by the difference between the strains, the difference of culture media or the sample derivatization. Palomino-Schätzlein et al. [10] have profiled *S. cerevisiae* global metabolism by NMR. Apart from the analytical technique, the differences in the used strains, the

culture media or the extraction method are the reason why we determined different secondary metabolites and less molecules from primary metabolism, such as nucleotides and all the amino-acids. The same observation is done by comparing the work by Farrés et al. [9] who studied *S. cerevisiae* metabolism by LC-MS; the primary amino-acids L-Tryptophan or L-Methionine are also found in our work but not displayed as they were found in the culture medium too.

The methodology described here, summarized in Fig. 1, is very versatile and could also be applied to many types of biological extracts. Globally, more features were described by positive ionization mode analysis, and more masses were detected in the first extract (DE extract), whether by visualizing FTICR-MS results or UHPLC-MS/MS molecular networks. The solvent mixture permitted to extract a wide range of polar, non-polar and averagely polar metabolites. The extraction step is crucial and has to be optimized for every type of analyzed sample. Treatment steps are also essential; the optimization of process parameters is a crucial point to care about for each analysis batch. Finally, decisive steps in this kind of study are manual investigation and data interpretation after the automated processes.

Table 1: Proposed metabolites from *Saccharomyces cerevisiae* analysis. 128 compounds from DE extract and M extract analysed in positive and negative mode are present. Concerning features detected in both ionisation modes, the observed m/z value comes from the UHPLC-MS/MS analysis in positive (POS) and/or negative mode (NEG).

Peak ID	RT (min)	Extract	Proposed metabolite	Type of ion	Observed m/z	Molecular formula	Annotation level
1	1.07	M	Unknown	Positive	381.0779	Unknown	4
2	1.07	M	Choline glycerophosphate	M + H	258.1101	C ₈ H ₂₀ NO ₆ P	2
3	1.10	DE	Alkylamine	M + H	102.1283	C ₆ H ₁₅ N	3
4	1.13	M	Unknown	M + H	365.1074	C ₁₈ H ₂₀ O ₆ S	4
5	1.13	M	β-D-Glucopyranose-3-amino-3-deoxy-β-D-glucopyranose (1:1)	M + H	360.1511	C ₁₂ H ₂₅ NO ₁₁	2
6	1.13	DE	Amino acid	M + H	176.0911	C ₇ H ₁₃ NO ₄	3
7	1.24	DE	N-Acetylmethionine sulfoxide	M + H M - H	208.0637 206.0485	C ₇ H ₁₃ NO ₄ S	2
8	1.28	M	Unknown	Positive	308.0913	Unknown	4
9	1.75	DE; M	Amino-acid	M + H	208.0636	C ₇ H ₁₃ NO ₄ S	3
10	3.45	DE	Lipids	M - H	145.0503	C ₆ H ₁₀ O ₄	3
11	5.48	DE	Unknown	Negative	327.0581	Unknown	4

12	5.48	DE	2-Amino-3-hydroxybenzoic acid	M + H	154.0504	C ₇ H ₇ NO ₃	2
13	5.58	DE	Indole-3-acetaldehyde	M + H	160.0760	C ₁₀ H ₉ NO	2
14	6.07	DE	Carboxylic acid	M – H	117.0550	C ₅ H ₁₀ O ₃	3
15	6.66	DE	Hydroxylated sulfated ester	M – H	193.0532	C ₇ H ₁₄ O ₄ S	3
16	6.70	DE	Carboxylic acid	M – H	161.0814	C ₇ H ₁₄ O ₄	3
17	6.86	DE	Unknown	Positive	293.9688	Unknown	4
18	6.99	DE; M	N-Acetyl-L-methionine	M + H M – H	192.0693 190.0539	C ₇ H ₁₃ NO ₃ S	2
19	7.04	M	4,8-Dihydroxyquinoline-2-carboxylic acid	M + H M – H	206.0446 204.0298	C ₁₀ H ₇ NO ₄	2
20	7.37	DE	Carboxylic acid	M – H	194.0445	C ₉ H ₉ NO ₄	3
21	7.40	DE M	4-Hydroxyquinoline-2-carboxylic acid	M + H M – H	190.0503 188.0356	C ₁₀ H ₇ NO ₃	1
22	7.44	DE M	Carboxylic acid	M – H	175.0609	C ₇ H ₁₂ O ₅	3
23	7.98	M	1-Methyl-1,2,3,4-tetrahydro-β-carboline-3-carboxylic acid	M + H M – H	231.1126 229.0974	C ₁₃ H ₁₄ N ₂ O ₂	2
24	8.12	DE	2-Benzamidoacetic acid	M – H	178.0500	C ₉ H ₉ NO ₃	2
25	8.21	DE	2-Isopropylmaleic acid	M – H	157.0502	C ₇ H ₁₀ O ₄	2
26	8.24	DE	5-Hydroxytryptophol	M + H	178.0864	C ₁₀ H ₁₁ NO ₂	2
27	8.24	DE M	N-Propionylmethionine	M + H M – H	206.0848 204.0693	C ₈ H ₁₅ NO ₃ S	2
28	8.40	DE	Carboxylic acid	M – H	131.0709	C ₆ H ₁₂ O ₃	3
29	8.71	DE	Alkaloid	M + H	160.0765	C ₁₀ H ₉ NO	3
30	8.84	DE	N-Acetyl-DL-leucine	M – H	172.0973	C ₈ H ₁₅ NO ₃	2
31	8.86	DE	(1H-Indol-3-yl)methanol	M + H	148.0761	C ₉ H ₉ NO	2
32	8.91	DE	Alkaloid	M + H	233.0925	C ₁₂ H ₁₂ N ₂ O ₃	3
33	8.91	DE	Alkaloid	M + H M – H	251.1023 249.0865	C ₁₂ H ₁₄ N ₂ O ₄	3
34	9.11	DE	Alkylphenylketone	M – H	209.0826	C ₁₁ H ₁₄ O ₄	3
35	9.23	DE	3-Phenyllactic acid	M – H	165.0554	C ₉ H ₁₀ O ₃	1
36	9.39	DE	Amino-acid	M + H	166.0867	C ₉ H ₁₁ NO ₂	3
37	9.45	DE	N-Acetyl-L-phenylalanine	M – H	206.0808	C ₁₁ H ₁₃ NO ₃	1
38	9.69	DE	1-Formyl-L-tryptophan	M + H	233.0922	C ₁₂ H ₁₂ N ₂ O ₃	2
39	9.72	DE	2-Hydroxy-3-(1H-indol-3-yl)propanoic acid	M – H	204.0665	C ₁₁ H ₁₁ NO ₃	2

40	9.81	DE	N-Acetyl-DL-tryptophan	M + H M - H	247.1080 245.0921	C ₁₃ H ₁₄ N ₂ O ₃	2
41	9.86	DE	Unknown	M - H	277.0816	C ₁₃ H ₁₄ N ₂ O ₅	4
42	9.95	DE	1H-Indazole-3-carbaldehyde	M + H	147.0555	C ₈ H ₆ N ₂ O	2
43	10.51	DE	Alkaloid	M - H	220.0965	C ₁₂ H ₁₅ NO ₃	3
44	10.57	DE	Alkaloid	M + H	291.1488	C ₁₉ H ₁₈ N ₂ O	3
45	10.57	DE	Indole-3-acetic acid	M + H M - H	176.0705 174.0557	C ₁₀ H ₉ NO ₂	1
46	10.57	DE	Tryptophol	M + H	162.0920	C ₁₀ H ₁₁ NO	1
47	10.57	DE	Unknown	M + H	319.1443	C ₂₀ H ₁₈ N ₂ O ₂	4
48	10.63	DE	Alkaloid	M + H	335.1385	C ₂₀ H ₁₈ N ₂ O ₃	3
49	10.65	DE	Sugar	M - H	327.0957	C ₁₀ H ₂₁ N ₂ O ₈ P	3
50	10.71	DE	N-Acetyl-5-methyl-DL-tryptophan	M + H M - H	261.1233 259.1078	C ₁₄ H ₁₆ N ₂ O ₃	2
51	10.89	DE	Unknown	M - H	337.1240	C ₁₆ H ₂₂ N ₂ O ₄ S	4
52	10.98	DE	3-Acetylindole	M + H	160.0764	C ₁₀ H ₉ NO	1
53	11.12	DE	N-(5-Methyl-3-oxohexyl)alanine	M - H	200.1283	C ₁₀ H ₁₉ NO ₃	2
54	11.18	DE	Amino-acid	M + H M - H	234.1155 232.1000	C ₁₀ H ₁₉ NO ₃ S	3
55	11.19	DE	Alkaloid	M + H	188.1091	C ₁₂ H ₁₃ NO	3
56	11.32	DE	3-(1H-Indol-3-yl)propanoic acid	M + H	190.0864	C ₁₁ H ₁₁ NO ₂	2
57	11.57	DE	Unknown	M - H	323.0666	C ₁₇ H ₁₂ N ₂ O ₅	4
58	11.63	DE	Alkaloid	M - H	234.1125	C ₁₃ H ₁₇ NO ₃	3
59	11.66	DE	Amino-acid	M + H M - H	268.1006 266.0849	C ₁₃ H ₁₇ NO ₃ S	3
60	11.74	DE	Alkaloid	M + H	379.1102	C ₂₁ H ₁₈ N ₂ O ₃ S	3
61	11.80	DE	6-Methylquinoline	M + H	144.0807	C ₁₀ H ₉ N	2
62	11.86	DE	Ethyl 3-methyl-9H-carbazole-9-carboxylate	M + H	254.1176	C ₁₆ H ₁₅ NO ₂	2
63	12.21	DE	1,2,3,5-Tetrahydroxy-10H-acridin-9-one	M + H	260.0558	C ₁₃ H ₉ NO ₅	2
64	12.27	DE	Unknown	Positive	234.9864	Unknown	4
65	12.41	DE	Alkaloid	M + H M - H	285.1229 283.1077	C ₁₆ H ₁₆ N ₂ O ₃	3
66	12.48	DE	Alkaloid	M + H	254.1176	C ₁₆ H ₁₅ NO ₂	3
67	12.56	DE	2,4-Dihydroxydodecanoic acid	M - H	231.1583	C ₁₂ H ₂₄ O ₄	2

68	12.76	DE	6-(3-Methylbut-2-enyl)-1H-indole-3-carboxylic acid	M + H	230.1176	C ₁₄ H ₁₅ NO ₂	2
69	12.76	DE	Paspalic acid	M + H	269.1290	C ₁₆ H ₁₆ N ₂ O ₂	2
70	12.95	DE	Alkaloid	M – H	374.1151	C ₂₁ H ₁₇ N ₃ O ₄	3
71	12.97	DE	Methanone, (6-hydroxy-9H-pyrido[3,4-b]indol-1-yl)(5-methoxy-1H-indol-3-yl)-	M + H	358.1188	C ₂₁ H ₁₅ N ₃ O ₃	2
72	13.03	DE	Methyl 2-(1H-indol-3-yl)acetate	M + H	190.0868	C ₁₁ H ₁₁ NO ₂	2
73	13.14	DE	Alkaloid	M + H	188.1075	C ₁₂ H ₁₃ NO	3
74	13.22	DE	Alkaloid	M + H	243.0761	C ₁₃ H ₁₀ N ₂ O ₃	3
75	13.22	DE	(2S,5R)-5-(1H-Indol-3-yl)-4-oxo-2,3,5,6-tetrahydro-1H-azepino[4,5-b]indole-2-carboxylic acid	M + H M – H	360.1339 358.1188	C ₂₁ H ₁₇ N ₃ O ₃	2
76	13.36	DE	Alkaloid	M + H	218.0812	C ₁₂ H ₁₁ NO ₃	3
77	13.36	DE	Alkaloid	M + H	335.1385	C ₂₀ H ₁₈ N ₂ O ₃	3
78	13.39	DE	Alkaloid	M – H	289.1345	C ₁₉ H ₁₈ N ₂ O	3
79	13.42	DE	1-Acetylindole	M + H	160.0767	C ₁₀ H ₉ NO	2
80	13.42	DE	Quinoline-2-carboxylic acid	M + H	174.0557	C ₁₀ H ₇ NO ₂	2
81	13.45	DE	Alkaloid	M – H	303.1130	C ₁₉ H ₁₆ N ₂ O ₂	3
82	13.49	DE	2,2-Di(1H-indol-3-yl)propanoic acid	M + H	305.1292	C ₁₉ H ₁₆ N ₂ O ₂	2
83	13.49	DE	Alkaloid	M + H	407.1971	C ₂₄ H ₂₆ N ₂ O ₄	3
84	13.57	DE	5-(3-Methyl-2-butenyl)-1H-indole-3-carboxylic acid	M + H	230.1180	C ₁₄ H ₁₅ NO ₂	2
85	13.63	DE	Alkaloid	M + H	333.1236	C ₂₀ H ₁₆ N ₂ O ₃	3
86	13.69	DE	Alkaloid	M + H	289.1332	C ₁₉ H ₁₆ N ₂ O	3
87	13.83	DE	Alkaloid	M + H	243.0765	C ₁₃ H ₁₀ N ₂ O ₃	3
88	13.83	DE	Alkaloid	M + H M – H	360.1348 358.1189	C ₂₁ H ₁₇ N ₃ O ₃	3
89	13.89	DE	Unknown	Negative	564.2285	Unknown	4
90	13.89	DE	Unknown	M + H	520.2361	C ₃₆ H ₂₉ N ₃ O	4
91	14.04	DE	Alkaloid	M + H	190.0871	C ₁₁ H ₁₁ NO ₂	3
92	14.31	DE	3,4-Bis(1H-indole-3-yl)-2,5-dihydrofuran-2-one	M + H	315.1129	C ₂₀ H ₁₄ N ₂ O ₂	2
93	14.39	DE	Alkaloid	M + H	188.1071	C ₁₂ H ₁₃ NO	3
94	14.41	M	Tauroursodeoxycholic acid	M + H M – H	500.3040 498.2892	C ₂₆ H ₄₅ NO ₆ S	2

95	14.43	DE	Unknown	M – H	347.1758	C ₂₂ H ₂₄ N ₂ O ₂	4
96	14.51	DE	3-[[4-(3-Carboxyanilino)-3,6-dioxocyclohexa-1,4-dien-1-yl]amino]benzoic acid	M + H	379.0925	C ₂₀ H ₁₄ N ₂ O ₆	2
97	14.68	DE	Alkaloid	M – H	287.1180	C ₁₉ H ₁₆ N ₂ O	3
98	14.72	DE	2-Phenyl-1-(9H-pyrido[3,4-b]indol-1-yl)ethanone	M + H	287.1178	C ₁₉ H ₁₄ N ₂ O	2
99	14.72	DE	Alkaloid	M + H M – H	333.1232 331.1078	C ₂₀ H ₁₆ N ₂ O ₃	3
100	14.86	DE	Alkaloid	M + H M – H	291.1489 289.1337	C ₁₉ H ₁₈ N ₂ O	3
101	14.86	DE	Alkaloid	M + H	305.1288	C ₁₉ H ₁₆ N ₂ O ₂	3
102	14.86	DE	N-(Phenylacetyl)-L-tryptophan	M + H	323.1396	C ₁₉ H ₁₈ N ₂ O ₃	2
103	14.86	DE	Unknown	Positive	310.1227	Unknown	4
104	14.88	DE	Lipids	M – H	259.1902	C ₁₄ H ₂₈ O ₄	3
105	15.01	DE	1-(1H-Indol-3-yl)butan-1-one	M + H	188.1067	C ₁₂ H ₁₃ NO	2
106	15.21	DE	Alkaloid	M + H	315.1133	C ₂₀ H ₁₄ N ₂ O ₂	3
107	15.79	DE	2-Quinolinemethanethiol acetate	M – H	216.0481	C ₁₂ H ₁₁ NOS	2
108	15.82	DE	Unknown	Positive	616.1765	Unknown	4
109	16.07	DE	Unknown	M – H	315.1131	C ₂₀ H ₁₆ N ₂ O ₂	4
110	16.09	DE	9H-Carbazole-1,3-diol	M + H	200.0691	C ₁₂ H ₉ NO ₂	2
111	16.28	DE	Alkaloid	M – H	359.1389	C ₂₂ H ₂₀ N ₂ O ₃	3
112	16.34	DE	2-Methylquinoline	M + H	144.0812	C ₁₀ H ₉ N	2
113	16.49	DE	(2S,3R)-2-aminoctadecane-1,3-diol	M + H	302.3047	C ₁₈ H ₃₉ NO ₂	2
114	16.57	DE	Alkaloid	M + H	289.1330	C ₁₉ H ₁₆ N ₂ O	3
115	16.79	DE	Alkaloid	M + H	188.1070	C ₁₂ H ₁₃ NO	3
116	17.17	DE	Polycyclic alcohol	M – H	253.1801	C ₁₅ H ₂₆ O ₃	3
117	17.52	DE	1-(1-Ethyl-1H-indol-3-yl)-ethanone	M + H	188.1073	C ₁₂ H ₁₃ NO	2
118	18.04	DE	Lipids	M + H	330.3366	C ₂₀ H ₄₃ NO ₂	3
119	20.21	DE	Unknown	M – H	342.2635	C ₁₉ H ₃₇ NO ₄	4
120	20.58	DE	Alkene	M + H	205.1944	C ₁₅ H ₂₄	3
121	21.09	DE	Amino-acid	M – H	357.1886	C ₁₄ H ₂₆ N ₆ O ₅	3
122	21.27	DE	Unknown	Positive	343.2023	Unknown	4
123	21.37	DE	Lipids	M + H M – H	255.2309 253.2163	C ₁₆ H ₃₀ O ₂	3

124	21.44	DE	Lipids	M + H	319.2843	C ₁₈ H ₃₈ O ₄	3
125	23.00	DE	N-[(2S,3R)-1,3-dihydroxyoctadecan-2-yl]dodecanamide	M + H	484.4716	C ₃₀ H ₆₁ NO ₃	2
126	23.85	DE	N-[(2S,3R)-1,3-dihydroxyoctadecan-2-yl]tetradecanamide	M + H	512.5035	C ₃₂ H ₆₅ NO ₃	2
127	24.63	DE	N-[(2S,3R)-1,3-dihydroxyoctadecan-2-yl]hexadecanamide	M + H	540.5356	C ₃₄ H ₆₉ NO ₃	2
128	25.51	DE	N-[(2S,3R)-1,3-dihydroxyoctadecan-2-yl]octadecanamide	M + H	568.5648	C ₃₆ H ₇₃ NO ₃	2

4. Conclusion

Throughout this study, *Saccharomyces cerevisiae* S288c strain cultures were extracted then dereplicated via a combination of mass spectrometry techniques. UHPLC-MS/MS-feature-based molecular networking provides accurate and reliable annotations, with the advantage of ionization suppression reduction and discrimination of isomers. In addition, FTICR-MS represents a valuable technology to complement our results by throwing away doubts within compound annotation, by means of accurate molecular formula determination. FTICR-MS data also brought another dimension to the metabolite mapping through a global molecular representation.

The most important challenge in most of metabolomics studies is the spectral data interpretation. Thanks to the large polarity extraction range, the data processing steps and the multiple available databases, 128 *S. cerevisiae* exclusive features were detected, with 55 % of them annotated in positive ion mode, and 44 % in negative (level 1, 2 and 3 annotations). The annotation propagation was permitted by the molecular networks that were built, with bigger networks in positive ion mode. A rich molecular network was obtained for the indole and quinoline-type alkaloids.

The direct injection FTICR-MS analysis detected many more features, due to its great sensitivity. It also largely enhanced the annotation levels of the UHPLC-MS/MS analysis, going up to 89 % annotations in

positive ion mode, and to 87 % in negative ion mode. This analysis brought van Krevelen diagrams as global metabolic profiling representations and demonstrating the richness of the yeast extracts in term of metabolites.

Among the annotated features, 58 % of alkaloids were found. Our work revealed many alkaloids produced by this strain that could expand fungal databases. The overall representation shows the many alkaloids and amino-acid derived molecules that *S. cerevisiae* can produce when cultivated on the studied medium.

As for *S. cerevisiae* strain S288c, our work provided an effective mapping of compounds by complementing previous studies. This methodology represents the laid foundations for a fast exploration of the metabolites from other yeast strains or species, and can also be applied to various biological samples, as plants or soils.

Availability of data and material

The metabolomics and metadata reported in this paper are available via MetaboLights [35] on www.ebi.ac.uk/metabolights/ MTBLS1701 study identifier MTBLS1701.

Funding

This work was partially supported by Normandie Université (NU), the Centre National de la Recherche Scientifique (CNRS), Université de Rouen Normandie (URN), INSA Rouen Normandie; the European Regional Development Fund (ERDF) N° HN0001343, the European Union's Horizon 2020 Research Infrastructures program (Grant Agreement 731077), the Région Normandie, and the Laboratoire d'Excellence (LabEx) SynOrg (ANR-11-LABX-0029), France.

CRedit authorship contribution statement

Olivier Perruchon: Methodology, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Isabelle Schmitz-Afonso**: Conceptualization, Methodology, Investigation, Resources, Writing - review & editing. **Cécile Grondin**: Conceptualization, Resources,

Writing - review & editing. **Serge Casaregola**: Conceptualization, Resources, Writing - review & editing. **Carlos Afonso**: Conceptualization, Investigation, Resources, Writing - review & editing, Funding acquisition. **Abdelhakim Elomri**: Conceptualization, Methodology, Investigation, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no conflict of interest.

Acknowledgements

Access to a CNRS FTICR research infrastructure (FR3624) is gratefully acknowledged. Authors would like to thank the CIRM-Levures (Micalis Institute, INRAE, Université Paris-Saclay) for the yeast strain cultures and scientific partnership; the COBRA team for all the support.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi: <https://doi.org/10.1016/j.jpba.2020.113857>.

References

- [1] C.P. Kurtzman, J.W. Fell, T. Boekhout, *The Yeasts: A Taxonomic Study*, 5th edition, 2011.
- [2] A. Goffeau, B.G. Barrell, H. Bussey, R.W. Davis, B. Dujon, H. Feldmann, F. Galibert, J.D. Hoheisel, C. Jacq, M. Johnston, E.J. Louis, H.W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, S.G. Oliver, *Life with 6000 genes*, *Science* (80-.) 274 (1996) 546–567, <http://dx.doi.org/10.1126/science.274.5287.546>.
- [3] J.L. Legras, D. Merdinoglu, J.M. Cornuet, F. Karst, *Bread, beer and wine: Saccharomyces cerevisiae diversity reflects human history*, *Mol. Ecol.* 16 (2007) 2091–2102, <http://dx.doi.org/10.1111/j.1365-294X.2007.03266.x>.
- [4] J. Nielsen, *Production of biopharmaceutical proteins by yeast*, *Bioengineered.* 4 (2013) 207–211, <http://dx.doi.org/10.4161/bioe.22856>.

- [5] W. Jilg, M. Schmidt, G. Zoulek, B. Lorbeer, B. Wilske, F. Deinhardt, *Clinical evaluation of a recombinant hepatitis B vaccine*, *Lancet*. 324 (1984) 1174–1175, [http://dx.doi.org/10.1016/S0140-6736\(84\)92740-5](http://dx.doi.org/10.1016/S0140-6736(84)92740-5).
- [6] K.J. Hofmann, J.C. Cook, J.G. Joyce, D.R. Brown, L.D. Schultz, H.A. George, M. Rosolowsky, K.H. Fife, K.U. Jansen, *Sequence determination of human papillomavirus type 6a and assembly of virus-like particles in Saccharomyces cerevisiae*, *Virology*. 209 (1995) 506–518, <http://dx.doi.org/10.1006/viro.1995.1283>.
- [7] M.J. Al-Jassani, G.J. Mohammed, I.H. Hameed, *Secondary metabolites analysis of Saccharomyces cerevisiae and evaluation of antibacterial activity*, *Int. J. Pharm. Clin. Res.* 8 (2016) 303–314.
- [8] A. Tejero Rioseras, D. Garcia Gomez, B.E. Ebert, L.M. Blank, A.J. Ibáñez, P.M.L. Sinues, *Comprehensive real-time analysis of the yeast volatilome*, *Sci. Rep.* 7 (2017) 14236, <http://dx.doi.org/10.1038/s41598-017-14554-y>.
- [9] M. Farrés, B. Pina, R. Tauler, *Chemometric evaluation of Saccharomyces cerevisiae metabolic profiles using LC–MS*, *Metabolomics*. 11 (2014) 210–224, <http://dx.doi.org/10.1007/s11306-014-0689-z>.
- [10] M. Palomino-Schätzlein, M.M. Molina-Navarro, M. Tormos-Pérez, S. Rodríguez-Navarro, A. Pineda-Lucena, *Optimised protocols for the metabolic profiling of S. Cerevisiae by 1H-NMR and HRMAS spectroscopy*, *Anal. Bioanal. Chem.* 405 (2013) 8431–8441, <http://dx.doi.org/10.1007/s00216-013-7271-9>.
- [11] T. Kind, H. Tsugawa, T. Cajka, Y. Ma, Z. Lai, S.S. Mehta, G. Wohlgemuth, D.K. Barupal, M.R. Showalter, M. Arita, O. Fiehn, *Identification of small molecules using accurate mass MS/MS search*, *Mass Spectrom. Rev.* 37 (2018) 513–532, <http://dx.doi.org/10.1002/mas.21535>.
- [12] J.L. Wolfender, J.M. Nuzillard, J.J.J. Van Der Hooft, J.H. Renault, S. Bertrand, *Accelerating Metabolite Identification in Natural Product Research: Toward an Ideal Combination of Liquid Chromatography-High-Resolution Tandem Mass Spectrometry and NMR Profiling, in Silico Databases, and Chemometrics*, *Anal. Chem.* 91 (2019) 704–742, <http://dx.doi.org/10.1021/acs.analchem.8b05112>.
- [13] T. Pluskal, S. Castillo, A. Villar-Briones, M. Orešič, *MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data*, *BMC Bioinformatics* 11 (2010) 395, <http://dx.doi.org/10.1186/1471-2105-11-395>.
- [14] F. Olivon, G. Grelier, F. Roussi, M. Litaudon, D. Touboul, *MZmine 2 data-preprocessing to enhance molecular networking reliability*, *Anal. Chem.* 89 (2017) 7836–7840, <http://dx.doi.org/10.1021/acs.analchem.7b01563>.
- [15] T. Jewison, C. Knox, V. Neveu, Y. Djombou, A.C. Guo, J. Lee, P. Liu, R. Mandal, R. Krishnamurthy, I. Sinelnikov, M. Wilson, D.S. Wishart, *YMDB: The yeast metabolome database*, *Nucleic Acids Res.* 40 (2012) 815–820, <http://dx.doi.org/10.1093/nar/gkr916>.
- [16] M. Ramirez-Gaona, A. Marcu, A. Pon, A.C. Guo, T. Sajed, N.A. Wishart, N. Karu, Y.D. Feunang, D. Arndt, D.S. Wishart, *YMDB 2.0: a significantly expanded version of the yeast metabolome database*, *Nucleic Acids Res.* 45 (2017) D440–D445, <http://dx.doi.org/10.1093/nar/gkw1058>.
- [17] D.S. Wishart, D. Tzur, C. Knox, R. Eisner, A.C. Guo, N. Young, D. Cheng, K. Jewell, D. Arndt, S. Sawhney, C. Fung, L. Nikolai, M. Lewis, M.A. Coutouly, I. Forsythe, P. Tang, S. Shrivastava, K. Jeroncic, P. Stothard, G. Amegbey, D. Block, D.D. Hau, J. Wagner, J. Miniaci, M. Clements, M. Gebremedhin, N. Guo, Y. Zhang, G.E. Duggan, G.D. MacInnis, A.M. Weljie, R. Dowlatabadi, F. Bamforth, D. Clive, R. Greiner, L. Li, T. Marrie, B.D. Sykes, H.J. Vogel, L. Querengesser, *HMDB: The human metabolome database*, *Nucleic Acids Res.* 35 (2007) 521–526, <http://dx.doi.org/10.1093/nar/gkl923>.

- [18] H. Horai, M. Arita, S. Kanaya, Y. Nihei, T. Ikeda, K. Suwa, Y. Ojima, K. Tanaka, S. Tanaka, K. Aoshima, Y. Oda, Y. Kakazu, M. Kusano, T. Tohge, F. Matsuda, Y. Sawada, M.Y. Hirai, H. Nakanishi, K. Ikeda, N. Akimoto, T. Maoka, H. Takahashi, T. Ara, N. Sakurai, H. Suzuki, D. Shibata, S. Neumann, T. Iida, K. Tanaka, K. Funatsu, F. Matsuura, T. Soga, R. Taguchi, K. Saito, T. Nishioka, *MassBank: a public repository for sharing mass spectral data for life sciences*, *J. Mass Spectrom.* 45 (2010) 703–714, <http://dx.doi.org/10.1002/jms.1777>.
- [19] T. Kind, O. Fiehn, *Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry*, *BMC Bioinformatics* 8 (2007) 105, <http://dx.doi.org/10.1186/1471-2105-8-105>.
- [20] M. Ghaste, R. Mistrik, V. Shulaev, *Applications of fourier transform ion cyclotron resonance (FT-ICR) and orbitrap based high resolution mass spectrometry in metabolomics and lipidomics*, *Int. J. Mol. Sci.* 17 (2016) 816, <http://dx.doi.org/10.3390/ijms17060816>.
- [21] C. Roullier-Gall, J. Signoret, D. Hemmler, M.A. Witting, B. Kanawati, B. Schäfer, R.D. Gougeon, P. Schmitt-Kopplin, *Usage of FT-ICR-MS metabolomics for characterizing the chemical signatures of barrel-aged whisky*, *Front. Chem.* 6 (2018) 29, <http://dx.doi.org/10.3389/fchem.2018.00029>.
- [22] E. Chekmeneva, G. Dos Santos Correia, M. Gómez-Romero, J. Stamler, Q. Chan, P. Elliott, J.K. Nicholson, E. Holmes, *Ultra-performance liquid chromatography-high-Resolution mass spectrometry and direct infusion-high-Resolution mass spectrometry for combined exploratory and targeted metabolic profiling of human urine*, *J. Proteome Res.* 17 (2018) 3492–3502, <http://dx.doi.org/10.1021/acs.jproteome.8b00413>.
- [23] T. Nagao, D. Yukihiro, Y. Fujimura, K. Saito, K. Takahashi, D. Miura, H. Wariishi, *Power of isotopic fine structure for unambiguous determination of metabolite elemental compositions: In silico evaluation and metabolomic application*, *Anal. Chim. Acta* 813 (2014) 70–76, <http://dx.doi.org/10.1016/j.aca.2014.01.032>.
- [24] D.W. van Krevelen, *Graphical-statistical method for the study of structure and reaction processes of coal*, *Fuel* 29 (1950) 269–284.
- [25] S. Kim, R.W. Kramer, P.G. Hatcher, *Graphical method for analysis of ultrahigh-resolution broadband mass spectra of natural organic matter, the Van krevelen diagram*, *Anal. Chem.* 75 (2003) 5336–5344, <http://dx.doi.org/10.1021/ac034415p>.
- [26] M. Wang, J.J. Carver, V.V. Phelan, L.M. Sanchez, N. Garg, Y. Peng, D.D. Nguyen, J. Watrous, C.A. Kaponov, T. Luzzatto-Knaan, C. Porto, A. Bouslimani, A.V. Melnik, M.J. Meehan, W.T. Liu, M. Crüsemann, P.D. Boudreau, E. Esquenazi, M. Sandoval-Calderón, R.D. Kersten, L.A. Pace, R.A. Quinn, K.R. Duncan, C.C. Hsu, D.J. Floros, R.G. Gavilan, K. Kleigrew, T. Northen, R.J. Dutton, D. Parrot, E.E. Carlson, B. Aigle, C.F. Michelsen, L. Jelsbak, C. Sohlenkamp, P. Pevzner, A. Edlund, J. McLean, J. Piel, B.T. Murphy, L. Gerwick, C.C. Liaw, Y.L. Yang, H.U. Humpf, M. Maansson, R.A. Keyzers, A.C. Sims, A.R. Johnson, A.M. Sidebottom, B.E. Sedio, A. Klitgaard, C.B. Larson, C.A.P. Boya, D. Torres-Mendoza, D.J. Gonzalez, D.B. Silva, L.M. Marques, D.P. Demarque, E. Pociute, E.C. O'Neill, E. Briand, E.J.N. Helfrich, E.A. Granatosky, E. Glukhov, F. Ryffel, H. Houson, H. Mohimani, J.J. Kharbush, Y. Zeng, J.A. Vorholt, K.L. Kurita, P. Charusanti, K.L. McPhail, K.F. Nielsen, L. Vuong, M. Elfeki, M.F. Traxler, N. Engene, N. Koyama, O.B. Vining, R. Baric, R.R. Silva, S.J. Mascuch, S. Tomasi, S. Jenkins, V. Macherla, T. Hoffman, V. Agarwal, P.G. Williams, J. Dai, R. Neupane, J. Gurr, A.M.C. Rodríguez, A. Lamsa, C. Zhang, K. Dorrestein, B.M. Duggan, J. Almaliti, P.M. Allard, P. Phapale, L.F. Nothias, T. Alexandrov, M. Litaudon, J.L. Wolfender, J.E. Kyle, T.O. Metz, T. Peryea, D.T. Nguyen, D. VanLeer, P. Shinn, A. Jadhav, R. Müller, K.M. Waters, W. Shi, X. Liu, L. Zhang, R. Knight, P.R. Jensen, B. Palsson, K. Pogliano, R.G. Linnington, M. Gutiérrez, N.P. Lopes, W.H. Gerwick, B.S. Moore, P.C. Dorrestein, N. Bandeira, *Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking*, *Nat. Biotechnol.* 34 (2016) 828–837, <http://dx.doi.org/10.1038/nbt.3597>.

- [27] J. Watrous, P. Roach, T. Alexandrov, B.S. Heath, J.Y. Yang, R.D. Kersten, M. van der Voort, K. Pogliano, H. Gross, J.M. Raaijmakers, B.S. Moore, J. Laskin, N. Bandeira, P.C. Dorrestein, *Mass spectral molecular networking of living microbial colonies*, Proc. Natl. Acad. Sci. U.S.A. 109 (2012) E1743–E1752, <http://dx.doi.org/10.1073/pnas.1203689109>.
- [28] R.A. Quinn, L.F. Nothias, O. Vining, M. Meehan, E. Esquenazi, P.C. Dorrestein, *Molecular networking As a drug discovery, drug metabolism, and precision medicine strategy*, Trends Pharmacol. Sci. 38 (2017) 143–154, <http://dx.doi.org/10.1016/j.tips.2016.10.011>.
- [29] L.W. Sumner, A. Amberg, D. Barrett, M.H. Beale, R. Beger, C.A. Daykin, T.W.-M. Fan, O. Fiehn, R. Goodacre, J.L. Griffin, T. Hankemeier, N. Hardy, J. Harnly, R. Higashi, J. Kopka, A.N. Lane, J.C. Lindon, P. Marriott, A.W. Nicholls, M.D. Reily, J.J. Thaden, M.R. Viant, *Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI)*, Metabolomics. 3 (2007) 211–221, <http://dx.doi.org/10.1007/s11306-007-0082-2>.
- [30] O.D. Myers, S.J. Sumner, S. Li, S. Barnes, X. Du, *One step forward for reducing false positive and false negative compound identifications from mass spectrometry metabolomics data: new algorithms for constructing extracted ion chromatograms and detecting chromatographic peaks*, Anal. Chem. 89 (2017) 8696–8703, <http://dx.doi.org/10.1021/acs.analchem.7b00947>.
- [31] J. Maillard, J. Ferey, C.P. Rüger, I. Schmitz-Afonso, S. Bekri, T. Gautier, N. Carrasco, C. Afonso, A. Tebani, *Optimization of ion trajectories in a dynamically harmonized Fourier-transform ion cyclotron resonance cell using a design of experiments strategy*, Rapid Commun. Mass Spectrom. 34 (2020), <http://dx.doi.org/10.1002/rcm.8659>.
- [32] R.D. Gougeon, M. Lucio, M. Frommberger, D. Peyron, D. Chassagne, H. Alexandre, F. Feuillat, A. Voilley, P. Cayot, I. Gebefügi, N. Hertkorn, P. Schmitt-Kopplin, *The chemodiversity of wines can reveal a metaboecography expression of cooperage oak wood*, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 9174–9179, <http://dx.doi.org/10.1073/pnas.0901100106>.
- [33] T. Correia, N. Grammel, I. Ortel, U. Keller, P. Tudzynski, *Molecular cloning and analysis of the ergopeptide assembly system in the ergot fungus claviceps purpurea*, Chem. Biol. 10 (2003) 1281–1292, <http://dx.doi.org/10.1016/j.chembiol.2003.11.013>.
- [34] B. Irlinger, A. Bartsch, H.J. Krämer, P. Mayser, W. Steglich, *New tryptophan metabolites from cultures of the lipophilic yeast Malassezia furfur*, Helv. Chim. Acta 88 (2005) 1472–1485, <http://dx.doi.org/10.1002/hlca.200590118>.
- [35] K. Haug, K. Cochrane, V.C. Nainala, M. Williams, J. Chang, K.V. Jayaseelan, C. O'Donovan, *MetaboLights: A resource evolving in response to the needs of its scientific community*, Nucleic Acids Res. 48 (2020) D440–D444, <http://dx.doi.org/10.1093/nar/gkz1019>.

Chapitre IV : Application de la méthode de déréplication à de nouvelles souches de levures.

I. Ajout de nouveaux outils de traitement de données

Les méthodes de traitement de données en métabolomique sont en constante évolution et de nouveaux outils apparaissent régulièrement, tandis que ceux existants bénéficient d'amélioration permettant aux utilisateurs de pousser plus loin les travaux de recherche et l'exploration des produits naturels.

De ce fait, après la mise en place d'une méthode fiable et efficace de déréplication pour l'analyse des extraits de levures réalisés durant ce projet de thèse, nous avons décidé de mettre à l'essai certaines fonctions et certains algorithmes afin d'affiner notre processus général.

1. Utilisation de modules supplémentaires de MZMine2

Une étape a également été ajoutée au pré-traitement des données UHPLC-MS/MS par le logiciel MZMine2. Il s'agit de l'étape de *gap-filling*, réalisée sur les listes de *features* des échantillons alignés. Le module utilisé compare les signaux des différentes variables et recherche les rapports m/z et les temps de rétention de celles-ci dans les échantillons où les composés n'ont pas été définis comme présents par les étapes de construction de chromatogrammes puis l'alignement des analyses. Il est en effet possible, occasionnellement, que ces derniers modules « passent à côté » de la détection des composés en question. L'application du *gap-filling* permet de combler ce manque et d'améliorer l'alignement des analyses.

2. Application d'outils intégrés à la plateforme GNPS

La plateforme GNPS s'est agrémenté, pendant la durée de ces travaux de thèse, de plusieurs outils en évolution permanente et permettant d'aller plus en profondeur dans l'interprétation des réseaux construits.

L'outil molDiscovery est une méthode de recherche sur base de données spectrales s'appuyant sur un algorithme de génération *in silico* de fragmentations MS/MS combiné à un apprentissage par la machine d'un modèle probabiliste reliant les petites molécules aux spectres expérimentaux [186]. Cet outil permet donc l'amélioration de l'étape d'annotation des composés en fournissant plus d'annotations que GNPS seul.

L'outil *Network Annotation Propagation* (NAP) permet également l'attribution de nouvelles annotations mais en analysant les réseaux créés par une tâche de GNPS [187]. Il utilise la fragmentation *in silico* des composés des bases de données pour dresser une liste de candidats possibles pour l'annotation d'une molécule d'un réseau, tout en prenant en compte les candidats des molécules environnantes dans ce même réseau.

L'outil MS2LDA, intégré également à la plateforme GNPS pour une utilisation simplifiée, a été essayé dans le cadre du traitement de données des souches *S. reginensis* et *S. kourouensis*. Cet outil utilise les données de fragmentation acquises au cours d'expériences MS/MS pour en extraire de manière non-supervisée des motifs structuraux ayant une importance biochimique [188]. Le procédé automatisé s'est révélé moins informatif que les combinaisons d'algorithmes déjà utilisées, ou du moins aucune nouvelle information ne fut apportée. Pour ces raisons, l'outil MS2LDA n'est pas intégré au processus de traitement des données d'expériences UHPLC-MS/MS.

II. Résultats de l'application de la méthode finalisée à deux souches récemment découvertes : *Starmerella reginensis* et *Starmerella kourouensis*

Ces optimisations de la méthodologie ont été mises au service de l'exploration de deux nouvelles souches sélectionnées pour leur découverte récente et pour leur appartenance à un genre intéressant, le genre *Starmerella*. Ainsi, des cultures de *Starmerella reginensis* et *Starmerella kourouensis* ont subi le protocole d'extraction pour donner des extraits destinés à être analysés par UHPLC-MS/MS et par FTICR-MS. L'investigation des données spectrales acquises grâce à ces deux techniques complémentaires a permis la détection et l'annotation de 439 *features* pour *S. reginensis* et de 513 *features* pour *S. kourouensis*.

L'utilisation de molDiscovery pour l'exploration de *S. reginensis* et *S. kourouensis* a permis d'annoter 75 composés (65 annotations de niveau 2 et 10 annotations de niveau 3), ce qui en fait un outil très utile pour la cartographie des métabolites des levures. L'utilisation du NAP pour l'investigation des *features* détectées dans les analyses de *S. reginensis* et *S. kourouensis* ont permis d'annoter 16 composés.

L'exploration de ces deux nouvelles souches inédites a permis de mettre en évidence 439 métabolites pour *S. reginensis* et 513 métabolites pour *S. kourouensis*. Respectivement 73 % et 78 % de ces

annotations sont de niveau 1, 2 et 3, ce qui reste un très bon niveau pour une étude métabolomique. Les profils dressés sont similaires pour les deux souches tropicales, avec de nombreux alcaloïdes annotés (respectivement 114 et 171 pour *S. reginensis* et *S. kourouensis*), ainsi que de multiples lipides, y compris des stéroïdes et des terpénoïdes, des acides aminés et des acides organiques. Enfin, 45 *features* sont retrouvées en commun avec notre étude sur la levure *S. cerevisiae*.

Des répliques biologiques ont été analysés afin de vérifier la reproductibilité de la méthode, tout comme lors des analyses de *S. cerevisiae*, avec une moyenne de 89 % de *features* retrouvées entre les différentes analyses réalisées. Les résultats de ces expériences ont montré de grandes similitudes entre les extraits provenant de cultures différentes des mêmes souches du genre *Starmerella*, permettant ainsi une seconde validation de la méthode expérimentale.

Ces résultats sur la déréplication de *S. reginensis* et *S. kourouensis* ont conduit à une publication dans le *Journal of Agricultural and Food Chemistry (ACS Publications)* [189]. Cette publication est jointe à partir de la page suivante ; tout comme la première, le format est adapté au style du manuscrit de thèse et les figures sont insérées dans le corps du texte. Les références sont numérotées selon l'ordre de l'article, et les informations supplémentaires sont fournies dans la partie Annexes.

Référence : O. Perruchon et al., *J. Agric. Food. Chem.* (2021).

<https://doi.org/10.1021/acs.jafc.1c03201>

Metabolome Exploration by High-Resolution Mass Spectrometry Methodologies of Two New Yeast Species: *Starmerella reginensis* and *Starmerella kourouensis*

Olivier Perruchon^a, Isabelle Schmitz-Afonso^a, Cécile Grondin^{bc}, Jean-Luc Legras^c, Carlos Afonso^a, Abdelhakim Elomri^a

^aNormandie Univ, UNIROUEN, INSA Rouen, CNRS, COBRA (UMR 6014), Rouen 76000, France

^bMicalis Institute, INRAE, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas 78352, France

^cSPO, Univ Montpellier, INRAE, Institut Agro, Montpellier 34000, France

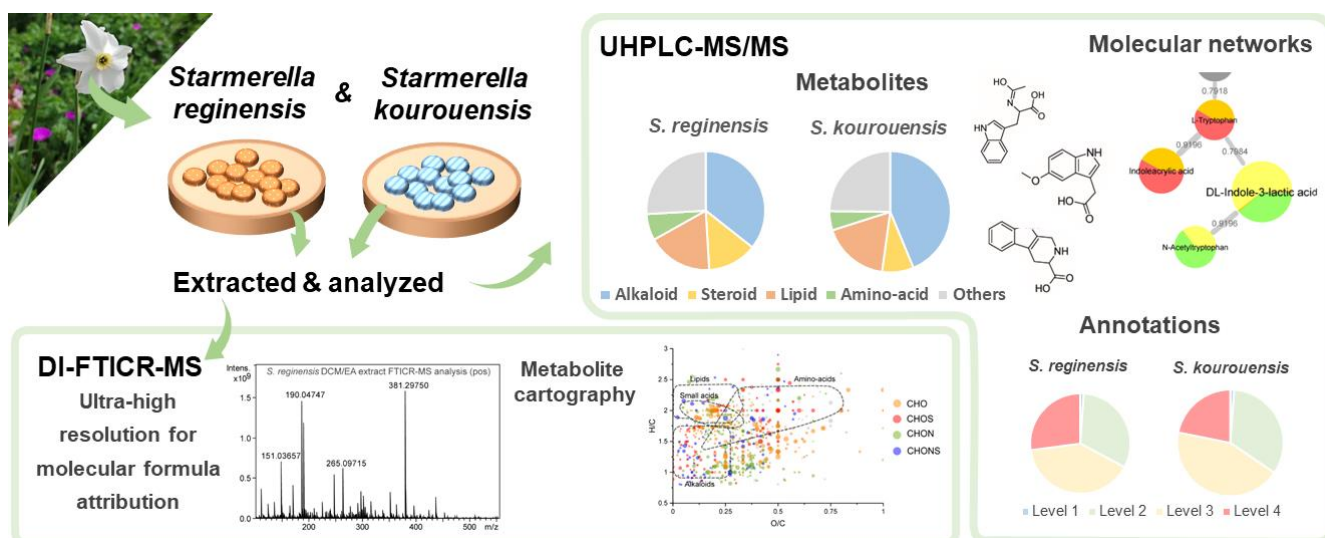
Corresponding author:

E-mail address: hakim.elomri@univ-rouen.fr (A. Elomri).

Keywords:

Untargeted metabolomics; Metabolic profiling; *Starmerella* yeast strains; Natural products; UHPLC-MS/MS; Molecular networking; FTICR-MS

Received: 28 May 2021. / Published online: 16 September 2021.



Abstract

Nature is harnessed since ancient times to fulfill human needs, and yeast culture has been mastered for bakery, brewery, or the preparation of beverages. In this context, the two recently discovered yeast species *Starmarella reginensis* and *Starmarella kourouensis*, belonging to a genus related to fermentative activities in the literature, were explored via untargeted metabolomics approaches. Ultrahigh-performance liquid chromatography hyphenated with tandem mass spectrometry and a deep investigation of molecular networks and spectral data allowed the annotation of, respectively, 439 and 513 metabolites for *S. reginensis* and *S. kourouensis*, with approximately 30% compound annotations and 40% chemical class annotations for both yeast strains. These analyses and Fourier transform ion cyclotron resonance mass spectrometry accurate metabolic profiles unveiled a rich content of alkaloids, lipids, amino acids, and terpenoids for *S. reginensis*. *S. kourouensis* presents a similar profile with more sulfated compounds. In short, these results enrich the current knowledge about *Starmarella* yeast secondary metabolites and reveal their significant structural diversity of small molecules.

Introduction

Yeasts have been cultivated and used from ancient times in human activities, for instance to make bread and fermented beverages such as beer and wine. In the contemporary era, they are also used as biotechnological tools.^{1,2} Yeasts are widely present in nature, and only few of them have already been explored. Analyses of their chemical composition showed an important capacity of adaptation to their culture medium and their ability to produce a very large number of molecules with a significant structural diversity, both as volatile aromatic molecules and nonvolatile substances. Yeasts are widely used in the food industry to diversify aromas and flavors of fermented foods and beverages, especially *Saccharomyces cerevisiae*, the well-known yeast used in brewery, bakery, or winemaking, to name just a few. Besides, it is used as a model system for eukaryotic biology and belongs to the order of *Saccharomycetales*.

The yeast genus *Starmerella* was introduced in 1998 by Rosa and Lachance.³ The first studied *Starmerella*, *Starmerella bombicola*, was a teleomorph, that is to say, a synonym of *Candida bombicola*, able to produce sophorolipids useable as biosurfactants. *Starmerella* yeasts have been described as osmotolerant and fermentative, and they use few carbon compounds to live and grow and may be collected from flowers and associated beetles.^{3,4} These non-*Saccharomyces* yeasts are considered useful in the preparation of alcoholic beverages, such as wine, to enhance sensorial complexity and the aroma profile or to reduce potential ethanol content.⁵ For example, *Starmerella bacillaris* is one member of the fermenting coffee berry microflora; its impact on wine aroma profiles has been explored via analyses by chromatographic techniques or mass spectrometry.⁶⁻⁸ Some strains of *S. bombicola* have also been isolated from an Italian cake sourdough.⁹ The production of sophorolipids by *Starmerella* genus yeasts has been reported by several studies and is very interesting for industrial applications as it can serve as a basis for molecule synthesis, as a detergent, as an antimicrobial agent, or as an anticancer agent.¹⁰⁻¹² *Starmerella* yeasts have also been isolated from kombucha tea, an Asian fermented tea with reported antioxidant, antimicrobial, antistress, antidiabetic, and other effects.^{13,14} Nowadays, research is ongoing, and more than 48 *Starmerella* species have been described in the literature.

Two *Starmerella* genus yeasts, *Starmerella reginensis* (CLIB 1634) and *Starmerella kourouensis* (CLIB 1707), have recently been discovered by Amoikon et al., who also provided their taxonomic description. Both *S. reginensis* and *S. kourouensis* have been isolated from flowers in the tropical environment of French Guiana, near the towns of Regina and Kourou, respectively.⁴ These two *Starmerella* species are

new, and this originality, combined with the industrial potential of *Starmerella* yeasts, makes the exploration of what these yeasts can produce very interesting.

There are more and more studies on the organoleptic properties of microorganisms for food and health, so a method to screen all secondary metabolites, or natural products, of a microorganism is of great interest. Among the approximate 200,000 compounds recorded in the Dictionary of Natural Products, only 9.99% are from fungal organisms, making fungal species an original and interesting biological source to explore.¹⁵ Untargeted metabolomics approaches are often used to study small molecules extracted from biological matrices. In metabolomics studies, ultrahigh-performance liquid chromatography (UHPLC) is commonly used to separate metabolites according to their polarity. UHPLC hyphenated with tandem mass spectrometry (UHPLC–MS/MS) allows a sensitive and specific analysis by providing a second separation dimension, separating molecules according to the mass-to-charge ratio of corresponding ions (m/z ratio).¹⁶ Moreover, tandem mass spectrometry leading to the fragmentation of the detected compounds brings structural information on these metabolites.^{17,18}

As these methodologies produce big data, automated data treatment is a crucial step and multiple tools in development are blooming in the field. These tools substantially accelerate MS data deciphering and reduce the human error by the use of algorithms. MZMine2 illustrates the capacity of data treatment software as it converts LC–MS data in the list of features (described by their m/z ratio and their retention time).¹⁹ The accurate mass measurements of detected features are compared with spectral databases (e.g., Human Metabolome Database and MassBank^{20,21}) to annotate the compounds. MZMine2- preprocessed MS/MS data may be used to build molecular networks on the Global Natural Product Social (GNPS) molecular networking platform (<https://gnps.ucsd.edu>) as structurally related molecules share similar fragmentation patterns.^{22,23} Additional tools, such as Network Annotation Propagation (NAP) or MolDiscovery, use all the capacities of spectral databases and in silico fragmentation principles to accelerate the dereplication process and facilitate the manual investigation of acquired data.^{24,25} All the annotation provided by this treatment may be gauged by the annotation level system of the Metabolomics Standard Initiative.²⁶ This system uses four levels covering fully identified molecules, putatively annotated compounds, putatively annotated compound chemical classes, and unknown metabolites.

Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS) is a suitable additional method for metabolomics as it enhances the LC–MS/MS results on one side, thanks to its high mass measurement accuracy, and its ultrahigh resolution, leading to attribution of a unique molecular

formula.^{27,28} On the other side, it allows metabolic profiling, thanks to the automated attribution of numerous molecular formulae to the signal crowd provided by this very sensitive mass spectrometry technology. From these molecular formulae, hydrogen/carbon and oxygen/carbon ratio calculation (H/C and O/C) brings a metabolite cartography via van Krevelen diagrams.^{29,30} The small molecules are then mapped according to a chemical meaning.

In this work, *S. reginensis* and *S. kourouensis* were studied for their nonvolatile metabolomes and, more particularly, their secondary metabolites. The two yeast species were cultivated; then, metabolites were extracted and analyzed by two complementary mass spectrometry techniques: UHPLC–MS/MS and FTICR–MS. The complete analysis and data treatment workflow for the dereplication of biological extracts have proven its efficiency and are explained in the Results and Discussion section and summarized in Figure 1. The goal was to observe both original species production capacities, as they never have been explored before, compare them with each other, with other *Starmerella* yeasts, and also with *S. cerevisiae* (which also belongs to the order of *Saccharomycetales*) from a previous study.³¹

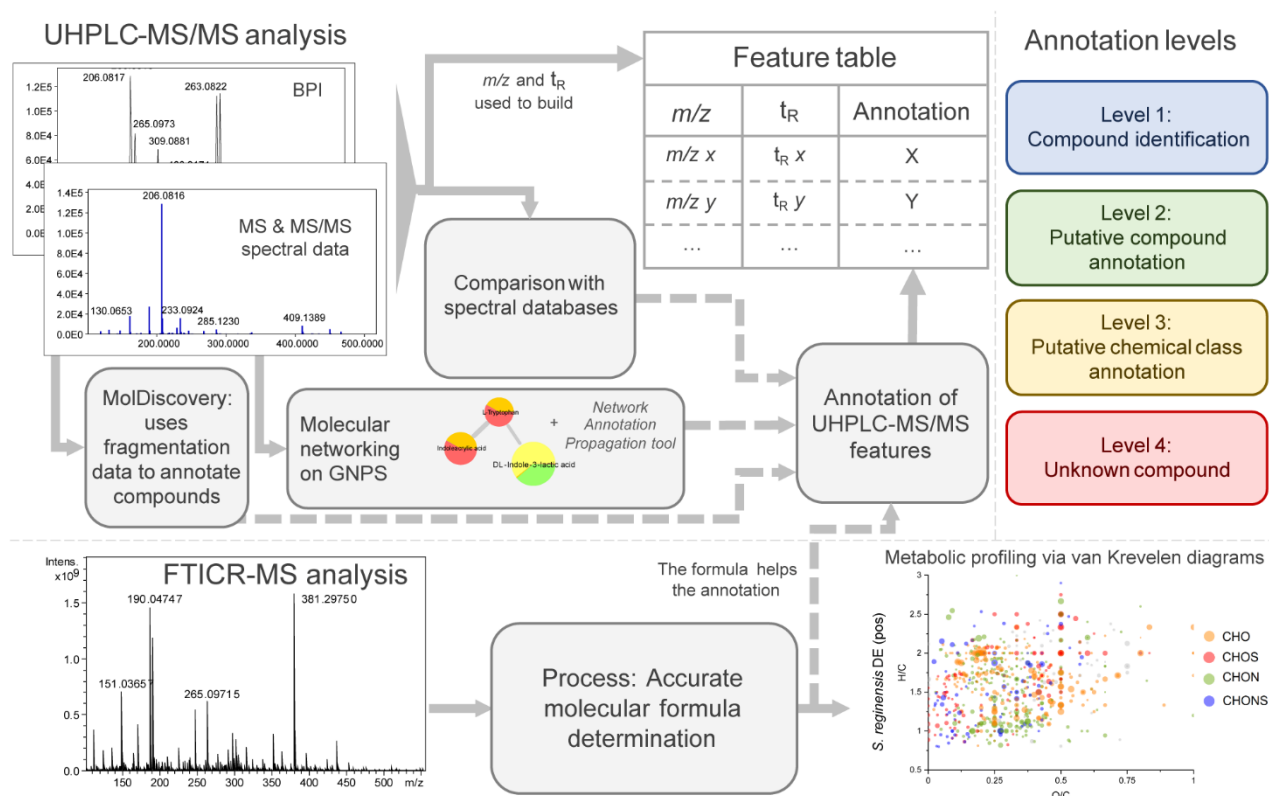


Figure 1. Untargeted metabolomics dereplication workflow for the analysis of biological extracts. The organization chart depicts the analysis, the automated, and the manual investigation steps of acquired spectral data. The joint use of UHPLC–MS/MS and FTICR–MS led to metabolic profiling and a feature table. Annotations are attributed to the compounds following the Metabolomics Standards Initiative, with four levels from fully identified compounds to unknown compounds.

Materials and methods

Yeast Culture

S. reginensis (CLIB 1634 = CBS 15247) and *S. kourouensis* (CLIB 1707 = CBS 15257) strains are preserved at the Biological Resource Centre CIRM-Levures (SPO, INRAE Montpellier, France; <https://www6.inrae.fr/cirm/Levures>). Yeasts were routinely grown on modified YPD agar medium (1% yeast extract, 1% peptone, 1% glucose, and 1.4% agar) at 26.5 °C. ⁴ Strains were streaked from an overnight culture on a 10 cm diameter Petri dish on a solidified synthetic medium: YNB agar medium (6.7% yeast nitrogen base, 1% peptone, 1% glucose, and 1.4% agar) and then incubated for 3 days at 26.5 °C. Culture medium samples without yeasts were also prepared as control samples. Metabolites from yeast cultures and molecules from culture medium samples were solvent based-extracted. Three cultures of each strain were used as biological replicates.

Metabolite Extraction

The extraction protocol was already used in previous studies and was chosen for its efficiency in the extraction of various ranges of intracellular small molecules, including secondary metabolites. ³¹ For each yeast culture (or control culture medium), the agar medium with or without the yeast colonies was placed in a glass bottle. Intracellular and extracellular metabolites were extracted with HPLC-grade dichloromethane, ethyl acetate, and methanol. Then, 15 mL of 1:1 v/v dichloromethane and ethyl acetate were added to the cultures. The cultures were sonicated for 30 min to disrupt cells (Branson B-5200-54, 450 W). The solvent was collected and filtered (sintered glass funnel porosity from 16 to 40 µm), and then, 10 mL of dichloromethane and ethyl acetate (1:1, v/v) were added to the medium. The culture was sonicated 10 min, and then, the solvent was filtered again and pooled with the previous one. The dichloromethane/ethyl acetate (DCM/EA) extracts were evaporated under vacuum at 40 °C and stored at -20 °C (*S. reginensis* dry DCM/EA extract: 5.2 mg; *S. kourouensis* dry DCM/EA extract: 6.1 mg). The remaining metabolites from culture medium residues were extracted with 10 mL of methanol, sonicated for 5 min, and filtered, and the methanol extracts (M extract) were evaporated under vacuum at 40 °C and stored at -20 °C (*S. reginensis* dry M extract: 100.1 mg; *S. kourouensis* dry M extract: 124.5 mg).

Sample Preparation

Methanol, water, and acetonitrile used in this step are of Fisher Chemical Optima LC–MS grade. Each DCM/ EA extract was suspended in methanol, and each M extract was suspended in water to a 50 mg/mL concentration. All samples were filtered through PTFE 0.45 µm filters. The samples (and culture medium controls) were diluted to 5 mg/mL with water/acetonitrile (ACN; 1:1 v/v) for LC–MS/MS analysis, and 10 µL was taken to be diluted at 1/10,000 in water/ACN (1:1 v/v) for FTICR–MS analysis. Analytical blanks were made of water/ACN (1:1 v/v). Pure chemical standards (Table S1) were individually dissolved in water/acetonitrile (1:1 v/v) and mixed, then diluted to 0.01 mg/mL, and analyzed by UHPLC–MS/MS.

UHPLC-MS/MS Analysis

Yeast extract metabolites were separated according to their polarity through a C18 silica-based column and heated at 40 °C (Acquity UPLC HSS T3 1.8 µm × 1.0 mm × 100 mm, prefilter of 0.2 µm, Waters Corporation, Milford, MA, USA) with a UHPLC system (Dionex Ultimate 3000 UPLC+, Thermo Scientific, San Jose, CA, USA). The samples were loaded on the autosampler that kept them at 4 °C, and then, 1 µL was injected. The mobile-phase solvents were water + 0.1% formic acid (solvent A) and acetonitrile + 0.1% formic acid (solvent B); the solvents were of Fisher Chemical Optima LC–MS grade; formic acid is LiChropur (Merck) for LC–MS. With a 0.1 mL/min flow, the gradient goes from 99% A/1% B from 0 to 1 min and then linearly to 100% B at 23 min. 100% B was maintained for 2 min; then, the gradient returned to initial conditions (99% A/1% B) at 25.2 min, and it was stabilized to 30 min, which is the end of the run.

Separated metabolites were detected through electrospray ionization-quadrupole time-of-flight tandem mass spectrometry (ESIQTOF, SYNAPT G2 HDMS, Waters MS Technologies, Manchester, UK; MassLynx 4.1). The samples were analyzed in both positive and negative ionization modes, with a m/z range of 50–2000 and a data-dependent scan method (Waters Survey Analysis). Ionization parameters and acquisition parameters are displayed in Table S2. The mass spectrometer was externally calibrated before analyses (sodium formate: 10% formic acid/0.1 M NaOH/acetonitrile in 1:1:8, v/v/v) and internally calibrated during analyses (Leucine Enkephalin solution: 2 ng/µL in water/isopropanol (1:1,

v/v); Lock Mass Waters injection system with ions m/z 556.2771 and m/z 554.2615 in the positive and negative ionization mode, respectively).

Data Processing

Centroided raw data (MassLynx 4.1 Automatic peak detection module) were imported into MZMine2. Masses were detected (exact mass algorithm) and built into chromatograms (ADAP chromatogram builder³²). Chromatograms from feature tables, characterized by the m/z ratio, retention time in minutes, intensity, and MS spectral data, were deconvoluted (Wavelet ADAP algorithm) to remove background noise, discriminated isomer, and link MS2 spectral data to features. Isotopes were removed; duplicated features were filtered. Feature lists from yeast samples and culture medium controls were aligned. Molecular formulae were attributed (formula prediction module) with parameters following the seven golden rules, which are heuristic rules used to avoid aberrant formula attribution (number of elements restrictions, respect of chemical rules, isotopic patterns, hydrogen/carbon ratios, elemental ratios against carbon, element ratio probabilities, and the presence of trimethylsilylated compounds).³³ Fragments and adducts were annotated by comparing MS1 and MS2 data. Detailed parameters are displayed in Table S3. Biological replicates were treated via MZMine2 and used to confirm the similarity between the cultures.

Exported data from MZMine2 were imported on the GNPS platform to build molecular networks. A molecular network was created with the Feature-Based Molecular Networking (FBMN) workflow on GNPS (<https://gnps.ucsd.edu>). MS/MS spectra were window-filtered by choosing only the top six fragment ions in the ± 50 Da window throughout the spectrum. The precursor ion mass tolerance was set to 0.02 Da and the MS/MS fragment ion tolerance to 0.02 Da. A molecular network was then created, where edges were filtered to have a cosine score above 0.7 and more than 3 matched peaks. Furthermore, edges between two nodes were kept in the network if and only if each of the nodes appeared in each other's respective top 10 most similar nodes. Finally, the maximum size of a molecular family was set to 100, and the lowest scoring edges were removed from molecular families until the molecular family size was below this threshold. The spectra in the network were then searched against GNPS spectral libraries. The library spectra were filtered in the same manner as the input data. All matches kept between network spectra, and library spectra were required to have a score above 0.7 and at least 3 matched peaks. The molecular networks were visualized using Cytoscape software.

After experimenting several bioinformatic tools on metabolomic datasets, we chose to apply the Network Annotation Propagation (NAP) and the MolDiscovery tool for their ease of use and for the most exhaustive and rapid exploration possible.^{24,25}

FTICR-MS Analysis

Yeast extracts and culture medium controls were analyzed on a FTICR-MS instrument (Solarix XR FTMS, Bruker Daltonics) equipped with a 12 T superconducting magnet and a dynamically harmonized ICR cell. 1/10,000 diluted samples were directly injected and ionized with an electrospray ionization source operated in positive and in negative ionization modes. Previous studies inspired the source and analyzer parameters that guarantee sensitive and robust analyses^{31,34} (Table S1). A resolution of 600,000 at m/z 200 was reached.

Data Analysis 5.0 (Bruker) was used to treat FTICR-MS spectral data. Spectral data were internally calibrated, thanks to the known ions (internal calibration list shown in Table S1) that were added before the analysis. Then, the m/z ratios from analytical blanks and culture medium were subtracted from the sample analyses to exclude media components and artifacts ("Xpose" algorithm, ratio = 5). Molecular formulae were attributed considering the $C_{1-100}H_{0-200}N_{0-8}O_{0-20}S_{0-1}$ species (then P_{0-1} in a second time) with respect to the seven golden rules, with a tolerance set to 0.3 ppm, rings and double-bond equivalents from 0 to 80, and H/C ratio from 0 to 3.³³ Considered adducts were $M + H$, Na, and K for the positive mode and $M - H$ and Cl for the negative mode. Finally, molecular formulae were manually verified to discard outliers and then exported to determine oxygen/carbon and hydrogen/carbon (O/C and H/C) ratios for each feature in order to build van Krevelen diagrams.

Results and Discussion

*Metabolic Profiling of *S. reginensis* and *S. kourouensis* by FTICR-MS*

S. reginensis and *S. kourouensis* DCM/EA and M extracts were analyzed in FTICR-MS in both positive and negative ionization modes. The FTICR-MS technology is a very sensitive method, so the acquired data count numerous signals. After internal calibrations and culture medium ion subtraction, accurate

molecular formula attribution to the signals was automated, with parameters that respect the seven golden rules of formula attribution.³³ The molecular formula lists were manually and closely checked to discard outliers and discriminate the most plausible formula when there were several possibilities. In case of ambiguity, the isotopic fine structure was checked, demonstrating an advantage of ultrahigh-resolution mass spectrometry. Table S4 summarizes the feature numbers determined in the different extracts.

From the molecular formula lists, O/C ratios were plotted against H/C ratios to build van Krevelen diagrams. As an example, the FTICR-MS mass spectrum of *S. reginensis* DCM/ EA extract analysis in the positive ionization mode is displayed in Figure 2A. The associated van Krevelen diagram used for metabolic profiling is displayed in Figure 2B. Figure 2C compares the four major compound classes for *S. reginensis* and *S. kourouensis* extracts [full van Krevelen diagrams are detailed in Figure S1 (vK) and S2 (vK-N)]. Thanks to our previous work, several areas can be attributed to chemical classes based on the elemental composition of the detected features. More features are detected in the positive ionization mode, whereas the two yeast strains show very few differences in the negative ionization mode, which may attest a similar content of acids. *S. reginensis* shows a rich content in alkaloids and a very rich content in lipids and amino-acid derivatives. CHON compounds are also numerous. *S. kourouensis* shows less features but many sulfated alkaloids, sulfated amino acids and derivatives (CHONS compounds in the alkaloid and amino-acid area), and CHO lipids. The methanol extract also contains many CHON molecules, such as in the alkaloid area. As it can be seen in Figure 2C, the metabolic profiles of the two *Starmerella* species present many similarities; the distribution of chemical composition is close except that *S. reginensis* contains generally a higher number of detected features.

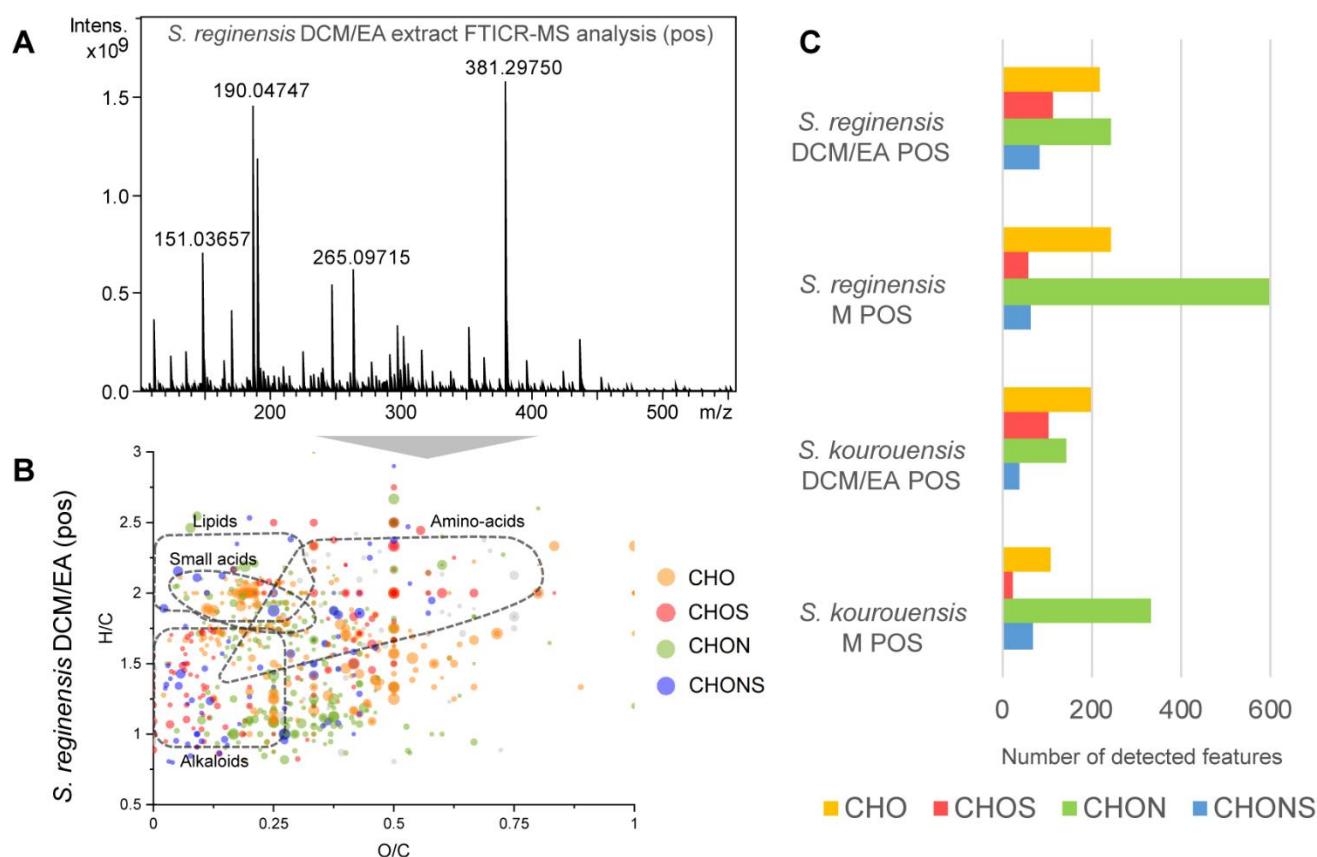


Figure 2. Results of the *S. reginensis* and *S. kourouensis* FTICR-MS metabolic profiling. (A) Example of the mass spectrum of *S. reginensis* DCM/EA extract FTICR-MS analysis in the positive ionization mode. The displayed m/z range goes from m/z 100 to m/z 550. (B) van Krevelen diagram (H/C vs O/C) generated after data treatment of the A spectrum. The node color corresponds to the chemical composition in C, H, O, and N (CH, CHN, CHNS, and molecules with phosphorus are displayed in gray). (C) Distribution (in number of features) of the four major chemical compositions (CHO, CHOS, CHON, and CHONS) among the four analyses in the positive mode of *Starmerella* yeast extracts.

Metabolic Exploration of *S. reginensis* and *S. kourouensis* by LC-MS/MS

S. reginensis and *S. kourouensis* extracts were successively injected and separated according to their polarity through UHPLC and then analyzed in ESI-QTOF-MS/MS in both positive and negative ionization modes. The base peak chromatograms of these analyses are shown in Figure S3. A large amount of data acquired was preprocessed within the MZMine2 software to detect masses, build chromatograms gathered in a feature table, and discard outlier features. The yeast sample analyses were aligned with each other and also with culture medium analyses to discard medium features and focus on yeast exclusive molecules. The yeast extract analyses were validated, thanks to biological triplicates whose

data were aligned with MZMine2. The generated feature table was the basis of manual investigation of results, and molecular formulae were accurately attributed to features by manually comparing with FTICR–MS analysis. Annotations were provided following the Metabolomics Standard Initiative, wherein level 1 are identifications, level 2 are putative annotations, level 3 are putative chemical class annotations, and level 4 are unknown molecules. Annotations were provided by the comparison of physical standards (with retention time and fragmentation correspondence; for level 1 identifications; see used standards in Table S5) or by comparison with online databases (HMDB, Dictionary of Natural Products, MassBank, GNPS).

Molecular networks were built on the basis of preprocessed spectral data on the Global Natural Product Social (GNPS) platform, with its feature-based molecular networking process. This GNPS database search tool (fragmentation comparison with multiple databases) allowed 18 level 2 annotations, which is not a large number of annotations but is a good starting point. The major intake of using the GNPS molecular networking platform is the fact that it accelerates the manual process of data investigation because the user may exploit the networks to propagate annotations among the nodes. As structurally similar compounds share fragmentation patterns and cluster together, one network often corresponds to a specific chemical class, and level 3 chemical class annotations are attributed when the feature is in a known network (e.g., ID 216, ID 265, and ID 301). The molecular networks generated are shown (after treatment and annotation workflow) in Figure S4. The application of the NAP tool (applied from the GNPS hub) enhanced the data processing by using spectral data within the networks and in silico fragmentation of database molecules to annotate compounds, thanks to library spectral matching or to annotated chemical classes (ID 67 for instance). The MolDiscovery tool (also applied from the GNPS hub) is another database search method that uses generated MS fragmentation and a probabilistic model to match molecules. This tool was applied and provided annotations, particularly level 2 and 3 annotations. The use of these tools and the deep manual interpretation resulted in the outcomes depicted in Figure 3, with 439 annotations for *S. reginensis* and 513 for *S. kourouensis*. These annotations are gathered in Table S6. With such a large amount of data generated by this kind of analysis, one could spend infinite time to investigate fragmentation data and networks in order to attribute compound names to the detected feature while always going further in the interpretation of signals. However, infinite time is not available, so the major interest toward the combined use of these analytical and bioinformatic tools is to go as far as possible in a limited period of time.

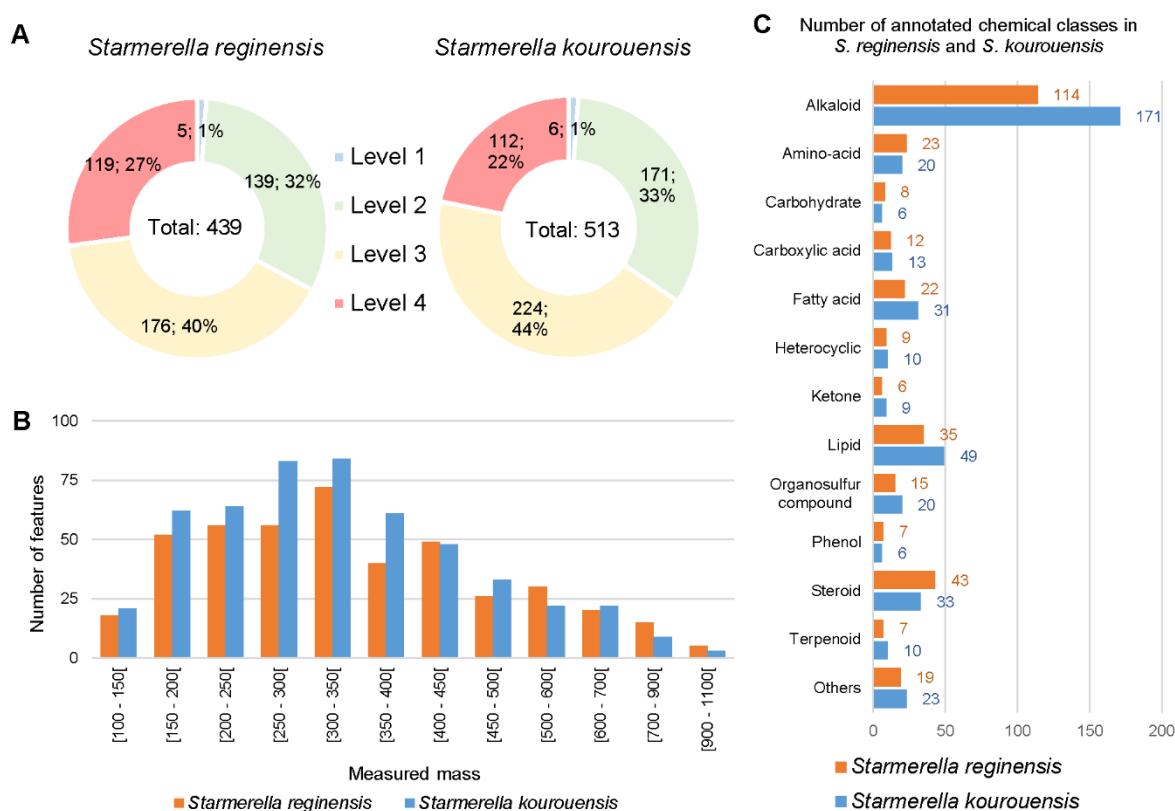


Figure 3. Results of the *S. reginensis* and *S. kourouensis* dereplication via UHPLC–MS/MS. (A) Distribution of annotation levels attributed, according to the Metabolomics Standard Initiative. Blue parts are level 1 identifications, green parts are level 2 putative compound annotations, yellow parts are level 3 putative chemical class annotations, and red parts are level 4 unknown compounds. (B) Molecular weight distribution (in m/z) of *S. reginensis* and *S. kourouensis* detected features. (C) Chemical class distribution of *S. reginensis* and *S. kourouensis* level 2 and 3 annotated metabolites. The class named “Others” corresponds to nucleosides, alcohols, esters, nitrogenous bases, amines, benzenoids, naphthols, and polyketides.

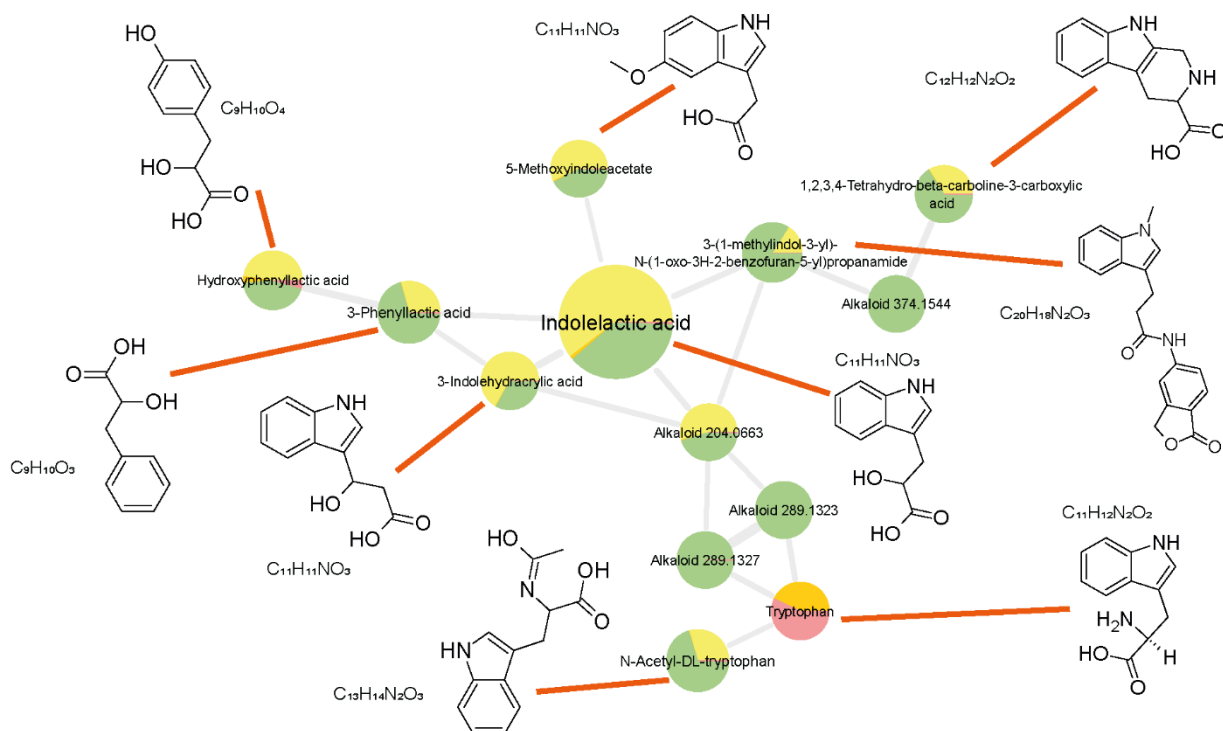
S. reginensis- and *S. kourouensis*-Type Strain Metabolites

As seen in Figure 3A, the whole analysis and annotation workflow is efficient, as shown by the 439 features detected in *S. reginensis* and the 513 features detected in *S. kourouensis*. 33 and 32% of these annotations are level 2 putative annotations, and 40 and 44% are level 3 chemical class hypotheses. This means that approximately a quarter of the detected features remains completely unknown, which is relatively low for a metabolomics study. Among the annotations (two species combined, Table

S6), 75 have been attributed, thanks to the MolDiscovery tool (65 level 2 and 10 level 3), and 16 have been attributed, thanks to NAP result investigation. These two tools brought a significant impact on the reliability and the speed of the used metabolomics workflow, enough to be integrated in the classical biological extract analysis protocol.

Concerning the two studied *Starmerella*-type strain metabolites, as seen in Figure 3B, a large majority of detected features were under m/z 450, 343 for *S. reginensis* and 423 for *S. kourouensis*. The distribution of detected masses is similar in *S. reginensis* and *S. kourouensis*. The same report is made by looking at the chemical classes of the level 2 and 3 annotated compounds (Figure 3C). All the annotated chemical classes are similarly distributed among the two species, except the fact that the analysis of *S. kourouensis* led to more annotations than *S. reginensis*. A large number of alkaloids were annotated (114 for *S. reginensis* and 171 for *S. kourouensis*), which is interesting as alkaloids may have a strong impact in the pharmaceutical or food industry. Among the 179 alkaloids in the two strains, 52 annotated indole alkaloids are reported in our study. It is also important to note that, to the best of our knowledge, it is the first time that such numerous alkaloid molecules are annotated from a metabolomics analysis of a yeast culture extract. Then, many lipids, including fatty acids, fatty acid derivatives, steroid and steroid derivatives, amino acids, amino acid derivatives, and carboxylic acids are the most represented chemical classes. However, these results must be weighted as the technique is semi-quantitative and not fully quantitative and also due to the fact that the compounds may be interesting but may still be hard to physically isolate from structurally close molecules. Figure 4 shows an extract from the generated molecular networks annotated with molecular formulae and molecular structures from compounds.

Example of generated molecular network (alkaloid)



Example of generated molecular networks (lipids)

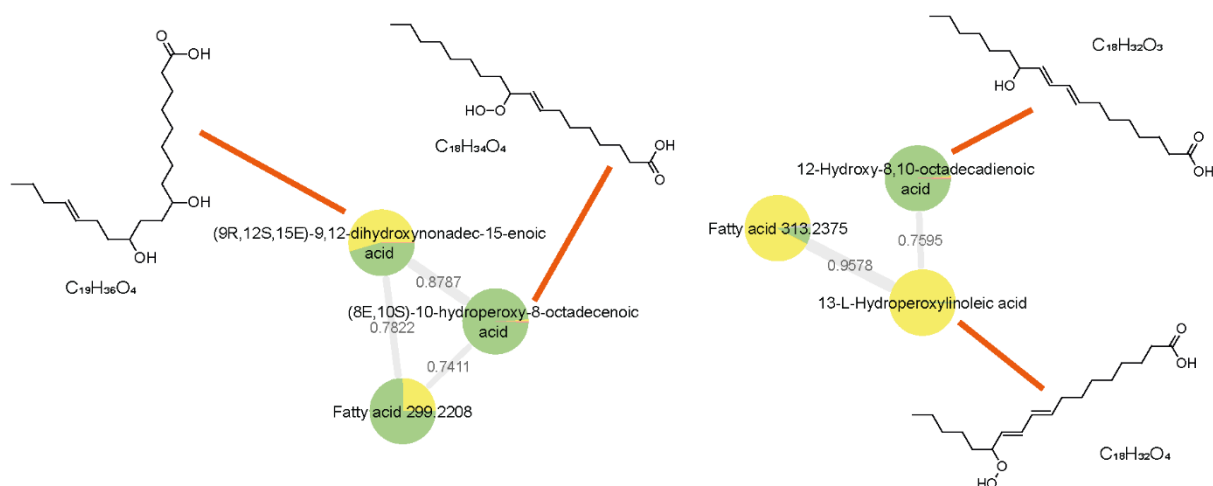


Figure 4. Extract from the generated molecular networks after *S. reginensis* and *S. kourouensis* UHPLC–MS/MS analysis and data treatment. Molecular formulae and structures of level 2 annotated compounds are displayed.

Among the 445 metabolites annotated at level 2 and 3 in the complete metabolite table (Table S6), 264 of them are common to the two yeast strains. As seen in Figure 3C, 51 compounds are exclusive to *S. reginensis* (with 14 steroids being the most represented class in these exclusives) and 131 are exclusive to *S. kourouensis* (with 66 exclusive alkaloids and 32 exclusive lipids, including fatty acids). Concerning

the 147 unknown features, 41 of them have an attributed molecular formula, thanks to the accuracy and the sensibility of the high-resolution mass spectrometry tools. A total of 84 of the unknown annotations are common to the two yeast species, while 35 are exclusive to *S. reginensis* and 28 are exclusive to *S. kourouensis*. It means that although the two yeasts have close small-molecule contents, each strain produces its specific metabolic profile, as seen previously (Figure 3), and *S. kourouensis* seems to have a slightly richer diversity of metabolites. They also have their own diversity of molecules that could not be annotated more than level 4 in this work and that are left to be explored further.

As an illustration of the diversity of annotated molecules and their biological implication in other matrices, 3-phenyllactic acid (ID 148) has been annotated, and this metabolite, also produced by lactic acid bacteria, has a reported fungicidal activity in sourdough breads, for example.³⁵ 3-Phenyllactic acid was reported as being a major component of honey and may have an impact on honey taste.³⁶ Hydroxyphenyllactic acid (ID 57) has been isolated from the bacteria *Klebsiella pneumonia*, *Escherichia coli*, and *Staphylococcus aureus*. Hydroxyphenyllactic acid is derived from phenyllactic acid and has a reported antioxidant activity.^{37,38} Tetrahydroharman-3-carboxylic acid (ID 100), a carboline alkaloid, has been isolated from fruits, especially from citrus species, and from chocolate and cocoa.^{39,40} It may have different effects on the central nervous system as it is potentially neurotoxic and also an antioxidant. Tryptophol (ID 197) and 5-hydroxytryptophol (ID 104) are alkaloids that have been isolated from plants such as *Matteucia struthiopteris*, bacteria such as *Azospirillum brasilense*, and fungi such as *Neurospora crassa*. They are linked to tryptophan metabolic pathways and are found in alcoholic beverages. They are fermentation products, so finding them in yeast extracts seems coherent. Tryptophol has several biological activities that differ according to the considered biological source; it is involved in communication, used as an antibiotic, and presents a drowsy effect. They are also characterized in the case of excessive alcohol intake.⁴¹ As a last example, 2-quinolinecarboxylate (ID 288; also called quinaldic acid) has already been isolated in natural samples, including the plant *Ephedra pachyclada*. This compound as well as its derivatives are potential antidiabetic agents.⁴²

Some annotated metabolites have already been characterized in fungi or yeasts from other genera, and here, are a few examples. 5-Hydroxy-L-tryptophan (ID 26) is produced by *S. cerevisiae* and other non-*Saccharomyces* yeast strains during fermentation of wine. These yeasts are able to synthesize indolic compounds from tryptophan, and 5-hydroxy-L-tryptophan has an impact on melatonin pathways.⁴³ 4-Methyl-4-sulfanylpentan-2-one (ID 84) is a volatile thiol that has an impact on wine aroma and which has also been isolated from wines after fermentation by *S. cerevisiae* or non-*Saccharomyces* yeasts. Fermentation with *Metschnikowia pulcherrima* led to a high amount of this compound.⁴⁴

6-Methoxymellein (ID 40) was isolated from a marine-derived fungus collected from a marine sponge.⁴⁵ Kynurenic acid (ID 74) is an interesting metabolite that presents an antioxidant and neuroprotective activity. It is also produced by the yeast *Yarrowia lipolytica* and is involved in the tryptophan metabolic pathway.⁴⁶ Indole-3-pyruvic acid (ID 89) is a precursor in plant growth hormone synthesis, which was detected in the basidiomycetous yeast *Rhodospiridium paludigenum*.⁴⁷ Finally, malassezione (ID 338) is an indole alkaloid, a metabolite from the tryptophan pathways which has been isolated from the cultures of the human pathogenic yeast *Malassezia furfur*.⁴⁸

We found that these strains of new *Starmerella* species produce a higher number of metabolites in comparison with our previous work on the known *S. cerevisiae* S288c (CLIB 338), with, respectively, 439 and 513 detected features for *S. reginensis* and *S. kourouensis* and 128 detected features for *S. cerevisiae*.³¹ The analytical workflow was identical, but the analysis and interpretation workflow have been improved on the instrumentation part (ionization parameters) on the FTICR-MS data treatment and in the annotation process (by the addition of the NAP and the MolDiscovery tools, the refinement of annotation investigation, among others). The proportion of level 2 and 3 annotations is also more important. Concerning the distribution of chemical classes annotated among the detected features, the three yeast strains have a similar metabolic profile, with a large number of annotated alkaloids and some lipids, amino acids, and other small acids. However, the two *Starmerella* strains have more annotated lipids, especially terpenoids and steroids. It is interesting, especially for the food industry, for example, as terpenoids have already been reported to impact the organoleptic properties of wine after *S. cerevisiae* fermentation.^{49,50} Additionally, 46 features from the S288c metabolite table were also detected and annotated in the *Starmerella* table. None of the unknown molecules of the *Starmerella* strains have been detected in S288c, and only one unknown in S288c metabolite table was reported in the *Starmerella* yeast table (an alkaloid, ID 208, level 3). Roullier-Gall et al. also analyzed a *Starmerella* yeast, *S. bacillaris*, with a mass spectrometry approach.⁸ They profiled the exometabolome of a red wine fermented by a *S. bacillaris* strain by FTICR-MS. However, the sample nature is so distinct from ours that it is difficult to compare. Despite these clear differences, the *S. bacillaris* exometabolome fingerprinting shows many sugars, amino acids, anthocyanins, and many CHO molecules, which is a very different profile from those presented here. It is important to note here that the results of a metabolic profiling study on a microorganism extract are very dependent on the culture medium, and these results can be compared across studies only if the media are standardized.

In this work, extracts from *S. reginensis*- and *S. kourouensis*-type strains have been analyzed in parallel by FTICR-MS and UHPLC-MS/MS. The deep data treatment, investigation, and interpretation processes

have led to metabolic cartography represented by van Krevelen diagrams and a rich metabolite table with 439 attributed annotations for *S. reginensis* and 513 attributed annotations for *S. kourouensis*. The two explored strains showed a large content of small molecules, especially in terms of alkaloids and lipids, and the metabolic profiles are richer than the one we found in our previous study on the *S. cerevisiae* strain S288c. The present work contributes to increase the knowledge about yeast metabolites in these strains. This exploration on original strains opens the way and will serve as a reference to explore other *Starmerella* species.

Data Availability

Raw data are available on the MassIVE hub (<https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp>) under the data set number MSV000088013.

Associated Content

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jafc.1c03201>.

Brief description of analytical parameters for DFTICR–MS and UHPLC–MS/MS analyses, processing parameters for MZMine2 software, FTICR–MS metabolic profiling results in feature numbers per analysis and in van Krevelen diagrams, base peak ion chromatograms, molecular networks, standards that were used for level 1 identifications, complete annotated metabolite table for UHPLC-MS/MS analysis results, and mirror matches from the GNPS annotation process (PDF).

Author Information

Corresponding Author

Abdelhakim Elomri – Normandie Université, UNIROUEN, INSA Rouen, CNRS, COBRA (UMR 6014), Rouen 76000, France; orcid.org/0000-0001-5026-5558; Email: hakim.elomri@univ-rouen.fr

Authors

Olivier Perruchon – Normandie Université, UNIROUEN, INSA Rouen, CNRS, COBRA (UMR 6014), Rouen 76000, France

Isabelle Schmitz-Afonso – Normandie Université, UNIROUEN, INSA Rouen, CNRS, COBRA (UMR 6014), Rouen 76000, France

Cécile Grondin – Micalis Institute, INRAE, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas 78352, France; SPO, Univ Montpellier, INRAE, Institut Agro, Montpellier 34000, France

Jean-Luc Legras – SPO, Univ Montpellier, INRAE, Institut Agro, Montpellier 34000, France

Carlos Afonso – Normandie Université, UNIROUEN, INSA Rouen, CNRS, COBRA (UMR 6014), Rouen 76000, France; orcid.org/0000-0002-2406-5664

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.jafc.1c03201>

Funding

This work has been partially supported by the University of Rouen Normandy, INSA Rouen Normandy, the Centre National de la Recherche Scientifique (CNRS), European Regional Development Fund (ERDF), Labex SynOrg (ANR11-LABX-0029), Carnot Institut I2C, the graduate school for research XI-Chem (ANR-18-EURE-0020 XL CHEM), and by Region Normandie. This work was supported by the European Union's Horizon 2020 Research Infrastructures program (grant agreement 731077). Access to a CNRS FTICR research infrastructure (FR3624) is gratefully acknowledged.

Notes

The authors declare no competing financial interest.

Acknowledgements

The authors would like to thank Serge Casaregola, PhD., for his advice, Noémie Vignolles, assistant engineer, for the yeast collection, and the COBRA team for all the support.

Abbreviations

HPLC, high-performance liquid chromatography; LC–MS, liquid chromatography; MS2 or MS/MS, tandem mass spectrometry (fragmentation); UHPLC–MS/MS, ultrahigh-performance liquid chromatography hyphenated with tandem mass spectrometry; ACN, acetonitrile; DCM/EA extract, dichloromethane/ethyl acetate extract; M extract, methanol extract; FTICR–MS, Fourier-transform ion cyclotron resonance mass spectrometry; ESI-QTOF, electrospray ionization quadrupole-time of flight; FBMN, feature-based molecular networking; GNPS, Global Natural Product Social Molecular Networking; O/C and H/C ratios, oxygen/carbon and hydrogen/carbon ratios

References

- (1) Legras, J.-L.; Merdinoglu, D.; Cornuet, J.-M.; Karst, F. Bread, Beer and Wine: *Saccharomyces Cerevisiae* Diversity Reflects Human History. *Mol. Ecol.* **2007**, *16*, 2091–2102.
- (2) Nielsen, J. Production of Biopharmaceutical Proteins by Yeast. *Bioengineered* **2013**, *4*, 207–211.
- (3) Rosa, C. A.; Lachance, M.-A. The Yeast Genus *Starmerella* Gen. Nov. and *Starmerella Bombicola* Sp. Nov., the Teleomorph of *Candida Bombicola* (Spencer, Gorin & Tullock) Meyer & Yarrow. *Int. J. Syst. Bacteriol.* **1998**, *48*, 1413–1417.
- (4) Amoikon, T. L. S.; Grondin, C.; Djéni, T. N. D.; Jacques, N.; Casaregola, S. *Starmerella reginensis* f.a., sp. nov. and *Starmerella kourouensis* f.a., sp. nov., isolated from flowers in French Guiana. *Int. J. Syst. Evol. Microbiol.* **2018**, *68*, 2299–2305.
- (5) Gschaedler, A. Contribution of Non-Conventional Yeasts in Alcoholic Beverages. *Curr. Opin. Food Sci.* **2017**, *13*, 73–77.
- (6) De Bruyn, F.; Zhang, S. J.; Pothakos, V.; Torres, J.; Lambot, C.; Moroni, A. V.; Callanan, M.; Sybesma, W.; Weckx, S.; De Vuyst, L. Exploring the Impacts of Postharvest Processing on the Microbiota and Metabolite Profiles during Green Coffee Bean Production. *Appl. Environ. Microbiol.* **2017**, *83*, 1–16.
- (7) Englezos, V.; Pollon, M.; Rantsiou, K.; Ortiz-Julien, A.; Botto, R.; Río Segade, S.; Giacosa, S.; Rolle, L.; Cocolin, L. *Saccharomyces Cerevisiae*-*Starmerella Bacillaris* Strains Interaction Modulates Chemical and Volatile Profile in Red Wine Mixed Fermentations. *Food Res. Int.* **2019**, *122*, 392–401.
- (8) Roullier-Gall, C.; David, V.; Hemmler, D.; Schmitt-Kopplin, P.; Alexandre, H. Exploring Yeast Interactions through Metabolic Profiling. *Sci. Rep.* **2020**, *10*, 6073.
- (9) Garofalo, C.; Silvestri, G.; Aquilanti, L.; Clementi, F. PCRDGGE Analysis of Lactic Acid Bacteria and Yeast Dynamics during the Production Processes of Three Varieties of Panettone. *J. Appl. Microbiol.* **2008**, *105*, 243–254.
- (10) Kurtzman, C. P.; Price, N. P. J.; Ray, K. J.; Kuo, T.-M. Production of Sophorolipid Biosurfactants by Multiple Species of the *Starmerella* (*Candida*) *Bombicola* Yeast Clade. *FEMS Microbiol. Lett.* **2010**, *311*, 140–146.

- (11) Alimadadi, N.; Soudi, M. R.; Wang, S.-A.; Wang, Q.-M.; Talebpour, Z.; Bai, F.-Y. *Starmerella Orientalis* f.a., Sp. Nov., an Ascomycetous Yeast Species Isolated from Flowers. *Int. J. Syst. Evol. Microbiol.* **2016**, *66*, 1476–1481.
- (12) Chen, J.; Song, X.; Zhang, H.; Qu, Y. Production, Structure Elucidation and Anticancer Properties of Sophorolipid from *Wickerhamiella Domercqiae*. *Enzyme Microb. Technol.* **2006**, *39*, 501–506.
- (13) Tu, C.; Hu, W.; Tang, S.; Meng, L.; Huang, Z.; Xu, X.; Xia, X.; Azi, F.; Dong, M. Isolation and Identification of *Starmerella Davenportii* Strain Do18 and Its Application in Black Tea Beverage Fermentation. *Food Sci. Hum. Wellness* **2020**, *9*, 355–362.
- (14) Bhattacharya, S.; Gachhui, R.; Sil, P. C. Effect of Kombucha, a Fermented Black Tea in Attenuating Oxidative Stress Mediated Tissue Damage in Alloxan Induced Diabetic Rats. *Food Chem. Toxicol.* **2013**, *60*, 328–340.
- (15) Chassagne, F.; Cabanac, G.; Hubert, G.; David, B.; Marti, G. The Landscape of Natural Product Diversity and Their Pharmacological Relevance from a Focus on the Dictionary of Natural Products. *Phytochem. Rev.* **2019**, *18*, 601–622.
- (16) Stark, T. D.; Lösch, S.; Wakamatsu, J.; Balemba, O. B.; Frank, O.; Hofmann, T. UPLC-ESI-TOF MS-Based Metabolite Profiling of the Antioxidative Food Supplement *Garcinia Buchananii*. *J. Agric. Food Chem.* **2015**, *63*, 7169–7179.
- (17) Kind, T.; Tsugawa, H.; Cajka, T.; Ma, Y.; Lai, Z.; Mehta, S. S.; Wohlgemuth, G.; Barupal, D. K.; Showalter, M. R.; Arita, M.; Fiehn, O. Identification of Small Molecules Using Accurate Mass MS/MS Search. *Mass Spectrom. Rev.* **2018**, *37*, 513–532.
- (18) Wolfender, J.-L.; Nuzillard, J.-M.; Van Der Hooft, J. J. J.; Renault, J.-H.; Bertrand, S. Accelerating Metabolite Identification in Natural Product Research: Toward an Ideal Combination of Liquid Chromatography-High-Resolution Tandem Mass Spectrometry and NMR Profiling, in Silico Databases, and Chemometrics. *Anal. Chem.* **2019**, *91*, 704–742.
- (19) Pluskal, T.; Castillo, S.; Villar-Briones, A.; Oresič, M. MZmine 2: Modular Framework for Processing, Visualizing, and Analyzing Mass Spectrometry-Based Molecular Profile Data. *BMC Bioinformatics* **2010**, *11*, 395.
- (20) Wishart, D. S.; Tzur, D.; Knox, C.; Eisner, R.; Guo, A. C.; Young, N.; Cheng, D.; Jewell, K.; Arndt, D.; Sawhney, S.; Fung, C.; Nikolai, L.; Lewis, M.; Coutouly, M.-A.; Forsythe, I.; Tang, P.; Shrivastava, S.; Jeroncic, K.; Stothard, P.; Amegbey, G.; Block, D.; Hau, D. D.; Wagner, J.; Miniaci, J.; Clements, M.; Gebremedhin, M.; Guo, N.; Zhang, Y.; Duggan, G. E.; MacInnis, G. D.; Weljie, A. M.; Dowlatabadi, R.; Bamforth, F.; Clive, D.; Greiner, R.; Li, L.; Marrie, T.; Sykes, B. D.; Vogel, H. J.; Querengesser, L. HMDB: The Human Metabolome Database. *Nucleic Acids Res.* **2007**, *35*, D521–D526.
- (21) Horai, H.; Arita, M.; Kanaya, S.; Nihei, Y.; Ikeda, T.; Suwa, K.; Ojima, Y.; Tanaka, K.; Tanaka, S.; Aoshima, K.; Oda, Y.; Kakazu, Y.; Kusano, M.; Tohge, T.; Matsuda, F.; Sawada, Y.; Hirai, M. Y.; Nakanishi, H.; Ikeda, K.; Akimoto, N.; Maoka, T.; Takahashi, H.; Ara, T.; Sakurai, N.; Suzuki, H.; Shibata, D.; Neumann, S.; Iida, T.; Tanaka, K.; Funatsu, K.; Matsuura, F.; Soga, T.; Taguchi, R.; Saito, K.; Nishioka, T. MassBank: A Public Repository for Sharing Mass Spectral Data for Life Sciences. *J. Mass Spectrom.* **2010**, *45*, 703–714.
- (22) Wang, M.; Carver, J. J.; Phelan, V. V.; Sanchez, L. M.; Garg, N.; Peng, Y.; Nguyen, D. D.; Watrous, J.; Kaponov, C. A.; Luzzatto-Knaan, T.; Porto, C.; Bouslimani, A.; Melnik, A. V.; Meehan, M. J.; Liu, W.-T.; Crüsemann, M.; Boudreau, P. D.; Esquenazi, E.; Sandoval-Calderón, M.; Kersten, R. D.; Pace, L. A.; Quinn, R. A.; Duncan, K. R.; Hsu, C.-C.; Floros, D. J.; Gavilan, R. G.; Kleigrew, K.; Northen, T.; Dutton, R. J.; Parrot, D.; Carlson, E. E.; Aigle, B.; Michelsen, C. F.; Jelsbak, L.; Sohlenkamp, C.; Pevzner, P.; Edlund, A.; McLean, J.; Piel, J.; Murphy, B. T.; Gerwick,

L.; Liaw, C.-C.; Yang, Y.-L.; Humpf, H.- U.; Maansson, M.; Keyzers, R. A.; Sims, A. C.; Johnson, A. R.; Sidebottom, A. M.; Sedio, B. E.; Klitgaard, A.; Larson, C. B.; Boya P, C. A.; Torres-Mendoza, D.; Gonzalez, D. J.; Silva, D. B.; Marques, L. M.; Demarque, D. P.; Pociute, E.; O'Neill, E. C.; Briand, E.; Helfrich, E. J. N.; Granatosky, E. A.; Glukhov, E.; Ryffel, F.; Houson, H.; Mohimani, H.; Kharbush, J. J.; Zeng, Y.; Vorholt, J. A.; Kurita, K. L.; Charusanti, P.; McPhail, K. L.; Nielsen, K. F.; Vuong, L.; Elfeki, M.; Traxler, M. F.; Engene, N.; Koyama, N.; Vining, O. B.; Baric, R.; Silva, R. R.; Mascuch, S. J.; Tomasi, S.; Jenkins, S.; Macherla, V.; Hoffman, T.; Agarwal, V.; Williams, P. G.; Dai, J.; Neupane, R.; Gurr, J.; Rodríguez, A. M. C.; Lamsa, A.; Zhang, C.; Dorrestein, K.; Duggan, B. M.; Almaliti, J.; Allard, P.-M.; Phapale, P.; Nothias, L.-F.; Alexandrov, T.; Litaudon, M.; Wolfender, J.-L.; Kyle, J. E.; Metz, T. O.; Peryea, T.; Nguyen, D.-T.; VanLeer, D.; Shinn, P.; Jadhav, A.; Müller, R.; Waters, K. M.; Shi, W.; Liu, X.; Zhang, L.; Knight, R.; Jensen, P. R.; Palsson, B. Ø.; Pogliano, K.; Lington, R. G.; Gutiérrez, M.; Lopes, N. P.; Gerwick, W. H.; Moore, B. S.; Dorrestein, P. C.; Bandeira, N. Sharing and Community Curation of Mass Spectrometry Data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **2016**, 34, 828–837.

(23) Lyu, Q.; Wen, X.; Liu, Y.; Sun, C.; Chen, K.; Hsu, C.-C.; Li, X. Comprehensive Profiling of Phenolic Compounds in White and Red Chinese Bayberries (*Morella Rubra* Sieb. et Zucc.) and Their Developmental Variations Using Tandem Mass Spectral Molecular Networking. *J. Agric. Food Chem.* **2021**, 69, 741–749.

(24) da Silva, R. R.; Wang, M.; Nothias, L.-F.; van der Hooft, J. J. J.; Caraballo-Rodríguez, A. M.; Fox, E.; Balunas, M. J.; Klassen, J. L.; Lopes, N. P.; Dorrestein, P. C. Propagating Annotations of Molecular Networks Using in Silico Fragmentation. *PLoS Comput. Biol.* **2018**, 14, No. e1006089.

(25) Cao, L.; Guler, M.; Tagirdzhanov, A.; Lee, Y.; Gurevich, A.; Mohimani, H. MolDiscovery: Learning Mass Spectrometry Fragmentation of Small Molecules. **2020**, bioRxiv:2020.11.28.401943.

(26) Sumner, L. W.; Amberg, A.; Barrett, D.; Beale, M. H.; Beger, R.; Daykin, C. A.; Fan, T. W.-M.; Fiehn, O.; Goodacre, R.; Griffin, J. L.; Hankemeier, T.; Hardy, N.; Harnly, J.; Higashi, R.; Kopka, J.; Lane, A. N.; Lindon, J. C.; Marriott, P.; Nicholls, A. W.; Reilly, M. D.; Thaden, J. J.; Viant, M. R. Proposed Minimum Reporting Standards for Chemical Analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* **2007**, 3, 211–221.

(27) Ghaste, M.; Mistrik, R.; Shulaev, V. Applications of Fourier Transform Ion Cyclotron Resonance (FT-ICR) and Orbitrap Based High Resolution Mass Spectrometry in Metabolomics and Lipidomics. *Int. J. Mol. Sci.* **2016**, 17, 816.

(28) Roullier-Gall, C.; Signoret, J.; Hemmler, D.; Witting, M. A.; Kanawati, B.; Schäfer, B.; Gougeon, R. D.; Schmitt-Kopplin, P. Usage of FT-ICR-MS Metabolomics for Characterizing the Chemical Signatures of Barrel-Aged Whisky. *Front. Chem.* **2018**, 6, 29.

(29) van Krevelen, D. W. Graphical-Statistical Method for the Study of Structure and Reaction Processes of Coal. *Fuel* **1950**, 29, 269–284.

(30) Kim, S.; Kramer, R. W.; Hatcher, P. G. Graphical Method for Analysis of Ultrahigh-Resolution Broadband Mass Spectra of Natural Organic Matter, the Van Krevelen Diagram. *Anal. Chem.* **2003**, 75, 5336–5344.

(31) Perruchon, O.; Schmitz-Afonso, I.; Grondin, C.; Casaregola, S.; Afonso, C.; Elomri, A. Combination of UHPLC-MS/MS-Molecular Networking Approach and FTICR-MS for the Metabolic Profiling of *Saccharomyces Cerevisiae*. *J. Pharm. Biomed. Anal.* **2021**, 195, 113857.

(32) Goncalves, C.; Gonç, alves, P. Multilayered Horizontal Operon, Transfers from Bacteria Reconstruct a Thiamine Salvage Pathway in Yeasts. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, 116, 22219–22228.

- (33) Kind, T.; Fiehn, O. Seven Golden Rules for Heuristic Filtering of Molecular Formulas Obtained by Accurate Mass Spectrometry. *BMC Bioinformatics* **2007**, *8*, 105.
- (34) Maillard, J.; Ferey, J.; Rüger, C. P.; Schmitz-Afonso, I.; Bekri, S.; Gautier, T.; Carrasco, N.; Afonso, C.; Tebani, A. Optimization of Ion Trajectories in a Dynamically Harmonized Fourier-Transform Ion Cyclotron Resonance Cell Using a Design of Experiments Strategy. *Rapid Commun. Mass Spectrom.* **2020**, *34*, No. e8659.
- (35) Lavermicocca, P.; Valerio, F.; Visconti, A. Antifungal Activity of Phenyllactic Acid against Molds Isolated from Bakery Products. *Appl. Environ. Microbiol.* **2003**, *69*, 634–640.
- (36) Wilkins, A. L.; Lu, Y.; Molan, P. C. Extractable Organic Substances from New Zealand Unifloral Manuka (*Leptospermum Scoparium*) Honeys. *J. Apicult. Res.* **1993**, *32*, 3–9.
- (37) Beloborodov, N. V.; Khodakova, A. S.; Bairamov, I. T.; Olenin, A. Y. Microbial Origin of Phenylcarboxylic Acids in the Human Body. *Biochemistry* **2009**, *74*, 1350–1355.
- (38) Beloborodova, N.; Bairamov, I.; Olenin, A.; Shubina, V.; Teplova, V.; Fedotcheva, N. Effect of Phenolic Acids of Microbial Origin on Production of Reactive Oxygen Species in Mitochondria and Neutrophils. *J. Biomed. Sci.* **2012**, *19*, 89.
- (39) Herraiz, T. 1-Methyl-1,2,3,4-Tetrahydro- β -Carboline-3-Carboxylic Acid and 1,2,3,4-Tetrahydro- β -Carboline-3-Carboxylic Acid in Fruits. *J. Agric. Food Chem.* **1999**, *47*, 4883–4887.
- (40) Herraiz, T. Tetrahydro- β -Carbolines, Potential Neuroactive Alkaloids, in Chocolate and Cocoa. *J. Agric. Food Chem.* **2000**, *48*, 4900–4904.
- (41) Palmieri, A.; Petrini, M. Tryptophol and Derivatives: Natural Occurrence and Applications to the Synthesis of Bioactive Compounds. *Nat. Prod. Rep.* **2019**, *36*, 490–530.
- (42) Lee, H.-W.; Yang, J.-Y.; Lee, H.-S. Quinoline-2-Carboxylic Acid Isolated from *Ephedra pachyclada* and Its Structural Derivatives Show Inhibitory Effects against α -Glucosidase and α -Amylase. *J. Korean Soc. Appl. Biol. Chem.* **2014**, *57*, 441–444.
- (43) Fernández-Cruz, E.; Álvarez-Fernández, M. A.; Valero, E.; Troncoso, A. M.; García-Parrilla, M. C. Melatonin and Derived L-Tryptophan Metabolites Produced during Alcoholic Fermentation by Different Wine Yeast Strains. *Food Chem.* **2017**, *217*, 431–437.
- (44) Zott, K.; Thibon, C.; Bely, M.; Lonvaud-Funel, A.; Dubourdieu, D.; Masneuf-Pomarede, I. The Grape Must Non-Saccharomyces Microbial Community: Impact on Volatile Thiol Release. *Int. J. Food Microbiol.* **2011**, *151*, 210–215.
- (45) Höller, U.; König, G. M.; Wright, A. D. Three New Metabolites from Marine-Derived Fungi of the Genera *Coniothyrium* and *Microsphaeropsis*. *J. Nat. Prod.* **1999**, *62*, 114–118.
- (46) Wróbel-Kwiatkowska, M.; Turski, W.; Kocki, T.; Rakicka-Pustułka, M.; Rymowicz, W. An Efficient Method for Production of Kynurenic Acid by *Yarrowia lipolytica*. *Yeast* **2020**, *37*, 541–547.
- (47) Nutaratat, P.; Srisuk, N.; Arunrattiyakorn, P.; Limtong, S. Indole-3-Acetic Acid Biosynthetic Pathways in the Basidiomycetous Yeast *Rhodospidium paludigenum*. *Arch. Microbiol.* **2016**, *198*, 429–437.
- (48) Irlinger, B.; Bartsch, A.; Krämer, H.-J.; Mayser, P.; Steglich, W. New Tryptophan Metabolites from Cultures of the Lipophilic Yeast *Malassezia furfur*. *Helv. Chim. Acta* **2005**, *88*, 1472–1485.

(49) Alves, Z.; Melo, A.; Figueiredo, A. R.; Coimbra, M. A.; Gomes, A. C.; Rocha, S. M. Exploring the *Saccharomyces Cerevisiae* Volatile Metabolome: Indigenous versus Commercial Strains. *PLoS One* **2015**, 10, No. e0143641.

(50) Binati, R. L.; Lemos Junior, W. J. F.; Luzzini, G.; Slaghenaufi, D.; Ugliano, M.; Torriani, S. Contribution of Non-*Saccharomyces* Yeasts to Wine Volatile and Sensory Diversity: A Study on *Lachancea Thermotolerans*, *Metschnikowia* Spp. and *Starmerella Bacillaris* Strains Isolated in Italy. *Int. J. Food Microbiol.* **2020**, 318, 108470.

III. Perspectives : *Kurtzmaniella hittingeri* comme support de l'amélioration de la gamme de composés détectés

Une souche d'une autre nouvelle espèce de levure, *Kurtzmaniella hittingeri*, a été sélectionnée et utilisée au cours de ce projet de thèse à des fins d'optimisation de certaines étapes analytiques. La comparaison de plusieurs analyses d'extraits de ces cultures a été menée pour valider la méthode de conservation des échantillons entre l'extraction des métabolites et l'analyse par spectrométrie de masse. L'UHPLC-MS/MS est donc ici mise à contribution pour comparer des expériences espacées d'un an entre elles (contre une validation d'analyses espacées de six mois pour *S. cerevisiae*).

Cette souche de *K. hittingeri* représente également une matrice intéressante pour les perspectives du projet. D'un côté, il s'agit d'une nouvelle espèce à explorer par le processus de déréplication mise en place et amélioré pendant cette thèse, et d'un autre elle permettrait l'ajout de nouvelles fonctionnalités ou étapes afin de pousser le protocole analytique encore plus loin.

1. Confirmation de l'efficacité de conservation d'extraits secs

La stabilité du procédé de conservation des échantillons à $-25\text{ }^{\circ}\text{C}$ au congélateur a été vérifiée lors d'une étude réalisée sur des extraits de levure après 12 mois de conservation. Des extraits de la levure *K. hittingeri* ont été analysés par UHPLC-MS/MS après une période de conservation des échantillons longue de plus d'une année. Les données acquises lors de ces expériences furent comparées avec celles d'analyses d'extraits provenant de cultures différentes de la même souche de *K. hittingeri*, ces analyses ayant été menées un an auparavant, directement après extraction des métabolites, au cours du projet de thèse.

Les résultats de l'étude de stabilité sur les extraits DCM/AE et M de la levure en question ont montré moins de 10 % de perte moyenne en nombre de *features* détectées entre les analyses de 2019 et 2020, après pré-traitement grâce au logiciel MZMine2. Ces résultats sont à pondérer avec l'existence de *features* détectées exclusivement dans les analyses plus récentes. Les légères différences observables sont explicables par la variabilité biologique des deux cultures différentes analysées, mais aussi par le fait que la méthode d'analyse ait été améliorée entre les analyses datant de 2019 et celles de 2020. Ainsi, ces optimisations peuvent expliquer les *features* détectées en 2020 et non lors des analyses précédentes. Les améliorations du processus d'analyse impliquent également une fragmentation en source moindre, c'est-à-dire la création de fragments dès l'ionisation, avant la sélection par les

quadrupôles. Les fragments en source provenant de métabolites sont alors considérés comme des *features* à part entière, et la méthode de 2020 limite ce nombre, pouvant ainsi expliquer une partie des différences observables entre les données de 2019 et celles de 2020.

En conclusion, la méthode de conservation à – 25 °C dans un petit pilulier stocké au congélateur a été validée et adoptée dans le protocole final de préparation des extraits de levures au cours du projet.

2. Perspectives : améliorations analytiques par emploi de différentes sources d'ionisation

K. hittingeri est également l'espèce qui représente actuellement les perspectives de ce projet de cartographie des métabolites secondaires de levures. L'objectif serait d'utiliser cette souche pour l'utilisation de nouveaux modes d'ionisation en plus de la source ESI lors des analyses UHPLC-MS/MS et éventuellement FTICR-MS : les sources APCI et APPI. Il ne s'agit pas de remplacer les analyses par LC-ESI-MS/MS par d'autres mais bien de compléter les expériences réalisées par de nouvelles, avec la même séparation chromatographique (selon le même gradient) mais une analyse par spectrométrie de masse utilisant une autre technologie d'ionisation. De nouvelles expériences utilisant l'ionisation par APCI ou l'ionisation par APPI pourraient ainsi être réalisées.

L'objectif de l'ajout au protocole de nouvelles sources d'ionisation est l'amélioration de l'exhaustivité de l'exploration des extraits de levures par l'élargissement de la gamme de métabolites détectés. En effet, les sources APCI et APPI permettent la mise en évidence de composés plus apolaires, ayant pu échapper aux analyses par LC-ESI-MS/MS. Il peut s'agir par exemple de composés qui auraient pu être analysés par GC-MS. Les analyses pourront s'imbriquer simplement dans le processus de traitement de données, car les données de LC-MS peuvent être importées dans MZMine2, quelle que soit la source d'ionisation utilisée. Les données seront ensuite alignées, les *features* en commun regroupées, ne générant pas plus de temps de traitement supplémentaire que celui nécessaire à l'annotation des signaux détectés exclusivement grâce aux sources nouvellement introduites dans le protocole analytique.

Une série de premières analyses ont été réalisées par UHPLC-MS/MS sur des extraits de *K. hittingeri*, avec une source d'ionisation APPI. La différence majeure avec les analyses utilisant l'ESI comme source est l'ajout d'un flux de solvant utilisé comme dopant et aidant à l'ionisation (remplace l'ajout d'acide formique qui soutient le processus d'ionisation en ESI). Ici, un flux d'acétone a été sélectionné pour faciliter l'ionisation des composés du mélange. Malheureusement, les analyses ont montré la présence

d'une très grande pollution dans la sonde utilisée pour la source, et les résultats n'ont pas pu être exploités comme attendus.

Malgré ces expériences peu informatives, l'évocation de nouvelles méthodes reste prometteuse et représente une voie intéressante pour la poursuite du projet et pour l'évolution de la méthodologie mise au point au cours de cette thèse.

Conclusion

Cette thèse s'inscrit dans le domaine de la chimie des substances naturelles. Elle présente deux aspects qui ont pu évoluer conjointement. La méthodologie de déréplication d'extraits naturels a été développée avec pour but l'application à des souches de levures et la cartographie des métabolites, principalement secondaires. La méthode implique la chromatographie en phase liquide en couplage avec la spectrométrie de masse à haute résolution, les outils bio-informatiques de traitement de données et la spectrométrie de masse à ultra-haute résolution. Elle a pu ainsi être appliquée à l'exploration de souches de levures récemment découvertes en Guyane française.

La première phase porte sur le développement de la méthodologie, en commençant par la méthode d'extraction des métabolites. Le protocole vise à extraire une large gamme de métabolites en termes de polarité grâce à l'utilisation de solvants différents. L'extraction mise au point est répétable et les analyses permettent l'acquisition de résultats informatifs amenant à une interprétation des données répondant aux attentes du projet.

Les travaux réalisés sur la levure modèle *Saccharomyces cerevisiae* ont permis la mise en place d'une méthode analytique combinée à un traitement de données fiable et précis. La méthode est basée sur des analyses UHPLC-MS/MS avec des mesures de masse précises. La séparation chromatographique amène à la résolution des isomères, et les données de fragmentation ouvrent la porte à l'utilisation des réseaux moléculaires. Le procédé d'analyse a été complété par l'utilisation de la FTICR-MS. Celle-ci complémente idéalement l'UHPLC-MS/MS en levant les quelques ambiguïtés retrouvées avec cette dernière. Elle apporte une mesure de masse ultra-précise nécessaire à l'attribution de formules moléculaires uniques, améliorant drastiquement l'étape d'annotation des composés. La construction de diagrammes de van Krevelen permet la cartographie des métabolites détectés selon leurs rapports O/C et H/C.

L'étude de la souche S288c de l'espèce *S. cerevisiae*, qui devait initialement servir de modèle pour le développement de notre méthodologie, s'est révélée être une source de résultats inédits. Ceux-ci prouvent l'efficacité de la combinaison des deux techniques de spectrométrie de masse. Au total, 128 *features* exclusives à *S. cerevisiae* ont été détectées. 55 % d'entre elles sont annotées avec une structure ou une classe chimique en mode d'ionisation positif, contre 44 % en mode d'ionisation négatif après investigation des données acquises en UHPLC-MS/MS. L'ajout de la FTICR-MS est venu améliorer ces chiffres en les faisant passer respectivement à 89 % et 87 % tout en fournissant des cartographies par diagrammes de van Krevelen, servant ainsi l'objectif de profilage métabolique de la souche.

Ces résultats démontrent la richesse de *S. cerevisiae* en termes de métabolites. Cultivée sur un milieu riche, elle produit de nombreux composés parmi lesquels une large proportion d'alcaloïdes. Aussi bien indoliques que quinoléiques, ils représentent 58 % des composés annotés. On retrouve également des acides aminés et dérivés pour 17 % des annotations, ainsi que des lipides pour 11 % d'entre elles. De nombreux composés sont annotés pour la première fois chez *S. cerevisiae*, avec pour exemples l'acide paspalique ou le malassezindole A. À notre connaissance, il s'agit d'une première utilisation sur la levure de cette combinaison des deux méthodes de spectrométrie de masse conjointement avec l'utilisation d'outils de traitement avancés tels que la création de réseaux moléculaires. Les métabolites annotés et les cartographies générées sont des réalisations originales ayant permis la publication d'un article durant la thèse dans le *Journal of Pharmaceutical and Biomedical Analysis*, intitulé « Combination of UHPLC-MS/MS-molecular networking approach and FTICR-MS for the metabolic profiling of *Saccharomyces cerevisiae* ».

Cette première phase posa les fondations d'une méthodologie destinée à l'exploration rapide d'extraits naturels. Notre approche reçut néanmoins, après sa mise en place, des améliorations en lien direct avec les nouveaux outils en développement dans le domaine de la métabolomique basée sur l'analyse par spectrométrie de masse. Deux souches de deux nouvelles espèces de levures ont permis la mise en œuvre de ces améliorations. *Starmerella reginensis* et *Starmerella kourouensis* nous servirent à l'application des outils Network Annotation Propagation et MolDiscovery à partir de la plateforme GNPS, et firent office de matrices naturelles inédites à explorer.

L'investigation des métabolites des deux souches récemment découvertes par combinaison d'UHPLC-MS/MS et FTICR-MS, assistée par les nouveaux outils décrits précédemment, a permis de détecter et d'annoter 439 composés pour *S. reginensis* et 513 composés pour *S. kourouensis*, avec respectivement 73 % et 78 % d'annotations de composés ou de classes chimiques. Cela représente de très bonnes proportions d'annotation pour une étude métabolomique. L'ajout de nouveaux modules au procédé d'interprétation des données spectrales a aidé à attribuer des structures ou des classes chimiques à 75 *features* pour MolDiscovery et 16 autres *features* pour le cas du NAP. Les profils dressés pour les deux nouvelles souches sont similaires et présentent de nombreux alcaloïdes (114 annotations pour *S. reginensis* et 171 pour *S. kourouensis*). Un tel contenu en alcaloïdes est, à notre connaissance, inédit dans l'analyse métabolomique d'un extrait de levures. De nombreux lipides ont été également reportés, notamment des stéroïdes et des terpénoïdes, ainsi que des acides aminés et des acides organiques. Chaque souche a cependant ses propres métabolites, avec 51 composés exclusifs pour *S. reginensis* et 131 pour *S. kourouensis*.

On observe à partir de ces résultats une nette amélioration de la méthode, du point de vue analytique comme du côté de l'investigation des résultats. Ces résultats originaux ont fait l'objet d'une publication dans le *Journal of Agricultural and Food Chemistry* et intitulée « Metabolome Exploration by High-Resolution Mass Spectrometry Methodologies of Two New Yeast Species: *Starmerella reginensis* and *Starmerella kourouensis* ».

Nous avons pu, au cours de cette thèse, mettre en évidence des résultats d'exploration de sources naturelles originaux et complétant les connaissances actuelles sur les métabolites de la flore levure. La méthodologie mise au point et appliquée aux souches de levure est tout aussi intéressante. Les analyses permettent une déréplication efficace des échantillons extraits de sources naturelles, dans le sens où elles permettent l'annotation de nombreux composés sans avoir la nécessité de les isoler du mélange complexe. Elle se révèle efficace pour la cartographie des composés connus au sein des extraits analysés tout comme pour repérer les organismes susceptibles de produire des composés inconnus d'intérêt.

De multiples perspectives sont ouvertes pour la suite de cette thèse. Comme cela a pu être démontré le long de ce recueil, la méthode développée est évolutive et suit l'actualisation des technologies utilisées dans les études sur les substances naturelles. Elle pourra poursuivre son amélioration en fonction des avancées technologiques de l'instrumentation et du développement des outils informatisés de traitement de données. La perspective de l'utilisation d'autres sources d'ionisation lors des analyses par spectrométrie de masse témoigne des possibilités d'évolution existantes, dans le cas présent pour élargir la gamme de polarité des métabolites détectés. La déréplication des matrices naturelles peut se poursuivre avec l'étude de davantage de nouvelles souches de levures, afin d'approfondir les connaissances sur les métabolites de ces microorganismes. Il serait également possible de se concentrer sur les extraits disposant d'activités biologiques, et ce grâce à des tests, afin de déterminer rapidement quelles molécules produites par les levures sont intéressantes pour certains secteurs industriels. Enfin, la méthode développée au cours de ces travaux de thèse est devenue quasiment une routine pour l'exploration de produits naturels et d'extraits biologiques au sein du laboratoire, notamment dans l'analyse d'extraits de plantes. Elle s'avère ainsi, comme imaginée initialement, adaptable à l'exploration de nombreux types d'extraits biologiques et de matrices naturelles.

Matériel et méthodes

Bien que les protocoles analytiques soient détaillés dans les parties correspondantes, de plus amples détails concernant les paramètres des différentes expériences sont regroupés dans cette rubrique.

Matériel biologique

Les souches de levures utilisées dans ces travaux sont *Saccharomyces cerevisiae* S288c (CLIB 338), *Starmerella reginensis* (CLIB 1634 = CBS 15247), *Starmerella kourouensis* (CLIB 1707 = CLIB 15257) et *Kurtzmaniella hittingeri* (CLIB 1609 = CBS 13469).

Culture de levures

Toutes les souches de levures étudiées ont été cultivées suivant un même protocole. Une culture primaire des souches est réalisée à 26,5 °C sur un milieu gélifié YPD (Yeast Peptone Dextrose ; composition : 1 % d'extrait de levure, 1 % de peptone, 1 % de glucose et 1,4 % d'agar).

Des levures de la culture primaire sont prélevées et ensemencées par stries dans une boîte de Petri de 10 cm de diamètre sur un milieu gélifié YNB (6,7 % Yeast Nitrogen Base, 1 % de peptone, 1 % de glucose et 1,4 % d'agar) ; ces cultures secondaires sont incubées pendant 3 jours à l'étuve à 26,5 °C. Les milieux de culture utilisés comme témoins sont préparés de la même manière que les cultures secondaires.

Protocole final d'extraction de métabolites :

Après incubation, la gélose portant les colonies de levures est découpée en morceaux et placée dans un flacon en verre. Un volume de 15 mL de mélange dichlorométhane / acétate d'éthyle (1:1, v/v, solvants HPLC-grade) est ajouté dans le flacon, puis celui-ci subit un traitement de 30 minutes aux ultrasons (Branson B-5200-54, 450 W). Le solvant est collecté et filtré au travers d'un entonnoir à fritté dont la porosité est comprise entre 16 et 40 µm, afin d'éliminer les plus gros débris cellulaires et de milieu de culture. Le milieu de culture est repris dans 10 mL de mélange DCM / AE puis passé aux ultrasons pendant 10 min pour récupérer les métabolites qui n'auraient pas été extraits lors du premier contact avec le solvant. Le mélange est filtré et regroupé avec l'extrait précédent, puis évaporé sous vide en maintenant la température du bain à 40°C. L'évaporation fournit un extrait sec d'environ 10 mg

nommé extrait DCM/AE. L'extrait est ensuite conservé au congélateur à une température inférieure à $-20\text{ }^{\circ}\text{C}$.

En parallèle, 10 mL de méthanol (HPLC-grade) sont ajoutés aux restes de milieu de culture pour extraire les métabolites plus polaires. Le flacon est soumis à 5 min de traitement aux ultrasons, puis le mélange est filtré et collecté pour donner l'extrait M, à la texture pâteuse. La masse des extraits M est comprise en moyenne entre 100 et 200 mg. Ils sont ensuite conservés au congélateur à une température inférieure à $-20\text{ }^{\circ}\text{C}$.

Préparation des échantillons pour analyse

Les échantillons à analyser sont préparés avec des solvants LC-MS-grade. Les extraits DCM/AE sont solubilisés dans du méthanol et les extraits M dans de l'eau ; ils sont ensuite dilués dans un mélange eau / acétonitrile (1:1, v/v) jusqu'à une concentration finale d'environ 5 mg/mL. Les échantillons sont filtrés au travers de filtres PTFE (Polytétrafluoroéthylène) de porosité 0,45 μm . Ils sont ensuite analysés par UHPLC-MS/MS. Des standards purs sont dissouts dans le même mélange eau/ACN puis dilués jusqu'à une concentration de 0,01 mg/mL (Tableau 5).

Tableau 5 : standards physiques utilisés pour les analyses UHPLC-MS/MS pour les éventuelles annotations de niveau 1.

Molecular formula	Compound	CAS	Fournisseur
C ₁₁ H ₁₃ NO ₃	N-Acetyl-L-phenylalanine	2018-61-3	TCI
C ₉ H ₁₀ O ₃	3-Phenyllactic acid	703-80-0	Aldrich
C ₁₀ H ₉ NO	3-Acetylindole	87-51-4	Acros Organics
C ₁₀ H ₉ N	6-Methylquinoline	495-69-2	Schuhschand München
C ₁₀ H ₇ NO ₃	4-Hydroxyquinoline-2-carboxylic acid	526-55-6	Sigma
C ₁₀ H ₉ NO ₂	Indole-3-acetic acid	20312-36-1	Alfa Aesar
C ₁₀ H ₁₁ NO	Tryptophol	91-62-3	Acros Organics

Un volume de 10 µL est prélevé de chaque extrait précédemment analysé en UHPLC-MS/MS et dilué à 1/10000 dans le mélange eau/ACN. Ceux-ci sont ensuite analysés par FTICR-MS.

Les blancs analytiques sont préparés avec le mélange eau/ACN (1:1, v/v, LC-MS-grade).

Analyse des extraits par CCM

Les premiers extraits réalisés à partir de cultures de *S. cerevisiae* ont été analysés par de rapides expériences de CCM afin de vérifier l'efficacité des extractions.

Une fraction des extraits DCM/AE et M secs est reprise dans du méthanol et déposée sur une plaque de silice. Un mélange DCM/AE (50/50, v/v) est utilisé comme phase mobile pour comparer les extraits de levure DCM/AE et M d'un côté, puis pour comparer les extraits DCM/AE de levure et de milieu de culture d'autre part. Les composés extraits exclusifs aux levures sont séparés. La révélation est réalisée dans un premier temps sous lampe UV à 254 nm, puis après pulvérisation de la vanilline sulfurique suivie du chauffage.

Analyses des extraits par UHPLC-MS/MS

La chromatographie liquide est réalisée avec un système de chromatographie UHPLC Dionex Ultimate 3000 UPLC+ (Thermo Scientific, San Jose, CA, USA). Les échantillons sont chargés sur un échantillonneur automatique thermostaté à 4 °C. 1 µL de chaque échantillon est injecté selon les séquences définies.

Les métabolites extraits de levures sont séparés selon leur polarité au travers d'une colonne à base de silice greffée de C18 (Acquity UPLC HSS T3 1.8 µm × 1.0 mm × 100 mm, Waters Corporation, Milford, MA, USA), chauffée à 40 °C et équipée d'un préfiltre de diamètre 0,2 µm.

Les solvants utilisés comme phases mobiles sont un mélange eau + 0,1 % d'acide formique (noté en voie A) et un mélange acétonitrile + 0,1 % d'acide formique (noté en voie B). Les solvants et l'acide formique sont de qualité LC-MS-grade. Le flux de solvant est réglé à 0,1 mL/min. Le gradient débute à 99 % A / 1 % B. Après 1 min, le gradient se déplace de façon linéaire jusqu'à 100 % B à 23 min ; cette condition est maintenue pendant 2 min. Le gradient retourne à la situation initiale (99 % A / 1 % B) à 25,2 min puis se stabilise jusqu'à la fin de l'expérience à 30 min.

Les métabolites séparés sont détectés à l'aide d'un spectromètre de masse hybride quadrupôle-temps de vol équipé d'une source ESI (SYNAPT G2 HDMS, Waters MS Technologies, Manchester, UK; contrôlé par le logiciel MassLynx 4.1). Les échantillons sont analysés en mode d'ionisation positif et négatif, pour une gamme de m/z 50-2000 et selon une acquisition dépendante des données (Waters Survey Analysis). Les paramètres de la source d'ionisation et de l'acquisition en mode Survey sont indiqués en annexe, en page 225.

L'étalonnage en masse est réalisé en externe avant les analyses à l'aide d'un mélange de formiate de sodium (10 % d'acide formique / 0,1 M de NaOH / acétonitrile en proportions 1:1:8, v/v/v). L'étalonnage est également réalisé en interne et en temps réel pendant les analyses avec une solution de leucine-enképhaline (à 2 ng/mL dans l'eau / isopropanol 1:1, v/v) injectée dans le système de *Lock Mass* du système Waters (étalonnage grâce aux ions à m/z 556,2771 et m/z 554,2615 respectivement en mode positif et négatif).

Traitement des données UHPLC-MS/MS

Les données acquises en UHPLC-MS/MS sont centroïdées grâce au module *Automatic peak detection* de MassLynx 4.1 avant d'être importées sous MZMine2. Les masses sont détectées via l'algorithme *Exact mass* puis utilisées pour la construction de chromatogrammes via l'algorithme *ADAP chromatogram builder* [180]. Ces chromatogrammes créés forment des listes de *features* caractérisées par les rapports m/z , les temps de rétention et les données spectrales de fragmentation associées. Ils sont ensuite déconvolués (via l'algorithme *Wavelet ADAP*) afin d'éliminer les signaux du bruit de fond, séparer les isomères et lier les données de MS/MS avec les *features* correspondantes. Les isotopes et les composés détectés en doubles sont retirés.

Les listes de *features* des analyses d'extraits de levures sont comparées avec celles de milieu de culture. Les formules chimiques sont attribuées automatiquement via le module *Formula prediction* selon des paramètres suivant les sept règles d'or [172]. Les fragments en source et adduits sont annotés avec les modules correspondants, qui comparent les données de MS1 et de MS2.

Les paramètres détaillés et valeurs numériques correspondantes sont donnés en annexe à la page 226. Après export, les données sont importées sur la plateforme GNPS pour création de réseaux moléculaires selon l'algorithme FBMN. Voici le paragraphe fourni automatiquement par le site web après construction de réseaux, avec les paramètres numériques indiqués :

"A molecular network was created with the Feature-Based Molecular Networking (FBMN) workflow on GNPS (<https://gnps.ucsd.edu>). MS/MS spectra were window-filtered by choosing only the top six fragment ions in the ± 50 Da window throughout the spectrum. The precursor ion mass tolerance was set to 0.02 Da and the MS/MS fragment ion tolerance to 0.02 Da. A molecular network was then created, where edges were filtered to have a cosine score above 0.7 and more than 3 matched peaks. Furthermore, edges between two nodes were kept in the network if and only if each of the nodes appeared in each other's respective top 10 most similar nodes. Finally, the maximum size of a molecular family was set to 100, and the lowest scoring edges were removed from molecular families until the molecular family size was below this threshold. The spectra in the network were then searched against GNPS spectral libraries. The library spectra were filtered in the same manner as the input data. All matches kept between network spectra, and library spectra were required to have a score above 0.7 and at least 3 matched peaks."

Analyses des extraits par FTICR-MS

Les extraits de cultures de levures et de milieux de culture sont analysés sur un instrument SolariX XR FTMS (Bruker Daltonics) équipé d'un aimant supraconducteur de 12 Tesla et d'une cellule ICR harmonisée dynamiquement. Les échantillons (pour rappel, dilués 10 000 fois par rapport aux analyses LC-MS) sont injectés directement, sans séparation, et ionisés par une source ESI, tantôt en mode d'ionisation positif tantôt en mode d'ionisation négatif. Les paramètres sélectionnés pour l'ionisation et l'analyse, détaillés en page 240, ont permis une résolution de 600 000 à m/z 200. Pour réaliser un étalonnage interne, le mélange Agilent TuneMix est utilisé (liste des ions en page 240). La liste d'ions d'étalonnage interne est complétée par des ions spécifiques des extraits analysés pour garantir la précision de mesure (en général inférieure à 0,1 ppm, étalonnage en mode quadratique).

Traitement des données FTICR-MS

Le logiciel Data Analysis 5.0 (Bruker) est utilisé pour traiter les données acquises. Les rapport m/z mesurés dans l'analyse des extraits de levure sont comparés à ceux des analyses de milieu de culture afin de soustraire ces derniers aux premières analyses ; il ne reste ainsi que les signaux exclusifs aux analyses des microorganismes. Ceci est réalisé grâce à l'algorithme *Xpose*, paramétré sur un ratio de 5.

Les formules moléculaires sont attribuées selon les nombres d'éléments suivants : $C_{1-100}H_{0-200}N_{0-8}O_{0-20}S_{0-1}$ dans un premier temps, puis les formules possibles avec $C_{1-100}H_{0-200}N_{0-8}O_{0-20}P_{0-1}$ sont ajoutées. Les paramètres respectent les sept règles d'or de l'attribution de formule moléculaire, afin d'éviter les attributions aberrantes (tolérance à 0,3 ppm, cycles et équivalents double-liaisons entre 0 et 80 et ratio H/C compris entre 0 et 3). Les adduits considérés sont M + H, Na et K en mode positif et M – H et + Cl en mode négatif.

Enfin, les listes de formules moléculaires générées puis vérifiées manuellement afin d'éliminer les formules aberrantes sont exportées, les rapports O/C et H/C sont calculés et utilisés pour générer des diagrammes de van Krevelen.

Analyses des extraits par GC-MS

Un extrait DCM/AE d'une culture de *S. cerevisiae* a été préparé pour analyse par GC-MS. Les composés de l'échantillon sont séparés sur un appareil GC System 7890A (Agilent Technologies). La séparation chromatographique dure 30 min avec un débit de gaz réglé à 1 mL/min, en mode *splitless*. Ils sont ensuite analysés grâce à un spectromètre de masse AccuTOF GCv 4G (JEOL). Celui-ci est équipé d'une source EI et d'un analyseur TOF, et est étalonné en externe avant l'expérience grâce à du perfluorokérosène. Les paramètres d'ionisation sont réglés à 70 V pour la tension d'ionisation, 300 mA pour le courant d'ionisation. L'analyse est réglée avec un délai de solvant à 3 minutes (pour ne pas acquérir le front de solvant du début d'analyse) et un détecteur à 2200 V. La gamme de m/z mesurée est fixée à 35-800.

Références

- [1] B. Petrovska, Historical review of medicinal plants' usage, *Pharmacogn. Rev.* 6 (2012) 1. <https://doi.org/10.4103/0973-7847.95849>.
- [2] D.H. Williams, M.J. Stone, P.R. Hauck, S.K. Rahman, Why Are Secondary Metabolites (Natural Products) Biosynthesized?, *J. Nat. Prod.* 52 (1989) 1189–1208. <https://doi.org/10.1021/np50066a001>.
- [3] A. Mithöfer, W. Boland, Plant Defense Against Herbivores: Chemical Aspects, *Annu. Rev. Plant Biol.* 63 (2012) 431–450. <https://doi.org/10.1146/annurev-arplant-042110-103854>.
- [4] M. Zaynab, M. Fatima, S. Abbas, Y. Sharif, M. Umair, M.H. Zafar, K. Bahadar, Role of secondary metabolites in plant defense against pathogens, *Microb. Pathog.* 124 (2018) 198–202. <https://doi.org/10.1016/j.micpath.2018.08.034>.
- [5] M.L. Cipollini, D.J. Levey, Secondary Metabolites Of Fleshy Vertebrate-Dispersed Fruits: Adaptive Hypotheses And Implications For Seed Dispersal, *Am. Nat.* 150 (1997) 346–372. <https://doi.org/10.1086/286069>.
- [6] M. González-Teuber, M. Heil, Nectar chemistry is tailored for both attraction of mutualists and protection from exploiters, *Plant Signal. Behav.* 4 (2009) 809–813. <https://doi.org/10.4161/psb.4.9.9393>.
- [7] E. Pichersky, D.R. Gang, Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective, *Trends Plant Sci.* 5 (2000) 439–445. [https://doi.org/10.1016/S1360-1385\(00\)01741-6](https://doi.org/10.1016/S1360-1385(00)01741-6).
- [8] J.N. Kabera, E. Semana, A.R. Mussa, X. He, Plant Secondary Metabolites: Biosynthesis, Classification, Function and Pharmacological Properties, *J. Pharm. Pharmacol.* 2 (2014) 377–392.
- [9] F. Chassagne, G. Cabanac, G. Hubert, B. David, G. Marti, The landscape of natural product diversity and their pharmacological relevance from a focus on the Dictionary of Natural Products®, *Phytochem. Rev.* 18 (2019) 601–622. <https://doi.org/10.1007/s11101-019-09606-2>.
- [10] R. Verpoorte, Exploration of nature's chemodiversity: the role of secondary metabolites as leads in drug development, *Drug Discov. Today.* 3 (1998) 232–238. [https://doi.org/10.1016/S1359-6446\(97\)01167-7](https://doi.org/10.1016/S1359-6446(97)01167-7).
- [11] M.M. Caldwell, J.F. Bornman, C.L. Ballar, S.D. Flint, G. Kulandaivelu, Terrestrial ecosystems, increased solar ultraviolet radiation, and interactions with other climate change factors, *Photochem. Photobiol. Sci.* 6 (2007) 252. <https://doi.org/10.1039/b700019g>.
- [12] D. Lucas-Barbosa, P. Sun, A. Hakman, T.A. Beek, J.J.A. Loon, M. Dicke, Visual and odour cues: plant responses to pollination and herbivory affect the behaviour of flower visitors, *Funct. Ecol.* 30 (2016) 431–441. <https://doi.org/10.1111/1365-2435.12509>.
- [13] K. Lundquist, J. Parkås, Different types of phenolic units in lignins, *BioResources.* 6 (2011) 920–926. <https://doi.org/10.15376/biores.6.2.920-926>.
- [14] E.-S. Park, W.-S. Moon, M.-J. Song, M.-N. Kim, K.-H. Chung, J.-S. Yoon, Antimicrobial activity of phenol and benzoic acid derivatives, *Int. Biodeterior. Biodegradation.* 47 (2001) 209–214. [https://doi.org/10.1016/S0964-8305\(01\)00058-0](https://doi.org/10.1016/S0964-8305(01)00058-0).
- [15] D.A. Evans, C.H. Mitch, Studies directed towards the total synthesis of morphine alkaloids, *Tetrahedron Lett.* 23 (1982) 285–288. [https://doi.org/10.1016/S0040-4039\(00\)86810-0](https://doi.org/10.1016/S0040-4039(00)86810-0).
- [16] C.A. Haller, P. Jacob, N.L. Benowitz, Pharmacology of ephedra alkaloids and caffeine after single-dose dietary supplement use, *Clin. Pharmacol. Ther.* 71 (2002) 421–432. <https://doi.org/10.1067/mcp.2002.124523>.
- [17] J.W. Fairbairn, A. Paterson, G. Wassel, The alkaloids of *Papaver somniferum* L.—II., *Phytochemistry.* 3 (1964) 577–582. [https://doi.org/10.1016/S0031-9422\(00\)82930-1](https://doi.org/10.1016/S0031-9422(00)82930-1).
- [18] E.J. Staba, A.C. Chung, Quinine and quinidine production by cinchona leaf, root and unorganized cultures, *Phytochemistry.* 20 (1981) 2495–2498. [https://doi.org/10.1016/0031-9422\(81\)83079-8](https://doi.org/10.1016/0031-9422(81)83079-8).
- [19] G.A. Strobel, A. Stierle, W.M. Hess, Taxol formation in yew — *Taxus*, *Plant Sci.* 92 (1993) 1–12. [https://doi.org/10.1016/0168-9452\(93\)90060-D](https://doi.org/10.1016/0168-9452(93)90060-D).

- [20] J. Gershenzon, N. Dudareva, The function of terpene natural products in the natural world, *Nat. Chem. Biol.* 3 (2007) 408–414. <https://doi.org/10.1038/nchembio.2007.5>.
- [21] G. Styger, B. Prior, F.F. Bauer, Wine flavor and aroma, *J. Ind. Microbiol. Biotechnol.* 38 (2011) 1145–1159. <https://doi.org/10.1007/s10295-011-1018-4>.
- [22] F. Bourgaud, A. Gravot, S. Milesi, E. Gontier, Production of plant secondary metabolites: A historical perspective, *Plant Sci.* 161 (2001) 839–851. [https://doi.org/10.1016/S0168-9452\(01\)00490-3](https://doi.org/10.1016/S0168-9452(01)00490-3).
- [23] C.P. Kurtzman, J.W. Fell, T. Boekhout, *The Yeasts: A Taxonomic Study* 5th Edition, 2011.
- [24] R. Mortimer, M. Polsinelli, On the origins of wine yeast, *Res. Microbiol.* 150 (1999) 199–204. [https://doi.org/10.1016/S0923-2508\(99\)80036-9](https://doi.org/10.1016/S0923-2508(99)80036-9).
- [25] I. Stefanini, L. Dapporto, J.-L. Legras, A. Calabretta, M. Di Paola, C. De Filippo, R. Viola, P. Capretti, M. Polsinelli, S. Turillazzi, D. Cavalieri, Role of social wasps in *Saccharomyces cerevisiae* ecology and evolution, *Proc. Natl. Acad. Sci.* 109 (2012) 13398–13403. <https://doi.org/10.1073/pnas.1208362109>.
- [26] C.C. Buser, R.D. Newcomb, A.C. Gaskett, M.R. Goddard, Niche construction initiates the evolution of mutualistic interactions, *Ecol. Lett.* 17 (2014) 1257–1264. <https://doi.org/10.1111/ele.12331>.
- [27] A. Goffeau, B.G. Barrell, H. Bussey, R.W. Davis, B. Dujon, H. Feldmann, F. Galibert, J.D. Hoheisel, C. Jacq, M. Johnston, E.J. Louis, H.W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, S.G. Oliver, Life with 6000 Genes, *Science* (80-.). 274 (1996) 546–567. <https://doi.org/10.1126/science.274.5287.546>.
- [28] J.L. Legras, D. Merdinoglu, J.M. Cornuet, F. Karst, Bread, beer and wine: *Saccharomyces cerevisiae* diversity reflects human history, *Mol. Ecol.* 16 (2007) 2091–2102. <https://doi.org/10.1111/j.1365-294X.2007.03266.x>.
- [29] M. Parapouli, A. Vasileiadi, A.-S. Afendra, E. Hatziloukas, *Saccharomyces cerevisiae* and its industrial applications, *AIMS Microbiol.* 6 (2020) 1–32. <https://doi.org/10.3934/microbiol.2020001>.
- [30] S. Kwak, Y.-S. Jin, Production of fuels and chemicals from xylose by engineered *Saccharomyces cerevisiae*: a review and perspective, *Microb. Cell Fact.* 16 (2017) 82. <https://doi.org/10.1186/s12934-017-0694-9>.
- [31] L. Favaro, T. Jansen, W.H. van Zyl, Exploring industrial and natural *Saccharomyces cerevisiae* strains for the bio-based economy from biomass: the case of bioethanol, *Crit. Rev. Biotechnol.* 39 (2019) 800–816. <https://doi.org/10.1080/07388551.2019.1619157>.
- [32] W. Jilg, M. Schmidt, G. Zoulek, B. Lorbeer, B. Wilske, F. Deinhardt, Clinical Evaluation of a Recombinant Hepatitis B Vaccine, *Lancet.* 324 (1984) 1174–1175. [https://doi.org/10.1016/S0140-6736\(84\)92740-5](https://doi.org/10.1016/S0140-6736(84)92740-5).
- [33] K.J. Hofmann, J.C. Cook, J.G. Joyce, D.R. Brown, L.D. Schultz, H.A. George, M. Rosolowsky, K.H. Fife, K.U. Jansen, Sequence Determination of Human Papillomavirus Type 6a and Assembly of Virus-like Particles in *Saccharomyces cerevisiae*, *Virology.* 209 (1995) 506–518. <https://doi.org/10.1006/viro.1995.1283>.
- [34] J. Nielsen, Production of biopharmaceutical proteins by yeast, *Bioengineered.* 4 (2013) 207–211. <https://doi.org/10.4161/bioe.22856>.
- [35] J.C. du Preez, M. Bosch, B.A. Prior, The fermentation of hexose and pentose sugars by *Candida shehatae* and *Pichia stipitis*, *Appl. Microbiol. Biotechnol.* 23 (1986) 228–233. <https://doi.org/10.1007/BF00261920>.
- [36] A. Soden, I.L. Francis, H. Oakey, P.A. Henschke, Effects of co-fermentation with *Candida stellata* and *Saccharomyces cerevisiae* on the aroma and composition of Chardonnay wine, *Aust. J. Grape Wine Res.* 6 (2000) 21–30. <https://doi.org/10.1111/j.1755-0238.2000.tb00158.x>.
- [37] C.A. Rosa, M.-A. Lachance, The yeast genus *Starmerella* gen. nov. and *Starmerella bombicola* sp. nov., the teleomorph of *Candida bombicola* (Spencer, Gorin & Tullock) Meyer & Yarrow, *Int. J. Syst. Bacteriol.* 48 (1998) 1413–1417. <https://doi.org/10.1099/00207713-48-4-1413>.
- [38] C. Garofalo, G. Silvestri, L. Aquilanti, F. Clementi, PCR-DGGE analysis of lactic acid bacteria and yeast dynamics during the production processes of three varieties of Panettone, *J. Appl. Microbiol.* 105 (2008) 243–254. <https://doi.org/10.1111/j.1365-2672.2008.03768.x>.
- [39] C.P. Kurtzman, N.P.J. Price, K.J. Ray, T.M. Kuo, Production of sophorolipid biosurfactants by multiple species of the *Starmerella* (*Candida*) *bombicola* yeast clade, *FEMS Microbiol. Lett.* 311 (2010) 140–146.

<https://doi.org/10.1111/j.1574-6968.2010.02082.x>.

- [40] J. Chen, X. Song, H. Zhang, Y. Qu, Production, structure elucidation and anticancer properties of sophorolipid from *Wickerhamiella domercqiae*, *Enzyme Microb. Technol.* 39 (2006) 501–506. <https://doi.org/10.1016/j.enzmictec.2005.12.022>.
- [41] F. De Bruyn, S.J. Zhang, V. Pothakos, J. Torres, C. Lambot, A. V. Moroni, M. Callanan, W. Sybesma, S. Weckx, L. De Vuyst, Exploring the Impacts of Postharvest Processing on the Microbiota and Metabolite Profiles during Green Coffee Bean Production, *Appl. Environ. Microbiol.* 83 (2017) 1–16. <https://doi.org/10.1128/AEM.02398-16>.
- [42] S. Bhattacharya, R. Gachhui, P.C. Sil, Effect of Kombucha, a fermented black tea in attenuating oxidative stress mediated tissue damage in alloxan induced diabetic rats, *Food Chem. Toxicol.* 60 (2013) 328–340. <https://doi.org/10.1016/j.fct.2013.07.051>.
- [43] C. Tu, W. Hu, S. Tang, L. Meng, Z. Huang, X. Xu, X. Xia, F. Azi, M. Dong, Isolation and identification of *Starmerella davenportii* strain Do18 and its application in black tea beverage fermentation, *Food Sci. Hum. Wellness.* 9 (2020) 355–362. <https://doi.org/10.1016/j.fshw.2020.04.010>.
- [44] A. Gschaedler, Contribution of non-conventional yeasts in alcoholic beverages, *Curr. Opin. Food Sci.* 13 (2017) 73–77. <https://doi.org/10.1016/j.cofs.2017.02.004>.
- [45] V. Englezos, M. Pollon, K. Rantsiou, A. Ortiz-Julien, R. Botto, S. Río Segade, S. Giacosa, L. Rolle, L. Cocolin, *Saccharomyces cerevisiae*-*Starmerella bacillaris* strains interaction modulates chemical and volatile profile in red wine mixed fermentations, *Food Res. Int.* 122 (2019) 392–401. <https://doi.org/10.1016/j.foodres.2019.03.072>.
- [46] C. Roullier-Gall, V. David, D. Hemmler, P. Schmitt-Kopplin, H. Alexandre, Exploring yeast interactions through metabolic profiling, *Sci. Rep.* 10 (2020) 6073. <https://doi.org/10.1038/s41598-020-63182-6>.
- [47] M.-A. Lachance, W.T. Starmer, *Kurtzmaniella* gen. nov. and description of the heterothallic, haplontic yeast species *Kurtzmaniella cleridarum* sp. nov., the teleomorph of *Candida cleridarum*, *Int. J. Syst. Evol. Microbiol.* 58 (2008) 520–524. <https://doi.org/10.1099/ijs.0.65460-0>.
- [48] C.P. Kurtzman, C.J. Robnett, Description of *Kuraishia piskuri* f.a., sp. nov., a new methanol assimilating yeast and transfer of phylogenetically related *Candida* species to the genera *Kuraishia* and *Nakazawaea* as new combinations, *FEMS Yeast Res.* 14 (2014) 1028–1036. <https://doi.org/10.1111/1567-1364.12192>.
- [49] H.-M. Daniel, M.-A. Lachance, C.P. Kurtzman, On the reclassification of species assigned to *Candida* and other anamorphic ascomycetous yeast genera based on phylogenetic circumscription, *Antonie Van Leeuwenhoek.* 106 (2014) 67–84. <https://doi.org/10.1007/s10482-014-0170-z>.
- [50] L. Belleggia, I. Ferrocino, A. Reale, F. Boscaino, T. Di Renzo, M.R. Corvaglia, L. Cocolin, V. Milanović, F. Cardinali, C. Garofalo, F. Clementi, L. Aquilanti, A. Osimani, Portuguese cacholeira blood sausage: A first taste of its microbiota and volatile organic compounds, *Food Res. Int.* 136 (2020) 109567. <https://doi.org/10.1016/j.foodres.2020.109567>.
- [51] L. Belleggia, V. Milanović, I. Ferrocino, L. Cocolin, M.N. Haouet, S. Scuota, A. Maoloni, C. Garofalo, F. Cardinali, L. Aquilanti, M. Mozzon, R. Foligni, M. Pasquini, M.F. Trombetta, F. Clementi, A. Osimani, Is there any still undisclosed biodiversity in Ciauscolo salami? A new glance into the microbiota of an artisan production as revealed by high-throughput sequencing, *Meat Sci.* 165 (2020) 108128. <https://doi.org/10.1016/j.meatsci.2020.108128>.
- [52] J. Vicente, J. Ruiz, I. Belda, I. Benito-Vázquez, D. Marquina, F. Calderón, A. Santos, S. Benito, The Genus *Metschnikowia* in Enology, *Microorganisms.* 8 (2020) 1038. <https://doi.org/10.3390/microorganisms8071038>.
- [53] D. Spadaro, A. Ciavarella, Z. Dianpeng, A. Garibaldi, M.L. Gullino, Effect of culture media and pH on the biomass production and biocontrol efficacy of a *Metschnikowia pulcherrima* strain to be used as a biofungicide for postharvest disease control, *Can. J. Microbiol.* 56 (2010) 128–137. <https://doi.org/10.1139/W09-117>.
- [54] E.C. Horning, M.G. Horning, Metabolic Profiles: Gas-Phase Methods for Analysis of Metabolites, *Clin. Chem.* 17 (1971) 802–809. <https://doi.org/10.1093/clinchem/17.8.802>.
- [55] S. Alseekh, A.R. Fernie, Metabolomics 20 years on: what have we learned and what hurdles remain?, *Plant J.* 94 (2018) 933–942. <https://doi.org/10.1111/tpj.13950>.
- [56] S.G. Oliver, M.K. Winson, D.B. Kell, F. Baganz, Systematic functional analysis of the yeast genome, *Trends Biotechnol.* 16 (1998) 373–378. [https://doi.org/10.1016/S0167-7799\(98\)01214-1](https://doi.org/10.1016/S0167-7799(98)01214-1).

- [57] L.W. Sumner, P. Mendes, R.A. Dixon, Plant metabolomics: large-scale phytochemistry in the functional genomics era, *Phytochemistry*. 62 (2003) 817–836. [https://doi.org/10.1016/S0031-9422\(02\)00708-2](https://doi.org/10.1016/S0031-9422(02)00708-2).
- [58] E. Jellum, Profiling of human body fluids in healthy and diseased states using gas chromatography and mass spectrometry, with special reference to organic acids, *J. Chromatogr. B Biomed. Sci. Appl.* 143 (1977) 427–462. [https://doi.org/10.1016/S0378-4347\(00\)81792-2](https://doi.org/10.1016/S0378-4347(00)81792-2).
- [59] S.C. Gates, N. Dendramis, C.C. Sweeley, Automated metabolic profiling of organic acids in human urine. I. Description of methods., *Clin. Chem.* 24 (1978) 1674–1679. <https://doi.org/10.1093/clinchem/24.10.1674>.
- [60] C. Ning, T. Kuhara, Y. Inoue, C.-H. Zhang, M. Matsumoto, T. Shinka, T. Furumoto, K. Yokota, I. Matsumoto, Gas chromatographic-mass spectrometric metabolic profiling of patients with fatal infantile mitochondrial myopathy with de Toni-Fanconi-Debré syndrome, *Pediatr. Int.* 38 (1996) 661–666. <https://doi.org/10.1111/j.1442-200X.1996.tb03727.x>.
- [61] H. Sauter, M. Lauer, H. Fritsch, Metabolic Profiling of Plants, in: *ChemInform*, 1991: pp. 288–299. <https://doi.org/10.1021/bk-1991-0443.ch024>.
- [62] E.L. Avery, R.H. Dunstan, J.A. Nell, The Use of Lipid Metabolic Profiling to Assess the Biological Impact of Marine Sewage Pollution, *Arch. Environ. Contam. Toxicol.* 35 (1998) 229–235. <https://doi.org/10.1007/s002449900371>.
- [63] O. Fiehn, J. Kopka, P. Dörmann, T. Altmann, R.N. Trethewey, L. Willmitzer, Metabolite profiling for plant functional genomics, *Nat. Biotechnol.* 18 (2000) 1157–1161. <https://doi.org/10.1038/81137>.
- [64] U. Roessner, A. Luedemann, D. Brust, O. Fiehn, T. Linke, L. Willmitzer, A.R. Fernie, Metabolic Profiling Allows Comprehensive Phenotyping of Genetically or Environmentally Modified Plant Systems, *Plant Cell*. 13 (2001) 11–29. <https://doi.org/10.1105/tpc.13.1.11>.
- [65] G.A. Cordell, Y.G. Shin, Finding the needle in the haystack. the dereplication of natural product extracts, *Pure Appl. Chem.* 71 (1999) 1089–1094. <https://doi.org/10.1351/pac199971061089>.
- [66] S.G. Villas-Bôas, Sampling and Sample Preparation, in: *Metabolome Anal.*, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2006: pp. 39–82. <https://doi.org/10.1002/9780470105511.ch3>.
- [67] K. Sasidharan, T. Soga, M. Tomita, D.B. Murray, A Yeast Metabolite Extraction Protocol Optimised for Time-Series Analyses, *PLoS One*. 7 (2012) 1–10. <https://doi.org/10.1371/journal.pone.0044283>.
- [68] A.B. Canelas, A. ten Pierick, C. Ras, R.M. Seifar, J.C. van Dam, W.M. van Gulik, J.J. Heijnen, Quantitative Evaluation of Intracellular Metabolite Extraction Techniques for Yeast Metabolomics, *Anal. Chem.* 81 (2009) 7379–7389. <https://doi.org/10.1021/ac900999t>.
- [69] M. Faijes, A.E. Mars, E.J. Smid, Comparison of quenching and extraction methodologies for metabolome analysis of *Lactobacillus plantarum*, *Microb. Cell Fact.* 6 (2007) 1–8. <https://doi.org/10.1186/1475-2859-6-27>.
- [70] B. Gonzalez, J. François, M. Renaud, A rapid and reliable method for metabolite extraction in yeast using boiling buffered ethanol, *Yeast*. 13 (1997) 1347–1356. [https://doi.org/10.1002/\(SICI\)1097-0061\(199711\)13:14<1347::AID-YEA176>3.0.CO;2-O](https://doi.org/10.1002/(SICI)1097-0061(199711)13:14<1347::AID-YEA176>3.0.CO;2-O).
- [71] N. Tambellini, V. Zaremborg, R. Turner, A. Weljie, Evaluation of Extraction Protocols for Simultaneous Polar and Non-Polar Yeast Metabolite Analysis Using Multivariate Projection Methods, *Metabolites*. 3 (2013) 592–605. <https://doi.org/10.3390/metabo3030592>.
- [72] F. Puig-Castellví, I. Alfonso, B. Piña, R. Tauler, A quantitative ¹H NMR approach for evaluating the metabolic response of *Saccharomyces cerevisiae* to mild heat stress, *Metabolomics*. 11 (2015) 1612–1625. <https://doi.org/10.1007/s11306-015-0812-9>.
- [73] B. Xie, Y. Wang, D.R. Jones, K.K. Dey, X. Wang, Y. Li, J.H. Cho, T.I. Shaw, H. Tan, J. Peng, Isotope Labeling-Assisted Evaluation of Hydrophilic and Hydrophobic Liquid Chromatograph-Mass Spectrometry for Metabolomics Profiling, *Anal. Chem.* 90 (2018) 8538–8545. <https://doi.org/10.1021/acs.analchem.8b01591>.
- [74] I. Kerzaon, Métabolites bioactifs d'Ascomycètes marins : dérégulation, isolement, identification et étude de production, Thèse Dr. Ec. Dr. VENAM. (2009).
- [75] E. Ortiz-Villanueva, F. Benavente, B. Piña, V. Sanz-Nebot, R. Tauler, J. Jaumot, Knowledge integration strategies for untargeted metabolomics based on MCR-ALS analysis of CE-MS and LC-MS data, *Anal. Chim. Acta.* 978 (2017) 10–

23. <https://doi.org/10.1016/j.aca.2017.04.049>.

- [76] A. Klitgaard, A. Iversen, M.R. Andersen, T.O. Larsen, J.C. Frisvad, K.F. Nielsen, Aggressive dereplication using UHPLC-DAD-QTOF: Screening extracts for up to 3000 fungal secondary metabolites, *Anal. Bioanal. Chem.* 406 (2014) 1933–1943. <https://doi.org/10.1007/s00216-013-7582-x>.
- [77] X. Duportet, R.B.M. Aggio, S. Carneiro, S.G. Villas-Bôas, The biological interpretation of metabolomic data can be misled by the extraction method used, *Metabolomics*. 8 (2012) 410–421. <https://doi.org/10.1007/s11306-011-0324-1>.
- [78] S.P. Cuellar-Bermudez, I. Aguilar-Hernandez, D.L. Cardenas-Chavez, N. Ornelas-Soto, M.A. Romero-Ogawa, R. Parra-Saldivar, Extraction and purification of high-value metabolites from microalgae: Essential lipids, astaxanthin and phycobiliproteins, *Microb. Biotechnol.* 8 (2015) 190–209. <https://doi.org/10.1111/1751-7915.12167>.
- [79] D. Kang, G.H. Son, H.M. Park, J. Kim, J.N. Choi, H.Y. Kim, S. Lee, S.B. Hong, C.H. Lee, Culture condition-dependent metabolite profiling of *Aspergillus fumigatus* with antifungal activity, *Fungal Biol.* 117 (2013) 211–219. <https://doi.org/10.1016/j.funbio.2013.01.009>.
- [80] S. Khoomrung, J.L. Martinez, S. Tippmann, S. Jansa-Ard, M.F. Buffing, R. Nicastro, J. Nielsen, Expanded metabolite coverage of *Saccharomyces cerevisiae* extract through improved chloroform/methanol extraction and tert-butylidimethylsilyl derivatization, *Anal. Chem. Res.* 6 (2015) 9–16. <https://doi.org/10.1016/j.ancr.2015.10.001>.
- [81] T. Pluskal, T. Nakamura, A. Villar-Briones, M. Yanagida, Metabolic profiling of the fission yeast *S. pombe*: Quantification of compounds under different temperatures and genetic perturbation, *Mol. Biosyst.* 6 (2009) 182–198. <https://doi.org/10.1039/b908784b>.
- [82] S. Kumala, K. DwiYuliani, P. Simanjuntak, Antimicrobial activity of secondary metabolites produced by endophytic fungi isolated from stems of Jati tree (*Tectonagrandis* L.F), *Int. J. Pharm. Sci. Res.* 6 (2015) 2349–2353. [https://doi.org/10.13040/IJPSR.0975-8232.6\(6\).2349-53](https://doi.org/10.13040/IJPSR.0975-8232.6(6).2349-53).
- [83] A.B. Canelas, J.J. Heijnen, R.M. Seifar, C. Ras, W.M. van Gulik, A. ten Pierick, J.C. van Dam, Quantitative Evaluation of Intracellular Metabolite Extraction Techniques for Yeast Metabolomics, *Anal. Chem.* 0 (2009) 7379–7389. <https://doi.org/10.1021/ac900999t>.
- [84] K.S. Howell, D. Cozzolino, E.J. Bartowsky, G.H. Fleet, P.A. Henschke, Metabolic profiling as a tool for revealing *Saccharomyces* interactions during wine fermentation, *FEMS Yeast Res.* 6 (2006) 91–101. <https://doi.org/10.1111/j.1567-1364.2005.00010.x>.
- [85] J.C. Touchstone, History of Chromatography, *J. Liq. Chromatogr.* 16 (1993) 1647–1665. <https://doi.org/10.1080/10826079308021679>.
- [86] L. Cai, Thin Layer Chromatography, *Curr. Protoc. Essent. Lab. Tech.* 8 (2014) 6.3.1-6.3.18. <https://doi.org/10.1002/9780470089941.et0603s08>.
- [87] K.D. Bartle, P. Myers, History of gas chromatography, *TrAC Trends Anal. Chem.* 21 (2002) 547–557. [https://doi.org/10.1016/S0165-9936\(02\)00806-3](https://doi.org/10.1016/S0165-9936(02)00806-3).
- [88] H. Kato, Y. Izumi, T. Hasunuma, F. Matsuda, A. Kondo, Widely targeted metabolic profiling analysis of yeast central metabolites, *J. Biosci. Bioeng.* 113 (2012) 665–673. <https://doi.org/10.1016/j.jbiosc.2011.12.013>.
- [89] C.L. Richter, A.D. Kennedy, L. Guo, N. Dokoozlian, Metabolomic Measurements at Three Time Points of a Chardonnay Wine Fermentation with *Saccharomyces cerevisiae*, *Am. J. Enol. Vitic.* 66 (2015) 294–301. <https://doi.org/10.5344/ajev.2015.14062>.
- [90] M.S.H. Akash, K. Rehman, Chapter 14: High Performance Liquid Chromatography, Springer Singapore, Singapore, 2020. <https://doi.org/10.1007/978-981-15-1547-7>.
- [91] G. Chawla, C. Ranjan, Principle, Instrumentation, and Applications of UPLC: A Novel Technique of Liquid Chromatography, *Open Chem. J.* 3 (2016) 1–16. <https://doi.org/10.2174/1874842201603010001>.
- [92] S.C. Moldoveanu, V. David, Basic Information about HPLC, in: *Essentials Mod. HPLC Sep.*, Elsevier, 2013. <https://doi.org/10.1016/B978-0-12-385013-3.00001-X>.
- [93] C. Wan, M. Zhang, Q. Fang, L. Xiong, X. Zhao, T. Hasunuma, F. Bai, A. Kondo, The impact of zinc sulfate addition on the dynamic metabolic profiling of *Saccharomyces cerevisiae* subjected to long term acetic acid stress treatment

and identification of key metabolites involved in the antioxidant effect of zinc, *Metallomics*. 7 (2015) 322–332. <https://doi.org/10.1039/C4MT00275J>.

- [94] J. Kim, K.H. Kim, Effects of minimal media vs. complex media on the metabolite profiles of *Escherichia coli* and *Saccharomyces cerevisiae*, *Process Biochem.* 57 (2017) 64–71. <https://doi.org/10.1016/j.procbio.2017.04.003>.
- [95] S. Nishino, N. Okahashi, F. Matsuda, H. Shimizu, Absolute quantitation of glycolytic intermediates reveals thermodynamic shifts in *Saccharomyces cerevisiae* strains lacking PFK1 or ZWF1 genes, *J. Biosci. Bioeng.* 120 (2015) 280–286. <https://doi.org/10.1016/j.jbiosc.2015.01.012>.
- [96] I. Mendes, I. Sanchez, R. Franco-Duarte, C. Camarasa, D. Schuller, S. Dequin, M.J. Sousa, Integrating transcriptomics and metabolomics for the analysis of the aroma profiles of *Saccharomyces cerevisiae* strains from diverse origins, *BMC Genomics*. 18 (2017) 455. <https://doi.org/10.1186/s12864-017-3816-1>.
- [97] R. Petry, M. Schmitt, J. Popp, Raman Spectroscopy-A Prospective Tool in the Life Sciences, *ChemPhysChem*. 4 (2003) 14–30. <https://doi.org/10.1002/cphc.200390004>.
- [98] M.L. Olson, J. Johnson, W.F. Carswell, L.H. Reyes, R.S. Senger, K.C. Kao, Characterization of an evolved carotenoids hyper-producer of *Saccharomyces cerevisiae* through bioreactor parameter optimization and Raman spectroscopy, *J. Ind. Microbiol. Biotechnol.* 43 (2016) 1355–1363. <https://doi.org/10.1007/s10295-016-1808-9>.
- [99] M.A. Mohamed, J. Jaafar, A.F. Ismail, M.H.D. Othman, M.A. Rahman, Fourier Transform Infrared (FTIR) Spectroscopy, in: *Membr. Charact.*, Elsevier, 2017: pp. 3–29. <https://doi.org/10.1016/B978-0-444-63776-5.00001-2>.
- [100] M.A. Mohamed, J. Jaafar, A.F. Ismail, M.H.D. Othman, M.A. Rahman, Fourier Transform Infrared (FTIR) Spectroscopy, in: *Membr. Charact.*, Elsevier, 2017: pp. 3–29. <https://doi.org/10.1016/B978-0-444-63776-5.00001-2>.
- [101] A. Kohler, U. Böcker, V. Shapaval, A. Forsmark, M. Andersson, J. Warringer, H. Martens, S.W. Omholt, A. Blomberg, High-Throughput Biochemical Fingerprinting of *Saccharomyces cerevisiae* by Fourier Transform Infrared Spectroscopy, *PLoS One*. 10 (2015) e0118052. <https://doi.org/10.1371/journal.pone.0118052>.
- [102] P.N.S. Sampaio, C.R.C. Calado, Comparative analysis of different transformed *Saccharomyces cerevisiae* strains based on high-throughput Fourier transform infrared spectroscopy, *J. Biotechnol.* 260 (2017) 1–10. <https://doi.org/10.1016/j.jbiotec.2017.08.020>.
- [103] D. Cozzolino, L. Flood, J. Bellon, M. Gishen, M. De Barros Lopes, Combining near infrared spectroscopy and multivariate analysis as a tool to differentiate different strains of *Saccharomyces cerevisiae*: a metabolomic study, *Yeast*. 23 (2006) 1089–1096. <https://doi.org/10.1002/yea.1418>.
- [104] S. Correa-García, M. Bermúdez-Moretti, A. Travo, G. Délérís, I. Forfar, FTIR spectroscopic metabolome analysis of lyophilized and fresh *Saccharomyces cerevisiae* yeast cells, *Anal. Methods*. 6 (2014) 1855. <https://doi.org/10.1039/c3ay42322k>.
- [105] H.H. Nieuwoudt, I.S. Pretorius, F.F. Bauer, D.G. Nel, B.A. Prior, Rapid screening of the fermentation profiles of wine yeasts by Fourier transform infrared spectroscopy, *J. Microbiol. Methods*. 67 (2006) 248–256. <https://doi.org/10.1016/j.mimet.2006.03.019>.
- [106] A. Marsico, R. Perniola, M. Cardone, M. Velenosi, D. Antonacci, V. Alba, T. Basile, Study of the Influence of Different Yeast Strains on Red Wine Fermentation with NIR Spectroscopy and Principal Component Analysis, *J.* 1 (2018) 133–147. <https://doi.org/10.3390/j1010013>.
- [107] B. Diehl, Principles in NMR Spectroscopy, in: *NMR Spectrosc. Pharm. Anal.*, Elsevier, 2008: pp. 1–41. <https://doi.org/10.1016/B978-0-444-53173-5.00001-9>.
- [108] J. Schripsema, Application of NMR in plant metabolomics: techniques, problems and prospects, *Phytochem. Anal.* 21 (2010) 14–21. <https://doi.org/10.1002/pca.1185>.
- [109] C.K. Larive, G.A. Barding, M.M. Dinges, NMR Spectroscopy for Metabolomics and Metabolic Profiling, *Anal. Chem.* 87 (2015) 133–146. <https://doi.org/10.1021/ac504075g>.
- [110] E.A. Mahrous, M.A. Farag, Two dimensional NMR spectroscopic approaches for exploring plant metabolome: A review, *J. Adv. Res.* 6 (2015) 3–15. <https://doi.org/10.1016/j.jare.2014.10.003>.
- [111] J.G. Bundy, B. Papp, R. Harmston, R.A. Browne, E.M. Clayson, N. Burton, R.J. Reece, S.G. Oliver, K.M. Brindle, Evaluation of predicted network modules in yeast metabolism using NMR-based metabolite profiling, *Genome Res.*

17 (2007) 510–519. <https://doi.org/10.1101/gr.5662207>.

- [112] S.S.W. Szeto, S.N. Reinke, B.D. Sykes, B.D. Lemire, Mutations in the *Saccharomyces cerevisiae* Succinate Dehydrogenase Result in Distinct Metabolic Phenotypes Revealed Through ¹H NMR-Based Metabolic Footprinting, *J. Proteome Res.* 9 (2010) 6729–6739. <https://doi.org/10.1021/pr100880y>.
- [113] M. Palomino-Schätzlein, M.M. Molina-Navarro, M. Tormos-Pérez, S. Rodríguez-Navarro, A. Pineda-Lucena, Optimised protocols for the metabolic profiling of *S. cerevisiae* by ¹H-NMR and HRMAS spectroscopy, *Anal. Bioanal. Chem.* 405 (2013) 8431–8441. <https://doi.org/10.1007/s00216-013-7271-9>.
- [114] F. Puig-Castellví, C. Bedia, I. Alfonso, B. Piña, R. Tauler, Deciphering the Underlying Metabolomic and Lipidomic Patterns Linked to Thermal Acclimation in *Saccharomyces cerevisiae*, *J. Proteome Res.* 17 (2018) 2034–2044. <https://doi.org/10.1021/acs.jproteome.7b00921>.
- [115] E. Cusano, L.R. Cagliani, R. Consonni, B. Simonato, G. Zapparoli, NMR-based metabolic profiling of different yeast fermented apple juices, *LWT.* 118 (2020) 108771. <https://doi.org/10.1016/j.lwt.2019.108771>.
- [116] W.Y. Kang, S.H. Kim, Y.K. Chae, Stress adaptation of *Saccharomyces cerevisiae* as monitored via metabolites using two-dimensional NMR spectroscopy, *FEMS Yeast Res.* 12 (2012) 608–616. <https://doi.org/10.1111/j.1567-1364.2012.00811.x>.
- [117] O. Perruchon, I. Schmitz-afonso, C. Afonso, A. Elomri, State-of-the-art in analytical methods for metabolic profiling of *Saccharomyces cerevisiae*, *Microchem. J.* 170 (2021) 106704. <https://doi.org/10.1016/j.microc.2021.106704>.
- [118] E. de Hoffmann, V. Stroobant, *Mass Spectrometry: Principles and Applications*, 3rd Edition, 2007. <https://www.wiley.com/en-us/Mass+Spectrometry%3A+Principles+and+Applications%2C+3rd+Edition-p-9780470033104>.
- [119] F. Xian, C.L. Hendrickson, A.G. Marshall, High Resolution Mass Spectrometry, *Anal. Chem.* 84 (2012) 708–719. <https://doi.org/10.1021/ac203191t>.
- [120] B.A. Mamyryn, Time-of-flight mass spectrometry (concepts, achievements, and prospects), *Int. J. Mass Spectrom.* 206 (2001) 251–266. [https://doi.org/10.1016/S1387-3806\(00\)00392-4](https://doi.org/10.1016/S1387-3806(00)00392-4).
- [121] A. Garcia, C. Barbas, Gas Chromatography-Mass Spectrometry (GC-MS)-Based Metabolomics, in: *Metab. Profiling*, 2011: pp. 191–204. https://doi.org/10.1007/978-1-61737-985-7_11.
- [122] H. Li, M.-L. Ma, S. Luo, R.-M. Zhang, P. Han, W. Hu, Metabolic responses to ethanol in *Saccharomyces cerevisiae* using a gas chromatography tandem mass spectrometry-based metabolomics approach, *Int. J. Biochem. Cell Biol.* 44 (2012) 1087–1096. <https://doi.org/10.1016/j.biocel.2012.03.017>.
- [123] S. Shao, T. Zhou, B.D. McGarvey, Comparative metabolomic analysis of *Saccharomyces cerevisiae* during the degradation of patulin using gas chromatography–mass spectrometry, *Appl. Microbiol. Biotechnol.* 94 (2012) 789–797. <https://doi.org/10.1007/s00253-011-3739-8>.
- [124] E. Grondin, A. Shum Cheong Sing, Y. Caro, M. Raheerimandimby, A.L. Randrianierenana, S. James, C. Nueno-Palop, J.M. François, T. Petit, A comparative study on the potential of epiphytic yeasts isolated from tropical fruits to produce flavoring compounds, *Int. J. Food Microbiol.* 203 (2015) 101–108. <https://doi.org/10.1016/j.ijfoodmicro.2015.02.032>.
- [125] M.J. Al-Jassani, G.J. Mohammed, I.H. Hameed, Secondary metabolites analysis of *Saccharomyces cerevisiae* and evaluation of antibacterial activity, *Int. J. Pharm. Clin. Res.* 8 (2016) 303–314.
- [126] A.B. Canelas, C. Ras, A. ten Pierick, J.C. van Dam, J.J. Heijnen, W.M. van Gulik, Leakage-free rapid quenching technique for yeast metabolomics, *Metabolomics.* 4 (2008) 226–239. <https://doi.org/10.1007/s11306-008-0116-4>.
- [127] J.-Y. Jung, T.-Y. Kim, C.-Y. Ng, M.-K. Oh, Characterization of GCY1 in *Saccharomyces cerevisiae* by metabolic profiling, *J. Appl. Microbiol.* 113 (2012) 1468–1478. <https://doi.org/10.1111/jam.12013>.
- [128] J.R.B. de Souza, K.C. Kupper, F. Augusto, In vivo investigation of the volatile metabolome of antiphytopathogenic yeast strains active against *Penicillium digitatum* using comprehensive two-dimensional gas chromatography and multivariate data analysis, *Microchem. J.* 141 (2018) 204–209. <https://doi.org/10.1016/j.microc.2018.05.036>.
- [129] K. Schneider, J. Krömer, C. Wittmann, I. Alves-Rodrigues, A. Meyerhans, J. Diez, E. Heinzle, Metabolite profiling studies in *Saccharomyces cerevisiae*: an assisting tool to prioritize host targets for antiviral drug screening, *Microb.*

Cell Fact. 8 (2009) 12. <https://doi.org/10.1186/1475-2859-8-12>.

- [130] Z. Alves, A. Melo, A.R. Figueiredo, M.A. Coimbra, A.C. Gomes, S.M. Rocha, Exploring the *Saccharomyces cerevisiae* Volatile Metabolome: Indigenous versus Commercial Strains, *PLoS One*. 10 (2015) e0143641. <https://doi.org/10.1371/journal.pone.0143641>.
- [131] M. Sadoudi, R. Tourdot-Maréchal, S. Rousseaux, D. Steyer, J.-J. Gallardo-Chacón, J. Ballester, S. Vichi, R. Guérin-Schneider, J. Caixach, H. Alexandre, Yeast–yeast interactions revealed by aromatic profile analysis of Sauvignon Blanc wine fermented by single or co-culture of non-*Saccharomyces* and *Saccharomyces* yeasts, *Food Microbiol.* 32 (2012) 243–253. <https://doi.org/10.1016/j.fm.2012.06.006>.
- [132] R.L. Binati, W.J.F. Lemos Junior, G. Luzzini, D. Slaghenaufi, M. Ugliano, S. Torriani, Contribution of non-*Saccharomyces* yeasts to wine volatile and sensory diversity: A study on *Lachancea thermotolerans*, *Metschnikowia* spp. and *Starmerella bacillaris* strains isolated in Italy, *Int. J. Food Microbiol.* 318 (2020) 108470. <https://doi.org/10.1016/j.ijfoodmicro.2019.108470>.
- [133] G.A. Theodoridis, H.G. Gika, E.J. Want, I.D. Wilson, Liquid chromatography–mass spectrometry based global metabolite profiling: A review, *Anal. Chim. Acta.* 711 (2012) 7–16. <https://doi.org/10.1016/j.aca.2011.09.042>.
- [134] B. Zhou, J.F. Xiao, L. Tuli, H.W. Ransom, LC-MS-based metabolomics, *Mol. Biosyst.* 8 (2012) 470–481. <https://doi.org/10.1039/C1MB05350G>.
- [135] S. Jenkins, S. Fischer, L. Chen, T. Sana, Global LC/MS Metabolomics Profiling of Calcium Stressed and Immunosuppressant Drug Treated *Saccharomyces cerevisiae*, *Metabolites.* 3 (2013) 1102–1117. <https://doi.org/10.3390/metabo3041102>.
- [136] M. Farrés, B. Piña, R. Tauler, LC-MS based metabolomics and chemometrics study of the toxic effects of copper on *Saccharomyces cerevisiae*, *Metallomics.* 8 (2016) 790–798. <https://doi.org/10.1039/C6MT00021E>.
- [137] A. Lafaye, J. Labarre, J.-C. Tabet, E. Ezan, C. Junot, Liquid Chromatography–Mass Spectrometry and ¹⁵N Metabolic Labeling for Quantitative Metabolic Profiling, *Anal. Chem.* 77 (2005) 2026–2033. <https://doi.org/10.1021/ac048657g>.
- [138] M.R. Mashego, L. Wu, J.C. Van Dam, C. Ras, J.L. Vinke, W.A. Van Winden, W.M. Van Gulik, J.J. Heijnen, MIRACLE: mass isotopomer ratio analysis of U-¹³C-labeled extracts. A new method for accurate quantification of changes in concentrations of intracellular metabolites, *Biotechnol. Bioeng.* 85 (2004) 620–628. <https://doi.org/10.1002/bit.10907>.
- [139] C. Wittmann, M. Hans, W.A. van Winden, C. Ras, J.J. Heijnen, Dynamics of intracellular metabolites of glycolysis and TCA cycle during cell-cycle-related oscillation in *Saccharomyces cerevisiae*, *Biotechnol. Bioeng.* 89 (2005) 839–847. <https://doi.org/10.1002/bit.20408>.
- [140] M. Farrés, B. Piña, R. Tauler, Chemometric evaluation of *Saccharomyces cerevisiae* metabolic profiles using LC–MS, *Metabolomics.* 11 (2015) 210–224. <https://doi.org/10.1007/s11306-014-0689-z>.
- [141] Y. Ito, T. Hirasawa, H. Shimizu, Metabolic engineering of *Saccharomyces cerevisiae* to improve succinic acid production based on metabolic profiling, *Biosci. Biotechnol. Biochem.* 78 (2014) 151–159. <https://doi.org/10.1080/09168451.2014.877816>.
- [142] R.H. Nugroho, K. Yoshikawa, H. Shimizu, Metabolomic analysis of acid stress response in *Saccharomyces cerevisiae*, *J. Biosci. Bioeng.* 120 (2015) 396–404. <https://doi.org/10.1016/j.jbiosc.2015.02.011>.
- [143] K. Hayakawa, F. Matsuda, H. Shimizu, Metabolome analysis of *Saccharomyces cerevisiae* and optimization of culture medium for S-adenosyl-l-methionine production, *AMB Express.* 6 (2016) 38. <https://doi.org/10.1186/s13568-016-0210-3>.
- [144] M.J. Mert, S.H. Rose, D.C. la Grange, T. Bamba, T. Hasunuma, A. Kondo, W.H. van Zyl, Quantitative metabolomics of a xylose-utilizing *Saccharomyces cerevisiae* strain expressing the *Bacteroides thetaiotaomicron* xylose isomerase on glucose and xylose, *J. Ind. Microbiol. Biotechnol.* 44 (2017) 1459–1470. <https://doi.org/10.1007/s10295-017-1969-1>.
- [145] E. Ortiz-Villanueva, F. Benavente, B. Piña, V. Sanz-Nebot, R. Tauler, J. Jaumot, Knowledge integration strategies for untargeted metabolomics based on MCR-ALS analysis of CE-MS and LC-MS data, *Anal. Chim. Acta.* 978 (2017) 10–23. <https://doi.org/10.1016/j.aca.2017.04.049>.

- [146] R. Ramautar, G.W. Somsen, G.J. de Jong, CE-MS for metabolomics: Developments and applications in the period 2014–2016, *Electrophoresis*. 38 (2017) 190–202. <https://doi.org/10.1002/elps.201600370>.
- [147] C. Roullier-Gall, M. Witting, R.D. Gougeon, P. Schmitt-Kopplin, High precision mass measurements for wine metabolomics, *Front. Chem.* 2 (2014) 102. <https://doi.org/10.3389/fchem.2014.00102>.
- [148] C. Roullier-Gall, J. Signoret, D. Hemmler, M.A. Witting, B. Kanawati, B. Schäfer, R.D. Gougeon, P. Schmitt-Kopplin, Usage of FT-ICR-MS Metabolomics for Characterizing the Chemical Signatures of Barrel-Aged Whisky, *Front. Chem.* 6 (2018) 29. <https://doi.org/10.3389/fchem.2018.00029>.
- [149] M. Gotthardt, B. Kanawati, F. Schmidt, S. Asam, R. Hammerl, O. Frank, T. Hofmann, P. Schmitt-Kopplin, M. Rychlik, Comprehensive Analysis of the *Alternaria Mycobolome* Using Mass Spectrometry Based Metabolomics, *Mol. Nutr. Food Res.* 64 (2020). <https://doi.org/10.1002/mnfr.201900558>.
- [150] J.I. Castrillo, A. Hayes, S. Mohammed, S.J. Gaskell, S.G. Oliver, An optimized protocol for metabolome analysis in yeast using direct infusion electrospray mass spectrometry, *Phytochemistry*. 62 (2003) 929–937. [https://doi.org/10.1016/S0031-9422\(02\)00713-6](https://doi.org/10.1016/S0031-9422(02)00713-6).
- [151] J. Højer-Pedersen, J. Smedsgaard, J. Nielsen, The yeast metabolome addressed by electrospray ionization mass spectrometry: Initiation of a mass spectral library and its applications for metabolic footprinting by direct infusion mass spectrometry, *Metabolomics*. 4 (2008) 393–405. <https://doi.org/10.1007/s11306-008-0132-4>.
- [152] A. Tejero Rioseras, D. Garcia Gomez, B.E. Ebert, L.M. Blank, A.J. Ibáñez, P.M.L. Sinues, Comprehensive Real-Time Analysis of the Yeast Volatilome, *Sci. Rep.* 7 (2017) 14236. <https://doi.org/10.1038/s41598-017-14554-y>.
- [153] D.W. van Krevelen, Graphical-statistical method for the study of structure and reaction processes of coal, *Fuel*. 29 (1950) 269–284.
- [154] S. Kim, R.W. Kramer, P.G. Hatcher, Graphical Method for Analysis of Ultrahigh-Resolution Broadband Mass Spectra of Natural Organic Matter, the Van Krevelen Diagram, *Anal. Chem.* 75 (2003) 5336–5344. <https://doi.org/10.1021/ac034415p>.
- [155] J. Watrous, P. Roach, T. Alexandrov, B.S. Heath, J.Y. Yang, R.D. Kersten, M. van der Voort, K. Pogliano, H. Gross, J.M. Raaijmakers, B.S. Moore, J. Laskin, N. Bandeira, P.C. Dorrestein, Mass spectral molecular networking of living microbial colonies, *Proc. Natl. Acad. Sci.* 109 (2012) E1743–E1752. <https://doi.org/10.1073/pnas.1203689109>.
- [156] K.S. Howell, D. Cozzolino, E.J. Bartowsky, G.H. Fleet, P.A. Henschke, Metabolic profiling as a tool for revealing *Saccharomyces* interactions during wine fermentation, *FEMS Yeast Res.* 6 (2006) 91–101. <https://doi.org/10.1111/j.1567-1364.2005.00010.x>.
- [157] M. Mashego, M. Jansen, J. Vinke, W. Vangulik, J. Heijnen, Changes in the metabolome of associated with evolution in aerobic glucose-limited chemostats, *FEMS Yeast Res.* 5 (2005) 419–430. <https://doi.org/10.1016/j.femsyr.2004.11.008>.
- [158] T. Kind, H. Tsugawa, T. Cajka, Y. Ma, Z. Lai, S.S. Mehta, G. Wohlgemuth, D.K. Barupal, M.R. Showalter, M. Arita, O. Fiehn, Identification of small molecules using accurate mass MS/MS search, *Mass Spectrom. Rev.* 37 (2018) 513–532. <https://doi.org/10.1002/mas.21535>.
- [159] J.L. Wolfender, J.M. Nuzillard, J.J.J. Van Der Hooft, J.H. Renault, S. Bertrand, Accelerating Metabolite Identification in Natural Product Research: Toward an Ideal Combination of Liquid Chromatography-High-Resolution Tandem Mass Spectrometry and NMR Profiling, in Silico Databases, and Chemometrics, *Anal. Chem.* 91 (2019) 704–742. <https://doi.org/10.1021/acs.analchem.8b05112>.
- [160] C.A. Smith, E.J. Want, G. O’Maille, R. Abagyan, G. Siuzdak, XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification, *Anal. Chem.* 78 (2006) 779–787. <https://doi.org/10.1021/ac051437y>.
- [161] R. Tautenhahn, G.J. Patti, D. Rinehart, G. Siuzdak, XCMS Online: A Web-Based Platform to Process Untargeted Metabolomic Data, *Anal. Chem.* 84 (2012) 5035–5039. <https://doi.org/10.1021/ac300698c>.
- [162] M. Katajamaa, J. Miettinen, M. Oresic, MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data, *Bioinformatics*. 22 (2006) 634–636. <https://doi.org/10.1093/bioinformatics/btk039>.
- [163] T. Pluskal, S. Castillo, A. Villar-Briones, M. Orešič, MZmine 2: Modular framework for processing, visualizing, and

analyzing mass spectrometry-based molecular profile data, *BMC Bioinformatics*. 11 (2010) 395. <https://doi.org/10.1186/1471-2105-11-395>.

- [164] T. Jewison, C. Knox, V. Neveu, Y. Djoumbou, A.C. Guo, J. Lee, P. Liu, R. Mandal, R. Krishnamurthy, I. Sinelnikov, M. Wilson, D.S. Wishart, YMDB: The yeast metabolome database, *Nucleic Acids Res.* 40 (2012) 815–820. <https://doi.org/10.1093/nar/gkr916>.
- [165] M. Ramirez-Gaona, A. Marcu, A. Pon, A.C. Guo, T. Sajed, N.A. Wishart, N. Karu, Y.D. Feunang, D. Arndt, D.S. Wishart, YMDB 2.0: A significantly expanded version of the yeast metabolome database, *Nucleic Acids Res.* 45 (2017) D440–D445. <https://doi.org/10.1093/nar/gkw1058>.
- [166] D.S. Wishart, D. Tzur, C. Knox, R. Eisner, A.C. Guo, N. Young, D. Cheng, K. Jewell, D. Arndt, S. Sawhney, C. Fung, L. Nikolai, M. Lewis, M.A. Coutouly, I. Forsythe, P. Tang, S. Shrivastava, K. Jeroncic, P. Stothard, G. Amegbey, D. Block, D.D. Hau, J. Wagner, J. Miniaci, M. Clements, M. Gebremedhin, N. Guo, Y. Zhang, G.E. Duggan, G.D. MacInnis, A.M. Weljie, R. Dowlatabadi, F. Bamforth, D. Clive, R. Greiner, L. Li, T. Marrie, B.D. Sykes, H.J. Vogel, L. Querengesser, HMDB: The human metabolome database, *Nucleic Acids Res.* 35 (2007) 521–526. <https://doi.org/10.1093/nar/gkl923>.
- [167] H. Horai, M. Arita, S. Kanaya, Y. Nihei, T. Ikeda, K. Suwa, Y. Ojima, K. Tanaka, S. Tanaka, K. Aoshima, Y. Oda, Y. Kakazu, M. Kusano, T. Tohge, F. Matsuda, Y. Sawada, M.Y. Hirai, H. Nakanishi, K. Ikeda, N. Akimoto, T. Maoka, H. Takahashi, T. Ara, N. Sakurai, H. Suzuki, D. Shibata, S. Neumann, T. Iida, K. Tanaka, K. Funatsu, F. Matsuura, T. Soga, R. Taguchi, K. Saito, T. Nishioka, MassBank: A public repository for sharing mass spectral data for life sciences, *J. Mass Spectrom.* 45 (2010) 703–714. <https://doi.org/10.1002/jms.1777>.
- [168] M. Wang, J.J. Carver, V. V. Phelan, L.M. Sanchez, N. Garg, Y. Peng, D.D. Nguyen, J. Watrous, C.A. Kaponov, T. Luzzatto-Knaan, C. Porto, A. Bouslimani, A. V. Melnik, M.J. Meehan, W.T. Liu, M. Crüsemann, P.D. Boudreau, E. Esquenazi, M. Sandoval-Calderón, R.D. Kersten, L.A. Pace, R.A. Quinn, K.R. Duncan, C.C. Hsu, D.J. Floros, R.G. Gavilan, K. Kleigrewe, T. Northen, R.J. Dutton, D. Parrot, E.E. Carlson, B. Aigle, C.F. Michelsen, L. Jelsbak, C. Sohlenkamp, P. Pevzner, A. Edlund, J. McLean, J. Piel, B.T. Murphy, L. Gerwick, C.C. Liaw, Y.L. Yang, H.U. Humpf, M. Maansson, R.A. Keyzers, A.C. Sims, A.R. Johnson, A.M. Sidebottom, B.E. Sedio, A. Klitgaard, C.B. Larson, C.A.P. Boya, D. Torres-Mendoza, D.J. Gonzalez, D.B. Silva, L.M. Marques, D.P. Demarque, E. Pociute, E.C. O'Neill, E. Briand, E.J.N. Helfrich, E.A. Granatosky, E. Glukhov, F. Ryffel, H. Houson, H. Mohimani, J.J. Kharbush, Y. Zeng, J.A. Vorholt, K.L. Kurita, P. Charusanti, K.L. McPhail, K.F. Nielsen, L. Vuong, M. Elfeki, M.F. Traxler, N. Engene, N. Koyama, O.B. Vining, R. Baric, R.R. Silva, S.J. Mascuch, S. Tomasi, S. Jenkins, V. Macherla, T. Hoffman, V. Agarwal, P.G. Williams, J. Dai, R. Neupane, J. Gurr, A.M.C. Rodríguez, A. Lamsa, C. Zhang, K. Dorrestein, B.M. Duggan, J. Almaliti, P.M. Allard, P. Phapale, L.F. Nothias, T. Alexandrov, M. Litaudon, J.L. Wolfender, J.E. Kyle, T.O. Metz, T. Peryea, D.T. Nguyen, D. VanLeer, P. Shinn, A. Jadhav, R. Müller, K.M. Waters, W. Shi, X. Liu, L. Zhang, R. Knight, P.R. Jensen, B. Palsson, K. Pogliano, R.G. Lington, M. Gutiérrez, N.P. Lopes, W.H. Gerwick, B.S. Moore, P.C. Dorrestein, N. Bandeira, Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking, *Nat. Biotechnol.* 34 (2016) 828–837. <https://doi.org/10.1038/nbt.3597>.
- [169] F. Olivon, N. Elie, G. Grelier, F. Roussi, M. Litaudon, D. Touboul, MetGem Software for the Generation of Molecular Networks Based on the t-SNE Algorithm, *Anal. Chem.* 90 (2018) 13900–13908. <https://doi.org/10.1021/acs.analchem.8b03099>.
- [170] R.A. Quinn, L.F. Nothias, O. Vining, M. Meehan, E. Esquenazi, P.C. Dorrestein, Molecular Networking As a Drug Discovery, Drug Metabolism, and Precision Medicine Strategy, *Trends Pharmacol. Sci.* 38 (2017) 143–154. <https://doi.org/10.1016/j.tips.2016.10.011>.
- [171] Q. Lyu, X. Wen, Y. Liu, C. Sun, K. Chen, C.-C. Hsu, X. Li, Comprehensive Profiling of Phenolic Compounds in White and Red Chinese Bayberries (*Morella rubra* Sieb. et Zucc.) and Their Developmental Variations Using Tandem Mass Spectral Molecular Networking, *J. Agric. Food Chem.* 69 (2021) 741–749. <https://doi.org/10.1021/acs.jafc.0c04117>.
- [172] T. Kind, O. Fiehn, Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry, *BMC Bioinformatics*. 8 (2007) 105. <https://doi.org/10.1186/1471-2105-8-105>.
- [173] T. Nagao, D. Yukihiro, Y. Fujimura, K. Saito, K. Takahashi, D. Miura, H. Wariishi, Power of isotopic fine structure for unambiguous determination of metabolite elemental compositions: In silico evaluation and metabolomic application, *Anal. Chim. Acta.* 813 (2014) 70–76. <https://doi.org/10.1016/j.aca.2014.01.032>.
- [174] E. Bae, I.J. Yeo, B. Jeong, Y. Shin, K.-H. Shin, S. Kim, Study of Double Bond Equivalents and the Numbers of Carbon and Oxygen Atom Distribution of Dissolved Organic Matter with Negative-Mode FT-ICR MS, *Anal. Chem.* 83 (2011)

4193–4199. <https://doi.org/10.1021/ac200464q>.

- [175] Y. Cho, J.E. Birdwell, M. Hur, J. Lee, B. Kim, S. Kim, Extension of the Analytical Window for Characterizing Aromatic Compounds in Oils Using a Comprehensive Suite of High-Resolution Mass Spectrometry Techniques and Double Bond Equivalence versus Carbon Number Plot, *Energy & Fuels*. 31 (2017) 7874–7883. <https://doi.org/10.1021/acs.energyfuels.7b00962>.
- [176] C.A. Hughey, C.L. Hendrickson, R.P. Rodgers, A.G. Marshall, K. Qian, Kendrick mass defect spectrum: A compact visual analysis for ultrahigh-resolution broadband mass spectra, *Anal. Chem.* 73 (2001) 4676–4681. <https://doi.org/10.1021/ac010560w>.
- [177] M.R. Lopes, A.R.O. Santos, J.D. Moreira, R. Santa-Brígida, M.B. Martins, F.O. Pinto, P. Valente, P.B. Morais, N. Jacques, C. Grondin, S. Casaregola, M.-A. Lachance, C.A. Rosa, *Kurtzmaniella hittingeri* f.a., sp. nov., isolated from rotting wood and fruits, and transfer of three *Candida* species to the genus *Kurtzmaniella* as new combinations, *Int. J. Syst. Evol. Microbiol.* 69 (2019) 1504–1508. <https://doi.org/10.1099/ijsem.0.003337>.
- [178] T.L.S. Amoikon, C. Grondin, T. N'Dédé Djéni, N. Jacques, S. Casaregola, *Starmerella reginensis* f.A., sp. Nov. and *Starmerella kourouensis* f.A., sp. Nov., isolated from flowers in French Guiana, *Int. J. Syst. Evol. Microbiol.* 68 (2018) 2299–2305. <https://doi.org/10.1099/ijsem.0.002829>.
- [179] F. Olivon, G. Grelier, F. Roussi, M. Litaudon, D. Touboul, MZmine 2 Data-Preprocessing to Enhance Molecular Networking Reliability, *Anal. Chem.* 89 (2017) 7836–7840. <https://doi.org/10.1021/acs.analchem.7b01563>.
- [180] O.D. Myers, S.J. Sumner, S. Li, S. Barnes, X. Du, One Step Forward for Reducing False Positive and False Negative Compound Identifications from Mass Spectrometry Metabolomics Data: New Algorithms for Constructing Extracted Ion Chromatograms and Detecting Chromatographic Peaks, *Anal. Chem.* 89 (2017) 8696–8703. <https://doi.org/10.1021/acs.analchem.7b00947>.
- [181] L.-F. Nothias, D. Petras, R. Schmid, K. Dührkop, J. Rainer, A. Sarvepalli, I. Protsyuk, M. Ernst, H. Tsugawa, M. Fleischauer, F. Aicheler, A.A. Aksenov, O. Alka, P.-M. Allard, A. Barsch, X. Cachet, A.M. Caraballo-Rodríguez, R.R. Da Silva, T. Dang, N. Garg, J.M. Gauglitz, A. Gurevich, G. Isaac, A.K. Jarmusch, Z. Kameník, K. Bin Kang, N. Kessler, I. Koester, A. Korf, A. Le Gouellec, M. Ludwig, C. Martin H., L.-I. McCall, J. McSayles, S.W. Meyer, H. Mohimani, M. Morsy, O. Moyne, S. Neumann, H. Neuweger, N.H. Nguyen, M. Nothias-Esposito, J. Paolini, V. V. Phelan, T. Pluskal, R.A. Quinn, S. Rogers, B. Shrestha, A. Tripathi, J.J.J. van der Hooft, F. Vargas, K.C. Weldon, M. Witting, H. Yang, Z. Zhang, F. Zubeil, O. Kohlbacher, S. Böcker, T. Alexandrov, N. Bandeira, M. Wang, P.C. Dorrestein, Feature-based molecular networking in the GNPS analysis environment, *Nat. Methods*. 17 (2020) 905–908. <https://doi.org/10.1038/s41592-020-0933-6>.
- [182] R. Chaleckis, I. Meister, P. Zhang, C.E. Wheelock, Challenges, progress and promises of metabolite annotation for LC–MS-based metabolomics, *Curr. Opin. Biotechnol.* 55 (2019) 44–50. <https://doi.org/10.1016/j.copbio.2018.07.010>.
- [183] L.W. Sumner, A. Amberg, D. Barrett, M.H. Beale, R. Beger, C.A. Daykin, T.W.M. Fan, O. Fiehn, R. Goodacre, J.L. Griffin, T. Hankemeier, N. Hardy, J. Harnly, R. Higashi, J. Kopka, A.N. Lane, J.C. Lindon, P. Marriott, A.W. Nicholls, M.D. Reilly, J.J. Thaden, M.R. Viant, Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI), *Metabolomics*. 3 (2007) 211–221. <https://doi.org/10.1007/s11306-007-0082-2>.
- [184] M. Ghaste, R. Mistrik, V. Shulaev, Applications of fourier transform ion cyclotron resonance (FT-ICR) and orbitrap based high resolution mass spectrometry in metabolomics and lipidomics, *Int. J. Mol. Sci.* 17 (2016) 816. <https://doi.org/10.3390/ijms17060816>.
- [185] O. Perruchon, I. Schmitz-Afonso, C. Grondin, S. Casaregola, C. Afonso, A. Elomri, Combination of UHPLC-MS/MS-molecular networking approach and FTICR-MS for the metabolic profiling of *Saccharomyces cerevisiae*, *J. Pharm. Biomed. Anal.* 195 (2021) 113857. <https://doi.org/10.1016/j.jpba.2020.113857>.
- [186] L. Cao, M. Guler, A. Tagirdzhanov, Y. Lee, A. Gurevich, H. Mohimani, MolDiscovery: Learning Mass Spectrometry Fragmentation of Small Molecules, *BioRxiv*. (2020). <https://doi.org/https://doi.org/10.1101/2020.11.28.401943>.
- [187] R.R. da Silva, M. Wang, L.F. Nothias, J.J.J. van der Hooft, A.M. Caraballo-Rodríguez, E. Fox, M.J. Balunas, J.L. Klassen, N.P. Lopes, P.C. Dorrestein, Propagating annotations of molecular networks using in silico fragmentation, *PLoS Comput. Biol.* 14 (2018) 1–26. <https://doi.org/10.1371/journal.pcbi.1006089>.

- [188] J.J.J. van der Hooft, J. Wandy, M.P. Barrett, K.E. V. Burgess, S. Rogers, Topic modeling for untargeted substructure exploration in metabolomics, *Proc. Natl. Acad. Sci.* 113 (2016) 13738–13743. <https://doi.org/10.1073/pnas.1608041113>.
- [189] O. Perruchon, I. Schmitz-Afonso, C. Grondin, J.-L. Legras, C. Afonso, A. Elomri, Metabolome Exploration by High-Resolution Mass Spectrometry Methodologies of Two New Yeast Species: *Starmerella reginensis* and *Starmerella kourouensis*, *J. Agric. Food Chem.* (2021). <https://doi.org/10.1021/acs.jafc.1c03201>.

Annexe 1

Supplementary Information: Combination of UHPLC-MS/MS-molecular networking approach and FTICR-MS for the metabolic profiling of *Saccharomyces cerevisiae*

Olivier Perruchon^a, Isabelle Schmitz-Afonso^a, Cécile Grondin^b, Serge Casaregola^b, Carlos Afonso^a, Abdelhakim Elomri^{a*}.

^a Normandie Univ, UNIROUEN, INSA Rouen, CNRS, COBRA (UMR 6014), Rouen 76000, France

^b Université Paris-Saclay, INRAE, AgroParisTech, Micalis Institute, CIRM-Levures, 78350, Jouy-en-Josas, France

*Corresponding author: hakim.elomri@univ-rouen.fr

Submission manuscript for “Journal of Pharmaceutical and Biomedical Analysis”

Table S1: pure chemical standards used for level 1 identification of six features.

Molecular formula	Proposed metabolite	CAS	Fournisseur	Correlated feature ID
C ₁₁ H ₁₃ NO ₃	N-Acetyl-L-phenylalanine	2018-61-3	TCI	37
C ₉ H ₁₀ O ₃	3-Phenyllactic acid	703-80-0	Aldrich	35
C ₁₀ H ₉ NO	3-Acetylindole	87-51-4	Acros Organics	29
C ₁₀ H ₇ NO ₃	4-Hydroxyquinoline-2-carboxylic acid	526-55-6	Sigma	21
C ₁₀ H ₉ NO ₂	Indole-3-acetic acid	20312-36-1	Alfa Aesar	45
C ₁₀ H ₁₁ NO	Tryptophol	91-62-3	Acros Organics	46

Table S2: ionization source parameters and data dependent scan acquisition parameters of UHPLC-ESI-QToF-MS/MS experiment.

Ionization parameter	Positive Ionization mode	Negative Ionization mode
Capillary voltage	3.0 kV	2.2 kV
Sampling cone voltage	25 V	25 V
Extraction cone voltage	5 V	3 V
Source temperature	120 °C	120 °C
Desolvation gas temperature	400 °C	400 °C
Desolvation gas flow	600 L/h	700 L/h
Cone gas flow	30 L/h	30 L/h
Lockspray capillary voltage	3.5 kV	2.2 kV

Acquisition parameter	Value
Intensity threshold	600
Fragmented ion for each scan	3
Number of MS/MS scans for each ion	4
Exclusion time after MS/MS scan	10 s
Low mass Collision energy ramp	10 V to 25 V
High mass Collision energy ramp	25 V to 50 V

Table S3: MZMine2 spectral data processing parameters for each module.

MZMine2 module	Parameters
Mass detection: Exact mass	Noise level 20 for MS level 1 and MS level 2
ADAP Chromatogram Builder	Min group size in number of scans 5; group intensity threshold 10^3 ; min highest intensity 10^3 ; m/z tolerance 0.006 m/z or 20 ppm
Chromatogram deconvolution: Wavelet (ADAP)	S/N Threshold 4; S/N estimator Intensity window SN; min feature height 1.000; coefficient/area threshold 20; Peak duration range 0.05 to 1.00; RT wavelet range 0.05 to 0.10
MS2 Scan pairing	m/z range (Da) 0.02; RT range (min) 0.5
Isotopic peak grouping	m/z tolerance 0.004 m/z or 20 ppm; RT tolerance 0.2 min; monotonic shape checked; maximum charge 2; Representative isotope most intense
Filtering: Duplicate peak filter	Filter mode new average; m/z tolerance 0.006 m/z or 20 ppm; RT tolerance 0.5 min
Alignment: Join Aligner	m/z tolerance 0.004 m/z or 5 ppm; Weight for m/z 2; RT tolerance 0.5 min; Weight for RT 1
Identification: Formula prediction	Corresponding ionization; m/z tolerance 0.001 m/z or 5 ppm; Max formulas 10; up to 100 carbons, 100 hydrogens, 30 oxygens, 8 nitrogens, 1 sulphur and 1 phosphorus; H/C ratio, NOPS/C ratios and Multiple element counts checked in Element count heuristics; RDBE range from - 1 to 40, it must be an integer; Isotope pattern filter with m/z tolerance 0.001 m/z or 5 ppm, minimum absolute intensity 10^3 and minimum score 60 %

Table S4: major ionization source parameters and analysis parameters of DI-FTICR-MS experiment.

Ionization Parameter	Positive Ionization mode	Negative Ionization mode
Source temperature	370 °C	
Nebulizer gas flow	1.5 bar	1 bar
Capillary voltage	4,500 V	4,100 V

Analysis Parameter	Positive Ionization mode	Negative Ionization mode
Transient length	1.12 s or 4 million points	
Mass range	m/z 50-1500	
External calibration	Sodium trifluoroacetate solution	
Internal calibration (Agilent Technologies tuning mix)	Ions: m/z 118.0863; 322.0481; 622.0290; 922.0098	Ions: m/z 112.9856; 431.9823; 601.9790; 1033.9881

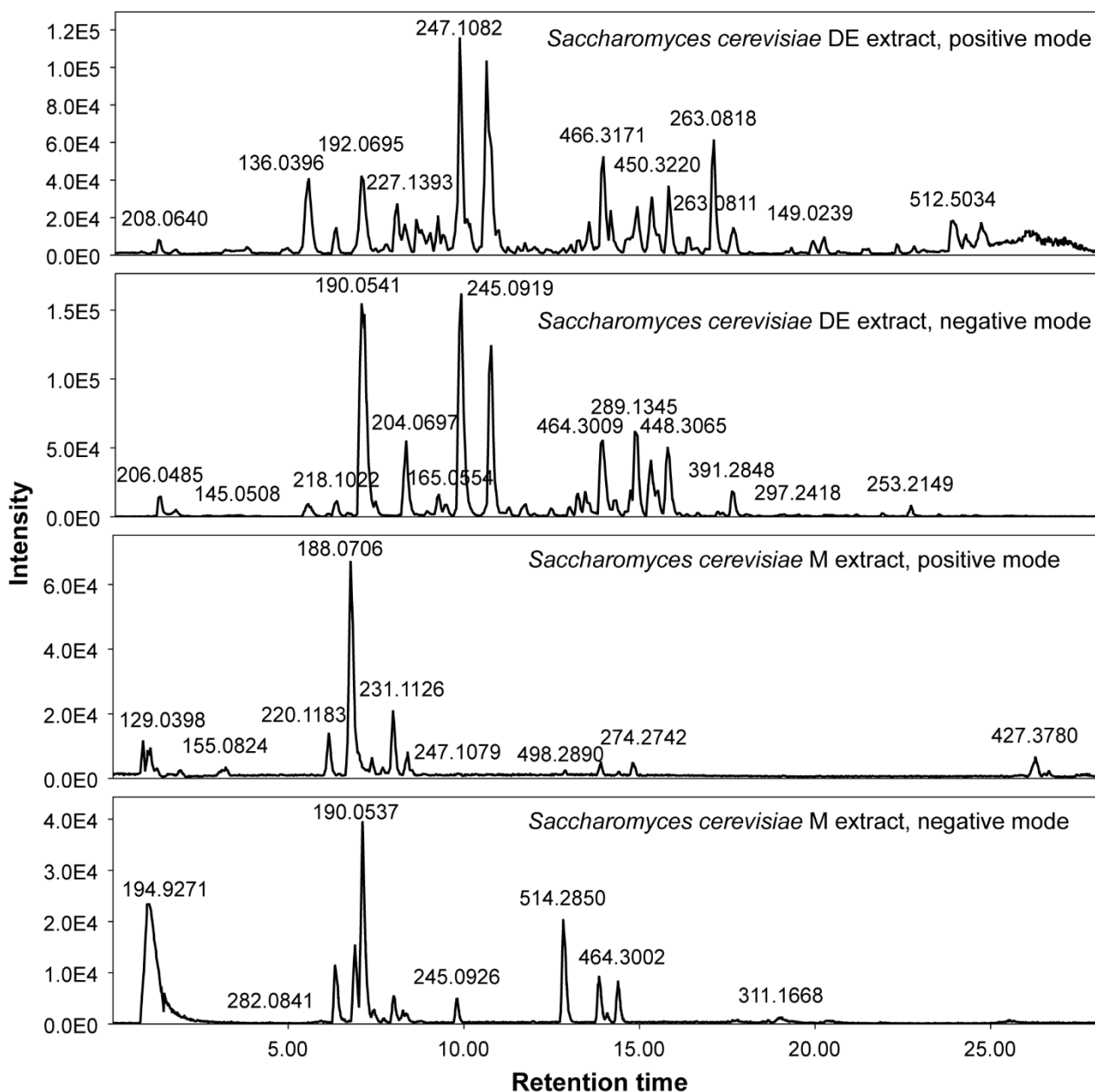
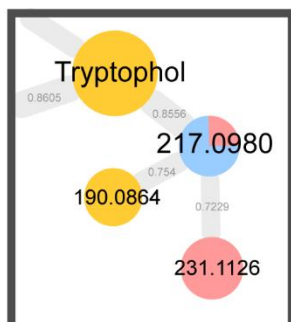
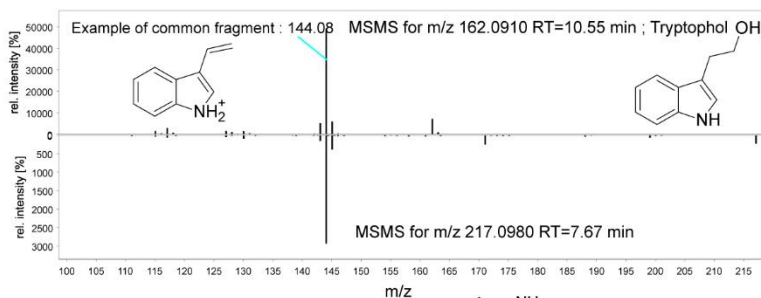


Figure S1: Base peak chromatograms for *Saccharomyces cerevisiae* DE extracts and M extracts after UHPLC-MS/MS analysis, in positive and in negative ionization mode. m/z ratio are displayed on top of chromatographic peaks. Retention time goes from 0.03 min to 28.00 min; mass range is m/z 50-2000.

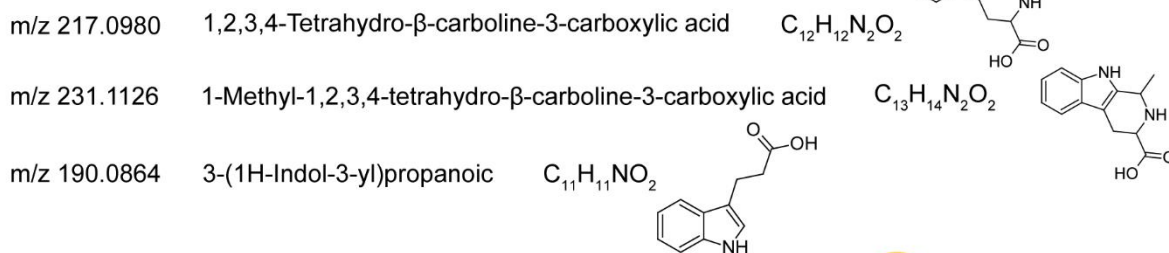
a Starting from molecular network :



Comparing fragmentation data in MS2 :



Hypotheses:



b

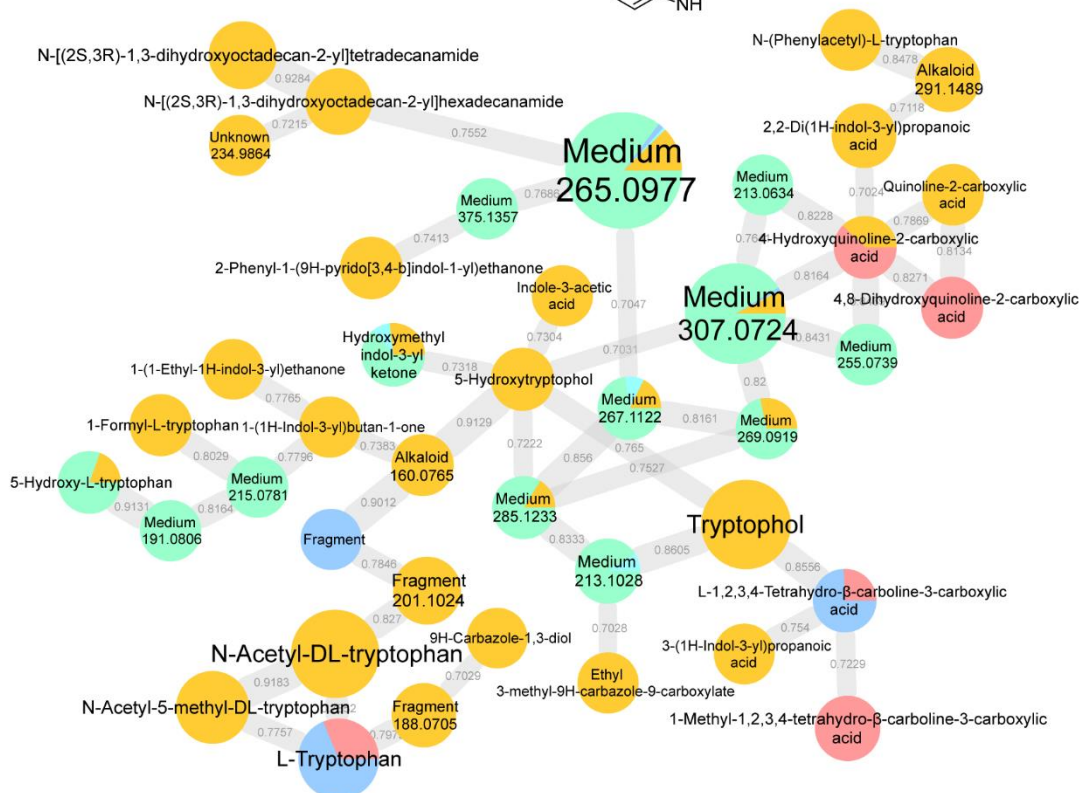
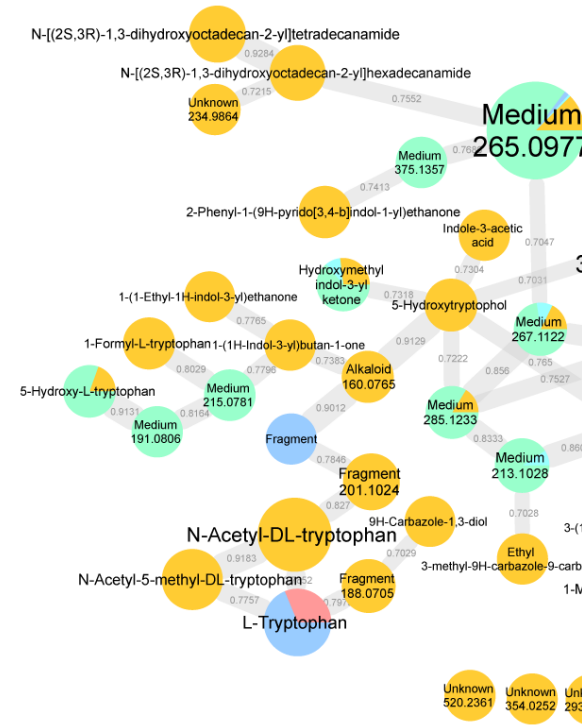


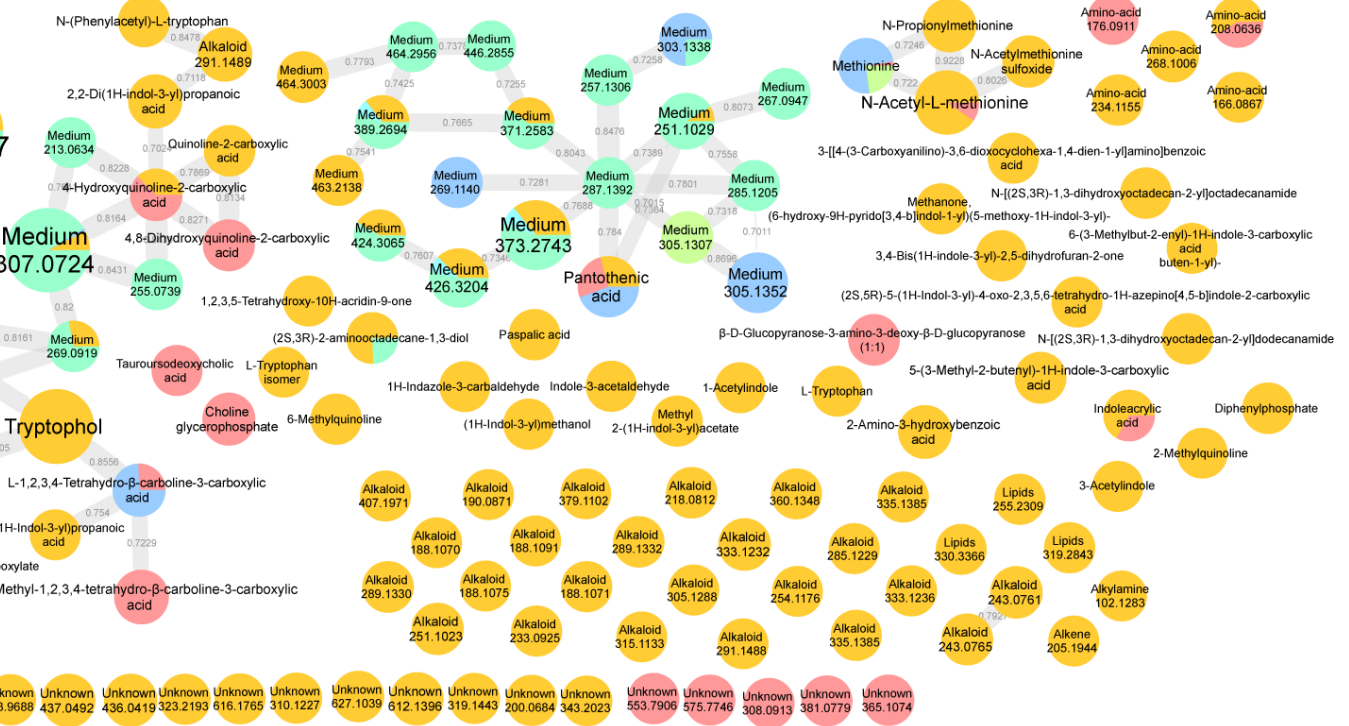
Figure S2: Investigation through molecular networks. a: Example of annotation propagation, the molecular network is visualized with Cytoscape. Based on the analysis in positive mode of DE and M extracts aligned together. Orange nodes: DE extract compounds. Red nodes: M extract compounds. Blue and green nodes: culture medium compounds. b: The alkaloid molecular network obtained after feature-based molecular networking GNPS workflow and visualized on Cytoscape. Process parameters are found in Material and methods section. Node labels display the proposed metabolite, or the

compound class with the m/z ratio. Node size is proportional to the precursor intensity, edge size to the cosine score between two nodes.

Positive mode



Cholic acid related medium compounds



Negative mode

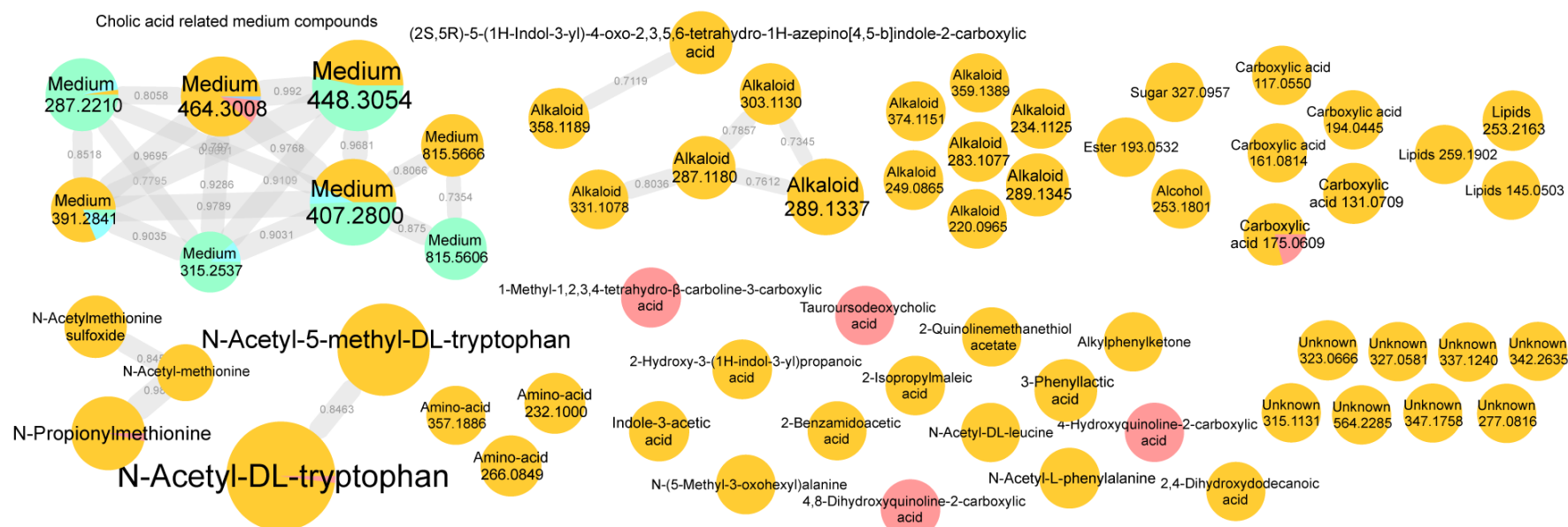
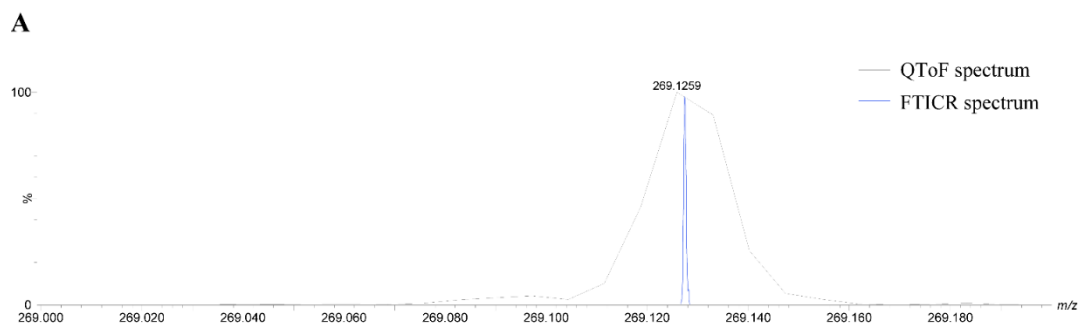


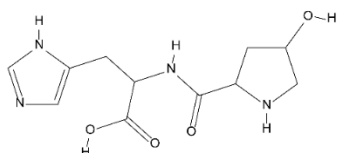
Figure S3: Molecular networks obtained after feature-based molecular networking GNPS workflow and visualized on Cytoscape. Process parameters are found in Material and methods section. Node labels display the proposed metabolite, or the compound class with the m/z ratio. Orange nodes: DE extract compounds. Red nodes: M extract compounds. Blue and green nodes: culture medium compounds. Culture medium compounds that are not in a network are not displayed. Node size is proportional to the precursor intensity, edge size to the cosine score between two nodes.



B

First formula proposal :
 $C_{16}H_{16}N_2O_2 / C_{21}H_{16} / C_{11}H_{14}N_4O_4$

First hypothesis :
 $C_{11}H_{14}N_4O_4$: Hydroxypropylhistidine



After FTICR analysis :
 Only $C_{16}H_{16}N_2O_2$

Hypothesis :
 Paspalic acid

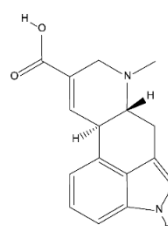


Figure S4: Comparison of the Paspalic acid (Retention time 12.76 min) between UHPLC-MS/MS QToF analysis and FTICR-MS analysis. a: overlaid spectra of m/z 269.000 to 269.200 range. Gray line is the QToF spectra and blue is the FTICR spectra. b: illustration of a hypothesis refinement made during the data interpretation.

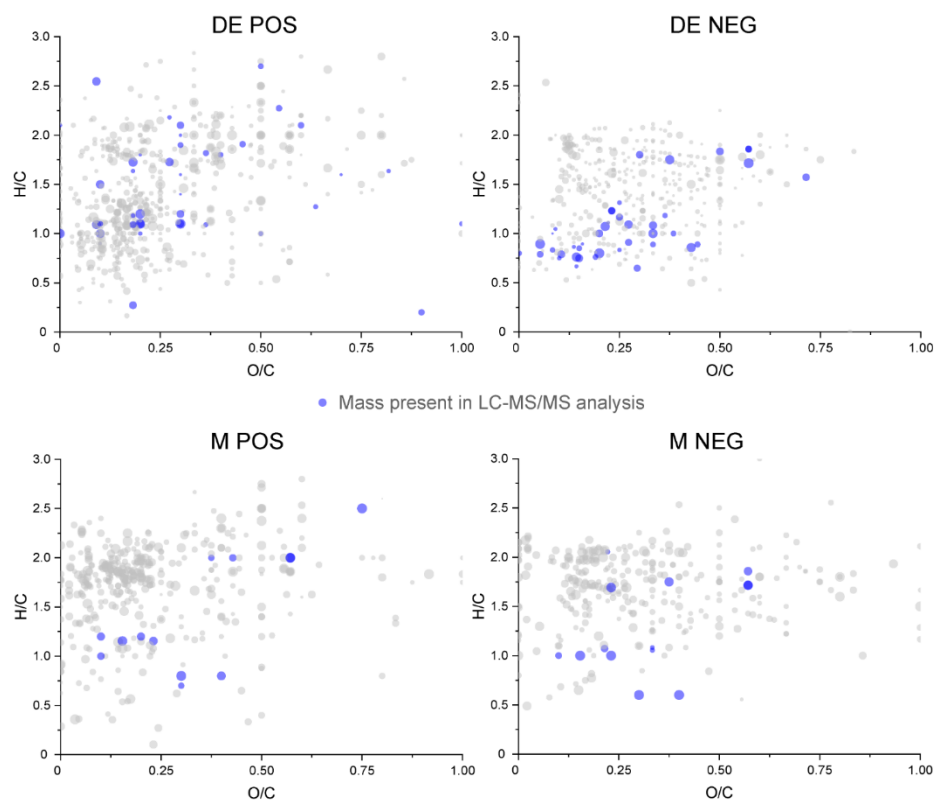


Figure S5: O/C vs. H/C Van Krevelen diagrams of extract FTICR-MS analyses in positive and negative ionization mode. Blue dots are masses detected both in UHPLC-MS/MS and in FTICR-MS.

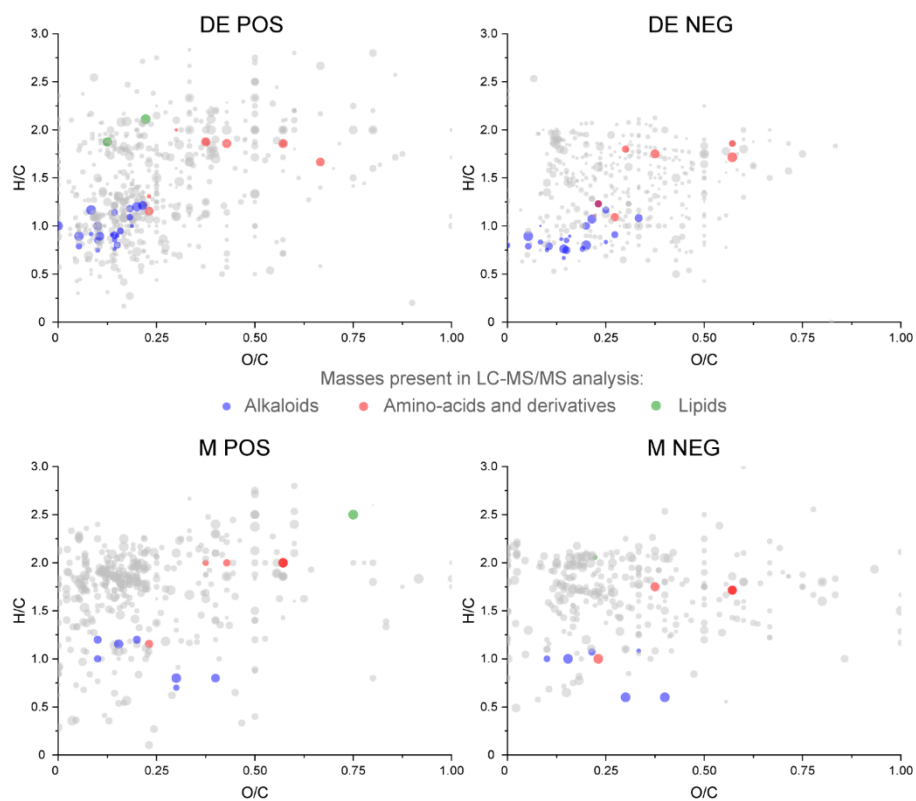


Figure S6: O/C vs. H/C Van Krevelen diagrams of extract FTICR-MS analyses in positive and negative ionization mode. The color corresponds to the molecular class of masses detected both in UHPLC-MS/MS and in FTICR-MS.

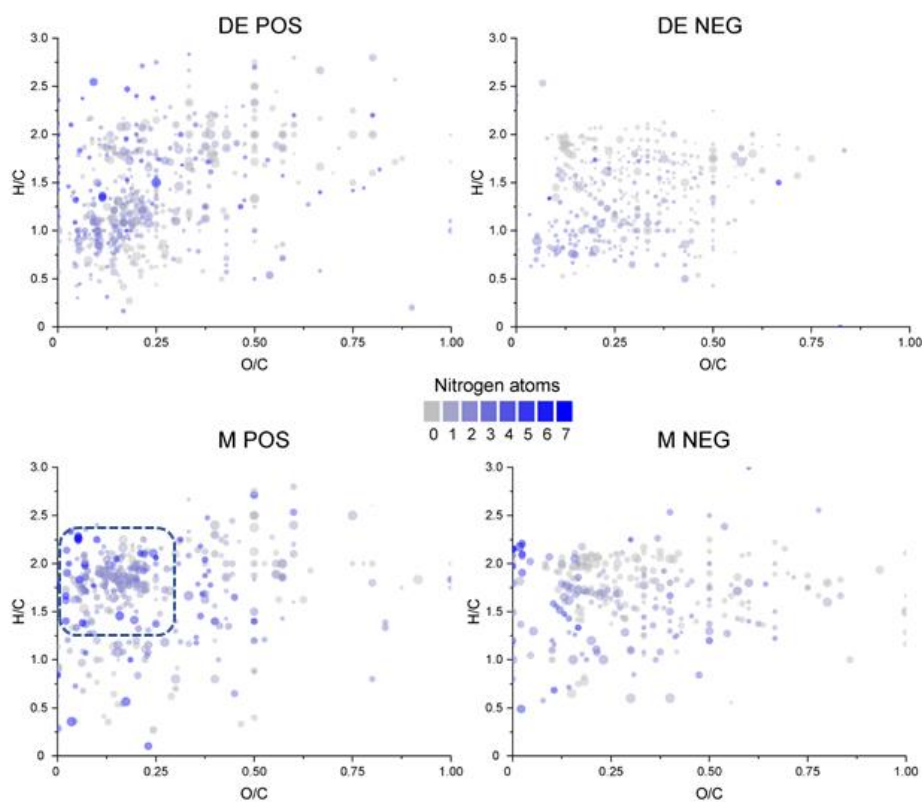


Figure S7: O/C vs. H/C Van Krevelen diagrams of extract FTICR-MS analyses in positive and negative ionization mode. The color corresponds to the number of nitrogen atoms in the compounds. The overlay between alkaloid region and lipid region is indicated.

Table S5: Proposed metabolites from *Saccharomyces cerevisiae* analysis. 128 compounds from DE extract and M extract analysed in positive and negative mode are present. Concerning features detected in both ionization modes, the observed m/z value comes from the positive and/or negative mode. Molecular formula of features without FTICR m/z error have been determined by UHPLC-MS/MS analysis.

Peak ID	RT (min)	Extract	Proposed metabolite	Type of ion	Observed m/z (QToF)	m/z Error (QToF - ppm)	Observed m/z (FTICR)	m/z Error (FTICR - ppm)	Molecular formula	Annotation level
1	1.07	M	Unknown	Positive	381.0779				Unknown	4
2	1.07	M	Choline glycerophosphate	M + H	258.1101	- 0.1	258.11009	0.06	C ₈ H ₂₀ NO ₆ P	2
3	1.10	DE	Alkylamine	M + H	102.1283	- 5.7	102.12773	- 0.08	C ₆ H ₁₅ N	3
4	1.13	M	Unknown	M + H	365.1074	- 5.6	365.10543	- 0.25	C ₁₈ H ₂₀ O ₆ S	4
5	1.13	M	β-D-Glucopyranose-3-amino-3-deoxy-β-D-glucopyranose (1:1)	M + H	360.1511	- 3.0	360.15002	0.05	C ₁₂ H ₂₅ NO ₁₁	2
6	1.13	DE	Amino acid	M + H	176.0911	3.6	176.09173	0.01	C ₇ H ₁₃ NO ₄	3
7	1.24	DE	N-Acetylmethionine sulfoxide	M + H M - H	208.0637 206.0485	0.4	208.06379	0.06	C ₇ H ₁₃ NO ₄ S	2
8	1.28	M	Unknown	Positive	308.0913				Unknown	4
9	1.75	DE; M	Amino-acid	M + H	208.0636	0.9	208.06379	- 0.1	C ₇ H ₁₃ NO ₄ S	3
10	3.45	DE	Lipids	M - H	145.0503	2.2	145.05061	0.18	C ₆ H ₁₀ O ₄	3
11	5.48	DE	Unknown	Negative	327.0581				Unknown	4
12	5.48	DE	2-Amino-3-hydroxybenzoic acid	M + H	154.0504	- 3.7	154.04986	0.08	C ₇ H ₇ NO ₃	2
13	5.58	DE	Indole-3-acetaldehyde	M + H	160.0760	- 1.9	160.07569	0	C ₁₀ H ₉ NO	2
14	6.07	DE	Carboxylic acid	M - H	117.0550	6.3	117.05571	0.09	C ₅ H ₁₀ O ₃	3
15	6.66	DE	Hydroxylated sulfated ester	M - H	193.0532	1.2	-	-	C ₇ H ₁₄ O ₄ S	3
16	6.70	DE	Carboxylic acid	M - H	161.0814	3.2	161.08192	0.1	C ₇ H ₁₄ O ₄	3
17	6.86	DE	Unknown	Positive	293.9688				Unknown	4
18	6.99	DE; M	N-Acetyl-L-methionine	M + H M - H	192.0693 190.0539	- 1.9	192.06887	0.1	C ₇ H ₁₃ NO ₃ S	2
19	7.04	M	4,8-Dihydroxyquinoline-2-carboxylic acid	M + H M - H	206.0446 204.0298	0.8	206.04477	0.05	C ₁₀ H ₇ NO ₄	2
20	7.37	DE	Carboxylic acid	M - H	194.0445	7.3	194.04587	0.05	C ₉ H ₉ NO ₄	3
21	7.40	DE; M	4-Hydroxyquinoline-2-carboxylic acid	M + H M - H	190.0503 188.0356	- 2.1	190.04987	- 0.02	C ₁₀ H ₇ NO ₃	1
22	7.44	DE; M	Carboxylic acid	M - H	175.0609	1.7	175.06118	0.12	C ₇ H ₁₂ O ₅	3
23	7.98	M	1-Methyl-1,2,3,4-tetrahydro-β-carboline-3-carboxylic acid	M + H M - H	231.1126 229.0974	0.8	231.1128	0	C ₁₃ H ₁₄ N ₂ O ₂	2
24	8.12	DE	2-Benzamidoacetic acid	M - H	178.0500	5.4	178.05095	0.09	C ₉ H ₉ NO ₃	2
25	8.21	DE	2-Isopropylmaleic acid	M - H	157.0502	2.9	157.05061	0.16	C ₇ H ₁₀ O ₄	2
26	8.24	DE	5-Hydroxytryptophol	M + H	178.0864	- 0.9	178.08625	0	C ₁₀ H ₁₁ NO ₂	2
27	8.24	DE; M	N-Propionylmethionine	M + H M - H	206.0848 204.0693	- 1.2	206.08454	0.02	C ₈ H ₁₅ NO ₃ S	2
28	8.40	DE	Carboxylic acid	M - H	131.0709	3.8	131.07134	0.18	C ₆ H ₁₂ O ₃	3
29	8.71	DE	3-Acetyldole	M + H	160.0765	- 5.3	160.07569	0	C ₁₀ H ₉ NO	1

30	8.84	DE	N-Acetyl-DL-leucine	M - H	172.0973	3.7	172.0979	0.11	C ₈ H ₁₅ NO ₃	2
31	8.86	DE	(1H-Indol-3-yl)methanol	M + H	148.0761	- 3.0	148.07569	0	C ₉ H ₉ NO	2
32	8.91	DE	Alkaloid	M + H	233.0925	- 1.8	233.09204	0.1	C ₁₂ H ₁₂ N ₂ O ₃	3
33	8.91	DE	Alkaloid	M + H M - H	251.1023 249.0865	1.4	251.10259	0.16	C ₁₂ H ₁₄ N ₂ O ₄	3
34	9.11	DE	Alkylphenylketone	M - H	209.0826	- 3.3	209.08192	0.06	C ₁₁ H ₁₄ O ₄	3
35	9.23	DE	3-Phenyllactic acid	M - H	165.0554	2.0	165.0557	0.09	C ₉ H ₁₀ O ₃	1
36	9.39	DE	Amino-acid	M + H	166.0867	- 2.9	166.08624	0.09	C ₉ H ₁₁ NO ₂	3
37	9.45	DE	N-Acetyl-L-phenylalanine	M - H	206.0808	7.1	206.08225	0.07	C ₁₁ H ₁₃ NO ₃	1
38	9.69	DE	1-Formyl-L-tryptophan	M + H	233.0922	- 0.5	233.09204	0.1	C ₁₂ H ₁₂ N ₂ O ₃	2
39	9.72	DE	2-Hydroxy-3-(1H-indol-3-yl)propanoic acid	M - H	204.0665	0.4	204.0666	0.1	C ₁₁ H ₁₁ NO ₃	2
40	9.81	DE	N-Acetyl-DL-tryptophan	M + H M - H	247.1080 245.0921	- 1.1	247.10769	0.1	C ₁₃ H ₁₄ N ₂ O ₃	2
41	9.86	DE	Unknown	M - H	277.0816	5.1	277.083	0	C ₁₃ H ₁₄ N ₂ O ₅	4
42	9.95	DE	1H-Indazole-3-carbaldehyde	M + H	147.0555	- 1.6	147.05528	0.07	C ₈ H ₆ N ₂ O	2
43	10.51	DE	Alkaloid	M - H	220.0965	6.3	220.0979	0.09	C ₁₂ H ₁₅ NO ₃	3
44	10.57	DE	Alkaloid	M + H	291.1488	1.3	291.14915	0.12	C ₁₉ H ₁₈ N ₂ O	3
45	10.57	DE	Indole-3-acetic acid	M + H M - H	176.0705 174.0557	0.5	176.07059	0.08	C ₁₀ H ₉ NO ₂	1
46	10.57	DE	Tryptophol	M + H	162.0920	- 3.9	162.09134	- 0.01	C ₁₀ H ₁₁ NO	1
47	10.57	DE	Unknown	M + H	319.1443	- 0.8	319.14411	- 0.03	C ₂₀ H ₁₈ N ₂ O ₂	4
48	10.63	DE	Alkaloid	M + H	335.1385	1.5	335.139	0.06	C ₂₀ H ₁₈ N ₂ O ₃	3
49	10.65	DE	Sugar	M - H	327.0957	1.8	327.09628	- 0.02	C ₁₀ H ₂₁ N ₂ O ₈ P	3
50	10.71	DE	N-Acetyl-5-methyl-DL-tryptophan	M + H M - H	261.1233 259.1078	0.1	261.12336	0.03	C ₁₄ H ₁₆ N ₂ O ₃	2
51	10.89	DE	Unknown	M - H	337.1240	- 3.8	337.12272	0.1	C ₁₆ H ₂₂ N ₂ O ₄ S	4
52	10.98	DE	Alkaloid	M + H	160.0764	- 4.3	160.07569	0	C ₁₀ H ₉ NO	3
53	11.12	DE	N-(5-Methyl-3-oxohexyl)alanine	M - H	200.1283	4.7	200.12922	0	C ₁₀ H ₁₉ NO ₃	2
54	11.18	DE	Amino-acid	M + H M - H	234.1155 232.1000	1.3	234.11586	- 0.1	C ₁₀ H ₁₉ NO ₃ S	3
55	11.19	DE	Alkaloid	M + H	188.1091	- 11.1	188.10699	- 0.01	C ₁₂ H ₁₃ NO	3
56	11.32	DE	3-(1H-Indol-3-yl)propanoic acid	M + H	190.0864	- 0.9	190.08626	- 0.01	C ₁₁ H ₁₁ NO ₂	2
57	11.57	DE	Unknown	M - H	323.0666	2.4	323.06733	0.06	C ₁₇ H ₁₂ N ₂ O ₅	4
58	11.63	DE	Alkaloid	M - H	234.1125	4.7	234.11356	0.01	C ₁₃ H ₁₇ NO ₃	3
59	11.66	DE	Amino-acid	M + H M - H	268.1006 266.0849	- 1.6	268.10022	- 0.11	C ₁₃ H ₁₇ NO ₃ S	3
60	11.74	DE	Alkaloid	M + H	379.1102	2.4	379.11109	0.01	C ₂₁ H ₁₈ N ₂ O ₃ S	3
61	11.80	DE	6-Methylquinoline	M + H	144.0807	0.4	144.08077	0.03	C ₁₀ H ₉ N	2
62	11.86	DE	Ethyl 3-methyl-9H-carbazole-9-carboxylate	M + H	254.1176	0.0	254.11752	0.13	C ₁₆ H ₁₅ NO ₂	2
63	12.21	DE	1,2,3,5-Tetrahydroxy-10H-acridin-9-one	M + H	260.0558	- 1.6	260.05529	0.22	C ₁₃ H ₉ NO ₅	2
64	12.27	DE	Unknown	Positive	234.9864				Unknown	4
65	12.41	DE	Alkaloid	M + H M - H	285.1229 283.1077	1.6	285.12333	0.13	C ₁₆ H ₁₆ N ₂ O ₃	3
66	12.48	DE	Alkaloid	M + H	254.1176	0.0	254.11752	0.13	C ₁₆ H ₁₅ NO ₂	3
67	12.56	DE	2,4-Dihydroxydodecanoic acid	M - H	231.1583	8.0	231.16018	0.01	C ₁₂ H ₂₄ O ₄	2
68	12.76	DE	6-(3-Methylbut-2-enyl)-1H-indole-3-carboxylic acid	M + H	230.1176	- 0.3	230.11755	0.03	C ₁₄ H ₁₅ NO ₂	2
69	12.76	DE	Paspalic acid	M + H	269.1290	- 1.9	269.12844	0.07	C ₁₆ H ₁₆ N ₂ O ₂	2

70	12.95	DE	Alkaloid	M - H	374.1151	- 1.3	374.11462	0.03	C ₂₁ H ₁₇ N ₃ O ₄	3
71	12.97	DE	Methanone, (6-hydroxy-9H-pyrido[3,4-b]indol-1-yl)(5-methoxy-1H-indol-3-yl)-	M + H	358.1188	- 0.5	358.11854	0.23	C ₂₁ H ₁₅ N ₃ O ₃	2
72	13.03	DE	Methyl 2-(1H-indol-3-yl)acetate	M + H	190.0868	- 2.8	190.08626	- 0.01	C ₁₁ H ₁₁ NO ₂	2
73	13.14	DE	Alkaloid	M + H	188.1075	- 2.7	188.10699	- 0.01	C ₁₂ H ₁₃ NO	3
74	13.22	DE	Alkaloid	M + H	243.0761	1.1	243.07642	0	C ₁₃ H ₁₀ N ₂ O ₃	3
75	13.22	DE	(2S,5R)-5-(1H-Indol-3-yl)-4-oxo-2,3,5,6-tetrahydro-1H-azepino[4,5-b]indole-2-carboxylic acid	M + H M - H	360.1339 358.1188	1.1	360.13421	0.16	C ₂₁ H ₁₇ N ₃ O ₃	2
76	13.36	DE	Alkaloid	M + H	218.0812	0.0	218.08114	0.12	C ₁₂ H ₁₁ NO ₃	3
77	13.36	DE	Alkaloid	M + H	335.1385	1.5	335.139	0.06	C ₂₀ H ₁₈ N ₂ O ₃	3
78	13.39	DE	Alkaloid	M - H	289.1345	0.6	289.13462	0.06	C ₁₉ H ₁₈ N ₂ O	3
79	13.42	DE	1-Acetylmethylethylamine	M + H	160.0767	- 6.2	160.07569	0	C ₁₀ H ₉ NO	2
80	13.42	DE	Quinoline-2-carboxylic acid	M + H	174.0557	- 4.1	174.05493	0.12	C ₁₀ H ₇ NO ₂	2
81	13.45	DE	Alkaloid	M - H	303.1130	2.8	303.11391	- 0.01	C ₁₉ H ₁₆ N ₂ O ₂	3
82	13.49	DE	2,2-Di(1H-indol-3-yl)propanoic acid	M + H	305.1292	- 2.5	305.12843	0.06	C ₁₉ H ₁₆ N ₂ O ₂	2
83	13.49	DE	Alkaloid	M + H	407.1971	- 1.3	407.19646	0.18	C ₂₄ H ₂₆ N ₂ O ₄	3
84	13.57	DE	5-(3-Methyl-2-butenyl)-1H-indole-3-carboxylic acid	M + H	230.1180	- 1.9	230.11755	0.03	C ₁₄ H ₁₅ NO ₂	2
85	13.63	DE	Alkaloid	M + H	333.1236	- 0.7	333.12336	0.02	C ₂₀ H ₁₆ N ₂ O ₃	3
86	13.69	DE	Alkaloid	M + H	289.1332	1.0	289.13351	0.12	C ₁₉ H ₁₆ N ₂ O	3
87	13.83	DE	Alkaloid	M + H	243.0765	- 0.3	243.07642	0	C ₁₃ H ₁₀ N ₂ O ₃	3
88	13.83	DE	Alkaloid	M + H M - H	360.1348 358.1189	- 1.4	360.13421	0.16	C ₂₁ H ₁₇ N ₃ O ₃	3
89	13.89	DE	Unknown	Negative	564.2285				Unknown	4
90	13.89	DE	Unknown	M + H	520.2361	- 1.7	-	-	C ₃₆ H ₂₉ N ₃ O	4
91	14.04	DE	3-(1H-Indol-3-yl)propanoic acid	M + H	190.0871	- 4.2	190.08626	- 0.01	C ₁₁ H ₁₁ NO ₂	2
92	14.31	DE	3,4-Bis(1H-indole-3-yl)-2,5-dihydrofuran-2-one	M + H	315.1129	- 0.5	315.11284	- 0.12	C ₂₀ H ₁₄ N ₂ O ₂	2
93	14.39	DE	Alkaloid	M + H	188.1071	- 0.4	188.10699	- 0.01	C ₁₂ H ₁₃ NO	3
94	14.41	M	Tauroursodeoxycholic acid	M + H M - H	500.3040 498.2892	- 0.3	-	-	C ₂₆ H ₄₅ NO ₆ S	2
95	14.43	DE	Unknown	M - H	347.1758	2.2	347.17649	0.04	C ₂₂ H ₂₄ N ₂ O ₂	4
96	14.51	DE	3-[[4-(3-Carboxyanilino)-3,6-dioxocyclohexa-1,4-dien-1-yl]amino]benzoic acid	M + H	379.0925	- 0.2	379.09238	0.22	C ₂₀ H ₁₄ N ₂ O ₆	2
97	14.68	DE	Alkaloid	M - H	287.1180	- 0.4	287.11788	0.03	C ₁₉ H ₁₆ N ₂ O	3
98	14.72	DE	2-Phenyl-1-(9H-pyrido[3,4-b]indol-1-yl)ethanone	M + H	287.1178	0.3	287.11788	0.03	C ₁₉ H ₁₄ N ₂ O	2
99	14.72	DE	Alkaloid	M + H M - H	333.1232 331.1078	0.4	333.12336	0.02	C ₂₀ H ₁₆ N ₂ O ₃	3
100	14.86	DE	Alkaloid	M + H M - H	291.1489 289.1337	1.0	291.14915	0.12	C ₁₉ H ₁₈ N ₂ O	3
101	14.86	DE	Alkaloid	M + H	305.1288	- 1.0	305.12843	0.06	C ₁₉ H ₁₆ N ₂ O ₂	3
102	14.86	DE	N-(Phenylacetyl)-L-tryptophan	M + H	323.1396	- 1.9	323.13897	0.15	C ₁₉ H ₁₈ N ₂ O ₃	2
103	14.86	DE	Unknown	Positive	310.1227				Unknown	4
104	14.88	DE	Lipids	M - H	259.1902	5.1	259.19149	- 0.02	C ₁₄ H ₂₈ O ₄	3
105	15.01	DE	1-(1H-Indol-3-yl)butan-1-one	M + H	188.1067	1.4	188.10699	- 0.01	C ₁₂ H ₁₃ NO	2
106	15.21	DE	Alkaloid	M + H	315.1133	- 1.6	315.11284	- 0.12	C ₂₀ H ₁₄ N ₂ O ₂	3
107	15.79	DE	2-Quinolinemethanethiol acetate	M - H	216.0481	3.6	216.04883	0.14	C ₁₂ H ₁₁ NOS	2
108	15.82	DE	Unknown	Positive	616.1765				Unknown	4
109	16.07	DE	Unknown	M - H	315.1131	2.4	315.11388	0.06	C ₂₀ H ₁₆ N ₂ O ₂	4

110	16.09	DE	9H-Carbazole-1,3-diol	M + H	200.0691	- 7.7	-	-	C ₁₂ H ₉ NO ₂	2
111	16.28	DE	Alkaloid	M - H	359.1389	3.3	359.14004	0.2	C ₂₂ H ₂₀ N ₂ O ₃	3
112	16.34	DE	2-Methylquinoline	M + H	144.0812	- 2.6	144.08077	0.03	C ₁₀ H ₉ N	2
113	16.49	DE	(2S,3R)-2-aminooctadecane-1,3-diol	M + H	302.3047	2.1	302.30532	0.11	C ₁₈ H ₃₉ NO ₂	2
114	16.57	DE	Alkaloid	M + H	289.1330	1.8	289.13351	0.12	C ₁₉ H ₁₆ N ₂ O	3
115	16.79	DE	Alkaloid	M + H	188.1070	- 0.3	188.10699	- 0.01	C ₁₂ H ₁₃ NO	3
116	17.17	DE	Polycyclic alcohol	M - H	253.1801	3.1	253.18092	0	C ₁₅ H ₂₆ O ₃	3
117	17.52	DE	1-(1-Ethyl-1H-indol-3-yl)-ethanone	M + H	188.1073	- 1.4	188.10699	- 0.01	C ₁₂ H ₁₃ NO	2
118	18.04	DE	Lipids	M + H	330.3366	0.2	330.33667	- 0.05	C ₂₀ H ₄₃ NO ₂	3
119	20.21	DE	Unknown	M - H	342.2635	4.2	342.26502	- 0.12	C ₁₉ H ₃₇ NO ₄	4
120	20.58	DE	Alkene	M + H	205.1944	3.1	205.19505	0.13	C ₁₅ H ₂₄	3
121	21.09	DE	Amino-acid	M - H	357.1886	1.7	357.18918	0.03	C ₁₄ H ₂₆ N ₆ O ₅	3
122	21.27	DE	Unknown	Positive	343.2023				Unknown	4
123	21.37	DE	Lipids	M + H	255.2309	3.7	255.23186	0	C ₁₆ H ₃₀ O ₂	3
				M - H	253.2163					
124	21.44	DE	Lipids	M + H	319.2843	0.0	319.28426	0.09	C ₁₈ H ₃₈ O ₄	3
125	23.00	DE	N-[(2S,3R)-1,3-dihydroxyoctadecan-2-yl]dodecanamide	M + H	484.4716	1.7	484.47234	0.17	C ₃₀ H ₆₁ NO ₃	2
126	23.85	DE	N-[(2S,3R)-1,3-dihydroxyoctadecan-2-yl]tetradecanamide	M + H	512.5035	0.4	512.5036	0.23	C ₃₂ H ₆₅ NO ₃	2
127	24.63	DE	N-[(2S,3R)-1,3-dihydroxyoctadecan-2-yl]hexadecanamide	M + H	540.5356	0.5	-	-	C ₃₄ H ₆₉ NO ₃	2
128	25.51	DE	N-[(2S,3R)-1,3-dihydroxyoctadecan-2-yl]octadecanamide	M + H	568.5648	- 2.5	-	-	C ₃₆ H ₇₃ NO ₃	2

Annexe 2

Supplementary Information: Metabolome exploration by high-resolution mass spectrometry methodologies of two new yeast species: *Starmerella reginensis* and *Starmerella kourouensis*

Olivier Perruchon^a, Isabelle Schmitz-Afonso^a, Cécile Grondin^{bc}, Jean-Luc Legras^c, Carlos Afonso^a, Abdelhakim Elomri^{a*}.

^aNormandie Univ, UNIROUEN, INSA Rouen, CNRS, COBRA (UMR 6014), Rouen 76000, France

^bMicalis Institute, INRAE, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas 78352, France

^cSPO, Univ Montpellier, INRAE, Institut Agro, Montpellier 34000, France

*Corresponding author: hakim.elomri@univ-rouen.fr

Submission manuscript for “Journal of Agricultural and Food Chemistry”

Table S1: major ionization source parameters and analysis parameters of DI-FTICR-MS experiment.

Ionization Parameter	Positive Ionization mode	Negative Ionization mode
Source temperature	370 °C	
Nebulizer gas flow	1.5 bar	1 bar
Capillary voltage	4,500 V	4,100 V

Analysis Parameter	Positive Ionization mode	Negative Ionization mode
Transient length	1.12 s or 4 million points	
Mass range	<i>m/z</i> 50-1500	
External calibration	Sodium trifluoroacetate solution	
Internal calibration (Agilent Technologies tuning mix)	Ions: <i>m/z</i> 118.0863; 322.0481; 622.0290; 922.0098	Ions: <i>m/z</i> 112.9856; 431.9823; 601.9790; 1033.9881

Table S2: ionization source parameters and data dependent scan acquisition parameters of UHPLC-ESI-QTOF-MS/MS experiment.

Ionization parameter	Positive Ionization mode	Negative Ionization mode
Capillary voltage	3.0 kV	2.2 kV
Sampling cone voltage	25 V	25 V
Extraction cone voltage	5 V	3 V
Source temperature	120 °C	120 °C
Desolvation gas temperature	400 °C	400 °C
Desolvation gas flow	600 L/h	700 L/h
Cone gas flow	30 L/h	30 L/h
Lockspray capillary voltage	3.5 kV	2.2 kV

Acquisition parameter	Value
Intensity threshold	600
Fragmented ion for each scan	3
Number of MS/MS scans for each ion	4
Exclusion time after MS/MS scan	10 s
Low mass Collision energy ramp	10 V to 25 V
High mass Collision energy ramp	25 V to 50 V

Table S3: MZMine2 spectral data processing parameters for each module.

MZMine2 module	Parameters
Mass detection: Exact mass	Noise level 20 for MS level 1 and MS level 2
ADAP Chromatogram Builder	Min group size in number of scans 3; group intensity threshold 1000 min highest intensity 1000 m/z tolerance 0.006 m/z or 20 ppm
Chromatogram deconvolution: Wavelet (ADAP)	S/N Threshold 4; S/N estimator Intensity window SN; min feature height 700; coefficient/area threshold 20; Peak duration range 0.05 to 2.00; RT wavelet range 0.04 to 0.10
MS2 Scan pairing	m/z range (Da) 0.02; RT range (min) 0.5
Isotopic peak grouper	m/z tolerance 0.004 m/z or 20 ppm; RT tolerance 0.2 min; monotonic shape checked; maximum charge 2; Representative isotope most intense
Filtering: Duplicate peak filter	Filter mode new average; m/z tolerance 0.006 m/z or 20 ppm; RT tolerance 0.2 min
Alignment: Join Aligner	m/z tolerance 0.004 m/z or 10 ppm; Weight for m/z 2; RT tolerance 0.2 min; Weight for RT 1
Identification: Formula prediction	Corresponding ionization; m/z tolerance 0.001 m/z or 5 ppm; Max formulas 10; up to 100 carbons, 200 hydrogens, 20 oxygens, 8 nitrogens, 1 sulphur and 1 phosphorus; H/C ratio, NOPS/C ratios and Multiple element counts checked in Element count heuristics; RDBE range from - 1 to 40, it must be an integer; Isotope pattern filter with m/z tolerance 0.001 m/z or 5 ppm, minimum absolute intensity 1000 and minimum score 60 %
Gap-filling	Same m/z and RT range; m/z tolerance = 0.004 m/z or 10 ppm.
Fragment search	RT tolerance = 0.1 min; m/z tolerance 0.004 m/z or 10 ppm; max fragment peak height = 100 %; Min MS2 peak height = 10.
Adduct search	RT tolerance = 0.1 min; m/z tolerance 0.004 m/z or 10 ppm; max relative adduct peak height = 50 %; adducts = [M+Na] and [M+K] for positive, [M+H ₃ PO ₄], [M+H ₂ SO ₄] and [M+H ₂ CO ₃] for negative.
Complex search	RT tolerance = 0.1 min; m/z tolerance 0.001 m/z or 5 ppm; Max complex peak height 50 %.

Table S4: Results of FTICR-MS analysis of *S. reginensis* of *S. kourouensis* extracts in number of attributed formulae (discarding culture medium features, automated SmartFormula process and manual discarding of outliers).

Sample	Analysis		Internal calibration standard deviation (in ppm)	Number of attributed formulae
		Ionization polarity		
<i>Starmerella reginensis</i> DCM/EA extract		Positive	0.141	717
		Negative	0.068	106
<i>Starmerella reginensis</i> M extract		Positive	0.045	1042
		Negative	0.052	194
<i>Starmerella kourouensis</i> DCM/EA extract		Positive	0.140	524
		Negative	0.065	166
<i>Starmerella kourouensis</i> M extract		Positive	0.043	572
		Negative	0.057	167

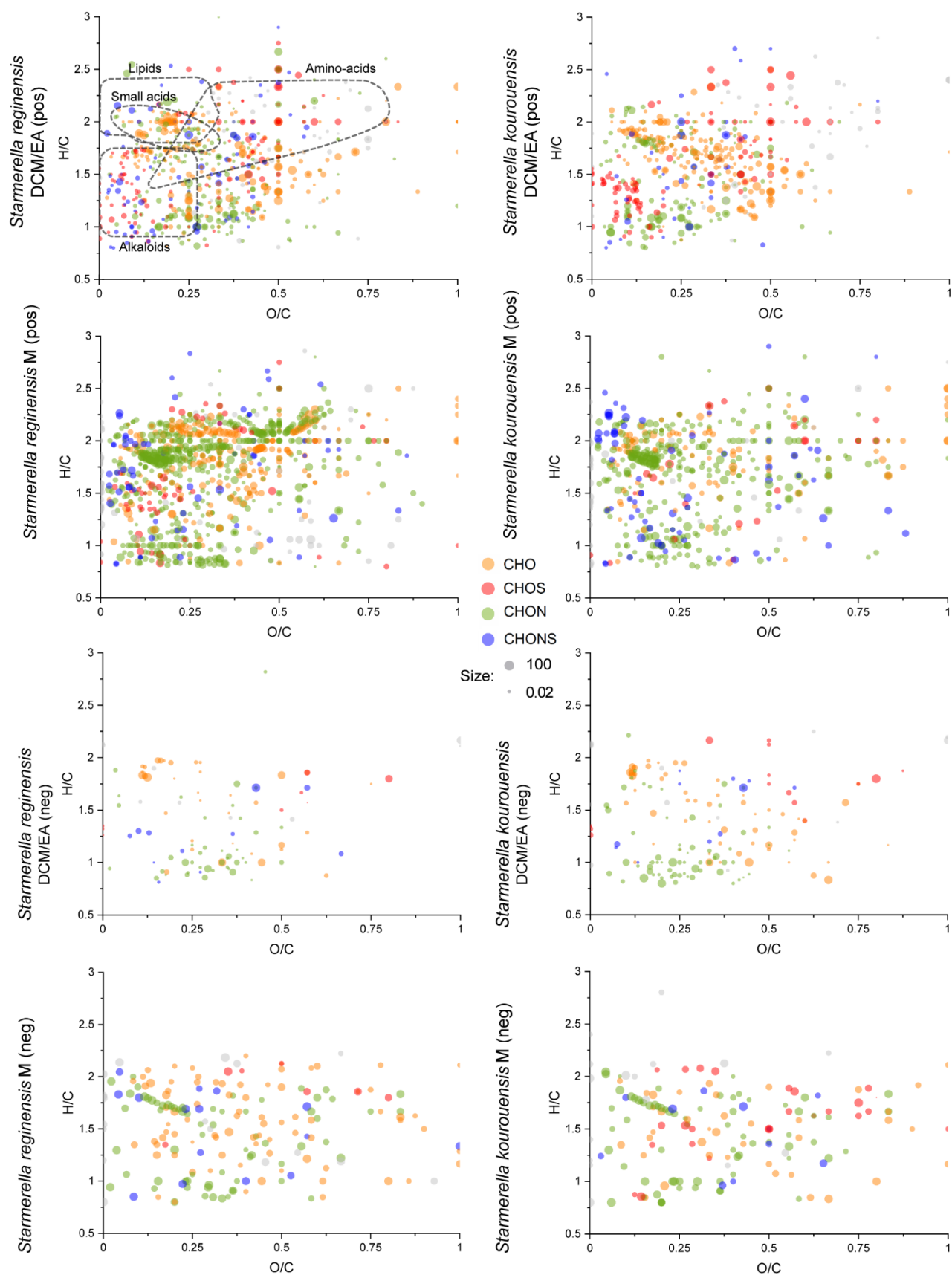


Figure S1: O/C vs. H/C Van Krevelen diagrams of FTICR-MS analyses of *S. reginensis* and *S. kourouensis* in positive and negative ionization mode. Brown dots are CHO features; red dots are CHOS; green dots are CHON; blue dots are CHONS.

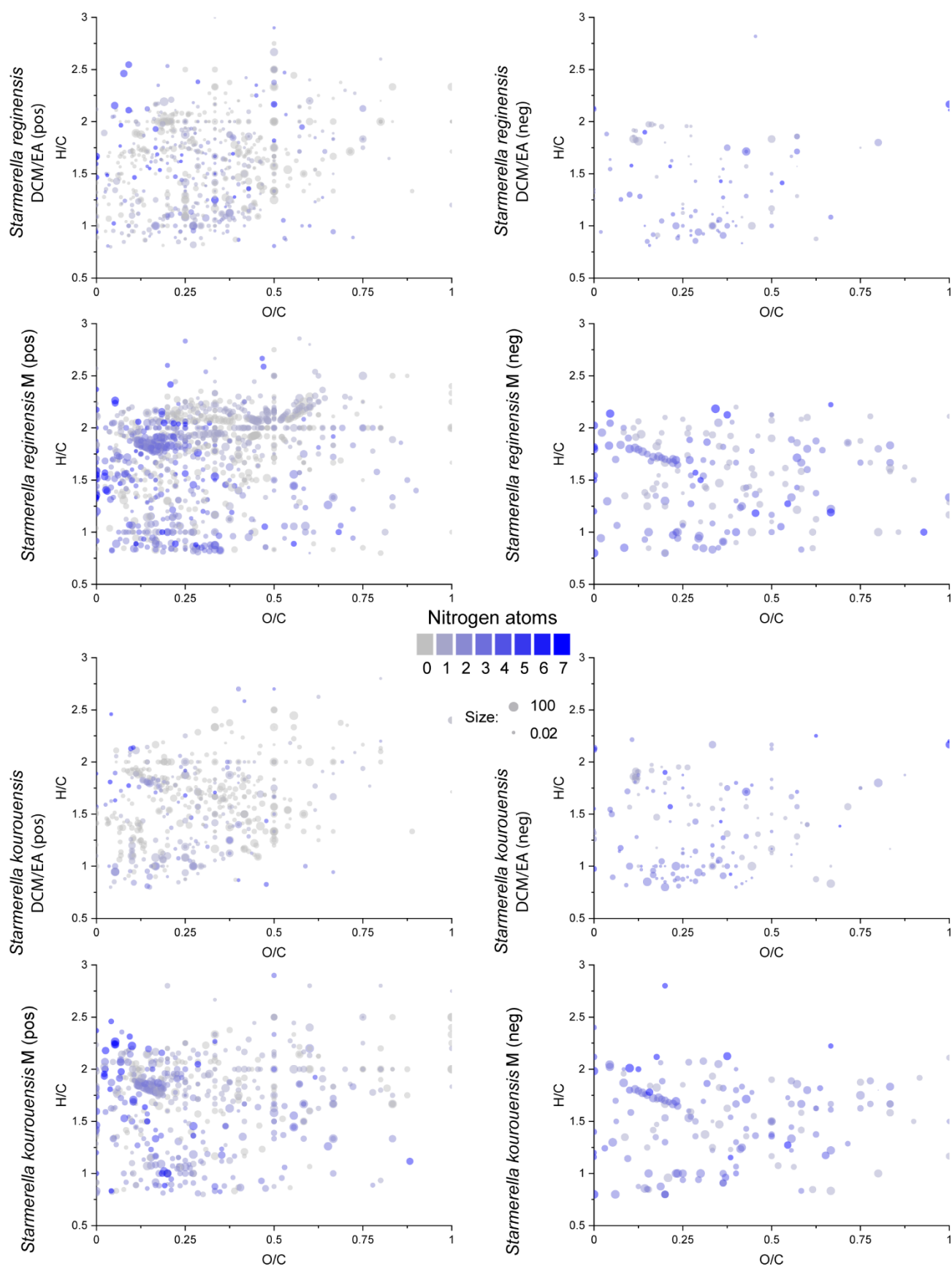


Figure S2: O/C vs. H/C Van Krevelen diagrams of FTICR-MS analyses of *S. reginensis* and *S. kourouensis* in positive and negative ionization mode. The color corresponds to the number of nitrogen atoms in the compounds.

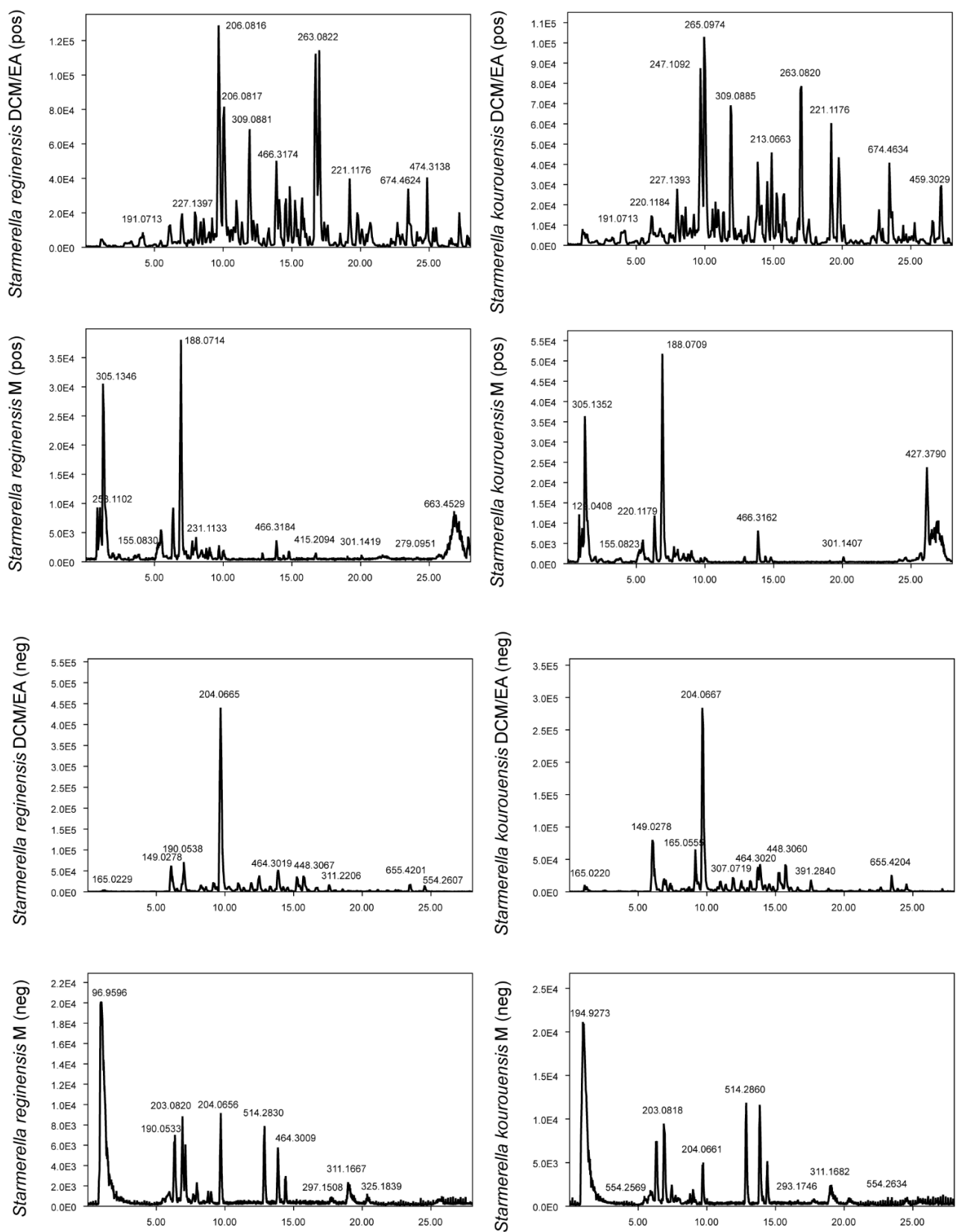
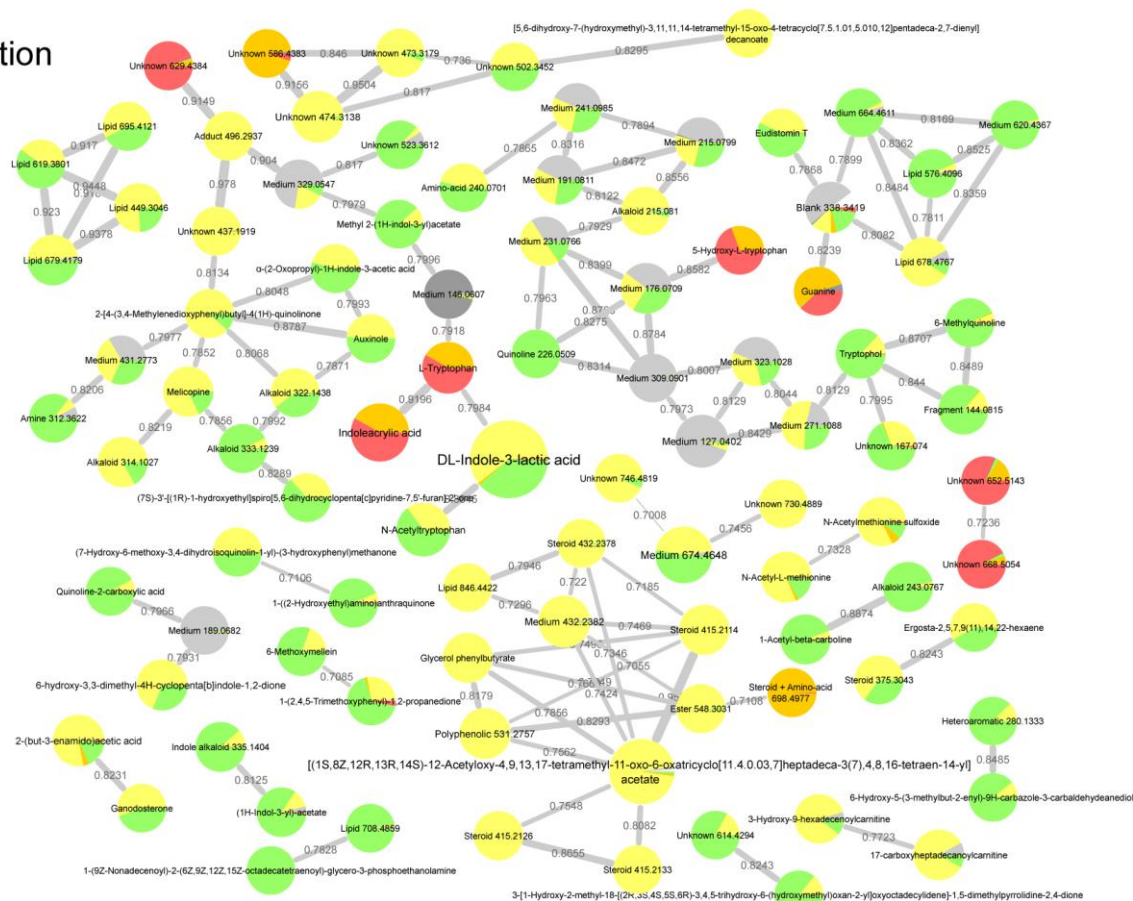


Figure S3: Base peak chromatograms for *Stammerella reginensis* and *Stammerella kourouensis* DCM/EA extracts and M extracts after UHPLC-MS/MS analysis, in positive and in negative ionization mode. m/z ratio are displayed on top of chromatographic peaks. Retention time goes from 0.03 min to 28.00 min; mass range is m/z 50-2000.

Positive ionization



Negative ionization

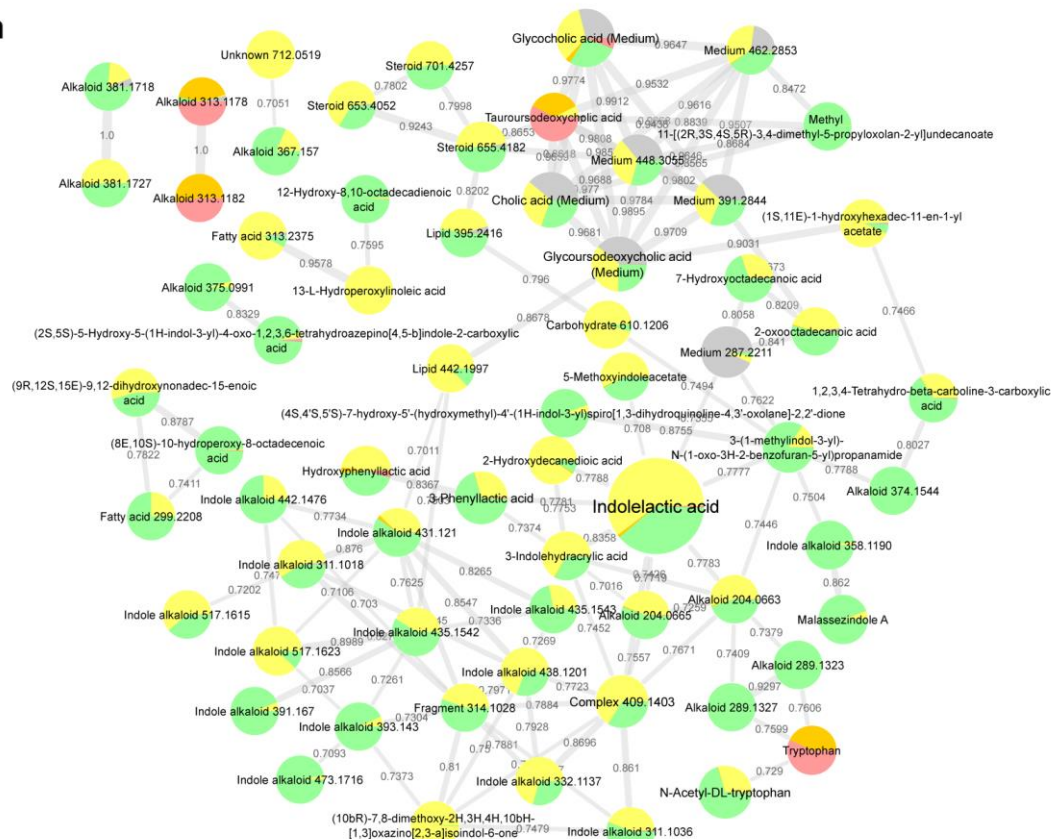


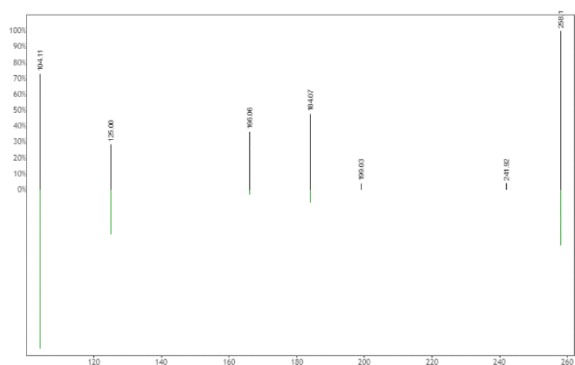
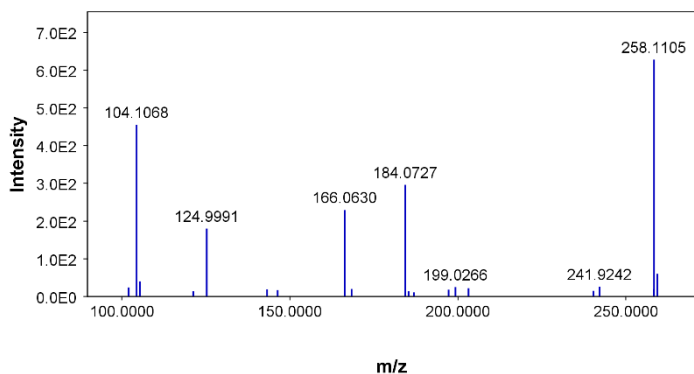
Figure S4: Molecular networks obtained after feature-based molecular networking GNPS workflow and visualized on Cytoscape. Process parameters are found in Material and methods section. Node labels display the proposed metabolite, or the compound class with the m/z ratio. Yellow: *S. reginensis* DCM/EA. Green: *S. kourouensis* DCM/EA. Orange: *S. reginensis* M. Red: *S. kourouensis* M. Gray: culture medium and solvent blank compounds. Non-significant culture medium compounds and singletons that are not in a network are not displayed. Node size is proportional to the precursor intensity, edge size to the cosine score between two nodes.

Table S5: pure chemical standards used for level 1 identification of six features.

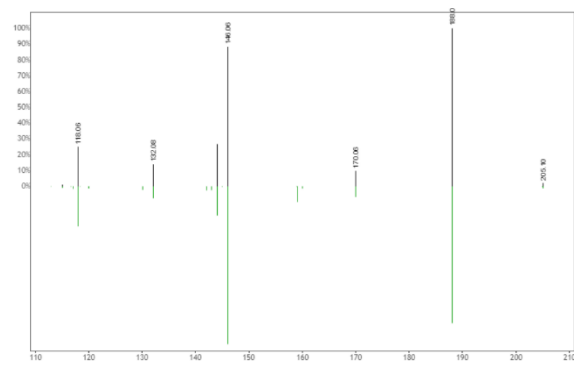
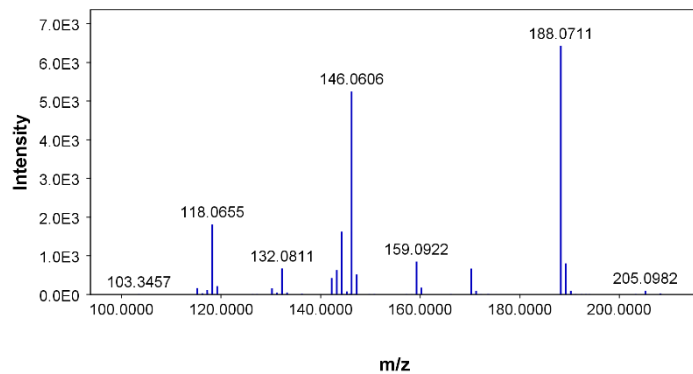
Molecular formula	Proposed metabolite	CAS	Fournisseur	Correlated feature ID
C ₁₁ H ₁₃ NO ₃	N-Acetyl-L-phenylalanine	2018-61-3	TCI	160
C ₉ H ₁₀ O ₃	3-Phenyllactic acid	703-80-0	Aldrich	154
C ₁₀ H ₉ N	6-Methylquinoline	495-69-2	Schuhschand München	332
C ₁₀ H ₇ NO ₃	4-Hydroxyquinoline-2-carboxylic acid	526-55-6	Sigma	79
C ₁₀ H ₉ NO ₂	Indole-3-acetic acid	20312-36-1	Alfa Aesar	206
C ₁₀ H ₁₁ NO	Tryptophol	91-62-3	Acros Organics	203

Figure S5: MS/MS spectra from various compounds and GNPS annotation process: mirror matches (Positive Ionisation mode). Black line: spectral library; Green line: experimental spectrum. The indicated annotation ID corresponds with the metabolite ID in the **Table S6**.

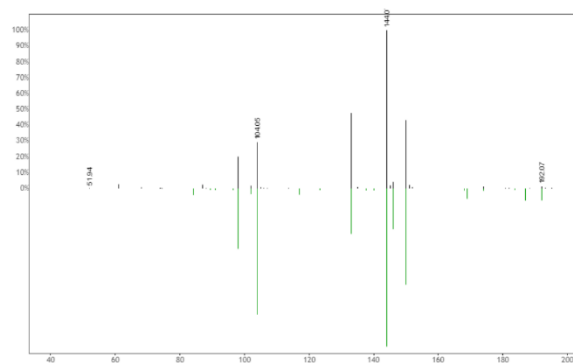
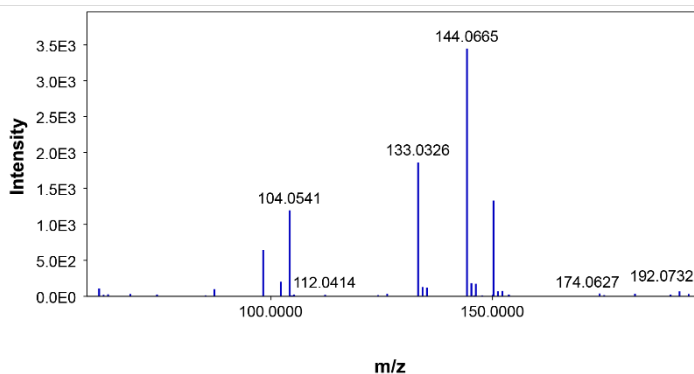
ID 1: Glycerophosphocholine



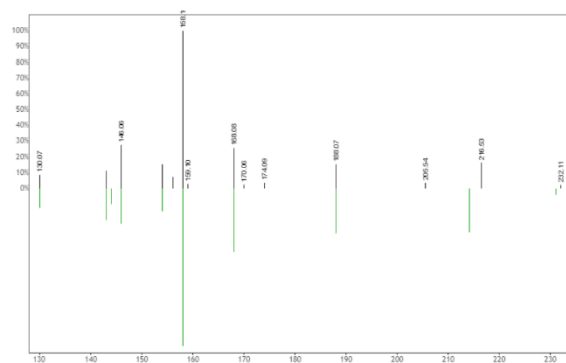
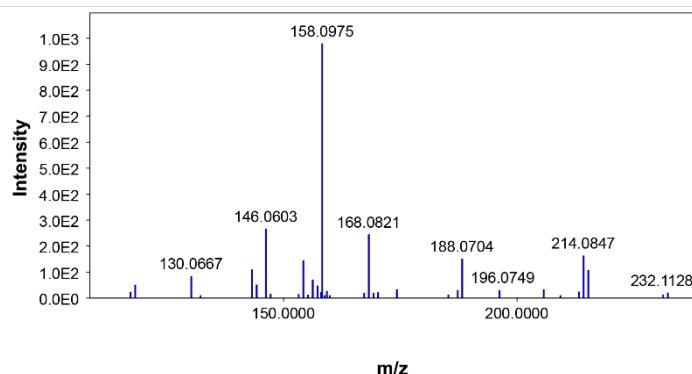
ID 56: L-Tryptophan



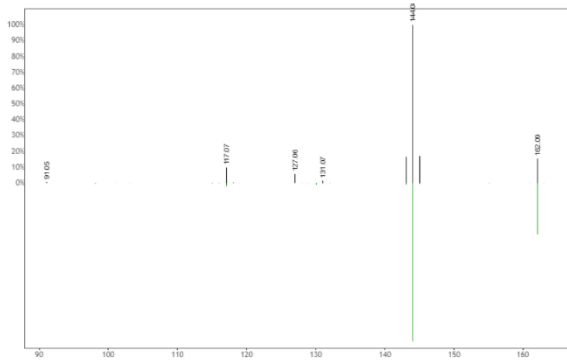
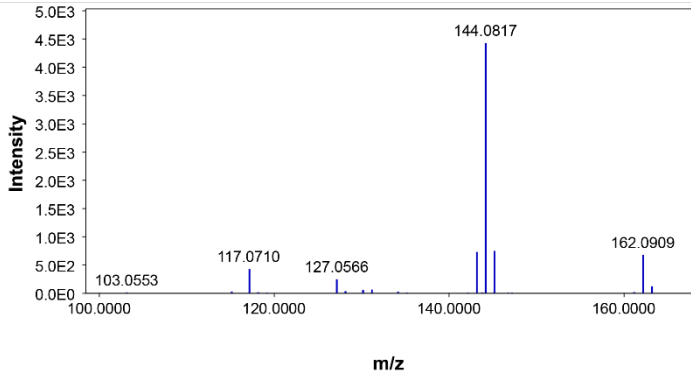
ID 66: N-Acetyl-L-methionine



ID 100: Tetrahydroharman-3-carboxylic acid



ID 197: Tryptophol



ID 288: 2-Quinolinecarboxylate

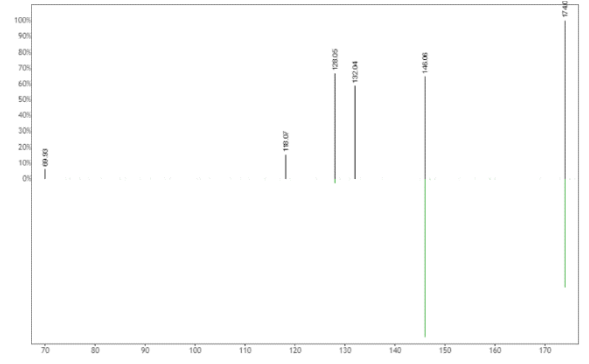
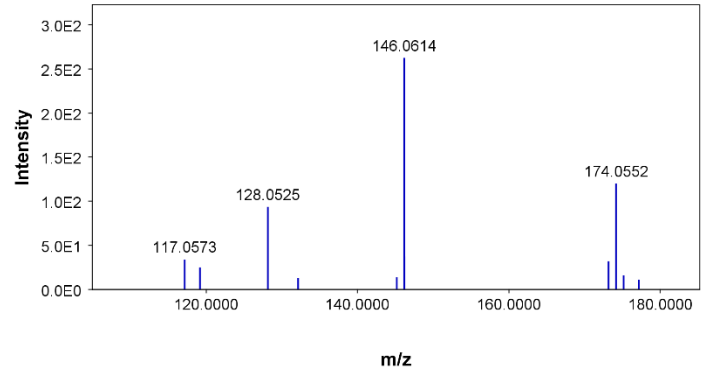
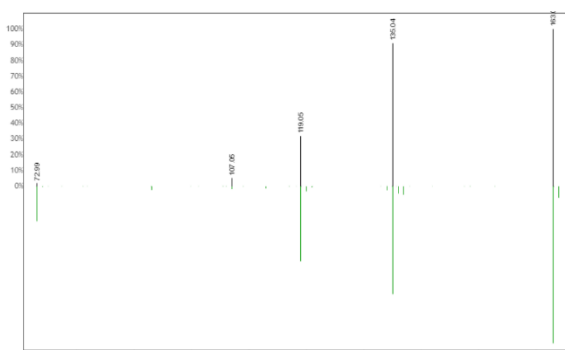
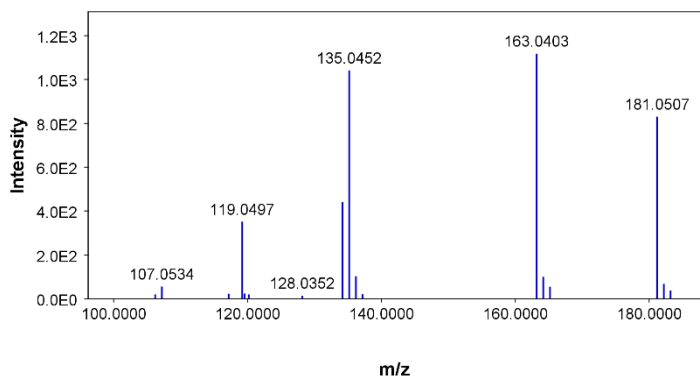
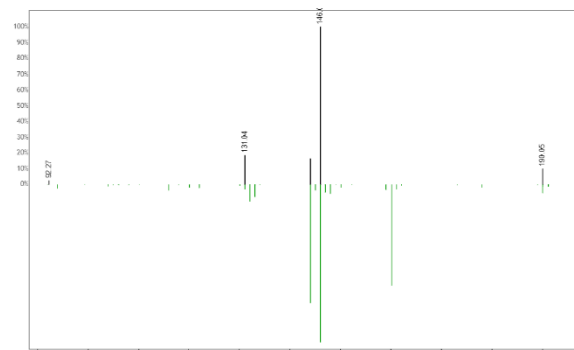
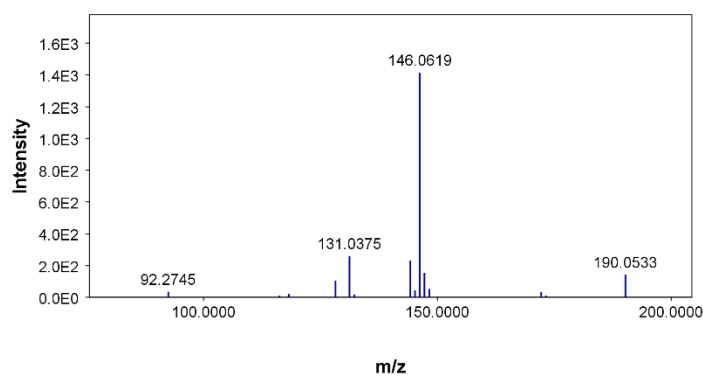


Figure S6: MS/MS spectra from various compounds and GNPS annotation process: mirror matches (Negative Ionisation mode). Black line: spectral library ; Green line: experimental spectrum. The indicated annotation ID corresponds with the metabolite ID in the **Table S6**.

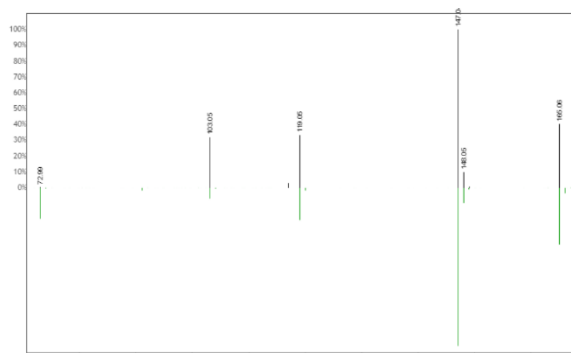
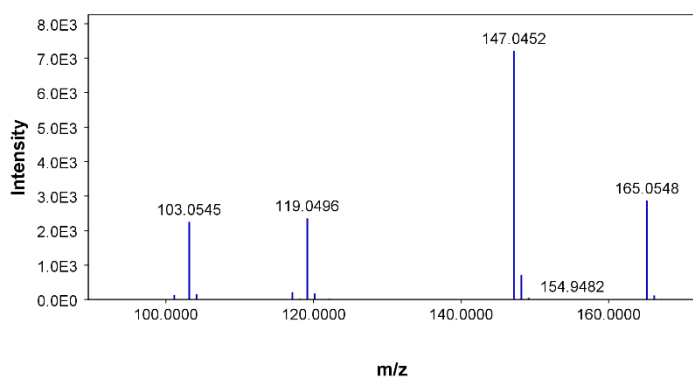
ID 57: Hydroxyphenyllactic acid



ID 132: Hydroxyindoleacetic acid



ID 148: 3-Phenyllactic acid



ID 177: N-acetyltryptophan

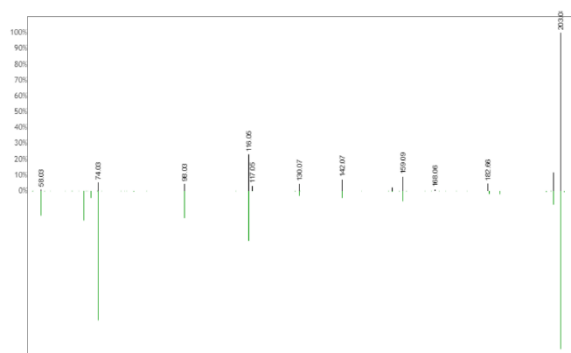
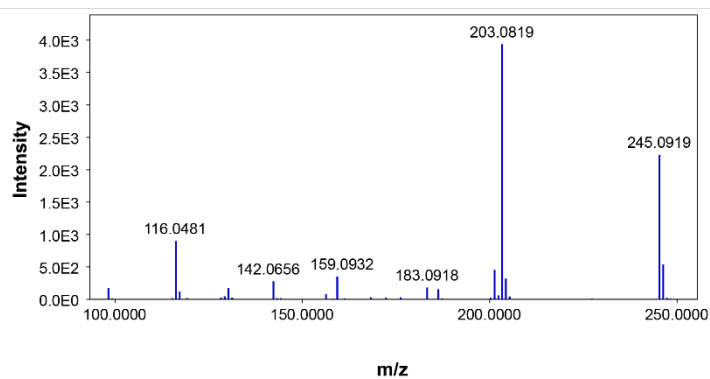


Table S6: Proposed metabolites from *Starmerella reginensis* and *Starmerella kourouensis* analyses. The compounds from dichloromethane/ethyl acetate extract (DCM/EA) and methanol extract (M) analyzed in positive and negative mode are present. Molecular formula of features without FTICR m/z error have been determined by UHPLC-MS/MS analysis. The last column indicates if the metabolite was also annotated in a previous work on *Saccharomyces cerevisiae*.

Peak ID	Observed m/z (QToF)	Retention time (min)	Ionization mode	Extract	Observed m/z (FTICR)	Error m/z (FTICR, ppm)	Molecular formula	Error m/z (ToF, ppm)	Proposed metabolite	Annotation level	Species	Found in <i>Saccharomyces cerevisiae</i>
1	258.1109	1.04	(+)	M	258.1101	0.13	C ₈ H ₂₀ NO ₆ P	3.00	Glycerophosphocholine	2	SR, SK	Yes (level 2)
2	339.0397	1.10	(-)	DCM/EA	-	-	-	-	Unknown	4	SK	
3	399.1440	1.12	(+)	M	-	-	C ₂₂ H ₂₂ O ₇	0.40	Carbohydrate	3	SR, SK	
4	371.0137	1.13	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
5	167.0387	1.16	(+)	DCM/EA	-	-	C ₆ H ₆ N ₄ S	0.50	6-Methylmercaptapurine	2	SR, SK	
6	189.0199	1.21	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
7	165.0228	1.23	(-)	DCM/EA	165.0227	0.17	C ₅ H ₁₀ O ₄ S	0.70	Sulfuric acid monoester	3	SR, SK	
8	208.0645	1.24	(+); (-)	DCM/EA	208.0638	-0.09	C ₇ H ₁₃ NO ₄ S	3.20	N-Acetylmethionine sulfoxide	2	SR, SK	Yes (level 2)
9	152.0584	1.29	(+)	M	-	-	C ₅ H ₅ N ₅ O	11.20	Guanine	2	SR, SK	
10	136.0626	1.31	(+)	M	136.0618	-0.02	C ₅ H ₅ N ₅	6.00	Adenine	2	SR, SK	
11	167.0740	1.43	(+)	DCM/EA	167.0736	-	-	-	Unknown	4	SR, SK	
12	208.0646	1.77	(+)	DCM/EA	208.0638	-0.09	C ₇ H ₁₃ NO ₄ S	3.70	Amino-acid	3	SR, SK	Yes (level 3)
13	167.0748	2.05	(+)	DCM/EA	167.0736	0.01	C ₆ H ₁₄ O ₃ S	6.80	5-(Methylsulfonyl)-1-pentanol	2	SR, SK	
14	141.0184	2.69	(-)	DCM/EA	141.0193	0.23	C ₆ H ₆ O ₄	-6.40	2,3-Methylenesuccinic acid	2	SR, SK	
15	228.0667	3.09	(+)	DCM/EA	228.0665	0.06	C ₈ H ₁₅ NO ₃ S	5.10	Amino-acid	3	SK	
16	297.0950	3.27	(+)	DCM/EA	297.0944	-0.14	C ₁₈ H ₁₆ O ₂ S	3.40	Aromatic sulfone	3	SR, SK	
17	147.0661	3.70	(-)	DCM/EA	147.0663	0.23	C ₆ H ₁₂ O ₄	-1.10	Carboxylic acid	3	SR, SK	
18	228.0666	3.72	(+)	DCM/EA	228.0665	0.06	C ₈ H ₁₅ NO ₃ S	4.70	N-Propionyl-L-methionine	2	SK	
19	192.0737	4.16	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
20	111.0451	4.18	(+)	DCM/EA	111.0441	0.03	C ₆ H ₆ O ₂	9.30	Hydroquinone	2	SR, SK	
21	191.0709	4.19	(+)	DCM/EA	191.0703	-0.05	C ₁₁ H ₁₀ O ₃	3.20	2,3-Dihydro-5-hydroxy-2-methyl-1,4-naphthoquinone	2	SR, SK	
22	297.0946	4.65	(+)	DCM/EA	297.0944	-0.14	C ₁₈ H ₁₆ O ₂ S	0.70	Aromatic sulfone	3	SK	
23	125.0243	4.76	(+)	DCM/EA	125.0234	-0.33	C ₆ H ₄ O ₃	7.70	2-Hydroxy-1,4-benzoquinone	2	SR, SK	
24	269.1134	5.25	(+)	M	269.1132	-0.03	C ₁₂ H ₁₆ N ₂ O ₅	0.70	Seryltyrosine	2	SR, SK	
25	215.0757	5.46	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
26	221.0929	5.66	(+)	M	-	-	C ₁₁ H ₁₂ N ₂ O ₃	3.70	5-Hydroxy-L-tryptophan	2	SR, SK	
27	138.0557	5.69	(-)	DCM/EA	138.0561	0.26	C ₇ H ₉ NO ₂	-2.40	Ethyl 2-pyrrolicarboxylate	2	SR	
28	222.0783	5.72	(+)	DCM/EA	-	-	C ₁₁ H ₁₁ NO ₄	9.90	1-(2-Hydroxyphenyl)-5-oxopyrrolidine-3-carboxylic acid	2	SR, SK	
29	297.0954	5.91	(+)	DCM/EA	297.0944	-0.14	C ₁₈ H ₁₆ O ₂ S	2.00	Aromatic sulfone	3	SR, SK	
30	321.0437	6.00	(-)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
31	339.0241	6.01	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
32	487.0439	6.03	(-)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
33	165.0225	6.09	(-)	DCM/EA	165.0227	0.17	C ₅ H ₁₀ O ₄ S	-1.10	[(3-Methylbut-2-en-1-yl)oxy]sulfonic acid	2	SR, SK	
34	310.0001	6.11	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
35	187.0402	6.11	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
36	356.0480	6.11	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
37	149.0275	6.14	(-)	DCM/EA	149.0278	-	C ₅ H ₁₀ O ₃ S	-3.10	2-Hydroxy-4-(methylthio)butyric acid	2	SR, SK	

38	168.0670	6.20	(+)	DCM/EA	168.0655	-0.05	C ₈ H ₉ NO ₃	8.70	Pyridoxal	2	SR, SK	
39	390.1199	6.23	(-)	DCM/EA	-	-	C ₁₉ H ₂₁ NO ₈	1.20	Glycoperine	2	SR, SK	
40	209.0809	6.31	(+)	DCM/EA	209.0809	-0.12	C ₁₁ H ₁₂ O ₄	0.20	6-Methoxymellein	2	SR, SK	
41	191.0716	6.33	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
42	239.0927	6.33	(+)	DCM/EA	239.0914	0.00	C ₁₂ H ₁₄ O ₅	5.40	1-(2,4,5-Trimethoxyphenyl)-1,2-propanedione	2	SR, SK	
43	149.0273	6.34	(-)	DCM/EA ; M	149.0278	-	C ₅ H ₁₀ O ₃ S	-3.10	Thia fatty acid	3	SR, SK	
44	151.0753	6.40	(+)	DCM/EA	151.0754	-0.12	C ₉ H ₁₀ O ₂	-0.50	1-(2-methoxyphenyl)ethan-1-one	2	SR, SK	
45	193.0531	6.57	(-)	DCM/EA	193.0540	0.19	C ₇ H ₁₄ O ₄ S	-4.50	Hydroxylated sulfated ester	3	SR, SK	Yes (level 3)
46	161.0813	6.60	(-)	DCM/EA	161.0819	0.01	C ₇ H ₁₄ O ₄	-3.80	2-(1-Ethoxyethoxy)propanoic acid	2	SR, SK	Yes (level 3)
47	145.0509	6.63	(-)	DCM/EA	-	-	C ₆ H ₁₀ O ₄	2.00	Methylglutaric acid	2	SR	
48	279.0858	6.75	(+)	DCM/EA	-	-	C ₁₄ H ₁₄ O ₆	-1.90	Methyl 3-(7,8-dihydroxy-3-methyl-1-oxoisochromen-6-yl)propanoate	2	SR, SK	
49	280.0911	6.75	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
50	261.1333	6.83	(-)	DCM/EA	261.1344	-0.01	C ₁₂ H ₂₂ O ₆	-4.00	5,8,12-trihydroxy-2-oxododecanoic acid	2	SK	
51	367.1521	6.84	(+)	M	367.1500	0.01	C ₁₇ H ₂₂ N ₂ O ₇	5.70	Tryptophan 2-C-mannoside	2	SR, SK	
52	181.0862	6.85	(+)	DCM/EA; M	181.0859	0.09	C ₁₀ H ₁₂ O ₃	1.40	4-(3,4-dihydroxyphenyl)butan-2-one	2	SK	
53	385.0888	6.87	(-)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
54	188.0715	6.88	(+)	M	188.0706	-0.07	C ₁₁ H ₉ NO ₂	4.70	Indoleacrylic acid	2	SR, SK	
55	429.1525	6.90	(-)	M	-	-	-	-	Unknown	4	SR, SK	
56	205.0977	6.90	(+); (-)	M	205.0972	-0.01	C ₁₁ H ₁₂ N ₂ O ₂	2.60	L-Tryptophan	2	SR, SK	
57	181.0501	6.93	(-)	DCM/EA; M	181.0506	0.01	C ₉ H ₁₀ O ₄	-2.80	Hydroxyphenyllactic acid	2	SR, SK	
58	403.0972	6.98	(-)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
59	436.0436	6.98	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
60	144.0667	7.02	(+)	DCM/EA	144.0655	-0.08	C ₆ H ₉ NO ₃	8.10	2-(but-3-enamido)acetic acid	2	SR, SK	
61	174.0583	7.02	(+)	DCM/EA	174.0583	-0.02	C ₇ H ₁₁ NO ₂ S	-	Unknown	4	SR, SK	
62	612.1379	7.02	(+); (-)	DCM/EA	-	-	-	-	Carbohydrate	3	SR	
63	380.0863	7.04	(+)	DCM/EA	-	-	-	-	Unknown	4	SK	
64	488.2357	7.06	(+)	M	-	-	C ₂₀ H ₃₃ N ₅ O ₉	1.20	Amino-acid	3	SR, SK	
65	310.0357	7.07	(+)	DCM/EA	-	-	-	-	Unknown	4	SK	
66	192.0699	7.08	(+); (-)	DCM/EA; M	192.0689	0.03	C ₇ H ₁₃ NO ₃ S	5.20	N-Acetyl-L-methionine	2	SR, SK	Yes (level 2)
67	170.0532	7.08	(+)	DCM/EA	-	-	-	-	Unknown	4	SK	
68	214.0514	7.09	(+)	DCM/EA	-	-	C ₅ H ₇ N ₇ OS	3.70	Unknown	4	SR, SK	
69	133.0680	7.11	(+)	DCM/EA	133.0682	-0.02	C ₆ H ₁₂ OS	-1.40	Alkylthiol	3	SK	
70	242.1035	7.14	(+)	DCM/EA	242.1022	0.24	C ₁₁ H ₁₅ NO ₅	4.90	Alkaloid	3	SR, SK	
71	161.0815	7.14	(-)	DCM/EA	161.0819	0.01	C ₇ H ₁₄ O ₄	-2.50	Carboxylic acid	3	SR, SK	
72	247.0587	7.15	(+)	DCM/EA	-	-	C ₁₃ H ₁₀ O ₅	-5.80	Coumarin	3	SR, SK	
73	373.1105	7.37	(-)	DCM/EA	-	-	C ₂₃ H ₁₉ ClN ₂ O	-2.10	Chlorinated alkaloid	3	SK	
74	190.0506	7.39	(+)	DCM/EA; M	190.0498	0.26	C ₁₀ H ₇ NO ₃	3.70	Kynurenic acid	1	SR, SK	Yes (level 1)
75	199.0571	7.40	(+)	DCM/EA	199.0577	-0.37	C ₁₃ H ₁₀ S	-2.60	Thiophene derivative	3	SR, SK	
76	175.0604	7.41	(-)	DCM/EA; M	175.0612	0.12	C ₇ H ₁₂ O ₅	-4.40	2,3-Dimethyl-3-hydroxyglutaric acid	2	SR, SK	
77	354.0617	7.45	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
78	245.0867	7.45	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
79	764.3568	7.46	(+)	M	-	-	-	-	Unknown	4	SR, SK	
80	365.0928	7.47	(+)	DCM/EA	-	-	-	-	Unknown	4	SK	
81	190.0503	7.48	(-)	DCM/EA	190.0510	0.09	C ₁₀ H ₉ NO ₃	-3.40	Indole alkaloid	3	SR, SK	
82	170.0530	7.49	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	

83	310.0359	7.49	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
84	133.0688	7.52	(+)	DCM/EA	133.0682	-0.02	C ₆ H ₁₂ OS	4.60	4-Methyl-4-sulfanylpentan-2-one	2	SR, SK	
85	173.0621	7.52	(+)	DCM/EA	173.0607	-0.04	C ₆ H ₁₃ NaO ₂ S	-	Unknown	4	SR, SK	
86	385.1028	7.54	(+)	DCM/EA	-	-	C ₁₉ H ₁₆ N ₂ O ₇	-0.70	Alkaloid	3	SR, SK	
87	189.0763	7.60	(-)	DCM/EA	189.0769	0.22	C ₈ H ₁₄ O ₅	-2.80	6-(2-Hydroxyethoxy)-6-oxohexanoic acid	2	SR	
88	215.0814	7.70	(-)	M	215.0826	0.25	C ₁₂ H ₁₂ N ₂ O ₂	-5.40	Alkaloid	3	SR, SK	
89	204.0652	7.75	(+)	DCM/EA	204.0655	0.02	C ₁₁ H ₉ NO ₃	-1.70	Indole-3-pyruvic acid	2	SR, SK	
90	217.0973	7.77	(+)	M	217.0971	0.08	C ₁₂ H ₁₂ N ₂ O ₂	0.60	Alkaloid	3	SR, SK	
91	321.1596	7.77	(+)	DCM/EA	-	-	C ₂₀ H ₂₀ N ₂ O ₂	-0.60	Indole alkaloid	3	SR, SK	
92	220.0605	7.79	(-)	DCM/EA	220.0615	0.19	C ₁₁ H ₁₁ NO ₄	-4.60	Alkaloid	3	SR, SK	
93	259.0710	7.91	(-)	M	259.0724	0.00	C ₁₃ H ₁₂ N ₂ O ₄	-5.40	1,2,3,4-Tetrahydro-b-carboline-1,3-dicarboxylic acid	2	SR, SK	
94	245.0884	7.96	(+)	DCM/EA	-	-	C ₈ H ₁₂ NaO ₅	1.40	Nucleoside	3	SR	
95	285.0792	8.05	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
96	151.0763	8.05	(+)	DCM/EA	151.0754	-0.12	C ₉ H ₁₀ O ₂	6.10	Benzenoid	3	SK	
97	239.0922	8.06	(+)	DCM/EA	239.0914	0.00	C ₁₂ H ₁₄ O ₅	3.30	Benzenoid	3	SK	
98	209.0811	8.09	(+)	DCM/EA	209.0809	-0.11	C ₁₁ H ₁₂ O ₄	1.20	3-(3-Methoxy-4,5-methylenedioxyphenyl)-2-propen-1-ol	2	SR, SK	
99	293.0994	8.09	(+)	DCM/EA	293.0995	-0.20	C ₁₉ H ₁₆ OS	-0.30	Sulfoxide	3	SR, SK	
100	231.1133	8.15	(+)	M	231.1128	-0.03	C ₁₃ H ₁₄ N ₂ O ₂	2.10	Tetrahydroharman-3-carboxylic acid	2	SR, SK	Yes (level 2)
101	248.0918	8.19	(+)	DCM/EA	-	-	C ₁₃ H ₁₃ NO ₄	0.20	Alkaloid	3	SR, SK	
102	157.0504	8.21	(-)	DCM/EA	-	-	C ₇ H ₁₀ O ₄	-1.30	Isopropylmaleic acid	2	SR, SK	Yes (level 2)
103	190.0500	8.22	(+)	DCM/EA; M	190.0498	0.26	C ₁₀ H ₇ NO ₃	0.60	8-Hydroxyquinoline-2-carboxylic acid	2	SK	
104	178.0872	8.22	(+)	DCM/EA	182.0863	-0.20	C ₁₀ H ₁₁ NO ₂	5.20	5-Hydroxytryptophol	2	SK	Yes (level 2)
105	334.1279	8.25	(+)	DCM/EA	334.1261	-0.12	C ₂₁ H ₁₉ NOS	5.60	Thiophene derivative	3	SR, SK	
106	204.0669	8.26	(-)	DCM/EA	204.0666	-0.10	C ₁₁ H ₁₁ NO ₃	1.50	Alkaloid	3	SR, SK	
107	235.0659	8.26	(-)	DCM/EA	235.0645	0.13	C ₉ H ₁₆ O ₅ S	5.70	Sulfated carboxylic acid	3	SK	
108	220.0611	8.27	(-); (+)	DCM/EA; M	220.0615	0.19	C ₁₁ H ₁₁ NO ₄	-1.80	Methyl 5-hydroxyoxindole-3-acetate	2	SR, SK	
109	266.1018	8.28	(+); (-)	DCM/EA	-	-	C ₁₃ H ₁₅ NO ₅	-2.00	Amino-acid	3	SR, SK	
110	142.0870	8.28	(+)	DCM/EA	142.0863	0.01	C ₇ H ₁₁ NO ₂	5.10	Alkaloid	3	SR, SK	
111	192.0668	8.31	(+)	DCM/EA	-	-	C ₁₀ H ₉ NO ₃	6.60	4-Oxo-2,3-dihydro-1H-quinoline-2-carboxylic acid	2	SR, SK	
112	285.1282	8.34	(-)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
113	131.0709	8.39	(-)	DCM/EA	131.0714	0.16	C ₆ H ₁₂ O ₃	-3.40	2-Hydroxyhexanoic acid	2	SR, SK	Yes (level 3)
114	244.0601	8.39	(+)	DCM/EA	-	-	C ₁₃ H ₉ NO ₄	-1.50	Alkaloid	3	SR, SK	
115	260.1132	8.43	(-)	DCM/EA	-	-	C ₁₁ H ₁₉ NO ₆	-2.80	Carbohydrate	3	SR, SK	
116	241.1181	8.52	(-)		241.1193	0.16	C ₁₁ H ₁₈ N ₂ O ₄	-5.20	Amino-acid	3	SR, SK	
117	196.0945	8.56	(+)	DCM/EA	-	-	-	-	Unknown	4	SR	
118	128.1091	8.56	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
119	174.1128	8.57	(+); (-)	DCM/EA	174.1125	-0.02	C ₈ H ₁₅ NO ₃	1.80	N-Acetyl-L-leucine	2	SR, SK	Yes (level 2)
120	382.1177	8.57	(+)	DCM/EA	-	-	C ₂₃ H ₁₅ N ₃ O ₃	-2.50	Alkaloid	3	SR, SK	
121	233.0835	8.57	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
122	153.0555	8.58	(+)	DCM/EA	153.0546	0.06	C ₈ H ₈ O ₃	5.60	2-hydroxy-1-(4-hydroxyphenyl)ethan-1-one	2	SR, SK	
123	191.0733	8.58	(+)	DCM/EA	-	-	C ₈ H ₁₄ O ₃ S	-1.90	Lipid	3	SR, SK	
124	269.0334	8.58	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
125	175.0757	8.59	(+)	DCM/EA	-	-	C ₁₁ H ₁₀ O ₂	1.80	4-Methoxy-1-naphthol	2	SR, SK	
126	235.0642	8.59	(-)	DCM/EA	235.0645	0.13	C ₉ H ₁₆ O ₅ S	-1.50	Sulfated carboxylic acid	3	SR, SK	
127	367.1848	8.60	(-)	DCM/EA	-	-	-	-	Unknown	4	SR	
128	343.1300	8.60	(+)	DCM/EA	343.1298	-	-	-	Unknown	4	SR	

129	111.0448	8.66	(+)	DCM/EA	111.0441	0.03	C ₆ H ₆ O ₂	6.50	1,3-Benzenediol	2	SR, SK	
130	214.0487	8.70	(+)	DCM/EA	-	-	-	-	Unknown	4	SK	
131	174.0557	8.73	(+)	DCM/EA	-	-	C ₁₀ H ₇ NO ₂	4.20	Alkaloid	3	SK	
132	190.0526	8.73	(-)	DCM/EA	190.0510	0.09	C ₁₀ H ₉ NO ₃	-3.40	5-Hydroxyindoleacetic acid	2	SR, SK	
133	193.0487	8.76	(-)	DCM/EA	-	-	-	-	Unknown	4	SR	
134	313.1182	8.77	(-)	M	313.1194	-0.02	C ₁₇ H ₁₈ N ₂ O ₄	-3.70	Alkaloid	3	SR, SK	
135	178.0863	8.78	(+)	DCM/EA	182.0863	-0.20	C ₁₀ H ₁₁ NO ₂	0.10	Alkaloid	3	SR, SK	
136	125.0242	8.79	(+)	DCM/EA	125.0234	-0.33	C ₆ H ₄ O ₃	6.90	Quinone	3	SR, SK	
137	191.0719	8.79	(+)	DCM/EA	-	-	-	-	Unknown	4	SK	
138	207.0275	8.80	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
139	127.0762	8.82	(-)	DCM/EA	127.0765	-0.02	C ₇ H ₁₂ O ₂	-1.80	Ethyl (2E)-2-methylbut-2-enoate	2	SR	
140	172.0983	8.86	(-)	DCM/EA	-	-	C ₈ H ₁₅ NO ₃	2.30	Amino-acid	3	SR, SK	
141	189.0749	8.92	(-)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
142	313.1178	8.94	(-)	M	313.1194	-0.02	C ₁₇ H ₁₈ N ₂ O ₄	-5.00	Alkaloid	3	SR, SK	
143	297.1244	8.95	(+)	M	297.1234	0.00	C ₁₇ H ₁₆ N ₂ O ₃	3.40	Indole alkaloid	3	SR, SK	
144	239.0741	9.07	(+)	DCM/EA	-	-	C ₁₂ H ₁₄ O ₃ S	1.80	Unknown	4	SR, SK	
145	267.0351	9.08	(-)	DCM/EA	-	-	-	-	Unknown	4	SR	
146	257.0859	9.12	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
147	297.0768	9.12	(+)	DCM/EA	297.0767	-0.96	C ₁₀ H ₁₂ N ₆ O ₃ S	1.20	Sulfated amine	3	SR, SK	
148	165.0551	9.19	(-)	DCM/EA	165.0557	0.04	C ₉ H ₁₀ O ₃	-3.60	3-Phenyllactic acid	1	SR, SK	Yes (level 2)
149	535.1269	9.19	(-)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
150	371.0802	9.20	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
151	253.0931	9.21	(+)	DCM/EA	-	-	C ₁₀ H ₁₂ N ₄ O ₄	-0.20	Nucleoside	3	SR	
152	293.0863	9.23	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
153	334.1280	9.23	(+)	DCM/EA	334.1261	-0.12	C ₂₁ H ₁₉ NOS	5.90	Thiophene derivative	3	SR, SK	
154	206.0817	9.41	(-)	DCM/EA	206.0823	0.09	C ₁₁ H ₁₃ NO ₃	-2.60	N-Acetyl-L-phenylalanine	1	SR, SK	Yes (level 1)
155	332.1137	9.43	(-); (+)	DCM/EA	332.1139	0.09	C ₁₇ H ₁₉ NO ₆	-0.70	Indole alkaloid	3	SR, SK	
156	247.0723	9.46	(-)	DCM/EA	247.0724	0.09	C ₁₂ H ₁₂ N ₂ O ₄	-0.40	5-Hydroxyindoleacetyl glycine	2	SR	
157	209.0800	9.51	(+)	DCM/EA	209.0809	-0.12	C ₁₁ H ₁₂ O ₄	-4.10	Heterocyclic	3	SK	
158	291.0708	9.60	(+)	DCM/EA	-	-	-	-	Unknown	4	SR	
159	215.0810	9.62	(+)	DCM/EA	215.0815	-0.04	C ₁₂ H ₁₀ N ₂ O ₂	-2.50	Alkaloid	3	SR	
160	336.1674	9.63	(+)	DCM/EA	336.1666	0.08	C ₁₅ H ₂₁ N ₅ O ₄	2.20	N ⁶ -(3-Methyl-2-butenyl)adenosine	2	SR, SK	
161	248.0909	9.64	(-)	DCM/EA	248.0928	0.02	C ₁₃ H ₁₅ NO ₄	-7.70	(10bR)-7,8-dimethoxy-2H,3H,4H,10bH-[1,3]oxazino[2,3-a]isoindol-6-one	2	SR	
162	465.0746	9.68	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
163	449.1036	9.69	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
164	217.0965	9.69	(+); (-)	DCM/EA	217.0972	0.00	C ₁₂ H ₁₂ N ₂ O ₂	-3.10	1,2,3,4-Tetrahydro-beta-carboline-3-carboxylic acid	2	SR, SK	
165	431.1210	9.69	(-)	DCM/EA	-	-	-	-	Indole alkaloid	3	SR, SK	
166	407.1240	9.70	(-); (+)	DCM/EA	407.1249	-0.17	C ₂₂ H ₂₀ N ₂ O ₆	-2.00	Indole alkaloid	2	SR, SK	
167	204.0658	9.70	(-); (+)	DCM/EA; M	204.0666	-0.10	C ₁₁ H ₁₁ NO ₃	-3.90	Indolelactic acid	2	SR, SK	Yes(level 2)
168	238.0711	9.71	(+)	DCM/EA	-	-	C ₁₁ H ₁₁ NO ₅	0.30	Alkaloid	3	SR, SK	
169	668.1363	9.73	(-); (+)	DCM/EA	-	-	C ₃₀ H ₂₇ N ₃ O ₁₅	-0.90	Amino-acid	3	SR	
170	389.1218	9.73	(+)	DCM/EA	389.1207	-0.23	C ₂₄ H ₂₀ O ₃ S	3.10	Unknown	4	SR, SK	
171	221.0703	9.75	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
172	228.0641	9.77	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
173	217.1071	9.80	(-)	DCM/EA	217.1081	0.16	C ₁₀ H ₁₈ O ₅	-4.70	2-Hydroxydecanedioic acid	2	SR, SK	

174	472.1487	9.81	(-)	DCM/EA	-	-	-	-	Unknown	4	SK	
175	490.1294	9.83	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
176	513.1757	9.84	(-)	DCM/EA	-	-	C ₂₇ H ₃₀ O ₁₀	-1.70	1-(1-hydroxy-6-methoxy-3-methyl-8-[(2R,3R,4S,5S,6R)-3,4,5-trihydroxy-6-[(1R)-1-hydroxy-2-(3-hydroxyphenyl)ethyl]oxan-2-yl)oxy)naphthalen-2-yl)ethanone	2	SR, SK	
177	245.0925	9.85	(-); (+)	DCM/EA	245.0932	0.00	C ₁₃ H ₁₄ N ₂ O ₃	-2.60	N-Acetyl-DL-tryptophan	2	SR, SK	Yes (level 2)
178	293.1144	9.88	(+)	DCM/EA	293.1150	-0.41	C ₁₂ H ₂₁ O ₆ P	-	Unknown	4	SR	
179	332.1137	9.90	(-); (+)	DCM/EA	332.1139	0.09	C ₁₇ H ₁₉ NO ₆	-0.70	Alkaloid	3	SR, SK	
180	356.1123	9.92	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
181	317.1112	9.93	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
182	256.0639	9.95	(-)	DCM/EA	256.0648	0.29	C ₁₁ H ₁₅ NO ₄ S	-3.80	N-(Phenylsulfonyl)-L-valine	2	SR, SK	
183	316.1192	10.00	(+)	DCM/EA	316.1179	0.03	C ₁₇ H ₁₇ NO ₅	3.90	Alkaloid	3	SR, SK	
184	191.0816	10.02	(+)	DCM/EA	191.0815	-0.08	C ₁₀ H ₁₀ N ₂ O ₂	0.40	6-Hydroxy-1H-indole-3-acetamide	2	SR, SK	
185	186.1127	10.16	(-)	DCM/EA	186.1135	0.15	C ₉ H ₁₇ NO ₃	-4.50	8-Amino-7-oxononanoic acid	2	SR	
186	272.1283	10.19	(+)	DCM/EA	272.1281	0.16	C ₁₆ H ₁₇ NO ₃	0.60	N-[2-hydroxy-2-(4-methoxyphenyl)ethyl]benzamide	2	SK	
187	355.0245	10.27	(-)	DCM/EA	-	-	-	-	Unknown	4	SR	
188	712.0519	10.27	(-)	DCM/EA	-	-	-	-	Unknown	4	SR	
189	232.0963	10.30	(+)	DCM/EA	-	-	C ₁₃ H ₁₃ NO ₃	-2.40	α-(2-Oxopropyl)-1H-indole-3-acetic acid	2	SR, SK	
190	204.0661	10.30	(-)	DCM/EA	204.0666	-0.10	C ₁₁ H ₁₁ NO ₃	-2.40	3-Indolehydracrylic acid	2	SR, SK	
191	354.0958	10.39	(+)	DCM/EA	-	-	C ₁₉ H ₁₅ NO ₆	-4.10	Alkaloid	3	SR, SK	
192	314.1017	10.42	(+)	DCM/EA	-	-	C ₁₇ H ₁₅ NO ₅	2.00	Alkaloid	2	SR, SK	
193	259.0841	10.56	(-)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
194	204.0663	10.61	(-)	DCM/EA	204.0666	-0.10	C ₁₁ H ₁₁ NO ₃	-1.40	Alkaloid	3	SR, SK	
195	230.0805	10.61	(+)	DCM/EA	230.0812	-0.03	C ₁₃ H ₁₁ NO ₃	-3.00	Quinoline alkaloid	3	SR, SK	
196	178.0861	10.62	(+)	DCM/EA	182.0863	-0.20	C ₁₀ H ₁₁ NO ₂	-1.00	Alkaloid	3	SR, SK	
197	162.0918	10.62	(+)	DCM/EA	162.0913	0.04	C ₁₀ H ₁₁ NO	2.70	Tryptophol	1	SR, SK	Yes (level 1)
198	130.0648	10.63	(-)	DCM/EA	-	-	-	-	Unknown	4	SK	
199	160.0767	10.63	(+)	DCM/EA	160.0757	-0.06	C ₁₀ H ₉ NO	6.20	Alkaloid	3	SR, SK	
200	174.0565	10.63	(-); (+)	DCM/EA	174.0560	0.24	C ₁₀ H ₉ NO ₂	2.70	(1H-Indol-3-yl)-acetate	1	SR, SK	Yes (level 1)
201	144.0450	10.65	(-)	DCM/EA	144.0455	0.04	C ₉ H ₇ NO	-3.20	Indole-3-carboxaldehyde	2	SK	
202	288.1227	10.70	(-); (+)	DCM/EA	288.1241	-0.04	C ₁₆ H ₁₉ NO ₄	-4.90	Alkaloid	3	SK	
203	263.1060	10.75	(+)	DCM/EA	-	-	C ₁₀ H ₁₈ N ₂ O ₄ S	-0.10	Methionyl-Hydroxyproline	2	SR	
204	295.0684	10.76	(-)	DCM/EA	-	-	-	-	Unknown	4	SK	
205	247.1088	10.79	(+)	DCM/EA	-	-	C ₁₃ H ₁₄ N ₂ O ₃	4.30	(1R)-1-(hydroxymethyl)-2,3,4,9-tetrahydro-1H-pyrido[3,4-b]indole-3-carboxylic acid	2	SR, SK	
206	245.0900	10.80	(+)	DCM/EA	-	-	-	-	Unknown	4	SR	
207	314.1031	10.89	(-); (+)	DCM/EA; M	314.1033	0.19	C ₁₇ H ₁₇ NO ₅	-0.90	4,5-Dihydroxy-3-methoxy-4-(4-methoxyphenyl)-1,3-dihydroquinolin-2-one	2	SR, SK	
208	337.1210	10.90	(-)	DCM/EA	-	-	C ₁₉ H ₁₈ N ₂ O ₄	4.90	Alkaloid	3	SR	Yes (level 4)
209	298.1079	10.97	(+)	DCM/EA	298.1081	-0.14	C ₁₀ H ₁₅ N ₇ O ₂ S	-0.70	Unknown	4	SR, SK	
210	313.1186	10.97	(+); (-)	DCM/EA	313.1183	0.11	C ₁₇ H ₁₆ N ₂ O ₄	0.90	Quinoline alkaloid	3	SR, SK	
211	160.0760	10.98	(+)	DCM/EA	160.0757	-0.06	C ₁₀ H ₉ NO	1.80	Indoleacetaldehyde	2	SR, SK	Yes (level 2)
212	338.1017	11.00	(+)	DCM/EA	-	-	C ₁₉ H ₁₅ NO ₅	-1.80	2-[(1E)-1-(4-Hydroxyphenyl)-3-methoxy-3-oxo-1-propen-2-yl]-1H-indole-3-carboxylic acid	2	SR, SK	
213	438.1201	11.12	(-)	DCM/EA	-	-	-	-	Alkaloid	3	SR, SK	

214	314.1026	11.35	(-); (+)	DCM/EA; M	314.1033	0.19	C ₁₇ H ₁₇ NO ₅	-2.50	5-hydroxy-1-(1-hydroxyethyl)-2-(2-hydroxyethyl)-3-methylbenzo[f]isoindole-4,9-dione	2	SR, SK	
215	190.0863	11.35	(+)	DCM/EA	190.0863	-0.04	C ₁₁ H ₁₁ NO ₂	0.10	2-(1H-Indol-3-yl)propanoic acid	2	SR, SK	Yes (level 2)
216	287.1189	11.36	(+)	DCM/EA	287.1180	-0.21	C ₁₉ H ₁₄ N ₂ O	3.40	Eudistomin T	2	SR, SK	
217	442.1476	11.45	(-)	DCM/EA	-	-	-	-	Indole alkaloid	3	SR, SK	
218	407.1228	11.48	(-)	DCM/EA	407.1249	-0.17	C ₂₂ H ₂₀ N ₂ O ₆	-5.00	Alkaloid	3	SR, SK	
219	318.1336	11.48	(-)	DCM/EA	-	-	C ₁₇ H ₂₁ NO ₅	-3.40	Alkaloid	3	SR, SK	
220	272.0925	11.50	(+)	DCM/EA	272.0951	-0.04	C ₁₂ H ₁₇ NO ₄ S	-9.70	2-Benzenesulfonamido-3-methylpentanoic acid	2	SR, SK	
221	440.1337	11.54	(-)	DCM/EA	440.1351	-0.06	C ₂₃ H ₂₃ NO ₈	-3.10	Heterocyclic	3	SR, SK	
222	314.1110	11.55	(+)	DCM/EA	-	-	C ₁₁ H ₁₅ N ₅ O ₆	4.70	8-Methoxyguanosine	2	SR, SK	
223	313.1192	11.58	(+); (-)	DCM/EA	313.1183	0.11	C ₁₇ H ₁₆ N ₂ O ₄	2.90	Quinoline alkaloid	3	SR, SK	
224	285.0880	11.64	(-)	DCM/EA	285.0881	0.05	C ₁₅ H ₁₄ N ₂ O ₄	-0.20	Alkaloid	3	SR, SK	
225	314.1027	11.69	(-)	DCM/EA	314.1033	0.19	C ₁₇ H ₁₇ NO ₅	-2.10	Alkaloid	3	SR, SK	
226	284.0917	11.74	(-)	DCM/EA	284.0929	-0.05	C ₁₆ H ₁₅ NO ₄	-3.90	Acridone alkaloid	3	SK	
227	268.0979	11.75	(+)	DCM/EA	-	-	C ₁₆ H ₁₃ NO ₃	4.00	1-((2-Hydroxyethyl)amino)anthraquinone	2	SK	
228	298.1084	11.76	(+)	DCM/EA	-	-	C ₁₇ H ₁₅ NO ₄	3.30	(7-Hydroxy-6-methoxy-3,4-dihydroisoquinolin-1-yl)-(3-hydroxyphenyl)methanone	2	SR, SK	
229	342.1228	11.77	(+)	DCM/EA	342.1237	0.14	C ₂₁ H ₁₅ N ₃ O ₂	-2.70	Carboline alkaloid	3	SK	
230	254.1174	11.87	(+)	DCM/EA	254.1175	0.14	C ₁₆ H ₁₅ NO ₂	-0.70	Ethyl 3-methyl-9H-carbazole-9-carboxylate	2	SK	Yes (level 2)
231	271.0722	12.00	(+)	DCM/EA	-	-	C ₁₅ H ₁₀ O ₅	-	Emodin	2	SK	
232	377.1144	12.04	(+)	DCM/EA	-	-	C ₂₁ H ₁₆ N ₂ O ₅	3.10	Alkaloid	3	SK	
233	232.0962	12.04	(+)	DCM/EA	-	-	C ₁₃ H ₁₃ NO ₃	-2.80	Alkaloid	3	SR, SK	
234	440.1348	12.04	(-)	DCM/EA	440.1351	-0.06	C ₂₃ H ₂₃ NO ₈	-0.60	Alkaloid	3	SR	
235	204.0671	12.04	(-)	DCM/EA	204.0666	-0.10	C ₁₁ H ₁₁ NO ₃	2.50	5-Methoxyindoleacetate	2	SR, SK	
236	314.1028	12.07	(+)	DCM/EA	-	-	C ₁₇ H ₁₅ NO ₅	1.50	Alkaloid	3	SR	
237	314.1027	12.15	(-); (+)	DCM/EA	314.1033	0.19	C ₁₇ H ₁₇ NO ₅	-2.10	Alkaloid	3	SR, SK	
238	313.1184	12.19	(+); (-)	DCM/EA	313.1183	0.11	C ₁₇ H ₁₆ N ₂ O ₄	0.30	Quinoline alkaloid	3	SR, SK	
239	240.0701	12.21	(+)	DCM/EA	-	-	C ₁₁ H ₁₃ NO ₃ S	5.00	Amino-acid	3	SR, SK	
240	307.1076	12.21	(+)	DCM/EA	307.1077	0.10	C ₁₈ H ₁₄ N ₂ O ₃	-0.50	Alkaloid	3	SK	
241	279.0860	12.21	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
242	375.1323	12.26	(+)	DCM/EA	375.1339	0.22	C ₂₂ H ₁₈ N ₂ O ₄	-4.40	Unknown	4	SR, SK	
243	230.0814	12.27	(+); (-)	DCM/EA	230.0812	-0.03	C ₁₃ H ₁₁ NO ₃	0.90	6-hydroxy-3,3-dimethyl-4H-cyclopenta[b]indole-1,2-dione	2	SR, SK	
244	235.0578	12.30	(+)	DCM/EA	235.0577	-0.26	C ₁₆ H ₁₀ S	0.80	Thiophene derivative	3	SK	
245	442.1495	12.30	(-)	DCM/EA	442.1508	-0.09	C ₂₃ H ₂₅ NO ₈	-2.70	Alkaloid	3	SR, SK	
246	153.0553	12.32	(+)	DCM/EA	153.0546	0.06	C ₈ H ₈ O ₃	4.30	Phenol	3	SK	
247	517.1615	12.34	(-)	DCM/EA	-	-	C ₂₈ H ₂₆ N ₂ O ₈	-0.20	Indole alkaloid	3	SR, SK	
248	232.0963	12.36	(+)	DCM/EA	-	-	C ₁₃ H ₁₃ NO ₃	-2.40	Alkaloid	3	SR, SK	
249	206.1185	12.37	(+)	DCM/EA	206.1176	-0.02	C ₁₂ H ₁₅ NO ₂	4.50	Alkaloid	3	SR, SK	
250	541.1590	12.42	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
251	254.1185	12.42	(+)	DCM/EA	254.1175	0.14	C ₁₆ H ₁₅ NO ₂	3.60	Alkaloid	3	SK	Yes (level 3)
252	257.0921	12.42	(+)	DCM/EA	-	-	C ₁₄ H ₁₂ N ₂ O ₃	0.00	7-Hydroxy-beta-carboline-1-propionic acid	2	SR, SK	
253	314.1017	12.44	(+)	DCM/EA	-	-	C ₁₇ H ₁₅ NO ₅	-2.00	Alkaloid	3	SR, SK	
254	315.1052	12.45	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
255	272.1279	12.47	(+)	DCM/EA	272.1281	0.16	C ₁₆ H ₁₇ NO ₃	-0.90	3,6-Dimethoxy-1,2-dimethyl-9H-carbazol-4-ol	2	SK	
256	333.1239	12.48	(+)	DCM/EA	333.1234	-0.08	C ₂₀ H ₁₆ N ₂ O ₃	1.50	Alkaloid	3	SK	
257	262.0710	12.49	(+)	DCM/EA	-	-	C ₁₃ H ₁₁ NO ₅	-0.10	1-(Ethoxycarbonyl)-4-oxo-4H-quinolizine-3-carboxylic acid	2	SK	

258	333.1236	12.52	(-)	DCM/EA	333.1244	0.18	C ₂₀ H ₁₈ N ₂ O ₃	-2.50	Alkaloid	3	SR, SK	
259	670.1526	12.52	(+); (-)	DCM/EA	-	-	C ₃₀ H ₂₇ N ₃ O ₁₅	1.60	Enterobactin	2	SR	
260	190.0872	12.53	(+)	DCM/EA	190.0863	-0.04	C ₁₁ H ₁₁ NO ₂	4.90	4-Methyl-2-(3-methyl-1,2-oxazol-5-yl)phenol	2	SR, SK	
261	426.1019	12.54	(-)	DCM/EA	426.1017	0.03	C ₂₂ H ₂₁ NO ₆ S	0.60	Unknown	4	SR	
262	174.0562	12.54	(+)	DCM/EA	-	-	C ₁₀ H ₇ NO ₂	7.10	Quinoline-4-carboxylic acid	2	SK	
263	377.1144	12.54	(-); (+)	DCM/EA	377.1143	-0.02	C ₂₁ H ₁₈ N ₂ O ₅	0.30	(4S,4'S,5'S)-7-hydroxy-5'-(hydroxymethyl)-4'-(1H-indol-3-yl)spiro[1,3-dihydroquinoline-4,3'-oxolane]-2,2'-dione	2	SK	
264	384.0987	12.60	(-); (+)	DCM/EA	-	-	C ₂₂ H ₁₅ N ₃ O ₄	-0.70	Alkaloid	3	SK	
265	435.1543	12.61	(-)	DCM/EA	-	-	-	-	Indole alkaloid	3	SR, SK	
266	347.1042	12.65	(-); (+)	DCM/EA	347.1038	-0.14	C ₂₀ H ₁₆ N ₂ O ₄	1.40	Alkaloid	3	SR, SK	
267	431.1717	12.69	(-); (+)	DCM/EA	431.1725	-0.03	C ₂₄ H ₂₄ N ₄ O ₄	-1.70	Alkaloid	3	SR, SK	
268	533.3254	12.85	(+)	M	-	-	C ₃₄ H ₄₄ O ₅	-1.50	Unknown	4	SR, SK	
269	375.0991	12.86	(-); (+)	DCM/EA	374.0987	-0.04	C ₂₁ H ₁₆ N ₂ O ₅	1.30	Alkaloid	3	SK	
270	514.2847	12.88	(-); (+)	M	514.2844	-0.06	C ₂₆ H ₄₅ NO ₇ S	0.60	Taurocholic acid	2	SR, SK	
271	388.1424	12.89	(+)	DCM/EA	-	-	-	-	Unknown	4	SK	
272	359.1047	12.89	(+)	DCM/EA	-	-	C ₂₁ H ₁₄ N ₂ O ₄	5.70	Alkaloid	3	SK	
273	582.2714	12.90	(-)	M	-	-	C ₃₃ H ₃₇ N ₅ O ₅	-1.30	Alkaloid	3	SR, SK	
274	341.0913	12.92	(-); (+)	DCM/EA	-	-	C ₂₁ H ₁₄ N ₂ O ₃	-5.40	Alkaloid	3	SK	
275	269.1296	12.93	(+)	DCM/EA	269.1284	0.17	C ₁₆ H ₁₆ N ₂ O ₂	4.20	Ethyl beta-carboline-1-propionate	2	SR, SK	Yes (level 2)
276	133.0688	12.95	(+)	DCM/EA	133.0682	-0.02	C ₆ H ₁₂ OS	4.60	2-Methyl-3-sulfanylpentanal	2	SR, SK	
277	374.1146	12.98	(-)	DCM/EA	374.1146	-0.03	C ₂₁ H ₁₇ N ₃ O ₄	0.00	(2S,5S)-5-Hydroxy-5-(1H-indol-3-yl)-4-oxo-1,2,3,6-tetrahydroazepino[4,5-b]indole-2-carboxylic acid	2	SK	Yes (level 3)
278	297.1236	13.00	(+); (-)	DCM/EA	297.1234	0.00	C ₁₇ H ₁₈ N ₂ O ₃	0.70	3-[(3-(2-Carboxyethyl)-4-methylpyrrol-2-yl)methylene]-2-indolinone	2	SR, SK	
279	426.1068	13.13	(-)	DCM/EA	-	-	-	-	Unknown	4	SK	
280	243.0776	13.19	(+)	DCM/EA	243.0764	0.10	C ₁₃ H ₁₀ N ₂ O ₃	4.80	1-Acetyl-beta-carboline	2	SK	Yes (level 3)
281	358.1184	13.20	(-); (+)	DCM/EA	358.1197	0.01	C ₂₁ H ₁₇ N ₃ O ₃	-3.60	Malassezindole A	2	SK	Yes (level 2)
282	270.1102	13.22	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
283	473.1716	13.23	(-)	DCM/EA	473.1719	-0.20	C ₂₇ H ₂₆ N ₂ O ₆	-0.40	Indole alkaloid	3	SK	
284	314.1027	13.29	(+)	DCM/EA	-	-	C ₁₇ H ₁₅ NO ₅	1.20	Alkaloid	3	SR, SK	
285	334.2739	13.30	(+)	DCM/EA	334.2741	0.01	C ₂₁ H ₃₅ NO ₂	-0.50	Unknown	4	SR	
286	541.1561	13.31	(+)	DCM/EA	-	-	-	-	Unknown	4	SR	
287	536.2018	13.32	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
288	174.0559	13.34	(+)	DCM/EA	-	-	C ₁₀ H ₇ NO ₂	5.30	Quinoline-2-carboxylic acid	2	SK	Yes (level 2)
289	503.1826	13.38	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
290	358.1190	13.38	(+)	DCM/EA	-	-	C ₂₁ H ₁₅ N ₃ O ₃	1.00	Chromozepinone A	3	SK	
291	517.1623	13.38	(-)	DCM/EA	-	-	C ₂₈ H ₂₆ N ₂ O ₈	-0.20	Indole alkaloid	3	SR, SK	
292	227.1261	13.39	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
293	190.0868	13.39	(+)	DCM/EA	190.0863	-0.04	C ₁₁ H ₁₁ NO ₂	2.80	Methyl 2-(1H-indol-3-yl)acetate	2	SR, SK	Yes (level 2)
294	435.1542	13.40	(-)	DCM/EA	-	-	-	-	Indole alkaloid	3	SR, SK	
295	289.1327	13.41	(-)	DCM/EA	289.1346	0.09	C ₁₉ H ₁₈ N ₂ O	-6.60	Alkaloid	3	SK	Yes (level 3)
296	289.1345	13.41	(+)	DCM/EA	289.1335	0.03	C ₁₉ H ₁₆ N ₂ O	3.30	Alkaloid	3	SK	Yes (level 3)
297	374.1544	13.42	(-)	DCM/EA	-	-	C ₂₇ H ₂₁ NO	-1.60	Alkaloid	3	SK	
298	218.0805	13.42	(+)	DCM/EA	218.0811	0.17	C ₁₂ H ₁₁ NO ₃	-3.20	Alkaloid	3	SK	Yes (level 3)
299	303.1128	13.45	(-); (+)	DCM/EA	303.1139	0.06	C ₁₉ H ₁₆ N ₂ O ₂	-3.50	2,2-Bis(1H-indol-3-yl)propanoic acid	2	SK	Yes (level 3)
300	232.0971	13.46	(+)	DCM/EA	-	-	C ₁₃ H ₁₃ NO ₃	1.10	(7S)-3'-[(1R)-1-hydroxyethyl]spiro[5,6-dihydrocyclopenta[c]pyridine-7,5'-furan]-2'-one	2	SR, SK	

301	393.1430	13.48	(-)	DCM/EA	-	-	-	-	Indole alkaloid	3	SK	
302	375.0977	13.49	(-); (+)	DCM/EA	375.0987	-0.04	C ₂₁ H ₁₆ N ₂ O ₅	-2.50	Alkaloid	3	SK	
303	149.0643	13.52	(+)	DCM/EA	149.0631	-0.02	C ₆ H ₁₂ O ₂ S	8.10	Sulfated ester	3	SK	
304	545.1939	13.58	(+)	DCM/EA	-	-	C ₃₁ H ₂₄ N ₆ O ₄	1.30	Alkaloid	3	SR, SK	
305	226.0509	13.64	(+)	DCM/EA	-	-	C ₁₃ H ₇ NO ₃	4.50	Quinoline alkaloid	3	SK	
306	343.1092	13.64	(+)	DCM/EA	-	-	-	-	Unknown	4	SK	
307	693.2300	13.67	(-)	DCM/EA	-	-	-	-	Unknown	4	SK	
308	462.2877	13.69	(-)	DCM/EA	462.2861	0.09	C ₂₆ H ₄₁ NO ₆	3.50	Unknown	4	SR, SK	
309	333.1235	13.70	(+)	DCM/EA	333.1234	-0.08	C ₂₀ H ₁₆ N ₂ O ₃	0.30	Alkaloid	3	SR, SK	Yes (level 3)
310	130.0660	13.71	(+)	DCM/EA	130.0651	0.18	C ₉ H ₇ N	6.60	Quinoline	2	SR, SK	
311	333.1234	13.71	(-); (+)	DCM/EA	333.1245	0.01	C ₂₀ H ₁₈ N ₂ O ₃	-3.10	3-(1-methylindol-3-yl)-N-(1-oxo-3H-2-benzofuran-5-yl)propanamide	2	SR, SK	Yes (level 3)
312	349.1186	13.73	(+)	DCM/EA	-	-	C ₂₀ H ₁₆ N ₂ O ₄	0.80	Alkaloid	3	SR, SK	
313	218.0798	13.74	(+)	DCM/EA	-	-	C ₁₂ H ₁₁ NO ₃	-6.40	3-(Indol-3-yl)-2-oxobutyric acid	2	SR, SK	
314	473.1720	13.77	(-)	DCM/EA	473.1719	-0.20	C ₂₇ H ₂₆ N ₂ O ₆	0.50	Indole alkaloid	3	SR, SK	
315	515.1826	13.79	(+)	DCM/EA	-	-	C ₂₉ H ₂₆ N ₂ O ₇	2.50	Alkaloid	3	SR, SK	
316	314.1034	13.85	(+)	DCM/EA	-	-	C ₁₇ H ₁₅ NO ₅	3.40	Alkaloid	3	SR, SK	
317	358.1190	13.86	(-); (+)	DCM/EA	358.1197	0.01	C ₂₁ H ₁₇ N ₃ O ₃	-1.90	Indole alkaloid	3	SK	Yes (level 3)
318	270.1127	13.87	(+)	DCM/EA	270.1125	0.08	C ₁₆ H ₁₅ NO ₃	0.80	Alkaloid	3	SK	
319	243.0767	13.91	(+)	DCM/EA	243.0764	0.10	C ₁₃ H ₁₀ N ₂ O ₃	1.10	Alkaloid	3	SK	Yes (level 3)
320	825.4455	13.91	(+)	DCM/EA	-	-	C ₄₇ H ₆₀ N ₄ O ₉	2.60	Alkaloid	3	SR, SK	
321	427.1305	14.03	(-)	DCM/EA	427.1300	-0.08	C ₂₅ H ₂₀ N ₂ O ₅	1.40	Alkaloid	3	SK	
322	498.2891	14.11	(-)	M	498.2895	-0.03	C ₂₆ H ₄₅ NO ₆ S	-0.70	Steroid	3	SR, SK	
323	232.0964	14.14	(+)	DCM/EA	-	-	C ₁₃ H ₁₃ NO ₃	-1.90	Alkaloid	3	SK	
324	391.1670	14.16	(-)	DCM/EA	391.1663	0.11	C ₂₃ H ₂₄ N ₂ O ₄	1.80	Indole alkaloid	3	SK	
325	359.1038	14.17	(-); (+)	DCM/EA	359.1037	0.06	C ₂₁ H ₁₆ N ₂ O ₄	0.30	Indole alkaloid	3	SK	
326	144.0816	14.27	(+)	DCM/EA	144.0808	-0.09	C ₁₀ H ₉ N	5.60	6-Methylquinoline	1	SK	Yes (level 2)
327	500.3013	14.42	(+); (-)	M	-	-	C ₂₆ H ₄₅ NO ₆ S	-5.50	Tauroursodeoxycholic acid	2	SR, SK	Yes (level 2)
328	244.0611	14.46	(+)	DCM/EA	-	-	C ₁₃ H ₉ NO ₄	2.60	Alkaloid	3	SK	
329	200.0714	14.47	(+)	DCM/EA	200.0706	0.00	C ₁₂ H ₉ NO ₂	3.90	Alkaloid	3	SK	
330	188.0723	14.59	(+)	DCM/EA	188.0706	-0.07	C ₁₁ H ₉ NO ₂	8.90	Alkaloid	3	SK	
331	333.1240	14.60	(+)	DCM/EA	333.1234	-0.08	C ₂₀ H ₁₆ N ₂ O ₃	1.80	Alkaloid	3	SK	Yes (level 3)
332	287.1172	14.66	(+)	DCM/EA	287.1180	-0.21	C ₁₉ H ₁₄ N ₂ O	-2.50	2-Phenyl-1-(9H-pyrido[3,4-b]indol-1-yl)ethanone	2	SK	Yes (level 2)
333	269.0917	14.67	(+)	DCM/EA	269.0920	0.27	C ₁₅ H ₁₂ N ₂ O ₃	-1.50	1-Acetyl-3-methoxycarbonyl-beta-carboline	2	SR, SK	
334	303.1131	14.68	(-)	DCM/EA	303.1139	0.06	C ₁₉ H ₁₆ N ₂ O ₂	-2.60	Indole alkaloid	3	SK	Yes (level 3)
335	358.1206	14.71	(+)	DCM/EA	-	-	C ₂₁ H ₁₅ N ₃ O ₃	5.50	Protoviolaceinic acid	2	SK	
336	204.0665	14.72	(-)	DCM/EA	204.0666	-0.10	C ₁₁ H ₁₁ NO ₃	-0.50	Alkaloid	3	SR, SK	
337	322.1438	14.72	(+)	DCM/EA	-	-	C ₂₀ H ₁₉ NO ₃	0.00	Alkaloid	3	SR, SK	
338	289.1357	14.80	(+); (-)	DCM/EA	289.1335	0.03	C ₁₉ H ₁₆ N ₂ O	7.40	Malassezione	2	SK	Yes (level 3)
339	174.0915	14.81	(+)	DCM/EA	174.0913	0.09	C ₁₁ H ₁₁ NO	0.80	4-Amino-2-methyl-1-naphthol	2	SK	
340	265.1423	14.81	(-)	DCM/EA	265.1446	-0.05	C ₁₅ H ₂₂ O ₄	-	Sesquiterpenoid	3	SK	
341	289.1323	14.85	(-); (+)	DCM/EA	289.1346	0.09	C ₁₉ H ₁₈ N ₂ O	-8.00	Indole alkaloid	3	SK	Yes (level 3)
342	269.1309	14.91	(-)	DCM/EA	269.1311	0.25	C ₁₄ H ₂₃ O ₃ P	-1.00	Unknown	4	SR, SK	
343	267.1591	15.12	(-)	DCM/EA	267.1601	0.18	C ₁₅ H ₂₄ O ₄	-3.90	Sesquiterpenoid	3	SK	
344	309.1693	15.27	(-)	DCM/EA	309.1707	0.15	C ₁₇ H ₂₆ O ₅	-4.60	Terpenoid	3	SR, SK	

345	508.3282	15.43	(+); (-)	DCM/EA	-	-	C ₂₈ H ₄₅ NO ₇	2.60	19-(2-Acetamido-2-deoxy-alpha-D-glucopyranosyloxy)isopimara-7,15-dien-3beta-ol	2	SK	
346	298.1080	15.89	(+)	DCM/EA	298.1081	-0.14	C ₁₀ H ₁₅ N ₇ O ₂ S	-0.30	Sulfated purine	3	SR, SK	
347	472.1736	15.90	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
348	311.1857	15.92	(+); (-)	DCM/EA	311.1853	0.03	C ₁₇ H ₂₆ O ₅	1.20	Valdiate	2	SR, SK	
349	311.1829	15.96	(-)	DCM/EA	-	-	-	-	Unknown	4	SK	
350	415.2126	16.03	(+)	DCM/EA	415.2115	0.07	C ₂₄ H ₃₀ O ₆	-	Steroid	3	SR	
351	280.1333	16.04	(+)	DCM/EA	-	-	C ₁₈ H ₁₇ NO ₂	0.30	6-Hydroxy-5-(3-methylbut-2-enyl)-9H-carbazole-3-carbaldehyde	2	SK	
352	341.1983	16.08	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
353	285.1335	16.10	(+)	DCM/EA	-	-	C ₁₄ H ₂₀ O ₆	0.70	2-Phenylethyl beta-D-glucopyranoside	2	SR, SK	
354	395.2416	16.11	(-)	DCM/EA	395.2440	-0.16	C ₂₂ H ₃₆ O ₆	-5.80	Lipid	3	SR, SK	
355	419.2385	16.13	(+)	DCM/EA	-	-	C ₁₉ H ₃₄ N ₂ O ₈	-0.80	Unknown	4	SR, SK	
356	322.1438	16.18	(+)	DCM/EA	-	-	C ₂₀ H ₁₉ NO ₃	0.00	Auxinole	2	SR, SK	
357	309.1691	16.18	(-)	DCM/EA	309.1707	0.15	C ₁₇ H ₂₆ O ₅	-5.20	Terpenoid	3	SK	
358	280.1333	16.42	(+)	DCM/EA	-	-	C ₁₈ H ₁₇ NO ₂	0.30	Alkaloid	3	SK	
359	289.1329	16.53	(+); (-)	DCM/EA	289.1335	0.03	C ₁₉ H ₁₆ N ₂ O	-2.30	Alkaloid	3	SK	Yes (level 3)
360	418.1644	16.55	(-)	DCM/EA	-	-	C ₂₅ H ₂₅ NO ₅	-3.80	Alkaloid	3	SK	
361	352.1544	16.58	(-)	DCM/EA	352.1554	0.15	C ₂₁ H ₂₃ NO ₄	-2.90	Alkaloid	3	SR, SK	
362	481.2126	16.58	(-)	DCM/EA	481.2134	-0.19	C ₃₀ H ₃₀ N ₂ O ₄	-1.40	Alkaloid	3	SK	
363	525.2026	16.59	(-)	DCM/EA	-	-	-	-	Unknown	4	SR	
364	292.1338	16.59	(+)	DCM/EA	-	-	C ₁₉ H ₁₇ NO ₂	2.00	Quinone	3	SK	
365	359.1037	16.60	(+)	DCM/EA	-	-	C ₂₁ H ₁₄ N ₂ O ₄	2.90	Alkaloid	3	SK	
366	278.1537	16.60	(+)	DCM/EA	278.1539	0.08	C ₁₉ H ₁₉ NO	-1.00	Quinoline alkaloid	3	SK	
367	269.1071	16.60	(-)	DCM/EA	-	-	-	-	Unknown	4	SK	
368	322.1433	16.60	(+)	DCM/EA	-	-	C ₂₀ H ₁₉ NO ₃	-1.50	Alkaloid	3	SR, SK	
369	293.1751	16.66	(-); (+)	DCM/EA	293.1758	0.20	C ₁₇ H ₂₆ O ₄	-2.40	7-(Hydroxymethyl)-2-(2-hydroxypropan-2-yl)-6-pentyl-2,3-dihydro-1-benzofuran-3-ol	2	SR, SK	
370	459.2008	16.72	(-)	DCM/EA	-	-	C ₂₅ H ₃₂ O ₈	-3.50	Polyketide	3	SR	
371	531.2752	16.73	(+)	DCM/EA	-	-	C ₃₃ H ₃₈ O ₆	2.00	Glycerol phenylbutyrate	2	SR	
372	434.1869	16.73	(+)	DCM/EA	-	-	-	-	Unknown	4	SR	
373	415.2114	16.74	(+)	DCM/EA	415.2115	0.09	C ₂₄ H ₃₀ O ₆	-	[(1S,8Z,12R,13R,14S)-12-Acetyloxy-4,9,13,17-tetramethyl-11-oxo-6-oxatricyclo[11.4.0.0.3,7]heptadeca-3(7),4,8,16-tetraen-14-yl] acetate	2	SR, SK	
374	311.1849	16.89	(-)	DCM/EA	311.1863	0.27	C ₁₇ H ₂₈ O ₅	-4.70	Unknown	4	SR, SK	
375	295.1912	16.89	(+)	DCM/EA	295.1903	0.15	C ₁₇ H ₂₆ O ₄	2.70	Sesquiterpenoid	3	SK	
376	437.1919	16.92	(+)	DCM/EA	-	-	C ₂₁ H ₂₈ N ₂ O ₈	0.10	Unknown	4	SR	
377	846.4422	17.08	(+)	DCM/EA	-	-	-	-	Lipid	3	SR	
378	269.1744	17.26	(+)	DCM/EA	-	-	C ₁₅ H ₂₄ O ₄	-1.30	Hymenoic acid	2	SR, SK	
379	415.2114	17.31	(+)	DCM/EA	415.2115	0.09	C ₂₄ H ₃₀ O ₆	-	Steroid	3	SR	
380	851.3958	17.34	(+)	DCM/EA	-	-	C ₆₃ H ₅₀ N ₂ O	-4.50	Unknown	4	SR	
381	432.2378	17.35	(+)	DCM/EA	432.2380	0.04	C ₂₄ H ₃₃ NO ₆	-	Steroid	3	SR	
382	493.2580	17.44	(+)	DCM/EA	-	-	C ₃₀ H ₃₆ O ₆	-1.00	Polyphenolic	3	SK	
383	293.1749	17.48	(-); (+)	DCM/EA	293.1758	0.20	C ₁₇ H ₂₆ O ₄	-3.10	3-dodecanoyl-4-hydroxy-5-methylidene-furan-2-one	2	SR, SK	
384	278.1546	17.51	(+)	DCM/EA	278.1539	0.08	C ₁₉ H ₁₉ NO	2.30	Quinoline alkaloid	3	SK	
385	322.1445	17.63	(+)	DCM/EA	-	-	C ₂₀ H ₁₉ NO ₃	2.20	Alkaloid	3	SR, SK	
386	277.1798	17.76	(-)	DCM/EA	277.1809	0.14	C ₁₇ H ₂₆ O ₃	-3.90	Fatty alcohol	3	SK	

387	492.3325	17.82	(+)	DCM/EA	-	-	C ₂₈ H ₄₅ NO ₆	1.00	Steroid	3	SK	
388	437.1921	18.17	(+)	DCM/EA	-	-	C ₂₁ H ₂₈ N ₂ O ₈	0.50	Unknown	4	SR	
389	415.2101	18.19	(+)	DCM/EA	415.2115	0.09	C ₂₄ H ₃₀ O ₆	-	Steroid	3	SR	
390	457.2214	18.42	(+)	DCM/EA	-	-	C ₂₆ H ₃₂ O ₇	-1.50	Steroid	3	SR	
391	322.1438	18.50	(+)	DCM/EA	-	-	C ₂₀ H ₁₉ NO ₃	0.00	2-[4-(3,4-Methylenedioxyphenyl)butyl]-4(1H)-quinolinone	2	SR	
392	426.1719	18.51	(-)	DCM/EA	-	-	-	-	Unknown	4	SR	
393	311.2218	18.60	(-)	DCM/EA	311.2228	-0.05	C ₁₈ H ₃₂ O ₄	-3.10	13-L-Hydroperoxylinoleic acid	2	SR	
394	333.1233	18.71	(+)	DCM/EA	333.1234	-0.08	C ₂₀ H ₁₆ N ₂ O ₃	0.00	Alkaloid	3	SR, SK	
395	295.2273	18.71	(-); (+)	DCM/EA	295.2279	-0.01	C ₁₈ H ₃₂ O ₃	-1.80	12-Hydroxy-8,10-octadecadienoic acid	2	SK	
396	279.2320	18.86	(+)	DCM/EA	-	-	C ₁₈ H ₃₀ O ₂	0.40	(E)-Octadec-9-en-6-ynoic acid	2	SK	
397	278.1535	19.00	(+)	DCM/EA	278.1539	0.08	C ₁₉ H ₁₉ NO	-1.70	Quinoline alkaloid	3	SK	
398	309.1713	19.04	(-)	DCM/EA; M	309.1707	0.15	C ₁₇ H ₂₆ O ₅	1.90	Terpenoid	3	SR, SK	
399	311.1679	19.05	(-)	DCM/EA; M	311.1686	0.17	C ₁₇ H ₂₈ O ₃ S	-2.30	N-Undecylbenzenesulfonic acid	2	SR, SK	
400	229.0988	19.07	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
401	415.2133	19.14	(+)	DCM/EA	415.2115	0.09	C ₂₄ H ₃₀ O ₆	-	Steroid	3	SR	
402	297.2415	19.17	(-)	DCM/EA	297.2435	-0.01	C ₁₈ H ₃₄ O ₃	-6.70	(15,11E)-1-hydroxyhexadec-11-en-1-yl acetate	2	SR	
403	575.3341	19.22	(+)	DCM/EA	575.3342	-0.04	C ₄₀ H ₄₆ OS	-0.20	Unknown	4	SR, SK	
404	149.0641	19.35	(+)	DCM/EA	149.0631	-0.02	C ₆ H ₁₂ O ₂ S	6.70	Sulfated ester	3	SR, SK	
405	752.5162	19.37	(+)	DCM/EA	-	-	C ₄₂ H ₇₄ NO ₈ P	-8.40	1-(9Z-Nonadecenoyl)-2-(6Z,9Z,12Z,15Z-octadecatetraenoyl)-glycero-3-phosphoethanolamine	2	SK	
406	708.4859	19.41	(+)	DCM/EA	-	-	-	-	Lipid	3	SK	
407	133.0691	19.45	(+)	DCM/EA	133.0682	-0.02	C ₆ H ₁₂ OS	6.90	Alkylthiol	3	SK	
408	576.4096	19.51	(+)	DCM/EA	-	-	C ₃₂ H ₅₁ N ₅ O ₅	0.10	Lipid	3	SK	
409	442.1997	19.52	(-)	DCM/EA	-	-	-	-	Lipid	3	SR	
410	444.3310	19.53	(+)	DCM/EA	444.3319	0.19	C ₂₄ H ₄₅ NO ₆	-2.20	Unknown	4	SK	
411	356.2599	19.53	(+)	DCM/EA	-	-	C ₂₃ H ₃₃ NO ₂	4.10	Alkaloid	3	SK	
412	322.1438	19.53	(+)	DCM/EA	-	-	C ₂₀ H ₁₉ NO ₃	0.00	Alkaloid	3	SR, SK	
413	296.1515	19.55	(+)	DCM/EA	-	-	-	-	Unknown	4	SK	
414	488.3595	19.55	(+)	DCM/EA	488.3582	-0.09	C ₂₆ H ₄₉ NO ₇	2.70	Unknown	4	SK	
415	312.3622	19.63	(+)	DCM/EA	-	-	C ₂₁ H ₄₅ N	-1.00	Amine	3	SK	
416	313.2375	19.64	(-)	DCM/EA	313.2384	0.12	C ₁₈ H ₃₄ O ₄	-2.90	Fatty acid	3	SR	
417	269.2117	19.66	(-)	DCM/EA	269.2122	0.08	C ₁₆ H ₃₀ O ₃	-1.80	Oxohexadecanoic acid	2	SR, SK	
418	416.3729	19.77	(-)	DCM/EA	-	-	C ₂₄ H ₅₁ NO ₄	-3.90	Fatty alcohol	3	SR, SK	
419	281.1730	19.78	(+)	DCM/EA	-	-	C ₁₂ H ₂₀ N ₆ O ₂	3.30	Unknown	4	SK	
420	259.1911	19.79	(+)	DCM/EA	259.1903	0.38	C ₁₄ H ₂₆ O ₄	2.70	Bis(2-methylpropyl) hexanedioate	2	SK	
421	450.2632	19.81	(+)	DCM/EA	-	-	C ₂₈ H ₃₅ NO ₄	-1.60	7-Deoxy-cytochalasin Z9	2	SK	
422	185.1165	19.85	(+)	DCM/EA	-	-	C ₁₀ H ₁₆ O ₃	-4.00	Unknown	4	SK	
423	269.2476	19.86	(+)	DCM/EA; M	269.2475	0.05	C ₁₇ H ₃₂ O ₂	0.30	Mono-unsaturated fatty acid	3	SR, SK	
424	263.1272	19.97	(+)	DCM/EA	-	-	-	-	Unknown	4	SR	
425	397.2008	19.98	(+)	DCM/EA	397.2009	0.00	C ₂₄ H ₂₈ O ₅	-0.40	4-Methoxy-2-(6-methoxy-2,2-dimethylchromen-8-yl)oxy-6-(3-methylbut-2-enyl)phenol	2	SR	
426	355.2838	20.22	(+)	DCM/EA	-	-	C ₂₁ H ₃₈ O ₄	-1.40	Glycerolipid	3	SR, SK	
427	299.2580	20.35	(-)	DCM/EA	299.2592	-0.05	C ₁₈ H ₃₆ O ₃	-3.80	13-Hydroxyoctadecanoic acid	2	SR, SK	
428	448.2470	20.46	(+)	DCM/EA	-	-	C ₂₈ H ₃₃ NO ₄	-2.80	Steroid	3	SK	
429	279.2318	20.48	(+)	DCM/EA	-	-	C ₁₈ H ₃₀ O ₂	-0.30	(11E,13E,15Z)-Octadeca-11,13,15-trienoic acid	2	SR, SK	
430	722.5033	20.54	(+)	DCM/EA	722.5049	-0.01	C ₃₇ H ₇₁ NO ₁₂	-2.30	Unknown	4	SR	

431	766.5309	20.55	(+)	DCM/EA	766.5312	-0.05	C ₃₉ H ₇₅ NO ₁₃	-0.30	Unknown	4	SR	
432	375.3043	20.55	(+)	DCM/EA	-	-	C ₂₈ H ₃₈	-0.90	Steroid	3	SR, SK	
433	678.4767	20.56	(+)	DCM/EA	678.4788	-0.09	C ₃₅ H ₆₇ NO ₁₁	-	Lipid	3	SR, SK	
434	491.3380	20.58	(-)	DCM/EA	-	-	C ₂₉ H ₄₈ O ₆	0.40	Lipid	3	SR, SK	
435	342.2440	20.62	(+)	DCM/EA	-	-	C ₂₂ H ₃₁ NO ₂	3.60	Steroid	3	SK	
436	299.2581	20.63	(-)	DCM/EA	299.2592	-0.05	C ₁₈ H ₃₆ O ₃	-3.50	10-Hydroxyoctadecanoic acid	2	SR, SK	
437	458.3482	20.73	(+)	DCM/EA	458.3476	0.06	C ₂₅ H ₄₇ NO ₆	1.20	17-carboxyheptadecanoylcarnitine	2	SR, SK	
438	507.3311	20.74	(+)	DCM/EA	-	-	C ₂₉ H ₄₆ O ₇	-1.10	Steroid	3	SR	
439	414.3206	20.76	(+)	DCM/EA	214.3214	0.13	C ₂₃ H ₄₃ NO ₅	-	3-Hydroxy-9-hexadecenoylcarnitine	2	SR, SK	
440	463.2999	20.79	(+)	DCM/EA	-	-	-	-	Unknown	4	SR	
441	370.2958	20.80	(+)	DCM/EA	370.2951	0.13	C ₂₁ H ₃₉ NO ₄	1.60	Fatty amide	3	SR, SK	
442	441.3204	20.80	(+)	DCM/EA	-	-	C ₂₅ H ₄₄ O ₆	-1.60	Blinoside B	2	SR	
443	369.1247	21.04	(+)	DCM/EA	-	-	C ₂₃ H ₁₆ N ₂ O ₃	3.60	Alkaloid	3	SR, SK	
444	269.2123	21.11	(-)	DCM/EA	269.2122	0.08	C ₁₆ H ₃₀ O ₃	0.40	Fatty acid	3	SR	
445	434.2693	21.19	(+)	DCM/EA	-	-	C ₂₈ H ₃₅ NO ₃	0.70	Cytochalasin J3	2	SK	
446	339.2901	21.39	(+)	DCM/EA	339.2893	0.16	C ₂₁ H ₃₈ O ₃	2.10	3-Hexadecyl-5-methyl-5-hydroxyfuran-2(5H)-one	2	SK	
447	403.2326	21.68	(+)	DCM/EA	403.2326	0.07	C ₂₀ H ₃₄ O ₈	-0.20	(E,5S)-5-hydroxy-7-[(1S,2S,5S,6R)-5-[(3S)-3-hydroxyoctyl]-3,4,7,8-tetraoxabicyclo[4.2.1]nonan-2-yl]hept-6-enoic acid	2	SR, SK	
448	425.2147	21.69	(+)	DCM/EA	425.2145	-0.12	C ₂₆ H ₃₂ O ₃ S	0.40	Unknown	4	SR	
449	297.2422	21.69	(+)	DCM/EA	297.2424	0.22	C ₁₈ H ₃₂ O ₃	-0.80	5-[(E)-7-Hydroxytetradec-8-enyl]oxolan-2-one	2	SR, SK	
450	299.2585	21.85	(-)	DCM/EA	299.2592	-0.05	C ₁₈ H ₃₆ O ₃	-2.10	7-Hydroxyoctadecanoic acid	2	SR, SK	
451	339.2907	21.89	(-)	DCM/EA	-	-	C ₂₁ H ₄₀ O ₃	0.70	Methyl 11-[(2R,3S,4S,5R)-3,4-dimethyl-5-propyloxolan-2-yl]undecanoate	2	SK	
452	375.3028	21.92	(+)	DCM/EA	-	-	C ₂₈ H ₃₈	-4.90	Ergosta-2,5,7,9(11),14,22-hexaene	2	SR, SK	
453	491.3383	21.96	(-)	DCM/EA	-	-	C ₂₉ H ₄₈ O ₆	1.00	Lipid	3	SR, SK	
454	427.3184	22.00	(+)	DCM/EA	-	-	C ₂₈ H ₄₂ O ₃	-5.40	Secocitreoanthrasteroid	2	SR, SK	
455	695.4121	22.09	(+)	DCM/EA	695.4129	-0.09	C ₄₅ H ₅₈ O ₄ S	-1.10	Unknown	4	SR, SK	
456	395.2437	22.09	(-)	DCM/EA	395.2440	-0.16	C ₂₂ H ₃₆ O ₆	-0.50	Lipid	3	SR, SK	
457	690.4583	22.12	(+)	DCM/EA	-	-	C ₃₉ H ₆₃ NO ₉	1.00	Unknown	4	SR, SK	
458	351.3253	22.29	(+)	DCM/EA	-	-	C ₂₃ H ₄₂ O ₂	-1.40	(14Z,17Z)-Tricoso-14,17-dienoic acid	2	SK	
459	368.2583	22.41	(+)	DCM/EA	-	-	C ₂₄ H ₃₃ NO ₂	-0.40	3-[(1S)-2-(1H-Indole-3-ylmethyl)-2beta,3alpha-dimethyl-6-isopropylidene-cyclohexane-1beta-yl]propionic acid methyl ester	2	SK	
460	409.3107	22.46	(+)	DCM/EA	-	-	C ₂₈ H ₄₀ O ₂	1.40	Ganodosterone	2	SR, SK	
461	706.4529	22.47	(+)	DCM/EA	-	-	C ₃₉ H ₆₃ NO ₁₀	0.60	beta-1-Chaconine	2	SR, SK	
462	443.3146	22.48	(-); (+)	DCM/EA	-	-	C ₂₈ H ₄₄ O ₄	-4.60	Steroid	3	SR, SK	
463	711.4069	22.50	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
464	489.3205	22.50	(-)	DCM/EA	-	-	C ₂₉ H ₄₆ O ₆	-3.30	Lipid	3	SR, SK	
465	297.2419	22.52	(-)	DCM/EA	297.2435	-0.01	C ₁₈ H ₃₄ O ₃	-5.30	2-Oxo-octadecanoic acid	2	SR, SK	
466	331.2862	22.64	(+)	DCM/EA	-	-	C ₁₉ H ₃₈ O ₄	5.70	Glycerolipid	3	SR, SK	
467	299.2208	22.65	(-)	DCM/EA	299.2228	-0.04	C ₁₇ H ₃₂ O ₄	-6.50	Fatty acid	3	SR, SK	
468	785.6318	22.65	(-)	DCM/EA	-	-	C ₄₉ H ₈₆ O ₇	2.20	Lipid	3	SR, SK	
469	455.3520	22.66	(-); (+)	DCM/EA	455.3531	-0.04	C ₃₀ H ₄₈ O ₃	-2.30	2-(7-Hydroxy-5a,5b,8,8,11a,13b-hexamethyl-1,2,3,3a,4,5,6,7,7a,9,10,11,11b,12,13,13a-hexadecahydrocyclopenta[a]chrysen-3-yl)prop-2-enoic acid	2	SR, SK	
470	253.2163	22.66	(-)	DCM/EA	253.2173	0.21	C ₁₆ H ₃₀ O ₂	-3.90	Palmitoleic acid	2	SR, SK	Yes (level 2)
471	311.2944	22.67	(+)	DCM/EA	311.2945	0.00	C ₂₀ H ₃₈ O ₂	-0.30	Hexadecenyl butyrate	2	SR, SK	

472	313.2727	22.68	(+)	DCM/EA	313.2737	-0.01	C ₁₉ H ₃₆ O ₃	-	Glycidyl Palmitate	2	SR, SK
473	323.2182	22.68	(+)	DCM/EA	323.2193	-0.83	C ₁₅ H ₂₆ N ₆ O ₂	-2.60	Unknown	4	SR, SK
474	353.1416	22.69	(-)	DCM/EA	-	-	-	-	Unknown	4	SR, SK
475	393.3156	22.72	(+)	DCM/EA	-	-	C ₂₈ H ₄₀ O	1.00	Steroid	3	SR, SK
476	256.2631	22.73	(+)	DCM/EA	256.2635	0.07	C ₁₆ H ₃₃ NO	-1.60	Hexadecanamide	2	SR, SK
477	395.3297	22.76	(+)	DCM/EA	-	-	C ₂₈ H ₄₂ O	-3.00	(3beta,22E,24R)-Ergosta-4,6,8(14),22-tetraen-3-ol	2	SR, SK
478	491.3383	22.81	(-)	DCM/EA	-	-	C ₂₉ H ₄₈ O ₆	1.00	Lipid	3	SR, SK
479	695.4121	22.83	(+)	DCM/EA	695.4129	-0.09	C ₄₅ H ₅₈ O ₄ S	-1.10	Lipid	3	SR, SK
480	690.4565	22.83	(+)	DCM/EA	-	-	C ₃₉ H ₆₃ NO ₉	-1.60	Unknown	4	SR, SK
481	429.3358	22.85	(+)	DCM/EA	-	-	C ₂₈ H ₄₄ O ₃	-1.30	25-hydroxy-1alpha-hydroxymethyl-16,17-didehydrocholecalciferol	2	SR, SK
482	671.4138	22.88	(-)	DCM/EA	-	-	-	-	Unknown	4	SR, SK
483	375.3055	22.91	(+)	DCM/EA	-	-	C ₂₈ H ₃₈	2.30	Steroid	3	SR
484	434.2684	22.94	(+); (-)	DCM/EA	-	-	C ₂₈ H ₃₅ NO ₃	-1.40	Cytochalasin J2	2	SK
485	683.5439	22.98	(+)	DCM/EA	683.5432	-0.13	C ₄₄ H ₇₄ O ₃ S	1.10	Unknown	4	SR, SK
486	515.3948	22.99	(+)	DCM/EA	-	-	-	-	Lipid	3	SR, SK
487	524.3817	23.01	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK
488	331.2844	23.01	(+)	DCM/EA; M	331.2843	0.00	C ₁₉ H ₃₈ O ₄	0.30	2-Palmitoylglycerol	2	SR, SK
489	459.3084	23.04	(+)	DCM/EA	-	-	C ₂₈ H ₄₂ O ₅	-4.90	Chaxine B	2	SR, SK
490	303.2314	23.05	(-)	DCM/EA	-	-	C ₂₀ H ₃₂ O ₂	-5.00	Diterpenoid	3	SR, SK
491	297.2428	23.14	(-)	DCM/EA	297.2435	0.11	C ₁₈ H ₃₄ O ₃	-2.30	10-Oxo-octadecanoic acid	2	SR, SK
492	481.2950	23.14	(+)	DCM/EA	-	-	C ₂₅ H ₄₀ N ₂ O ₇	8.60	Amino-acid	3	SR, SK
493	333.2418	23.22	(+)	DCM/EA	-	-	C ₂₁ H ₃₂ O ₃	-1.90	Incisterol	2	SR, SK
494	299.2587	23.24	(-)	DCM/EA	299.2592	-0.05	C ₁₈ H ₃₆ O ₃	-1.50	5-Hydroxyoctadecanoic acid	2	SR, SK
495	727.4402	23.26	(-)	DCM/EA	-	-	-	-	Unknown	4	SR
496	746.4819	23.26	(+)	DCM/EA	-	-	-	-	Unknown	4	SR
497	411.3270	23.28	(+)	DCM/EA	-	-	C ₂₈ H ₄₂ O ₂	3.00	(6alpha,22E)-6-Hydroxy-4,7,22-ergostatrien-3-one	2	SR, SK
498	375.3057	23.31	(+)	DCM/EA	-	-	C ₂₈ H ₃₈	2.80	Steroid	3	SR, SK
499	393.3150	23.31	(+)	DCM/EA	-	-	C ₂₈ H ₄₀ O	-0.60	Steroid	3	SR
500	653.4052	23.41	(-); (+)	DCM/EA	-	-	-	-	Steroid	3	SR, SK
501	706.4529	23.46	(+)	DCM/EA	-	-	C ₃₉ H ₆₃ NO ₁₀	0.60	Steroid	3	SR, SK
502	1078.5662	23.48	(+)	DCM/EA	-	-	-	-	Unknown	4	SR
503	1080.5735	23.48	(+)	DCM/EA	-	-	-	-	Unknown	4	SR
504	701.4257	23.50	(-)	DCM/EA	-	-	-	-	Steroid	3	SR, SK
505	464.3580	23.50	(+)	DCM/EA	-	-	C ₂₄ H ₄₉ NO ₇	-0.40	Glycosylsphingoid	3	SK
506	548.3031	23.52	(+)	DCM/EA	-	-	C ₃₄ H ₃₇ N ₅ O ₂	2.00	Ester	3	SR
507	679.4179	23.53	(+)	DCM/EA	679.4181	-0.47	C ₃₇ H ₅₅ N ₆ O ₆	-	Lipid	3	SR, SK
508	531.2757	23.53	(+)	DCM/EA	-	-	C ₃₃ H ₃₈ O ₆	2.90	Polyphenolic	3	SR
509	655.4182	23.54	(-)	DCM/EA	-	-	-	-	Steroid	3	SR, SK
510	377.3209	23.62	(+)	DCM/EA	-	-	C ₂₈ H ₄₀	1.60	Steroid	3	SR
511	695.4062	23.62	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK
512	367.1570	23.63	(-)	DCM/EA	-	-	C ₂₃ H ₂₀ N ₄ O	1.60	Alkaloid	3	SR, SK
513	333.2285	23.64	(-)	DCM/EA	-	-	C ₁₇ H ₃₄ O ₆	0.80	Lipid	3	SR
514	267.2318	23.64	(-)	DCM/EA	267.2329	0.19	C ₁₇ H ₃₂ O ₂	-4.20	Mono-unsaturated fatty acid	3	SR, SK
515	313.2372	23.64	(-)	DCM/EA	313.2384	0.12	C ₁₈ H ₃₄ O ₄	-3.80	(8E,10S)-10-hydroperoxy-8-octadecenoic acid	2	SK

516	429.3361	23.65	(+)	DCM/EA	-	-	C ₂₈ H ₄₄ O ₃	-0.60	(22E)-1alpha,25-dihydroxy-26-methyl-22,23-didehydrocholecalciferol	2	SR, SK	
517	690.4577	23.66	(+)	DCM/EA	-	-	C ₃₉ H ₆₃ NO ₉	0.20	Unknown	4	SR, SK	
518	347.3166	23.68	(+)	DCM/EA	347.3156	0.00	C ₂₀ H ₄₂ O ₄	2.90	1,2,3,4-Eicosanetetrol	2	SR, SK	
519	591.3510	23.70	(+)	DCM/EA	591.3503	-0.12	C ₃₇ H ₅₀ O ₄ S	1.20	Steroid	3	SR, SK	
520	586.3964	23.75	(+)	DCM/EA	-	-	C ₃₁ H ₅₅ NO ₉	2.40	3-[1-Hydroxy-2-methyl-18-[(2R,3S,4S,5S,6R)-3,4,5-trihydroxy-6-(hydroxymethyl)oxan-2-yl]oxyoctadecylidene]-1,5-dimethylpyrrolidine-2,4-dione	2	SR, SK	
521	445.3323	23.86	(+)	DCM/EA	-	-	C ₂₈ H ₄₄ O ₄	2.30	(25S)-1alpha,25,26-trihydroxyergocalciferol	2	SR, SK	
522	381.3373	23.89	(+)	DCM/EA	-	-	C ₂₄ H ₄₄ O ₃	2.50	Lipid	3	SK	
523	299.2587	23.92	(-)	DCM/EA	299.2592	-0.05	C ₁₈ H ₃₆ O ₃	-1.50	2-Hydroxyoctadecanoic acid	2	SK	
524	531.2762	24.00	(+)	DCM/EA	-	-	C ₃₃ H ₃₈ O ₆	3.90	Polyphenolic	3	SR	
525	455.3359	24.00	(+)	DCM/EA	455.3343	-0.14	C ₃₀ H ₄₆ OS	3.70	Unknown	4	SR, SK	
526	393.3157	24.09	(+)	DCM/EA	-	-	C ₂₈ H ₄₀ O	1.20	Citreoanthrasteroid B	2	SR, SK	
527	427.3216	24.10	(+)	DCM/EA	-	-	C ₂₈ H ₄₂ O ₃	2.10	(3beta,5alpha,6alpha,7beta,22E,24R)-5,6-Epoxyergosta-8(14),22-diene-3,7-diol	2	SR, SK	
528	432.2544	24.11	(+)	DCM/EA	-	-	C ₂₈ H ₃₃ NO ₃	2.40	Ketocytocalasin	2	SK	
529	461.3598	24.11	(+)	DCM/EA	-	-	C ₂₉ H ₄₈ O ₄	-6.00	1alpha-hydroxy-18-(5-hydroxy-5-methylhexyloxy)-23,24,25,26,27-pentanorcholecalciferol	2	SK	
530	449.3046	24.14	(+)	DCM/EA	-	-	-	-	Lipid	3	SR, SK	
531	437.4344	24.16	(+)	DCM/EA	-	-	C ₂₉ H ₅₆ O ₂	-2.10	Beta-diketone	3	SK	
532	381.3372	24.18	(-)	DCM/EA	-	-	C ₂₄ H ₄₆ O ₃	-0.50	Fatty alcohol	3	SK	
533	313.2372	24.19	(-)	DCM/EA	313.2384	0.12	C ₁₈ H ₃₄ O ₄	-3.80	Fatty acid	3	SR, SK	
534	359.3161	24.25	(+)	DCM/EA	359.3156	0.01	C ₂₁ H ₄₂ O ₄	1.40	Glycerolipid	3	SK	
535	341.3063	24.26	(+)	DCM/EA	341.3050	-0.01	C ₂₁ H ₄₀ O ₃	3.40	Glycidyl stearate	2	SK	
536	603.3870	24.36	(-)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
537	377.3202	24.40	(+)	DCM/EA	-	-	C ₂₈ H ₄₀	-0.30	Steroid	3	SR, SK	
538	375.3040	24.40	(+)	DCM/EA	-	-	C ₂₈ H ₃₈	-1.70	Steroid	3	SR	
539	411.3226	24.42	(+); (-)	DCM/EA	411.3216	0.23	C ₂₃ H ₄₂ N ₂ O ₄	2.10	Amino-acid	3	SR, SK	
540	458.3053	24.45	(+)	DCM/EA	-	-	C ₃₁ H ₃₉ NO ₂	-0.20	Lipid	3	SK	
541	309.2425	24.53	(-)	DCM/EA	-	-	C ₁₉ H ₃₄ O ₃	-3.20	Fatty alcohol	3	SR, SK	
542	381.1727	24.53	(-)	DCM/EA	-	-	C ₂₄ H ₂₂ N ₄ O	1.70	Alkaloid	3	SR, SK	
543	491.3394	24.54	(-)	DCM/EA	-	-	C ₂₉ H ₄₈ O ₆	3.30	Lipid	3	SK	
544	281.2481	24.56	(-); (+)	DCM/EA	281.2486	0.16	C ₁₈ H ₃₄ O ₂	-1.70	Mono-unsaturated fatty acid	3	SR, SK	
545	313.2377	24.56	(-)	DCM/EA	313.2384	0.12	C ₁₈ H ₃₄ O ₄	-2.30	Fatty acid	3	SR, SK	
546	327.2524	24.56	(-)	DCM/EA	327.2541	0.01	C ₁₉ H ₃₆ O ₄	-5.10	(9R,12S,15E)-9,12-dihydroxynonadec-15-enoic acid	2	SR, SK	
547	395.2767	24.63	(+)	DCM/EA	395.2768	-0.16	C ₂₇ H ₃₈ S	-0.10	Sulfated carbohydrate	3	SR, SK	
548	531.2747	24.63	(+)	DCM/EA	-	-	C ₃₃ H ₃₈ O ₆	1.10	Polyphenolic	3	SR	
549	608.4883	24.65	(+)	M	-	-	-	-	Unknown	4	SK	
550	475.3456	24.66	(+)	DCM/EA	-	-	-	-	Lipid	3	SK	
551	355.2848	24.68	(+)	DCM/EA	-	-	C ₂₁ H ₃₈ O ₄	1.40	Glycerolipid	3	SR	
552	433.3639	24.69	(+)	DCM/EA	433.3651	-0.41	C ₂₄ H ₄₄ N ₆ O	-2.50	Unknown	4	SK	
553	393.3713	24.70	(+)	DCM/EA	-	-	C ₂₆ H ₄₈ O ₂	-3.60	15,19-Hexacosadienoic acid	2	SK	
554	451.3221	24.70	(+)	DCM/EA	-	-	C ₃₀ H ₄₂ O ₃	3.10	Steroid	3	SR	
555	429.3348	24.73	(+)	DCM/EA	-	-	C ₂₈ H ₄₄ O ₃	-3.60	Steroid	3	SR	
556	310.3116	24.74	(+)	DCM/EA	310.3104	-0.05	C ₂₀ H ₃₉ NO	3.70	1-Pyrrolidin-1-ylhexadecan-1-one	2	SR, SK	
557	411.3248	24.74	(+)	DCM/EA	-	-	C ₂₈ H ₄₂ O ₂	-2.40	Steroid	3	SR, SK	

558	427.3200	24.75	(+)	DCM/EA	-	-	C ₂₈ H ₄₂ O ₃	-1.60	Lipid	3	SR, SK	
559	964.6442	24.85	(+)	DCM/EA	-	-	C ₅₆ H ₈₉ N ₃ O ₈ S	-0.10	Unknown	4	SR	
560	969.6000	24.85	(+)	DCM/EA	-	-	C ₆₂ H ₇₆ N ₆ O ₄	-0.10	Unknown	4	SR	
561	474.3138	24.86	(+)	DCM/EA	-	-	C ₂₇ H ₄₃ N ₃ O ₂ S	-2.30	Unknown	4	SR	
562	491.3380	24.88	(+)	DCM/EA	-	-	C ₃₀ H ₄₂ N ₄ O ₂	-0.20	Alkaloid	3	SR	
563	437.4363	24.88	(+)	DCM/EA	-	-	C ₂₉ H ₅₆ O ₂	2.20	Beta-diketone	3	SK	
564	353.3415	24.89	(+)	DCM/EA	353.3415	-0.23	C ₂₃ H ₄₄ O ₂	0.20	Fatty acid	3	SK	
565	359.3161	24.91	(+)	DCM/EA	359.3156	0.01	C ₂₁ H ₄₂ O ₄	1.40	2-(1-Hydroxyheptadecyl)-1,3-dioxolane-4-methanol	2	SK	
566	284.2959	24.93	(+)	DCM/EA	284.2948	-0.09	C ₁₈ H ₃₇ NO	3.80	Octadecanamide	2	SR, SK	
567	281.2458	24.95	(-)	DCM/EA	-	-	C ₁₄ H ₃₀ N ₆	-0.30	Unknown	4	SR, SK	
568	730.4889	25.05	(+)	DCM/EA	-	-	-	-	Unknown	4	SR	
569	463.3039	25.06	(+)	DCM/EA	-	-	C ₂₇ H ₄₂ O ₆	-3.30	Steroid	3	SR, SK	
570	506.3386	25.35	(+)	DCM/EA	-	-	C ₃₁ H ₄₃ N ₃ O ₃	1.70	Unknown	4	SR, SK	
571	339.2892	25.37	(+)	DCM/EA	339.2893	0.16	C ₂₁ H ₃₈ O ₃	-0.60	5-(8-Heptadecenyl)-3-hydroxydihydro-2(3H)-furanone	2	SK	
572	619.3801	25.38	(+)	DCM/EA	-	-	-	-	Lipid	3	SR, SK	
573	414.3452	25.41	(+)	DCM/EA	-	-	-	-	Unknown	4	SR	
574	614.4294	25.46	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
575	411.3249	25.51	(+)	DCM/EA	-	-	C ₂₈ H ₄₂ O ₂	-2.10	(3beta,22E,24R)-3-Hydroxyergosta-5,8,22-trien-7-one	2	SR, SK	
576	624.4832	25.70	(+)	DCM/EA; M	624.4831	0.45	C ₃₆ H ₆₅ NO ₇	-0.30	Lipid	3	SK	
577	629.4384	25.74	(+)	M	-	-	-	-	Unknown	4	SK	
578	423.3830	25.82	(+)	DCM/EA	-	-	C ₂₇ H ₅₀ O ₃	-0.70	(5Z,9Z)-2-Methoxy-5,9-hexacosadienoic acid	2	SK	
579	375.3443	26.03	(+)	DCM/EA	-	-	C ₁₈ H ₄₂ N ₆ O ₂	0.20	Unknown	4	SR, SK	
580	523.3612	26.03	(+)	DCM/EA	523.3604	-0.43	C ₂₆ H ₄₆ N ₆ O ₅	-	Unknown	4	SK	
581	413.3401	26.09	(+)	DCM/EA	-	-	C ₂₈ H ₄₄ O ₂	-3.20	Steroid	3	SR, SK	
582	673.4600	26.10	(+)	M	-	-	-	-	Unknown	4	SR, SK	
583	668.5054	26.12	(+)	M	-	-	-	-	Unknown	4	SK	
584	427.3788	26.16	(+)	M	427.3781	0.15	C ₂₆ H ₅₀ O ₄	1.40	Fatty ester	3	SK	
585	444.4044	26.17	(+)	M	-	-	C ₂₆ H ₅₃ NO ₄	-0.80	N-Octanoylphyto-sphingosine	2	SK	
586	481.4239	26.37	(+)	DCM/EA	-	-	C ₃₀ H ₅₆ O ₄	-2.60	Fatty acid	3	SK	
587	591.3250	26.78	(+)	DCM/EA	-	-	C ₂₇ H ₄₂ N ₈ O ₇	0.10	Amino-acid	3	SR	
588	395.3882	26.88	(+)	DCM/EA	395.3883	0.07	C ₂₆ H ₅₀ O ₂	-0.50	Tetrahydro-6-henicosyl-2H-pyran-2-one	2	SK	
589	652.5143	26.96	(+)	M	-	-	-	-	Unknown	4	SK	
590	473.3179	26.99	(+)	DCM/EA	-	-	-	-	Unknown	4	SR	
591	586.4383	27.11	(+)	M	-	-	-	-	Unknown	4	SR	
592	457.2850	27.16	(-)	DCM/EA	-	-	C ₃₀ H ₃₈ N ₂ O ₂	-2.20	Alkaloid	3	SK	
593	939.5782	27.17	(+)	DCM/EA	-	-	C ₄₉ H ₈₂ N ₂ O ₁₅	-0.70	Lactone polyketide	4	SK	
594	936.6298	27.20	(+)	DCM/EA	-	-	-	-	Unknown	4	SK	
595	950.6357	27.23	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
596	465.4265	27.43	(+)	DCM/EA	-	-	-	-	Lipid	3	SK	
597	698.4977	27.82	(+)	M	698.4979	0.16	C ₄₂ H ₆₇ NO ₇	-	Steroid + Amino-acid	3	SR	
598	502.3452	27.88	(+)	DCM/EA	-	-	-	-	Unknown	4	SR, SK	
599	503.3368	27.89	(+)	DCM/EA	-	-	C ₃₀ H ₄₆ O ₆	0.10	[5,6-dihydroxy-7-(hydroxymethyl)-3,11,11,14-tetramethyl-15-oxo-4-tetracyclo[7.5.1.01,5.010,12]pentadeca-2,7-dienyl] decanoate	2	SR	

Résumé

Cette thèse s'intitule « Identification et cartographie des métabolites secondaires de nouvelles souches de levures tropicales ». L'objectif principal est le développement d'une méthodologie de déréplication pour l'exploration des métabolites secondaires extraits de cultures de champignons unicellulaires. La méthode développée fait intervenir la chromatographie liquide à ultra-haute performance couplée à la spectrométrie de masse avec l'outil des réseaux moléculaires en complémentarité de la spectrométrie de masse à ultra-haute résolution. La méthodologie permet d'annoter les métabolites détectés par le procédé analytique, c'est-à-dire formuler des hypothèses sur l'identité et la structure de nombreuses molécules produites par les nouvelles espèces de levures sur lesquelles elle est appliquée. De plus, des cartographies et profils métaboliques sont dressés grâce aux diagrammes et réseaux moléculaires construits. L'application de la méthodologie à de nouvelles souches de levures a permis l'annotation de plusieurs centaines de métabolites, parmi lesquels une large proportion d'alkaloïdes mais aussi des acides aminés, des terpénoïdes et des stérols.

Mots-clés : déréplication ; métabolomique non ciblée ; UHPLC-MS/MS ; réseaux moléculaires ; FTICR-MS.

Summary

This PhD thesis project is named "Identification and cartography of secondary metabolites from new tropical yeast strains". The main objective is the development of a dereplication methodology and the application of it to the exploration of the secondary metabolites from unicellular fungal extracts. The developed method implies ultra-high-performance liquid chromatography hyphenated with mass spectrometry with the use of the molecular networking tool in combination with the ultra-high-resolution mass spectrometry. The methodology allows the annotation of the detected metabolites from acquired data, that is to say the formulation of hypotheses about the structure and identity of the numerous molecules produced by the analyzed yeast. Moreover, cartography and metabolic profiles are provided thanks to the built diagrams and molecular networks. The application of the methodology to new yeast strains led to the annotation of hundreds of metabolites, with a rich content in alkaloids, but also in amino-acids, terpenoids and sterols.

Keywords: dereplication; untargeted metabolomics; UHPLC-MS/MS; molecular networking; FTICR-MS.