université PARIS-SACLAY

NNT : 2018SACLE025

Thèse de doctorat

Exploration de la face cachée du métabolisme bactérien chez *Acinetobacter baylyi* ADP1

Thèse de doctorat de l'Université Paris-Saclay,
préparée à l'Université d'Evry Val d'Essonne

École doctorale n°577
Structure et dynamique des systèmes vivants (SDSV)

Spécialité de doctorat
Science de la Vie et de la Santé – Chimie analytique

Thèse présentée et soutenue à Évry, le 04 Octobre 2018, par

Marion THOMAS

Composition du Jury :

Florence GONNET Professeur, Université d'Évry Val d'Essonne	Président du Jury
Carlos AFONSO Professeur, Université de Rouen	Rapporteur
Olivier LEQUIN Professeur, Sorbonne Université	Rapporteur
Anne Emmanuelle HAY DE BETTIGNIES Maître de conférences, Université Claude Bernard (Lyon)	Examinateur
Alain PERRET Chargé de recherches, CEA- Genoscope	Co-directeur de thèse
Marcel SALANOUBAT Directeur UMR 8030, CEA- Genoscope	Co-directeur de thèse

Remerciements

Au terme de ce parcours, c'est à mes deux directeurs de thèse que j'adresse mes premiers remerciements. Marcel, Alain, un grand merci pour votre bienveillance et votre soutien indéféctible, tout au long de ces trois ans et surtout pendant la rédaction de ce manuscrit. Vos conseils et encouragements m'ont accompagnés durant tout ce temps.

Je remercie vivement les membres de mon jury de thèse : Mme Florence GONNET, Professeure à l'Université d'Évry Val d'Essonne, Mme Anne-Emmanuelle HAY-DE BETTIGNIES, Maître de conférences à l'Université Claude Bernard de Lyon ainsi qu'à mes deux rapporteurs, Mr Olivier LEQUIN, Professeur à Sorbonne Université, et Mr Carlos AFONSO Professeur à l'Université de Rouen d'avoir accepté d'étudier et discuter de mon travail.

Merci à Jean Weissenbach, de m'avoir acceptée au Genoscope et de m'avoir permis de réaliser mon doctorat au sein du Laboratoire de Génomique et Biochimie du Métabolisme.

Un grand merci à Katya et Pierre Loïc : Katya, pour ta précision, ta douceur et ta gentillesse, Pierre-Loïc pour ta rigueur et ta motivation. Vous m'avez également tout deux beaucoup appris sur la spectrométrie de masse et la RMN.

J'aimerais adresser mes remerciements au Professeur Jean Claude TABET, avec qui j'ai eu l'occasion d'échanger lors du SMMAP et de l'IMSC.

Tout ce travail n'aurait pas pu être réalisé et présenté ici sans l'aide de Christophe et d'Émilie. Merci à Christophe pour tous ces moments où nous avons refait le monde, au soleil, café et thé en main. Grâce à tes conseils et tes encouragments, tu as été un soutien moral et professionnel tout au long de ces trois années. Émilie, ta rigueur et ta volonté sont précieuses ; elles ont été sans conteste un atout à ce projet. Mais surtout, merci pour ton sourire, ta joie de vivre et ta curiosité. Ton enthousiasme m'a fait beaucoup de bien. Une petite pensée pour Anthony, notre stagiaire sur le projet, qui s'est investi et dont la gentillesse et le répertoire musical improbable m'ont beaucoup marquée.

Cette thèse et surtout la rédaction du manuscrit n'aurait pas été la même sans Thomas et Ombeline. On galère ensemble, on s'énerve ensemble, on panique ensemble, mais on va y arriver ! Bon courage à Sophia, Chloé, et Oriane, que nous avons probablement un peu effrayées ces dernières semaines de rédaction… Mais vous le savez, la récompense est belle ! Merci également aux anciens : Tiffany, Tristan et Marion ainsi qu'à William, pour tous ces moments partagés.

Une pensée à mes acolytes de repas Loïc, notre ninja, Laurence et Nath' ainsi qu'à mes anciens partenaires de bureau : Nathalie, Peggy, Delphine et Christine, Jean-Louis, Pascal, Virginie, Adrien,

Aline. Merci à toutes les personnes du premier étage et à Catherine, pour votre accueil, votre bienveillance et votre bonne humeur quotidienne, elles sont un soutien précieux.

Merci à Pierre, Laurent, Pierre-Olivier, Soline, Martin, Imane, Ryma, Axel, Safae, Éric, Perrinouchette, Anne-Laure, Isa, Claire, Olga, mes amis du conservatoire : Thibaud, Benoit, Lise et Camille, Amandine, la famille Rousseau, les Gaudy (la Creuse, c'est la vie, on est d'accord)... Vous êtes toujours là même si je suis un peu frapadingue... J'ai également une pensée toute particulière pour Étienne, Monique et Mr Cholleton qui ont eu la patience de m'enseigner la musique toutes ces années. Ca me manque sérieusement…

Et enfin, un GRAND merci (certains sont loin) à toute ma famille, à Papa, Maman, Émeline, Ghislain et Rémi, (oserai-je citer Waldo ?) toujours présents. Mention spéciale à Cédric, qui a supporté les angoisses de mes derniers mois de thèse, qui m'a tant aidée et soutenue, même si je le réveille toujours trop tôt et que je ne mets pas toujours la vaisselle dans l'évier.

Table des matières

REMERCIEMENTS **3**

LISTE DES FIGURES **9**

LISTE DES TABLEAUX **12**

LISTE DES ANNEXES **14**

LISTE DES ABREVIATIONS **15**

1 INTRODUCTION **17**

1.1 Annotation fonctionnelle d'une cellule procaryote: état des lieux **19**

1.1.1 Le génome et son annotation 19

1.1.1.1 Le séquençage *19*
1.1.1.2 L'annotation in silico *21*
1.1.1.2.1 Homologie de séquences ou présence de motifs 21
1.1.1.2.2 Gènes dont la co-localisation est conservée au cours de l'évolution 22
1.1.1.2.3 Détermination et modélisation de la structure 3D des enzymes par homologie 22
1.1.1.3 Ressources biologiques *23*
1.1.1.3.1 Ressources génétiques 24
1.1.1.3.2 Utilisation d'un ORFéome 24
1.1.1.4 Ressources méthodologiques : la cascade des « omiques » *25*

1.1.2 Le métabolome 26

1.1.2.1 Le métabolisme : « terra incognita » ? *27*
1.1.2.2 Qu'est-ce que la métabolomique ? *28*
1.1.2.2.1 Définitions 28
1.1.2.2.2 La métabolomique, un défi pour la chimie analytique 29
1.1.2.3 La métabolomique, un outil fonctionnel *30*
1.1.2.3.1 Gènes orphelins de fonction 30
1.1.2.3.2 De nouvelles voies métaboliques 34

1.2 Comment effectuer une étude de métabolomique ? **38**

1.2.1 Préparation des échantillons 38

1.2.1.1 La collecte des cellules *39*
1.2.1.2 Le quenching *39*
1.2.1.3 Extraction des métabolites *40*

1.2.2 Choix de la méthode analytique 41

1.2.3 La spectrométrie de masse 43

1.2.3.1 Le Velos Pro Orbitrap Elite *43*
1.2.3.1.1 Source HESI II 44
1.2.3.1.2 Piège 2D (Velos Pro) 44

1.2.3.1.3 L'Orbitrap Elite 46
1.2.3.1.4 Caractéristiques des analyseurs 47
1.2.3.2 Traitement de données 49
1.2.3.3 Identification structurale 50

1.3 Acinetobacter baylyi ADP1 au laboratoire 54

1.3.1 Le génome 54

1.3.1.1 Construction de ressources pour l'annotation fonctionnelle et l'étude du métabolisme. 54
1.3.1.2 Banque de mutants de délétion 55
1.3.1.3 Création de l'ORFéome 56
1.3.1.4 Modèle informatique de la bactérie 56

1.3.2 Le métabolisme du quinate reconsidéré par des approches globales 57

1.3.2.1 Une réponse spécifique attendue 58
1.3.2.2 Une réponse globale inattendue 58
1.3.2.3 De nouveaux métabolites? 59

1.4 Explorer la face cachée du métabolisme d'Acinetobacter baylyi ADP1 61

2 MATÉRIEL ET MÉTHODES 63

2.1 Matériel 65

2.1.1 Milieux de culture 65

2.1.2 Phases mobiles 65

2.2 Protocoles de préparation de métabolomes 65

2.2.1 Protocole d'extraction de métabolome haut-débit 65

2.2.2 Préparation de métabolomes pour la purification du MO211 68

2.2.3 Production de métabolomes à grande échelle 69

2.3 Chromatographie liquide 71

2.3.1 Chromatographie liquide analytique 71

2.3.2 Chromatographie liquide préparative 71

2.4 Paramétrage du Velos Pro Orbitrap Elite 72

2.4.1 Source d'ionisation HESI 72

2.4.2 Analyseurs 73

2.4.3 Traitement des données. 73

2.5 Résonance Magnétique Nucléaire 74

2.6 Validation de la structure du MO211 75

3 RÉSULTATS ET DISCUSSIONS 77

3.1 A la recherche des gènes impliqués dans la synthèse des Métabolites Orphelins 79

3.1.1 Réduction du temps d'analyse LC-MS 79

3.1.2 Préparation des métabolomes 84

3.1.3 Choix des Métabolites Orphelins 86

3.1.3.1 Les MO identifiés, non prédits par le modèle métabolique de la bactérie 86

3.1.3.2 Les MO inconnus 88

3.1.4 Traitement des données 92

3.1.5 Analyse des données 97

3.1.6 Contrôles qualité 100

3.1.6.1 Suivi de la croissance cellulaire 101

3.1.6.2 Contrôle de la qualité de l'extraction 102

3.1.6.3 Contrôle de la qualité du signal LC-MS 104

3.1.7 Interprétation et validation des données obtenues 105

3.2 Élucidation structurale d'un métabolite inconnu : le MO211 107

3.2.1 Analyse du MO211 en HRMS 107

3.2.1.1 Propriétés générales 107

3.2.1.2 Détermination de la composition élémentaire 109

3.2.1.3 Échanges H/D 110

3.2.1.4 Plusieurs signaux pour le MO211 111

3.2.1.5 Spectres de fragmentation 114

3.2.2 Analyse du MO211 par RMN 119

3.2.2.1 Production, extraction et purification du MO211 119

3.2.2.2 Interprétation des spectres RMN 124

3.2.2.3 Validation de la structure du MO 211 130

3.2.3 D'autres métabolites orphelins ? 133

3.2.3.1 Analyse du MO193 en HRMS 134

3.2.3.1.1 Propriétés générales 134

3.2.3.1.2 Détermination de la composition élémentaire 135

3.2.3.1.3 Échanges H/D 136

3.2.3.1.4 Plusieurs signaux pour le MO193 137

3.2.3.1.5 Premières expériences de fragmentation 138

3.2.3.2 Protocole d'extraction du métabolome 139

3.2.3.2.1 Relation entre volume de culture et teneur en MO 140

3.2.3.2.2 Influence de la valeur de DO_{600} sur l'abondance des métabolites extraits 141

3.2.3.2.3 Influence du solvant utilisé pour l'extraction sur la teneur en MO193 ... 143

3.2.3.3 Purification du MO193 144

4 CONCLUSIONS 149

BIBLIOGRAPHIE 155

ANNEXES 163

Liste des Figures

Figure 1: Coût de séquençage d'un génome du début du siècle à aujourd'hui 20

Figure 2: Pourcentage de gènes annotés selon différents niveaux de confiance pour trois bactéries : *Acinetobacter baylyi* ADP1, *E. coli*. K12, et *Bacillus subtilis* 168. 21

Figure 3: La cascade des "omiques" 25

Figure 4: Reconstruction in vitro de la voie de dégradation de la trigonelline..... 35

Figure 5 : Étapes d'un protocole de préparation d'échantillon type pour une étude métabolomique 38

Figure 6: Protocole de préparation des échantillons à partir de culture solide sur filtre 40

Figure 7: Schéma du spectromètre de masse Velos Pro Orbitrap Elite 43

Figure 8: Le piège à ions 2D (Velos Pro) 45

Figure 9 : Schéma de l'analyseur Orbitrap (d'après http://planetorbitrap.com, juillet 2018)..... 46

Figure 10: Les différentes étapes de l'identification des métabolites par spectrométrie de masse..... 52

Figure 11: Méthode de construction de mutants de délétion d'ADP1 55

Figure 12 : Dégradation des composés aromatiques via les deux voies menant à la voie du b-cétoadipate chez ADP1..... 57

Figure 13: Métabolisme du quinate chez ADP1..... 58

Figure 14 : Nombre et classification fonctionnelle COG des gènes différentiellement exprimés 59

Figure 15: Structure précédemment proposée pour la forme neutre de le MO211 60

Figure 16: Protocole de culture des cellules et extraction des métabolomes à haut débit..... 66

Figure 17: Plans de correspondance des puits de la plaque 96 puits aux plaques 24 puits. 67

Figure 18: Protocole de préparation des métabolomes en plaque pour la purification du MO211 68

Figure 19: Protocole de préparation des métabolomes à grande échelle pour la production des MO... 70

Figure 20 La phase stationnaire zwitterionique de la colonne ZIC-pHILIC 80

Figure 21: Évaluation de la qualité chromatographique d'une analyse d'un métabolome en UPLC.... 80

Figure 22: Représentation graphique des deux types de méthodes chromatographiques analytiques utilisées au laboratoire..... 82

Figure 23: Représentation graphique de la méthode analytique utilisée pour le criblage de la banque de mutant lors de deux injections consécutives 83

Figure 24: Comparaison de la qualité chromatographique des deux méthodes chromatographiques analytiques utilisées au laboratoire..... 84

Figure 25:Comparaison de la qualité des différentes méthodes d'extraction..... 85

Figure 26: Schéma explicatif de l'analyse d'une fraction du « métabolome fractionné » sur deux méthodes analytiques différentes. 89

Figure 27: Caractérisation du MO193 en LC/MS 90

Figure 28: Caractérisation du MO211 en LC/MS 91

Figure 29: Capture d'écran d'une matrice de résultats issu d'XCMS Online. 93

Figure 30: Évaluation de la qualité du prétraitement de données par XCMS 95

Figure 31: Évaluation de la qualité de l'éintégration effectuée par XCMS 96

Figure 32: Plan d'une plaque 96 puits de la banque de mutants 97

Figure 33: Visualisation des *fold* et Z-score de quelques puits au travers de Business Object..... 100

Figure 34: Visualisation de la mesure de densité optique à 600 nm des cultures de la plaque 1 de la banque de mutant au travers de Business Object. 101

Figure 35: Illustration de la méthode d'analyse d'une plaque pour le MO211 au travers de Business Object.. 102

Figure 36: Illustration de la méthode d'analyse d'une plauque au travers de Business Object 103

Figure 37: Évaluation de l'intensité du signal en spectrométrie de masse lors d'une séquence d'injection du criblage de la banque..... 104

Figure 38 : Procédure d'analyse des résultats du criblage de la banque de mutants de délétion 105

Figure 39: Identification des gènes impliqués dans la synthèse des métabolites orphelins au travers du criblage de la banque de mutants de délétion d'ADP1..... 106

Figure 40: Illustration d'une analyse comparative entre le métabolome du mutant Δ3 et ce même métabolome complémenté par le produit du gène excisé chez ce mutant. 107

Figure 41: Détection en LC/HRMS du MO211 108

Figure 42: Spectre d'absorbance du MO211..... 108

Figure 43: Structures de l'aprobarbital (a) et de la 2,6-dihydroxypseudonicotine (b) 109

Figure 44: Formule de calcul du nombre d'instaurations 110

Figure 45: Détermination du nombre d'échanges H/D pour le MO211 111

Figure 46: EIC des différents ions semblant provenir du MO211. 112

Figure 47: Spectre de fragmentation MS^2 de l'ion de valeur m/z209,09245, forme oxydée du MO211 116

Figure 48: Détermination du nombre d'échanges H/D pour le m/z 209-ox à gauche et m/z 180 à droite 117

Figure 49 : Première étape de purification du MO211 par HPLC préparative..... 121

Figure 50: Deuxième étape de purification du MO211 par HPLC préparative..... 122

Figure 51: Évalutation de la qualité de purification du MO211..... 123

Figure 52: Extraits du spectres RMN ^{1}H du MO211 purifié, 600 MHz, solvant H_2O/D_2O (90:10) ... 124

Figure 53: Position relatives des protons sur le cycle aromatique 125

Figure 54: Spectre RMN ^{13}C du MO 211 purifié, (150 MHz, solvant H_2O / D_2O (90:10))..... 126

Figure 55: Extrait du spectre RMN ^{1}H-^{13}C HSQC du MO 211 purifié..... 127

Figure 56: Structure suggérée par l'interprétation des données du spectre RMN HMBC ^{1}H-^{15}N..... 128

Figure 57 : Extrait du spectre HMBC 1H-13C pour la zone aromatique..... 129

Figure 58: Six structures possibles pour le MO211 130

Figure 59: Extraits des spectres RMN ^{1}H du MO211 et des 5 composés synthétisés..... 131

Figure 60: Analyse LC/MS des 5 composés synthétisés..... 132

Figure 61: Analyse LC/MS d'un métabolome additionné du 3DAP-4HBA (50µM)..... 133

Figure 62: Détection en LC/HRMS du MO193 134

Figure 63: Spectre d'absorption du MO193..... 135

Figure 64: Structures des huit molécules proposées pour le MO193 par les bases de données KEGG, Chebi et METLIN 135

Figure 65: Détermination du nombre d'échanges H/D pour le MO193 137

Figure 66 : EIC des différents ions semblant provenir du MO193 138

Figure 67: Flacon Fernbach, avec fond à ailettes..... 140

Figure 68 : Évalutation de la rétention du MO193 dans différentes conditions de pH sur C18..... 145

Figure 69 : Évaluation de la rétention du MO193 dans différentes conditions chromatographiques sur C18 146

Figure 70: Première étape de purification du MO193..... 147

Liste des Tableaux

Tableau 1: Récapitulatif des performances de différents analyseurs utilisés en métabolomique..... 49

Tableau 2: Présentation des trois méthodes chromatographiques analytiques utilisées..... 71

Tableau 3 : Présentation des trois méthodes chromatographiques préparative utilisées..... 72

Tableau 4: Paramètres de source des méthodes utilisées en modes d'ionisation positif et négatif..... 72

Tableau 5: Paramètres XCMS utilisés pour le traitement des données du criblage de la banque de mutant 74

Tableau 6: Liste des métabolites identifiés, non prédits par le modèle informatique de la bactérie détectés sur la méthode analytique de 50 min. 87

Tableau 7: Liste des MO inconnus retenus pour le projet. 88

Tableau 8: Liste des MO inconnus dont le suivi est possible pour le criblage de la banque 92

Tableau 9: Liste des métabolites de références séléctionnés pour l'analyse de la banque 103

Tableau 10: Liste récapitulative des signaux observés lors de l'analyse de l'analyse d'un métabolome, au temps de rétention du MO211 (14,6 sur la colonne de 15cm). 113

Tableau 11: Liste récapitulative des isotopes observés lors de l'analyse d'un métabolome au temps de rétention du MO211 (14,6 min. sur la colonne de 15 cm)..... 114

Tableau 12: Produits de fragmentations résultants des expériences HCD-MS^2 sur le MO211 115

Tableau 13: Produits de fragmentations résultants des expériences CID-MS^3 sur le métabolite d'intérêt de valeur m/z 194 115

Tableau 14: Produits de fragmentations résultants des expériences HCD, MS^2 sur le métabolite d'intérêt, ... 117

Tableau 15: Produits de fragmentations résultants des expériences CID-MS3 sur le métabolite d'intérêt, de valeur m/z 165 118

Tableau 16: Comparaison des abondances relatives de différents ions selon le volume de culture pour un même volume d'extraction..... 119

Tableau 17 : Interprétation des données obtenues par RMN ^{1}H et COSY des protons aromatiques .. 125

Tableau 18: Interprétation des données obtenues par RMN 1H et COSY des protons aliphatiques... 125

Tableau 19: Interprétation des données obtenues par RMN 2D HMBC ^{1}H-^{13}C et ^{1}H-^{15}N HMBC..... 128

Tableau 20: Corrélations observées lors de l'expérience RMN HMBC 1H-13C HMBC pour les protons aromatiques 129

Tableau 21: Récapitulatif des fragments observé en HCD- MS^2 de chacune des 5 molécules synthétisées 133

Tableau 22 : Liste récapitulative des isotopes observés lors de l'analyse d'un métabolome au temps de

rétention du MO193 (4,6 min. sur la colonne de 15 cm) en mode d'ionisation positif..... 136

Tableau 23 : Produits de fragmentations résultant des expériences HCD-MS^2 sur le MO193, à une énergie de collision normalisée à 35%. 139

Tableau 24: Comparaison des abondances relatives de différents ions selon le type de verrerie de culture utilisée pour un même volume et protocole d'extraction..... 141

Tableau 25 : Comparaison des abondances relatives de différents ions selon la DO pour un même volume d'extraction 142

Tableau 26 : Comparaison des abondances relatives de différents ions selon le solvant de lyse utilisé pour un même volume d'extraction..... 144

Liste des Annexes

ANNEXE 1: Convertisseurs de méthode chromatographique. 165

ANNEXE 2: Protocole de préparation d'un métabolome d'ADP1 selon le protocole d'extraction utilisé précédemment par le laboratoire[105]..... 167

ANNEXE 3: Liste des 141 métabolites inconnus.. 169

ANNEXE 4: XIC générés par XCMS des différents m/z semblant provenir du MO211. 173

ANNEXE 5 : Spectre de masse MS avec détection FTMS du MO211 en mode d'ionisation positif (haut) et négatif (bas) 174

ANNEXE 6 : Spectres MS^n avec détection haute résolution du MO211,10704 en mode d'ionisation positif..... 175

ANNEXE 7: Spectres MS^n avec détection haute résolution de l'ion m/z 209,09336 en mode d'ionisation négatif..... 176

ANNEXE 8: Spectre RMN ^1H du MO211 solvant H_2O/D_2O (90:10), 600 MHz). 177

ANNEXE 9: Spectre COSY ^1H-^1H obtenu du MO211 178

ANNEXE 10 : Spectre RMN HSQC ^1H-^{13}C du MO211 ; solvant H_2O/D_2O (90:10) 179

ANNEXE 11 : Spectre RMN HMBC ^1H-^{13}C du MO211 ; solvant H_2O/D_2O (90:10) 180

ANNEXE 12: Spectre RMN HMBC ^1H-^{15}N du MO211 ; solvant H_2O/D_2O (90:10)..... 181

ANNEXE 13: Superposition des spectres RMN du MO211 et des composés synthétisés 182

ANNEXE 14 : Spectres de détection en haute résolution du MO211 dans le métabolome et des 5 composés synthétisés, en mode d'ionisation positif..... 183

ANNEXE 15: Spectre MS^2 avec détection haute résolution du MO211 et des 5 composés synthétisés en mode d'ionisation positif..... 184

ANNEXE 16: Extrait du spectre de masse MS avec détection FTMS du MO193 en mode d'ionisation positif (haut) et négatif (bas) 185

ANNEXE 17: XIC générés par XCMS des différents m/z semblant provenir du MO193 186

ANNEXE 18 : Spectres MS^n avec détection haute résolution de l'ion m/z 193,09727 en mode d'ionisation positif..... 187

ANNEXE 19: Spectres MSn avec détection haute résolution de l'ion m/z 193,09727 en mode d'ionisation négatif..... 188

Liste des abréviations

ADP1	*Acinetobacter baylyi* ADP1
AM1	*Methylobacterium extroquens*
amu	unités de masse atomique
B. subtilis	*Bacillus subtilis*
CID	Collision Induced Dissociation
CDS	Séquences Codantes
E. coli	*Escherichia coli*
EIC	Extracted Ion Chromatogram
ESI	Electrospray Ionisation
FTMS	Fourier Transform Mass Spectrometry
FT-ICR	Fourier Transorm Ion Cyclotron Resonance
FWHM	Full Width at Half Maximum
HCD	High-energy Collision induced Dissociation
HESI	Heated Electrospray Ionization
HILIC	Hydrophilic Interaction Liquid Chromatography
HMBC	Heteronuclear Multiple Bond Correlation
HPLC	High performance Liquid Chromatography
HRMS	High Resolution Mass Spectrometry
HSQC	Heteronuclear Single Quantum Correlation
LC/MS	Chromatographie Liquide couplée à la Spectrométrie de Masse
Mb	Mégabase
MO	Métabolite Orphelins
MS	Mass spectrometry
MS/MS	Spectrométrie de masse en tandem
MS^n	Multi-stage Mass Spectrometry
MSI	Metabolomics standard Initiative
ORF	Open Reading Frame
Ppm	Partie par million
Q-TOF	Quadrupole Time Of Flight
RF	Radiofréquence
RMN	Résonance Magnétique Nucléaire
RP	Reverse Phase
TIC	Total Ion Current
UPLC	Ultra Performance Liquide Chromatography
UV	UltraViolet radiation
ZIC	Zwitterionic

1 INTRODUCTION

1.1 Annotation fonctionnelle d'une cellule procaryote: état des lieux

La première partie de cette introduction présentera les différentes ressources/méthodes qui existent pour appréhender le fonctionnement d'une cellule procaryote. Le champ de cette problématique étant très vaste nous avons pris le parti d'illustrer le propos avec les problématiques et les outils dont nous disposons au laboratoire.

1.1.1 Le génome et son annotation

1.1.1.1 Le séquençage

La première séquence complète d'un organisme a été publiée en 1977, celle du phage phiX174 [1] dont la taille est inférieure à 5 kb. Par la suite, le séquençage de génomes de nombreux organismes procaryotes a été réalisé. On citera pour mémoire : en 1995, la séquence du premier organisme autonome (free living organism) *Haemophilus influenzae* [2], suivi en 1997 de celle de l'organisme modèle *Escherichia coli* (*E. coli*) [3] ; la taille de ces deux génomes étaient respectivement de 1830 et 4639 kb.

Chez les eucaryotes, le séquençage du premier chromosome d'un organisme (le chromosome III de la levure de 315 kb) a été finalisé en 1991 par un consortium de 35 laboratoires européens [4]. A la fin des années 1990, les séquençages massifs d'organismes se sont démocratisés (*Ralstonia solanacearum* [5] ou *Arabidopsis thaliana* [6]) et ont ouvert la voie au séquençage du génome humain, d'une longueur de plus de 3000 Mb. Au total, plus de 20 laboratoires de six pays différents (USA, Japon, Chine, Grande Bretagne, Allemagne et France) ont été impliqués dans ce projet, pour un budget avoisinant les 3 milliards de dollars.

Depuis l'achèvement du projet sur le génome humain en 2004 [7], des progrès extraordinaires ont été réalisés dans les technologies de séquençage du génome. En 2006, la société Solexa, rapidement rachetée par Illumina, révolutionne le séquençage avec la série des HI-seq, produisant 100 Gb pour les Hiseq 2000 et 1000 Gb pour les Hiseq 4000 en un seul run. La même société commercialise actuellement une machine encore plus performante, le NovaSeq, qui produit trois à quatre fois plus de séquences que le HIseq 4000. En 2014, Oxford Nanopore Technologies a lancé une plateforme de troisième génération, le séquenceur MinION. Il s'agit d'un dispositif portable, permettant un séquençage en temps réel de l'ADN et de l'ARN. L'avantage de ce type de séquençage par rapport au séquençage Illumina réside dans la possibilité de séquencer des régions de plusieurs dizaines voire centaines de kilobases. Une évolution du MinION, actuellement en version beta, résulte de la parallélisation de 48 MinIONs. Ce type de séquençage est particulièrement intéressant pour le séquençage *de novo* ou le phasage des haplotypes. Actuellement son inconvénient est un taux d'erreur élevé [8]. De plus, il faut noter qu'en

parallèle à l'augmentation du débit de séquençage, le prix de la séquence a évolué de façon inversement proportionnelle. Le séquençage d'un génome humain avec des Hi-seq X ou des NovaSeq (société Illumina) est estimé à environ un millier d'euros (Figure 1).

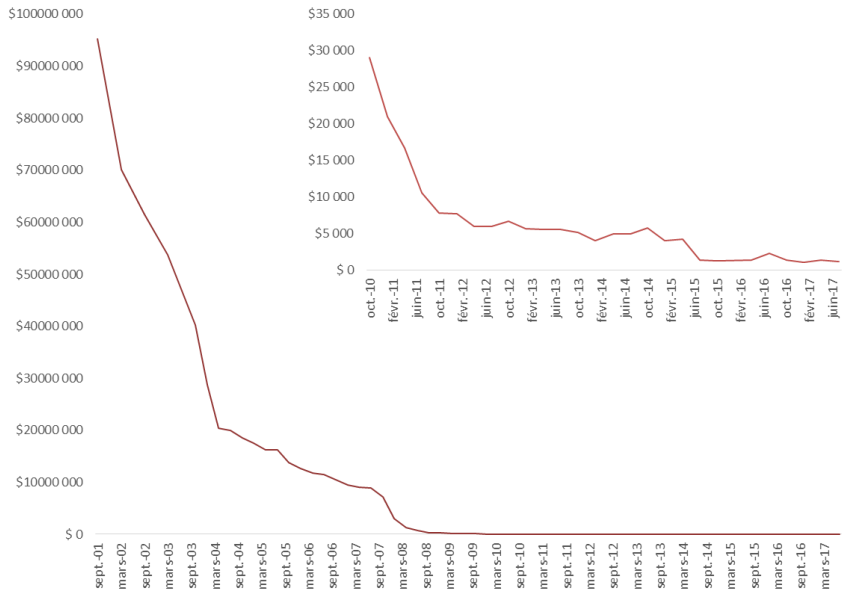


Figure 1: Coût de séquençage d'un génome du début du siècle à aujourd'hui (Source : NHGRI)

Aujourd'hui le séquençage n'est plus le principal challenge. Il est désormais devenu primordial de donner un sens biologique aux séquences génomiques générées. En effet, une séquence génomique est sans valeur si elle ne peut pas être déchiffrée. L'annotation fonctionnelle des données génétiques, c'est-à-dire la détermination des fonctions des régions codantes du génome, est considérée comme l'un des grands défis de la biologie. Elle est en effet essentielle pour développer une compréhension complète et détaillée de la physiologie cellulaire et est donc une préoccupation majeure dans la recherche biologique.

L'annotation concerne deux aspects : le premier consiste à repérer les régions où sont localisés les éléments génétiques (gènes, régulateurs, ARN non codants etc.), quand le second s'occupe d'associer des fonctions à ces éléments. Ce processus dont la phase initiale peut être totalement automatisée, nécessite, pour obtenir une annotation de qualité, une analyse experte et une validation expérimentale. Dans le cas d'*Acinetobacter baylyi* ADP1 (ADP1), organisme d'étude de notre laboratoire, seuls 5% des gènes ont une fonction déterminée expérimentalement. Pour 33% des gènes, des fonctions sont prédites grâce à une homologie de séquence avec un gène dont la fonction a été déterminée expérimentalement chez un autre organisme. 29% des gènes de la bactérie ne possèdent qu'une homologie partielle avec des gènes de fonction connue. Enfin, 33% des gènes n'ont aucune fonction associée. Même dans le cas d'organismes modèles comme *E. coli*, seuls 53% des gènes ont une fonction déterminée expérimentalement (Figure 2). De telles données sont révélatrices de l'ignorance actuelle sur

l'information fonctionnelle contenue dans les génomes.

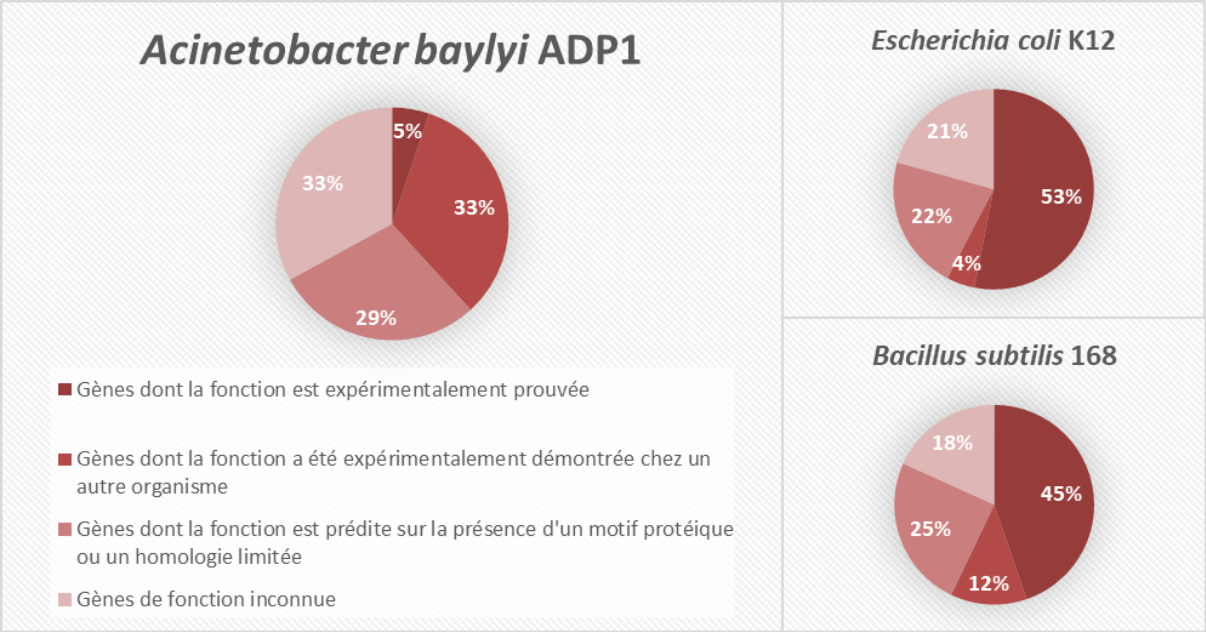


Figure 2: Pourcentage de gènes annotés selon différents niveaux de confiance pour trois bactéries : Acinetobacter baylyi *ADP1,* E. coli. *K12, et* Bacillus subtilis *168. D'après la plateforme Microscope* https://www.genoscope.cns.fr/agc/microscope/home/*).*

1.1.1.2 L'annotation in silico

L'ajout fréquent de nouveaux génomes dans des bases de données de séquences publiques permet un accès rapide à plus de 120 000 milliards de séquences protéiques. Cependant, la caractérisation expérimentale ne peut être effectuée que sur une infime fraction des séquences disponibles, et ce, malgré le développement de techniques à haut débit. En effet, de telles méthodes de travail ne peuvent être appliquées que sur - au maximum - quelques centaines voire quelques milliers de séquences de gènes ou de protéines à la fois. En conséquence, l'annotation des séquences repose très majoritairement sur des méthodes de calculs, développées pour prédire les fonctions moléculaires des millions de séquences de protéines qui n'ont pas pu être caractérisées expérimentalement. Au cours du temps, des algorithmes de prédiction de fonction de plus en plus efficaces ont été développés. Néanmoins bien que les meilleures méthodes soient suffisamment performantes pour guider les expériences, il y a un besoin important d'améliorer les outils actuellement disponibles [9]. Ce type de travail d'assignation de fonction s'appuie sur trois grands types de stratégies.

1.1.1.2.1 Homologie de séquences ou présence de motifs

Les homologies de séquences sont des méthodes basées sur l'homologie de séquence *in silico*. Elles ont été largement utilisées pour l'annotation de génome et le sont encore actuellement. La séquence de la protéine à annoter est comparée, grâce à des algorithmes de comparaison de séquences

nucléotidiques ou protéiques, tels BLAST [10], aux séquences présentes dans des bases de données telles que UNIPROT [11]. Si la similarité entre la séquence étudiée et une séquence dont la fonction est connue dépasse un seuil prédéfini, cette fonction sera associée au gène à annoter. Ces méthodes appliquent la règle suivante : deux séquences qui se ressemblent ont probablement la même fonction. Cette règle pose d'une part le problème du *cut-off* (en français, quantification du terme « ressemble à ») et d'autre part de l'application indue des règles de transitivité où « A ressemble à B » et « B ressemble à C », impliquerait « A ressemble à C ». Or cette propriété n'est pas exacte pour des relations d'homologie de séquence.

Dans le cas où aucune homologie « globale » de séquence n'est trouvée, il est toujours possible de considérer la présence de certains motifs « fonctionnels ». Ces motifs sont recensés dans des bases de données telles que PROSITE [12] ou PFAM [13]. De même que dans le cas de l'homologie de séquence, une fonction sera assignée au gène si celui-ci présente un ou plusieurs motifs pouvant le relier à une famille possédant déjà une annotation fonctionnelle. Ce genre d'annotation, qui a l'avantage de détecter des homologies faibles/lointaines, ne s'adresse souvent qu'à des régions de tailles limitées. Cependant, l'assignation de fonction peut être complexifiée dans le cas de protéines multi domaines.

1.1.1.2.2 Gènes dont la co-localisation est conservée au cours de l'évolution

L'annotation fonctionnelle peut s'appuyer sur l'existence de clusters de gènes dont la présence est conservée au cours de l'évolution. Cette notion est d'autant plus forte que la relation phylogénétique entre les organismes considérés est plus grande. La présence de ces clusters montre qu'il existe, au cours de l'évolution, une pression de sélection qui a amené les gènes correspondant à rester groupés sur le génome, voire à s'organiser en opérons. En effet, il a été observé que les opérons bactériens sont typiquement composés de gènes dont les produits contribuent à une même fonction métabolique [14]. Ainsi, la présence des gènes de fonctions connus dans le voisinage de gènes de fonction inconnue, permettra de proposer des hypothèses sur leur fonction [15]. Overbeek *et al*. [16] présentent notamment une méthode basée sur la conservation de la co-locatisation de gènes dans une vingtaine de génomes pour préciser la fonction de certains des gènes présent dans des clusters. Dans le monde bactérien, la part de gènes organisés en opéron est évaluée à 35%, ce qui en fait un outil particulièrement utile.

1.1.1.2.3 Détermination et modélisation de la structure 3D des enzymes par homologie

Enfin, il est également possible d'utiliser les propriétés tridimensionnelles des enzymes et plus particulièrement du site actif. Par exemple, le « docking » consiste à prédire la conformation et l'orientation du substrat dans la poche catalytique de l'enzyme. En effet, le ligand se fixe grâce à des interactions précises (Van der Waals, électrostatique, liaisons hydrogène) dans le site catalytique de l'enzyme. La modélisation de ces interactions pour différents substrats (par des logiciels tels que Dock, Fred ou AutoDock) permettra de déterminer les fonctions potentielles que peut remplir cette enzyme

[17].

Une autre méthode, développée au Genoscope, consiste à modéliser une enzyme par homologie, à partir d'enzymes de séquence homologue (>30% d'identité) dont les structures 3D sont connues. Cette technique, appelée l'ASMC (Active Site Modeling Clustering) [18], utilise l'alignement tridimensionnel des sites actifs d'enzymes pour les classer sous formes de groupes supposés iso-fonctionnels. Pour chaque groupe, il est possible de définir des acides aminés spécifiques. Après une caractérisation expérimentale de la fonction de certains représentants de chacun des groupes, il devient possible de relier les acides aminés spécifiques du site actif à une fonction. Par la suite, des règles logiques peuvent être définies et il est ainsi possible d'annoter de façon très fiable la fonction de chacun des membres de la famille [19, 20].

Malgré toutes les avancées algorithmiques et expérimentales il n'en reste pas moins que les annotations *in silico* sont souvent peu précises voire fausses. Récemment, il a été montré au laboratoire que, sur deux familles d'enzymes impliquées dans l'acétylation ou la succinylation de l'homosérine, première étape dans la voie de synthèse de la méthionine, 60% des 10 000 séquences présentes dans les bases de données avaient une annotation de fonction erronée [20]. De façon semblable, Schnoes *et al.* ont étudié les niveaux d'erreur d'annotation de fonctions dans quatre bases de données de séquences protéiques publiques [21]. Au total, 37 familles d'enzymes pour lesquelles des informations expérimentales sont disponibles ont été étudiées. Plusieurs types d'erreurs d'annotations ont ainsi pu être mis en évidence, la principale concernant la sur-annotation (« over annotation » en anglais) des gènes, c'est-à-dire l'annotation fonctionnelle de séquence de manière trop précise au vu des preuves apportées. Il est par ailleurs précisé que la propagation de ce type d'erreurs est problématique et n'est souvent pas prise en compte. Il s'agit là d'une menace sérieuse pour l'exactitude et la fiabilité des bases de données génomiques et protéiques.

Il est donc très important de prendre conscience qu'une annotation fonctionnelle pertinente doit être réalisée en deux étapes. Dans un premier temps, les outils bio-informatiques doivent être utilisés pour donner des pistes sur la fonction des gènes. Cependant, il est important de garder à l'esprit que cette étape ne fournit que des prédictions qui doivent être impérativement validées ou invalidées par un travail expérimental. Il est important de noter que, dans certains cas, des règles d'annotation peuvent être élaborées et propagées de façon informatique tout en gardant une très bonne qualité d'annotation fonctionnelles [19, 20].

1.1.1.3 Ressources biologiques

Le travail expérimental est très chronophage, mais aujourd'hui, le défi est d'essayer de réduire le fossé séparant la découverte de nouvelles protéines, qui s'accumulent de façon vertigineuse dans les bases de données, et la détermination de leur fonction. Pour cela, des ressources biologiques ont été

développées au fil des années pour aider à découvrir la fonction des gènes. Outre la séquence du génome et son annotation syntaxique et fonctionnelle qui est un prérequis, les ressources biologiques s'avèrent indispensables. Celles disponibles au laboratoire sont décrites ci-après.

1.1.1.3.1 Ressources génétiques

Les banques de mutants, et en particulier celles où les mutations correspondent à des délétions complètes de gènes, sont des outils de choix. Il s'agit d'une collection de souches dont un seul et unique gène a été excisé, idéalement de l'ATG au STOP. Une banque de mutants de délétion a été créée au laboratoire pour la souche ADP1 [22].

Celle-ci a été utilisée dans de nombreux contextes, notamment afin d'identifier les gènes impliqués dans une voie alternative de dégradation du D-glucarate [23]. ADP1 est capable d'utiliser le D-glutarate comme seule source de carbone et la voie de dégradation principale du D-glucarate est bien connue chez les bactéries. Cependant, une voie alternative a été observée chez ADP1 ainsi que chez d'autres organismes (tel *Bacillus subtilis*), et celle-ci n'était pas encore totalement caractérisée. Le laboratoire a alors étudié les capacités de croissance de chacun des mutants dans un milieu dont la seule source de carbone est le D-glucarate. Les mutants dont les gènes excisés interviennent dans la dégradation du D-glucarate ont ainsi été rapidement identifiés. Chacune des enzymes candidates pour chaque étape biochimique de la voie a été purifiée et leur fonction a été déterminée expérimentalement.

Cette année, Perchat et al. ont utilisé la banque de mutants d'ADP1 afin d'identifier des gènes impliqués dans la dégradation d'un métabolite : la trigonelline. Nous présenterons cette étude un peu plus loin (1.1.2.3.2.1)

Ces deux travaux basés sur la même banque de mutants de délétion montrent bien l'efficacité des approches pour contribuer à l'annotation fonctionnelle de ces organismes.

1.1.1.3.2 Utilisation d'un ORFéome

Un ORFéome est constitué de l'ensemble ou d'une grande partie des séquences codantes (CDS) d'une espèce, définies sur la base de l'annotation, clonées dans *E. coli* (dans le cas où l'on travaille avec des organismes procaryotes) [24]. Chaque enzyme d'intérêt peut alors être produite, et purifiée avant de procéder au travail de sa caractérisation. Par exemple, Liu *et al.* ont produit et purifié 1250 protéines de *Mycobacterium tuberculosis* [25]. Ces protéines ont été soumises à deux tests : le premier consiste à repérer les protéines impliquées dans l'immunité humorale en quantifiant les immunoglobulines G (IgG), le second permet de repérer les protéines antigéniques qui induisent la sécrétion d'une cytokine (IFN-γ), par les cellules sanguines. Au total, 29 des 1250 protéines soumises au premier test ont été détectées et décrites comme ayant une réponse antigène spécifique importante. Parmi ces 29 enzymes, une

douzaine d'entre elles n'avait jamais été décrite comme tel. Le second test révèle une antigénicité potentielle de 11 protéines, même si celle-ci est inférieure aux 2 antigènes commerciaux utilisés actuellement. Cette étude montre bien comment l'utilisation d'un ORFéome augmente le rythme de détection d'activités spécifiques des protéines et donc du gène associé.

Ce genre d'outil est donc particulièrement utile pour le haut débit. Dans le cas de la recherche de fonction enzymatique, tout ou partie de l'ORFéome peut être criblée pour trouver l'enzyme catalysant une fonction désirée, qu'elle soit métabolique, ou non. Chez ADP1 nous avons constitué un ORFéome de 2400 clones regroupant l'ensemble des gènes codant pour des enzymes non membranaires et les protéines de fonction inconnues.

En 2018, Perchat *et al*. ont étudié la place de la trigonelline dans le métabolisme de cette même bactérie. Dans ce travail, l'utilisation de la banque de mutant a permis d'identifier un groupe de gènes cataboliques, de transporteurs et de régulateurs [26] dont nous reparlerons plus bas (1.1.2.3.2.1).

Ces deux travaux basés sur la même banque de mutants de délétion montrent bien l'importance de l'étude des modifications génétiques sur un organisme. En multipliant les conditions de croissance de la banque, il serait alors possible de repérer facilement et rapidement par simple observation du phénotype les gènes impliqués dans telle ou telle activité catalytique.

1.1.1.4 Ressources méthodologiques : la cascade des « omiques »

Il s'agit là de replacer les gènes dans un contexte fonctionnel. Prenons l'exemple des gènes répondant de façon coordonnée à une perturbation. Cette réponse peut être analysée à trois niveaux possibles : le transcriptome (la transcriptomique), le protéome (la protéomique) ou le métabolome (la métabolomique). Cette analyse peut s'effectuer sur chaque niveau de façon indépendante, mais une étude intégrative, idéalement sur l'ensemble des niveaux, permet de mieux comprendre les mécanismes biologiques dans les systèmes complexes. Combinées à la génomique, ces techniques sont connues sous le nom : cascade des « omiques » [27] (Figure 3).

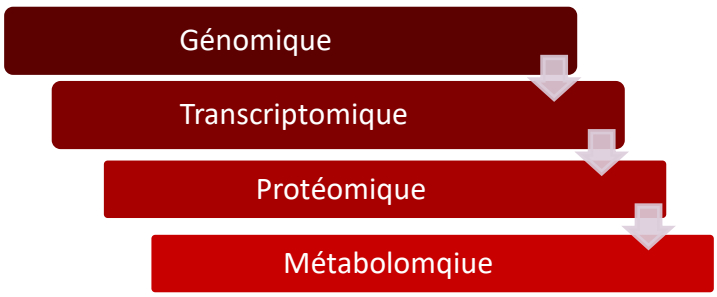


Figure 3: La cascade des "omiques"

Pour comprendre la relation entre le génome et le fonctionnement des cellules, il est essentiel d'étudier

les différents produits de ces gènes. Ce travail peut être fait au travers de l'étude de leur dynamique, lors de l'analyse globale de leur transcription, de leur traduction et *in fine* de l'ensemble de métabolites qui sont impliqués dans la chimie du vivant. Ce dernier aspect englobe le sujet de cette thèse et sera présenté indépendamment ci-après.

En étudiant les transcriptomes (ensemble des ARN issus de la transcription du génome), nous avons accès à des informations telles que le niveau de transcription des gènes en réponse à un stress biotique ou abiotique. En effet, le nombre de transcrits peut être quantifié donnant ainsi une idée des niveaux d'expression des gènes d'une population de cellules dans un état physiologique défini.

DeRisi *et al*. ont notamment montré l'intérêt de la transcriptomique chez *Saccharomyces cerevisiae* [28]. Ils ont mesuré la transcription des gènes de la levure dans deux conditions de cultures : l'une d'elle en anaérobie, l'autre en aérobie dans le but de suivre l'adaptation du métabolisme de la levure lors du passage d'une condition à l'autre (fermentation-respiration). D'un côté, l'étude des changements de niveaux d'expression des gènes entre les deux conditions a permis de détecter les voies métaboliques affectées par la transition fermentation-respiration, révélant comment l'organisme s'adapte à un changement environnemental. Par exemple, ils ont pu montrer que lorsque la cellule passe d'un état anaérobie à un état aérobie, l'organisme utilisera l'éthanol comme source de carbone et ceci déclenche notamment l'induction de deux gènes, impliqués dans la néoglucogenèse : FBP1 et PCK1. Ce travail a également mis en évidence d'autres modifications transcriptionnelles qui ne sont pas encore expliquées. En effet, près de 25% des gènes dont la transcription est modifiée entre les deux états physiologiques n'ont pas de fonction connue. Ceci montre, s'il en est encore besoin, que notre connaissance du métabolisme et de sa régulation est loin d'être complète même chez un organisme modèle.

La protéomique est l'étude de l'ensemble des protéines dans une cellule ou un tissu, et elle comprend des détails sur la quantité et la localisation dans les différents compartiments cellulaires et la diversité des protéines. L'utilisation d'ORFéomes, comme décrit précédemment, aide au déchiffrement de l'information contenue dans le code génétique. Cependant, le protéome peut ne pas raconter toute l'histoire d'une cellule. En effet, il n'existe pas de relations directes entre le génome et les protéines [29]. La protéomique est donc une science complémentaire de la génomique, n'éclaircissant qu'une partie du chemin suivi par la cellule au cours de sa vie.

1.1.2 Le métabolome

Déterminer la fonction des gènes, c'est mieux comprendre le fonctionnement d'un organisme, c'est-à-dire mieux discerner les arcanes de son métabolisme.

1.1.2.1 Le métabolisme : « terra incognita » ?

Le métabolisme est le résultat de l'ensemble des réactions chimiques se produisant dans les cellules d'un organisme. Elles sont catalysées par des enzymes, elles-mêmes codées à partir de l'information des gènes de l'organisme. En 1970, Donald Nicholson a compilé toutes les réactions métaboliques connues à l'époque en un seul graphique. Ce tableau, a fourni la première carte d'un métabolome cellulaire qu'il pensait complet. De fait, il ne met en évidence qu'une vingtaine de voies métaboliques [30]. Désormais, grâce au travail de la communauté scientifique, nous avons à notre disposition toujours plus de fonctions métaboliques aussi diverses les unes que les autres. Cependant, l'absence apparente de logique dans son organisation pourrait effrayer ceux qui s'aventureraient dans ce dédale de voies en inter-relation.

De manière générale, le métabolisme peut être défini en deux parties : le métabolisme primaire dit également central, et le métabolisme secondaire. Le premier est décrit par les fonctions de la cellule nécessaire à sa survie : la croissance, la reproduction et l'homéostasie métabolique. Le métabolisme secondaire s'occupe quant à lui des fonctions non essentielles et propres à chaque organisme en fonction de l'environnement dans lequel il se trouve.

De plus, ainsi que le décrivent Antoine Danchin et Agnieska Sekowska [31], le métabolisme résulte de contraintes très sélectives et est logiquement organisé. Il est constitué autour d'un set de blocs de construction précis et de molécules travaillant de concert pour la gestion de l'énergie. Les contraintes de bases conduisant aux processus tels que nous les connaissons sont :

- les atomes utilisés par la cellule : C, N, H et O sont les plus courants. Certains atomes tels que P, S, Na, Fe, Co, Cu etc. seront utilisés en tant que contre-ions dans les fluides extracellulaires ou bien dans les interactions électrostatiques du métabolisme de la cellule. Certains atomes ne seront utilisés que chez certains organismes, tels Si, As et Se, ainsi que certains des halogènes tels que F et Br.

- la stéréochimie des molécules produites et utilisées par la cellule. Par exemple les protéines sont composées de L-acides aminés. En revanche, les glucides dans les cellules sont le plus souvent basés sur un squelette de D-stéréoisomère.

- la plupart des réactions anaboliques sont à l'équilibre et peuvent être réversibles selon la disponibilité des substrats et produits.

Pourtant, bien qu'organisée autour de ces bases métaboliques, il existe encore une grande proportion de gènes et de protéines d'un organisme pour lesquels aucune fonction n'est associée. Par voie de conséquence, en supposant que 30% des gènes codent pour des enzymes, une grande partie du métabolisme reste inconnu. Une autre problématique a été soulevée au début des années 2000 : de

nombreuses activités enzymatiques, référencées grâce à leur numéro EC, ne sont pas associées à un gène de la cellule étudiée [32, 33]. Nommées activités enzymatiques orphelines (orphan enzyme) par Lespinet et al., elles pointent, elles aussi, le manque de connaissance partiel du métabolisme. Les conséquences de cet état de fait sont importantes. Comme aucune enzyme n'est associée à ces activités enzymatiques, elles sont systématiquement oubliées lors de l'annotation des génomes qui est essentiellement basé sur des relations d'homologie de séquence. Par voies de conséquence la biocatalyse et l'ingénierie métabolique ne peuvent donc pas utiliser ces activités enzymatiques.

Pour le métabolome, l'existence de gènes de fonction inconnue, et, dans une certaine mesure, d'activités métaboliques orphelines de gènes, entraine l'existence de métabolites inconnus ou non décrits dans le métabolisme, non associés à des enzymes ou à des voies métaboliques, que l'on nommera métabolites orphelins.

1.1.2.2 Qu'est-ce que la métabolomique ?

1.1.2.2.1 Définitions

Le métabolisme est défini par l'ensemble des réactions chimiques, catalysées par les enzymes synthétisées à partir des gènes de l'organisme. Le métabolome décrit le résultat de ces réactions chimiques. Il est l'ensemble des petites molécules (généralement de poids inférieur à 1000 Da) présentes dans un échantillon biologique (tissu, extrait cellulaire, urine etc.). De même que le terme «transcriptome» représente l'ensemble des transcrits et le «protéome» représente l'ensemble des protéines produites par un organisme, l'ensemble des métabolites synthétisés par un système biologique est le «métabolome».

Il représente des molécules avec une grande diversité de propriétés chimiques (poids moléculaire, polarité, solubilité) et présents dans une large gamme de concentration, allant du pmol au mmol [34]. La taille du métabolome varie considérablement en fonction de l'organisme étudié : la base de données du métabolome d'*E. coli*, ECMDB, contient plus de 2600 métabolites avec des liens vers environ 1500 gènes, y compris les enzymes et les transporteurs [35]. *Saccharomyces cerevisiae*, un autre organisme modèle, contiendrait environ 600 métabolites [36], alors que *Arabidopsis thaliana* en produirait plus de 2200 uniquement dans les feuilles et les racines [37]. Ceci peut refléter un contenu en gènes très différent entre les deux organismes, mais également, une richesse du métabolisme secondaire que l'on retrouve chez les bactéries, comme les streptomyces, ou chez les végétaux. Le métabolome humain se situe, lui, à une toute autre échelle : HMDB recense aujourd'hui plus de 114 000 molécules, comprenant des molécules de toute classe chimique, qu'elles soient abondants (> 1 µM) ou relativement rares (<1 nM). Cette complexité peut s'expliquer notamment par la diversité des tissus et fluides présents chez l'homme ainsi que sa grande complexité (réponses hormonale, immunitaire… à différents stimuli).

1.1.2.2.2 La métabolomique, un défi pour la chimie analytique

Bien que définie par le terme « métabolomique » pour la première fois en 1998 par Oliver *et al.* [38], l'étude d'un métabolome a été décrite pour la première fois en 1970 lorsque Horning et Horning ont utilisé les profils métaboliques (obtenus sur GC/MS) de sujets pour différencier les individus sains d'individus pathologiques [39]. Prenant véritablement son essor dans les années 2000, c'est en 2005 puis en 2007 que les deux premières bases de données de métabolites humains ont été construites : METLIN et The Human Database, respectivement. Le grand défi de la métabolomique réside dans la définition même du métabolome. Certaines études évaluent le nombre de métabolites d'un organisme entre 200 000 [40] et 1 000 000 [41], mais ces résultats sont à prendre avec précaution. En effet, dans un métabolome, un même métabolite peut se retrouver représenté un grand nombre de fois (adduits, fragmentation dans la source...) lors de l'analyse de l'échantillon. Ainsi, la détection, dans un extrait métabolique, de la présence de centaines, voire milliers, de signaux m/z ne signifie pas la présence du même nombre de métabolites correspondants. On se retrouve face à deux problèmes : le premier concernant le tri des signaux parasites, le second étant celui de l'identification des métabolites inconnus.

De manière générale, la biologie moléculaire et l'annotation fonctionnelle *in silico* partent du gène dans le but de lui associer une fonction qui, dans le cas qui nous intéresse, est une activité métabolique. Il est cependant possible de choisir comme point de départ l'élément le plus proche du phénotype : le métabolite. Il s'agit donc ici d'analyser le métabolome par une approche métabolomique sans *a priori* puis de remonter au gène par des approches de génétique, classiques dans l'esprit, mais qui nécessitent des ressources biologiques rarement disponibles. Cette approche métabolomique pose aussi des défis technologiques en termes de chimie analytique.

Raamsdonk *et al*. ont, parmi les premiers, montré en 2001 comment le profilage métabolique peut aider à l'annotation fonctionnelle [42]. En analysant les métabolomes de levures sauvages et mutantes ne montrant pas de phénotype de croissance particulier, les auteurs ont montré que malgré l'absence de différence phénotypique le métabolome des souches mutantes était modifié par rapport à la souche sauvage. En effet les auteurs ont mis en évidence une modification de concentrations de certains métabolites, dont certains impliqués dans la glycolyse (ATP, ADP, AMP, G6P, F6P et pyruvate), qui compensait l'effet de la mutation.

En 2009, Kind, Scholz et Fiehn [43] ont analysé deux métabolomes provenant de deux riz très différents génétiquement : *Oryza Japonica* et *Oryza Indica* dans le but d'établir une liste, la plus exhaustive possible, des métabolites présents chez le riz. Bon nombre de molécules détectées ne se retrouvait pas dans les bases de données et d'autres ont été assignées à des structures aberrantes chez ces organismes (pesticides par exemple). Les auteurs de cette étude ont alors conclu à la non exhaustivité de ces bases de données et proposent un changement dans les standards de publication contenant des structures

chimiques.

Aujourd'hui même si la métabolomique est utilisée en routine grâce aux innovations récentes des différentes méthodes analytiques, dans nombre d'applications de recherche ou dans des contrôles systématiques antidopage (chez l'homme ou le cheval), elle reste un challenge pour la détection et l'identification de nouveaux métabolites.

1.1.2.3 La métabolomique, un outil fonctionnel

La métabolomique permet de suivre l'évolution des transformations métaboliques d'une entité biologique et intervient dans deux domaines de recherche :

- l'assignation de fonctions à des gènes orphelins,
- la découverte de nouvelles enzymes et voies métaboliques

1.1.2.3.1 Gènes orphelins de fonction

1.1.2.3.1.1 Détermination de fonction in vitro

Un grand nombre de gènes restent dépourvus de fonctions précises. Il s'agit ici, dans le contexte du métabolisme, de définir a minima quels sont les substrats utilisés et les produits formés. Dans un certain nombre de cas, il existe des informations permettant de tester une série limitée de substrats, et il sera possible de tester les activités en suivant soit des observables indirectes (comme la réduction/oxydation de co-facteurs), soit l'apparition/disparition des produits/substrats par spectrométrie de masse [26, 44]. Dans d'autres cas, qui vont être illustrés ci-dessous, les informations obtenues ne permettent pas de poser des hypothèses quant aux substrats possibles des enzymes d'intérêts, mais d'autres stratégies ont été développées.

Nous présenterons tout d'abord ici, une approche utilisée pour élucider la fonction de deux activités enzymatiques chez *E. coli* [45]. Dans ce travail, un mélange complexe de métabolites, préparé à base de composés de culture de cellules (ex : extrait de levure), a été incubé avec des protéines candidates préalablement purifiées. Le mélange réactionnel a été ensuite analysé par électrophorèse capillaire couplée à la spectrométrie de masse (CE-MS). Comme preuve de concept Saito *et al*. [45], ont ainsi évalué *in vitro* la méthode sur plusieurs enzymes métaboliques d'*E. coli*, en suivant, les changements dans les niveaux de substrats et/ou produits. La puissance de cette méthode réside dans le fait qu'aucune connaissance *a priori* de l'activité protéique, ou de la composition du mélange de métabolites, est nécessaire.

Cependant, l'utilisation de substrats commerciaux en mélange, pose le problème de l'exhaustivité du mélange. Certains auteurs ont alors imaginé d'utiliser le métabolome d'un organisme comme substrat générique pour découvrir la fonction des enzymes chez ce même organisme. Le métabolome est un

substrat particulièrement pertinent car la quantité des métabolites présents dans le mélange est une image "fidèle" de l'état du métabolisme au moment de l'extraction. Enfin, il peut contenir des molécules inconnues ou non disponibles sur le marché. Cette approche alternative a été illustrée par Sévin *et al.* [46]. 1275 protéines non caractérisées issues de l'ORFéome ASKA de *E.coli* [24], ont été surexprimées et purifiées. Les enzymes ont alors été incubées dans un cocktail métabolique (métabolome extrait préalablement) et pour chaque enzyme les modifications des métabolites ont été analysées par LC/MS (Q-TOF Agilent 6520). Au total, l'analyse des modifications du mélange réactionnel de départ a révélé, parmi les 1250 enzymes purifiées, que 241 possédaient une activité enzymatique dans le mélange réactionnel. Seize de ces protéines ont été purifiées pour être incubées avec leurs substrats potentiels purs. Pour douze d'entre elles, l'activité enzymatique supposée a été validée. Cette méthode est généralisable à des lysats cellulaires bruts.

1.1.2.3.1.2 Détermination de fonction par marquage isotopique

Il est également possible, dans certains cas, d'utiliser des substrats marqués afin de faciliter l'identification des substrats et produits lors de l'analyse de données. Tang *et al.* [47] se sont penchés sur le cas de cytochromes P450 humains (P450s), dont les substrats étaient encore inconnus. Ces enzymes jouent un rôle important dans le métabolisme d'un grand nombre de composés, y compris les stérols, les acides gras. La stratégie utilisée est la suivante : chaque P450 est incubé dans un milieu réactionnel contenant en particulier une NADPH-P450 réductase et de l'extrait de foie sous un mélange de $^{18}O_2$ / $^{16}O_2$ (1: 1). Une analyse LC/MS du mélange réactionnel permet de suivre l'apparition des produits, ceux issu du marquage isotopique, apparaissant sous la forme de doublets M / M + 2. Ceux-ci peuvent être directement identifiés via un programme de traitement de données de spectrométrie de masse DoGEX [48]. Après validation de la méthode sur le P450 7A1, cette stratégie a été appliquée aux P450 humains 1A2, 2C8 et 2C9 et a permis de découvrir sept acides gras comme substrats de ces trois P450 hépatiques. La combinaison de la métabolomique LC/MS et du programme DoGEX permet une simplification considérable de la procédure d'identification de réactions dans des extraits tissulaires, de substrats potentiels et d'identification de produits de réactions pour des enzymes de fonction inconnue.

1.1.2.3.1.3 Détermination de fonction par profilage métabolique

Cependant, les manipulations *in vitro* des protéines se limitent aux protéines solubles et pouvant être purifiées. Or, le cœur de la régulation des réseaux biochimiques repose sur une multiplicité d'interactions complexes et dynamiques entre les différents acteurs enzymatiques et les métabolites impliqués [49]. Ainsi, l'observation de la réponse métabolomique à des perturbations biotiques ou abiotiques sur des cellules vivantes fournit un cadre plus physiologique pour la découverte de la fonction des enzymes. Ce travail évite également le besoin de sélectionner arbitrairement des substrats candidats pour l'analyse. Il est alors possible d'étudier l'enzyme dans son organisme d'origine. C'est ici que les techniques de modifications génétiques d'organismes, comme la construction de mutants, peuvent aider

à la détermination du rôle des enzymes.

Dans cette optique, Saghatelian *et al*. [50] décrivent une stratégie générale pour identifier les substrats endogènes d'enzymes. Dans ce but, des métabolomes issus de souris, où l'enzyme d'intérêt (une amidehydrolase d'acide gras : FAAH) a été inactivée, ont été comparés à des métabolomes issus de tissus non modifiés. Après des analyses en LC-MS, les auteurs ont découvert plusieurs lipides cérébraux régulés par la FAAH, dont une nouvelle famille de lipides présents dans le système nerveux : les acides gras taurins conjugués. En conclusion, les auteurs insistent sur le fait que l'activité hydrolytique de la FAAH pour les lipides *in vitro* ne pouvait permettre de prédire les substrats spécifiques de la FAAH *in vivo*.

Plus récemment, une stratégie a été développée, basée sur la surexpression transitoire d'un ensemble d'enzymes de fonction inconnue, dans un système cellulaire approprié. Le métabolome a ensuite été analysé en LC-MS. Une étude de comparaison des pics a révélé les métabolites modifiés en réponse à la surexpression spécifique d'une enzyme. L'équipe a appliqué cette stratégie à des études de profilage métabolique de lignées de cellules transfectées avec douze hydrolases non caractérisées, de la classe métabolique des sérines hydrolases. L'une d'entre elle, ABHD3 a été particulièrement étudiée et a pu être identifiée comme une lipase spécifique des phospholipides à chaîne moyenne et des phospholipides tronqués par oxydation [51]. Là encore, les auteurs soulignent que les essais biochimiques traditionnels n'auraient peut-être pas réussi à mettre en évidence l'activité inhabituelle de l'ABHD3 sur les phospholipides à chaîne moyenne.

1.1.2.3.1.4 Découverte de la fonction de gènes par une approche génétique

Dans le cas d'organismes modèles tels *E. coli* ou ADP1, il existe des banques de mutants de délétions. Une approche génétique peut alors être envisagée en analysant les métabolomes de mutants, afin de déterminer quel mutant induira une modification de l'activité métabolique d'intérêt. Le principal inconvénient d'un tel travail est qu'il nécessite un travail titanesque de criblage et d'analyses de données.

Messerli *et al*. [52] ont contourné ce problème en effectuant ce travail de criblage sur un nombre restreint de mutants. En partant de l'idée que des mutations dans les gènes impliqués dans une même voie métabolique conduient à un même phénotype, ils ont sélectionné des mutants défectueux ou surproducteurs d'amidon. Ils ont analysé les profils métaboliques de ces mutants et montré que chaque génotype présentait un métabolome unique. Par analyses statistiques (par exemple, analyse en composantes principale, PCA), il a été possible de regrouper sur la base du profilage métabolique, une logique biologique au regroupement des mutants. Par exemple, le premier groupe obtenu par une analyse HCA (Classification Ascendante Hiérarchique) regroupe deux mutants impliqués dans deux étapes consécutives d'une même voie métabolique, le deuxième regroupe deux mutants dont l'un est déficient dans la synthèse de l'amidon et l'autre surproducteur d'amidon. Malgré les deux phénotypes opposés,

les auteurs expliquent que leur regroupement provient d'un même état de manque de carbohydrates à la fin de la nuit, moment correspondant à l'heure de prélèvement des échantillons. Ces travaux mettent en évidence l'intérêt des analyses métabolomique non ciblées (« full scan ») dans le contexte de l'étude du métabolisme.

Le profilage métabolique appliqué à des organismes modifiés génétiquement, a également été utilisé afin de repérer les gènes impliqués dans telle ou telle activité métabolique. Par exemple, Baran, *et al*. [53] ont suivi la présence/absence d'une dizaine de métabolites extracellulaires (footprint) captés ou libérés par deux organismes : *E. coli* et *Shewanella oneidensis*. Les milieux de cultures de huit mille quarante-deux mutants (trois mille neuf cent un pour *E. coli* et quatre mille cent quarante et un pour *S. oneidensis*) délétés chacun pour un gène unique ont été criblés. Les exo-métabolomes (footprint) de chacun des mutants ont été analysés, et, sur la base des résultats obtenus et de l'accessibilité des métabolites, une dizaine ont été sélectionnés. En utilisant cette approche, les auteurs ont identifié des gènes de fonction connue ainsi que de nouveaux gènes (par exemple, transport de la citrulline, ABC transporteur, histidinase). Cette étude montre l'apport de l'utilisation du profilage métabolique sur des banques de mutants pour compléter la connaissance du métabolisme. Il faut toutefois noter que la redondance fonctionnelle est une limitation à ce type d'approche.

1.1.2.3.1.5 Combinaison de méthodes « omiques »

La combinaison d'études de génomique, de transcriptomique et de métabolomique peut s'avérer très efficace pour fournir des informations sur les réseaux de gènes. L'idée de départ est très simple et repose sur l'hypothèse que les gènes impliqués dans une voie de biosynthèse partagent des caractéristiques en commun. Par exemple, si un gène inconnu est co-exprimé avec des gènes connus comme impliqués dans un processus biologique particulier (plus concrètement, dans une voie métabolique particulière), ce gène inconnu est probablement également impliqué de façon directe ou indirecte dans ce même processus. Si cette analyse est couplée à des analyses génomiques, en utilisant le principe de « culpabilité par association » on peut distinguer facilement ceux qui participent à une même voie métabolique et se retrouvent co-localisés sur le génome [54] de ceux qui répondent par exemple aux mêmes facteurs de transcription.

L'intégration de la métabolomique et de la transcriptomique peut fournir des informations précises sur les relations gènes/métabolites. Tohge *et al*. [55] décrivent une analyse complète du métabolome et du transcriptome d'*Arabidopsis thaliana* en surexprimant le gène PAP1, un facteur de transcription, pour découvrir de nouvelles fonctions dans la voie de biosynthèse des flavonoïdes. L'analyse métabolomique ciblée sur les flavonoïdes a permis de mettre en évidence huit nouvelles anthocyanines à partir de 1800 « métabolites » détectés dans les plantes où PAP1 est surexprimé. L'analyse transcriptomique a mis en évidence, de son côté, la surexpression de 38 gènes dont plusieurs

gènes sans fonction connue ou avec des annotations fonctionnelles élusives. Deux gènes de glycosyltransférase putative (At5g17050 et At4g14090) sont confirmés être, respectivement, une flavonoïde 3-O-glucosyltransférase et une anthocyanine 5-O-glucosyltransférase. De plus, pris dans leur ensemble, les résultats de cette analyse intégrative des données du métabolome et du transcriptome suggèrent que PAP1 régule spécifiquement les gènes impliqués dans la synthèse des flavonoïdes. Cette approche représente donc une méthode efficace pour d'identifier de nouvelles fonctions du métabolisme des plantes mais probablement aussi chez les micro-organismes procaryotes.

1.1.2.3.2 De nouvelles voies métaboliques

Notre compréhension actuelle du métabolisme central a été influencée par des études réalisées dans un petit nombre d'organismes modèles dans un ensemble limité de conditions d'essai. On peut donc supposer que les voies réellement existantes sont mal représentées dans les cartes métaboliques et que de nombreuses voies métaboliques restent à découvrir. Par conséquent, la découverte de nouvelles voies métaboliques est un effort qui doit être continué et amplifié [26, 56].

1.1.2.3.2.1 Le profilage métabolique pour l'étude du métabolisme du carbone

Par exemple, comprendre comment les micro-organismes sont capables de convertir des sources de carbones uniques en biomolécules complexes est un intérêt majeur d'un point de vue fondamental mais aussi pour des applications dans les biotechnologies.

Le métabolisme central du carbone est, de l'avis général, connu de façon exhaustive. Cependant, la robustesse de la voie pourrait encore dissimuler des réactions non caractérisées. C'est pourquoi Nakahigashi *et al.* [57] ont, en 2009, analysé les voies métaboliques alternatives utilisées par *E. coli* en réponse à l'inactivation de gènes codant pour des enzymes primaires du métabolisme central du carbone. Pour découvrir ces voies cryptiques, les auteurs ont comparé les données expérimentales de croissance obtenues sur 12 sources de carbones différentes pour 2177 mutants multiples (deux mutations ou plus dans la même souche), aux prédictions issues du modèle métabolique de cet organisme. Dans 39 cas, il existait une contradiction entre les données expérimentales et la prédiction faite par le modèle. Parmi elles, les auteurs ont mis en évidence de nouvelles réactions catalysant la dégradation du sédoheptulose7-phosphate en érythrose 4 phosphate (E4P) et dihydroxyacétone (DHA). En présence de forte concentration de sédoheptulose7-phosphate, celui-ci peut être transformé en sédoheptulose-1,7-bisphosphate puis en E4P et DHA par l'action successive d'une fructokinase (PfkA) et une aldolase (FbpA) chez les mutants déficients en transaldolase. Il s'agit donc de la découverte d'une activité métabolique « promiscuitaire » de dégradation du sédoheptulose7-phosphate, en réponse au stress créé par la suppression de l'activité transaldolase « native ».

Chez cette même bactérie, la combinaison d'outils génétiques et de métabolomique a permis d'identifier

une nouvelle voie de dégradation de l'uridine monophosphate (UMP) [58] grâce à des souches d'*E. coli* déficientes en un régulateur de l'homéostasie pyrimidique. Dans ce cas, l'établissement du profil métabolomique a permis l'identification non seulement d'une nouvelle activité enzymatique (phosphatase UMP), mais également d'un mécanisme régulateur unique impliqué dans le catabolisme de la pyrimidine.

Une autre étude, évoquée précédemment dans cette introduction (1.1.1.3.2 et 1.1.1.3.1), a étudié la voie de dégradation de la trigonelline chez ADP1 [17]. Dans une étape initiale ils ont montré qu'ADP1 était capable d'assimiler la trigonelline comme source d'azote et de carbone. En observant les phénotypes de croissance en utilisant la trigonelline comme source de carbone de la banque de mutants de délétion d'ADP1, ils ont pu mettre en évidence un cluster de gènes impliqués dans la dégradation de la trigonelline. Après purification des enzymes codés par les gènes correspondants, ils ont testé la capacité de différents cocktails de ces enzymes à dégrader *in vitro* la trigonelline.

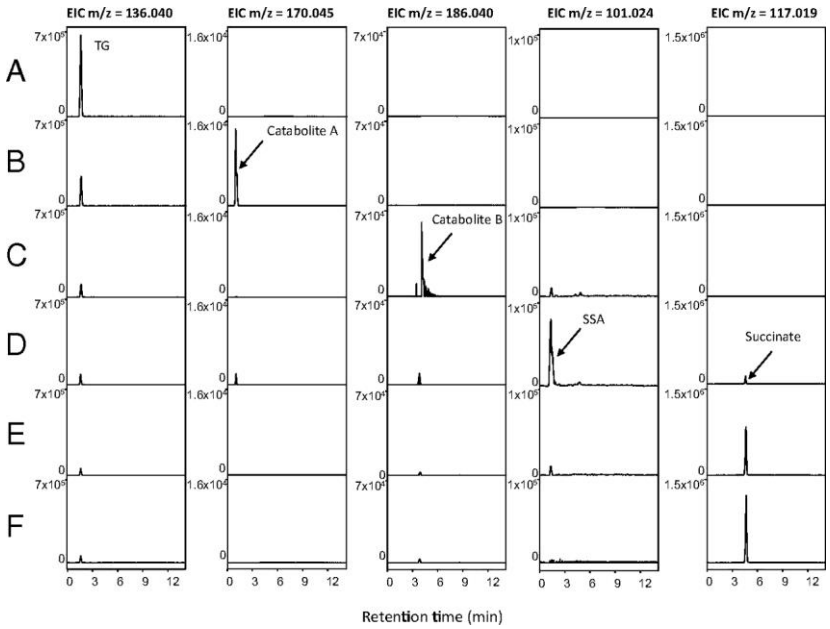


Figure 4: Reconstruction in vitro *de la voie de dégradation de la trigonelline. Deux microgrammes de chaque protéine ont été ajoutés. (A) Mélange réactionnel en l'absence d'enzyme. (B) Echantillon analysé 30 minutes après l'ajout de TgnA et TgnB. (C) Échantillon analysé 30 min après l'ajout de TgnAB et TgnC. (D) Échantillon analysé 30 minutes après l'ajout de TgnAB, TgnC et TgnD. (E) Échantillon analysé 30 minutes après l'ajout de TgnAB, TgnC, TgnD et TgnE. (F) Échantillon analysé 30 minutes après l'ajout de TgnAB, TgnC, TgnD, TgnD, TgnE et TgnF. Les échantillons ont été analysés par LC/MS en mode d'ionisation négative. Les chromatogrammes ioniques extraits (EIC) correspondent aux ions moléculaires ([M-H]-) de TG [m/z136,040 unités de masse atomique (amu)], catabolite A (m/z170,045 amu), catabolite B (m/z186,040 amu), Succinate Semialéhyde (SSA) (m/z101,024 amu) et succinate (m/z117,019 amu) avec une précision de 5 ppm. Issu de [26]*

Chacune des enzymes du cluster a ensuite été incubée dans un mélange contenant de la trigonelline, du NADH et du FMN. Comme montré ci-dessous (figure 4), la présence de TgnAB a déclenché la production d'un premier catabolite A et une diminution du pic de la trigonelline. Une fois la première enzyme identifiée le processus a été itéré en la combinant avec les enzymes restantes jusqu'à l'ordonnancement de toutes les enzymes dans la voie. Les enzymes, leur fonction et les métabolites associés ont été caractérisés et l'ordre des réactions de la voie identifiée. Deux métabolites nouveaux ont été découverts et leur structure a été déterminée.

Cette étude montre bien l'efficacité de cette utilisation du profilage métabolomique pour décrypter de nouvelles voies métaboliques quand il est utilisé en combinaison avec des approches génétiques, biochimique et de chimie analytique.

Varier les conditions de croissance, ou appliquer des stimuli abiotiques ou biotiques, peuvent également être couplées à du profilage métabolomique pour l'élucidation des voies métaboliques cryptiques.

Farag *et al*. [59] ont récemment utilisé une approche métabolomique pour découvrir plusieurs nouvelles voies de synthèse des isoflavonoïdes et des phénylpropanoïdes chez la légumineuse *Medicago truncatula*. Pour cela, des cellules en culture ont été soumises à une stimulation par le jasmonate de méthyle (MeJA) et l'« éliciteur de levure » (YE), deux agents qui induisent une reprogrammation génétique générale du métabolisme secondaire chez des plantes. Le profilage métabolomique, au-delà de la caractérisation de la cinétique d'induction de production de métabolites de la voie de synthèse des flavonoïdes par YE et MeJA, a permis de découvrir trois nouvelles isoflavones méthylées induites par YE, ainsi que leur origine métabolique par l'utilisation de traceurs marqués. De nombreux métabolites secondaires végétaux ont des propriétés pharmacologiques importantes, et il est donc crucial de comprendre leurs voies de biosynthèse et leur régulation. Comme le montre cette étude, les stratégies basées sur la métabolomique peuvent jouer un rôle central dans la réalisation de cet objectif.

1.1.2.3.2.2 L'utilisation de techniques de marquage isotopique

L'utilisation de traceurs marqués permet d'aider à l'identification de voies métaboliques. En effet, il est possible de suivre la présence d'isotopes le long d'une voie métabolique et de suivre les flux de certains métabolites et ainsi mettre en évidence la présence des voies de biosynthèse ou des schémas complexes de régulations de ces voies.

Le marquage isotopique a été appliqué pour démontrer que chez *Methylobacterium extroquens (AM1)*, la voie de l'éthylmalonyl-CoA [60] était utilisée pour l'assimilation de composés C1 (ex : le méthanol). La conversion de l'acétyl-CoA en glyoxylate chez les bactéries isocitrate lyase négatives utilise deux voies métaboliques : le Glycine Reductive Cycle et/ou la voie de l'éthylmalonyl-CoA qui implique

comme intermédiaires des dérivés de CoA de nature différente. Pour déterminer quelle(s) voie(s) sont utilisées chez AM1, les cellules ont été cultivées en présence de méthanol marqué au ^{13}C et la cinétique d'apparition rapide des dérivés de CoA marqués a été suivie par spectrométrie de masse à haute résolution (HRMS). Ce travail couplé à des études sur la distribution des ^{13}C-isotopomères de la glycine a permis de démontrer que la voie de l'éthylmalonyl-CoA était exclusivement utilisée par AM1 pour l'assimilation des C1. Outre l'importance de l'étude sur l'assimilation des C1 réduits, le présent travail met également l'accent sur les complémentarités de la spectrométrie de masse et de la Résonance Magnétique Nucléaire (RMN) pour mettre en évidence la topologie de réseaux métaboliques complexes à partir d'expériences de marquage au ^{13}C.

1.2 Comment effectuer une étude de métabolomique ?

1.2.1 Préparation des échantillons

De manière générale, une grande partie de la diversité du métabolisme est due à la présence d'un très large éventail de métabolites secondaires, dont le nombre est largement supérieur à celui des métabolites primaires. Dans tout système biologique, il existe une très grande diversité chimique de métabolites, présents dans une gamme dynamique de concentrations très vaste. Le type de molécules observé dépendra évidemment du type d'échantillons choisis (sang, urine, tissus, milieu intra ou extra cellulaire …). Un plan de travail détaillé [27] permettra l'acquisition de données fiables ainsi qu'une interprétation biologique pertinente.

Il peut être détaillé selon les étapes détaillées en Figure 5.

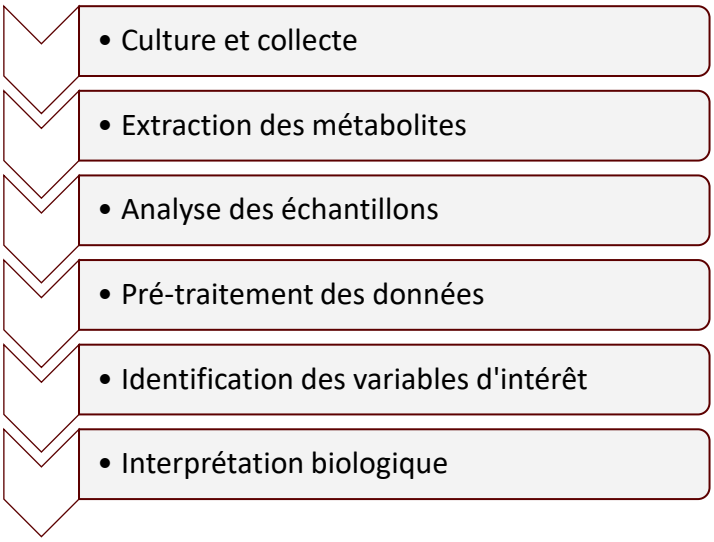


Figure 5 : Étapes d'un protocole de préparation d'échantillon type pour une étude métabolomique

Une méthode de préparation d'échantillons idéale doit être :

- la moins sélective possible (afin d'observer une diversité de métabolites la plus exhaustive possible)
- la plus simple et la plus rapide afin d'éviter la perte de molécules instables par exemple,
- reproductible, et si possible, contenir une étape de *quenching*.

Ces aspects ont déjà été développés en détail dans le travail de thèse de Lucille STUANI, publiée en Octobre 2014 [61], seul un résumé est présenté ci-dessous.

1.2.1.1 La collecte des cellules

Afin d'avoir des mesures reproductibles, les conditions de culture du matériel biologique à étudier doivent être aussi homogènes que possible. La lumière, la température, le taux d'humidité, les nutriments apportés ainsi que le temps d'échantillonnage peuvent influer sur la composition d'un échantillon.

Il existe plusieurs méthodes d'échantillonnage ou de collecte des cellules. Dans tous les cas, il est nécessaire que cette étape engendre le moins de stress possible pour l'échantillon (les cellules seront de préférence dans un état stationnaire par exemple). Il y a trois manières de procéder :

- la centrifugation
- la filtration rapide
- la collecte directe

La centrifugation soumet les cellules pendant plusieurs minutes à une force centrifuge de plusieurs milliers de g. Cette étape n'est utilisée en routine que dans les cas d'études de milieux extracellulaires ou de sang pour séparer le plasma des cellules. Cette technique a l'avantage de la simplicité mais ne permet pas d'obtenir une image exacte du métabolome "réel", les dérives dues au stress et au temps de centrifugation amènent inévitablement à des dérives de la composition en métabolites des échantillons [62] .

La filtration de la culture sous vide est simple et rapide, les cellules qui se retrouvent sur le filtre peuvent alors être rincées ou "quenchées" rapidement. Elle constitue une méthode de choix quand une quantité limitée de métabolomes doit être analysée. Dans nos mains et avec la technique mise au point au laboratoire une personne peut prendre en charge quatre échantillons par jour. Là encore comme pour la centrifugation, le temps de filtration va amener des biais quant à l'exhaustivité des métabolites extraits.

Enfin, la collecte directe permet de mélanger directement le milieu de culture des cellules avec la solution de *quenching.* Un des avantages de cette technique est la rapidité mais l'inconvénient principal est de ne pas permettre de différencier les métabolites intracellulaires et extracellulaires. Il est cependant possible avec une telle approche de suivre les cinétiques de métabolites en versant, à une fréquence prédéfinie, un volume donné du milieu de culture directement dans la solution de *quenching* [63].

1.2.1.2 Le quenching

Le *quenching* est une étape qui permet de figer le métabolisme de la cellule au moment de l'extraction. Le but est de limiter les activités enzymatiques à partir de cette étape afin d'éviter les variations de concentration de métabolites, biaisant inévitablement les interprétations biologiques [64]. Différentes méthodes peuvent être utilisées, en particulier un abaissement brutal de la température (ex

azote liquide) ou l'utilisation d'un solvant de *quenching* comme le méthanol froid (entre -20 et -60°C), ou encore, un changement de pH obtenu par ajout d'acide perchlorique ou de soude.

Parmi les différentes techniques présentées dans la littérature, celle développée par Rabinowitz chez *E. coli* sera présentée ici (Figure 6) car le laboratoire s'en est inspiré pour l'extraction de métabolites [65]. Il propose une alternative à la culture liquide et propose de faire la culture sur un filtre contenant les cellules, posé sur milieu solide. Ce protocole a l'avantage de minimiser les temps de transfert entre le milieu de culture et la solution de *quenching*.

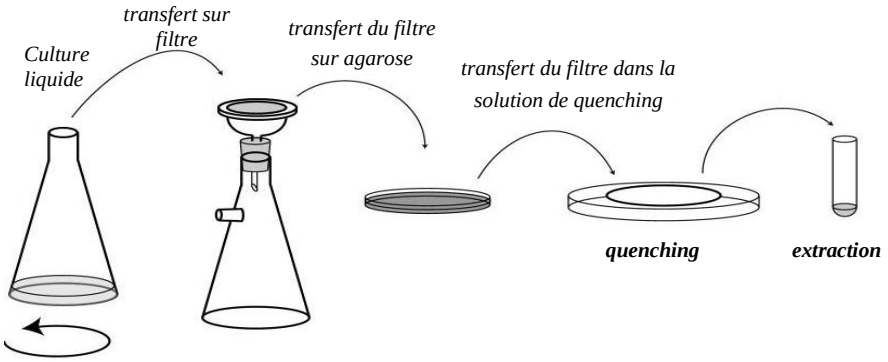


Figure 6: Protocole de préparation des échantillons à partir de culture solide sur filtre [65]

1.2.1.3 Extraction des métabolites

Le type et la quantité de métabolites extraits dépendent fortement de la méthode d'extraction choisie, mais il existe quelques règles communes à tous types d'extractions. Par exemple, en règle générale, plus il y a d'étapes dans la préparation de l'échantillon, plus la diversité chimique des composés présents sera modifiée par rapport à la réalité de la composition du métabolome.

En particulier pour les applications de profilage métabolique à "haut débit" (plusieurs centaines d'échantillons à extraire simultanément), une méthode d'extraction rapide, reproductible et surtout non sélective est préférée pour détecter la plus large gamme de métabolites possible, en évitant les modifications chimiques imprévues.

Une analyse des différentes méthodes d'extraction a été réalisée par Canelas *et al*. [66]. Dans le cadre d'une revue, les auteurs ont comparé la qualité de différentes extractions sur plusieurs souches (bactéries, levures ou plantes). Ils ont constaté notamment que le choix de la méthode d'extraction peut considérablement affecter le niveau des métabolites détectés. Ils notent en particulier que les méthodes d'extraction sont organisme-dépendantes et qu'il est crucial de développer un protocole *ad hoc* pour chaque organisme étudié. Enfin, parce que l'état biologique d'une cellule peut changer rapidement, et peut être très sensible à l'environnement, il est primordial d'être méticuleux au niveau des conditions de culture. De façon plus générale, l'interprétation biologique des variations de métabolomes doit se faire

à l'aune des conditions expérimentales qui en plus d'être reproductibles doivent être précisées en détails pour la publication des données. Son utilisation dans des méta-analyses provenant de différents expérimentateurs ou produites par différents laboratoires ne peut avoir une signification biologique que si les détails expérimentaux sont fournis avec les sets de données. Il existe donc un besoin de normes pour les métadonnées associées. Les normes rendent l'information plus accessible et permettent l'extraction de la valeur maximale des ensembles de données, en facilitant l'interrogation efficace, l'analyse précise des données et en facilitant l'échange de données.

Cet effort a été initié notamment par la Metabolomic Standard Initiative (MSI) [67]. La MSI est structurée en cinq groupes de travail, chacun s'occupant d'une étape de ce genre d'études :

- l'acquisition de métadonnées de contexte biologique (c'est-à-dire la conception expérimentale selon laquelle les échantillons biologiques sont générés chez les mammifères, les plantes, les cultures cellulaires, mais également, pour les études de conditions expérimentales),
- les analyses chimiques,
- le traitement des données et études statistiques,
- l'ontologie,
- l'échange de données.

Chaque étape de la préparation d'un échantillon à la publication des résultats est donc encadrée et peut répondre à un cahier des charges bien précis, validant la pertinence scientifique des conclusions qui en sont tirées.

1.2.2 Choix de la méthode analytique

Les échantillons préparés peuvent ensuite être analysés. Or ces échantillons présentent une très grande diversité chimique (diversité structurale, de concentrations et de propriétés physico-chimiques des molécules présentes), leur analyse par l'utilisation d'une unique méthode analytique ne serait pas exhaustive [27]. Plusieurs technologies ont été développées. On pourra citer pour les techniques les plus utilisées, la spectrométrie de masse [68] parfois couplée à des techniques séparatives (LC ou GC) et la Résonance Magnétique Nucléaire [69].

La RMN possède de nombreux avantages pour la métabolomique. Sa robustesse et sa reproductibilité sont très appréciées [70] ; elle est également quantitative et fournit de multiples informations pouvant mener à l'élucidation structurale des métabolites [71]. Son principal désavantage est son manque de sensibilité car elle se limite à la détection des métabolites les plus abondants ($\geqslant$1 µM). De plus, une analyse de routine peut nécessiter une quantité importante d'échantillon : jusqu'à 600 µL pour une sonde

de 5 mm.

La spectrométrie de masse possède, quant à elle, une sensibilité bien plus élevée (jusqu'au femtomolaire) pour des volumes échantillons analysés de l'ordre du microlitre. En revanche, cette instrumentation ne permet pas une détection non discriminante des composés : en effet, en mode d'ionisation positif, la molécule devra être basique en phase gazeuse pour être détectée, mais acide en mode d'ionisation négatif, afin de pouvoir libérer un proton lors de l'ionisation. De plus, l'analyse d'un mélange complexe de composés va être impactée par l'ionisation compétitive des analytes, introduits simultanément dans l'instrument. Cela peut conduire à l'ionisation préférentielle de certaines molécules et à une perte de sensibilité voire à la non-détection d'autres molécules, ce que l'on appelle suppression de signal, ou effet matrice [72]. Cela va alors réduire la quantité de métabolites détectés et créer une image incomplète du métabolome. Par conséquent, la chromatographie en phase liquide (LC) ou la chromatographie en phase gazeuse (GC) sont couramment utilisées couplées à la spectrométrie de masse afin de minimiser cet effet de suppression de signal.

Au cours des dernières années, les progrès techniques dans le domaine des aimants à champ élevé, des séquences d'impulsions et de la technologie des cryosondes ont permis d'améliorer sensiblement la sensibilité et la résolution des expériences de RMN [69]. Ceci, à son tour, a amélioré la qualité des données métabolomiques et a contribué à une augmentation des études métabolomiques basées sur la RMN. En dépit de ces progrès, l'exploitation des spectres RMN ^{1}H (1D) est encore limitée par une importante quantité de chevauchement des pics, même avec les aimants à champ élevé, en raison de la dispersion chimique limitée des métabolites [73].

Comme nous l'avons évoqué plus haut, la combinaison des deux méthodes analytiques peut s'avérer un choix judicieux dans une étude métabolomique. C'est notamment le cas lors de travaux d'élucidation structurale de métabolites inconnus. Cette année, Boiteau *et al*. [74] ont utilisé une approche basée sur l'utilisation simultanée de ces techniques analytiques afin de procéder à l'élucidation structurale d'un métabolite inconnu. Celui-ci est issu d'un extrait d'*Arabidopsis thaliana* et l'interrogation des bases de données pour son identification, à partir de sa masse précise et de sa signature MS^2, n'a pas donné de proposition structurale pertinente. Tout d'abord, la formule chimique du composé est déterminée grâce aux informations fournies par la spectrométrie de masse (masse exacte et profils isotopiques). Des expériences de fragmentation et une analyse RMN 2D ^{13}C-^{1}H HSQC (étape nécessitant une étape de purification du métabolite) ont été effectuées. Ces données ont alors été comparées à des simulations de spectres de fragmentation et de RMN 2D ^{13}C-^{1}H HSQC des structures envisagées, permettant de proposer une structure finale pour le métabolite. L'hypothèse de structure émise a été validée en comparant les spectres RMN du métabolite inconnu, à ceux du standard commercial acheté. Cette molécule, encore non référencée dans les bases de données a été nommée glucoraphanine. Ils montrent ainsi que l'élucidation de structure d'un produit naturel inconnu est

grandement facilitée en combinant les informations obtenues par RMN et par la spectrométrie de masse. Bien que la majorité des études métabolomiques utilise une seule méthode analytique, on comprend de mieux en mieux la valeur inhérente de la combinaison de la RMN et de la MS pour la métabolomique.

Au laboratoire, nous utilisons la RMN dans le but de procéder à l'élucidation structurale de métabolites conjointement à la spectrométrie de masse. Cependant, le développement de l'approche métabolomique du laboratoire s'est fait uniquement à partir de la spectrométrie de masse. Nous présenterons donc dans le paragraphe ci-après, les spécificités de cette instrumentation.

1.2.3 La spectrométrie de masse

La spectrométrie de masse est une technique d'analyse qui permet la détermination des masses moléculaires des composés analysés ainsi que leur identification. Elle est fondée sur la détection d'ions en phase gazeuse, formés à partir des molécules d'un échantillon dans une source d'ionisation. Le spectromètre de masse est composé d'une source où se produit l'ionisation des composés introduits, d'un analyseur qui sépare les ions formés selon leur rapport masse sur charge (m/z), et d'un détecteur comptabilisant l'abondance des ions. Un traitement informatique permet de visualiser les résultats sous forme de spectres de masse représentant l'abondance des ions en fonction de leur rapport m/z.

Il existe à ce jour plusieurs types de spectromètres de masse. Je présenterai ici l'instrumentation utilisée dans ce projet : le Velos Pro Orbitrap Elite.

1.2.3.1 Le Velos Pro Orbitrap Elite

L'instrument utilisé dans ce projet est un spectromètre de masse de type piège 2D couplé à un Orbitrap. Il se présente en deux parties : le Velos Pro MS, et l'analyseur Orbitrap (Figure 7).

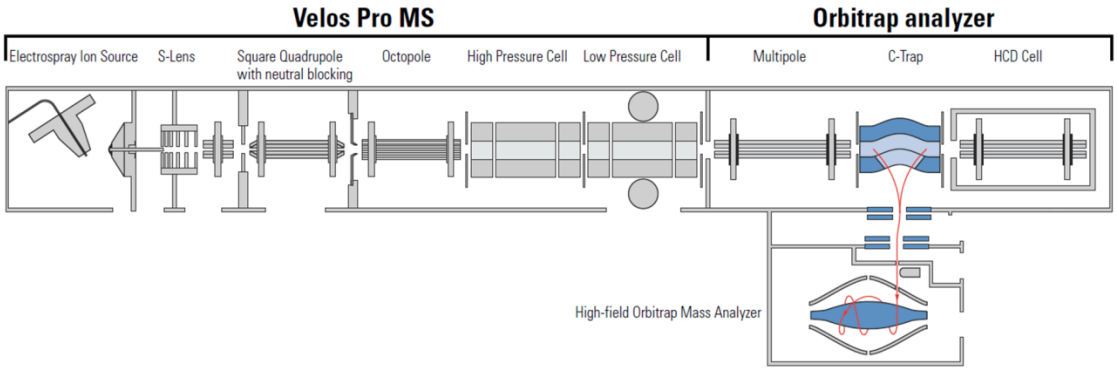


Figure 7: Schéma du spectromètre de masse Velos Pro Orbitrap Elite (d'après http://www.thermofisher.com/, juillet 2018)

1.2.3.1.1 Source HESI II

Depuis plusieurs années, les nouvelles techniques d'ionisations douces ont démocratisé l'utilisation des spectromètres de masse. Ces techniques à pression atmosphérique ont permis le couplage facile avec la chromatographique liquide (LC/MS) pour l'analyse de biomolécules. Le Velos Pro Orbitrap Elite possède une source ESI chauffée (HESI).

La source d'ionisation par électrospray ESI permet de protoner $[M+H]^+$ ou déprotoner $[M-H]^-$ les molécules selon la polarité de l'ionisation. Ce type d'ionisation conserve l'intégrité de la molécule de départ, sous forme d'ions pseudo-moléculaires et permet d'accéder à sa composition élémentaire. Cependant, la formation d'ions adduits et de dimères tels que $[2M+H]^+$,$[2M-H]^-$, $[M+Cl]^-$, $[M+Na]^+$, $[M+K]^+$, $[M+NH_4]^+$ ou encore $[M-H+Mg]^+$ et $[M-H+Ca]^+$ est à considérer attentivement [75]. Les solvants, tampons, pH utilisés ainsi que le pKa des analytes en solution et leur affinité protonique en phase gazeuse influencent la formation de ces dimères et ions adduits [76].

Le phénomène ESI se produit sous l'effet d'un fort champ électrique et peut se résumer en trois étapes :

Étape 1 : Production de gouttelettes chargées à partir de l'analyte en solution sous l'effet combiné d'un fort potentiel électrique entre le capillaire et la contre-électrode et d'un flux de gaz inerte (gaz gaine, « sheath gas ») ajouté le long du capillaire. Les charges s'accumulent à la surface du liquide pour former un fin brouillard de gouttelettes chargées.

Étape 2 : Fission des gouttelettes chargées par explosions coulombiennes. Les gouttelettes rétrécissent par évaporation du solvant assistée par un flux d'azote à contre-courant du nuage d'ions jusqu'au moment où les forces coulombiennes vont approcher le niveau des forces de cohésion des gouttelettes et ainsi provoquer leur explosion.

Étape 3 : Transfert des ions désolvatés en phase gazeuse vers l'analyseur.

1.2.3.1.2 Piège 2D (Velos Pro)

Les ions formés dans la source électrospray sont conduits par un capillaire de transfert vers une série de lentille (S-lens) afin de les focaliser et les concentrer *via* l'octopole jusqu'au piège à ions (Figure 8).

Ce dernier est composé de deux cellules quadripolaires linéaires identiques séparées par une lentille centrale. Chacune des deux cellules est composée de barres hyperboliques segmentées en trois sections : deux électrodes terminales (end caps) et une électrode centrale qui comporte quatre barres ayant chacune une fente permettant l'éjection radiale des ions. Il permet différents types d'analyses en commençant par le spectre de masse global, où tous les ions sont collectés dans le piège puis éjectés pour donner un spectre représentant l'abondance de chaque ion entré dans le piège.

Figure 8: Le piège à ions 2D (Velos Pro) (d'après http://www.thermofisher.com/, juillet 2018)

Chacune des deux cellules a une fonction particulière. La première cellule est à haute pression. Elle est utilisée à une pression d'hélium de 6,6 x 10^{-6} bar tandis que la seconde, la cellule à basse pression, est à une pression inférieure d'hélium (3,9 x 10^{-7} bar). Le piège à haute pression d'hélium est nécessaire pour l'injection, le piégeage, le refroidissement des ions, et la fragmentation (mode CID). Les ions d'intérêt, (ou les fragments) sont ensuite transférés dans la seconde cellule basse pression qui permettra l'éjection des ions, minimisant les collisions (et donc les fragmentations dues à celles-ci) durant ce processus. L'utilisation de ce système de double piège permet d'obtenir un isolement plus rapide et une capture plus efficace, que les pièges à ions linéaires classiques. Un gaz vecteur, l'hélium, est utilisé dans ces pièges pour refroidir, fragmenter, et piéger les ions.

Fragmentation CID et HDC

Comme nous l'avons précisé précédemment, il est possible d'effectuer des expériences MS^n séquentielles où certains ions seront soumis aux étapes d'isolement et de fragmentations successives : il s'agit de la fragmentation CID [77]. Il s'agit d'une excitation résonnante : les ions seront d'abord stockés sans discrimination de m/z, puis un seul ion sera isolé dans la première cellule pour être ensuite activé (application d'une tension radiofréquence spécifique à l'ion d'intérêt, faisant entrer l'ion en résonnance dans le piège) pour se dissocier en ions-produits lors de ses collisions avec le gaz vecteur (hélium). Ces ions-produits seront alors éjectés pour être détectés. La détection des ions fragments peut être réalisée directement dans le piège à ions (basse résolution et précision en masse), ou bien dans l'Orbitrap (haute résolution et précision en masse). En mode CID, il est possible de réaliser des fragmentations successives en ré-isolant l'un des fragments obtenus au fur et à mesure, afin de pouvoir procéder ensuite à leur propre fragmentation.

Il existe un autre mode de fragmentation avec l'Orbitrap Elite : le mode HCD (« Higher-energy Collision induced Dissociation ») [78] dont le mode d'excitation est « non résonnant » contrairement à la fragmentation dans le piège à ions. Cette fragmentation se produit dans une cellule de collision

spécifique (un octopole dénommé cellule HCD), et non dans le piège à ions. Les ions sont accélérés dans le piège grâce à l'application d'une tension DC (Direct Current), et leurs collisions avec les particules de gaz (azote) provoqueront leur fragmentation. Une fois fragmentés, ils sont alors transférés dans l'Orbitrap pour détection en haute résolution, à travers la C-trap. Contrairement à la fragmentation CID, les expériences de fragmentation MS^n séquentielles dans la cellule HCD sont impossibles car il n'est pas possible de procéder à l'isolement de fragments obtenus.

Dans le cas de fragmentations effectuées en mode CID, il existe un seuil de détection de masse en deçà duquel les fragments de masse plus faible ne seront pas détectés. Ce phénomène est appelé le « low-mass cut-off » et dépend de la région de stabilité des ions dans le piège. Les valeurs m/z les plus basses détectables sont d'environ 1/3 de la valeur m/z de l'ion parent. Cependant, ce phénomène n'existe pas dans la cellule HCD [78] : il est ainsi possible d'observer des fragments plus petits qu'en CID, ce qui permet d'obtenir plus d'informations structurales à partir des fragmentations d'un composé. Par ailleurs, le mode de fragmentation HCD utilise également des énergies de dissociation plus élevées que celles utilisées en CID, permettant une plus large gamme de voies de fragmentations.

1.2.3.1.3 L'Orbitrap Elite

L'Orbitrap est située après le piège à ions linéaire et la C-trap. Les ions sont transférés dans la C-trap, qui sert à concentrer et injecter les ions vers l'Orbitrap. Les ions doivent être injectés rapidement et de manière la plus compacte possible avant d'entrer dans l'analyseur Orbitrap pour être séparés selon le rapport masse/charge (m/z).

L'Orbitrap est composé de deux électrodes (Figure 9) : une électrode centrale en forme de fuseau et une électrode extérieure en forme de tonneau séparée en deux par un fin anneau.

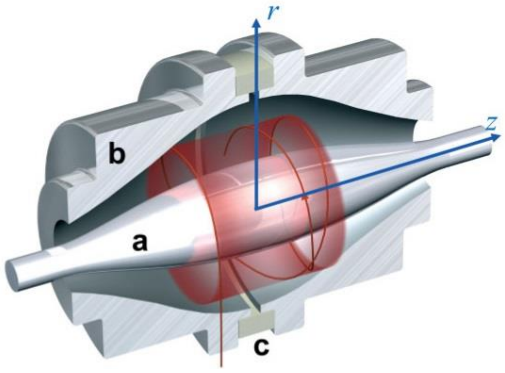


Figure 9 : Schéma de l'analyseur Orbitrap (d'après http://planetorbitrap.com, juillet 2018)

Leur trajectoire peut se décomposer en deux mouvements : un axial, le long de l'électrode central et un autre mouvement de rotation autour de cette même électrode [79]. L'application d'un potentiel

logarithmique par rapport au fuseau permet un piégeage orbital des ions dans le plan radial. Un potentiel quadripolaire par rapport au fuseau confine les ions axialement. La trajectoire de l'ion selon l'axe z suit une loi d'oscillateur harmonique dont la fréquence des oscillations dépend uniquement de la masse et de la charge de l'ion oscillant. Ainsi, à chaque passage devant les électrodes externes, les ions génèrent des courants induits sur lesquels une transformée de Fourier est appliquée pour déconvoluer les différentes fréquences d'oscillation axiale des ions ; elles seront finalement converties en rapports m/z précis. Le mouvement de rotation, permet quant à lui de piéger les ions dans le plan radial et ne dépend, pour un champ électrique donné, que de l'énergie cinétique de l'ion entrant.

Par rapport à l'Orbitrap XL, précédemment utilisé au laboratoire, l'Orbitrap Elite permet une montée en gamme en termes de précision en masse et de résolution. De façon courante nous obtenons des résolutions maximales à 240 000 à comparer avec les 100 000 de résolution de l'Orbitrap XL avec des fréquences d'acquisition plus courtes pour une même résolution.

1.2.3.1.4 Caractéristiques des analyseurs

Le spectromètre de masse se caractérise par ses performances analytiques. Celles-ci sont évaluées à partir de cinq paramètres principaux:

La résolution R : Le pouvoir de résolution de masse est la capacité d'un spectromètre de masse à séparer des rapports m/z proches. Il est défini de la manière suivante :

$$FWHM\ (full\ Width\ at\ Half\ Minimum) = \frac{m}{\delta m}$$

où m est le rapport masse/charge (m/z) mesuré d'un pic d'intérêt et δm la largeur à mi-hauteur de ce même pic. En fonction de cette valeur, les analyseurs seront placés dans le groupe des appareils à basse résolution, haute résolution (FWHM>10 000) voire à très haute résolution (R> 50 000).

La précision en masse (δppm) : correspond à la différence mesurée entre le m/z expérimental d'un ion (m/z_{exp}) et son m/z théorique (m/z_{th}) calculé à partir des masses atomiques précises de chaque atome ;

Elle se définit de la manière suivante :

$$\delta ppm = \frac{(m/z_{exp} - m/z_{th})}{m/z_{th}} \times 10^6$$

Exprimée en partie par million (ppm), elle dépendra uniquement des performances de l'instrument.

La gamme dynamique : est l'étendue des concentrations relatives détectables ; c'est-à-dire le rapport entre la plus grande et la plus petite concentration des composés détectés dans la matrice analysée. Ce

paramètre s'avère crucial pour la détection des ions des composés minoritaires.

Le temps d'acquisition : de l'analyseur correspond à la durée nécessaire à l'analyseur pour couvrir une gamme donnée de rapports m/z. Ce paramètre prend toute son importance lors d'un couplage avec une méthode séparative. En effet, l'échantillonnage d'un pic chromatographique est directement déterminé par le temps d'acquisition ; notamment dans le cadre d'analyses quantitatives, ce paramètre est un facteur crucial.

La gamme de m/z : détermine les limites des rapports m/z détectables par l'instrument. L'IUPAC définit aujourd'hui le rapport m/z sans unité [80].

Les informations présentées dans le tableau 1 montrent bien que chaque instrument présente ses avantages et ses inconvénients qu'il faut considérer en regard du type d'étude menée.

Analyseur	Résolution (FWHM)	Précision en masse (ppm)	Gamme dynamique	Gamme de m/z	Possibilité MSn
Triple quadrupôle (Q-trap 5500 ABSCIEX)	12000	100-500 ppm	4000	5-1250 m/z	MS^2
Piège à ions (Esquire 3000+ Brucker)	1650-27000 (amu/sec)	100-500 ppm	4000	50-3000 m/z	MS^n
Q-TOF (6545 Q-TOF, Agilent)	> 45000 (m/z 2722)	< 1ppm	500	50-1700 ; 100 - 3200 (quadrupole jusqu'à 40 000)	MS^2
Q-exactive (Thermo)	jusqu'à 140 000 (m/z 200)	< 3ppm (si calibration interne 1ppm)	5000	50-6000 m/z	MS^2
LTQ Orbitrap XL (Thermo)	7500 à plus de 100 000 (m/z 400)	< 3ppm (si calibration interne <1ppm)	> 4000	50 à2000 m/z; 200 à 4000 m/z	MS^n
Velos Pro Orbitrap Elite (Thermo)	15 000 à plus de 240 000 (m/z 400) jusqu'à 1 000 000 (m/z <300-400) [80]	< 3ppm si calibration interne < 1 ppm)	> 5000	50 à2000 m/z; 200 à 4000 m/z	MS^n
FT-ICR (15T solariX, Brucker)	30 000 à 10 000 000 (m/z 400)	< 2ppm (si calibration interne <0,25 ppm)	10^3-10^4	50-6000 m/z	MS^n

Tableau 1: Récapitulatif des performances de différents analyseurs utilisés en métabolomique (données constructeurs)

Dans cette étude, nous utilisons un Velos Pro Orbitrap Elite dont nous exploitons à la fois la très haute résolution, la précision en masse et les possibilités de fragmentation pour les différents aspects du projet (élucidation structurale de métabolites et analyse haut débit d'échantillons complexes).

1.2.3.2 Traitement de données

Nous utilisons un spectromètre de masse couplé à une chromatographie liquide. Nous présenterons donc ici les enjeux et méthode de traitement de données acquises sur ce genre d'appareillage.

A chaque instant de l'analyse d'un échantillon, le spectromètre de masse enregistre tous les rapports m/z des espèces ioniques détectées ainsi que l'intensité du signal associé. Toutes ces données peuvent être visualisées de deux manières différentes : le chromatogramme total, ou TIC (Total Ion Current) qui représente l'intensité du courant ionique total en fonction du temps de rétention sur la colonne chromatographique ; ou, extrayant les intensités de signal pour un unique m/z en fonction du temps, le chromatogramme EIC (Extracted Ion Current)

Chaque ion sera caractérisé par son rapport m/z, son abondance et son temps de rétention dans la colonne chromatographique. Le principal objectif du traitement de données est d'extraire de manière exhaustive l'information contenue dans ces fichiers et d'en retirer la redondance. Cet aspect est très bien décrit par

l'étude de Tautenhahn *et al*. [81] En injectant séparément 14 standards de référence sur UPLC/ESI-QTOF, ils obtiennent un total de 296 variables détectées et annotées en tant qu'adduits, isotopes et fragments. Ces résultats démontrent que l'acquisition des données en ESI-MS génère un grand nombre de fausses variables qui complexifient leur analyse. Par exemple les isotopes, adduits, dimères et fragments conduisent à une redondance des données car ils représentent le même métabolite.

Ce traitement de données se fait en plusieurs étapes :

- élimination du bruit de fond
- détection et extraction des pics
- correction des variations des temps de rétention
- identification des composantes : regroupement des signaux ayant un même temps de rétention : adduits, isotopes….)

Il est bien évident qu'il est impossible d'effectuer ce travail de façon manuelle et plusieurs logiciels ont été développés pour le traitement de données (notamment par les constructeurs : Metabolic Profiler pour Bruker ou Sieve pour Thermo). De son côté le SCRIPPS a développé un logiciel gratuit, XCMS Online [82], qui intègre régulièrement de nouvelles fonctionnalités. La plateforme logicielle est accessible en ligne, ne requiert ni installation, ni réelle expertise informatique. L'utilisation de ce logiciel, très intuitif, se résume en trois étapes : téléchargement des données, choix des paramètres - personnalisable selon le type d'instrumentation analytique utilisée - et visualisation des résultats. Cette dernière étape met à la disposition de l'utilisateur un panel d'analyses, de représentations statistiques ainsi qu'une matrice de résultats où chaque ligne correspond à une variable, décrite par son rapport m/z, son temps de rétention et à laquelle sont assignées les aires de pic extraites pour chaque injection.

XCMS Online intègre notamment un outil d'annotation du signal, CAMERA [83], appréciable pour l'analyse des résultats obtenus. En effet, nous avons évoqué précédemment que l'utilisation de sources d'ionisations dites douces provoque la création de fragments, adduits et isotopes, compliquant l'étape d'annotation des variables en métabolites. CAMERA permet de procéder au regroupement des variables observées et correspondant à un même métabolite en repérant les variables provenant d'un même évènement chromatographique, sur la base de la forme de pics chromatographiques. Ainsi, en corrélant les profils spectraux et les coïncidences de temps de rétention au travers des différents échantillons, il est possible de regrouper les variables en « composantes » décrivant tous les ions issus du même métabolite.

1.2.3.3 Identification structurale

Une fois le traitement de données effectué et les variables d'intérêt extraites, il faut, afin de pouvoir les placer dans un contexte biologique, les identifier. Nous traiterons ici de l'identification structurale par

LC/MS.

Les différentes étapes de l'identification structurale sont présentées en Figure 10. L'identification des variables débute par une interrogation des bases de données à partir de la masse précise mesurée (dont la précision sera choisie en fonction de l'instrumentation utilisée). Il en existe plusieurs qui peuvent recenser différents types de métabolites : par exemple LipidMaps [84], plateforme publique qui recense tous les lipides identifiés et HMDB [85], qui recensera les métabolites humains. En revanche, KEGG [86], METLIN [87] et PubChem [88] sont des bases de données recensant tout type de métabolites.

L'interrogation des bases de données pour l'identification des variables ne donne pas systématiquement une unique solution. Ainsi, afin de limiter le nombre de propositions, ou si aucune proposition n'a été faite, il est possible de générer des spectres de fragmentation MS/MS. En effet, certaines pertes peuvent être caractéristiques de la présence de groupes fonctionnels tels CO_2 en mode d'ionisation négative, ou NH_3 en mode d'ionisation positive. Les fragmentations se font le plus souvent en CID, dont les paramètres tels que les énergies de collision, les temps d'activation etc. peuvent être réglés pas l'utilisateur selon l'information recherchée.

Cependant, en LC-MS, l'utilisation de sources d'ionisation douces complique l'étape d'identification. En effet, il a été montré que différents paramètres cinétiques (énergie interne et temps d'activation) et instrumentaux (pression et nature du gaz de collision) influent sur les mécanismes de dissociation observés lors des expériences MS/MS en CID, ce qui rend la comparaison des données brutes au banques de données compliquée [89]. C'est pourquoi, certains laboratoires développent des bases de données internes dans lesquelles les spectres de fragmentation des échantillons sont acquis sur les mêmes instrumentations avec les mêmes paramètres de fragmentation que ceux des standards [90].

Il existe également des bases de données spectrales CID telles que MassBank [91], METLIN [87] ou HMDB [85]. Pour chacun de ces exemples aucun protocole de standardisation n'est appliqué ce qui peut mener à des identifications erronées [91].

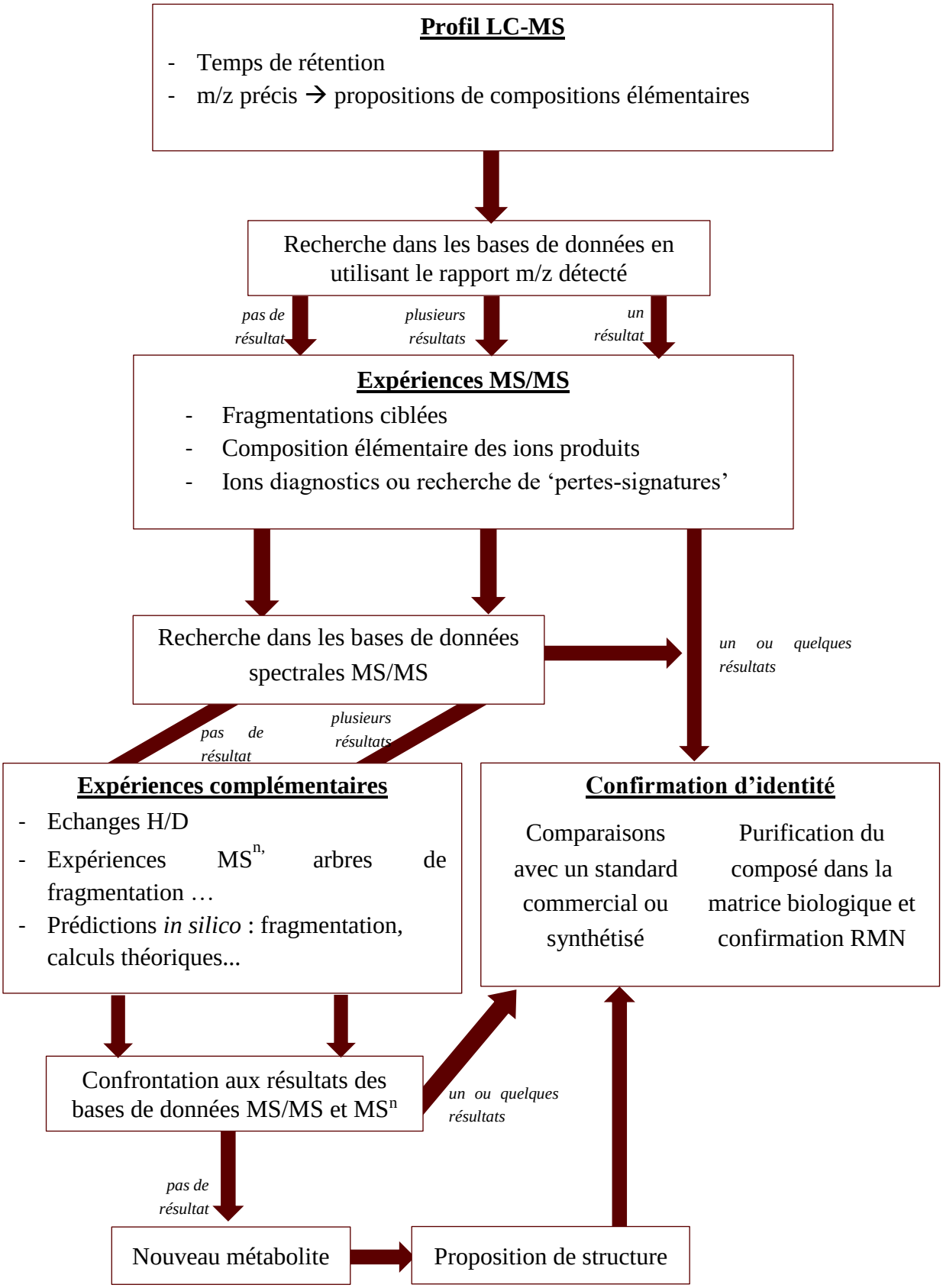


Figure 10: Les différentes étapes de l'identification des métabolites par spectrométrie de masse. Issu de [61]

Toutefois, afin d'augmenter le nombre d'empreintes spectrales disponibles et donc d'augmenter les chances de recoupement, METLIN et HMDB intègrent des spectres CID enregistrés à plusieurs énergies de collisions (faible, moyenne et forte). De plus, à l'initiative du Réseau Français de Métabolomique et de Fluxomique, une procédure de standardisation a été développée, adaptée à l'instrumentation et au mode de fragmentation choisi [91]. Grâce à cela, les capacités de comparaisons de données MS/MS entre laboratoire sont améliorées.

Par ailleurs, des logiciels de prédictions de fragmentation ont été créés, tels que MetFrag [92] (gratuit) ou MassFrontier (commercial). Ainsi, il est également possible de comparer les spectres MS/MS obtenus dans les données brutes avec ceux prédits *in silico* dans ces bases de données. Les deux logiciels possèdent deux approches différentes : les spectres prédits dans MetFrag sont créés à partir de la simulation de ruptures simples. Aucun réarrangement n'est pris en compte, or en phase gazeuse, cela arrive fréquemment [93]. Mass Frontier s'appuie sur les réactions de dissociation, déjà observées dans la littérature, ainsi que des règles générales de réarrangement et de fragmentation. Mais aucune de ces règles ou de ces algorithmes n'est explicitée.

Le partage et la standardisation des données a toute son importance car plus de données seront disponibles, plus l'identification des métabolites d'un échantillon sera possible et facilitée. En 2012, deux initiatives ont vu le jour. La première est Metabolights [94], une plateforme de dépôt de données de métabolomique. La seconde, COSMOS [95] regroupe 14 partenaires européens et a pour objectif de centraliser les données expérimentales en métabolomique sous un format libre et gratuit, notamment en utilisant Metabolights et en s'assurant que les recommandations de la MSI sont suivies.

L'identification d'un métabolite par LC/MS est un processus complexe. Pour s'assurer de la qualité des résultats, la MSI (Metabolomic Standard initiative) a proposé en 2007, une liste de renseignement à fournir dans le cas d'une publication de métabolomique, notamment pour l'identification de métabolites [67]. Quatre niveaux d'identifications y sont notamment proposés : du niveau 1, où deux propriétés physico-chimique indépendantes, obtenues dans les même conditions d'analyse entre le métabolite à identifier et un standard sont identiques (par exemple, en LC/MS, temps de rétention et masse précise mesurée), jusqu'au niveau 4 décrivant l'état d'identification d'un métabolite détecté et possiblement quantifié, mais où aucune correspondance avec les caractéristiques d'un métabolite ou une famille de métabolite n'a pu être obtenue. Cependant, comme le soulignent les auteurs en 2014, cette classification a ses limites, notamment lors de l'identification de différents stéréo-isomères, difficilement distinguables lors de profilage métabolique [96]. Plusieurs propositions sont encore actuellement débattues. L'une d'entre elles serait d'ajouter un score d'identification.

1.3 Acinetobacter baylyi ADP1 au laboratoire

1.3.1 Le génome

Acinetobacter baylyi ADP1 (ADP1) appartient à la classe des γ-protéobactéries, de l'ordre des Pseudomonadales et au genre *Acinetobacter*. Les espèces de ce genre sont Gram négatif, peuvent être pathogènes et multirésistantes, comme *Acinetobacter baumanii*, un pathogène bien connu du milieu hospitalier car responsable de maladies nosocomiales. ADP1 est en revanche un coccobacille du sol, non pathogène, aérobie strict dont la température d'incubation est comprise entre 30°C et 37°C.

Bien que plus de 80% des sujets de recherche sur ce genre concernent justement *A. baumanii*, ADP1 possède deux propriétés particulièrement utiles pour les manipulations génétiques : elle est naturellement compétente (elle peut intégrer des parties entières d'ADN disponible dans son environnement) et fait spontanément la recombinaison homologue (elle peut échanger et intégrer dans son propre génome des parties d'ADN identiques ou similaires). Il est donc possible d'éliminer, d'ajouter ou encore de muter un gène de cette souche. Ces qualités offertes par ADP1 pour sa manipulation génétique en font un modèle idéal pour l'annotation fonctionnelle et l'étude du métabolisme, ou encore comme châssis potentiel pour des applications en biologie synthétique.

Séquencée au Genoscope, elle fut le premier génome publié du genre Acinetobacter. Cette souche a également été annotée par une équipe de notre centre de recherche [97]. Son génome de 3,7 Mb, porté par un chromosome circulaire, présente un métabolisme plus simple et sans de multiples redondances de fonction. Ce dernier point est particulièrement important, car il rend possible le suivi de phénotypes suite à la délétion d'un ou plusieurs gènes, évitant de masquer l'adaptation de la bactérie par la présence d'un gène paralogue iso-fonctionnel.

Cette bactérie présente des capacités métaboliques importantes. En effet, 38% des gènes de fonction connue ou supposée codent pour des enzymes ou des transporteurs. Ainsi, plusieurs îlots génomiques dédiés au catabolisme d'une grande variété de composés organiques (cafféate, férulate, quinate…) ont été mis en évidence [98]. Grâce à cette caractéristique, ADP1 peut s'adapter à son milieu, et assimiler des produits de dégradation des végétaux.

1.3.1.1 Construction de ressources pour l'annotation fonctionnelle et l'étude du métabolisme.

Plusieurs ressources ont été créées au Genoscope afin de participer aux efforts de développement de méthodes pour l'annotation fonctionnelle et l'étude du métabolisme de la bactérie.

1.3.1.2 Banque de mutants de délétion

L'une des méthodes pour étudier la fonction d'un gène est d'observer les réactions (phénotypiques ou métaboliques) de la bactérie suite à une perturbation génétique (mutation ou suppression d'un ou plusieurs gènes). Ainsi en 2006, l'université de Keio a créé une banque complète de mutants de délétion chez *E. coli*.

De Berardinis et al. [22] ont créé en 2008 une collection complète de mutants de délétion chez ADP1. Le laboratoire a utilisé la capacité d'ADP1 à faire de la recombinaison homologue pour procéder à l'intégration dans la bactérie d'une cassette portant un gène de résistance à un antibiotique, la kanamycine. Le gène de résistance est flanqué de deux pieds de recombinaisons correspondant à environ 300 pb des régions entourant le gène spécifique à éliminer (Figure 11).

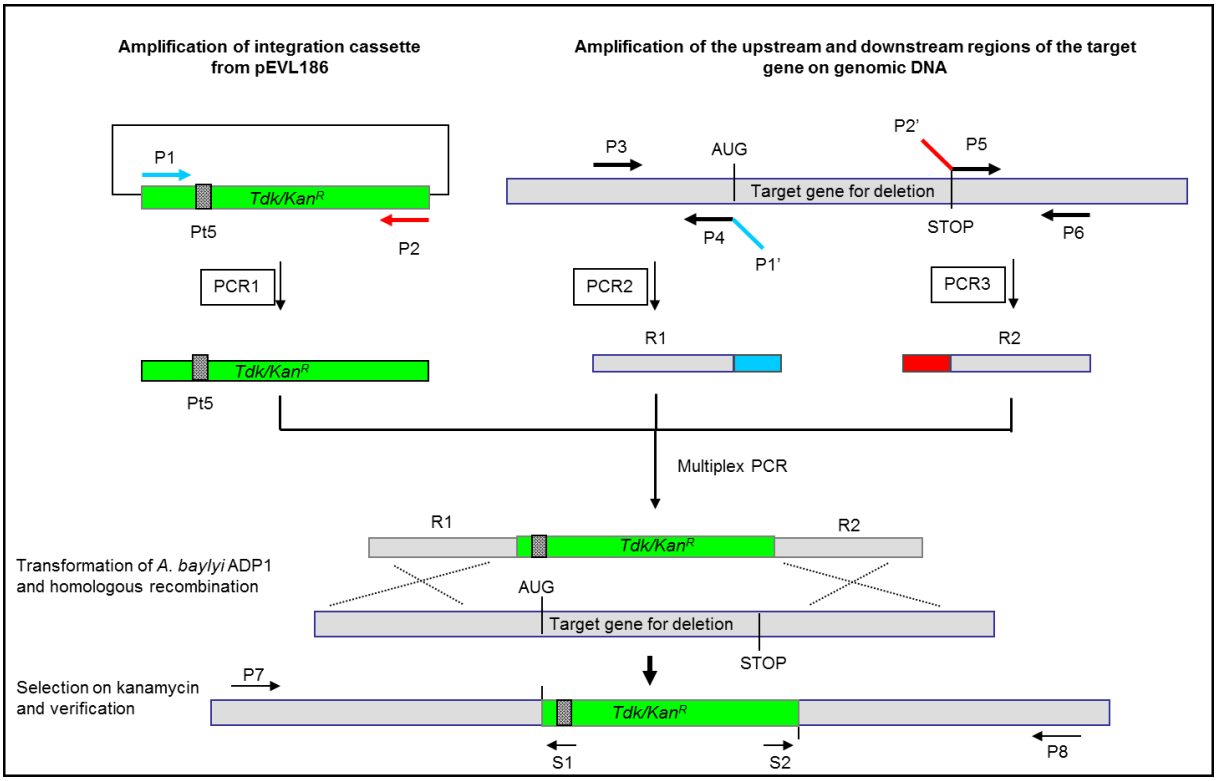


Figure 11: Méthode de construction de mutants de délétion d'ADP1 par substitution du gène cible par une cassette d'intégration contenant un gène de résistance à la kanamycine.

Au total, sur les 3197 CDS validés chez ADP1, 2594 mutants de délétion ont été obtenus. Parmi les gènes qui ont pu être délétés, plus d'un millier a une fonction totalement inconnue.

Il est important de noter que 499 gènes n'ont pas pu être éliminés chez ADP1 malgré plusieurs tentatives du fait de leur potentielle essentialité. Les mutants ayant été sélectionnés sur milieu minimum, le nombre de gènes essentiels n'aurait pas été aussi important si cette sélection avait été effectuée sur milieu riche.

En revanche, pour 41% des gènes prédits comme essentiels à la survie de la bactérie, les mutants ont pu

être obtenus. Il a alors été mis en évidence la présence de duplication du gène ailleurs dans le génome juste avant sa substitution par la cassette. Ce phénomène a été révélé par des PCR pour lesquelles les amorces avaient été choisies en dehors des régions potentielles de recombinaison. Ces mutants appelés « double bande » présentaient en effet deux produits de PCR : le premier pour la région où le gène cible avait été remplacé par la cassette, le second, étant une bande de même taille que la région sauvage.

Cette banque de mutants a déjà été utilisée à de nombreuses reprises. En effet, cette collection a été phénotypée sur un ensemble de sources de carbone (quinate[99, 100], 2-3 butanediol [98], glucose [98], glucarate/galactarate [23], trigonelline [26]) et d'azote [101] afin de mettre en évidence les gènes impliqués dans le catabolisme chez ADP1.

1.3.1.3 Création de l'ORFéome

Afin de mettre en place une approche complémentaire à la banque de mutants, un ORFéome a été construit. Les gènes d'ADP1 (hors protéines membranaires, ou enzymes liées à la réplication de l'ADN ou à la réplication cellulaire, soit près de 2200 gènes) ont été clonés dans un vecteur d'expression modifié de type pET22 avec une étiquette poly-histidine dans la partie N-terminale de la protéine d'intérêt, pour aider à la purification.

Ainsi, dans l'objectif d'aider à l'annotation fonctionnelle, il est possible grâce à cette ressource de tester biochimiquement la fonction de l'ensemble des protéines d'ADP1. Cette ressource peut être utilisée pour rechercher des candidats à une activité enzymatique d'intérêt. Les protéines candidates peuvent par exemple être incubées en présence de substrats définis et le suivi d'activité indique ainsi quelles sont les protéines qui catalysent la réaction d'intérêt. Le suivi peut se faire au format 96 puits, par spectrophotométrie ou LC/MS. Ce genre d'approches a notamment été utilisé par Perchat *et al*. où les protéines impliquées dans la voie de dégradation de la trigonelline avaient été purifiées, afin de pouvoir procéder à la reconstruction de la voie de dégradation de cette source de carbone et d'azote [26].

1.3.1.4 Modèle informatique de la bactérie

Enfin, pour compléter les outils à notre disposition au laboratoire, une dernière ressource a été développée. Bien que la fonction de nombreux gènes d'ADP1 soit inconnue, les efforts d'annotation et les études biochimiques publiées par divers laboratoires depuis plusieurs dizaines d'années ont contribué à améliorer significativement la connaissance du métabolisme de la bactérie. En combinant toutes ces données, un modèle mathématique simulant le métabolisme d'ADP1 a été construit en 2008 [102]. Ce dernier a été affiné avec les données expérimentales obtenues grâce au phénotypage de croissance de la banque de mutants. Des informations sur l'essentialité des gènes ou la capacité de croissance des mutants sur différentes sources de carbone ou d'azote ont été intégrées au modèle.

1.3.2 Le métabolisme du quinate reconsidéré par des approches globales

ADP1 peut utiliser une très large gamme de composés aromatiques [98]. Leur dégradation passe par la formation de catéchol ou de protocatéchuate avant d'emprunter la voie du β-cétoadipate et rejoindre le cycle de Krebs. A partir du catéchol et du protocatéchuate, deux voies parallèles convertissent ces composés en succinyl-CoA et acétyl-CoA (Figure 12). D'un côté, la plupart des gènes impliqués dans la dégradation des composés aromatiques qui aboutissent au catéchol (voie du catéchol) est colocalisée avec ceux qui dégradent le catéchol et forment un premier ilot catabolique (*ACIAD1424-1451*). D'un autre côté, la majorité des gènes qui participent au métabolisme des substrats aromatiques pour former du protocatéchuate (voie du protocatéchuate) est aussi colocalisée avec ceux qui dégradent le protocatéchuate et forment un autre ilot catabolique (*ACIAD1702-1728*).

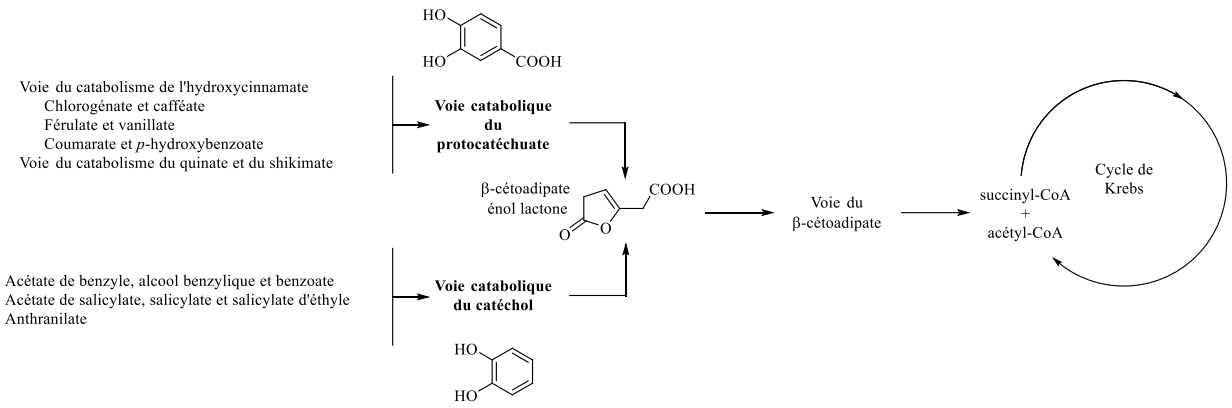


Figure 12 : Dégradation des composés aromatiques via les deux voies menant à la voie du β-cétoadipate chez ADP1 [103]

Le quinate est un composé alicyclique qui emprunte la branche du protocatéchuate *via* l'opéron *qui*. Au total, 14 gènes sont impliqués dans le catabolisme du quinate ; trois codent pour deux transporteurs et un régulateur, les onze autres pour des enzymes. Chacun des gènes et des intermédiaires de dégradation est connus [98] (Figure13).

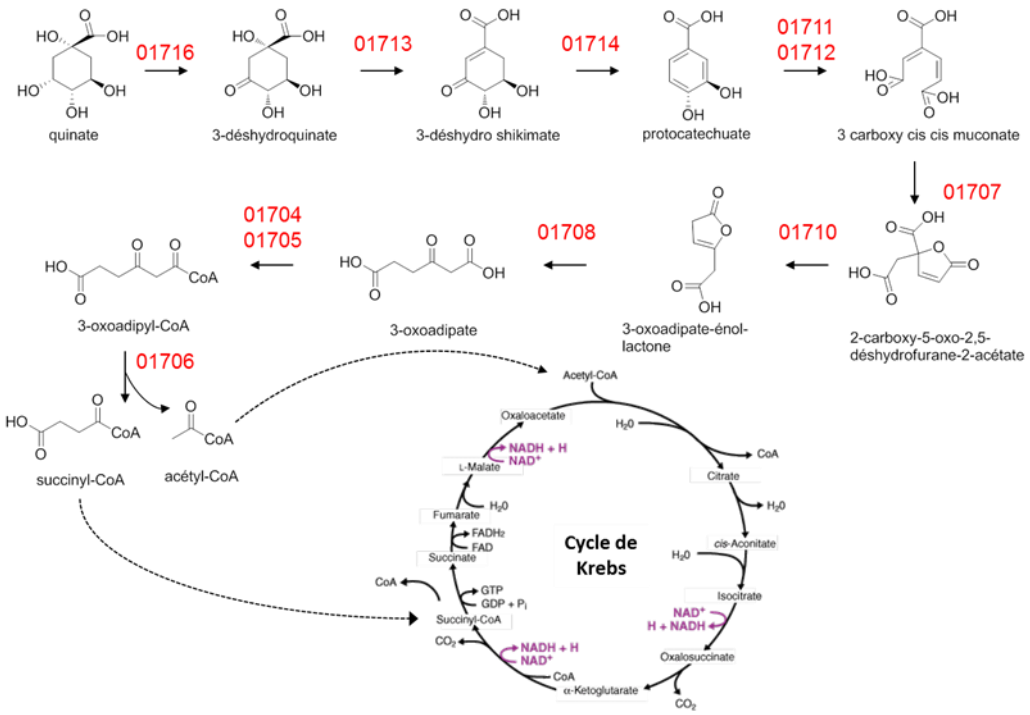


Figure 13: Métabolisme du quinate chez ADP1. Les chiffres en rouge indiquent le nom des gènes (ACIAD) chez ADP1. Issu de [61]

La voie de dégradation du quinate est donc bien documentée du point de vue génétique et biochimique. Cette voie métabolique n'avait cependant été étudiée que par des approches ciblées. Le laboratoire a supposé qu'une approche globale et sans a priori comme la métabolomique pourrait lui donner un éclairage nouveau. Le laboratoire s'est donc intéressé à suivre la réorganisation du métabolisme d'ADP1 après un changement de source de carbone (du succinate vers le quinate).

1.3.2.1 Une réponse spécifique attendue

L'analyse métabolomique a confirmé que les catabolites du quinate sont uniquement présents dans les métabolomes de cellules utilisant le quinate. Une analyse de transcriptomique par RNAseq avait également été menée au laboratoire quelques années auparavant sur des cellules d'ADP1 qui utilisent le succinate ou le quinate. Celle-ci a montré comme attendu que l'utilisation du quinate induit la surexpression des gènes impliqués dans sa dégradation.

De même l'analyse des métabolomes intracellulaires d'ADP1 en LC-MS dans les deux conditions (quinate et succinate) a montré la présence des intermédiaires de réaction de la dégradation du quinate.

Cependant, chacune des études a donné des résultats inattendus.

1.3.2.2 Une réponse globale inattendue

Le résultat marquant de cette étude est que ce changement de source de carbone a impacté en

profondeur le métabolisme global de la cellule. Près de 200 métabolites sont plus concentrés dans les cellules utilisant le quinate et une centaine d'entre eux a même un rapport de concentration supérieur à 10. Parallèlement, l'analyse transcriptomique globale a montré que l'expression de plus de 400 gènes (soit 12% du nombre total de gènes) est modifiée d'un facteur au moins supérieur à 3 (Figure 14). Une partie importante de ces gènes est de fonction inconnue.

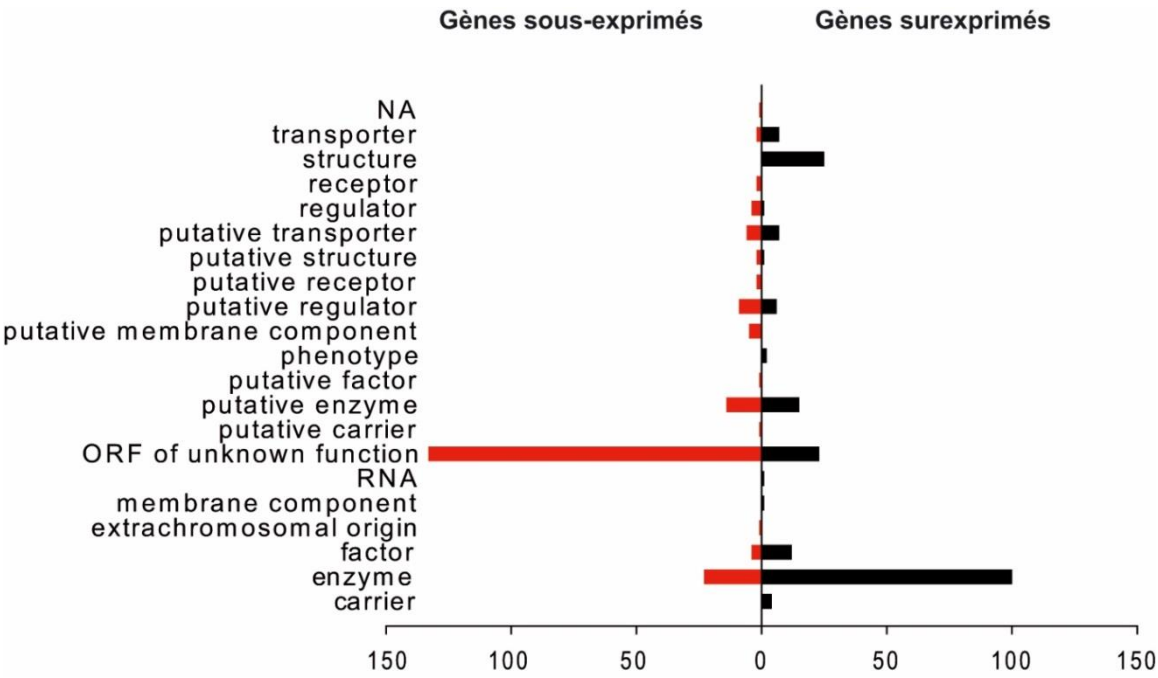


Figure 14 : Nombre et classification fonctionnelle COG des gènes différentiellement exprimés (rapport d'expression>3) lorsqu'ADP1 utilise du quinate comme source de carbone. Issu de [104]

1.3.2.3 De nouveaux métabolites?

L'analyse globale du métabolome d'ADP1 dans les deux conditions de culture a permis de détecter 450 métabolites potentiels. Une identité putative n'a pu être proposée de façon automatique (en interrogeant simplement des bases de données) que pour 250 d'entre eux. Pour 100 autres, leur masse précise ne correspond à aucun métabolite répertorié. Enfin, 105 métabolites ont été identifiés à l'aide de standards commerciaux (comparaison des masses précises, du temps de rétention chromatographique et du profil de fragmentation MS/MS).

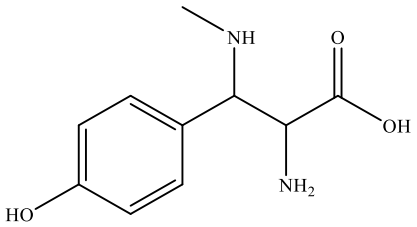
Une trentaine de composés non identifiés est produit exclusivement dans les cellules qui utilisent le quinate. Les deux tiers n'ont pas d'identité proposée par les bases de données. Il est raisonnable de supposer que lors de la réorganisation transcriptionnelle causée par l'utilisation du quinate, des activités enzymatiques insoupçonnées conduisent à la formation de ces métabolites nouveaux, non reliés à la connaissance actuelle du métabolisme.

Par ailleurs, parmi les métabolites identifiés, quinze ne sont pas prédits par le modèle métabolique

d'ADP1. Comme 30% des gènes d'ADP1 sont de fonction totalement inconnue, le modèle métabolique n'intègre probablement pas toutes les capacités enzymatiques de la cellule. Ces métabolites identifiés, également, ne sont pas reliés à la connaissance du métabolisme d'ADP1.

Caractérisation d'un métabolite orphelin

Le laboratoire a alors tenté d'identifier par spectrométrie de masse un des métabolites inattendus produits exclusivement sur quinate. Celui-ci est détectable à la fois en mode d'ionisation positif (m/z211,10704 forme protonée) et négatif (m/z209,09207 forme déprotonée). Le signal, intense en mode positif, est un avantage pour effectuer des expériences MS^n séquentielles, mais aussi pour utiliser son rapport isotopique (entre ^{12}C et ^{13}C) afin de déterminer son nombre d'atomes de carbone et aider à établir sa composition élémentaire. Celle-ci ($C_{10}H_{14}N_2O_3$) a été établie en se basant sur la précision en masse de l'Orbitrap, sa signature isotopique et l'utilisation des « Seven Golden Rules » [105]. Des expériences CID-MS^n séquentielles ont été effectuées sur cette molécule (jusqu'en MS^5) avec une détection à haute résolution. Leur interprétation, associée aux informations obtenues lors d'expériences d'échanges isotopiques H/D, a conduit à proposer comme structure une β-(N-méthylamine)-tyrosine (BMAT) (Figure 15).



$C_{10}H_{14}N_2O_3$
210,10044
BMAT β-(*N*-méthylamine)-tyrosine

Figure 15: Structure précédemment proposée pour la forme neutre du MO211 : β-(N-méthylamine)-tyrosine (BMAT).

Le standard a été synthétisé par voie chimique au Genoscope. Malheureusement, ni son temps de rétention sur la colonne chromatographique ni son profil de fragmentation ne correspondent à ceux de la molécule biologique, ce qui a invalidé son identité.

1.4 Explorer la face cachée du métabolisme d'Acinetobacter baylyi ADP1

Mon travail de thèse s'inscrit directement dans l'étude globale du métabolisme du quinate par ADP1. D'une part, celle-ci a montré que des métabolites non identifiés et non attendus sont produits majoritairement voire exclusivement dans les cellules utilisant le quinate. D'autre part, certains des métabolites identifiés ne sont pas prédits par le modèle métabolique d'ADP1. Ces deux groupes constituent donc des « Métabolites Orphelins » (MO) que nous nous proposons de rattacher à des gènes.

L'objectif de mon travail de thèse était de commencer à mettre au jour les chemins métaboliques impliquant ces composés. Ceci suppose *in fine* d'identifier les gènes impliqués et de préciser leur fonction, d'établir la structure des composés inconnus ainsi que celle de tous les précurseurs de synthèse des deux groupes de MO (métabolites d'identité connue et inconnue).

Mon travail comprend deux parties. La première a consisté à établir la structure du MO de valeur m/z 211,10704 (MO211). Son identification a impliqué sa purification par chromatographie liquide et sa caractérisation par RMN. L'interprétation des données de RMN a permis de proposer une identité potentielle. Afin de valider celle-ci, la synthèse du standard a été effectuée. L'identité du MO211 a pu être définitivement établie en évaluant d'une part les spectres RMN du standard avec ceux de la molécule biologique et d'autre part en comparant leur signature LC/MS/MS à très haute résolution. Nous travaillons actuellement à l'élucidation structurale d'un second MO, de valeur m/z =193,0966, détecté en mode d'ionisation positif.

Dans la seconde partie de ma thèse, toujours en cours, nous utilisons la collection de mutants de délétion d'ADP1 pour repérer les gènes impliqués dans la formation des MO. Les 2600 mutants sont cultivés en présence de quinate et leur métabolome est analysé par LC/MS. Nous regardons, pour chaque mutant, quels sont les MO présents et absents. L'absence d'un MO chez un mutant donné doit indiquer que le gène excisé est impliqué dans sa synthèse. L'analyse systématique de la banque permettra de dresser la liste de tous les gènes impliqués dans la synthèse de chaque métabolite d'intérêt.

2 MATÉRIEL ET MÉTHODES

2.1 Matériel

2.1.1 Milieux de culture

Milieu MA : 31 mM Na_2HPO_4, 25 mM KH_2PO_4, 18 mM NH_4Cl, 41 µM acide nitrilotriacétique, 2 mM $MgSO_4$, 0,45 mM $CaCl_2$, 3 µM $FeCl_3$, 1 µM $MnCl_2$, 1 µM $ZnCl_2$, 0,3 µM ($CrCl_3$, H_3BO_3, $CoCl_2$, $CuCl_2$, $NiCl_2$, Na_2MoO_2, Na_2SeO_3).

Le milieu de culture sera additionné d'une source de carbone dont la concentration sera de 0,02M (quinate ou succinate).

2.1.2 Phases mobiles

Plusieurs méthodes de chromatographie liquide ont été développées, avec différentes phases mobiles. Chacune d'entre elle est composée d'une phase aqueuse et d'une phase organique.

Phases aqueuses :

pH 9,9 ⇨ H_2O, carbonate d'ammonium ($(NH_4)_2CO_3$) 10 mM, pH 9.9 (ajusté avec NH_4OH)

pH 6,8 ⇨ H_2O, acétate d'ammonium (NH_4OAc) 10 mM, pH 6.8

pH 4 ⇨ H_2O, acétate d'ammonium (NH_4OAc) 10 mM, pH 4 (ajusté avec acide acétique)

Phases organiques utilisées

Acétonitrile (ACN)

Acétonitrile/Méthanol (ACN/MeOH) 50 : 50.

2.2 Protocoles de préparation de métabolomes

Trois protocoles différents ont été utilisés au cours de ce travail.

2.2.1 Protocole d'extraction de métabolome haut-débit

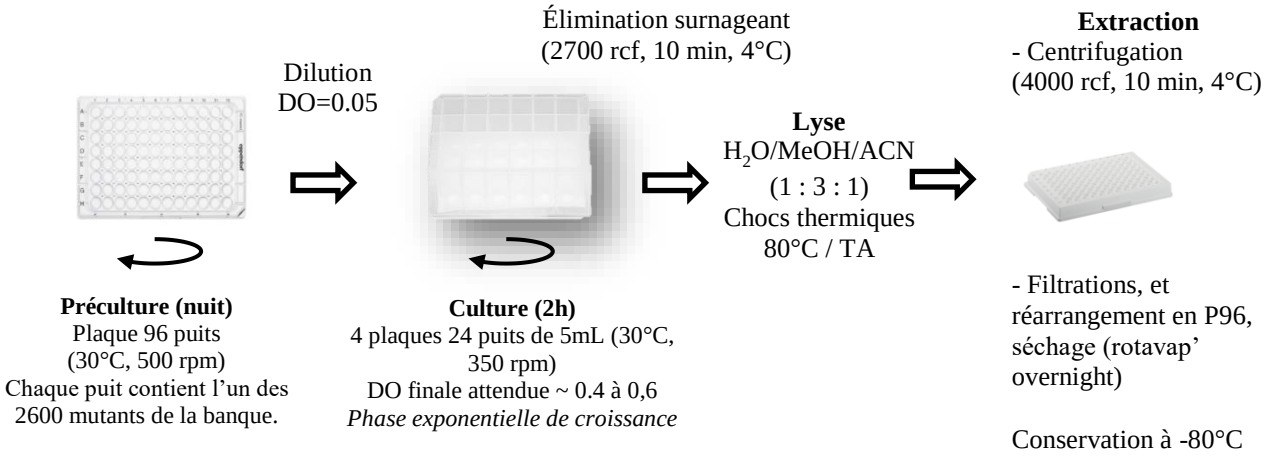
Un protocole adapté à l'extraction haut débit de métabolome a été adapté de Metabolomics Service Protocols Sample Preparation for Cells de l'université de Glasgow [106]. 94 métabolomes différents peuvent être préparés par jour. Jusqu'à trois plaques peuvent être préparées chaque semaine.

Pour notre utilisation, la banque de mutants a été préparée sous format plaque où chaque puits d'une plaque 96 puits contient un unique mutant. Pour cela, à partir des tubes glycérolés de la banque de mutant, dans chaque puit d'une macro-plaque 96 puits à fond rond de 2 mL, 5 µL d'une culture glycérolée d'un mutant provenant des cryotubes de la banque de mutants a été ajouté à 1 mL de milieu MA, additionné de succinate de concentration finale 20 mM (milieu MAS) et de kanamycine. Les puits H11 et H12 sont laissés vides. La plaque est recouverte d'un pad poreux, placée sous agitation (500 rpm) et incubée la nuit à 30°C dans un Kühner (Climo Shaker ISF1-X Kühner). La plaque est alors glycérolée et conservée à -80°C jusqu'à utilisation. 28 plaques sont alors obtenues. Par la suite, ces plaques seront nommées « plaques mère ». Chaque jour d'extraction, une plaque mère permet d'ensemencer rapidement la plaque de préculture de l'extraction du jour.

De plus, lors de la création des mutants, chaque gène excisé a été remplacé par une cassette de résistance à la kanamycine. Ainsi, lors de la préculture, afin d'éviter au maximum toute contamination par la souche sauvage, nous avons additionné le milieu de culture (MA + succinate à 20 mM) de kanamycine (62 µM). Cependant, la présence d'un antibiotique peut modifier drastiquement le métabolisme d'une bactérie [107] : nous avons donc choisi de ne pas mettre d'antibiotique dans le milieu de culture de nos souches à l'étape de la culture afin de conserver l'image du métabolome que nous connaissons.

Chaque jour d'extraction, nous suivons le protocole décrit ci-après (Figure 16).

Figure 16: Protocole de culture des cellules et extraction des métabolomes à haut débit



Préculture

Dans 94 puits d'une macro-plaque 96 puits à fond rond de 2 mL, on ensemence 5 µL d'une culture glycérolée provenant des plaques mères de la banque de mutant dans 1 mL de milieu minimum, additionné de succinate de concentration finale 20 mM (milieu MAS) et de kanamycine (62µM).

Pour le puits H11, on ensemence 5 µL d'une culture glycérolée d'ADP1 Wild Type (WT), dans 1 ml de milieu MAS (sans kanamycine). Le puits H12 n'est pas ensemencé. Il ne contient que du milieu de

culture et sert de témoin de contamination. La plaque est recouverte d'un pad poreux et placée sous agitation (500 rpm) et incubée la nuit à 30°C dans un Kühner (Climo Shaker ISF1-X Kühner).

Culture :

Après avoir mesuré la DO_{600} de chaque puits de préculture, ceux-ci sont dilués dans 5 ml de milieu MAQ à la DO_{600} de 0,05 dans les puits correspondants (Figure 17) de quatre macro-plaques 24 puits de 10 ml à fonds ronds (Whatman, référence 7701-5110). Les quatre macro-plaques sont recouvertes d'un pad poreux, et placées sous agitation (300 rpm) et incubées à 30°C jusqu'à atteindre une valeur de DO_{600} comprise entre 0,4 à 0,6 (~4h) dans un Kühner.

96	1	2	3	4	5	6	7	8	9	10	11	12
A	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
B	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12
C	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
D	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12
E	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12
F	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12
G	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12
H	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12

24	1	3	5	7	9	11
A	A1	A3	A5	A7	A9	A11
C	C1	C3	C5	C7	C9	C11
E	E1	E3	E5	E7	E9	E11
G	G1	G3	G5	G7	G9	G11

24	2	4	6	8	10	12
A	A2	A4	A6	A8	A10	A12
C	C2	C4	C6	C8	C10	C12
E	E2	E4	E6	E8	E10	E12
G	G2	G4	G6	G8	G10	G12

24	1	3	5	7	9	11
B	B1	B3	B5	B7	B9	B11
D	D1	D3	D5	D7	D9	D11
F	F1	F3	F5	F7	F9	F11
H	H1	H3	H5	H7	H9	H11

24	2	4	6	8	10	12
B	B2	B4	B6	B8	B10	B12
D	D2	D4	D6	D8	D10	D12
F	F2	F4	F6	F8	F10	F12
H	H2	H4	H6	H8	H10	H12

Figure 17: Plans de correspondance des puits de la plaque 96 puits aux plaques 24 puits.

Extraction

Les quatre macro-plaques sont centrifugées à 2700 rcf pendant 10 minutes à 4°C. Le surnageant est éliminé par retournement. 200 µl de solution d'extraction (H_2O/MeOH/ACN 1 : 3 : 1) sont ajoutés dans chacun des puits. Un couvercle est placé sur chaque plaque. Le fonds des puits est immergé dans un bain de carboglace/éthanol jusqu'à congélation du culot. Les macro-plaques sont laissées à température ambiante jusqu'à décongélation du culot. Cette étape congélation/décongélation est répétée deux fois. Les macro-plaques sont placées en chambre froide à 4°C, sur un agitateur orbital Bellco pendant 1h.

Réarrangement

Les macro-plaques sont alors centrifugées à 4100 rpm, à 4°C, pendant 10 minutes. Le surnageant de chaque puits est récupéré et transféré dans une plaque 96 puits à fond conique (Biorad, référence HSP9611) par un automate (Sciclone ALH3000 Perkin Elmer). Les plaques sont centrifugées à 4100 rpm, à 4°C, pendant 10 minutes et les surnageants sont transférés dans une nouvelle plaque 96 à fond conique. Ces deux étapes de centrifugations permettent d'éliminer la totalité des débris cellulaires.

Séchage

Les métabolomes sont séchés pendant la nuit à température ambiante au « SpeedVac™» (SpeedVac system AES2010 Thermo). La plaque est placée au congélateur à -80°C pour conservation jusqu'au moment de l'injection.

Préparation à l'injection

Chaque culot séché est repris dans 20 µl d'eau. Les plaques sont agitées, sur un agitateur Bellco pendant 1h à 4°C. Dans chaque puits on ajoute 42 µl de phase mobile (ACN / H_2O pH 9,9 80 :20). Les quatre macro-plaques sont centrifugées pendant 10 minutes à 4100 rcf à 4°C. Les surnageants de chaque puits sont transférés dans une plaque 96 puits, dont chacun des puits contient une membrane de filtration de 0,22 µm (AcroPrep Advance 96 Filter Plate 0,2µm, 350µL well). Le filtrat est récupéré une plaque 96 puits à fonds coniques. La filtration est effectuée par centrifugation à 1500 rcf pendant 5 min à 4°C. La plaque est finalement placée dans l'échantillonneur de notre système LC/MS, à 4°C.

2.2.2 Préparation de métabolomes pour la purification du MO211

La préparation de métabolome adaptée à la purification du MO211 a été adaptée du protocole développé pour le criblage de la banque (paragraphe 2.2.1) et peut se décomposer en 4 étapes (Figure 18). Le développement de cette méthode d'extraction a fait l'objet du stage de master d'Adrien CUSSON [108] au sein de notre laboratoire.).

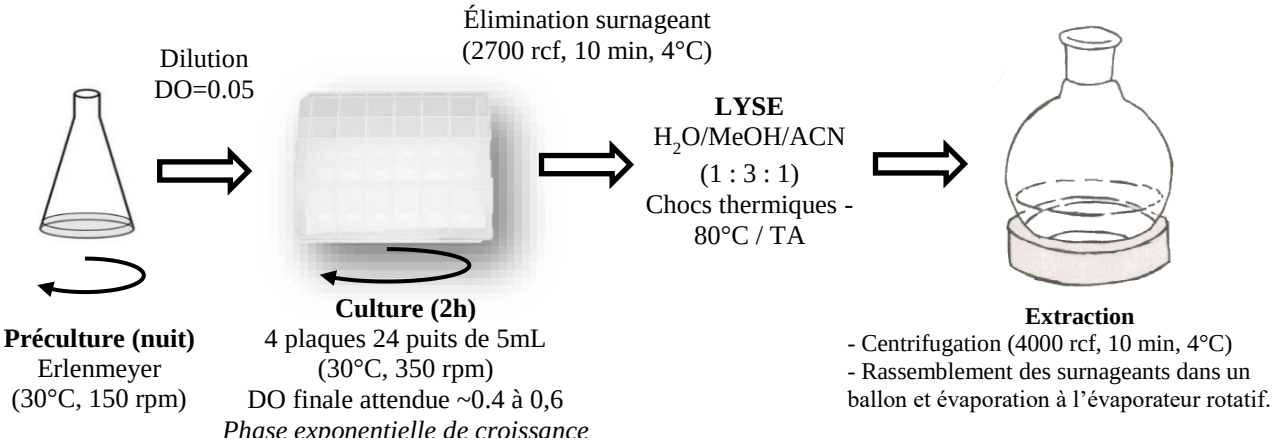


Figure 18: Protocole de préparation des métabolomes en plaque pour la purification du MO211

Préculture

5 µl d'une culture glycérolée d'ADP1 WT est ensemencé dans 100 mL de milieu MA contenant 20 mM

de quinate, dans un erlenmeyer de 500mL. Cet erlemeyer est placé sous agitation (150 rpm) et incubé la nuit à 30°C dans un Infors (Infors Multitron standard).

Culture

Après avoir mesuré la DO_{600} de le préculture, celle-ci sera diluée dans 5 ml du même milieu (MAQ) à la valeur de DO_{600} de 0,05 dans quatre macro-plaques 24 puits. Les quatre macro-plaques sont recouvertes d'un pad poreux, placées sous agitation (300 rpm) et incubées à 30°C dans un Kühner jusqu'à une valeur de DO_{600} comprise entre 0,4 et 0,6.

Extraction

Les quatre macro-plaques sont centrifugées à 2700 rcf pendant 10 minutes à 4°C. Le surnageant est éliminé par retournement. 200µL de solution d'extraction (H_2O/MeOH/ACN 1 : 3 : 1) sont ajoutés à chaque culot, un couvercle est placé sur chacune des plaques. Le fond des plaques est immergé dans un bain de carboglace/éthanol jusqu'à congélation du culot.

Les macro-plaques sont laissées à température ambiante jusqu'à décongélation du culot. Cette étape congélation/décongélation est répétée deux fois. Les macro-plaques sont placées en chambre froide à 4°C et agitées sur un agitateur Bellco pendant 1 heure.

Réarrangement et séchage

Le contenu de tous les puits de la plaque est rassemblé dans un tube Falcon de 50 ml. Les solvants sont évaporés au SpeedVac™ (Concentrator 5301, Eppendorf) à température ambiante. Le résidu sec (solide jaune orangé) est placé à -80°C pour être conservé jusqu'à utilisation. Ce protocole est répété trois fois (3 jours).

Préparation pour purification

Le résidu sec est repris dans 6 ml de phase mobile (ACN / H_2O pH 9,9 80 : 20) puis soumis aux ultrasons pendant 10 minutes. Les débris sont éliminés après centrifugation (4000 rcf, 4°C, 10 min). Le surnageant est ensuite filtré sur une membrane de 0,22 µm (Millipore Millex- GV 13 mm) qui est ensuite rincé avec 1 ml de cette même phase mobile phase mobile. Pour chaque injection, un volume de 200 µL sera prélevé et injecté.

2.2.3 Production de métabolomes à grande échelle

Un second protocole (Figure 19) de préparation de métabolomes à échelle préparative a été développé. Celui-ci permet notamment la culture des cellules en grand volumes (jusqu'à 600 mL) dans un erlenmeyer.

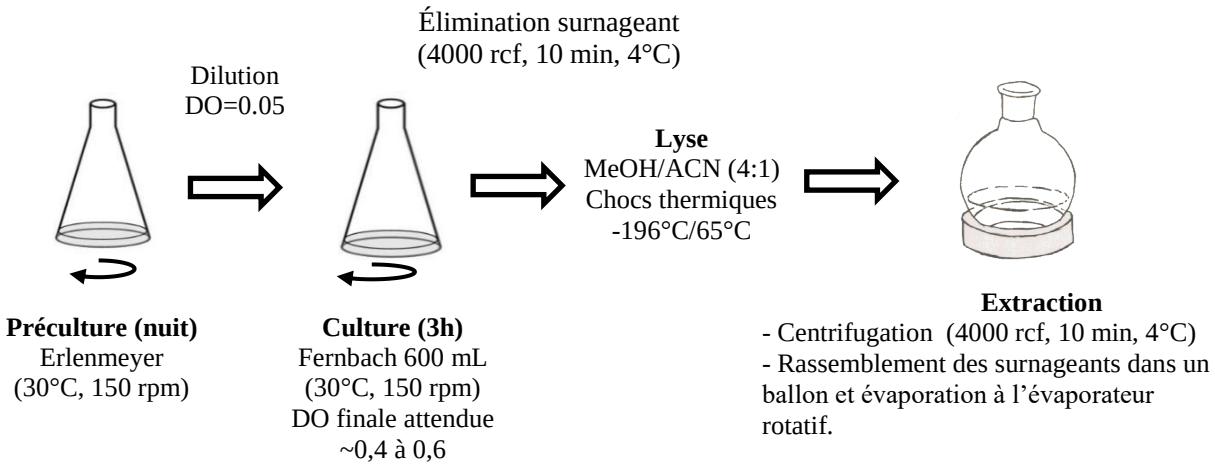


Figure 19: Protocole de préparation des métabolomes à grande échelle pour la production des MO

Préculture

On ensemence avec 5 µl d'une culture glycérolée d'ADP1 WT 100 ml de milieu MA contenant 20 mM de quinate dans un erlemeyer de 500 mL. La préculture est agitée à 150 rpm pendant toute la nuit à 30°C dans un agitateur Infors (Infors Multitron standard).

Culture liquide

La préculture est diluée dans 600 mL du même milieu minimum à une valeur de DO_{600} de 0,05 dans un flacon Fernbach à ailettes d'une contenance de 2,6 L. La culture est incubée sous agitation à 30°C jusqu'à obtenir une valeur de DO_{600} de minimum 0,5 (3h00-3h30 de culture).

Extraction

La culture est centrifugée à 20°C, 4000 rcf, pendant 10 minutes. Le surnageant est éliminé par retournement. Le culot est repris dans 300 mL de solution de lyse (ACN/MeOH 80 :20) et réparti dans 7 tubes Falcon de 50 mL. Le mélange subira six cycles de congélation/décongélation (azote liquide/ bain-marie à 65°C).

Evaporation des solvants

Les échantillons sont versés dans un ballon de 1L et les solvants sont évaporés grâce à un évaporateur rotatif. Le résidu sec (solide jaune orangé) est placé à -80°C pour être conservé jusqu'à utilisation.

2.3 Chromatographie liquide

2.3.1 Chromatographie liquide analytique

Les échantillons sont analysés sur une UHPLC Ultimate 3000 (Dionex). Trois méthodes chromatographiques (Tableau 2) sont utilisées.

Zic-*p*HILIC (Merck Sequant polymeric beads 5µm, 150 x 4.6 mm)					
Phase aqueuse	phase organique (B)	débit (mL/min)	volume d'injection (µL)	température colonne	
pH 9,9 ou pH 6,8	ACN	0,5	10	40°C	
Gradient (min)	80% B	80% B -> 40%B	40% B	40% B -> 80%B	80%B
	2	20	8	5	15

Zic-*p*HILIC (Merck Sequant polymeric beads 5µm, 50 x 2,1 mm)					
phase aqueuse	phase organique (B)	débit (mL/min)	vol. d'injection (µL)	température colonne	
pH 9,9	ACN	0,22	2	40°C	
Gradient (min)	80% B	80% B -> 40%B	40% B	40% B -> 80%B	80%B
	15 sec	3 min 30	1 min 30 sec	1 min 6 sec	2 min 30

Syncronis HILIC (Themo Scientific, 1,7µm, 50 x 2,1 mm)					
phase aqueuse	phase organique (B)	débit (mL/min)	vol. d'injection (µL)	température colonne	
pH 6,8	ACN	0,2	2	30°C	
Gradient (min)	80% B	80% B -> 40%B	40% B	40% B -> 80%B	80%B
	18 sec	1 min 30 sec	1 min 30 sec	30 sec	1 min 12 sec

Tableau 2: Présentation des trois méthodes chromatographiques analytiques utilisées dans ce projet pour trois colonnes analytiques : ZIC-pHILIC 15 et 5 cm et Syncronis HILIC de 5 cm.

2.3.2 Chromatographie liquide préparative

Les deux étapes de purification ont été faites sur une HPLC préparative couplée à un détecteur UV (Pompe LC-20AP, Détecteur UV SPD-20AV, Collecteur FRC-10A, Shimadzu). Les longueurs d'ondes utilisées pour visualiser l'évolution de l'élution sont 224 et 307 nm pour le MO211.

Nous disposons au laboratoire de deux colonnes adaptées aux purifications de composés : une colonne ZIC-HILIC (Merck Sequant polymeric beads 5µm, 250 x 10 mm), semi-préparative et une colonne C18 (XBridgeTM Prep 5µm, 150 x 30 mm), préparative. Les différentes méthodes chromatographiques

utilisées, sont décrites dans le Tableau 3.

ZIC-HILIC (Merck Sequant polymeric beads 5µm, 250 x 10 mm)					
phase aqueuse	phase organique (B)	débit (mL/min)	vol. d'injection (µL)	température colonne	
pH 6,8	ACN	4,6	200	Température ambiante (20°C)	
Gradient (min)	80% B	80% B -> 40%B	40% B	40% B -> 80%B	80%B
	5 min	30 min	10 min	8 min	20 min

ZIC-HILIC (Merck Sequant polymeric beads 5µm, 250 x 3 mm)					
phase aqueuse	phase organique (B)	débit (mL/min)	vol. d'injection (µL)	température colonne	
pH 6,8	ACN	4,6	200	Température ambiante (20°C)	
isocratique	80% B	80% B -> 40%B	40% B	40% B -> 80%B	80%B
	25 min	10 min	10 min	10 min	20 min

C18 (XBridge™ Prep 5µm, 150 x 30 mm)					
phase aqueuse	phase organique	débit (mL/min)	vol. d'injection (µL)	température colonne	
pH 4	ACN/MeOH 50 : 50	4,6	200	Température ambiante (20°C)	
isocratique	80% A	80% A -> 40%A	40% A	40% A -> 80%A	80%A
	25 min	10 min	10 min	10 min	20 min

Tableau 3 : Présentation des trois méthodes chromatographiques préparatives utilisées dans ce projet pour la colonne semi-préparative ZIC HILIC de 25 cm et la colonne préparative C18 de 15 cm.

2.4 Paramétrage du Velos Pro Orbitrap Elite

2.4.1 Source d'ionisation HESI

Les paramètres de source utilisés sont les suivants :

Paramètres	Mode d'ionisation positif	Mode d'ionisation négatif
Tension de cône (kV)	4,20	4,00
Lentille tubulaire de désolvatation : *Tube Lens* (V)	55	66
Gaz gaine – N_2 (unité arbitraire)	60	60
Gaz auxiliaire – N_2- (unité arbitraire)	44	50
Température de désolvatation (°C)	275	275

Tableau 4: Paramètres de source des méthodes utilisées en modes d'ionisation positif et négatif

2.4.2 Analyseurs

Pour les injections des criblages de la banque de mutant, la méthode d'analyse ne comprendra que l'acquisition des spectres de masse pour une gamme comprise entre 50 et 1000 avec une détection haute résolution de 60 000 à m/z 400.

Lors des travaux d'élucidation structurale du composé MO211, des expériences HCD et CID MS^n séquentielles sont menées. Le mode « *dependant scan* » est utilisé pour l'implémentation de liste d'ions précurseurs spécifiques à fragmenter en MS/MS (parent list) et en MS^3 (product list), ainsi que l'optimisation des énergies de collision. Pour toutes les expériences de fragmentation, la détection se fera dans l'Orbitrap, avec une résolution de 60 000 FWHM (m/z 400) ainsi qu'une activation Q de 0,250 ont été utilisées avec un temps d'activation de 10,0 ms pour CID et de 0,100 ms pour HCD. La largeur d'isolement et la gamme de balayage étaient toujours fixées à 1,0 m/z et [50;250] respectivement.

2.4.3 Traitement des données.

Les données LC/MS ont été acquises en fichiers .RAW en mode profil. Ils sont ensuite convertis et en mode centroïde au format universel mzXML avec MassMatrix File Conversion Tools (http://www.massmatrix.net). Les données ont ensuite été analysées en utilisant la plateforme web XCMS Online [82] version 3.7.0 avec l'algorithme centWave [81]. Le paramétrage par défaut « HPLC/Orbitrap II » a servi de base au paramétrage. Les paramètres ont été ajustés à la méthode analytique utilisée pour le criblage; ils sont récapitulés dans le Tableau 5.

La redondance du signal (isotopes, adduits, fragments …) est automatiquement annotée par XCMS et ces ions sont regroupés dans différentes « familles » formant des groupes d'ions correspondant potentiellement au même métabolite.

Paramètres	Criblage
ppm	3
Minimum peak width	5
Maximum peak width	20
S/N	3
mzdiff	0,001
Prefilter peak/ prefilter intensity	3/100
Prof step	1
Min frac	0,75
bw	5
mzwid	0,001
max	100
minsamp	1
Statistical test	Unpaired parametric t-test (Welsh t-test)
Fold change	1
p-value highly significant features	0,1
p-value significant features	0,5
Search for	Isotopes+adducts
ppm	5
m/z absolute error	0,015
ppm	10
adducts	$[M+H]^+$, $[M+NH_4]^+$,$[M+Na]^+$,$[M+K]^+$ / $[M-H]^-$
EIC width	200

Tableau 5: Paramètres XCMS utilisés pour le traitement des données du criblage de la banque de mutant

2.5 Résonance Magnétique Nucléaire

Les analyses RMN du MO211 ont été effectuées à l'Institut de Chimie des Substances Naturelles à Gif-sur-Yvette sur une RMN Bruker 600 MHz pour RMN ^{1}H avec cryosonde. L'échantillon contenant le métabolite purifié a été dissous dans de l'eau deutérée et analysé en RMN. Les spectres suivants ont été acquis :

- 1D : ^{1}H, ^{13}C ;
- 2D : ^{1}H-^{1}H-COSY, ^{1}H-^{13}C-HSQC, ^{1}H-^{13}C-HMBC et ^{1}H-^{15}N-HMBC.

2.6 Validation de la structure du MO211

Cinq des six structures chimiques supposées possibles pour le MO211 ont été faites synthétiser par la société Synthenova.

Chacun des cinq composés ayant pu être synthétisé a subi les mêmes analyses en LC-MS que le MO211 (paragraphe 2.3.3). Les comparaisons des temps de rétentions et des fragmentations ont permis de valider la structure du MO211.

Deux à trois milligrammes de chacune des molécules a également été dissous dans 0,75 mL de H_2O/D_2O 90 : 10 et analysées les unes après les autres en RMN. Les expériences RMN des composés synthétiques ont été réalisées à l'Université d'Évry sur un spectromètre Bruker 600 MHz pour RMN ^{1}H avec cryosonde. Les spectres suivants ont été acquis pour chacune des molécules synthétisées :

- 1D : ^{1}H, ^{13}C ;
- 2D : ^{1}H-^{1}H-COSY, ^{1}H-^{13}C-HSQC, ^{1}H-^{13}C-HMBC.

3 RÉSULTATS ET DISCUSSIONS

3.1 A la recherche des gènes impliqués dans la synthèse des Métabolites Orphelins

Comme exposé précédemment dans l'introduction, une partie de mon travail de thèse consiste à repérer les gènes impliqués dans la formation des MO. Ces MO sont :

- des métabolites identifiés (via des standards commerciaux), qui ne sont pas prédits par le modèle métabolique d'ADP1.

- des métabolites non identifiés et non attendus très majoritairement voire exclusivement produits dans les cellules utilisant le quinate (ANNEXE 3). Aucun de ces métabolites n'a pu être identifié à l'aide de standards commerciaux et, pour exemple, parmi les 28 métabolites non identifiés exclusivement présents sur quinate, 8 d'entre eux n'ont aucune proposition d'identité *via* les bases de données.

Dans la première partie de ma thèse, nous avons cherché à identifier les gènes impliqués dans la biosynthèse de certains de ces MO. Pour tenter de découvrir ces gènes, nous analysons le phénotype moléculaire, ou chémotype, de l'ensemble des 2600 mutants de délétion de notre bactérie, en recherchant l'absence d'un ou plusieurs MO dans chaque métabolome.

La préparation des 2600 métabolomes, leur analyse, puis le traitement des données a nécessité de repenser le protocole de chacune de ces étapes.

3.1.1 Réduction du temps d'analyse LC-MS

La méthode analytique des métabolomes utilisée précédemment par le laboratoire faisait intervenir une étape chromatographique de 50 minutes. Utiliser une telle méthode pour l'analyse de 2600 échantillons prendrait au minimum six mois (sans interruption) et monopoliserait la machine au détriment de tout autre projet. Pour cette même raison, nous avons restreint les injections de métabolomes à la durée du weekend exclusivement. Nous disposions donc d'une soixantaine d'heure chaque semaine pour procéder à nos analyses LC/MS. Nous avons dû développer une nouvelle méthode LC/MS.

Deux contraintes dictent la mise au point de la nouvelle méthode chromatographique : réduire au maximum le temps entre l'injection de deux échantillons tout en conservant pour les MO étudiés un temps de rétention compatible avec une bonne analyse chromatographique.

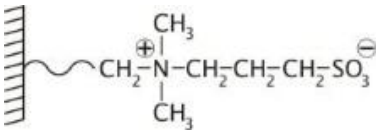


Figure 20: La phase stationnaire zwitterionique de la colonne ZIC-pHILIC

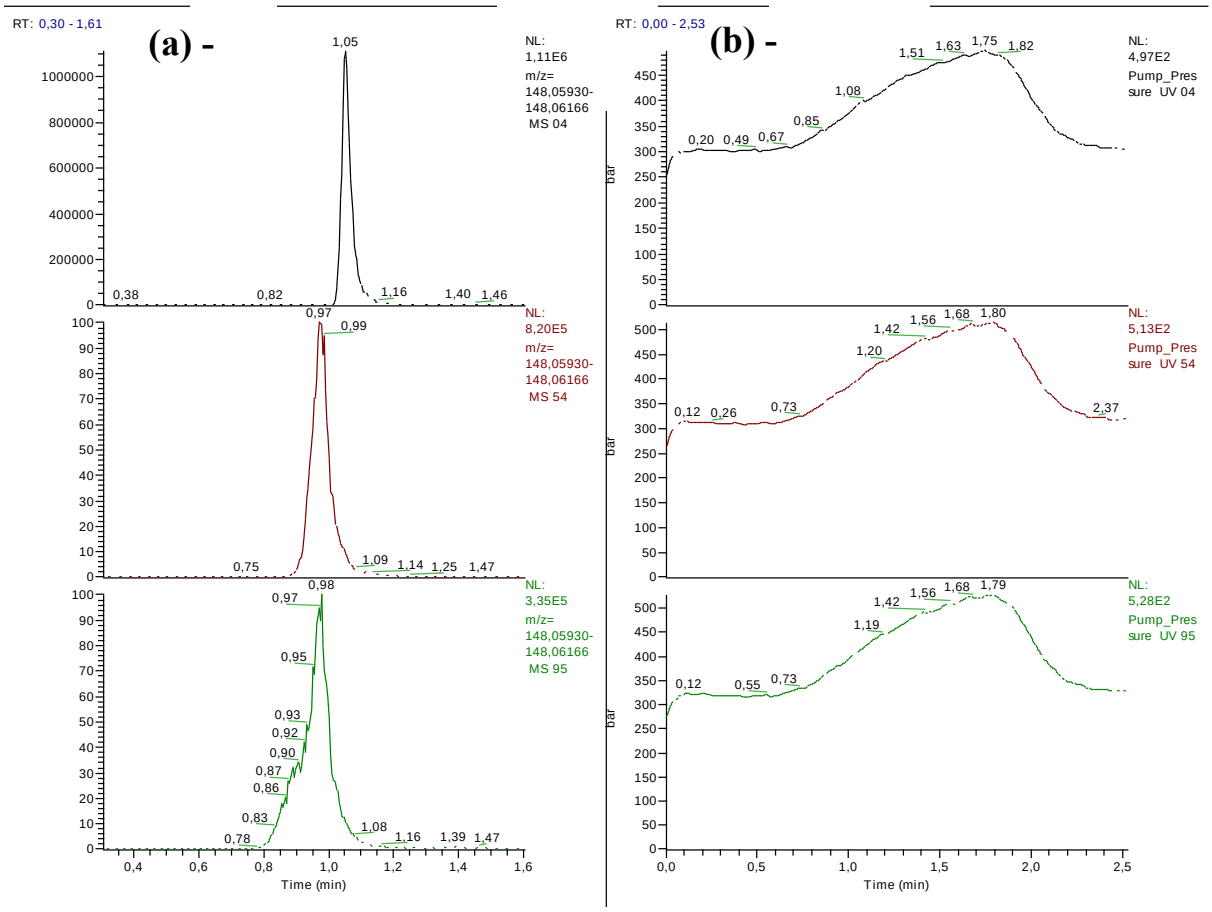


Figure 21: Évaluation de la qualité chromatographique d'une analyse d'un métabolome en UPLC. (a) EIC du glutamate (m/z148,06048, 5ppm) et (b) relevé de pression pour trois injections d'une séquence de 96 injections consécutives de métabolomes sur une colonne UPLC Syncronis HILIC 50 x 2,1 mm²; 1,7µm. Du haut vers le bas injection n° : 4, 54 puis 94 de la séquence.

Nous avons souhaité utiliser pour la nouvelle méthode chromatographique la même phase stationnaire que celle utilisée lors de l'analyse comparative quinate/succinate [104] car les métabolites d'intérêt y sont retenus et séparés de façon satisfaisante. Cette colonne (ZIC-*p*HILIC ; Merck) utilise comme phase stationnaire une sulfobétaïne zwitterionique (figure 20).

Afin de réduire le temps d'analyse, nous avons considéré une colonne plus courte (5 cm contre 15 cm pour la méthode originale) et d'un diamètre interne inférieur (2,1 mm contre 4,6).

Tout d'abord, l'utilisation de colonnes dédiées à l'UPLC (Ultra Performance Liquid Chromatography), comme les colonnes Syncronis HILIC 50 x 2,1 mm^2; 1,7 µm (Thermo) a été envisagée. Celles-ci utilisent la même phase stationnaire que la ZIC-*p*HILIC. L'UPLC repose sur l'emploi de phase stationnaire composée de particules < 2 µm (alors que les colonnes HPLC sont habituellement remplies de particules de 3 à 5 µm). De plus, nous avons un système pouvant supporter des pressions jusqu'à

1000 bars. Il nous a donc été possible d'augmenter le débit de la phase mobile, et donc la vitesse d'analyse, sans altérer les performances chromatographiques.

Cependant, il s'est avéré que la performance de ces colonnes était incompatible avec l'injection de nombreux échantillons complexes. A titre d'exemple, comme présenté en Figure 21, après une centaine d'injections seulement, la pression augmentait de plus de 20 bars (Figure 21-b). En un weekend soit 400 injections, la pression pouvait atteindre les limites maximales du système (1000 bars). De plus, comme présenté en Figure 21-a, la résolution chute drastiquement au fur et à mesure des injections.

Cette colonne a donc été écartée et nous avons reconsidéré les colonnes HPLC, de granulométrie conventionnelle.

Nous avons finalement choisi d'utiliser une colonne ZIC-*p*HILIC 50 x 2,1 mm^2 ; 5 µm. L'avantage d'utiliser une colonne courte est de réduire le temps d'analyse, mais cela est au détriment de la résolution. Cette baisse de résolution doit en partie être compensée par le diamètre réduit de la colonne (2,1 au lieu de 4,6 mm).

La méthode analytique a donc été réadaptée à partir de la méthode chromatographique de 50 minutes. Pour la mise en place de la nouvelle méthode chromatographique analytique, nous avons considéré un premier critère : la réduction du temps d'analyse. Afin de pouvoir effectuer toutes les injections prévues le weekend (2 plaques 96 puits dans les deux modes d'ionisations + contrôles qualités) soit 440 injections en maximum 2 jours et demi, nous devions mettre en place une méthode analytique de 9 minutes maximum.

Pour effectuer la conversion de méthode analytique d'une colonne à l'autre nous avons suivi les recommandations du fabricant (Merck). Nous avons également considéré les recommandations techniques proposées par deux convertisseurs de méthodes analytiques proposées (ANNEXE 1) par Thermo (https://www.thermofisher.com/content/dam/tfs/ATG/CMD/cmd-documents/oper/tech/chrom/lc/sys/RSLC-Method-Transfer-Tool-2-3.xlsm) et Sigma Aldrich (https://www.sigmaaldrich.com/analytical-chromatography/hplc/method-transfer-calculator.html).
Afin de garder une bonne résolution, nous avons choisi d'utiliser un débit de 0,22 mL/min. Nous avons également, à l'aide d'un convertisseur de méthode chromatographique fourni par le fabricant, adapté la méthode chromatographique de 50 minutes à la géométrie de la colonne utilisée pour le criblage (Annexe 1). Ainsi, la méthode chromatographique de l'injection jusqu'au retour aux conditions initiales dure 6,35 minutes. La méthode chromatographique finale est présentée en Figure 22.

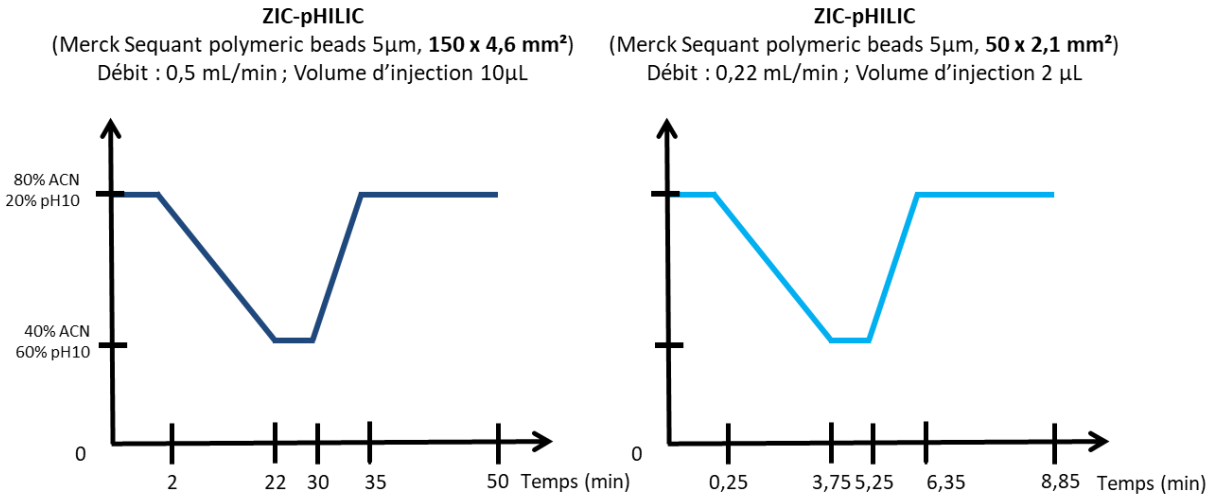


Figure 22: Représentation graphique de deux types de méthode chromatographique utilisée au laboratoire. A gauche, celle utilisée initialement par le laboratoire, incompatible avec le haut-débit (gradient de 50 min, ZIC p*HILIC 150 x 4,6 mm², 5 µm). A droite, méthode analytique développée pour le criblage de la banque de mutants (gradient de 8.85 minutes, ZIC* p*HILIC 50 x 2,1 mm², 5 µm).*

Cependant, les colonnes ZIC-*p*HILIC nécessitent un temps de rééquilibration de 7 à 8 volumes de colonne, soit 5 à 6 minutes (pour une colonne de 5 cm à un débit de 0,22 mL/min). Or, le temps de rééquilibration de la colonne est un « temps mort » dans le déroulé de la séquence. Nous avons donc réduit au maximum le temps de rééquilibration, en considérant qu'un retour à la pression de début d'analyse (pression au moment de l'injection de l'échantillon) est significatif d'une rééquilibration suffisante. En 2 minutes et 30 secondes à partir du retour à une phase mobile composée de 80% d'ACN et 20% de tampon pH 9,9, le système retrouvait la pression mesurée au moment de l'injection (environ 30 bars). Nous avons donc conservé 2 minutes et 30 secondes comme temps de rééquilibration. Au total, nous avons réduit de 50% le temps de rééquilibration. Finalement, entre deux injections consécutives, il y a 8 minutes et 51 secondes.

Enfin, le port d'injection de la chromatographie liquide effectue deux cycles de lavage à chaque injection : l'un avant, l'autre après l'injection. Ces deux lavages se font indépendamment du débit chromatographique. Pour gagner du temps, nous avons fait en sorte que la LC procède au nettoyage du port d'injection durant la phase de rééquilibration de la colonne. Cette astuce nous permet ainsi de gagner encore 2 minutes et 30 secondes, soit le temps de rééquilibration de la colonne. Cette astuce est présentée en Figure 23.

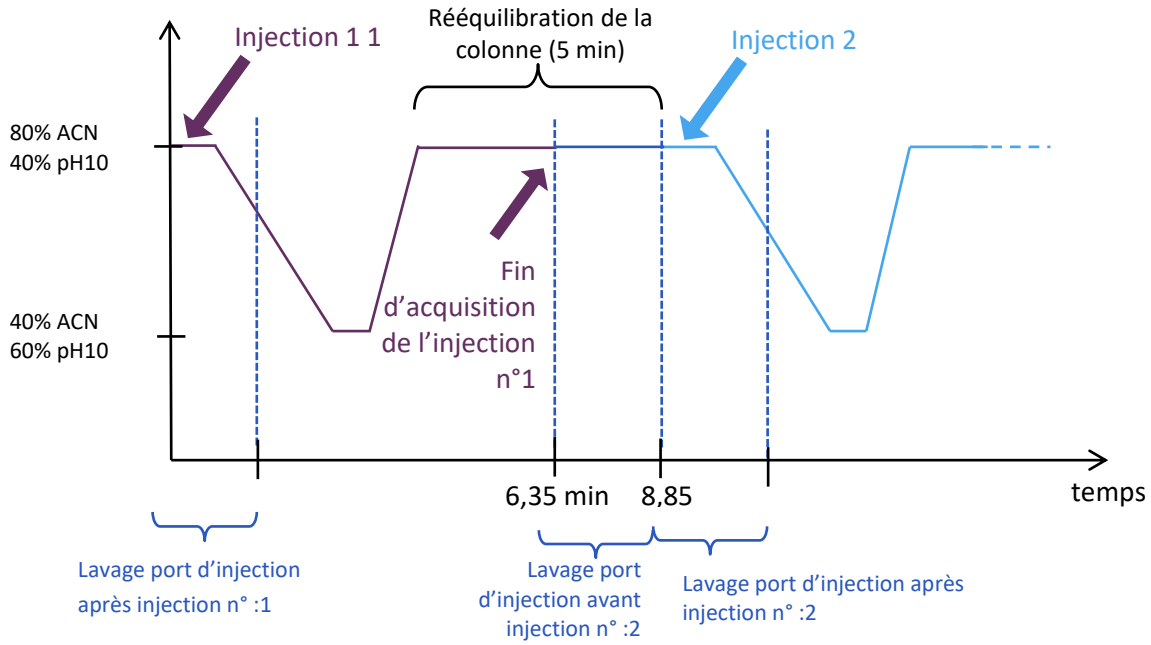


Figure 23: Représentation graphique de la méthode analytique utilisée pour le criblage de la banque de mutant lors de deux injections consécutives

Finalement, en considérant 9 minutes d'analyse par échantillon, 435 échantillons à injecter chaque weekend, nous occupons l'instrumentation environ 2 jours et 17h.

L'injection d'une plaque complète a permis d'éprouver la méthode chromatographique à l'épreuve du haut débit. Comme le présente la Figure 24, la résolution des pics est suffisante, et les temps de rétention de nos métabolites sont compatibles avec une bonne analyse chromatographique.

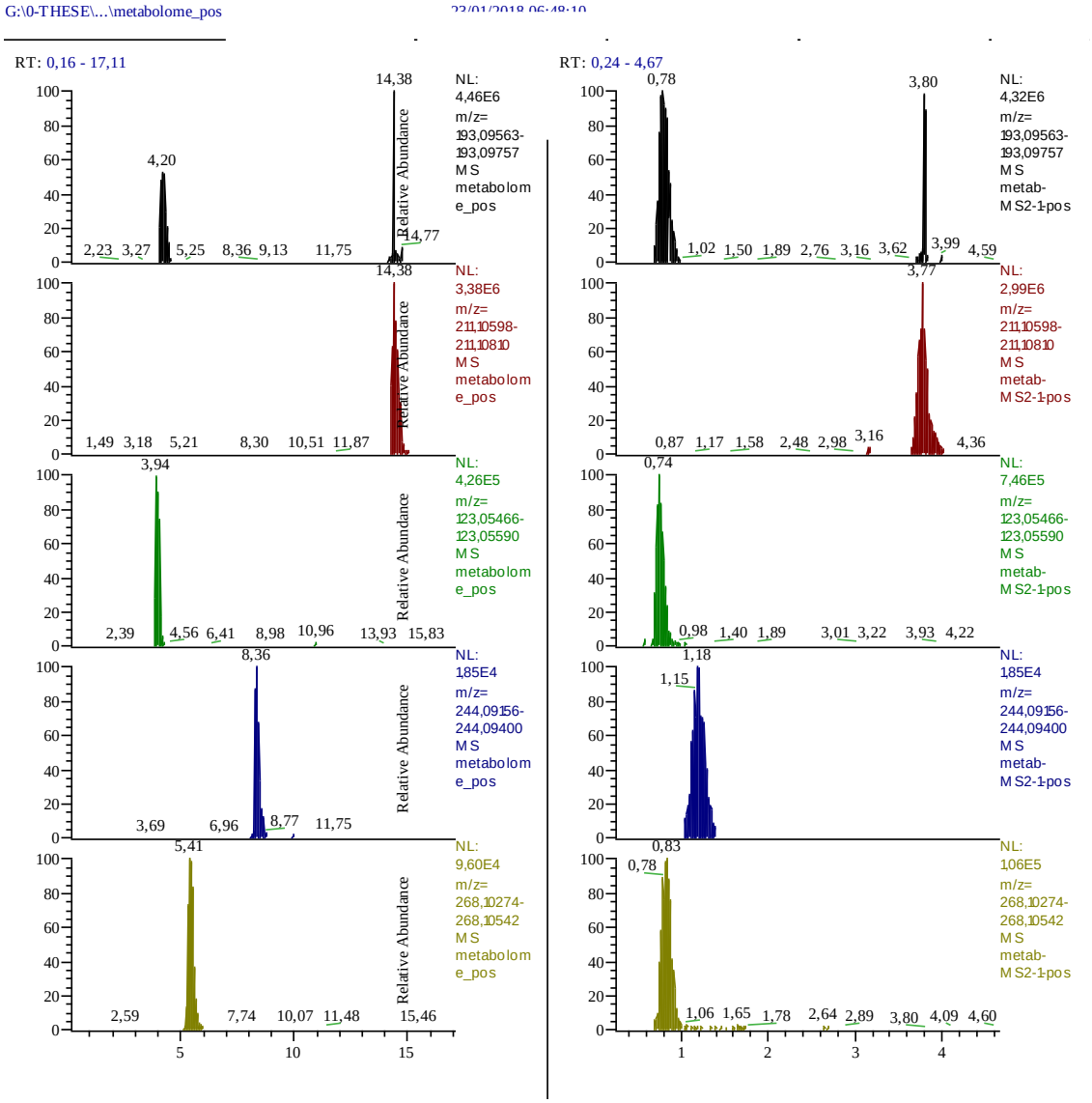


Figure 24: Comparaison de la qualité chromatographique des deux méthodes chromatographiques analytiques utilisées au laboratoire. EIC de quelques MO choisis du haut vers le bas : MO193, MO211, m/z123,05528 (nicotinamide), m/z =244,09278 (cytidine) et m/z268,10408 (adénosine), (5ppm, mode d'ionisation positif): (à gauche) sur la méthode analytique de 50 minutes, (à droite) sur la méthode analytique utilisée pour le criblage.

3.1.2 Préparation des métabolomes

Le protocole de préparation de métabolome du laboratoire [104] ne permettait la préparation que de quelques échantillons par jour. Ce débit n'étant pas adapté au criblage de milliers de métabolomes, un nouveau protocole a été développé. Celui-ci a fait l'objet d'un stage de M2 en 2015 [108]. Il permet la préparation d'une centaine de métabolomes par jour (2.2.1). Ce nouveau protocole permet également d'extraire un nombre de MO équivalent au nombre de MO extraits lors d'une extraction selon l'ancien protocole.

Les 2600 mutants de la banque ont été répartis et mis en culture dans des plaques au format 96 puits. Chacune de ces plaques contient 94 mutants distincts (puits A1 à H10 d'une plaque). Le protocole détaillé est décrit dans le chapitre Matériel et Méthodes (2.2.1).

Afin d'évaluer la qualité d'extraction du nouveau protocole, nous avons préparé deux types de métabolomes : l'un selon le protocole utilisé précédemment par le laboratoire (ANNEXE 2), l'autre selon le protocole développé pour le criblage de la banque. Au total 12 métabolomes ont été extraits selon le protocole plaque et 3 métabolomes ont été préparés selon l'ancien protocole. Nous avons alors analysé ces deux types de métabolomes sur une même méthode chromatographique : le gradient de 50 minutes sur ZIC-*p*HILIC 150 x 4,6, 5µm.

Nous avons alors comparé les abondances de différents métabolites, identifiés sur la colonne de 15 cm grâce à l'injection des standards commerciaux. Ce travail est illustré dans la Figure 25.

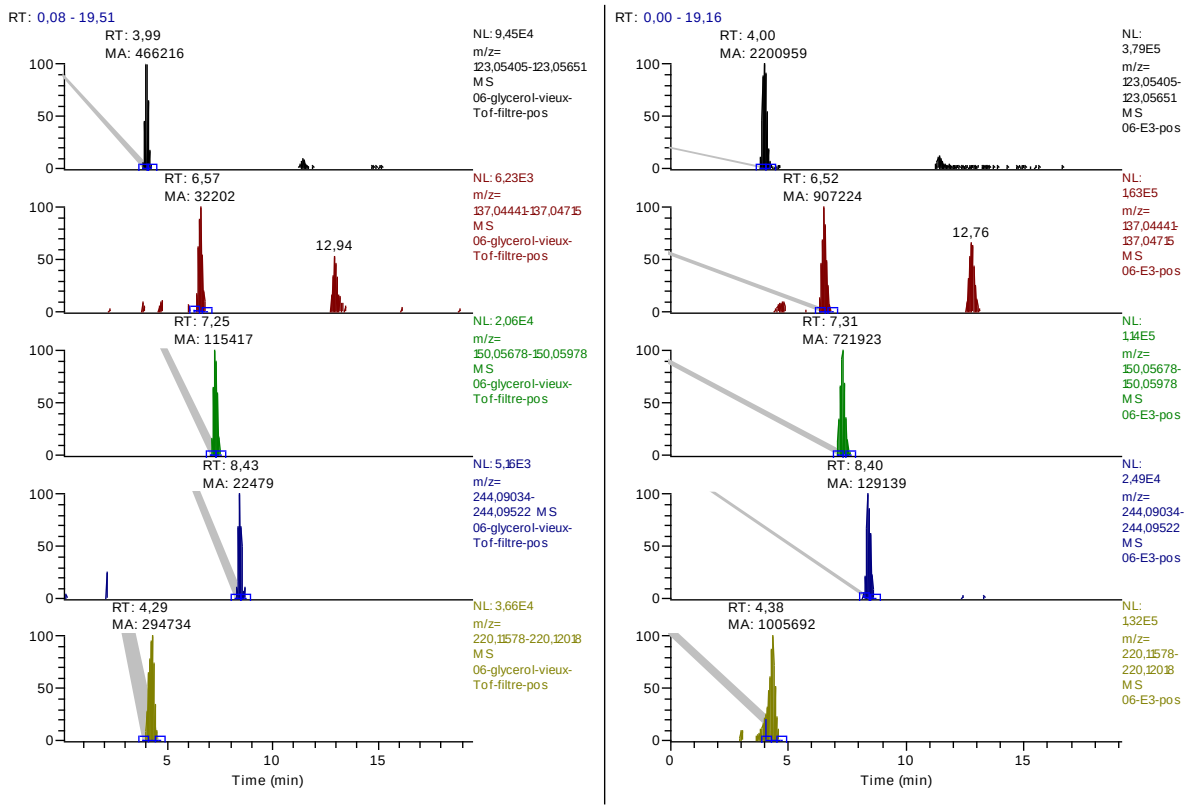


Figure 25: Comparaison de la qualité des différentes méthodes d'extraction au laboratoire. A gauche, EIC (5ppm) de métabolites extraits selon le protocole d'extraction des métabolites utilisé pour le criblage ; à droite, EIC (5ppm) de métabolites extraits selon le protocole anciennement utilisé par le laboratoire. Au-dessus de chaque pic se trouve le temps de rétention du métabolite ainsi que l'aire du pic, intégré manuellement. Du haut vers le bas : Nicotinamide (m/z123,05528, t_R = 4,19 min), Hypoxanthine (m/z137,04578, t_R=6,5 min), Méthionine (m/z150, 05828, t_R = 7,3 min), Cytidine (m/z244,09278, t_R = 8,79 min), Acide pantothénique (m/z 220,11798, t_R = 4,48 min).

La première différence observable d'un protocole d'extraction à l'autre est la différence d'intensité des

métabolites : en effet, ceux-ci sont plus abondants lors de la préparation des métabolomes *via* l'ancien protocole d'extraction utilisé par le criblage. On peut supposer que ce genre de différence est induit par les différences entre les deux protocoles d'extractions de métabolites. Plusieurs paramètres peuvent influencer sur l'abondance des métabolites : l'étape de lyse des cellules n'est pas identique (solvants et chocs thermiques différents), de plus, l'ancien protocole de préparation de métabolome contient une étape de « quenching ». Il s'agit d'une étape permettant de figer le métabolisme afin d'éviter de potentiellement perdre des métabolites de demi-vie très courte, sensibles aux stress métaboliques. Il est alors possible d'expliquer la baisse d'intensité de certains métabolites par la suppression de cette étape dans le nouveau protocole d'extraction de métabolites.

Cependant, cette baisse d'intensité n'est pas suffisante pour être problématique dans le cas de nos analyses. Nous avons donc considéré le nouveau protocole comme étant adapté au projet.

3.1.3 Choix des Métabolites Orphelins

Dans un premier temps, il s'agit de repérer (lorsque plusieurs pics correspondent à un même rapport m/z) les MO, détectés initialement sur la colonne de 15 cm, sur la version de 5 cm. La méthode d'analyse chromatographique étant modifiée, il a fallu déterminer leur temps de rétention sur la nouvelle colonne.

3.1.3.1 Les MO identifiés, non prédits par le modèle métabolique de la bactérie

Une partie des MO est celle des métabolites identifiés, qui ne sont pas prédits par le modèle métabolique d'ADP1. Ils sont au nombre de 18.

Nous possédons les standards commerciaux de ces composés. Il est donc aisé de connaitre leur temps de rétention sur colonne de 5 cm. Cependant, il est important de considérer le potentiel « effet matrice » (ou suppression d'ionisation évoquée en introduction) dans la détection de ces métabolites. Nous avons donc extrait une plaque 24 puits de métabolomes issus de la souche sauvage cultivée sur quinate. Un standard différent a été ajouté à 18 d'entre eux, à une concentration finale de 50 µM. Chacun des mélanges a été injecté sur le système LC-MS. Nous avons ainsi déterminé les temps de rétention de chacun des standards. Ceux-ci sont présentés dans le Tableau 6.

Tout d'abord, nous remarquons que la diminution du temps de la méthode chromatographique entraîne une superposition de certains pics, phénomène particulièrement problématique dans le cas de métabolites de même masse comme la tyrosine et β-tyrosine, les acides para et méta amino-benzoïque et les méthylxanthines. Ceux-ci ne sont pas distinguables sur la colonne de 5 cm et sont logiquement écartés de notre étude.

De plus, d'autres métabolites (en noir dans le Tableau 6) ne sont simplement pas détectés dans les métabolomes. Ces métabolites, probablement victimes de suppression d'ionisation, sont également écartés. Au total, nous pouvons suivre 8 MO identifiés dans les métabolomes de mutants (en vert dans le tableau 5).

			Temps de rétention (min) sur Zic-*p*HILIC 150 x 4,6 mm², 5 µm	**Temps de rétention (min) sur Zic-pHILIC 50 x 2,1 mm², 5 µm**	
Métabolite	**[M+H]+**	**[M-H]-**		**métabolome avec ajout de standard**	**métabolome sans ajout de standards**
γ-aminobutyric acid	104,07061	102,05605	12,17	2,84	**pos & neg**
3-methyl-2-oxobutanoic acid	**117,05462**	**115,04007**	**2,75**	**0,7**	**neg**
mono-Methyl hydrogen succinate	**133,04954**	**131,03498**	**3,13**	**0,73**	**neg**
oABA	**138,05496**	**136,0404**	**3,2**	**0,85**	**pos & neg**
pABA	138,05496	136,0404	5,35	1,0 (non distinguable)	x
mABA	138,05496	136,0404	8,55	1,0 (non distinguable)	x
Mevalonic acid	**149,08084**	**147,06628**	**3,25**	**Non**	**neg**
3-Hydroxy 3-methylglutaric acid	163,05337	161,04555	9,9	x	x
1-Methylxanthine	167,05635	165,0418	4,62	0,82 (non distinguable)	neg (faible)
7-Methylxanthine	167,05635	165,0418	4,95	0,82 (non distinguable)	neg (faible)
N-Acetyl-L-aspartic acid	**176,05535**	**174,0408**	**10,66**	**3,27**	**pos & neg (faible)**
N-formyl-L-methionine	178,05324	176,03869	3,13	0,74	x
Tyrosine	182,08117	180,06662	9,09	1,3 (non distinguable)	x
β-Tyrosine	182,08117	180,06662	8,47	1,3 (non distinguable)	x
Ala-Leu	203,13902	201,12447	3,13	0,67	x
Leu-Pro	229,15467	227,14012	3,12	0,71	x
Glu-Glu	**277,10303**	**275,08847**	**14,1**	**4,32**	**pos & neg (faible)**
5'-Deoxy-5'-(methyl-thio) adenosine	**298,09684**	**296,08228**	**3,9**	**0,72**	**pos & neg**

*Tableau 6: Liste des métabolites identifiés, non prédits par le modèle informatique de la bactérie détectés avec la méthode analytique de 50 min (ZIC-*pHILIC *510 x 4,6, 5µm). En vert, métabolites détectés avec la méthode analytique utilisée pour le criblage dans les métabolomes extraits selon le protocole d'extraction des métabolites utilisé pour le criblage. $[M+H]^+$: valeur m/z du métabolite en mode d'ionisation positif, $[M-H]^-$: valeur m/z du métabolite en mode d'ionisation négatif.*

Au total, nous pourrons suivre la présence ou l'absence de 8 métabolites identifiés mais non prédits par le modèle informatique de la bactérie dans les métabolomes de mutants (en vert dans le Tableau 6).

3.1.3.2 Les MO inconnus

Les MO non identifiés et non attendus, essentiellement présents dans des cellules qui utilisent le quinate (liste en ANNEXE 3) n'ont été jusqu'ici caractérisés que par leur temps de rétention sur la colonne de 15 cm et leur masse précise. Nous avons choisi de nous concentrer sur les MO dont les abondances (plus précisément les intensités de signaux) dans les « métabolomes quinate » sont les plus importantes. Cette liste de MO sélectionnés est présentée dans le Tableau 7.

POSITIF				NEGATIF			
Identifiant	**[M+H]+**	**masse neutre**	**Temps de rétention (min)**	**Identifiant**	**[M-H]-**	**Masse neutre**	**Temps de rétention (min)**
MO129	129,0656	128,05778	6,14	MO115T240	115,04048	116,0483	3,9
MO157	157,0968	156,08898	8,59	MO115T720	115,04052	116,04834	12
MO165	165,1019	164,09408	15,89	MO152	152,03586	153,04368	9,5
MO167	167,1176	166,10978	5,23	MO175	175,02527	176,03309	15,9
MO193	193,09664	192,08882	4,6	MO187T750	187,02516	188,03298	12,5
MO201	201,12271	200,1148916	23,5	MO187T702	187,02521	188,03303	11,7
MO204	204,08599	203,0781666	8,6	MO189	189,04079	190,04861	5,6
MO211	211,10704	210,0992216	14,5	MO203T1014	203,02005	204,02787	16,9
MO219	219,13317	218,1253514	23,2	MO203T828	203,02009	204,02791	13,8
MO221	221,09127	220,0834469	10,1	MO213	213,01736	214,02518	8,5
MO245T582	245,11231	244,1044878	9,7	MO219	219,01481	220,02263	17
MO245T300	245,11241	244,1045895	5	MO245	245,04368	246,0515	9,2
MO263	263,06877	262,0609488	9,8	MO314	314,03609	315,04391	9,7
MO265	265,11739	264,1095678	8,7	MO331	331,06194	332,06976	11,8
MO265	265,11742	264,1096041	8,3	MO345	345,0844	346,09222	9,6
MO313	313,08109	312,0732683	2,9	MO349	349,00375	350,01157	9,7
MO314	314,08921	313,081393	8,7	MO379	379,19529	380,20311	2,6
MO333	333,0742	332,06638	3.6				
MO339	339,15401	338,1461913	9,6				
MO353	353,13336	352,1255364	5,3				
MO452	452,27634	451,2685249	2,7				
MO851	851,34869	850,3408677	11,6				

Tableau 7: Liste des MO inconnus retenus pour le projet, avec leur masse précise dans les deux modes d'ionisation et leur temps de rétention chromatographique sur la colonne de 15 cm.

Afin de s'assurer des temps de rétention de ces ions sur la colonne de 5 cm, nous avons tout d'abord effectué une « purification analytique » du métabolome, en collectant manuellement l'éluât de la colonne analytique en plusieurs fractions. Un métabolome issu de la souche sauvage a été injecté sur la colonne de 15 cm et collecté en échantillonnant toutes les minutes. Nous avons obtenu ainsi 23 fractions que nous avons séchées, reprises dans 20 µL de phase mobile (ACN/pH 9,9 80 : 20) puis réinjectées sur chacune des deux colonnes analytiques (15 et 5 cm). L'analyse d'une fraction sur la colonne de 15 cm

indique sa composition. Son analyse sur la colonne de 5 cm permet d'attribuer les signaux observés. Cette approche est illustrée en Figure 26 au travers de l'exemple du MO129 détecté en mode d'ionisation positif.

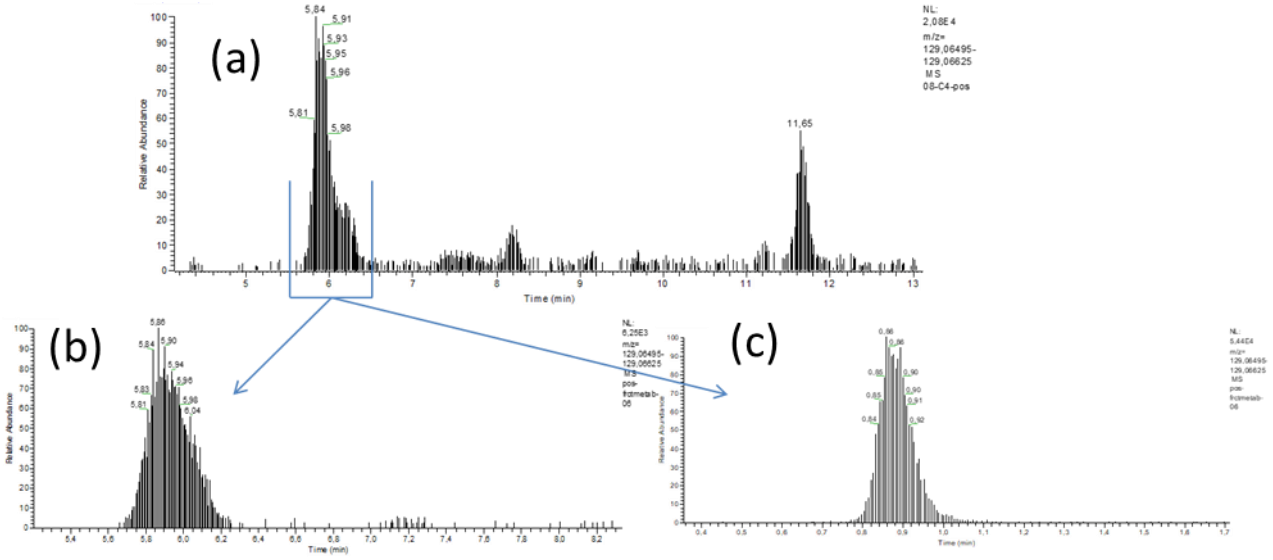


Figure 26: Schéma explicatif de l'analyse d'une fraction du « métabolome fractionné » sur deux méthodes analytiques différentes. (a) - EIC du MO129 (5ppm) issu de l'analyse du métabolome sur la colonne de 15cm ; (b) - EIC du MO129 (5 ppm) issu de l'analyse de la fraction n°5 (5,5 min à 6,5 min) sur la colonne de 15 cm ; (c) - EIC du MO 129 (5 ppm) issu de l'analyse de la fraction n°5 sur la colonne de 5 cm.

En analysant les correspondances entre les résultats d'analyses de ces fractions sur les deux colonnes, nous avons pu assigner les MO à leur signal. Cependant, certains MO n'ont toutefois pas pu être assignés avec cette approche. En effet, certains n'ont pas pu être détectés dans les fractions obtenues après la purification du métabolome. Ces métabolites peuvent être, par exemple, trop peu stables pour pouvoir être encore présents dans les fractions obtenues une fois séchées et reprises et réinjectées.

La liste des MO sélectionnés pour le projet est présentée dans le Tableau 8.

Afin de caractériser plus de métabolites sur la colonne de 5 cm, nous pourrions envisager des purifications sélectives des MO « perdus » lors de la transition entre les colonnes de 15 et 5 cm, mais à une échelle un peu plus importante (en injectant et collectant plusieurs fois un même métabolome et en regroupant les fractions correspondantes, par exemple). Cependant, les métabolites ne sont pas tous aisément purifiables, pour des raisons de trop faible concentration dans les métabolomes ou peut-être des problèmes de stabilité. Il est toutefois possible de caractériser les MO *via* leur signature MS/MS. Pour cela, il faut considérer les MO sur la colonne de 15 cm en enregistrant leur signature MS^2 et en comparant ainsi les masses précises et les signatures MS/MS sur les deux colonnes, nous pouvons attribuer certains MO aux signaux détectés sur la colonne de 5 cm.

Cette approche est illustrée au travers de deux exemples : le MO193 et le MO211 (tous deux observés en mode d'ionisation positif).

Le temps de rétention sur la colonne de 15 cm du MO193 est de 4,6 minutes. A la masse du MO193, nous observons deux pics chromatographiques sur chacune des colonnes : 4,6 et 14,6 minutes sur la colonne de 15 cm, contre 0,79 et 3,71 minutes sur la colonne de 5 cm. Afin de bien repérer le MO193 sur la colonne de 5 cm, nous avons comparé sa signature MS/MS sur chaque colonne (Figure 27). Ainsi, nous avons pu associer le pic m/z 193,0966 lié au temps de rétention 4,6 minutes sur la colonne de 15 cm au pic m/z 193,0966 lié au temps de rétention 0,79 min sur la colonne de 5 cm.

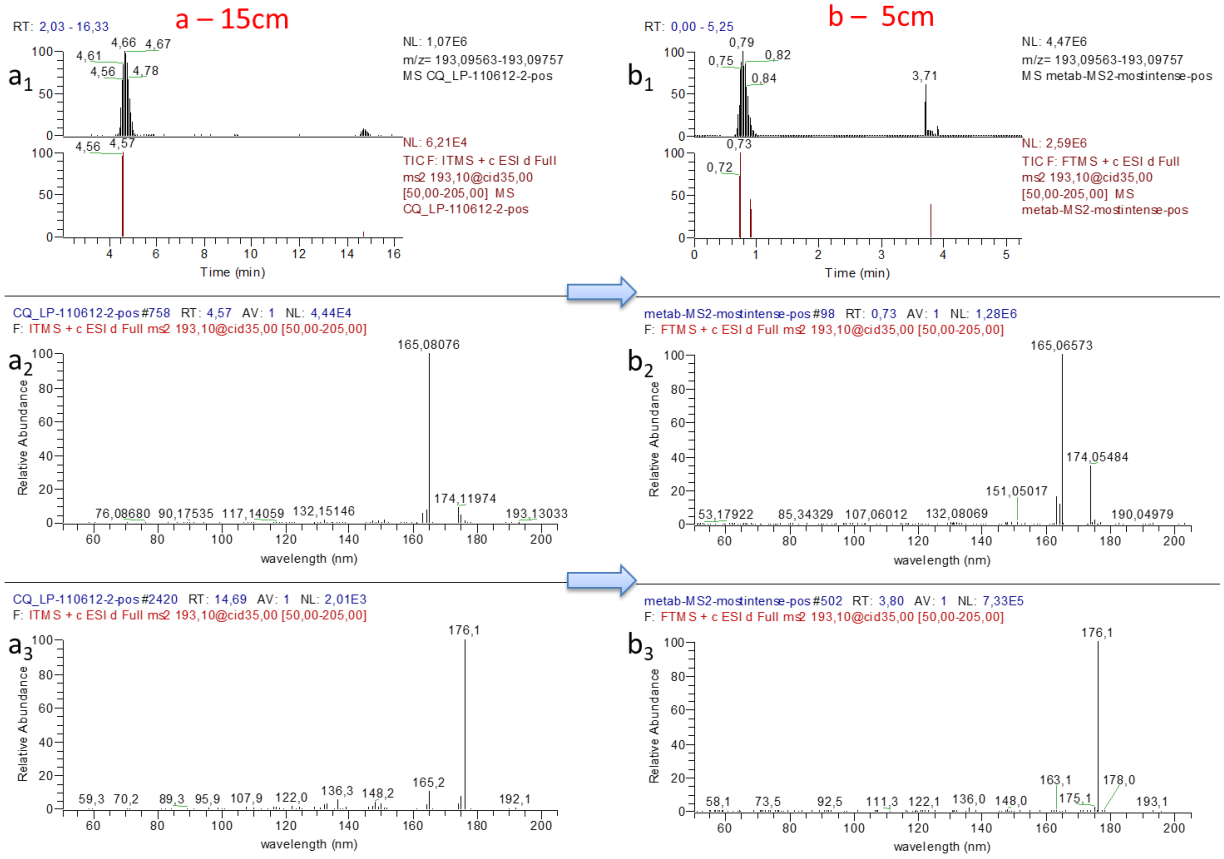


*Figure 27: Caractérisation du MO193 en LC/MS **(a)** sur la colonne de 15 cm, **(b)** sur la colonne de 5 cm. **(a₁)**: EIC du m/z 193,0966 (5ppm), **(a₂)**: Fragmentation CID 35% m/z= 193,0966 au temps t_R=4,57 min, **(a₃)** fragmentation CID 35%, m/z=193,0966 au temps t_R=14,69 min; **(b₁)**: EIC du m/z 193,0966 (5ppm) **(b₂)**: Fragmentation CID 35% m/z= 193,0966 au temps t_R=0,73 min, **(b₃)** fragmentation CID 35%, m/z=193,0966 au temps t_R=3,80 min. Les ions issus de la fragmentation en mode CID sont ici détectés dans le piège à ions.*

Dans le cas du MO211, (temps de rétention t_R = 14,6 minutes sur la colonne de 15 cm), sa fragmentation dans la source est singulière (paragraphe 3.1.2.4). Ce MO semble perdre un fragment de composition CH_5N (Δm = 31 u) et donne un pic à m/z 180, d'intensité équivalente à celle du parent (Figure 28). Il présente aussi un fragment à m/z 193,0966, issu d'une perte d'eau de l'ion parent (Δm = 18 u). Cet ion

correspond précisément au second pic détecté à la masse du MO193 (Figure 27). Enfin, la fragmentation du MO211 en mode CID montre la présence de deux ions m/z 193 et m/z 194. Cette signature en MS^2 ainsi que sa fragmentation dans la source suffisent à le caractériser. Ces deux types de données sont utilisés pour valider son suivi d'une colonne à l'autre.

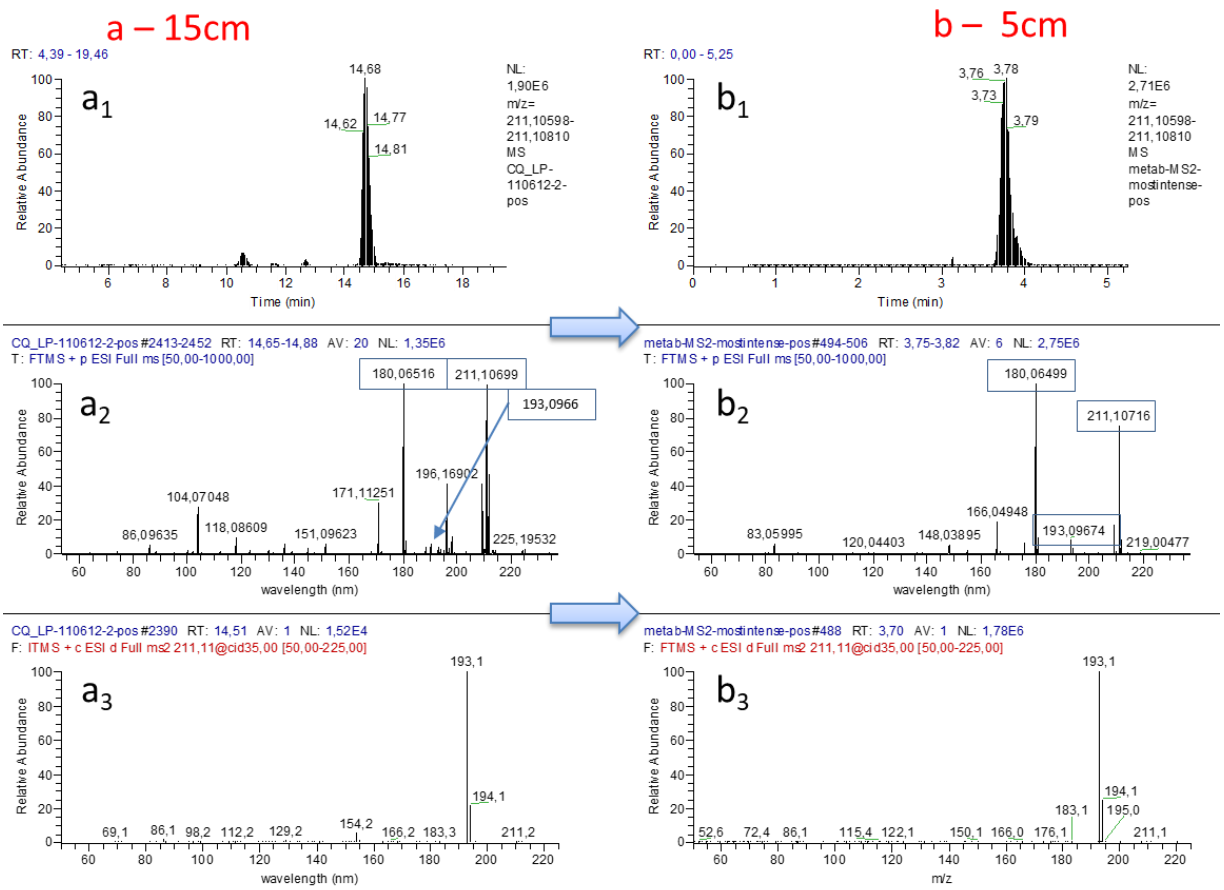


Figure 28: Caractérisation du MO211 en LC/MS (a) sur la colonne de 15 cm, (b) sur la colonne de 5 cm. ***(a_1)****: EIC du MO211 (5 ppm),* ***(a_2)****: détection en FTMS autour du temps de rétention du MO211,* ***(a_3)*** *fragmentation CID 35% m/z211,10704 au temps t_R=14,51min* ***(b_1)****: EIC du MO211 (5 ppm),* ***(b_2)****: détection en FTMS autour du temps de rétention du MO211,* ***(b_3)*** *fragmentation CID 35% m/z211,10704 au temps t_R=3,70 min. Les ions issus de la fragmentation en mode CID sont ici détectés dans le piège à ions.*

Au total, nous avons validé la correspondance entre les deux colonnes de 19 MO différents, chacun étant détectable préférentiellement (ou uniquement) dans un mode d'ionisation donné. La liste des MO est présentée dans le tableau 8.

Identifiant	[M+H]+	masse neutre	Formule brute	Temps de rétention (min) ZIC pHILIC (150 x 4,6, 5µm)	ZIC pHILIC (150 x 2,1, 5µm)
			POSITIF		
MO129	129,0656	128,05778	$C_5H_8O_2N_2$	6,14	0,87
MO157	157,0968	156,08898	$C_7H_{12}O_2N_2$	8,59	1,07
MO167	167,1176	166,10978	$C_9H_{14}ON_2$	5,23	0,81
MO193	193,09664	192,08882	$C_{10}H_{12}O_2N$	4,6	0,78
MO211	211,10704	210,09922	$C_{10}H_{14}O_3N_2$	14,5	3,7
MO219	219,13317	218,12535	$C_9H_{18}O_4N_2$	23,2	4,65
MO221	221,09127	220,08345	$C_{11}H_{10}O_3N_2$	10,1	1,51
MO245	245,11231	244,10449	$C_{10}H_{16}O_5N_2$	9,7	3,07
MO245T582	314,08921	313,08139	$C_{14}H_{13}O_4N_5$	8,7	1,07
MO452	452,27634	451,26852	non déterminée	2,7	0,59
			NEGATIF		
MO115T240	115,04048	116,0483	$C_5H_8O_3$	3,9	0,69
MO175	175,02527	176,03309	$C_6H_8O_6$	15,9	4,45
MO187T702	187,02521	188,03303	$C_7H_8O_6$	11,7	3,21
MO189	189,04079	190,04861	$C_7H_{10}O_6$	5,6	0,94
MO203T1014	203,02005	204,02787	$C_7H_8O_7$	16,9	4,78
MO203T828	203,02009	204,02791	$C_7H_8O_7$	13,8	3,9
MO314	314,03609	315,04391	non déterminée	9,7	1,95
MO345	345,0844	346,09222	non déterminée	9,6	1,3
MO349	349,00375	350,01157	non déterminée	9,7	1,95

Tableau 8: Liste des MO inconnus dont le suivi est possible pour le criblage de la banque de mutants.

3.1.4 Traitement des données

Chaque semaine, le criblage de la banque génère environ 400 fichiers «.raw » dont il faut extraire toute l'information utile. Le traitement manuel d'une telle quantité de données est inenvisageable. Pour aider à l'analyse, nous avons utilisé XCMS Online. Cet outil, présenté en introduction de ce manuscrit (1.2.3.2) nous permet d'obtenir en quelques heures, pour chaque plaque de 96 puits, un tableau récapitulatif des aires chromatographiques de tous les ions détectés dans chaque puits de la plaque. En d'autres termes, les fichiers correspondant à chaque plaque de 96 puits sont prétraités de façon indépendante par XCMS.

.

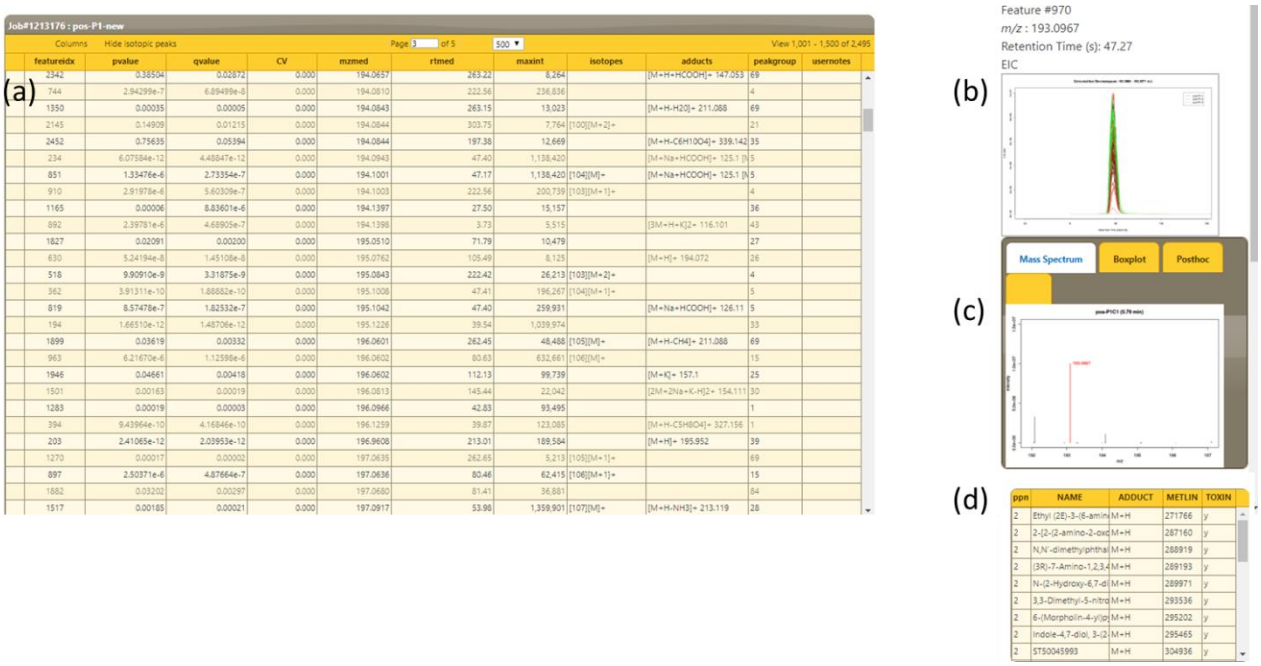


Figure 29: Capture d'écran d'une matrice de résultats issu d' XCMS Online comprenant : (a) la matrice de résultats interactive, (b) la superposition de tous les EIC correspondant à un même ion dans chacun des puits d'une plaque (c) les détails du spectre de masse de l'ion d'intérêt, (d) ses identifications potentielles selon la base de données METLIN.

Le tableau de résultats contient la totalité des informations nécessaires à notre analyse des données : la liste des masses précises de tous les ions détectés dans tous les échantillons, leur temps de rétention, ainsi que les aires de leurs pics chromatographiques dans chacun des puits de la plaque analysée (Figure 29a). Il permet aussi notamment d'avoir accès à l'annotation des isotopes et adduits proposée par CAMERA (Figure 29a) ainsi qu'aux propositions de structures proposées par la base de données METLIN (Figure 29 d). De plus, cette interface permet de visualiser pour chacune des masses présentées dans le tableau son EIC (figure 29 b), les propositions et d'autres informations complémentaires (Figure 29 c).

Plusieurs paramétrages par défaut sont disponibles selon l'instrumentation analytique utilisée (selon le type de chromatographie : UPLC ou HPLC et GC ; et selon le type de spectromètre de masse : Q-TOF, triple quadripôle, piège à ion, Orbitrap). Ces paramètres peuvent être modifiés et optimisés par l'utilisateur. Le prétraitement de données permet de visualiser les résultats en ligne, avec un panel d'analyses et de représentations statistiques (Figure 30). Nous avons utilisé le paramétrage proposé par XCMS Online « HPLC Orbitrap II », que nous avons adapté à notre projet.

Le paramétrage d'XCMS est présenté dans le Tableau 20 (2.4.3).

Le rapport signal/bruit (S/N threshold) proposé dans le paramétrage « HPLC Orbitrap II » est fixé à 6. Nous avons cependant choisi un rapport signal/bruit plus bas, à 3. Plus le rapport signal sur bruit est bas,

plus on peut observer l'apparition de pics « fantômes » par XCMS dans les tables de résultats. Le logiciel peut considérer dans le bruit des signaux qu'il intègre comme un pic, un métabolite « fantôme ». Dans ce cas, un travail de curage manuel serait nécessaire pour éliminer ces artefacts. Cependant, dans le cas de notre analyse, ciblée, il est essentiel de détecter tous les MO étudiés, même ceux à l'origine d'un signal peu intense. En choisissant un rapport signal/bruit bas, nous augmentons la probabilité de détecter les MO pour lesquels le signal peut être faible. D'un autre côté, ces variables « fantômes » ne perturbent pas notre analyse ciblée.

Pour des raisons similaires, un autre paramètre a été modifié : le « minfrac », paramètre correspondant à la valeur minimale de la fraction du nombre total des échantillons dans lesquels le signal correspondant à un métabolite donné doit être détecté pour être considéré en tant qu'ion dans la liste finale. Comme les mutants diffèrent les uns des autres par la délétion d'un seul gène, nous supposons que le métabolome de chacun est globalement proche de celui de la souche sauvage. Nous faisons donc l'hypothèse que les métabolomes des mutants sont globalement similaires. Ainsi, si un métabolite d'intérêt est absent du métabolome d'un mutant donné, il devrait être présent dans les autres métabolomes. Suivant ces hypothèses, nous aurions pu fixer le « min frac » à 0,9, par exemple. Cependant, le travail effectué ici est une analyse ciblée des métabolomes : l'apparition de pics moins pertinents que d'autres dans notre tableau de résultats n'est donc pas un problème car elle ne gêne pas l'analyse. En effet, nous savons quel métabolite rechercher dans ces résultats. Nous avons donc pris le parti d'utiliser un « minfrac » plus bas, à 0,75. Ainsi, un signal sera considéré comme un pic s'il est détecté dans au moins 75% des puits de la plaque analysée. Dans ces conditions, la probabilité de « perdre » un MO lors du prétraitement par XCMS doit être plus réduite.

Nous avons évalué la capacité d'XCMS à intégrer des pics de faible intensité. Ce travail est illustré dans la Figure 30 qui présente les EIC de deux MO, (MO193 et MO274) qui présentent des différences d'intensité de 2 à 3 ordres de grandeur. L'intensité du signal du MO193 varie de 10^6 à 10^7, alors que celle du MO274 est comprise entre 10^3 et 10^4. Nous remarquons que dans le second cas, même si le signal de l'ion est particulièrement faible, XCMS le détecte au travers de tous les échantillons et l'intègre de manière fiable.

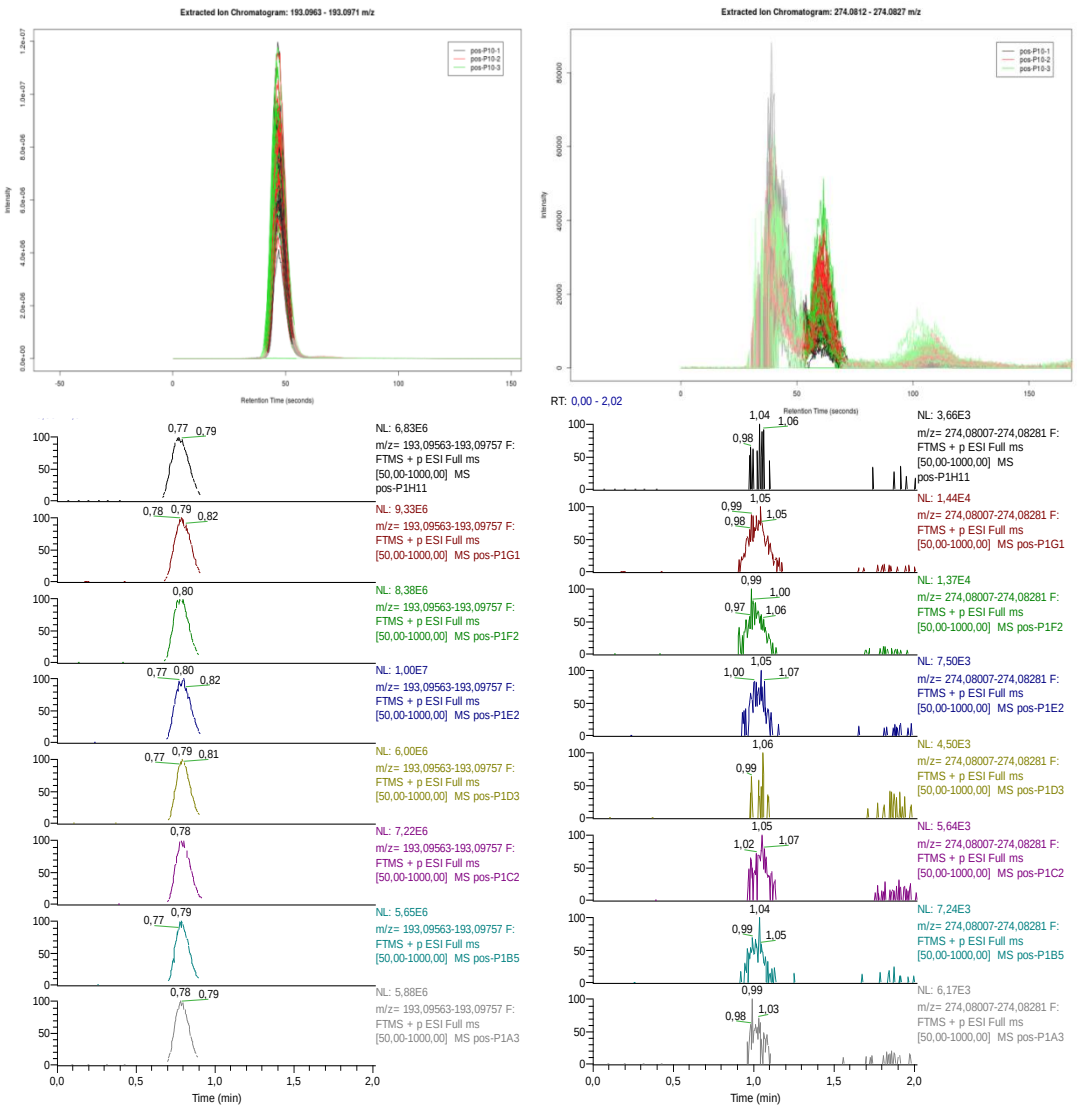
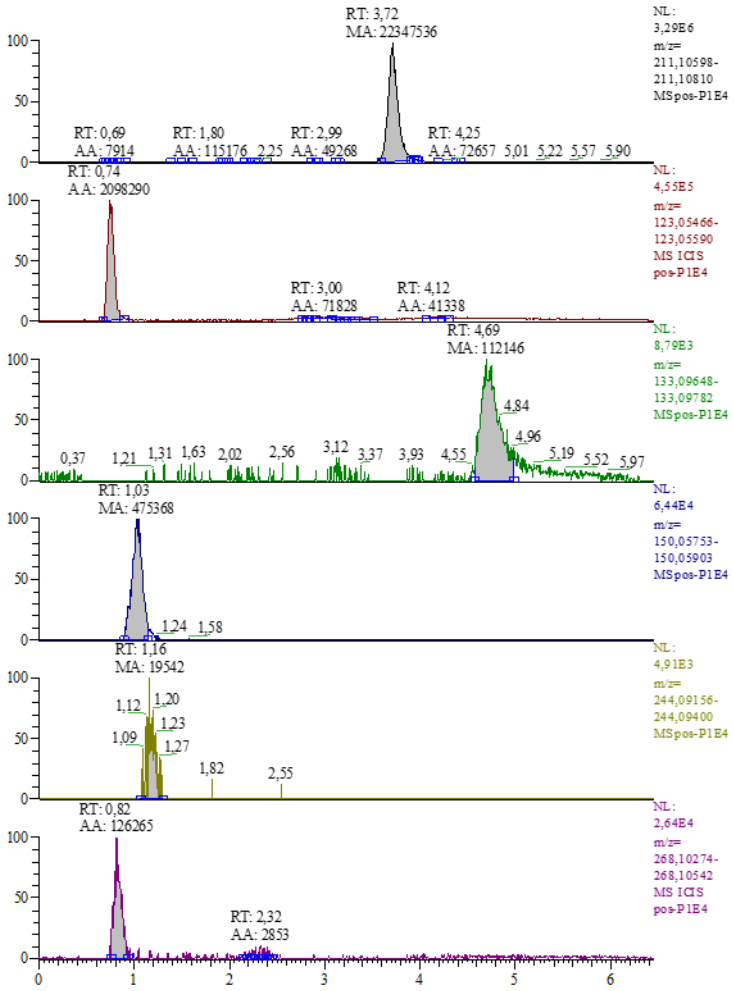


Figure 30: Évaluation de la qualité du prétraitement de données par XCMS (détection des pics et alignement des temps de rétentions) pour le MO193 (à gauche) et le MO274 (à droite) d'une plaque de 96 puits en mode d'ionisation positif. HAUT : superposition des EIC du MO obtenu dans XCMS. BAS : superposition d'EIC extraits des données brutes (.RAW). Les 8 chromatogrammes correspondent à huit injections de puits différents : du bas vers le haut : A3, B5, C2, D3, E2, F8, G1 et H11.

Afin d'évaluer son travail d'intégration, nous avons comparé les valeurs d'intégrations de plusieurs métabolites d'intensité variable au sein d'une plaque, déterminée par XCMS, avec celles obtenues manuellement. Ce travail effectué pour plusieurs puits, au sein de plusieurs plaques est illustré Figure 31 et les résultats montrent que les valeurs d'intégrations obtenues par XCMS sont particulièrement pertinentes.



Métabolite	[M+H]+	aire (calculée à la main)	aire (XCMS)
MO211	211,10704	2,23E+07	2,16E+07
Nicotinamide	123,05528	2,10E+06	2,07E+06
Ornithine	133,09715	1,12E+05	1,04E+05
Methionine	150,05828	4,75E+05	4,88E+05
Cytidine	244,09278	1,95E+04	2,91E+04
Adenosine	268,10408	1,26E+05	2,61E+05

Figure 31: Comparaison des données brutes aux valeurs de fold obtenues via XCMS ou calculées à la main, du puits P1E4.

Malgré tout, analyser manuellement les données du tableau de sortie d'XCMS Online serait peu efficace. Les données prétraitées par XCMS sont donc stockées dans une base de données (développée au Genoscope) puis interrogées par un système de requêtes réadapté à nos besoins, Business Object. Ce logiciel permet d'éditer des rapports pré formatés en ciblant les MO. Les données de chacune des plaques de 96 métabolomes sont traitées individuellement.

3.1.5 Analyse des données

L'analyse des données du criblage correspond à repérer dans les différentes plaques 96 puits, les puits (et donc les mutants qui y sont associés) dans lesquels un MO est absent, ou pour lequel la surface de son pic chromatographique est significativement différente de sa valeur habituelle (dans la souche sauvage).

La séquence d'injection dans un mode d'ionisation donné se déroule de la manière suivante (Figure 32) :

1) injection de deux « blancs (flacon) »

2) injection des puits de la première ligne de la plaque (A1 à A12)

3) injection d'un « blanc » (flacon)

4) injection des puits d'une ligne de la plaque (B1 à B12)

5) injection d'un « blanc » (flacon)

…

16) injection des puits H1 à H12

17) injection d'un « blanc » (flacon)

Ce processus est ensuite recommencé dans l'autre mode d'ionisation.

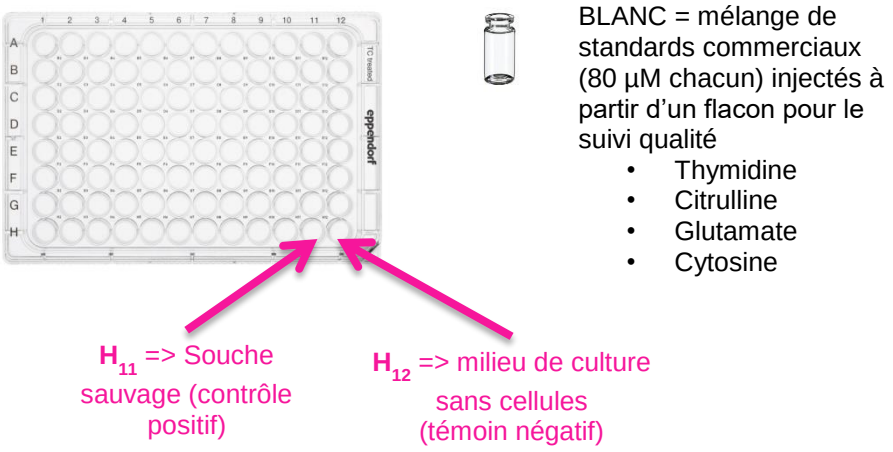


Figure 32: Plan d'une plaque 96 puits de la banque de mutants: Les puis A1 à H10 contiennent chacun un mutant. Le puits H11 est ensemencé avec la souche sauvage, le puits H12 ne contient que du milieu de culture sans cellules. Un flacon, indépendant de la plaque, est placé dans le passeur d'échantillons, contenant un mélange de standards commerciaux (« blanc »).

L'utilisation de « blancs » dans la séquence d'injections d'échantillons d'une plaque est discutée plus loin (paragraphe 3.1.6.2).

Deux outils : le *fold* et le Z-score

Afin de pouvoir repérer rapidement dans les différentes plaques les mutants dans lesquels un MO est impacté, nous avons considéré deux outils: le *fold* et le Z-score.

Le *fold* se défini de la manière suivante.

$$fold = \frac{AIRE_j^i}{AIRE_{H11}^i}$$

$AIRE_j^i$: Aire du pic du MOi dans le puits j de la plaque analysée

$AIRE_{H11}^i$: Aire du pic du MOi dans le puit H11(souche sauvage) de la plaque analysée

Il permet de comparer rapidement la valeur de l'aire du pic du MO d'intérêt dans un puits de la plaque à ce même paramètre dans le métabolome de référence de la plaque, à savoir celui issu de la souche sauvage : le puits H11.

Trois cas peuvent se présenter :

- *fold* environ égal à 1 : le MO est présent en quantité similaire dans le métabolome issu du mutant et dans celui issu du sauvage. Le mutant n'est pas affecté dans le métabolisme du MO.

- *fold*>1 : le métabolite s'accumule dans le métabolome du mutant. Ce résultat peut s'interpréter comme une accumulation du MO dans un mutant où le gène codant pour l'enzyme qui l'utilise comme substrat est excisé.

- *fold*<1 : le métabolite est largement moins présent dans le métabolome issu du mutant. Le MO ne serait pas formé car le gène impliqué dans sa synthèse est excisé.

Bien que cet outil soit simple à mettre en place, il possède un inconvénient. En effet, nous utilisons comme référence l'aire du pic du MO d'intérêt dans le puits H11. Or, il est possible que la culture cellulaire ou l'extraction des métabolites de la souche sauvage présente un problème. Dans ce cas, les métabolites de ce puits sont peu abondants. Ceci peut mener à la surévaluation de la valeur du *fold* du MO pour les autres puits et conduire à des conclusions erronées. Nous avons donc considéré également un autre outil pour s'affranchir de ce biais: le Z-score.

La standardisation par le Z-score permet, dans notre cas, de quantifier l'abondance de chaque ion dans un puits donné par rapport à son abondance dans tous les autres puits de la plaque.

Le Z-score se calcule de la manière suivante :

$$Z - score = \frac{AIRE_j^i - MOY_{95}^i}{S}$$

$AIRE_j^i$: Aire du pic du MOi dans le puits j de la plaque analysée

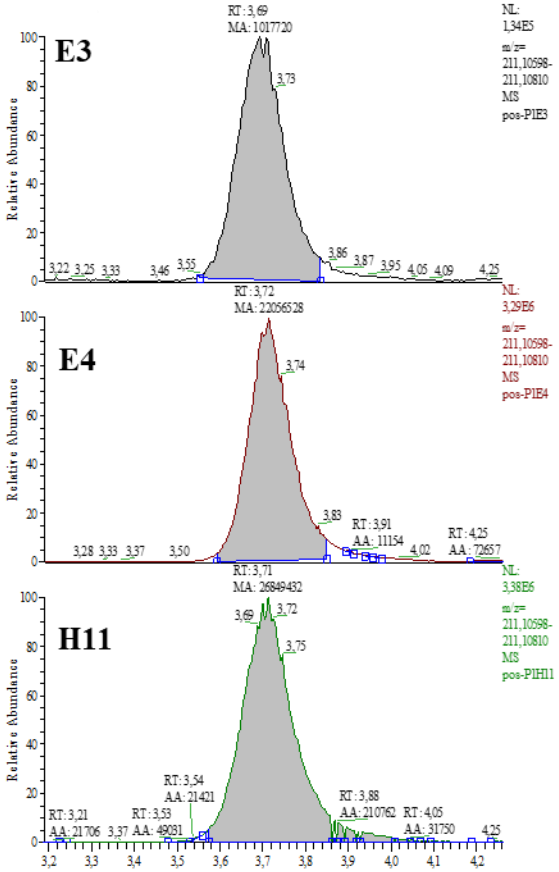
MOY_{95}^i : Moyenne des aires du MOi de chacun des 95 puits de la plaque (sauf H12)

S : déviation standard des valeurs à la moyenne

Le puits H12 ne contient pas de métabolome. L'intégrer au calcul du Z-score diminuerait la moyenne des valeurs de la plaque et biaiserait donc le calcul.

Deux cas peuvent se présenter : si │Z-score│> 'valeur seuil' (fixée arbitrairement), la teneur du MO est considérée comme différente dans le puits concerné par rapport aux autres puits de la plaque. Si │Z-score│< 'valeur seuil', la teneur du MO dans ce puits est considérée comme comparable à la moyenne des aires de ce même métabolite dans les autres puits de la plaque.

Nous avons comparé les valeurs de *fold* et Z-score de plusieurs métabolites dans les données brutes LC-MS (fichiers .RAW analysés manuellement) de plusieurs puits. Nous avons estimé que : si fold <0,3 ou fold >3 et/ou │Z-score│>2, le mutant est affecté par la délétion du gène excisé dans la production/consommation du métabolite d'intérêt.



PLAQUE 1 - 211,10704 (+)									
fold					Z-score				
	01	02	03	04		01	02	03	04
A	0,73	0,72	0,74	0,66	A	-0,22	-0,24	-0,20	-0,45
B	0,78	0,66	0,59	0,60	B	-0,06	-0,44	-0,68	-0,66
C	1,19	0,87	0,82	0,71	C	1,29	0,25	0,08	-0,28
D	1,07	0,71	0,75	0,22	D	0,90	-0,27	-0,15	-1,90
E	0,88	1,29	0,04	0,83	E	0,26	1,61	-2,49	0,12
F	1,18	1,24	0,90	0,13	F	1,26	1,45	0,35	-2,18
G	1,31	1,04	0,97	0,49	G	1,68	0,79	0,58	-1,00

x,xx 0,1 < fold < 0,3
2,5 < IZ-scorel < 5

x,xx fold < 0,1
IZ-scorel > 5

Figure 33: Visualisation des fold et Z-score de quelques puits au travers de Business Object. Haut : EIC du MO211 (5 ppm) (intégré manuellement) dans trois puits de la plaque : du haut vers le bas : E3, E4, et H11 (souche sauvage). Bas, visualisation des fold *et Z-score du MO211 au sein de quelques puits de la plaque P1 après traitement des données par Business Object.*

Nous remarquons, sur l'EIC du MO211 présenté en Figure 33 que celui-ci est présent dans le puits E3. Cependant, son intensité en E3 est inférieure d'un ordre de grandeur par rapport à celle mesurée en E4 et H11. Cette différence se retrouve dans le traitement de données par Business Object, à la fois dans la valeur du *fold* (0,04 en E3) et du Z-score (-2,49 en E3).

3.1.6 Contrôles qualité

Avant d'analyser nos résultats, nous avons mis en place un suivi de la qualité de nos injections. Trois

étapes clés sont considérées : la culture, l'extraction et la détection des ions en LC-MS. Nous avons mis en place pour chacune de ces étapes des contrôles permettant d'en évaluer la qualité.

3.1.6.1 Suivi de la croissance cellulaire

Afin de s'assurer de la croissance de chacune des souches au sein d'une plaque, nous suivons celle-ci au travers de sa densité optique (à 600 nm). Cette mesure nous permet, lors de l'analyse des données *via* Business Objet, de savoir si la faible intensité du signal d'un métabolite dans un échantillon peut être imputée à un problème de croissance cellulaire. De plus, pour chaque plaque, la lyse des cellules est effectuée si la moyenne des valeurs de DO de tous les puits est comprise entre 0,4 et 0,6. Lors de l'analyse au travers de la base de données, la valeur de la DO pour chaque puits est indiquée par un code couleur, si celle-ci est inférieure à 0,5. Ce code couleur est représenté sur la Figure 34.

DO plaque 1

	01	02	03	04	05	06	07	08	09	10	11	12
A	0,56	0,54	0,59	0,53	0,56	0,55	0,54	0,47	0,57	0,54	0,52	0,56
B	0,46	0,56	0,53	0,58	0,56	0,59	0,57	0,54	0,59	0,59	0,56	0,54
C	0,43	0,52	0,55	0,52	0,53	0,49	0,54	0,54	0,47	0,54	0,53	0,50
D	0,51	0,59	0,56	0,41	0,55	0,61	0,55	0,59	0,56	0,60	0,60	0,55
E	0,47	0,36	0,37	0,58	0,53	0,53	0,55	0,56	0,57	0,52	0,59	0,52
F	0,44	0,55	0,58	0,26	0,54	0,23	0,55	0,62	0,52	0,57	0,58	0,57
G	0,49	0,52	0,59	0,58	0,53	0,57	0,51	0,57	0,57	0,55	0,54	0,56
H	0,48	0,60	0,57	0,56	0,57	0,57	0,56	0,63	0,54	0,58	0,54	0,04

0,3 < DO <0,5

0,1 < DO < 0,3

DO < 0,1

Figure 34: Visualisation des mesure de densité optique à 600 nm des cultures de la plaque 1 de la banque de mutants au travers de Business Object. Les valeurs de DO inférieures à 0,5 sont indiquées par un code couleur.

Lorsqu'un *fold* est considéré comme discriminant, nous contrôlons au travers de la base de données la croissance du mutant, en vérifiant la valeur de DO obtenue avant la lyse des bactéries. Nous pouvons donc distinguer rapidement une absence de MO dans un puits due à un problème de croissance cellulaire ou à l'excision d'un gène. Un exemple d'analyse d'une plaque au travers des valeurs du *fold* et du Z-score est présenté sur la Figure 35.

PLAQUE 1 - 211,10704 (+)									
FOLD					Z-score				
	01	02	03	04		01	02	03	04
A	0,73	0,72	0,74	0,66	A	-0,22	-0,24	-0,20	-0,45
B	0,78	0,66	0,59	0,60	B	-0,06	-0,44	-0,68	-0,66
C	1,19	0,87	0,82	0,71	C	1,29	0,25	0,08	-0,28
D	1,07	0,71	0,75	0,22	D	0,90	-0,27	-0,15	-1,90
E	0,88	1,29	0,04	0,83	E	0,26	1,61	-2,49	0,12
F	1,18	1,24	0,90	0,13	F	1,26	1,45	0,35	-2,18
G	1,31	1,04	0,97	0,49	G	1,68	0,79	0,58	-1,00

0,3 < DO <0,5

0,1 < DO < 0,3

DO < 0,1

x,xx 0,1 < fold < 0,3
2,5 < |Z-score| < 5

x,xx fold < 0,1
|Z-score| > 5

Figure 35: Illustration de l'analyse d'une plaque pour le MO211 au travers de Business Object. Analyse du MO211 sur les puits A1 à G4 de la plaque 1 au travers de son fold (gauche) et de son Z-score (droite). Dans le puits entouré en rouge (fold et Z-score), le fold et le Z-score semblent indiquer une absence du MO211 dans ce puits. La valeur de DO dans ce puits est comprise entre 0,3 et 0,5.

Le puits E3 de la plaque 1 affiche une valeur de *fold* de 0,04 et de Z-score de -2,49 pour le MO211 (Figure 35). Pour ce puits, la croissance des cellules est normale, car la DO au moment de l'extraction est supérieure à 0,3 (indiqué par le code couleur relatif à la DO). Le mutant de ce puits est potentiellement impliqué dans la synthèse du MO211. Le puits F4 affiche également des valeurs de *fold* et de Z-score susceptibles de montrer une implication de la mutation dans la synthèse du MO211 (0,13 et -2,18, respectivement). Toutefois, la valeur de la DO des cellules est faible et comprise entre 0,1 et 0,3 (comme l'indique le code couleur de la DO). Dans ce cas, les valeurs de *fold* et de Z-score pourraient n'être que le reflet d'un défaut de croissance du mutant considéré.

Cet exemple est à relier au fait que certains mutants peuvent avoir un temps de génération significativement supérieur à celui de la souche sauvage. Dans ces conditions, ces mutants ne présenteront pas une bonne densité cellulaire au moment de l'extraction des métabolites de la plaque. Leurs métabolomes ne pourront pas être correctement analysés au travers de Business Object dans ces conditions. Tous ces mutants seront dans un second temps regroupés selon leur vitesse de croissance et analysés ultérieurement.

3.1.6.2 Contrôle de la qualité de l'extraction

Un point primordial de la préparation à haut débit de métabolomes est la vérification de la qualité de l'extraction des métabolites. Celle-ci doit être similaire pour chacun des mutants. Pour évaluer la phase d'extraction, nous avons considéré 5 métabolites, détectables chacun dans les deux modes d'ionisation, que nous considérons comme des métabolites intracellulaires de référence (Tableau 9).

Métabolite	masse neutre	[M+H]+	[M-H]-	RT (min)
Nicotinamide	122,048	123,05528	121,04072	0,75
Ornithine	132,08987	133,09715	131,08259	4,73
Methionine	149,051	150,05828	148,04372	1,03
Cytidine	243,0855	244,09278	242,07822	1,15
Adenosine	267,0968	268,10408	266,08952	0,82

Tableau 9: Liste des métabolites de référence sélectionnés pour l'analyse de la banque pour les contrôles qualités de l'extraction

Nous faisons l'hypothèse qu'en moyenne, la teneur de ces métabolites ne sera pas impactée par la mutation de chaque clone. Le suivi de leur abondance (aire du pic chromatographique) pour chaque mutant nous permet d'évaluer sommairement la qualité de l'extraction du puits considéré. Ces métabolites de référence ont été choisis en fonction de l'intensité de leur signal en LC/MS et la stabilité de leur abondance relative chez les différents mutants. Lors de l'analyse d'une plaque, nous considérons les valeurs de *fold* et de Z-score pour ces métabolites (Figure 36).

Plaque P2

A — **(a_1)** fold - 193,0966(+)

	01	02	03	04
A	0,67	0,83	0,00	1,05
B	0,64	0,66	0,13	0,56
C	0,76	1,00	1,05	0,66

(a_2) Z-score - 193,0966(+)

	01	02	03	04
A	-0,33	0,22	-2,53	0,95
B	-0,40	-0,36	-2,12	-0,68
C	-0,01	0,79	0,94	-0,37

B — **(b_1)** fold - Nicotinamide (+)

	01	02	03	04
A	0,85	0,81	0,10	1,09
B	1,04	0,84	0,27	0,86
C	0,82	0,92	0,81	0,85

(b_2) Z-score - Nicotinamide (+)

	01	02	03	04
A	-0,33	0,22	-2,53	0,95
B	-0,40	-0,36	-2,12	-0,68
C	-0,01	0,79	0,94	-0,37

C — **(c_1)** fold - Ornithine (+)

	01	02	03	04
A	0,67	0,83	0,00	1,05
B	0,64	0,66	0,13	0,56
C	0,76	1,00	1,05	0,66

(c_2) Z-score - Ornithine (+)

	01	02	03	04
A	-0,33	0,22	-2,53	0,95
B	-0,40	-0,36	-2,12	-0,68
C	-0,01	0,79	0,94	-0,37

Figure 36:Illustrattion de la méthode d'analyse complète d'une plaque au travers de Business Object. A – Analyse du MO193 sur les puits A1 à C4 de la plaque 2 au travers de son fold (a_1) et de son Z-score (a_2). B – Analyse du nicotinamide (métabolite dit de référence) sur les puits A1 à C4 de la plaque 2 au travers de son fold (b_1) et de son Z-score (b_2). C – Analyse de l'ornithine (métabolite dit de référence) sur les puits A1 à C4 de la plaque 2 au travers de son fold (c_1) et de son Z-score (c_2).

Le puits B3 semble être un puits où le mutant est affecté par sa délétion dans la production du MO193. Cependant les valeurs de fold et de Z-score observés pour les références d'extraction sont également discriminantes. Il n'est donc pas possible de conclure dans le cas de ce mutant sur sa capacité à synthétiser ou non le MO193. Tous les mutants présentant ce genre de caractéristiques seront dans un second temps regroupés selon leur vitesse de croissance et analysés ultérieurement.

3.1.6.3 Contrôle de la qualité du signal LC-MS

Au cours d'un weekend, nous injectons près de 400 matrices complexes. Une telle quantité de matériel est susceptible d'encrasser la source d'ionisation et conduire à une diminution du signal ionique. Pour minimiser cette éventualité, nous avons intégré à la fin du protocole d'extraction des métabolites deux étapes de centrifugation des échantillons afin d'éliminer la majeure partie des débris cellulaires et composants insolubles. La filtration sur 0,22 µm contribue également à éliminer des débris.

Cependant, afin nous assurer de la qualité de l'ionisation dans la source HESI, nous avons également mis en place un suivi du signal ionique. Un échantillon contenant 4 standards commerciaux à 80 µM chacun (thymidine, citrulline, glutamate, cytosine) solubilisés dans la phase mobile (ACN / H_2O pH 9,9 ; 80 : 20), est injecté deux fois au début de chaque séquence d'injections et après que tous les puits d'une ligne de chaque plaque ont été injectés. Nous pouvons alors évaluer l'intensité du signal pour ces quatre molécules tout au long de la séquence (Figure 37).

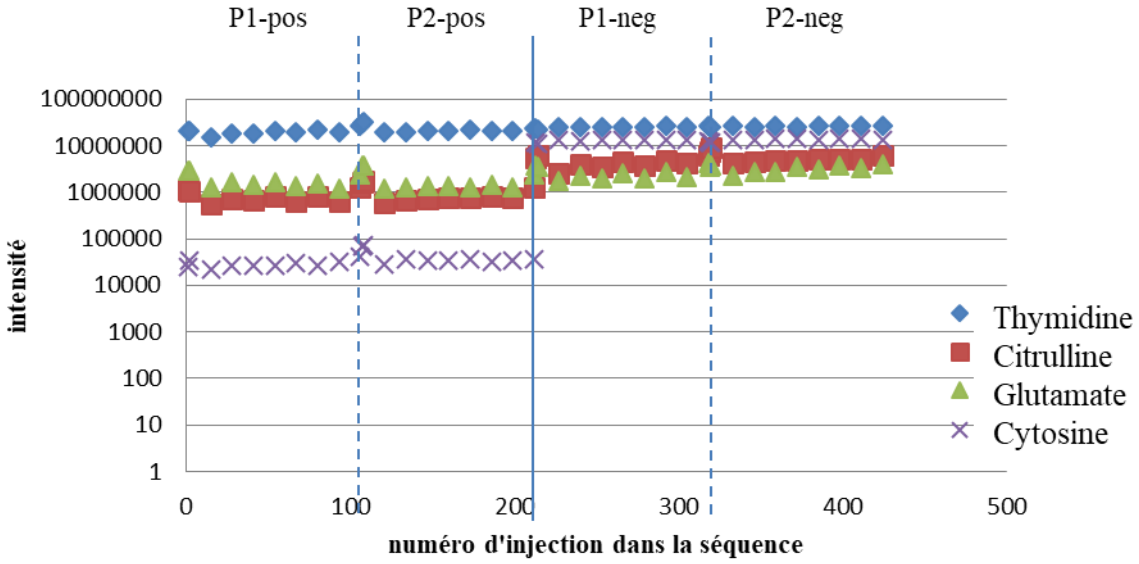


Figure 37: Évaluation de l'intensité de signal en spectrométrie de masse lors d'une séquence d'injection du criblage de la banque. Évolution de l'intensité du signal des 4 standards commerciaux au cours de l'injection d'échantillons de deux plaques de 96 puits (P1 et P2), dans les deux modes d'ionisation. P1-pos : injection des échantillons de la plaque 1 en mode positif. P2-pos : injection des échantillons de la plaque 2 en mode positif. P1-neg : injection des échantillons de la plaque 1 en mode négatif. P2-neg : injection des échantillons de la plaque 2 en mode négatif.

Nous injectons chaque weekend près de 400 matrices complexes. Il serait donc probable que le système, et plus particulièrement la source d'ionisation du spectromètre de masse, s'encrasse au fur et à mesure des injections. Nous nous attendions donc à observer une perte d'intensité de signal pour chacun des

standards commerciaux sélectionnés. Cependant, nous pouvons remarquer sur les résultats présentés en Figure 37 que l'intensité du signal des standards commerciaux injectés, reste du même ordre de grandeur tout au long de la séquence d'injection. Ceci montre bien que l'injection d'autant de matériel en un weekend ne cause pas de perte de signal. Il n'y aura donc aucune normalisation d'intensité à apporter aux résultats lors de leur analyse.

3.1.7 Interprétation et validation des données obtenues

La procédure d'analyse et d'interprétation des résultats du criblage est présentée sur la Figure 38.

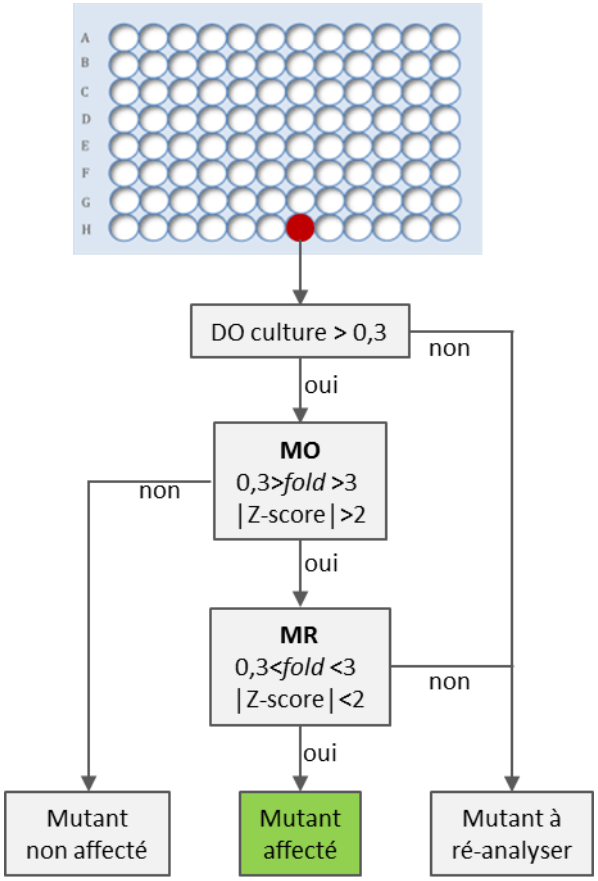


Figure 38 : Procédure d'analyse des résultats du criblage de la banque de mutants de délétion. MO : Métabolite Orphelin ; MR : Métabolites de Référence

Le criblage de la banque de mutants devrait se terminer à l'automne. Une fois les mutants d'intérêt repérés, des réplicas biologiques permettront de valider l'observation de la modification des métabolomes.

L'analyse systématique de la banque permettra *in fine* de dresser la liste de tous les gènes impliqués dans la synthèse de chaque MO. Une illustration des résultats à venir est donnée dans la Figure 39. A

l'issue du criblage, les résultats pourront être représentés sous forme de matrice (Figure 39-A). Dans cet exemple, le métabolite A est absent des métabolomes des mutants Δ1, Δ2 et Δ3 (dans lesquels les gènes 1, 2 et 3 ont été respectivement ôtés). Le métabolite B est quant à lui absent des métabolomes des mutants Δ1, Δ2, Δ3, Δ4 et Δ5. Une autre façon de représenter ces résultats est proposée sur la Figure 39-B. A partir d'un précurseur du métabolisme central les gènes 1, 2 et 3 interviennent dans la synthèse du métabolite A (en synthétisant les précurseurs a_1 et a_2), puis les gènes 4 et 5 dans la formation de B. Dans cet exemple théorique, les métabolites A et B participent à une même voie métabolique. Les résultats du criblage de la collection de mutants de délétion permettront ainsi de commencer à reconstituer ces voies métaboliques nouvelles.

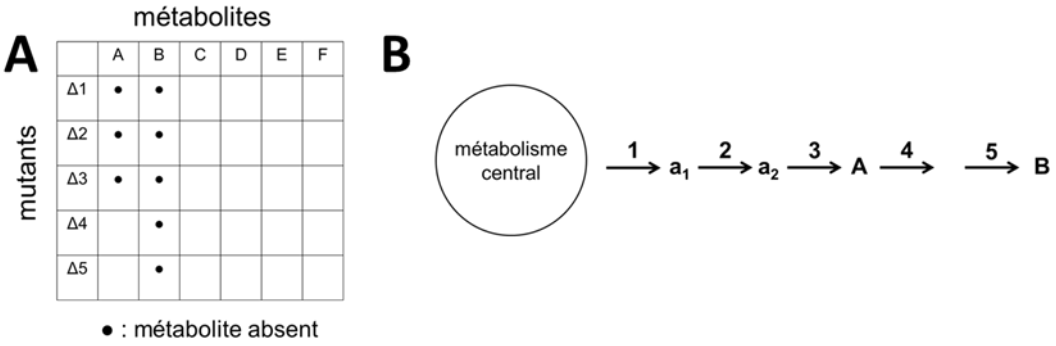


Figure 39: Identification des gènes impliqués dans la synthèse des métabolites orphelins au travers du criblage de la banque de mutants de délétion d'ADP1. A : représentation matricielle. B : représentation graphique

A ce stade, l'ordonnancement des gènes le long de la voie métabolique ne sera pas établi. Pour savoir dans quel ordre interviennent les gènes 1, 2 et 3 pour synthétiser le métabolite A, les protéines codées par ces gènes seront purifiées. Considérons le cas où le gène 3 est impliqué dans la synthèse directe du métabolite A. Le métabolome du mutant Δ3 ne contient pas A mais doit accumuler son précurseur immédiat (métabolite a2, Figure 39 - droite). L'addition dans ce métabolome de l'enzyme purifiée codée par le gène 3 doit conduire à la formation du métabolite A, qui sera détecté par LC/MS (Figure 40). Cette expérience validera le rôle du gène 3 dans la formation de A. De manière plus générale, une approche combinatoire permettra de déterminer quel est entre 1, 2 et 3, le gène impliqué dans la formation directe du composé A. En incubant les métabolomes des mutants Δ1, Δ2 et Δ3 avec chacune des protéines 1, 2 et 3, A ne sera détecté qu'en incubant le métabolome de Δ3 avec la protéine 3.

Une analyse comparative fine du métabolome du mutant Δ3 incubé en présence et en absence de l'enzyme 3, doit nous permettre de détecter la disparition (ou du moins la diminution) du signal (m/z) correspondant au précurseur a2, métabolisé par l'enzyme 3 (Figure 40). Une fois le métabolite a_2 détecté, la même stratégie peut être appliquée (sur les métabolomes Δ1 et Δ2) pour préciser le gène directement impliqué dans sa synthèse et détecter son précurseur (ici, le gène 2 et le métabolite a_1). Nous devrions,

de proche en proche, remonter jusqu'à un métabolite connu du métabolisme central et ainsi préciser une voie métabolique nouvelle.

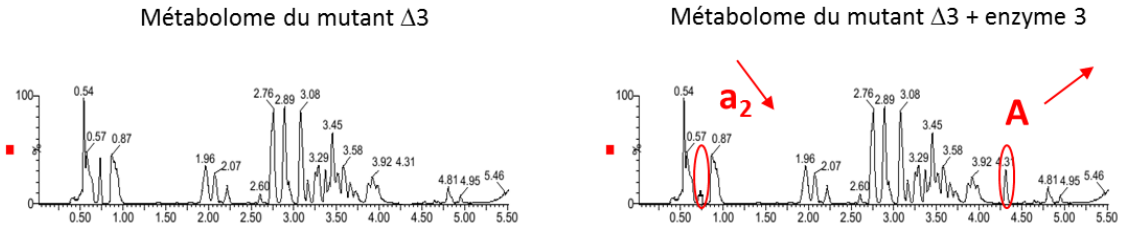


Figure 40: Illustration d'une analyse comparative entre le métabolome du mutant Δ3 et ce même métabolome complémenté par le produit du gène excisé chez ce mutant.

La mise en évidence d'une voie métabolique nécessite non seulement d'identifier les gènes impliqués et de préciser leur fonction, mais également d'établir la structure des MO ainsi synthétisés, ainsi que celle de leurs précurseurs.

En parallèle, un travail a été mené sur l'aspect structural d'un premier MO. Ce travail constitue la seconde partie de ma thèse de doctorat.

3.2 Élucidation structurale d'un métabolite inconnu : le MO211

Parmi les 19 MO suivis lors de du criblage de la banque de mutants, le MO211 avait déjà retenu l'attention du laboratoire, avant même que le criblage ne soit envisagé. Son signal est intense en mode d'ionisation positif (autour de 10^6 u.a.), ce qui est un avantage lorsque l'on souhaite effectuer des expériences MS^n séquentielles et utiliser les rapports isotopiques entre M et M+1. Plusieurs informations ont été obtenues par analyses LC/MS avant mon arrivée au laboratoire.

3.2.1 Analyse du MO211 en HRMS

L'étude de ce métabolite orphelin en spectrométrie de masse a été entamée lors du travail de thèse de Lucille STUANI [61]. Nous présenterons rapidement ici les premiers résultats de ses travaux, dont les données ont été obtenues sur un LTQ Orbitrap XL, puis nous discuterons autour des informations de fragmentation obtenues sur le Velos Pro Orbitrap Elite en mode CID et en mode HCD.

3.2.1.1 Propriétés générales

Le MO211 est nettement retenu sur la colonne chromatographique de 15 cm (Figure 41). Il est détecté à m/z211,107041 et m/z209,09359 dans les modes d'ionisations positif et négatif, respectivement. Sa

masse neutre calculée est de m/z210,09976.

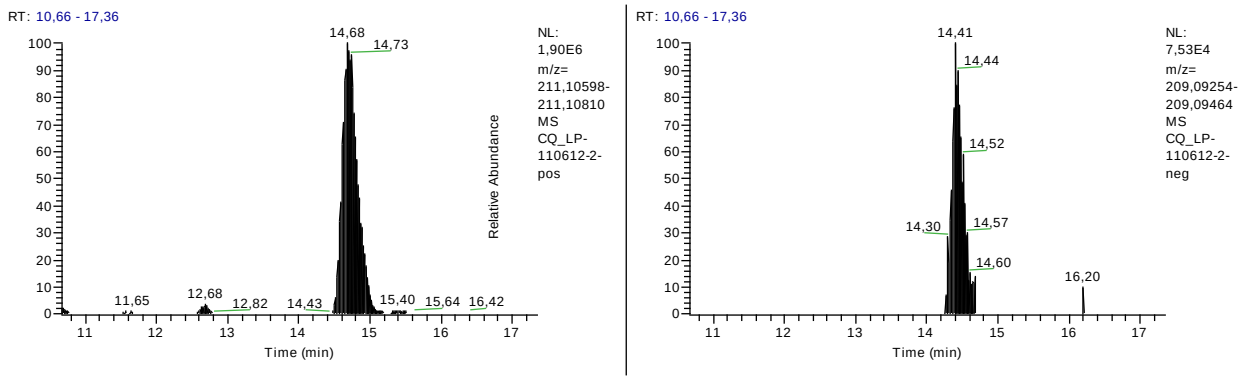


Figure 41: Détection en LC-HRMS du MO211. Le MO211 est détecté sur le gradient de 50 minute sur ZIC p-HILIC (150 x 4,6, 5µm), dans les deux modes d'ionisation sous forme de deux pics : à gauche, mode d'ionisation positif m/z =211,10704 ; à droite mode d'ionisation négatif m/z=209,09359 (5ppm).

Ces informations (temps de rétention élevé, détection dans les deux modes d'ionisation) suggèrent une structure polaire, avec probablement des groupes fonctionnels acides et basiques. Par ailleurs, ce composé présente un spectre d'absorption UV (Figure 42). Celui-ci montre trois maxima : 223, 259 et 302 nm. Les régions d'absorbance comprises entre 270 et 280 nm ainsi qu'entre 210 et 230 nm sont caractéristiques de composés organiques possédant un cycle aromatique tels que l'aniline, l'acide benzoïque, ou encore des dérivés benzéniques tel le phénol. Il est donc très probable que le MO211 comporte un cycle aromatique.

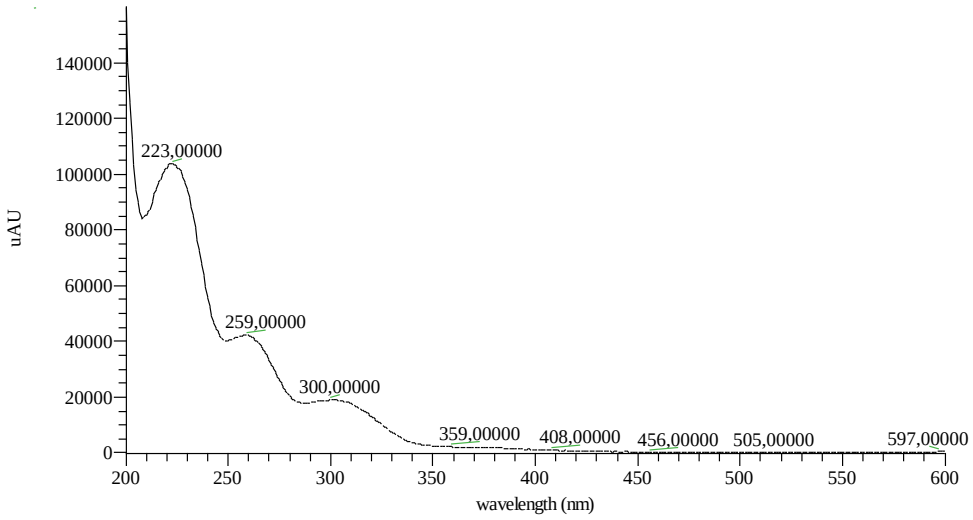


Figure 42: Spectre d'absorbance du MO211

L'interrogation des bases de données publiques en utilisant la masse précise (CheBI et KEGG) ou chargée (METLIN) du MO211 ne donne que deux résultats : l'aprobarbital et la 2,6-dihydroxypseudonicotine, dont les structures sont présentées sur la Figure 43.

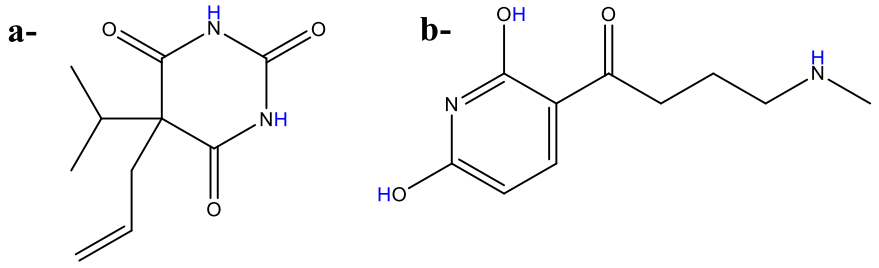


Figure 43: Structures de l'aprobarbital (a) et de la 2,6-dihydroxypseudonicotine (b) proposées pour le métabolite d'intérêt par les bases de données KEGG, Chebi et METLIN sur la base de la masse précise (210,09995 Da). Les atomes d'hydrogènes colorés en bleu sont susceptibles de s'échanger lors des expériences d'échanges isotopiques. Issu de [61]

3.2.1.2 Détermination de la composition élémentaire

A partir des spectres de masse acquis en haute résolution sur le LTQ Orbitrap XL (30 000k), la composition élémentaire du MO211 a été déterminée [61] à partir des données acquises en haute résolution et validée par l'utilisation d'une macro Excel, publique développée en 2007 par Tobias Kind et Oliver Fiehn : les « Sept Règles d'Or [105]. Plusieurs informations ont été utilisées afin de déterminer la composition élémentaire du MO211. Parmi eux, nous pouvons citer :

- la masse neutre du composé : 210,09995 Da
- la précision en masse et l'erreur sur le massif isotopique (10 ppm et 20% respectivement sur le LTQ Orbitrap XL)
- l'abondance de certains isotopes (^{13}C) dans le spectre de masse (M+1 = 8.6% et M+2= 0.08%) en mode d'ionisation positif observée à l'époque sur l' LTQ Orbitrap XL,
- l'absence d'isotopes du chlore dans le spectre de masse (dont l'abondance devrait atteindre les 30% du pic monoisotopique)

La proposition retenue fut donc $C_{10}H_{14}O_3N_2$. Le degré d'insaturation (calculé selon la formule présentée en Figure 44) d'une telle molécule est de 5.

$$Nb\ d'insaturation = \frac{2n_C + 2 - n_H + n_N - n_X}{2}$$

Avec :

- n_C: le nombre d'atomes de carbone
- n_H : le nombre d'hydrogène
- n_N : le nombre d'atomes d'azote
- n_X : le nombre d'atomes d'halogène

Figure 44: Formule de calcul du nombre d'insaturations pour la formule brute supposée du MO211 $C_{10}H_{14}O_3N_2$

3.2.1.3 Échanges H/D

Afin de déterminer le nombre de protons échangeables en solution et de le comparer avec ceux de l'aprobarbital et de la 2,6-dihydroxypseudonicotine, des expériences d'échanges H/D ont été effectuées sur le métabolome issu de la souche sauvage cultivée avec du quinate comme source de carbone.

Pour cela, le métabolome a été analysé sur une méthode analytique utilisant la chromatographie liquide avec une phase mobile dont l'eau du tampon a été remplacée par de l'eau deutérée. Pour des raisons d'économie de solvants deutérés, nous avons choisi d'utiliser une colonne chromatographique de diamètre inférieur à la colonne de 15 cm (2,1 au lieu de 4,6 mm). Cependant, nous souhaitions conserver une bonne séparation chromatographique des métabolites, nous avons donc opté pour une colonne ZIC-*p*HILIC 150 x 2,1 mm², pour laquelle nous avons réadapté le gradient (20 minutes au lieu de 30) et le débit (200 µL/min au lieu de 500 µL/min). Cette nouvelle méthode analytique de 30 minutes est présentée en Matériel et Méthodes (2.3.1).

Le but de cette expérience étant de remplacer la totalité des hydrogènes échangeables par des deutériums, il est important de limiter au maximum la présence d'eau au contact de l'échantillon. Le métabolome extrait et séché est repris dans un mélange 80% ACN et 20% d'eau lourde additionnée de carbonate d'ammonium 10 mM et dont le pH est ajusté à pH 9,9 grâce à NH_4OH). De la même manière, la colonne chromatographique a été rééquilibrée dans la phase mobile deutérée avant l'injection des échantillons.

Afin de déterminer le nombre d'hydrogènes échangeables présent dans la structure du MO211, nous avons calculé tout d'abord les rapports m/z possibles à détecter si le métabolite présente 1, 2, 3 etc. protons échangés par des deutériums. Nous avons alors extrait les EIC présentés en Figure 45 de chacun de ces m/z.

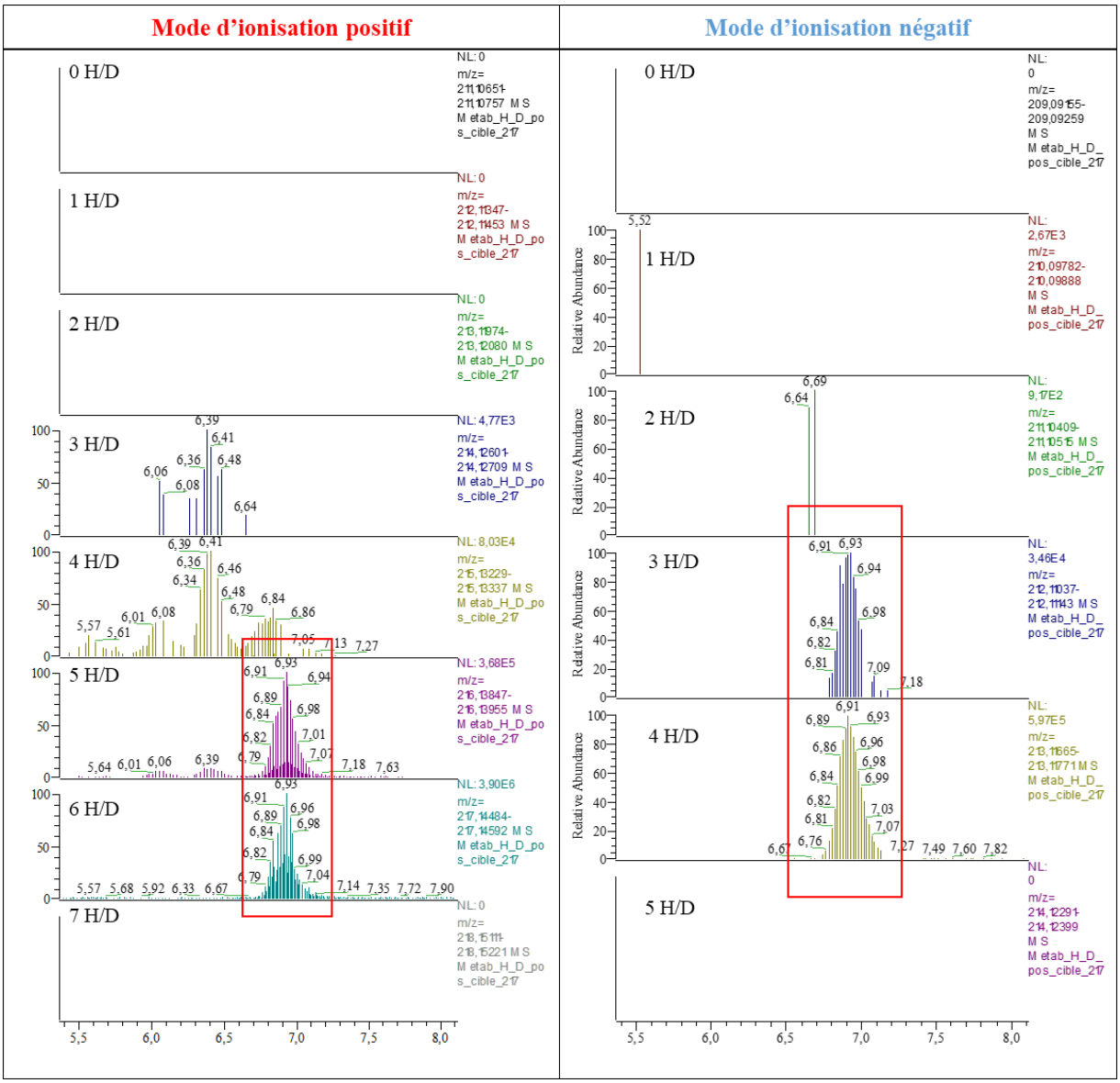


Figure 45: Détermination du nombre d'échanges H/D pour le MO211 dans les métabolomes deutérés dans les deux modes d'ionisation.

Au total, 6 échanges H/D en mode d'ionisation positives, et 4 en mode d'ionisation négatif ont été observés. Le métabolite neutre possède donc 5 protons labiles. Or les deux structures proposées par les bases de données possèdent 2 ou 3 protons labiles. Elles ne peuvent donc pas correspondre au métabolite orphelin. Il est donc été conclu que le MO211 est probablement un métabolite nouveau, non référencé dans les bases de données.

Il a donc été nécessaire de procéder à son élucidation structurale. Pour cela, nous avons employé deux méthodes analytiques orthogonales : la résonance magnétique nucléaire et la LC/MS.

3.2.1.4 Plusieurs signaux pour le MO211

Nous discuterons dans un premier temps, des résultats obtenus lors de la détection en FTMS du MO211

et de ses fragmentations en mode CID et HCD dans le paragraphe suivant.

Lors de l'étude de comparaison des métabolomes cultivés sur quinate aux métabolomes cultivés sur succinate [104], l'analyse des données par XCMS a mis en évidence, l'existence d'une corrélation entre plusieurs signaux m/z. Ces signaux, observés en mode d'ionisation positif, possèdent un profil chromatographique très similaire.

L'analyse du métabolome sur le Velos Pro Orbitrap Elite (couplé à l'HPLC, méthode analytique utilisant la colonne de 15 cm), nous a permis d'observer également en mode d'ionisation positif cette corrélation entre différents signaux dont les profils chromatographiques sont très similaires (Figure 46).

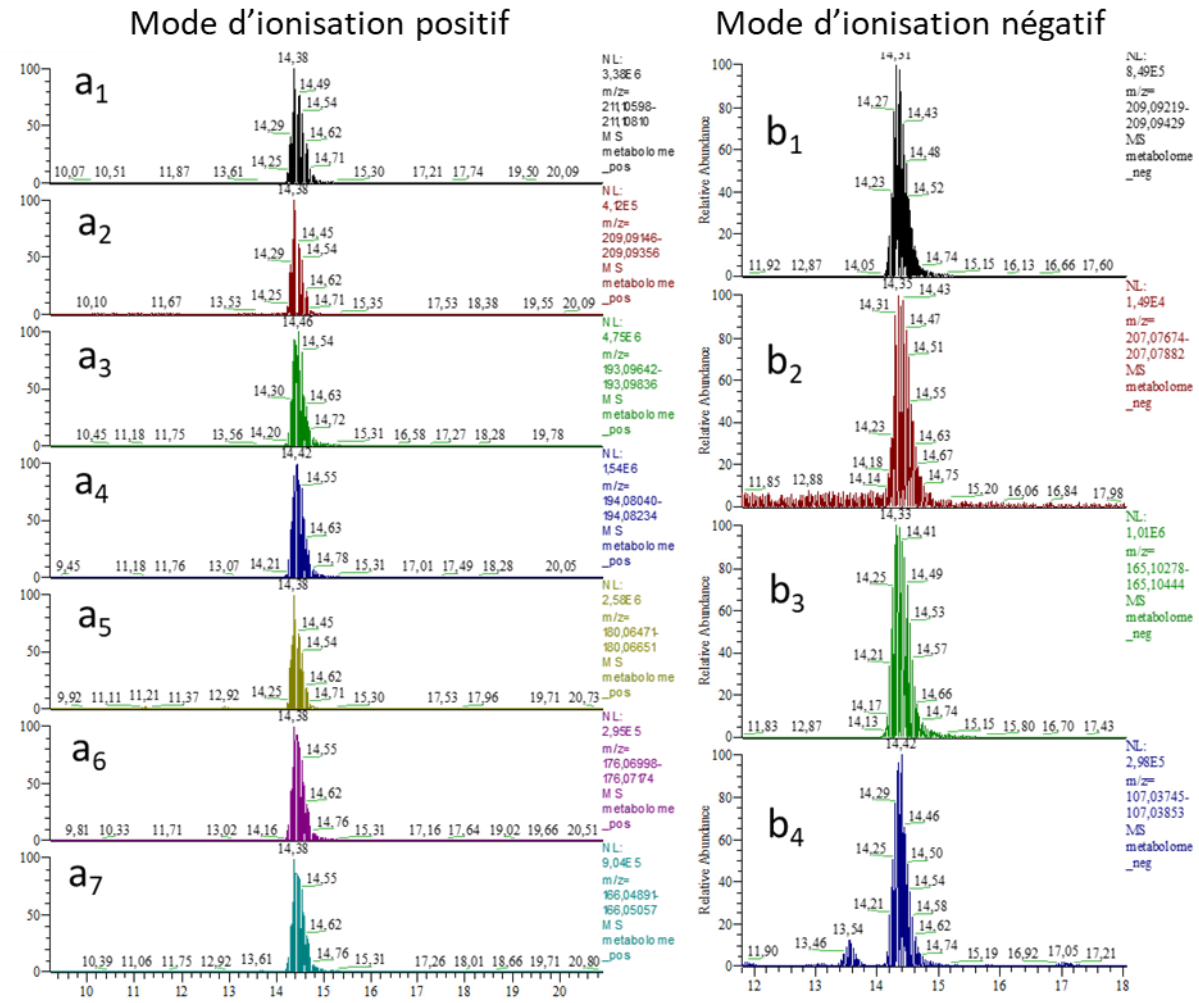


Figure 46: EIC des différents m/z semblant provenir du MO211. EIC extrait de l'analyse LC/MS d'un métabolome sur la colonne de 15 cm (5ppm) pour les valeurs m/z suivantes : (a_1) – m/z211,10704 ; (a_2) – m/z209,09211 ; (a_3) – m/z193,09703 ; (a_4) – m/z194,08104 ; (a_5) – m/z180,06533 ; (a_6) m/z176,07052 ; (a_7) m/z =166,04974 ; (b_1) - m/z209,09324, (b_2) – m/z 207,07778 ; (b_3) – m/z 165,0361 ; (b_4) – m/z107,03799.

Du fait de la meilleure sensibilité de l'Orbitrap Elite, nous avons pu observer en FTMS en mode

d'ionisation positif d'autres signaux, de plus faible intensité que les signaux m/z 180,06533 (m/z 180), m/z 193,09703 (m/z 193) et m/z 209,09211 (m/z 209-ox) corrélés au MO211: les signaux de rapports m/z166,04974 (m/z 166), m/z176,07052 (m/z 176) et m/z194,08104 (m/z 194). En mode d'ionisation négatif, trois signaux de même forme chromatographique que la forme $[M-H]^-$ du MO211 (le m/z209, 09324 ; m/z 209) sont observés : m/z207,0778 (m/z 207) ainsi que deux autres signaux de faible intensité : les ions de m/z165,10361 (m/z 165) et m/z=107,03799 (m/z 107).

L'analyse des données par XCMS révèle la même corrélation entre tous ces signaux, en mode d'ionisation positif (ANNEXE 4).

La similitude des profils chromatographiques suggère que ces ions seraient générés à partir de la fragmentation dans la source de l'ion parent m/z 211,10747 (MO211) en mode d'ionisation positif et m/z 209 en mode d'ionisation négatif.

Dans chacun des cas, la différence de masse entre l'ion parent, MO211 en mode d'ionisation positif ou m/z 209 en mode d'ionisation négatif, et l'un des autres signaux observés dans le même mode d'ionisation, est révélateur de la perte associée lors de la fragmentation dans la source de l'ion parent. Par exemple, la différence de masse observée entre MO211 et m/z 193 est de 18,01044, ce qui est compatible avec une perte d'eau. Nous avons procédé de même pour tous les fragments et les résultats sont présentés dans le Tableau 10.

m/z détectée	intensité relative	Composition élémentaire des ions produits	Perte associée	δ ppm
211,10747	100	$C_{10}H_{15}O_3N_2$		-0,269
209,09211	10,8	$C_{10}H_{13}O_3N_2$	forme oxydée : $-H_2$	0,031
194,08104	3,5	$C_{10}H_{12}O_3N$	$-NH_3$	-0,13
193,09703	6,6	$C_{10}H_{13}O_2N_2$	$-H_2O$	-0,124
180,06533	65,4	$C_9H_{10}O_3N$	$-CH_5N$	-0,21
176,07052	4,8	$C_{10}H_{10}O_2N$	$-C_2H_4$	-0,095
166,04974	18,8	$C_8H_8O_3N$	$-C_2H_7N$	-0,14

m/z détectée	intensité relative	Composition élémentaire des ions produits	Perte associée	δ ppm
209,09324	100,0	$C_{10}H_{13}O_3N_2$		1,17
207,0778	5,2	$C_{10}H_{11}O_3N_2$	Forme oxydée : $-H_2$	1,15
165,10361	3,6	$C_9H_{13}ON_2$	$-CO_2$	1,37
107,03799	3,6	$C_6H_5ON^{\bullet}$	$-CO_2 + C_3H_8N^{\bullet}$	1,42

Tableau 10: Liste récapitulative des signaux observés lors de l'analyse d'un métabolome, au temps de rétention du MO211 (14,6 sur la colonne de 15 cm). En haut, mode d'ionisation positif,

en bas, mode d'ionisation négatif.

Un signal un peu particulier a retenu notre attention : le m/z 209, 09211 (m/z 209-ox), détecté en mode d'ionisation positif : il correspond formellement à la perte de deux hydrogènes par rapport à l'ion parent MO211. Nous supposons qu'il correspond à une oxydation du métabolite dans la source.

Parmi ces signaux, nous avons également pu remarquer la présence des isotopes du carbone pour chacun d'entre eux (Tableau 11). Les ions observés sont donc des ions mono-chargés.

m/z détectée	Isotopes (^{13}C) observés	intensité relative	Composition élémentaire des ions produits	δ ppm
211,10747	212,11078	11,0	$C_9{}^{13}CH_{15}O_3N_2$	-0,314
	213,11199	0,4	$C_8{}^{13}C_2H_{15}O_3N_2$	-2,458
209,09211	210,09541	10,3	$C_9{}^{13}CH_{13}O_3N_2$	-0,034
	211,10028	1,5	$C_8{}^{13}C_2H_{13}O_3N_2$	1,502
194,08104	195,0844	10,3	$C_9{}^{13}CH_{12}O_3N$	0,125
193,09703	194,10034	10,0	$C_9{}^{13}CH_{13}O_2N_2$	-0,169
180,06533	181,06863	10,0	$C_8{}^{13}CH_{10}O_3N$	-0,245
	182,07222	0,2	$C_7{}^{13}C_2H_{10}O_3N$	-0,009
176,07052	177,07391	10,4	$C_9{}^{13}CH_{10}O_2N$	-0,05
166,04974	167,05311	8,5	$C_7{}^{13}CH_8O_3N$	-0,114
	168,05657	0,2	$C_6{}^{13}C_2H_8O_3N$	-0,009

m/z détectée	Isotopes (^{13}C) observés	intensité relative	Composition élémentaire des ions produits	δ ppm
209,09324	210,09654	10,6	$C_9{}^{13}CH_{13}O_3N_2$	1,16
165,10361	166,10706	9,7	$C_8{}^{13}CH_{13}ON_2$	1,46
107,03799	108,04136	5,2	$C_5{}^{13}CH_5ON$	1,44

Tableau 11: Liste récapitulative des isotopes observés lors de l'analyse d'un métabolome au temps de rétention du MO211 (14,6 min. sur la colonne de 15 cm). En haut, mode d'ionisation positif, en bas, mode d'ionisation négatif

3.2.1.5 Spectres de fragmentation

L'Orbitrap Elite nous permet d'effectuer des fragmentations en mode HCD. Ce mode de fragmentation est particulièrement intéressant ici car les fragments de petites masses y sont mieux observés (pas de « low mass cut-off »). Cependant, afin d'obtenir la filiation entre les différents fragments obtenus, des expériences MS^n peuvent s'avérer nécessaire. Des expériences de fragmentation MS^3 ont été réalisées en CID avec détection en FTMS.

Mode d'ionisation positif

Les spectres de fragmentation MS^2 du MO211 obtenus dans le mode d'ionisation positif (Tableau 12) présentent deux ions majoritaires : m/z 193,0967 (m/z 193) et m/z 194,0809 (m/z 194). Ils correspondent à une perte d'eau et d'ammoniac, respectivement. La perte d'ammoniac suggère la présence d'une amine primaire dans la molécule. En mode d'ionisation positif, la perte d'eau peut être reliée à la présence d'un groupement hydroxyle aliphatique ou aromatique ou encore d'un acide carboxylique.

ion précurseur	m/z	Intensité relative	Composition élémentaire des ions produits	δ ppm	perte associée
211,10822	194,08155	58,5	$C_{10}H_{12}O_3N$	0,38	- NH_3
	193,09764	100	$C_{10}H_{13}O_2N_2$	0,49	- H_2O
	176,07102	10,3	$C_{10}H_{10}O_2N$	0,42	- H_2O - NH_3
	166,05024	30,7	$C_8H_8O_3N$	0,37	- NH_3 - C_2H_4
	150,09167	2,4	$C_9H_{12}O$ N	0,33	- NH_3 - CO_2
	122,06023	1,0	C_7H_8ON	0,19	- NH_3 - C_2H_4 - CO_2

Tableau 12: Produits de fragmentations résultant des expériences HCD-MS^2 sur le métabolite d'intérêt, à une énergie de collision normalisée à 28%

Deux autres ions fragments moins abondants sont observés : m/z 166 et m/z 176. Ceux-ci sont liés à la perte de [C_2H_7N] et [NH_3+H_2O], respectivement. Le fragment neutre [C_2H_7N] correspond à l'alkyl amine insaturé. Le même ion à m/z 166 est observé dans le spectre MS^3 à partir de l'ion à m/z 194, généré par la perte d'ammoniac de l'ion parent (Tableau 13). Il est lié à la perte consécutive d'ammoniac puis d'un fragment C_2H_4. Le fragment neutre [C_2H_7N] représente donc probablement l'éthylamine plutôt que la diméthylamine.

ion précurseur	m/z	Intensité relative	Composition élémentaire des ions produits	δ ppm	perte associée
194,0815	176,07096	4,0	$C_{10}H_{10}O_2N$	0,355	- H_2O
	166,05014	100,0	$C_8H_8O_3N$	0,27	- C_2H_4
	150,09162	5,4	$C_9H_{12}O$ N	0,279	- CO_2
	122,06019	1,1	C_7H_8ON	0,15	-C_2H_4 - CO_2

Tableau 13: Produits de fragmentations résultant des expériences CID-MS^3 sur le métabolite d'intérêt de valeur m/z 194, à une énergie de collision normalisée à 23% pour les deux étapes d'activation

Enfin, dans le spectre MS^2 de l'ion parent, il a également été possible d'observer la présence de deux ions de très faible intensité : m/z 150 et m/z 122. Ces ions sont également observés lors de la

fragmentation en MS^3 de m/z 194. Ils représentent respectivement, la perte de CO_2 et de $CO_2 + C_2H_4$ à partir du m/z 194.

Nous pouvons remarquer que l'ion m/z 180, n'est observé dans aucun de nos spectres de fragmentations du MO211, qu'il s'agisse de fragmentation CID ou HCD. Cependant, il est particulièrement abondant dans les spectres de fragmentation de la forme oxydée du MO211, le m/z 209-ox (Figure 47).

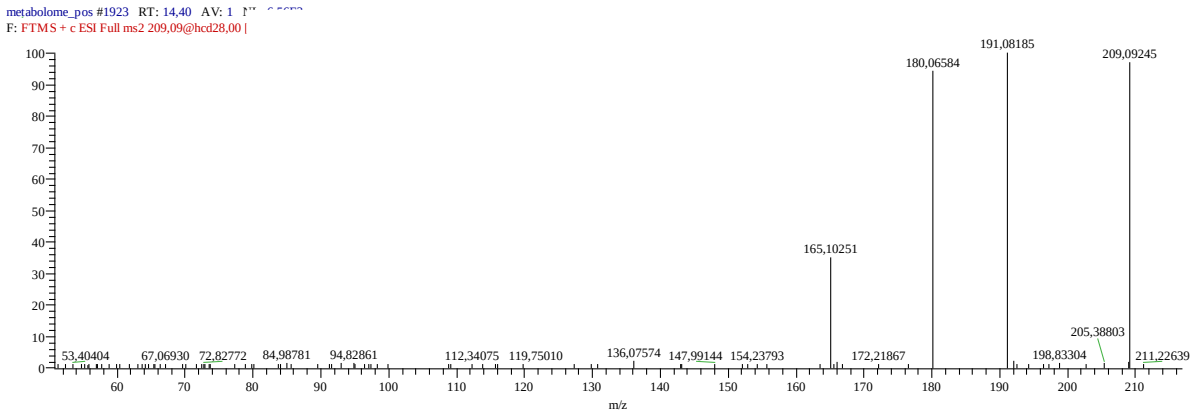


Figure 47: Spectre de fragmentation MS^2 de l'ion de valeur m/z209,09245, forme oxydée du MO211, en mode d'ionisation positif. (HCD, énergie d'activation normalisée : 28%, détection haute résolution)

De plus, nous avons pu observer en FTMS lors des expériences d'échanges H/D, que le m/z 209-ox possédait 4 hydrogènes échangeables en mode d'ionisation positif (Figure 48). Le m/z 180 présente deux formes possibles de l'ion marqué au deutérium : il peut échanger 3 ou 4 protons avec des deutériums. Cependant, comme présenté dans le spectre de masse associé, l'abondance de la forme m/z183,08435 (3 H/D) est plus importante que l'ion m/z 184,09063. Ainsi, la forme neutre du m/z 180 possède 2 hydrogènes échangeables. Il est donc plus probable que l'ion m/z 180 provienne de la fragmentation dans la source du m/z 209, en perdant CH_2NH (un hydrogène échangeable), plutôt que de MO211, en perdant CH_3NH_2 (deux hydrogènes échangeables).

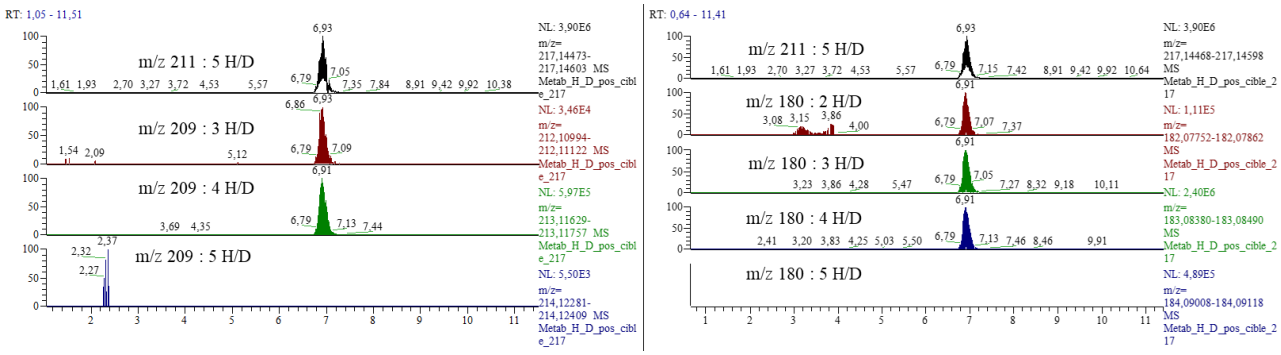


Figure 48: Détermination du nombre d'échanges H/D pour le m/z 209-ox à gauche et m/z 180 à droite dans les métabolomes deutérés en mode d'ionisation positif. Les EIC ont été générés à partir des données RAW (5ppm) et les rapports m/z calculés à partir des compositions élémentaires $C_{10}H_{13-x}D_xO_3N_2$ (pour m/z 209-ox) et $C_9H_{10-x}D_xO_3N$ (pour m/z 180).

Mode d'ionisation négatif

Les spectres de fragmentation MS^2 de l'ion parent m/z 209, présentent un ion fragment particulièrement abondant : m/z 165, correspondant formellement à une perte de CO_2 à partir de l'ion parent (Tableau 14). Cette perte majoritaire de CO_2 en mode négatif est caractéristique de la présence d'un groupement carboxylique dans la molécule.

ion précurseur	m/z	Intensité relative	Composition élémentaire des ions produits	δ ppm	perte associée
209,09319	165,10335	100,0	$C_9H_{13}ON_2$	1,46	- CO_2
	151,02798	17,2	$C_7H_5O_3N^{\bullet}$	1,58	- $C_3H_8N^{\bullet}$
	107,03784	51,3	$C_6H_5ON^{\bullet}$	1,5	- CO_2 - $C_3H_8N^{\bullet}$

Tableau 14: Produits de fragmentations résultant des expériences HCD-MS^2 sur le métabolite d'intérêt, à une énergie de collision normalisée à 75%.

Deux autres fragments, moins abondants, sont observés à m/z 107 et m/z 151. Ils ont comme composition élémentaire $C_6H_5ON^{\bullet}$ et $C_7H_5O_3N^{\bullet}$, respectivement. Ces deux ions ont des valeurs de m/z impaires et contiennent un seul atome d'azote. Selon la règle de l'azote, ils contiennent un nombre impair d'électrons et représentent donc les ions radicalaires $[C_6H_5ON]^{-\bullet}$ et $[C_7H_5O_3N]^{-\bullet}$, respectivement. L'ion m/z 151 correspond à la perte du radical $C_3H_8N^{\bullet}$ de l'ion parent m/z 209. L'ion m/z 107, également observé dans le spectre MS^3 de l'ion m/z 165 (Tableau 15), est lié à la perte consécutive du même radical $C_3H_8N^{\bullet}$ à partir de l'ion fragment m/z 165.

ion précurseur	m/z	Intensité relative	Composition élémentaire des ions produits	δ ppm	perte associée
165,10338	134,06133	7,4	C_8H_8ON	1,29	CH_3NH_2
	107,03793	100	$C_7H_6O_3N$	1,36	- $C_3H_8N^{\cdot}$

Tableau 15: Produits de fragmentations résultant des expériences CID-MS^3 sur le métabolite d'intérêt de valeur m/z 165, à une énergie de collision normalisée à 27% pour m/z 209,09319 et 30% pour m/z 165,10338

La combinaison des données de fragmentation dans les deux modes d'ionisation et des informations fournies par l'absorbance du composé en UV, nous permettent de procéder à plusieurs conclusions :

- la perte massive d'ammoniac en mode d'ionisation positif suggère la présence d'une amine primaire dans la molécule. Ceci explique la présence de 2 hydrogènes échangeables dans la molécule.

- la perte de CO_2 de l'ion parent en mode d'ionisation négatif est caractéristique de la présence d'un groupement carboxylique sur la molécule. Ce groupement est porteur d'un hydrogène échangeable dans la molécule.

- l'existence du fragment m/z 107 en mode d'ionisation négatif, suggère la présence d'un cycle aromatique dans la molécule, information confirmée par les longueurs d'ondes d'absorbance maximales du MO211 : 220 et 300 nm.

- la présence du fragment m/z 193 suggère la présence d'un groupement hydroxyle aliphatique ou aromatique, ou encore d'un acide carboxylique. Cependant, il ne reste qu'un atome d'oxygène à attribuer au sein de la molécule (les deux autres étant attribués au groupement carboxylique), il n'est donc pas possible d'avoir un autre groupement carboxylique dans la molécule. Cette perte d'eau est donc significative de la présence d'un groupement hydroxyle dans la structure du MO211. Dans le cas où ce groupement chimique serait placé sur la chaine aliphatique, nous observerions une seconde perte d'eau dans les spectres de fragmentation. Or, nous n'en observons qu'une : le groupement hydroxyle est donc très probablement fixé sur le cycle aromatique de la molécule. Ce groupement est porteur d'un proton labile.

En conclusion, en se basant sur le nombre total d'hydrogène échangeable (cinq sur la molécule neutre), la présence d'un groupement carboxylique (un hydrogène échangeable), d'un phénol probable (un hydrogène échangeable), d'un groupement amine primaire (deux hydrogènes échangeables) et la composition élémentaire du MO211, nous pouvons supposer que la molécule contient également une amine secondaire (un hydrogène échangeable).

3.2.2 Analyse du MO211 par RMN

Ces informations structurales sont importantes mais restent parcellaires et ne permettent pas de proposer d'identité. Nous avons donc décidé de procéder à l'élucidation structurale du composé par RMN.

3.2.2.1 Production, extraction et purification du MO211

Avant de procéder à l'analyse structurale du MO211 par RMN, il est nécessaire de disposer de plusieurs centaines de µg de la molécule pure. Dans ce but, nous avons établi un protocole d'extraction de ces métabolites à une échelle compatible avec la quantité de matériel purifié requise.

Jusqu'ici, l'extraction des métabolites se faisait à une échelle analytique, à partir de cultures de 5 mL. Des essais d'extraction à plus grande échelle ont été tentés antérieurement au laboratoire, lors d'un stage de M2 de Cécile Robert - Ansart, en 2015. Différents volumes de cultures (15, 50 et 500 mL) avaient été mis à incuber au même moment, en erlenmeyer, à partir de la même pré-culture, dans les mêmes conditions d'incubation (30°C et 150 rpm). Un volume de 5 mL pour chaque culture avait alors été prélevé afin d'appliquer le même protocole d'extraction. Les données obtenues après analyse LC-MS sont présentées dans le Tableau 16.

Volume de culture	**15 mL**	**50 mL**	**500 mL**
Proline	$3,35 \pm 0,93 .10^{7}$	$2,88 \pm 0,13 .10^{7}$	$1,62 \pm 0,25 .10^{7}$
Leucine	$7,33 \pm 2,20 .10^{6}$	$8,59 \pm 0,16 .10^{6}$	$5,09 \pm 0,48 .10^{6}$
MO211	$4,90 \pm 4,06 .10^{7}$	$7,71 \pm 6,59 .10^{6}$	$9,31 \pm 4,45 .10^{5}$

Tableau 16: Comparaison des abondances relatives de différents ions selon le volume de culture pour un même volume d'extraction

L'augmentation du volume de culture induit une baisse en concentration de métabolites, notamment du MO211 qui diminue d'un facteur 50 entre une culture de 15 mL et une culture de 50 mL. L'hypothèse alors privilégiée était que le métabolisme des cellules est perturbé lorsque le volume de culture augmente. Il a donc été décidé pour la culture des souches de multiplier les petits volumes de culture cellulaire.

Ainsi, un protocole de production et d'extraction de métabolome a été adapté à partir du protocole de production et d'extraction des métabolomes utilisé pour le criblage de la banque de mutants. Simplement, nous avons multiplié les cultures de 5 ml à partir desquelles l'extraction du MO211 est satisfaisante.

Nous avons effectué nos cultures dans 4 plaques de 24 puits, celles-là mêmes qui ont été utilisées pour l'extraction des métabolites à l'échelle analytique. Ce qui correspond à un volume total de culture proche de 500 mL. Dans chaque puits la culture est ensemencée avec la souche sauvage. La suite du protocole se déroule de la même manière que pour la préparation d'échantillons lors du criblage de la banque de mutants.

Afin de s'assurer une quantité de métabolome suffisante pour obtenir la quantité de métabolite nécessaire à l'analyse en RMN, nous avons répété ce protocole de culture et d'extraction trois fois, trois jours de suite. Au final, nous avons obtenu une quantité de métabolome équivalente à 1,5 L de volume de culture extrait.

Purification

Nous disposons au laboratoire d'une colonne dont le format et la phase stationnaire sont compatibles avec une purification à l'échelle semi-préparative : une colonne ZIC-HILIC de dimension 250 x 10 mm^2, 5 µm. Cette colonne est constituée de la même phase stationnaire que la ZIC-*p*HILIC, mais cette dernière possède des particules poreuses polymériques additionnelles qui lui confèrent une résistance mécanique supplémentaire et permettent d'étendre la gamme de pH à 2-10 contre 2-8 seulement pour une colonne ZIC-HILIC classique.

Dans un premier temps, avant d'utiliser la colonne ZIC-HILIC « préparative » comme première étape de purification, nous avons validé que le MO211 est bien retenu sur sa version « analytique » de dimension 150 x 4,6 mm^2 ; 5 µm. En adaptant la phase mobile avec un pH compatible avec cette colonne ZIC-HILIC (acétate d'ammonium 10 mM, pH 7 au lieu de carbonate d'ammonium 10 mM + NH_4OH , pH 10 pour la ZIC-*p*HILIC) et en utilisant le même gradient (2.3.1), le temps de rétention du MO211 est peu modifié (15 minutes contre 14,6 minutes). La forme chromatographique du pic reste également bien symétrique.

Nous avons ensuite adapté le gradient utilisé avec la colonne ZIC-HILIC analytique aux dimensions de la colonne semi-préparative en utilisant les convertisseurs de méthodes analytiques fournis par Thermo et Sigma Aldrich. La méthode « semi-préparative » utilisée est présentée au paragraphe 2.3.2. Le métabolome extrait et sec a été repris dans un total de 7 mL de phase mobile (80 : 20 ACN / pH6.8). Au total, 33 injections du filtrat de 200 µL ont été effectuées.

Nous avons suivi l'évolution de la purification grâce au détecteur UV-visible double faisceau de l'HPLC préparative aux longueurs d'onde d'absorbance maximales du MO211, soit de 224 nm et 307 nm (figure 49 (a)). Bien que le signal UV soit suffisant pour détecter le pic d'élution du MO211 (à 19,5 minutes) lors d'une injection, nous avons également suivi la composition des fractions par LC/MS avec une colonne Syncronis HILIC (50 x 2,1 mm^2 ; 1,7 µm). En effet, le MO211 avait tendance à « traîner » sur

les fractions suivant le pic (Figure 49 (b)). Comme nous souhaitions conserver un maximum de matériel pour les analyses suivantes, nous avons choisi de récupérer le maximum de fractions possible contenant le MO211. Pour chacune des injections, les fractions entourant le pic chromatographique du MO211 ont été systématiquement analysées en LC/MS.

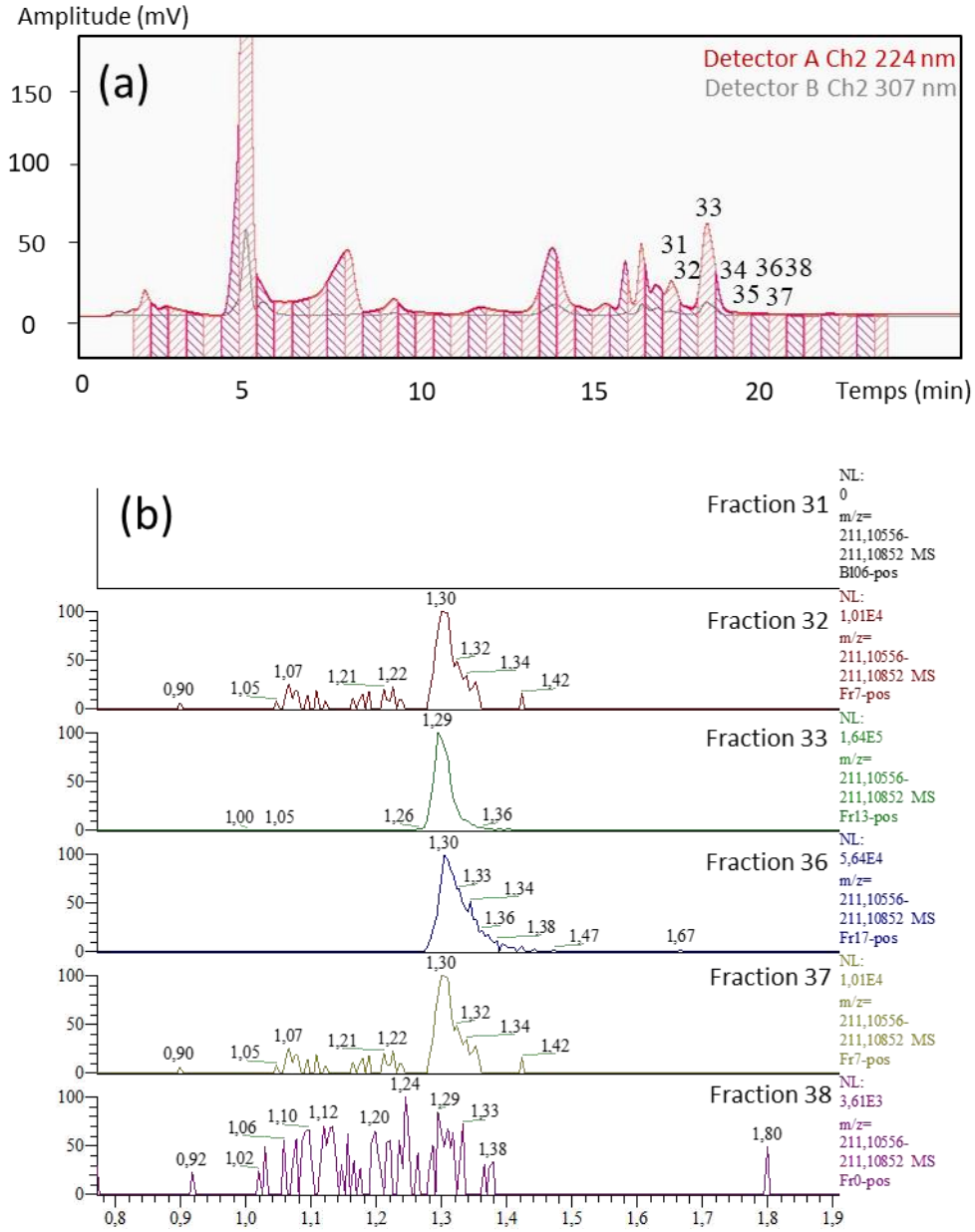


Figure 49 : Première étape de purification du MO211 par HPLC préparative. (a) Suivi en UV (224 nm) d'une injection de la purification du métabolome sur HPLC préparative (b) EIC du MO211 de l'analyse en LC/MS (gradient sur Syncronis HILIC 50 x 2,1, 1,7µm, paragraphe 2.3.1) sur chromatographie liquide de différentes fractions issues de cette injection pour la purification en HPLC préparative du métabolome.

Dans le cas de l'injection présentée en Figure 49, les fractions 32 à 37 ont alors été rassemblées dans un ballon, séchées à l'évaporateur rotatif (élimination de l'acétonitrile), puis lyophilisées pendant la nuit

suivant la purification (élimination de l'eau).

L'étape de purification de 1,5 L de culture bactérienne extraite a permis au total d'obtenir, après lyophilisation, 1,1 mg d'un solide jaune-marron pâle.

L'analyse de cet échantillon (gradient de 50 min sur ZIC pHILIC 150 x 4,6 mm², 5µm, présenté en paragraphe 2.3.1) a révélé la présence de plusieurs autres métabolites, contaminant le MO211 (Figure 51-b)

Nous avons donc développé à l'échelle analytique sur la colonne ZIC-HILIC (150 x 4,6mm², 5µm) une méthode isocratique permettant la séparation des contaminants du MO211. Un pourcentage de 72% d'acétonitrile et de 28% de tampon pH 6,8 (10 mM acétate d'ammonium) s'est avéré le plus efficace pour éliminer les autres métabolites. Cette méthode a été transposée à la colonne ZIC-HILIC (250 x 10mm², 5µm) préparative, grâce aux convertisseurs fournis par Thermo et Sigma Aldrich. La méthode « préparative » utilisée pour cette deuxième étape de purification est présentée au paragraphe 2.3.2.

Nous avons procédé à la purification du solide jaune-marron pâle. Au total, 15 injections du filtrat de 200 µL ont été effectuées. De même que précédemment, nous avons suivi l'évolution de la purification grâce au détecteur UV de l'HPLC préparative aux longueurs d'onde d'absorbance maximales du MO211, soit de 224 nm et 307 nm (Figure 50). Pour les même raisons que pour l'étape précédente de purification, nous avons criblé les fractions en LC/MS (gradient sur Syncronis HILIC 50 x 2,1, 1,7µm, paragraphe 2.3.1).

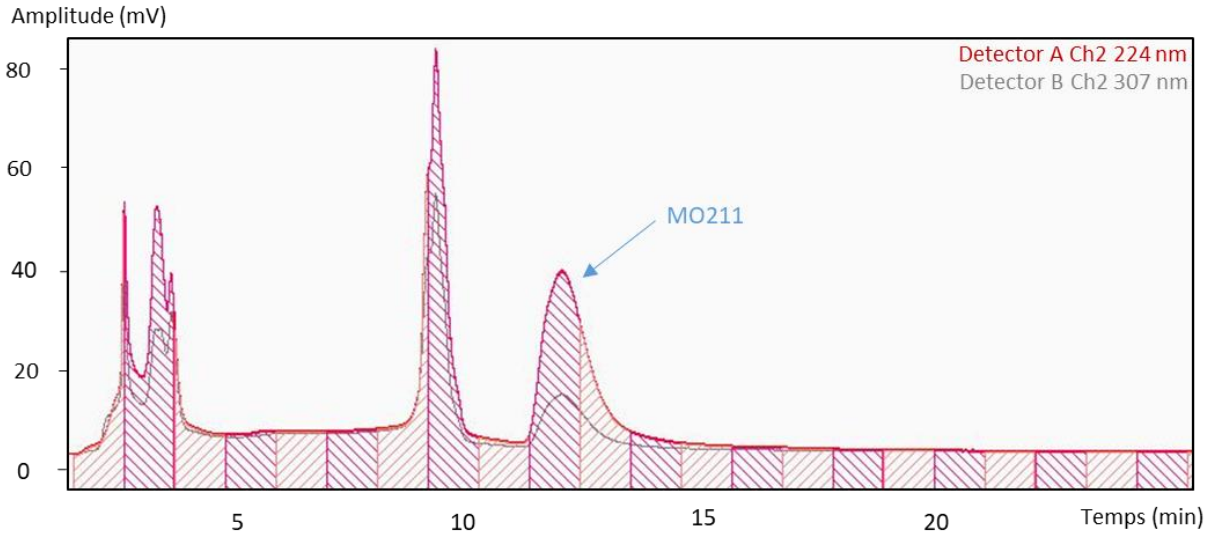


Figure 50: Deuxième étape de la purification du MO211 par HPLC préparative. Suivi en UV (224 et 304 nm) de la purification du métabolome sur HPLC préparative

Deux types de fractions ont été distingués : celles contenant le MO211, sur le pic chromatographique

visible à l'UV, et celles contenant le MO211 en plus faible quantité, présentes après le pic chromatographique. Chacun de ces deux types de fraction ont été rassemblés en deux échantillons différents qui ont été ensuite séchés à l'évaporateur rotatif puis lyophilisés.

Environ 100 µg d'un solide blanc ont ainsi été obtenus.

L'échantillon a alors été analysé par la méthode analytique de 50 minutes (colonne ZIC-*p*HILIC 150 x 4,6 mm², 5µm). Cette analyse a montré la présence d'un unique composé, le MO211 dans l'échantillon (Figure 51 (c)).

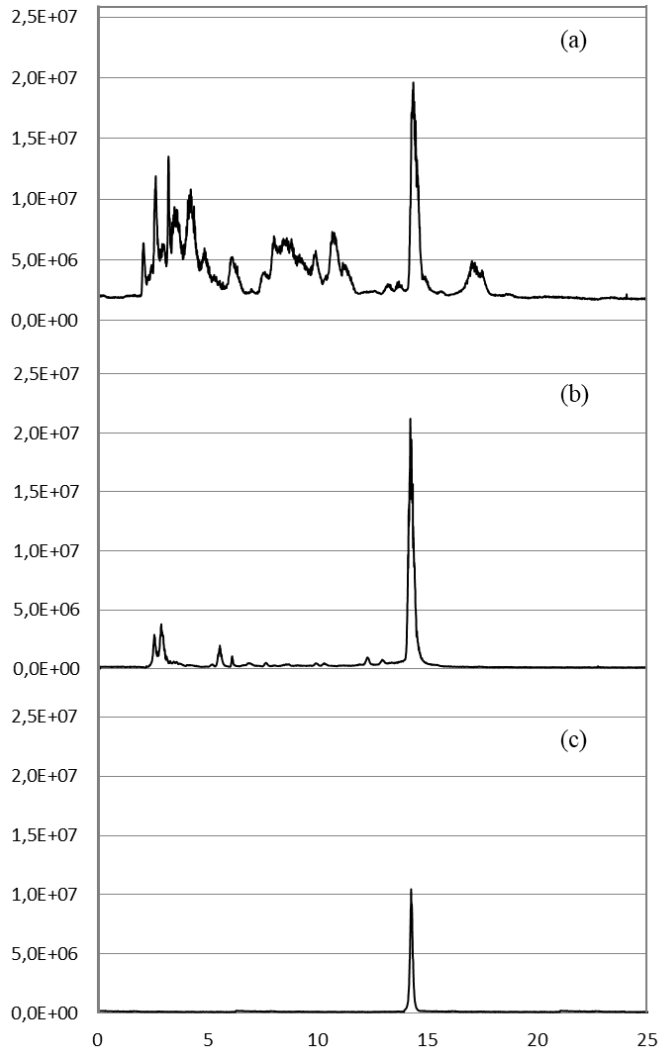


Figure 51: Évaluation de la qualité de purification du MO211. TIC issus de l'analyse en LC/MS (gradient de 50 minutes sur ZIC pHILIC 150 x 4,6 mm², 5µm, présenté paragraphe 2.3.1) de trois échantillons : (a) le métabolome avant purification, (b) le produit obtenu à l'issue de la première étape de purification, (c) échantillon pur obtenu à l'issue de la deuxième étape de purification, fractions de cœur.

3.2.2.2 Interprétation des spectres RMN

RMN ^{1}H et COSY

Une collaboration avec l'Institut de Chimie des Substances Naturelles du CNRS de Gif-sur-Yvette (Jean-François Gallard) nous a donné l'accès à un spectromètre RMN 600 MHz possédant une sonde cryogénique améliorant la sensibilité des acquisitions. La quantité du MO211 (environ 100 µg) solubilisé dans un mélange D_2O/H_2O (10 : 90) a ainsi pu être suffisante pour l'étude complète par RMN en réalisant des expériences mono- et bidimensionnelles (^{1}H, ^{13}C, ^{1}H-^{1}H-COSY, ^{1}H-^{13}C-HSQC, ^{1}H-^{13}C-HMBC et ^{1}H-^{15}N-HMBC).

RMN ^{1}H et COSY :

Sur le spectre ^{1}H du MO211 (Annexe 8), nous pouvons distinguer clairement deux familles de signaux, l'une dans la zone aromatique pour les protons H_1, H_2, et H_3 (Figure 52-a) et l'autre dans la zone aliphatique pour les protons H_4, H_5, et H_6 (Figure 52-b).

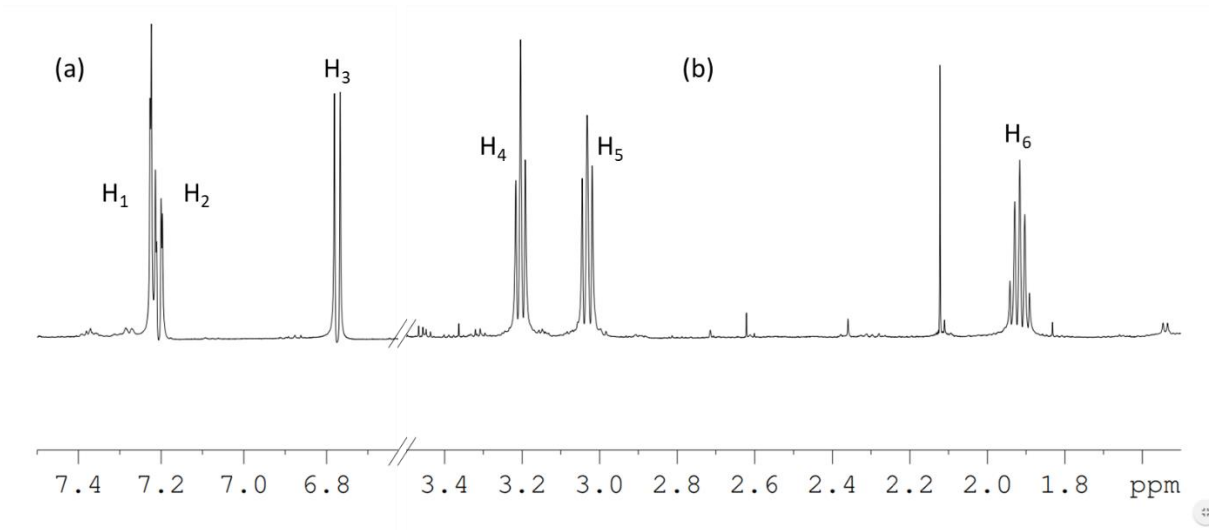


Figure 52: Extraits du spectre RMN ^{1}H du MO211 purifié, 600 MHz, solvant H_2O/D_2O (90:10)

La multiplicité des protons H_4, H_5, et H_6 ainsi que les corrélations observées entre eux sur le spectre COSY (Annexe 9) indiquent que le proton H_6 couple à la fois avec le proton H_4 et le proton H_5. Les valeurs des constantes de couplage (Tableau 17) permettent de placer les protons H_2 et H_3 en ortho l'un de l'autre et les protons H_1 et H_2 en méta l'un de l'autre sur le cycle aromatique.

Proton	Déplacement chimique (ppm)	Intégration	Multiplicité	Constante de couplage J(Hz)	Couplage scalaire
H1	7,22	1	d	1,9	J(H1, H2)
H2	7,20	1	dd	1,9 8,0	J(H2, H1) J(H2, H3)
H3	6,77	1	d	8,0	J(H3, H2)

Tableau 17 : Interprétation des données obtenues par RMN ^1H et COSY des protons aromatiques

Le proton H_3 est plus blindé que les protons H_2 et H_1, cela indique qu'il se situe en ortho ou para d'un groupement mésomère donneur tel que -OH ou -NH_2, et H_1 et H_2 sont en ortho ou para d'un groupement mésomère attracteur tel que -COOH.

Ainsi, nous pouvons proposer une première structure générique pour la partie aromatique présentée en Figure 53.

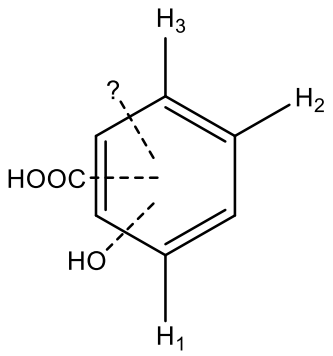


Figure 53: Position relatives des protons sur le cycle aromatique

Les déplacements chimiques des protons H_4, H_5 et H_6 indiquent qu'il s'agit de protons aliphatiques. La multiplicité des signaux (Tableau 18) ainsi que les corrélations observées sur le spectre COSY indiquent un motif –CH_2-CH_2-CH_2-.

Proton	Déplacement chimique (ppm)	Intégration	Multiplicité	Constante de couplage J(Hz)	Couplage scalaire
H4	3,2	2	t	6,5	J(H4, H6)
H5	3,0	2	t	7,0	J(H5, H6)
H6	1,09	2	quint	6,6	J(H6, H4) J(H6, H5)

Tableau 18: Interprétation des données obtenues par RMN ^1H et COSY des protons aliphatiques

Les valeurs de couplage proches de H_4 et H_5 permettent d'expliquer la présence d'un quintuplet pour H_6 au lieu du triplet de triplet attendu. Aucune corrélation entre protons aromatiques et aliphatiques n'est observable sur le spectre COSY ce qui complique l'élucidation structurale. Les déplacements chimiques des protons H_4 et H_5, nettement plus déblindés que H_6 démontrent la connexion à des groupements inductifs électroattracteurs tels - NH_2 ou - OH.

D'après les données de spectrométrie de masse, la formule brute neutre de la molécule est $C_{10}H_{14}O_3N_2$. Seuls 9 des 14 protons sont donc visibles sur les spectres ^{1}H et COSY de l'échantillon dissous dans un mélange H_2O/D_2O (90 : 10). Plusieurs autres essais ont été conduits en modifiant le solvant (utilisation de DMF-d7 et de DMSO-d_6), en acidifiant le milieu (ajout d'acide trifluoroacétique) et en abaissant la température, mais sans succès. Les 5 protons non détectés représentent des protons labiles précédemment mis en évidence lors des expériences d'échange H/D effectués en LC-MS (3.2.1.3).

RMN ^{13}C et RMN ^{1}H-^{13}C HSQC

Neuf signaux sont observables sur le spectre RMN ^{13}C (Figure 54): trois signaux correspondant à des atomes de carbone sp3 (26,8 ppm, 38,4 ppm et 41,6 ppm), 5 signaux à des atomes de carbone sp2 aromatiques et/ou vinyliques et un atome de carbone à 176,0 ppm appartenant à la fonction acide carboxylique mise en évidence lors de l'étude en LC-MS^n. Ceci ne correspond pas au nombre d'atomes de carbone déterminés précédemment dans la formule brute du MO211.

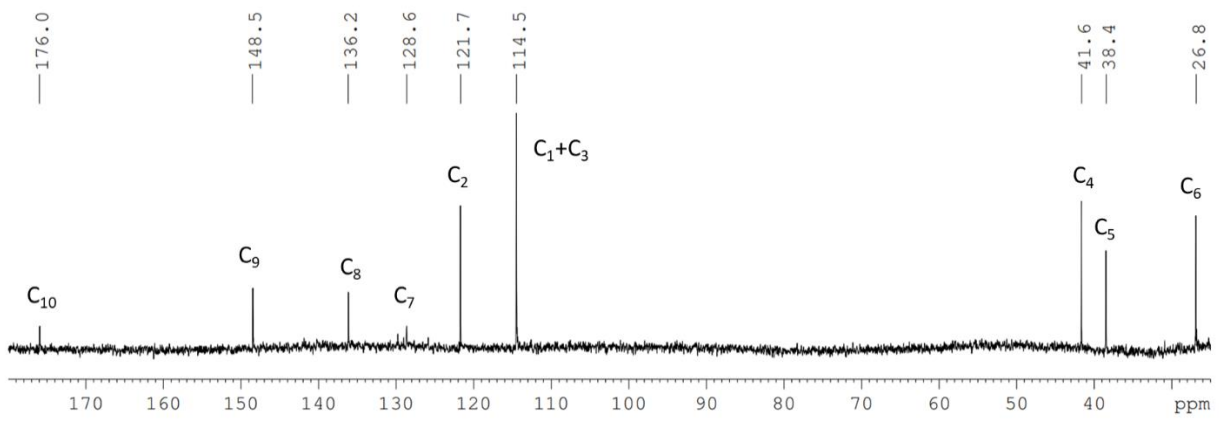


Figure 54: Spectre RMN ^{13}C du MO 211 purifié, (150 MHz, solvant H_2O / D_2O (90:10))

Pour lever cette apparente contradiction et relier les protons aux atomes de carbone auxquels ils sont liés, nous avons utilisé une expérience RMN bidimensionnelle HSQC (Heteronuclear Single Quantum Coherence) sélectionnant les corrélations entre protons et atomes de carbone à courte distance ($^1J_{C\text{-}H}$). Sur le spectre, il apparait que, les signaux des protons H_1 et H_3 couplent avec le même signal du spectre ^{13}C (δ= 114.5 ppm). En fait, il y a superposition des signaux des atomes de carbone C_1 et C_3 sur le spectre ^{13}C (Figure 55).

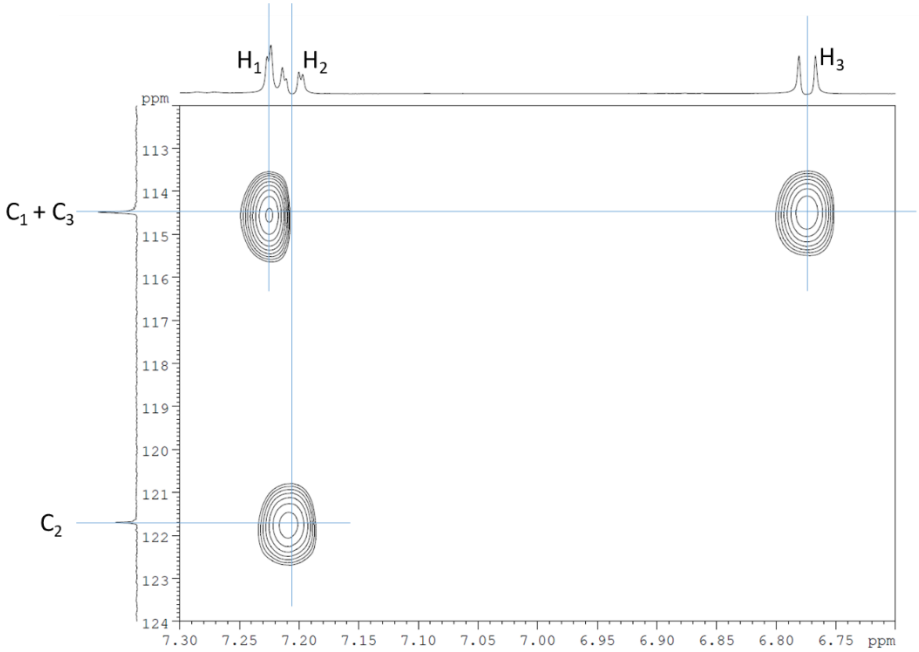


Figure 55: Extrait du spectre RMN ^{1}H-^{13}C HSQC du MO 211 purifié

Même si la séquence utilisée pour l'acquisition du spectre RMN ^{13}C n'est pas quantitative (non élimination de l'effet NOE ^{1}H-^{13}C et temps de retour à l'équilibre trop court par rapport à la relaxation de certains noyaux ^{13}C), la comparaison des intensités des signaux à 114,5 ppm et à 121,7 ppm (l'autre signal CH aromatique) corrobore bien la conclusion tirée de l'expérience HSQC.

Les autres tâches de corrélation sur le spectre HSQC (Annexe 10) permettent de lier les protons H_4, H_5 et H_6 aux atomes de carbone de déplacements chimiques respectifs 41,6, 38,4 et 26,8 ppm. Le déblindage des signaux des atomes de carbone C_4 et C_5 tendent à prouver la substitution du motif $–CH_2$-CH_2-CH_2- par des atomes d'azote plutôt que des atomes d'oxygène. Enfin, le déplacement chimique de l'atome de carbone quaternaire à 148,5 ppm conforte l'hypothèse de la présence d'un groupe phénolique, comme proposé à l'issu des analyses réalisées en spectrométrie de masse.

RMN HMBC ^{1}H-^{13}C et ^{1}H-^{15}N

Afin de confirmer les hypothèses précédentes et finaliser l'élucidation structurale, des expériences bidimensionnelles à longue distance HMBC (Heteronuclear Multiple Bond Connectivity) (ANNEXE 11 et 12) sont menées. Les résultats des couplages observés sont résumés dans le Tableau 19.

L'expérience ^{1}H-^{15}N HMBC sur l'échantillon du MO 211 avec l'isotope ^{15}N en abondance naturelle montre la présence de quatre tâches de corrélations (Annexe 12), résumées dans le Tableau 19. L'azote alpha couple avec les deux protons aromatiques H_1 et H_3 ainsi qu'avec les protons H_6 de la chaine aliphatique.

Protons	Atome de carbone avec couplage scalaire longue distance	Atome d'azote avec couplage scalaire longue distance
H_1	C_2, C_7, C_9 et C_{10}	N_α
H_2	C_1 et/ou C_3, C_9 et C_{10}	
H_3	C_7, C_8 et C_9	N_α
H_4	C_6, C_5 et C_8	
H_5	C_4 et C_6	
H_6	C_4 et C_5	N_α et N_β

Tableau 19: Interprétation des données obtenues par RMN 2D ^1H-^{13}C HMBC et ^1H-^{15}N HMBC

L'azote beta, quant à lui, ne présente qu'une seule tâche de corrélation visible : celle avec les protons H_6 (Tableau 19). Ces observations confirment la présence du motif 1,3-diaminopropyl, même si les tâches de corrélation issues des couplages $^2J_{N\alpha-H4}$ et $^2J_{N\beta\ -H5}$ restent invisibles. Dans la littérature, il apparait que les constantes de couplage à longue distance $^2J_{H-N}$, $^3J_{H-N}$, $^4J_{H-N}$ et $^5J_{H-N}$ sont généralement toutes comprises en valeur absolue entre 0 et 20 Hz [109, 110]. La faible concentration de l'échantillon et la valeur arbitraire du filtre à longue distance (5 Hz) sont sans doute à l'origine de ce phénomène. Les deux couplages supplémentaires observés ($J_{H1-N\alpha}$ et $J_{H3-N\alpha}$) indiquent que le branchement du motif diaminopropyl s'effectue *via* l'azote alpha. Du fait de l'impossibilité de couplage N-$H_{aromatique}$ à deux liaisons de distance, ces corrélations ont d'abord été interprétées comme des couplages ^3J (Figure 56). Cependant, les positions relatives des trois protons aromatiques déterminées précédemment est incompatible avec cette proposition de positionnement.

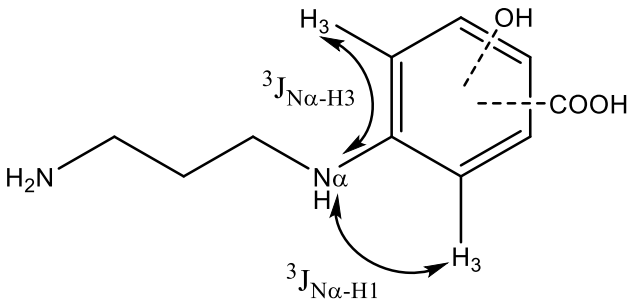


Figure 56: Structure suggérée par l'interprétation des données du spectre RMN HMBC ^1H-^{15}N

Ceci suggère donc plutôt des couplages à trois et quatre liaisons de distance, mais comme plusieurs références bibliographiques le montrent, des couplages ^1H-^{15}N à cinq liaisons de distance peuvent également être visibles [109, 110].

L'expérience ^1H-^{13}C HMBC confirme la connexion de la chaine aliphatique *via* l'azote alpha du fait de la présence de la tâche de corrélation entre les protons H_4 et l'atome de carbone C_8 aromatique.

Les corrélations visibles au sein du système aromatique (Figure 57, Tableau 20) ne permettent pas de

conclure quant au placement des différents substituants.

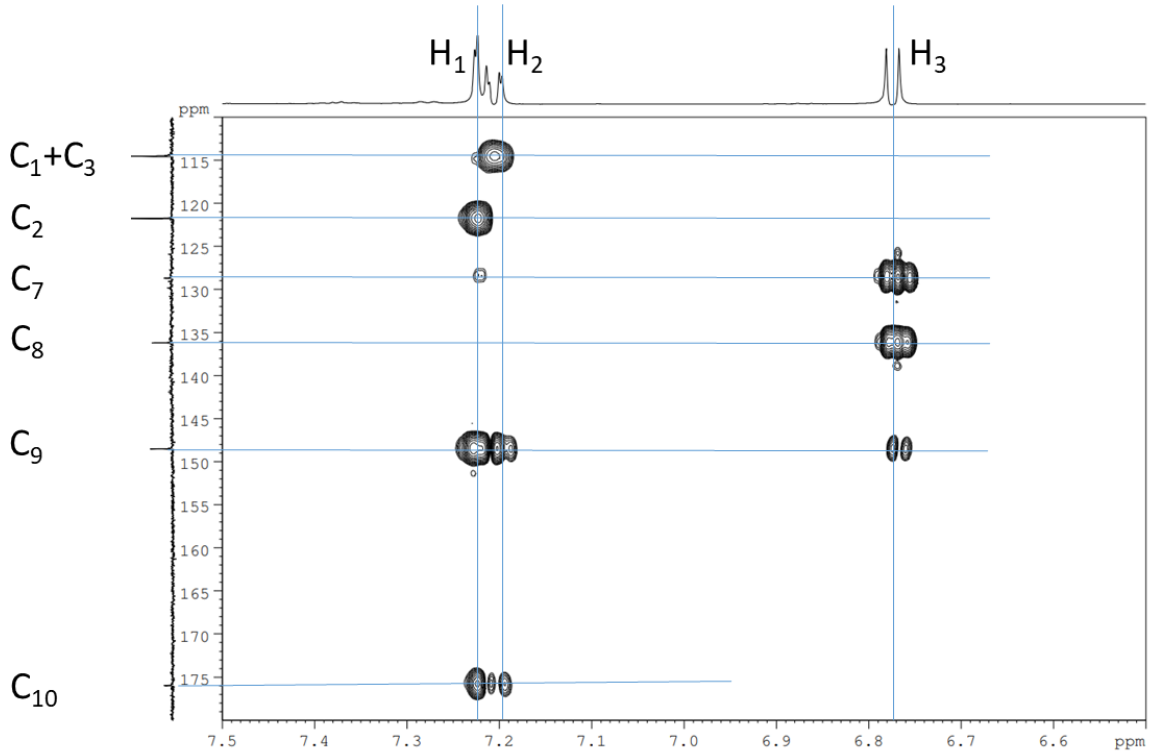


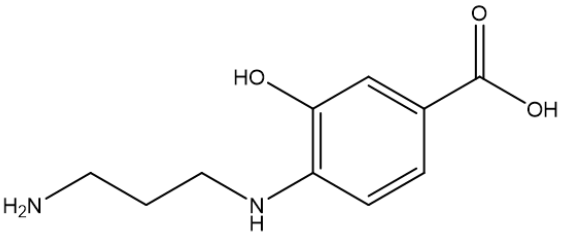
Figure 57 : Extrait du spectre HMBC ^{1}H-^{13}C pour la zone aromatique

Protons	Carbones avec couplage scalaire ^{2}J ou ^{3}J
H1	C_2, C_7, C_9 et C_{10}
H2	C_1+C_3, C_9 et C_{10}
H3	C_7, C_8 et C_9

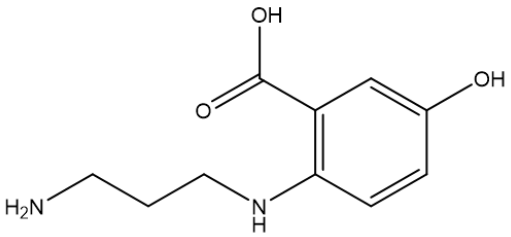
Tableau 20: Corrélations observées lors de l'expérience RMN HMBC ^{1}H-^{13}C pour les protons aromatiques

A l'issue des différentes expériences RMN, la position des différents substituants décorant le noyau aromatique n'a donc pas pu être déterminée avec certitude. Les fonctions chimiques amines et phénol présentent un état de protonation difficilement prédictible ce qui rend l'utilisation de logiciels de simulation de spectres RMN ^{13}C incertaine.

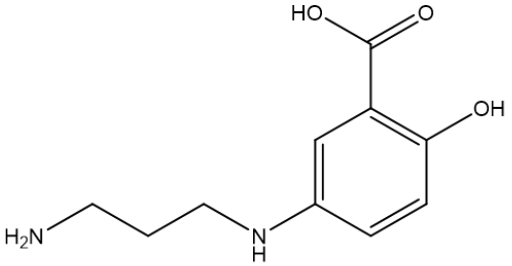
A ce stade, nous considérons donc que les six structures présentées en Figure 58 sont compatibles avec les données RMN et LC-MS^{n}.



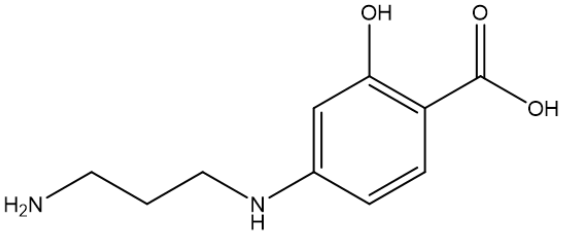
acide 4- (3- aminopropyl) amino- 3- hydroxybenzoïque
4DAP-3HBA



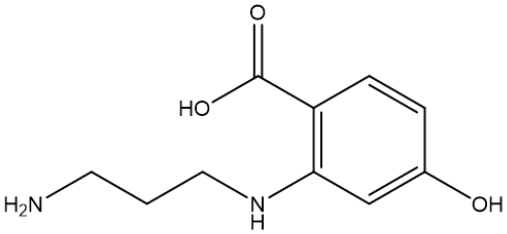
acide 2- (3- aminopropyl) amino- 5- hydroxybenzoïque
2DAP-5HBA



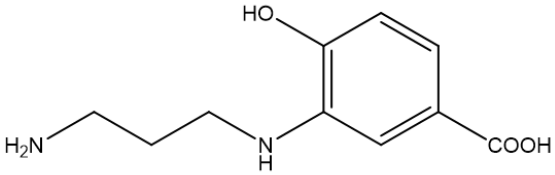
acide 5- (3- aminopropyl) amino- 2- hydroxybenzoïque
5DAP-2HBA



acide 4- (3- aminopropyl) amino- 2- hydroxybenzoïque
4DAP-2HBA



acide 2- (3- aminopropyl) amino- 4- hydroxybenzoïque
2DAP-4HBA



acide 3- (3- aminopropyl) amino- 4- hydroxybenzoïque
3DAP-4HBA

Figure 58: Six structures possibles pour le MO211

3.2.2.3 Validation de la structure du MO 211

Afin de finaliser l'élucidation structurale du MO211, nous avons décidé d'acheter les standards commerciaux. Seules cinq des six molécules ont pu être obtenues commercialement. Le 2DAP-4HBA se révélant être trop complexe à synthétiser.

Les cinq composés synthétisés ont été analysés en LC-MS et en RMN. En comparant les données obtenues, il s'est avéré que le 3DAP-4HBA semble correspondre le mieux au métabolite inconnu. En effet, les spectres RMN ^{1}H des composés deux composés (le MO211 purifié et le 3DAP-4HBA synthétisé) présentent les mêmes signaux. Seules les valeurs de déplacements chimiques varient légèrement (Figure 59). Nous avons donc choisi d'effectuer une analyse complémentaire en LC/MS afin de valider cette hypothèse.

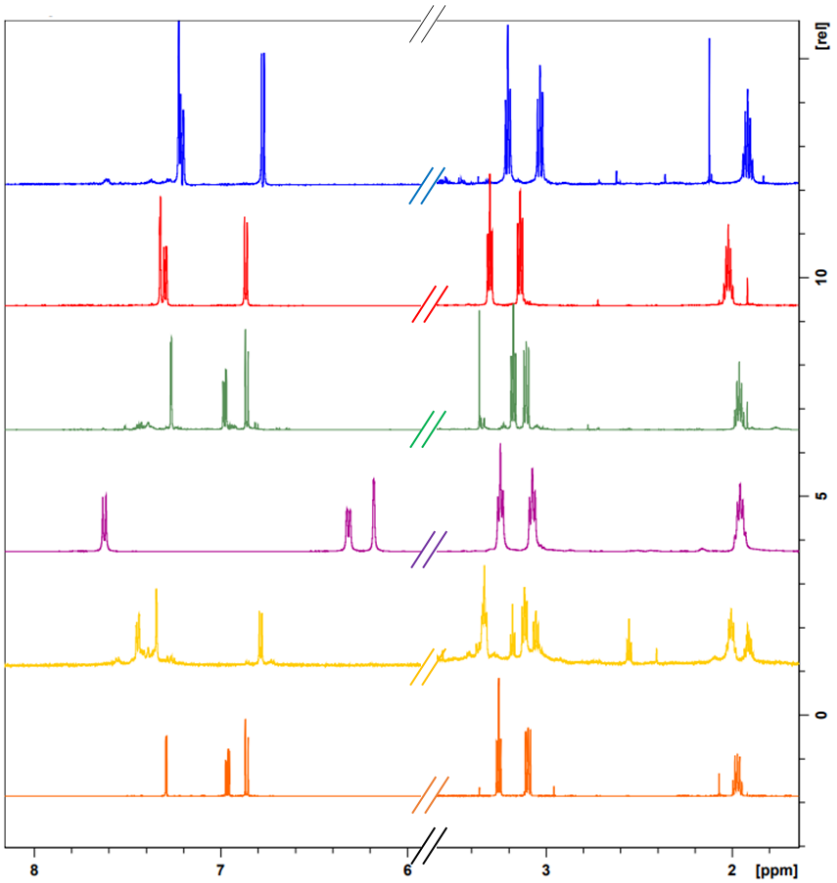


Figure 59: Extraits des spectres RMN ^{1}H du MO211 et des 5 composés synthétisés. (Spectres complets voir ANNEXE 13), 600 MHz, solvant H_2O/D_2O (90:10) de : (bleu) MO 211, (rouge) 3DAP-4HBA, (vert) 5DAP-2HBA, (violet) 4DAP-2HBA, (jaune) 4DAP-3HBA, (orange) 2DAP-5HBA.

L'analyse en LC/MS montre que le MO211 au sein du métabolome et 3DAP-4HBA pur présentent des temps de rétention quasi-identiques (Figure 60).

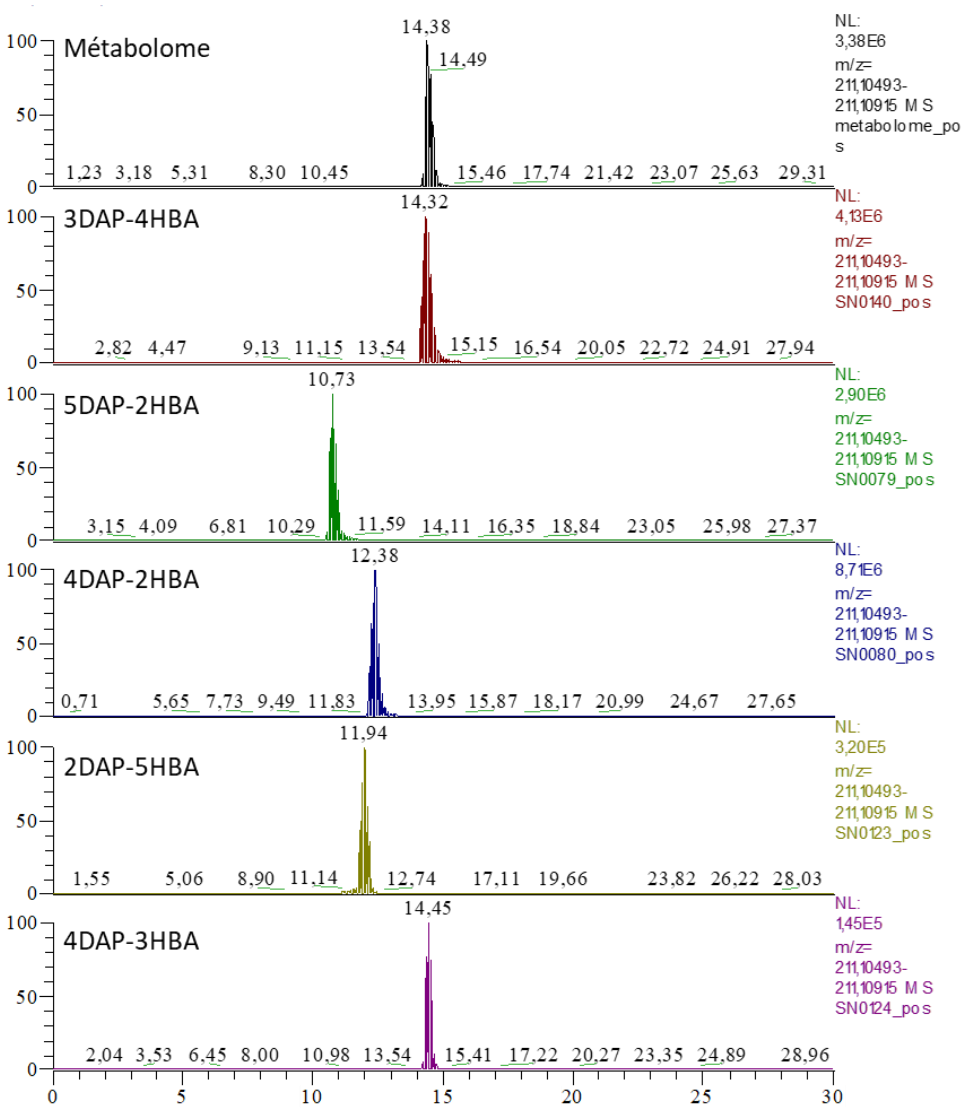


Figure 60: Analyse LC/MS des 5 composés synthétisés (5ppm, gradient de 50 minutes sur ZIC-pHILIC 150 x 4,6, 5 µm).

Une analyse complémentaire en dopant un échantillon de métabolome contenant du MO 211 par le 3DAP-4HBA ne conduit qu'à un seul pic chromatographique (Figure 61).

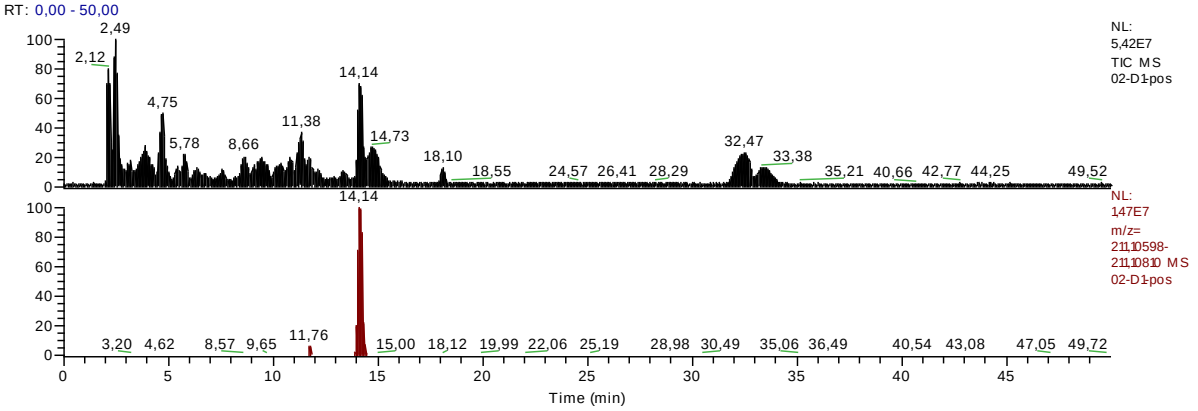


Figure 61: Analyse LC/MS d'un métabolome additionné du 3DAP-4HBA (50 µM) (gradient de 50 minutes sur ZIC-pHILIC 150 x 4,6, 5 µm). En haut, TIC, en bas, EIC (5 ppm) généré à partir des données .RAW pour la valeur m/z211,10704

Enfin, la fragmentation HCD de chacune des molécules synthétisées a été effectuée afin d'achever de les caractériser en spectrométrie de masse. Comme le montre le Tableau 21, chacune d'entre elle induit la production de fragments en différentes abondances (Annexe 14 et 15)

fragment formule brute perte correspondante	m/z 211 $C_{10}H_{14}O_3N_2$ ion parent	m/z 194 $C_{10}H_{12}O_3N^+$ NH_3	m/z 193 $C_{10}H_{13}O_2N_2^+$ H_2O	m/z 176 $C_{10}H_{10}O_2N^+$ $NH_3 + H_2O$	m/z 167 $C_9H_{15}O_3N^+$ CO_2	m/z 166 $C_8H_8ON_2^+$ $C_2H_4 + NH_3$	m/z 164 $C_9H_{10}O_2N^+$ $CH_2NH + H_2O$	m/z150 $C_9H_{12}ON^+$ $NH_3 + CO_2$	m/z 136 $C_7H_6O_2N^+$	m/z 134 $C_8H_8ON^+$	m/z 122 $C_7H_8O_2^+$ $NH_3 + H_2O + C_2H_4$	m/z 121 $C_7H_5O_2^+$	m/z 107 $C_7H_7O^+$
métabolome	46	59	100	11	x	30	x	2	0,2	x	1	x	x
3DAP-4HBA	45	54	100	10	x	28	x	2	0,2	x	1	x	x
5DAP-2HBA	100	61	22	20	x	85	x	3	0,5	x	0,5	x	x
4DAP-2HBA	10	8	100	4	0,6	2	x	0,5	0,1	x	x	0,1	0,1
2DAP-5HBA	100	40	34	2	36	28	40	15	0,14	0,21	10	x	x
4DAP-3HBA	84	100	0,3	2	0,5	60	x	3	x	x	12	x	x

Tableau 21: Récapitulatif des fragments observé en HCD-MS^2 de chacune des molécules synthétisées, à une énergie de collision normalisée à 30%.

Nous remarquons bien dans ce tableau que seul le 3DAP-HBA présente une fragmentation quasi-identique au MO211 : il s'agit donc bien de la bonne structure.

Les résultats présentés dans ce paragraphe (3.2) ont fait l'objet d'un manuscrit qui sera soumis au journal Metabolomics en septembre 2018.

3.2.3 D'autres métabolites orphelins ?

Afin de poursuivre l'effort d'identification structurale des métabolites inconnus, nous nous sommes penchés sur le cas d'un second métabolite, détecté également en mode d'ionisation positif : le MO193. De même que le MO211, le MO193 (m/z193,0966, t_R = 4.6 min sur Zic-*p*HILIC 150 x 4,6 mm², 5 µm) fait partie des 28 métabolites orphelins présents exclusivement dans les « métabolome quinate ». Son signal d'ionisation en mode positif est également intense et son rapport masse sur charge, identique à l'un des fragments MS^2 du MO211 (perte d'eau), nous ont logiquement amené à tenter d'identifier ce

métabolite.

3.2.3.1 Analyse du MO193 en HRMS

3.2.3.1.1 Propriétés générales

Le MO193 est nettement moins retenu sur la colonne chromatographique de 15 cm que le MO211 (Figure 41). Il est détecté à m/z193,09664 et m/z191,0826 dans les modes d'ionisations positif et négatif, respectivement. Sa masse neutre calculée est de m/z192,08936

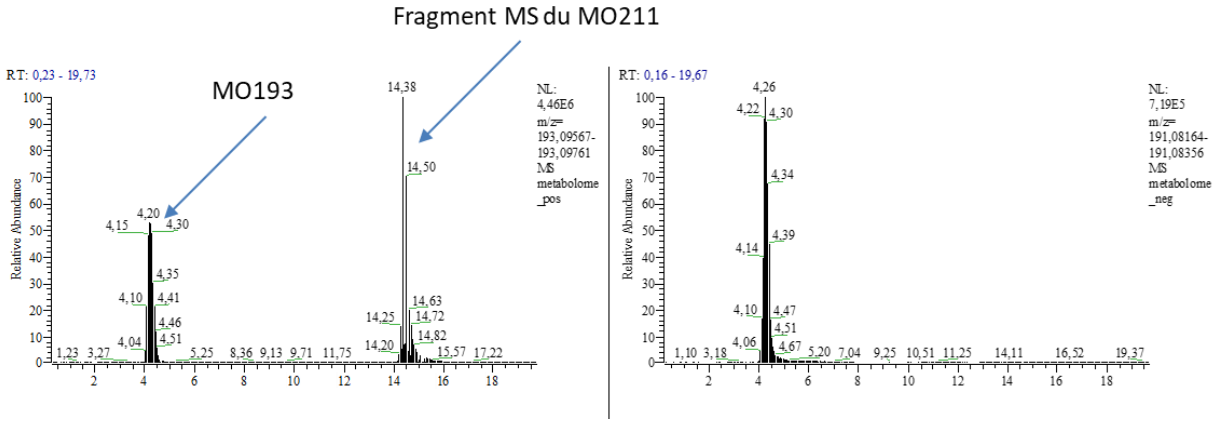


Figure 62: Détection en LC/HRMS du MO193. Le MO193 est détecté dans les deux modes d'ionisation sous forme de deux pics : à gauche, mode d'ionisation positif m/z =193,09664 ; à droite mode d'ionisation négatif m/z=191,0826. En mode d'ionisation positif, l'EIC extrait du m/z193,0966 présente un second pic, correspondant au fragment du MO211 (perte d'eau) formé dans la source à un temps de rétention de 14,6 min.

Ces informations (temps de rétention relativement bas et détection dans les deux modes d'ionisation) suggèrent une structure moins polaire que le MO211, mais possédant probablement des groupes fonctionnels acides et basiques. Par ailleurs, ce composé présente un spectre d'absorption UV (Figure 63) montrant trois maxima : 229, 270 et 310 nm. Les régions d'absorbance comprises entre 270 et 280 nm ainsi qu'entre 210 et 230 nm sont caractéristiques de composés organiques possédant un cycle aromatique. Il est donc très probable que le MO193 comporte un cycle aromatique.

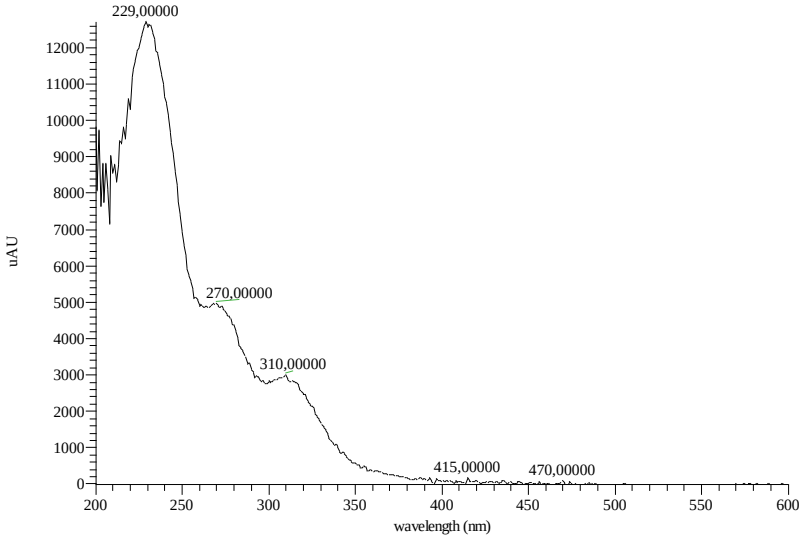


Figure 63: Spectre d'absorption du MO193

L'interrogation des bases de données publiques en utilisant la masse précise (CheBI et KEGG) ou chargée (METLIN) du MO193 propose huit résultats présentés en Figure 64.

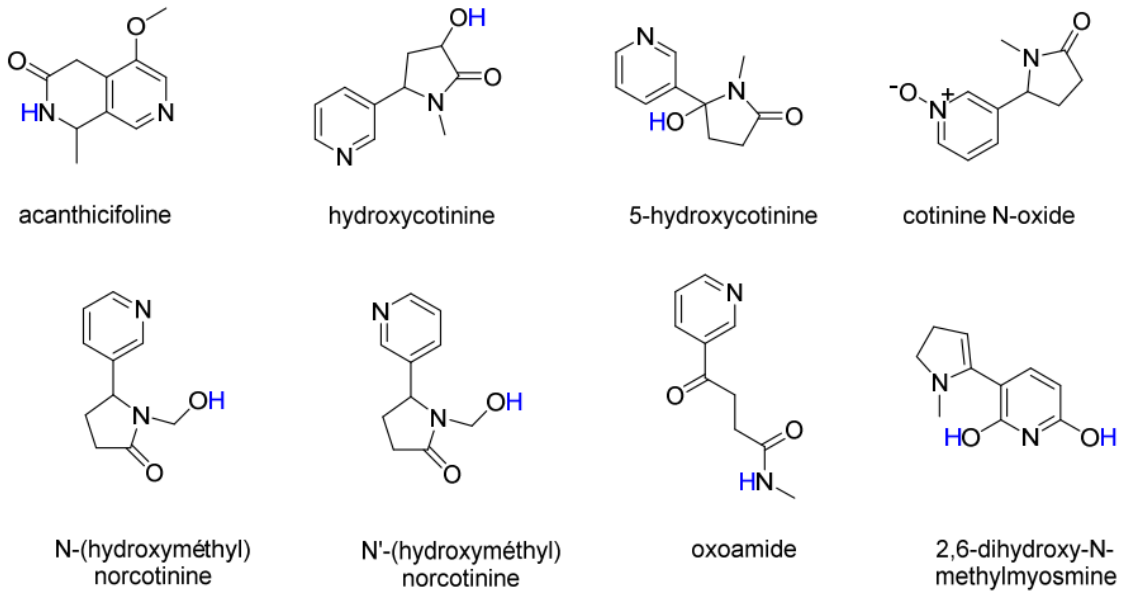


Figure 64: Structures des huit molécules proposées pour le MO193 par les bases de données KEGG, Chebi et METLIN sur la base de la masse précise (192,08936 Da). Les atomes d'hydrogènes colorés en bleu sont susceptibles de s'échanger lors des expériences d'échanges isotopiques.

3.2.3.1.2 Détermination de la composition élémentaire

A partir des spectres de masse acquis en haute résolution sur l'Orbitrap Elite (60 000k), la composition élémentaire du MO193 a été déterminée en utilisant, de même que dans le cas du MO211,

la macro Excel nommée les « Sept Règles d'Or [82] » . Afin de déterminer la formule brute du composé, nous avons intégré dans cet outil, plusieurs informations dont : la masse neutre du composé 192,08936 Da, ainsi que les abondances de certains isotopes dans le spectre de masse (M+1=10% et M+2=0.17%) (Tableau 22, Annexe 16).

Masse détectée	Isotopes (^{13}C) observés	intensité relative	Formule brute proposée par Xcalibur	δ ppm
193,09683	194,10025	10,9	$C_9{}^{13}CH_{15}O_2N_2$	-0,259
	195,10082	0,17	$C_8{}^{13}C_2H_{13}O_2N_2$	-0,406

Tableau 22 : Liste récapitulative des isotopes observés lors de l'analyse d'un métabolome au temps de rétention du MO193 (4,6 min. sur la colonne de 15 cm) en mode d'ionisation positif

Aucun isotope du chlore ou du phosphore n'est observé dans ce spectre de masse : nous pouvons donc supposer qu'il n'y a pas de chlore ou de phosphore dans la formule brute du MO193.

La seule formule brute proposée par les « 7 Golden Rules » $C_{10}H_{12}O_2N_2$, correspond en tout point aux caractéristiques énoncées ci-dessus. La formule brute du MO193 est donc : $C_{10}H_{12}O_2N_2$ et possède 6 insaturations.

3.2.3.1.3 Échanges H/D

Nous avons souhaité déterminer le nombre d'hydrogène échangeables en solution afin de pouvoir le comparer avec ceux des huit molécules proposées par les bases de données. Nous avions déjà analysé en LC/MS un métabolome repris dans le D_2O et analysé sur une colonne chromatographique équilibrée avec une phase chromatographique deutérée pour évaluer le nombre de protons labiles du MO211. Nous avons donc simplement extrait de ces données, les EIC de chacune des valeurs m/z correspondant au MO193 une fois marqué, deux fois marqué etc. Ces EIC sont présentés en Figure 65. Le temps de rétention du MO193 sur cette méthode chromatographique est de 2,1 minutes.

Au total, 4 échanges H/D en mode d'ionisation positive, et 2 en mode d'ionisation négatif ont été observés. Le métabolite neutre possède donc 3 protons labiles. Or les deux structures proposées par les bases de données possèdent 1 ou 2 protons labiles. Elles ne peuvent donc pas correspondre au métabolite orphelin. Il a donc été conclu que le MO193 est probablement un métabolite nouveau, non référencé dans les bases de données.

Il a donc été nécessaire de procéder à l'élucidation structurale de ce métabolite. Pour cela, nous avons choisi de procéder avec deux méthodes analytiques orthogonales : la RMN et la LC/MS.

Actuellement, nous travaillons à la caractérisation du MO193 en spectrométrie de masse, que nous présenterons dans les paragraphes 3.2.3.1.4 et 3.2.3.1.5. Afin de pouvoir procéder aux analyses RMN,

nous sommes également actuellement en train de développer des protocoles de purification, qui seront présentés au paragraphe 3.2.3.3.

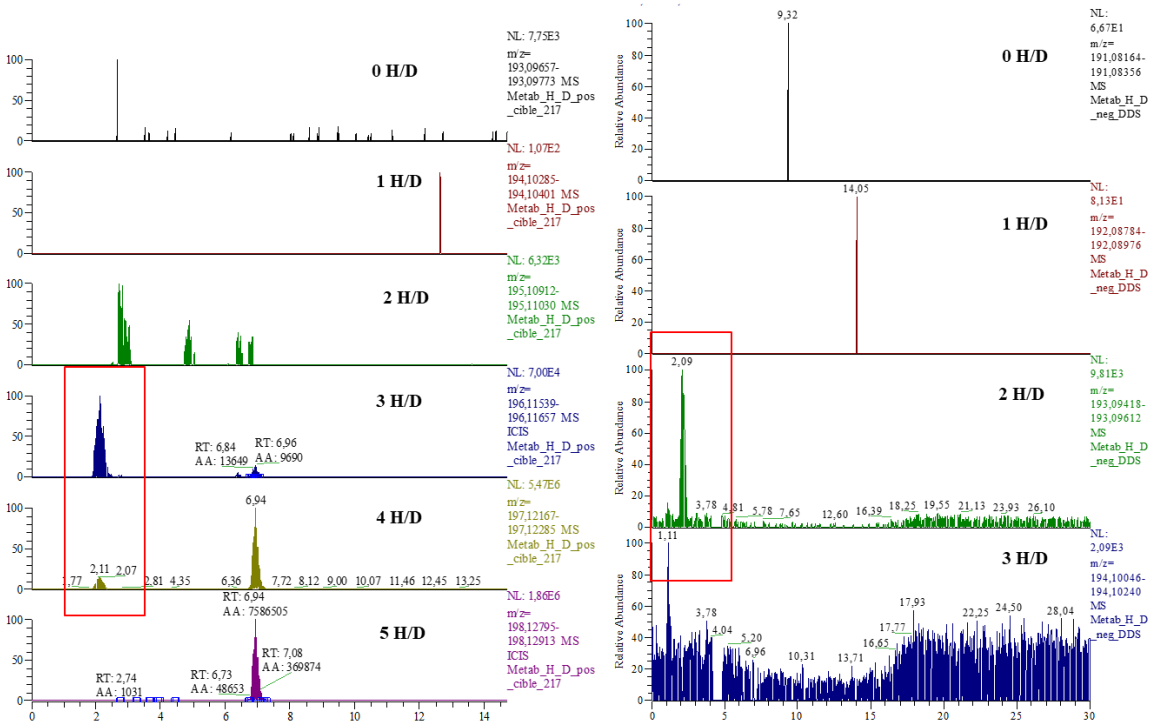


Figure 65: Détermination du nombre d'échanges H/D pour le MO193 dans les métabolomes deutérés dans les deux modes d'ionisation. On peut éliminer les propositions de bases de données

3.2.3.1.4 Plusieurs signaux pour le MO193

Nous discuterons dans un premier temps, des résultats obtenus lors de la détection en FTMS du MO193 et de ses fragmentations en mode CID et HCD dans le paragraphe suivant.

Lors de l'étude de comparaison des métabolomes cultivés sur quinate aux métabolomes cultivés sur succinate [104], l'analyse des données par XCMS a mis en évidence, l'existence d'une corrélation entre plusieurs signaux m/z. Ces signaux, observés en mode d'ionisation positif, possèdent un profil chromatographique très similaire.

L'analyse du métabolome sur le Velos Pro Orbitrap Elite (couplé à l'HPLC, méthode analytique utilisant la colonne de 15 cm), nous a permis d'observer également en mode d'ionisation positif cette corrélation entre différents signaux dont les profils chromatographiques sont très similaires (Figure 66, ANNEXE 17).

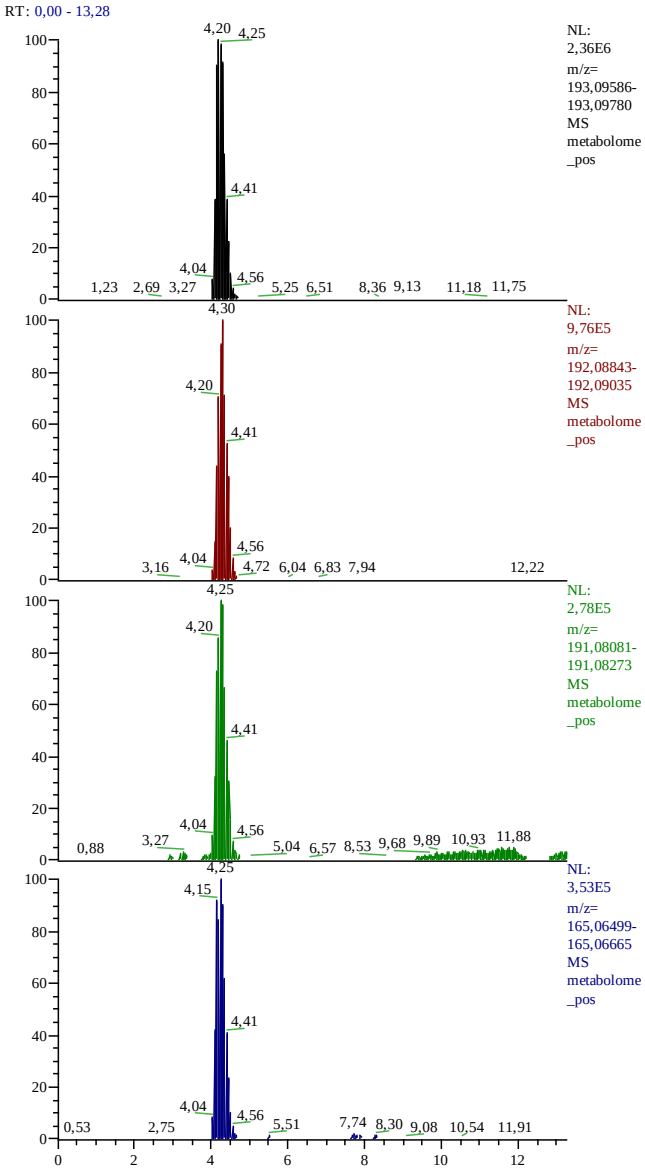


Figure 66 : EIC des différents ions semblant provenir du MO193. EIC extrait de l'analyse LC/MS d'un métabolome sur la colonne de 15 cm (5 ppm) pour les valeurs m/z suivantes : (a_1) – m/z193,0966, (a_2) – m/z192,08939, (a_3) – m/z191,08177, (a_4) – m/z165,06582

3.2.3.1.5 Premières expériences de fragmentation

Nous présenterons et discuterons ici des premières expériences de fragmentation en mode CID et HCD effectuées sur le MO193.

Les spectres de fragmentation MS^2 du MO193 obtenus dans le mode d'ionisation positif (Tableau 23) présentent un ion majoritaire : m/z 165,06582 (m/z 165). Il correspond à la perte de C_2H_4.

ion précurseur	m/z	Intensité relative	Composition élémentaire des ions produits	perte associée	δppm
193,09727	165,06587	58,5	$C_8H_9O_2N_2$	C_2H_4	0,016
	138,05502	100,0	$C_7H_8O_2N$	C_2H_4 + HCN	0,065
	120,04439	10,3	C_7H_6ON	C_2H_4 + HCN + H_2O	0,001
	94,06497	30,7	C_6H_8N	C_2H_4 + HCN + CO_2	0,156

Tableau 23 : Produits de fragmentations résultant des expériences HCD-MS^2 sur le MO193, à une énergie de collision normalisée à 35%.

Trois autres ions fragments moins abondants sont également observés : m/z138,05502 (m/z 138) ; m/z120,04439 (m/z 120) et m/z=94,06497 (m/z 94). Ils sont également observés dans lors des expériences de fragmentation MS^3, à partir de l'ion m/z 165 lui-même généré par la perte de C_2H_4 à partir de l'ion parent. Enfin, les ions m/z 120 et m/z 94 sont observés lors d'expériences MS^4, à partir de l'ion m/z 138 (ANNEXE 18).

En mode d'ionisation négatif, la perte majoritaire observée est une perte de CO_2 (m/z147,09290).

Plusieurs hypothèses peuvent être émises à la suite de ces analyses :

- les 6 insaturations attendues et la présence du fragment m/z 94 de composition élémentaire C_6H_8N laissent penser à la présence d'un cycle aromatique sur la molécule. Les longueures d'absorbance maximales du MO193 corroborent cette hypothèse.

- la molécule comporte vraisemblablement un groupement ester ou acide carboxyique (perte de CO_2 en mode d'ionisation positif et négatif), ainsi qu'une méthylamine ou methylimine (à l'origine de la perte d'HCN en mode d'ionisation positif) ainsi qu'une chaîne éthlyle (à l'origine de la perte de C_2H_4).

D'autres expériences de fragmentations sont encore à effectuer, notamment pour expliquer l'ordre des fragmentations que nous observons ici. En effet, il semble que dans le cas du MO193, la perte d'éthylène soit plus facile que la perte d'un groupement ester ou carboxylique, ou encore d'une amine. Il serait alors intéressant de comparer nos résultats de fragmentations à ceux obtenus pour des standards commerciaux, nous permettant de conclure sur la présence de certains groupements chimiques dans la molécule.

3.2.3.2 Protocole d'extraction du métabolome

Lors de la mise en place du travail de purification pour l'élucidation structurale du MO211 le protocole de production des métabolites a été réadapté à partir du protocole de production de métabolome utilisé pour le criblage. Cependant, l'extraction et la purification du MO211 a permis d'obtenir seulement 100 µg de composé à partir de 1,5L de culture bactérienne. Au total, une semaine

complète de travail a été nécessaire pour obtenir cette quantité.

Dans le cas de métabolites moins abondants, nous serions dans l'obligation de multiplier les extractions à partir de cultures en plaques 24 puits. Nous avons plutôt décidé de repenser le protocole d'extraction des métabolites. L'augmentation du volume de culture serait bien évidemment avantageuse car il permettrait d'augmenter la quantité de cellules produite et donc de maximiser la quantité de MO produite.

Ce travail de développement d'un nouveau protocole de production de métabolites sera illustré ici au travers du MO211, du MO193 détectés tous deux en mode d'ionisation positif, ainsi que de trois métabolites identifiés présents chez ADP1 dans les « métabolomes quinate » : la proline, la cytidine et l'adénosine.

Pour chacun des tests présentés ci-dessous, une préculture servant à ensemencer les différents volumes de culture est préparée la veille au soir. Pour cela, 5 µL d'une culture glycérolée d'ADP1 est versée dans 50 mL de MAQ contenu dans un erlenmeyer de 200 mL, puis placé, pendant la nuit, à 30°C, 150 rpm dans un incubateur (Infors). Le lendemain matin, les cultures sont ensemencées à DO_{600} = 0,05 à partir de cette préculture.

3.2.3.2.1 Relation entre volume de culture et teneur en MO

Comme nous l'avons précédemment évoqué dans le cas du MO211, l'augmentation des volumes de culture provoquait la chute de concentration de métabolite dans les métabolomes alors extraits (3.2.2.1). Une hypothèse peut expliquer le phénomène : à mesure que le volume augmente, une agitation fixée à 150 rpm est de moins en moins efficace. Ceci peut provoquer une mauvaise oxygénation des cellules, et donc mener à un état métabolique des cellules différentes. Afin de remédier à ce problème, nous avons donc envisagé une solution : l'utilisation d'une verrerie spécifique, tels les flacons Fernbach à ailettes. Ces ailettes brisent le mouvement circulaire du liquide en rotation dans l'agitateur et augmentent ainsi l'oxygénation des cellules (Figure 67).

Figure 67: Flacon Fernbach, avec fond à ailettes

Afin de comparer les verreries utilisables (erlenmeyer de 2 L ou Fernbach à ailette de 2,6 L), nous avons étudié l'abondance des métabolites dans deux conditions de culture différentes :

- 400 mL de MAQ (ensemencé à DO_{600} =0,05 comme précisé plus haut) ont été mis à incuber dans un erlenmeyer de 2L à 30°C, 150 rpm,

- 600 mL de MAQ (encemencé à DO_{600} =0,05 comme précisé plus haut) ont été mis à incuber dans un fernbach à ailette de 2,6L, à 30°C, 150 rpm.

Lorsque la culture atteint une DO_{600} = 0,5, 5 mL de ces milieux sont prélevés. Ils subissent alors le même protocole d'extraction (protocole d'extraction plaque, utilisé pour le criblage). Les métabolites extraits sont alors analysés sur la colonne de 15 cm. Les résultats sont présentés dans le Tableau 24.

Ce protocole d'extraction a été répété trois jours d'affilés. Ainsi chaque mode de culture a été testé 3 fois.

	MO211	**MO193**	**proline**	**cytidine**	**adénosine**
Protocole erlen	2,19 ± 0,65 E+05	7,12 ± 1,18 E+04	1,02 ± 0,13E+06	5,46 ± 0,5 E+04	2,08 ± 0,3 E+04
Protocole Fernbach	5,83 ± 1,69 E6	2,83 ± 1,04 E+6	1,46 ± 0.57 E+06	9,6 4± 3,51 E+04	5,97 ± 1,0 E+03

Tableau 24: Comparaison des abondances relatives de différents ions pseudo-moléculaires selon le type de verrerie de culture utilisée pour un même volume et protocole d'extraction. Les abondances sont indiquées en unités arbitraires de surface de pic.

La première observation fut la vitesse de croissance des cellules. En effet, la culture contenue dans le Fernbach a atteint la DO souhaitée (0,5) en 3h30 d'incubation seulement, contre 4h30 pour la culture en erlenmeyer. Ceci est explicable par la meilleure oxygénation que permettent les ailettes du Fernbach.

La seconde observation concerne l'abondance des métabolites orphelins. Comme le montre le Tableau 23, ceux – ci sont plus abondants lorsque l'oxygénation du milieu de culture est meilleure, dans le Fernbach à ailettes. Ainsi, nous avons choisi d'effectuer nos cultures bactériennes en Fernbach pour production de métabolites en vue d'une purification.

Nous possédons deux Fernbach de 2,6 L à ailette au laboratoire, ainsi en une journée, nous sommes en mesure de produire 1,2 L de culture bactérienne.

3.2.3.2.2 Influence de la valeur de DO_{600} sur l'abondance des métabolites extraits

La phase de croissance exponentielle des cellules chez ADP1 est comprise entre des valeurs de DO_{600} de 0,2 et 1,2). Durant cette phase, le métabolisme de la cellule est à l'état stationnaire, et doit donc être indépendant de la valeur de la DO.

On ne s'attend donc pas, dans cette zone de densité cellulaire, à une influence de la DO sur la teneur en

métabolites. Nous avons malgré tout, tenu à vérifier ce postulat. Une culture (600 mL en Fernbach à ailettes) a été ensemencée à DO= 0,05 à partir de la préculture ensemencée la veille, et mise à incuber (30°C, 150 rpm). Des échantillons de 5 mL sont prélevés dans cette culture à différentes temps (120, 180, 210, 240, 270, 285 et 300 minutes). Les valeurs de DO sont mesurées à chacun des prélèvements. Les métabolites sont extraits selon le protocole utilisé pour le criblage.

Ce protocole a été répété un second jour d'extraction. Les résultats obtenus pour ces deux jours d'extraction sont présentés dans le Tableau 25.

Temps de prélèvement	DO	MO211	MO193	proline	cytidine	adénosine
120	0,22	1,12E+05	4,30E+04	5,37E+05	1,08E+04	6,49E+03
180	0,37	6,04E+05	1,89E+05	9,85E+05	5,20E+04	3,38E+04
210	0,50	4,86E+06	2,63E+06	1,39E+06	1,05E+05	7,12E+04
240	0,59	4,70E+06	1,69E+06	1,42E+06	7,68E+04	7,39E+04
270	0,67	6,74E+06	3,44E+06	1,72E+06	6,83E+03	7,60E+04
285	0,80	5,00E+06	2,55E+06	1,81E+06	8,19E+04	2,62E+05
300	0,93	4,51E+06	3,12E+06	3,13E+06	1,83E+05	1,13E+05

Temps de prélèvement	DO	MO211	MO193	proline	cytidine	adénosine
120	0,18	5,05E+05	8,83E+04	6,44E+05	6,44E+03	6,62E+03
180	0,34	7,83E+05	1,59E+05	9,39E+05	5,50E+04	4,86E+04
210	0,53	4,72E+06	1,88E+06	1,20E+06	4,37E+04	3,42E+04
240	0,61	5,06E+06	1,39E+06	1,22E+06	1,01E+05	1,02E+05
270	0,74	4,91E+06	2,34E+06	1,34E+06	9,74E+04	5,86E+04
285	0,89	3,74E+06	2,90E+06	2,35E+06	8,21E+04	2,43E+05
300	1,09	2,40E+06	1,51E+06	2,41E+06	1,78E+05	3,20E+04

Tableau 25 : Comparaison des abondances relatives de différents ions selon la DO pour un même volume d'extraction. Les abondances sont indiquées en unités arbitraires de surface de pic. En haut, premier réplicas, en bas, second réplicas.

Les résultats présentés dans le Tableau 25 montrent l'évolution de l'abondance des métabolites en fonction de la DO à laquelle la culture a été stoppée. Nous pouvons remarquer que l'abondance pour les deux métabolites orphelins, MO211 et MO193, ainsi que les métabolites de référence augmente jusqu'à ce que la DO atteigne 0,5 puis se stabilise. L'abondance des MO dans les échantillons est ensuite relativement constante malgré l'augmentation de la quantité de cellules lysées. Ce genre de phénomène, que nous ne savons pas expliquer chez ADP1, avait déjà été observé lors de la mise en place du protocole d'extraction des métabolites pour le criblage de la banque (protocole plaque). Ainsi, dans le cas du protocole d'extraction en plaque, la DO optimale à laquelle les métabolites sont extraits avait été déterminée entre 0,4 et 0,6.

Lorsque nous souhaiterons préparer de grandes quantité de métabolomes pour la purification de

métabolite, nous procéderons donc à l'extraction des métabolites lorsque la DO de culture des cellules aura atteint et dépassé une DO_{600} =0,5.

3.2.3.2.3 Influence du solvant utilisé pour l'extraction sur la teneur en MO193

Une fois la quantité de cellules désirée obtenue, il faut adapter le protocole d'extraction des métabolites à de telles quantités. En effet, procéder à l'extraction de 1,2 L de culture cellulaire par aliquot de 5 mL est impensable. Nous avons alors envisagé deux options : les extractions *via* l'utilisation de solvants, ou l'utilisation d'une FrenchPress. Cette seconde solution paraissait particulièrement prometteuse, car elle permet la lyse des cellules sans l'ajout de solvants organiques pouvant entrainer la dégradation de certains métabolites. Cependant, l'utilisation de la FrenchPress n'a pas permis l'extraction de métabolites. Nous avons donc choisi d'opter pour l'extraction de métabolites par l'utilisation de solvants. Le laboratoire utilise couramment deux mélanges de solvants de lyse : le premier étant celui utilisé pour les extractions de métabolomes pour le criblage de la banque (1 : 3 : 1 H_2O/MeOH/ACN), le second étant utilisé dans le cadre d'un autre protocole d'extraction (ACN/MeOH 80 : 20 (Annexe 2). Nous avons donc choisi de tester ces deux solvants de lyse sur nos cultures.

A partir d'une culture de 600 ml réalisée en Fernbach à ailettes (DO = 0,4), deux aliquots de 200 mL ont été prélevés et centrifugés (4000 rcf, 4°C, 10 minutes) lorsque la DO_{600} de la culture eut atteint 0,6. Le surnageant est ensuite éliminé. Chacun des culots cellulaires a alors été extrait avec 50 mL de solution de lyse, soit avec le mélange H_2O/MeOH/ACN 1 : 3 : 1, soit avec le mélange ACN/MeOH 80 : 20.

Dans le cas du protocole d'extraction de métabolites utilisé pour le criblage de la banque, le mélange subit alors deux cycles de chocs thermiques (-50°C/température ambiante). Le but est ici de congeler rapidement le mélange afin de favoriser la rupture des parois cellulaires lors de la formation des cristaux de solvant. Cependant, la température de fusion du méthanol est de -98°C et celle de l'acétonitrile de -45°C. Afin d'assurer la congélation rapide des deux mélanges de lyse, nous avons donc choisi d'effectuer les chocs thermiques dans les conditions suivantes : azote liquide/bain marie à 65°C répété 6 fois.

Ce protocole a été répété trois fois, trois jours d'affilé et les résultats obtenus sont présentés dans le Tableau 26.

Nous remarquons qu'avec le mélange H_2O/MeOH/ACN 1 :3 :1, la teneur en métabolites orphelins et en adénosine extraits est nulle. Ce solvant de lyse n'est donc pas du tout adapté à l'extraction de telles quantités de cellules. En revanche, l'utilisation du mélange ACN/MeOH 80 : 20 en guise de solvant de lyse permet d'extraire correctement les métabolites.

	MO211	**MO193**	**proline**	**cytidine**	**adénosine**

ACN/MeOH/H_2O	0	0	1,55 ± 0,13 E+05	1,28 ± 0,7 E+04	0
ACN/MeOH	4,17 ± 1,18 E+06	1,98 ± 0,67 E+05	1,69 ± 0.28 E+06	4,70 ± 1,55 E+04	1,17 ± 0,45 E+03

Tableau 26 : Comparaison des abondances relatives de différents ions selon le solvant de lyse utilisé pour un même volume d'extraction. Les abondances sont indiquées en unités arbitraires de surface de pics.

Nous ne savons à ce jour pas expliquer ce phénomène. Nous pouvons émettre seulement deux hypothèses. La première concerne le volume de solvant de lyse par rapport au volume de cellules obtenu après centrifugation et élimination du milieu de culture. En effet, dans le cas du criblage de la banque, 200 µL de solution de lyse est ajouté sur le culot de 5 mL de culture à DO_{600} environ égale à 0,5. Dans le cas de notre expérience, nous ajoutons 50 mL sur le culot obtenu après centrifugation de 200 mL de culture à DO environ égal à 0,6. Nous ajoutons donc bien plus de solvant de lyse par rapport à la quantité de cellule à lyser dans le second cas par rapport au premier. Il est possible que ce paramètre affecte la lyse des cellules et limite donc l'extraction des métabolites. Nous pouvons également supposer que la vitesse de congélation entre en jeu. En effet, bien que nous utilisions ici de l'azote liquide, la congélation complète de 50 mL de ce genre de solvants nécessite près de deux minutes. Il est donc possible que la vitesse de congélation ne soit optimale pour ce solvant dans de telles quantités.

Nous utiliserons donc le mélange suivant en solvant de lyse : ACN/MeOH 80 : 20. Nous effectuerons également les chocs thermiques dans les conditions suivantes : azote liquide/bain marie à 65°C répété 6 fois. Nous avons choisi arbitrairement un volume de solution de lyse égal à 50 mL pour 200 mL de culture centrifugée. Au vu des hypothèses émises précédemment il serait utile d'étudier également l'influence du volume de lyse sur la qualité de l'extraction pour une même quantité de cellules à lyser. Pour cela, nous pourrions ensemencer 600 mL de MAQ dans un Fernbach. Une fois que la DO aura atteint 0,5, nous séparerions le volume de culture en 6 échantillons de 100 mL. Une fois centrifugés, et le surnageant éliminé, à chacun d'entre eux serait ajouté un volume différent de solution de lyse, par exemple : 5, 10, 25 ,50 et 75 et 100 mL, puis, tous ces échantillons suivraient le même protocole d'extraction jusqu'à l'injection pour analyse.

Nous disposons actuellement d'un protocole d'extraction de métabolites efficace, nous pouvons extraire en une journée 1,2 L de cultures cellulaires.

3.2.3.3 Purification du MO193

Les tests décrits nous ont permis d'obtenir des quantités relativement importantes de métabolome. Nous avons donc décidé d'entamer le développement de méthode pour la purification du MO193.

Comme, nous avons pu le présenter en première partie de ce paragraphe (3.2.3.1), le MO193 est moins polaire que le MO211. Nous avons donc envisagé une purification en HPLC préparative sur une colonne

de type C18. Nous disposons au laboratoire d'une colonne préparative de dimensions : 250 x 10, 5µm. Une telle géométrie de colonne permet l'injection de 1 à 5 mL d'échantillon et cette caractéristique est particulièrement intéressante dans notre cas. En effet, nous serions à même de devoir injecter de grandes quantités de métabolomes pour pouvoir obtenir une quantité de MO suffisante après purification.

Nous avons donc développé la méthode de purification la plus adéquate à ce métabolite, sur une colonne analytique C18 (150 x 2,1, 5µm). Quelques premiers essais ont été effectués. Tout d'abord, nous avons étudié le comportement du métabolite dans le cas d'une séparation isocratique ACN/ tampon aqueux 92 : 8. Trois types de phases aqueuses ont été testés :

- pH 9,9 : 10 mM NH_4CO_3, ajusté à pH 9,9 grâce à NH_4OH,
- pH 6,8 : 10 mM NH_4OAc
- pH 4 : 10 mM NH_4OAc, ajusté à pH 4 grâce à l'ajout d'acide acétique.

Les différents chromatogrammes obtenus sont présentés dans la Figure 68.

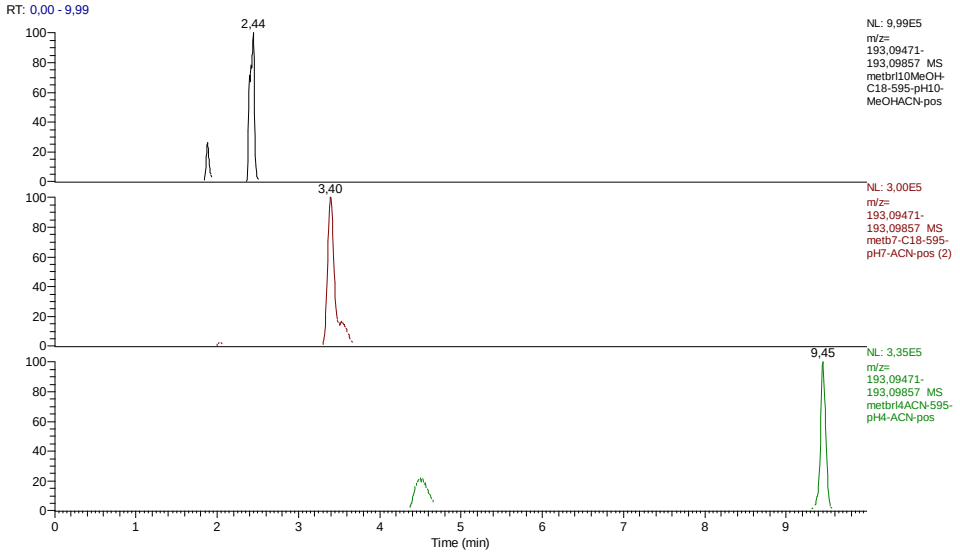


Figure 68 : Évaluation de la rétention du MO193 dans différentes conditions de pH. EIC extrait de 3 analyses d'un métabolome en isocratique sur la colonne C18 150 x 2,1, 5µm. Composition de la phase mobile : en haut, ACN/H_20 pH 9,9 (92 : 8), au centre ACN/ H_20 pH 6,8 (92 : 8), en bas ACN/ H_20 pH 4 (92 :8).

Nous pouvons remarquer la présence de deux pics chromatographiques pour le rapport m/z193, 0966. En effet, nous analysons ici un métabolome complet. Le premier pic détecté ($t_{=}$1,8 minutes dans le premier cas, t_R = 2 minutes dans le second et enfin t_R = 4,5 minutes dans le troisième cas) correspond à la détection du fragment correspondant à la perte d'eau dans la source du MO211. Il ne s'agit donc pas du MO193. Les pics les plus intenses à t_R = 2,44, 3,40 et 9,45 minutes sont le MO193.

La rétention du MO193 est très variable en fonction du pH utilisé, et c'est lorsque le tampon à pH 4 est

utilisé que la rétention du MO193 est la meilleure. Cependant, ce temps de rétention n'est pas encore optimal pour une purification en HPLC préparative (nous souhaitons un temps de rétention au minimum à 10-12 minutes pour s'assurer une bonne purification). Afin d'améliorer la rétention du MO193, sur la colonne analytique, nous avons alors choisi d'utiliser en lieu et place d'ACN, un mélange ACN/MeOH 50 : 50. En effet, le methanol est plus polaire que l'acétonitrile, nous supposons qu'il permettra de retenir un peu mieux le MO193 sur la colonne C18. Nous avons alors effectué plusieurs essais d'élution en utilisant la phase mobile suivante : x% ACN/MeOH 50 : 50, 100-x% pH4 (x = 5,7 et 8). Les résultats sont présentés dans la Figure 69.

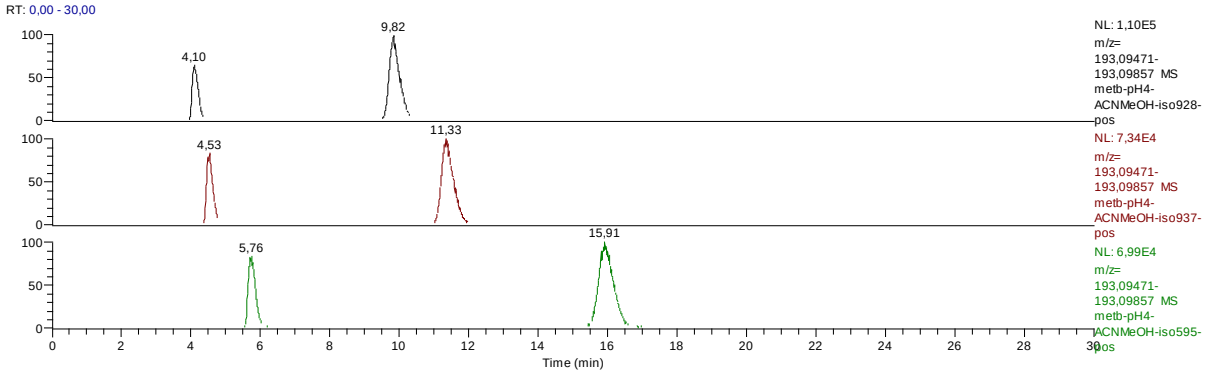


Figure 69 : Évaluation de la rétention du MO193 dans différentes conditions chromatographiques. EIC extrait de 3 analyses d'un métabolome en isocratique sur la colonne C18 150 x 2,1, 5µm. Composition de la phase mobile : en haut, ACN/$H_2$0 pH 4 (92 : 8), au centre ACN/$H_2$0 pH 4 (93 : 7), en bas ACN/$H_2$0 pH 4 (95 : 5)

Là encore, nous détectons le fragment dans la source du MO211 à des temps de rétention inférieurs à 6 minutes. Les signaux correspondant au MO193 sont ceux aux t_R = 9.82, 11,33 et 15,91 minutes (du haut vers le bas). Le meilleur temps de rétention pour la purification du MO193, t_R=15,9 minutes est atteint pour une phase mobile composée de 5% ACN/MeOH 50 : 50, 95% pH 4. De plus le pic garde une forme gaussienne chromatographique symétrique. Nous avons donc choisi d'effectuer une première purification du métabolome dans ces conditions.

Nous avons alors transposé la méthode analytique sélectionnée à l'échelle préparative à l'aide des convertisseurs fournis par les fabricants Thermo et Sigma Aldrich. Nous avons injecté 1 mL de métabolome dilué dans la même phase mobile que celle utilisée pour la purification, sur la colonne préparative C18 du laboratoire.

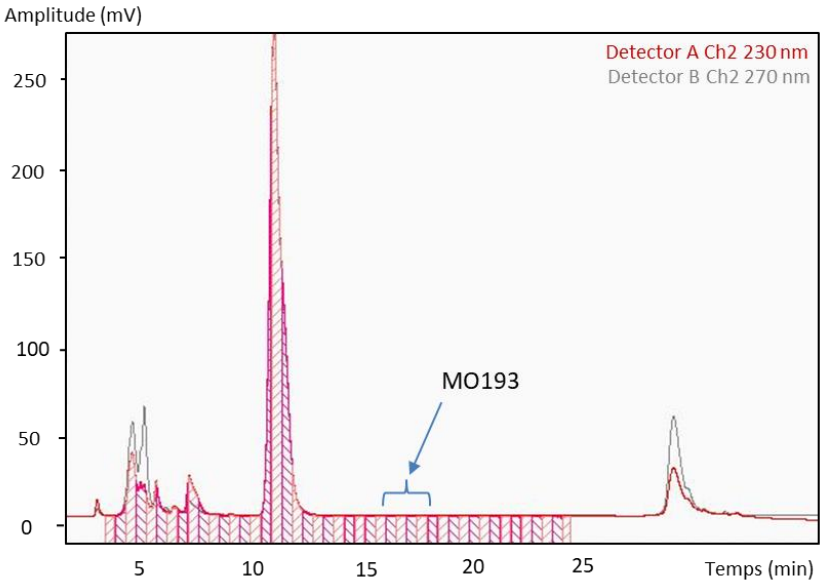


Figure 70: Première étape de purification du MO193. Suivi UV (230 et 270 nm) de la purification du métabolome sur HPLC préparative.

Cependant, la purification du MO193 s'avère plus ardue que celle du MO211. En effet, comme le montre la Figure 70, le métabolite est peu UV actif. Or, nous suivons l'élution de nos composés lors de leur purification par la mesure de deux longueurs d'ondes. Malgré l'utilisation de longueurs d'ondes optimales d'absorbance du MO193, aucun pic UV n'est visible lors de l'injection d'un échantillon sur l'HPLC préparative. Nous sommes donc contraints d'analyser la totalité des fractions en LC/MS afin de détecter en spectrométrie de masse la présence ou l'absence du MO193 dans chacune d'entre elle. Ce travail supplémentaire alourdit le travail de purification du MO193 à partir du métabolome.

Les fractions contenant le MO193 ont été rassemblées et séchées. Nous les avons alors analysées sur la colonne de 15 cm afin d'évaluer la qualité de la purification. Bien que de nombreux métabolites soient éliminés de l'échantillon, comme le MO211 par exemple, la pureté du MO193 ne parait pas suffisante pour procéder à l'analyse en RMN d'un tel échantillon pour procéder à l'élucidation structurale du MO. Ceci nous laisse donc penser qu'une seconde étape de purification sera nécessaire pour obtenir le MO193 pur.

Étant donné la lourdeur qu'impose l'HPLC préparative pour le suivi de la composition des fractions dans le cas du MO193 (nous devons analyser toutes les fractions en LC/MS pour détecter la présence ou l'absence du MO), nous envisageons d'autres moyens de purifications du MO193. Les extractions en phase solide (SPE) seraient, par exemple, un moyen rapide et aisé pour la séparation de composés de propriétés chimiques différentes. Le laboratoire disposant d'échantillons de quatre types de ces colonnes, l'objectif sera alors de trouver une cartouche permettant une pré-purification du MO193 par rapport aux autres métabolites présents dans le métabolome.

4 CONCLUSIONS

Les objectifs du travail de recherche auquel je me suis attachée pendant ces trois ans de doctorat étaient doubles :

- mettre en place un protocole de criblage et d'analyse métabolomique de la banque de mutants de la bactérie *Acinetobacter baylyi* ADP1 pour identifier des voies de biosynthèse impliquées dans la synthèse de métabolites orphelins ;
- procéder à l'élucidation structurale de MOs et en particulier de MO211.

Au cours de mon travail de thèse j'ai donc mis en place un protocole haut-débit permettant l'analyse des métabolomes des 2600 souches que compte la banque de mutants de délétion du laboratoire. Cette étape d'optimisation achevée nous avons procédé au criblage de la banque de mutants. Actuellement, nous avons collecté les données sur la moitié des plaques de la banque. Nous avons ensuite procédé au traitement des données brutes par XCMS avec intégration dans la base de donné Business Object. Les premiers résultats obtenus par l'interrogation multifactorielle (DO des cultures, score et Z score des métabolites de référence, stabilité du signal, suivi d'une validation manuelle) de la banque de données a permis de proposer un nombre limité de gènes candidats pour la voie de biosynthèse du MO211. De plus les gènes correspondant aux mutants ne produisant plus le MO211, contrairement à notre attente, ne sont pas co-localisés sur le génome. Ces résultats même s'ils n'étaient pas attendus sont encourageants pour plusieurs raisons qui sont décrites dans le cadre du MO211 mais qui sont applicables aux autres MO. Le nombre limité de mutants sélectionnés ne produisant pas le MO211 nous conforte dans l'idée que l'analyse que nous avons mise en place est « spécifique ».

Il est possible que le criblage de la banque de mutants ne nous donne pas accès à la totalité des gènes impliqués dans la production/consommation du métabolite d'intérêt. Plusieurs raisons peuvent être évoquées :

- certaines activités enzymatiques peuvent être catalysées par plusieurs enzymes (promiscuité de substrat). Il n'est possible d'identifier, par l'analyse des mutants, que les gènes codants pour des activités enzymatiques « uniques ». Cependant, nous souhaitons déterminer les voies de biosynthèses de nombreux métabolites. Il est donc peu probable qu'au moins une fonction impliquée dans la synthèse de chacun de ces MO soit dupliquée ;

- certaines réactions peuvent être non enzymatiques, mais chimiques ;

- enfin, nous pourrions être confrontés à un problème technique que nous n'aurions ni détecté, ni anticipé, entraînant une modification des résultats, ne nous permettant pas de détecter les mutants *bona fide* qui ne produiraient pas l'un des MO.

Il est également possible qu'une seule étape enzymatique permette de synthétiser le MO211 à partir d'un métabolite produit par la bactérie, auquel cas, nous n'obtiendrons une absence du MO211 que pour un

unique mutant.

De plus, nous pourrions ne pas analyser l'ensemble des mutants dans une même condition physiologique (effet pléiotropique de la mutation). Cela pourrait nous amener à détecter une absence d'un MO lors de l'analyse de la banque alors que celui-ci devrait être présent dans le métabolome étudié. Plusieurs raisons peuvent expliquer ce problème :

- chaque mutant diffère génétiquement des autres mutants. Il est donc probable qu'ils aient chacun une cinétique de croissance propre. Hors, nous avons pris le parti, lors de l'extraction des métabolites, de stopper la croissance des souches et déclencher la lyse lorsque la DO_{600} moyenne de la plaque est comprise entre 0,4 et 0,6. Nous aurions pu alors observer massivement des différences de pousses pour certains mutants, provoquant une absence artificielle d'un ou plusieurs MO. Cependant, les cinétiques de croissance des mutants semblent relativement similaires, et l'intégration des contrôles qualité nous permet de repérer rapidement ce genre de problèmes, afin de pouvoir y palier. Dans le cas d'un défaut de pousse, une deuxième analyse suffira à écarter ce problème ;

- il est également possible que lors du traitement des données, XCMS n'intègre pas convenablement les aires du pic correspondant au métabolite. Ce cas sera rapidement écarté car nous analysons manuellement systématiquement les fichiers .RAW des mutants semblant présenter un *fold* ou un Z-score discriminant pour un MO.

Ces hypothèses préliminaires permettent de donner des perspectives pour finaliser ce travail en abordant les points suivants. Nous procéderons dans un premier temps à la validation, avec des moyens d'analyse plus précis, de la présence/absence du MO211 dans les souches mutantes sélectionnées lors du premier criblage. De plus, il n'est pas nécessaire de détecter l'ensemble des gènes impliqués dans la voie de biosynthèse d'un métabolite orphelin. En effet, en n'en validant qu'un ou deux, nous pourrions compléter la voie par des méthodes de génomique comparative en appliquant le principe de « culpabilité par association », c'est-à-dire considérer qu'un gène impliqué dans un même processus physiologique se retrouve souvent physiquement co-localisé sur le génome. Si nous ne trouvons qu'un seul mutant validé qui ne produit pas un métabolite orphelin, nous pourrions alors cribler les mutants dont les gènes excisés sont physiquement voisins dans le génome même s'ils ont été non détectés au court de ce criblage, lors d'une analyse clone par clone de leur métabolome.

Dans un second temps, afin d'extraire un maximum d'informations métaboliques du criblage de la banque, nous pourrons comparer les métabolomes des mutants affectés au métabolome d'une souche sauvage de manière non-ciblée, afin de détecter les différences entre les deux types de métabolomes, pour pouvoir finalement, caractériser complètement l'impact de la délétion du gène sur le mutant.

Enfin, dans le cas favorable ou les gènes de la voie de synthèse d'un ou plusieurs MO sont découverts, il reste à reconstruire la voie de synthèse, c'est-à-dire décrire les réactions enzymatiques et déterminer leur ordre relatif. Cette étape, qui semblait difficile il y a encore quelques temps, a été résolue récemment lors d'un travail sur la détermination de la voie de dégradation d'un composé ubiquitaire issu de la dégradation des végétaux, la trigonelline [26] comme décrit dans l'introduction de cette thèse.

En parallèle je me suis intéressée dans cette thèse à l'élucidation structurale de métabolites orphelins et il est ici intéressant de tirer des conclusions sur l'expérience acquise durant ce travail de thèse. A ce jour, l'élucidation structurale de métabolites *de novo* est un véritable challenge et la combinaison de deux méthodes est couramment utilisée pour effectuer ce travail : la spectrométrie de masse et la RMN.

Nous avons procédé ici à l'élucidation structurale d'un MO : le MO211 par l'utilisation de ces deux méthodes.

Les expériences MS^n séquentielles, donnent accès à une grande quantité d'informations qui facilite l'élucidation structurale de métabolites inconnus. Le principal avantage de cette méthode est d'éviter les étapes de production et de purification des métabolites d'intérêt. Par exemple, en 2017, Zhang, et al. [111] ont réussi à proposer la structure d'un métabolite jusqu'ici inconnu, en procédant uniquement à l'interprétation des fragmentations du métabolite d'une façon semblable à celle que nous avons utilisée pour le MO211. Dans la conclusion, les auteurs insistent sur la difficulté d'aboutir à une détermination structurale avec ce type d'approche. En effet, des essais d'élucidation structurale du MO211 *via* l'analyse des mécanismes de fragmentation en spectrométrie de masse avaient été effectués, mais n'ont pas permis d'aboutir à la structure exacte du composé. L'analyse RMN s'était avérée indispensable.

Nous avons alors mis en place un protocole de production des métabolites, ainsi qu'un protocole de purification adapté au MO211 afin de pouvoir procéder à son analyse en RMN. Cependant, ces deux étapes sont particulièrement longues et fastidieuses et peuvent se révéler problématiques dans le cas de métabolites peu stables ou peu abondants (ce qui n'est pas le cas du MO211). De plus, l'analyse des spectres RMN du composé ne s'est pas avérée suffisante pour conclure à la structure du MO211. Seule une comparaison avec des standards commerciaux nous a permis d'obtenir avec certitude la structure finale. Nous n'avons réussi qu'à dresser une liste de six composés possibles, qu'il a fallu sélectionner par comparaison avec des standards synthétisés pour l'occasion. De plus, la synthèse de tels composés n'est pas toujours chose aisée. Par exemple, l'une des six structures chimiques envisagée n'a pas pu être obtenue.

Au total, l'élucidation structurale du MO211 aura pris plus de quatre ans. L'élucidation

structurale de métabolites inconnus est donc un domaine encore trop peu exploré mais qu'il est pourtant indispensable de découvrir afin de compléter nos connaissances sur le métabolisme.

Dans un premier temps, l'étude des mécanismes de fragmentation de molécules en phase gazeuse est importante, car elle permettra, à l'avenir, de s'affranchir des étapes de production et de purification des métabolites lors du travail d'élucidation structurale. Afin de participer à cet effort, le laboratoire travaille actuellement à l'étude des fragmentations des 6 molécules synthétisées. Le but ici est de déterminer des règles empiriques de fragmentation, pouvant aider à compléter les connaissances que nous avons actuellement sur les mécanismes de fragmentation en phase gazeuse de ce genre de molécules.

De plus, il est important de partager au maximum les données structurales obtenues (fragmentations et informations RMN) afin de compléter au mieux les bases de données de métabolites. Ainsi, ces bases de données pourront permettre l'identification par comparaisons d'un nombre de métabolites de plus en plus grand au sein d'un échantillon métabolique.

Enfin, l'apparition de nouvelles technologies permettra de faciliter et donc de démocratiser l'élucidation structurale de composés. L'apparition de la microscopie à force atomique s'avère prometteuse pour l'élucidation structurale de molécules inconnues [112]. Ce microscope sonde la surface d'un échantillon à l'aide d'une pointe placée au bout d'un levier et d'un système optique, équipé d'un laser, permettant de détecter les déviations de la pointe. Il est donc possible de visualiser des surfaces à des résolutions atomiques.

BIBLIOGRAPHIE

1. Sanger, F., et al., *Nucleotide sequence of bacteriophage φX174 DNA.* Nature, 1977. **265**: p. 687.
2. Fleischmann, R.D., et al., *Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.* Science, 1995. **269**(5223): p. 496.
3. Blattner, F.R., et al., *The Complete Genome Sequence of Escherichia coli K-12.* Science, 1997. **277**(5331): p. 1453.
4. Oliver, S.G., et al., *The complete DNA sequence of yeast chromosome III.* Nature, 1992. **357**: p. 38.
5. Salanoubat, M., et al., *Genome sequence of the plant pathogen Ralstonia solanacearum.* Nature, 2002. **415**: p. 497.
6. European Union Chromosome 3 Arabidopsis Genome Sequencing, C., R. The Institute for Genomic, and D.N.A.R.I. Kazusa, *Sequence and analysis of chromosome 3 of the plant Arabidopsis thaliana.* Nature, 2000. **408**: p. 820.
7. International Human Genome Sequencing, C., *Finishing the euchromatic sequence of the human genome.* Nature, 2004. **431**(7011): p. 931-45.
8. Jain, M., et al., *Nanopore sequencing and assembly of a human genome with ultra-long reads.* Nature Biotechnology, 2018. **36**: p. 338.
9. Radivojac, P., et al., *A large-scale evaluation of computational protein function prediction.* Nat Methods, 2013. **10**(3): p. 221-227.
10. Altschul, S.F., et al., *Basic local alignment search tool.* J Mol Biol, 1990. **215**(3): p. 403-10.
11. Apweiler, R., et al., *UniProt: the Universal Protein knowledgebase.* Nucleic Acids Res, 2004. **32**(Database issue): p. D115-9.
12. Bairoch, A., *PROSITE: a dictionary of sites and patterns in proteins.* Nucleic Acids Res, 1991. **19 Suppl**: p. 2241-5.
13. Sonnhammer, E.L., S.R. Eddy, and R. Durbin, *Pfam: a comprehensive database of protein domain families based on seed alignments.* Proteins, 1997. **28**(3): p. 405-20.
14. Jacob, F., *L'opéron 25 ans après.* Comptes Rendus de l'Académie des Sciences - Series III - Sciences de la Vie, 1997. **320**(3): p. 199-206.
15. Hanson, A.D., et al., *'Unknown' proteins and 'orphan' enzymes: the missing half of the engineering parts list--and how to find it.* Biochem J, 2009. **425**(1): p. 1-11.
16. Overbeek, R., et al., *The use of gene clusters to infer functional coupling.* Proc Natl Acad Sci U S A, 1999. **96**(6): p. 2896-901.
17. Zhao, S., et al., *Discovery of new enzymes and metabolic pathways using structure and genome context.* Nature, 2013. **502**(7473): p. 698-702.
18. de Melo-Minardi, R.C., K. Bastard, and F. Artiguenave, *Identification of subfamily-specific sites based on active sites modeling and clustering.* Bioinformatics, 2010. **26**(24): p. 3075-82.
19. Bastard, K., et al., *Revealing the hidden functional diversity of an enzyme family.* Nat Chem Biol, 2014. **10**(1): p. 42-9.
20. Bastard, K., et al., *Parallel evolution of non-homologous isofunctional enzymes in methionine biosynthesis.* Nat Chem Biol, 2017. **13**(8): p. 858-866.
21. Schnoes, A.M., et al., *Annotation error in public databases: misannotation of molecular function in enzyme superfamilies.* PLoS Comput Biol, 2009. **5**(12): p. e1000605.
22. de Berardinis, V., et al., *A complete collection of single-gene deletion mutants of Acinetobacter baylyi ADP1.* Molecular Systems Biology, 2008. **4**: p. 174-174.
23. Aghaie, A., et al., *New Insights into the Alternative d-Glucarate Degradation Pathway.* The Journal of Biological Chemistry, 2008. **283**(23): p. 15638-15646.
24. Kitagawa, M., et al., *Complete set of ORF clones of Escherichia coli ASKA library (a complete set of E. coli K-12 ORF archive): unique resources for biological research.*

DNA Res, 2005. **12**(5): p. 291-9.
25. Liu, L., et al., *Exploration of novel cellular and serological antigen biomarkers in the ORFeome of Mycobacterium tuberculosis.* Mol Cell Proteomics, 2014. **13**(3): p. 897-906.
26. Perchat, N., et al., *Elucidation of the trigonelline degradation pathway reveals previously undescribed enzymes and metabolites.* Proceedings of the National Academy of Sciences, 2018.
27. Dettmer, K., P.A. Aronov, and B.D. Hammock, *Mass spectrometry-based metabolomics.* Mass Spectrom Rev, 2007. **26**(1): p. 51-78.
28. DeRisi, J.L., V.R. Iyer, and P.O. Brown, *Exploring the Metabolic and Genetic Control of Gene Expression on a Genomic Scale.* Science, 1997. **278**(5338): p. 680.
29. Pandey, A. and M. Mann, *Proteomics to study genes and genomes.* Nature, 2000. **405**(6788): p. 837-46.
30. Dagley, S. and D.E. Nicholson, *An introduction to metabolic pathways*1970, Oxford: Blackwell Scientific Publications.
31. Danchin, A. and A. Sekowska, *The logic of metabolism.* Perspectives in Science, 2015. **6**: p. 15-26.
32. Lespinet, O. and B. Labedan, *Orphan enzymes?* Science, 2005. **307**(5706): p. 42.
33. Karp, P.D., *Call for an enzyme genomics initiative.* Genome Biol, 2004. **5**(8): p. 401.
34. Dunn, W.B., N.J. Bailey, and H.E. Johnson, *Measuring the metabolome: current analytical technologies.* Analyst, 2005. **130**(5): p. 606-25.
35. Guo, A.C., et al., *ECMDB: The E. coli Metabolome Database.* Nucleic Acids Res, 2013. **41**(Database issue): p. D625-D630.
36. Forster, J., et al., *Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network.* Genome Res, 2003. **13**(2): p. 244-53.
37. von Roepenack-Lahaye, E., et al., *Profiling of Arabidopsis Secondary Metabolites by Capillary Liquid Chromatography Coupled to Electrospray Ionization Quadrupole Time-of-Flight Mass Spectrometry.* Plant Physiol, 2004. **134**(2): p. 548-559.
38. Oliver, S.G., et al., *Systematic functional analysis of the yeast genome.* Trends in Biotechnology, 1998. **16**(9): p. 373-378.
39. Horning, E.C. and M.G. Horning, *Metabolic profiles: gas-phase methods for analysis of metabolites.* Clin Chem, 1971. **17**(8): p. 802-9.
40. Fiehn, O., *Metabolomics--the link between genotypes and phenotypes.* Plant Mol Biol, 2002. **48**(1-2): p. 155-71.
41. Ott, M.A. and G. Vriend, *Correcting ligands, metabolites, and pathways.* BMC Bioinformatics, 2006. **7**: p. 517.
42. Raamsdonk, L.M., et al., *A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations.* Nature Biotechnology, 2001. **19**: p. 45.
43. Kind, T., M. Scholz, and O. Fiehn, *How large is the metabolome? A critical analysis of data exchange practices in chemistry.* PLoS One, 2009. **4**(5): p. e5440.
44. Kreimeyer, A., et al., *Identification of the Last Unknown Genes in the Fermentation Pathway of Lysine.* Journal of Biological Chemistry, 2007. **282**(10): p. 7191-7197.
45. Saito, N., et al., *Metabolomics approach for enzyme discovery.* J Proteome Res, 2006. **5**(8): p. 1979-87.
46. Sevin, D.C., et al., *Nontargeted in vitro metabolomics for high-throughput identification of novel enzymes in Escherichia coli.* Nat Methods, 2017. **14**(2): p. 187-194.
47. Tang, Z., M.V. Martin, and F.P. Guengerich, *Elucidation of functions of human cytochrome P450 enzymes: identification of endogenous substrates in tissue extracts using metabolomic and isotopic labeling approaches.* Anal Chem, 2009. **81**(8): p. 3071-8.

48. Sanchez-Ponce, R. and F.P. Guengerich, *Untargeted analysis of mass spectrometry data for elucidation of metabolites and function of enzymes.* Anal Chem, 2007. **79**(9): p. 3355-62.
49. Remy, I. and S.W. Michnick, *Visualization of biochemical networks in living cells.* Proc Natl Acad Sci U S A, 2001. **98**(14): p. 7678-83.
50. Saghatelian, A., et al., *Assignment of endogenous substrates to enzymes by global metabolite profiling.* Biochemistry, 2004. **43**(45): p. 14332-9.
51. Long, J.Z., et al., *Metabolomics annotates ABHD3 as a physiologic regulator of medium-chain phospholipids.* Nat Chem Biol, 2011. **7**(11): p. 763-5.
52. Messerli, G., et al., *Rapid classification of phenotypic mutants of Arabidopsis via metabolite fingerprinting.* Plant Physiol, 2007. **143**(4): p. 1484-92.
53. Baran, R., et al., *Metabolic footprinting of mutant libraries to map metabolite utilization to genotype.* ACS Chem Biol, 2013. **8**(1): p. 189-99.
54. Wolfe, C.J., I.S. Kohane, and A.J. Butte, *Systematic survey reveals general applicability of "guilt-by-association" within gene coexpression networks.* BMC Bioinformatics, 2005. **6**: p. 227.
55. Tohge, T., et al., *Functional genomics by integrated analysis of metabolome and transcriptome of Arabidopsis plants over-expressing an MYB transcription factor.* The Plant Journal, 2005. **42**(2): p. 218-235.
56. Zhang, G.F., et al., *Metabolomics, pathway regulation, and pathway discovery.* J Biol Chem, 2011. **286**(27): p. 23631-5.
57. Nakahigashi, K., et al., *Systematic phenome analysis of Escherichia coli multiple-knockout mutants reveals hidden reactions in central carbon metabolism.* Mol Syst Biol, 2009. **5**: p. 306.
58. Reaves, M.L., et al., *Pyrimidine homeostasis is accomplished by directed overflow metabolism.* Nature, 2013. **500**(7461): p. 237-41.
59. Farag, M.A., et al., *Metabolomics reveals novel pathways and differential mechanistic and elicitor-specific responses in phenylpropanoid and isoflavonoid biosynthesis in Medicago truncatula cell cultures.* Plant Physiol, 2008. **146**(2): p. 387-402.
60. Peyraud, R., et al., *Demonstration of the ethylmalonyl-CoA pathway by using 13C metabolomics.* Proc Natl Acad Sci U S A, 2009. **106**(12): p. 4846-51.
61. Stuani, L., *Nouveaux aspects du métabolisme d'acinetobacter baylyi ADP1 : une approche métabolomique*, 2014.
62. Maharjan, R.P. and T. Ferenci, *Global metabolite analysis: the influence of extraction methodology on metabolome profiles of Escherichia coli.* Anal Biochem, 2003. **313**(1): p. 145-54.
63. Schaefer, U., et al., *Automated sampling device for monitoring intracellular metabolite dynamics.* Anal Biochem, 1999. **270**(1): p. 88-96.
64. Vuckovic, D., *Current trends and challenges in sample preparation for global metabolomics using liquid chromatography-mass spectrometry.* Anal Bioanal Chem, 2012. **403**(6): p. 1523-48.
65. Brauer, M.J., et al., *Conservation of the metabolomic response to starvation across two divergent microbes.* Proc Natl Acad Sci U S A, 2006. **103**(51): p. 19302-7.
66. Canelas, A.B., et al., *Quantitative evaluation of intracellular metabolite extraction techniques for yeast metabolomics.* Anal Chem, 2009. **81**(17): p. 7379-89.
67. Members, M.S.I.B., et al., *The metabolomics standards initiative.* Nat Biotechnol, 2007. **25**(8): p. 846-8.
68. Courant, F., et al., *Basics of mass spectrometry based metabolomics.* Proteomics, 2014. **14**(21-22): p. 2369-88.
69. Markley, J.L., et al., *The future of NMR-based metabolomics.* Curr Opin Biotechnol,

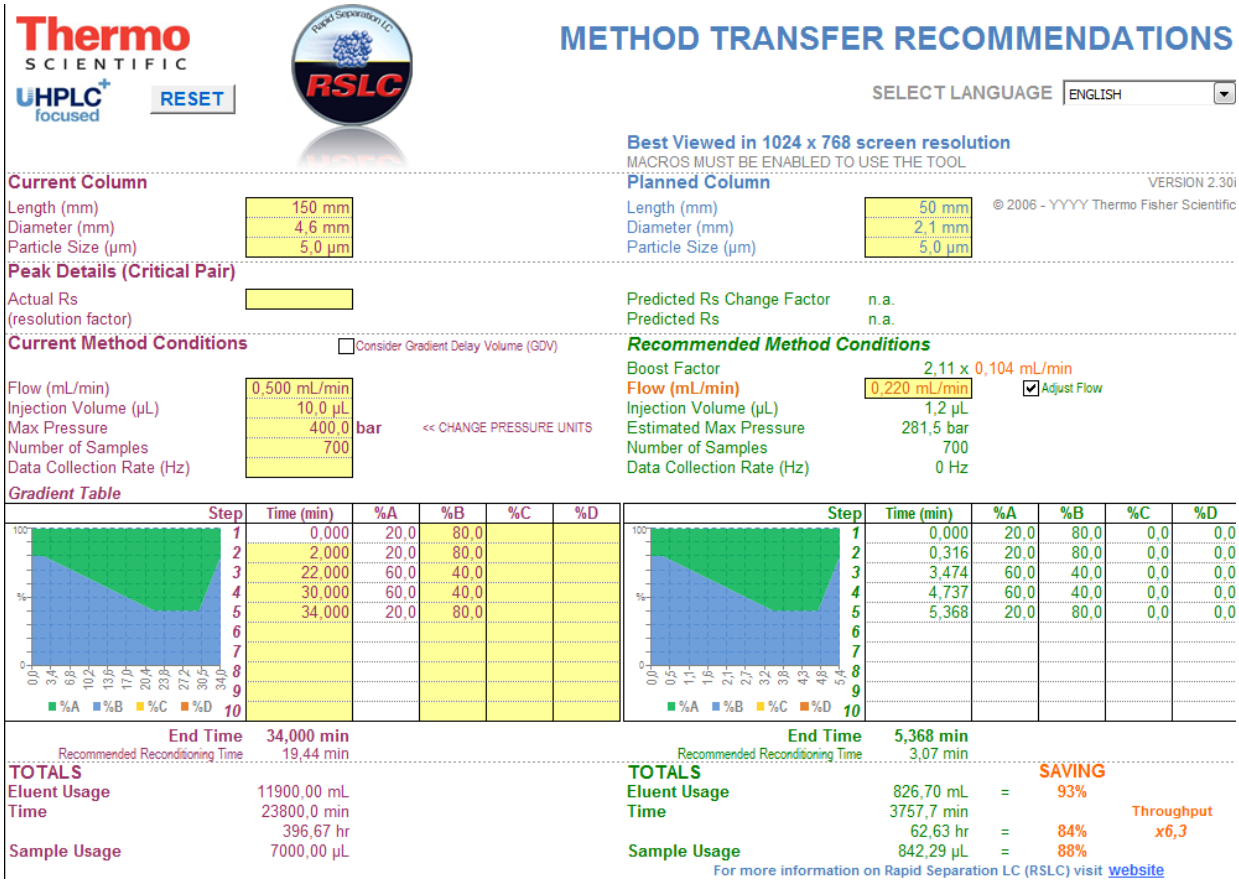
2017. **43**: p. 34-40.
70. Keun, H.C., et al., *Analytical reproducibility in (1)H NMR-based metabonomic urinalysis.* Chem Res Toxicol, 2002. **15**(11): p. 1380-6.
71. Halabalaki, M., et al., *Recent advances and new strategies in the NMR-based identification of natural products.* Curr Opin Biotechnol, 2014. **25**: p. 1-7.
72. Villas-Boas, S.G., et al., *Mass spectrometry in metabolome analysis.* Mass Spectrom Rev, 2005. **24**(5): p. 613-46.
73. Bingol, K. and R. Bruschweiler, *Multidimensional approaches to NMR-based metabolomics.* Anal Chem, 2014. **86**(1): p. 47-57.
74. Boiteau, R.M., et al., *Structure Elucidation of Unknown Metabolites in Metabolomics by Combined NMR and MS/MS Prediction.* Metabolites, 2018. **8**(1).
75. Mortier, K.A., et al., *Adduct formation in quantitative bioanalysis: effect of ionization conditions on paclitaxel.* J Am Soc Mass Spectrom, 2004. **15**(4): p. 585-92.
76. Schug, K. and H.M. McNair, *Adduct formation in electrospray ionization. Part 1: Common acidic pharmaceuticals.* Journal of Separation Science, 2002. **25**(12): p. 760-766.
77. Cooks, R.G., *Special feature: Historical. Collision-induced dissociation: Readings and commentary.* Journal of Mass Spectrometry, 1995. **30**(9): p. 1215-1221.
78. Olsen, J.V., et al., *Higher-energy C-trap dissociation for peptide modification analysis.* Nature Methods, 2007. **4**: p. 709.
79. Makarov, A., *Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis.* Analytical Chemistry, 2000. **72**(6): p. 1156-1162.
80. Price, P., *Standard definitions of terms relating to mass spectrometry : A report from the committee on measurements and standards of the American society for mass spectrometry.* J Am Soc Mass Spectrom, 1991. **2**(4): p. 336-48.
81. Tautenhahn, R., C. Bottcher, and S. Neumann, *Highly sensitive feature detection for high resolution LC/MS.* BMC Bioinformatics, 2008. **9**: p. 504.
82. Tautenhahn, R., et al., *XCMS Online: a web-based platform to process untargeted metabolomic data.* Anal Chem, 2012. **84**(11): p. 5035-9.
83. Kuhl, C., et al., *CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets.* Anal Chem, 2012. **84**(1): p. 283-9.
84. Fahy, E., et al., *LIPID MAPS online tools for lipid research.* Nucleic Acids Res, 2007. **35**(Web Server issue): p. W606-12.
85. Wishart, D.S., et al., *HMDB 3.0--The Human Metabolome Database in 2013.* Nucleic Acids Res, 2013. **41**(Database issue): p. D801-7.
86. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes.* Nucleic Acids Res, 2000. **28**(1): p. 27-30.
87. Smith, C.A., et al., *METLIN: a metabolite mass spectral database.* Ther Drug Monit, 2005. **27**(6): p. 747-51.
88. Wang, Y., et al., *PubChem BioAssay: 2014 update.* Nucleic Acids Res, 2014. **42**(Database issue): p. D1075-82.
89. Milman, B.L., *Towards a full reference library of MS(n) spectra. Testing of a library containing 3126 MS2 spectra of 1743 compounds.* Rapid Commun Mass Spectrom, 2005. **19**(19): p. 2833-9.
90. Fredenhagen, A., C. Derrien, and E. Gassmann, *An MS/MS library on an ion-trap instrument for efficient dereplication of natural products. Different fragmentation patterns for [M + H]+ and [M + Na]+ ions.* J Nat Prod, 2005. **68**(3): p. 385-91.
91. Ichou, F., *Mise en place d'une méthode de calibration pour construire une base de données MS/MS et développement d'un outil pour l'identification en ESI-HR-MS/MS de*

composés organophosphorés, 2013. p. 1 vol. (287 p.).
92. Wolf, S., et al., *In silico fragmentation for computer assisted identification of metabolite mass spectra.* BMC Bioinformatics, 2010. **11**: p. 148.
93. Qin, X.Z., *Tandem mass spectrum of a farnesyl transferase inhibitor--gas-phase rearrangements involving imidazole.* J Mass Spectrom, 2001. **36**(8): p. 911-7.
94. Salek, R.M., K. Haug, and C. Steinbeck, *Dissemination of metabolomics results: role of MetaboLights and COSMOS.* Gigascience, 2013. **2**(1): p. 8.
95. Steinbeck, C., et al., *MetaboLights: towards a new COSMOS of metabolomics data management.* Metabolomics, 2012. **8**(5): p. 757-760.
96. Creek, D., et al., *Metabolite identification: Are you sure? And how do your peers gauge your confidence?* Vol. 10. 2014.
97. Barbe, V., et al., *Unique features revealed by the genome sequence of Acinetobacter sp. ADP1, a versatile and naturally transformation competent bacterium.* Nucleic Acids Res, 2004. **32**(19): p. 5766-5779.
98. Young, D.M., D. Parke, and L.N. Ornston, *Opportunities for genetic investigation afforded by Acinetobacter baylyi, a nutritionally versatile bacterial species that is highly competent for natural transformation.* Annu Rev Microbiol, 2005. **59**: p. 519-51.
99. de Berardinis, V., et al., *A complete collection of single-gene deletion mutants of Acinetobacter baylyi ADP1.* Mol Syst Biol, 2008. **4**: p. 174.
100. Stuani, L., et al., *Novel metabolic features in ADP1 revealed by a multiomics approach.* Metabolomics, 2014. **10**(6): p. 1223-1238.
101. Michiel, M., et al., *Microbial urate catabolism: characterization of HpyO, a non-homologous isofunctional isoform of the flavoprotein urate hydroxylase HpxO.* Environ Microbiol Rep, 2012. **4**(6): p. 642-7.
102. Durot, M., et al., *Iterative reconstruction of a global metabolic model of Acinetobacter baylyi ADP1 using high-throughput growth phenotype and gene essentiality data.* BMC Syst Biol, 2008. **2**: p. 85.
103. de Berardinis, V., et al., *Acinetobacter baylyi ADP1 as a model for metabolic system biology.* Current Opinion in Microbiology, 2009. **12**(5): p. 568-576.
104. Stuani, L., et al., *Novel metabolic features in Acinetobacter baylyi ADP1 revealed by a multiomics approach.* Metabolomics, 2014. **10**(6): p. 1223-1238.
105. Kind, T. and O. Fiehn, *Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry.* BMC Bioinformatics, 2007. **8**: p. 105.
106. *Metabolomics Service Protocol, Sample Preparation for Cells.* Glasgow Polyomics, 2013.
107. Dörries, K., R. Schlueter, and M. Lalk, *Impact of Antibiotics with Various Target Sites on the Metabolome of Staphylococcus aureus.* Antimicrobial Agents and Chemotherapy, 2014. **58**(12): p. 7151-7163.
108. Cusson, A., *Mise en place de protocoles pour l'analyse systématique de la banque de mutants de la bactérie Acinetobacter baylyi ADP1 par une approche métabolomique*, 2015, CPE Lyon.
109. Martin, G.E. and C.E. Hadden, *Long-Range 1H−15N Heteronuclear Shift Correlation at Natural Abundance.* Journal of Natural Products, 2000. **63**(4): p. 543-585.
110. Martin, G. and A. Williams, *Utilizing Long-Range 1H-15N 2D NMR Spectroscopy in Chemical Structure Elucidation*2008.
111. Zhang, Q., et al., *Structure elucidation of metabolite x17299 by interpretation of mass spectrometric data.* Metabolomics, 2017. **13**(8): p. 92.
112. Gross, L., et al., *Atomic Force Microscopy for Molecular Structure Elucidation.* Angewandte Chemie International Edition, 2018. **57**(15): p. 3888-3908.

ANNEXES

ANNEXE 1: Convertisseurs de méthode chromatographique utilisés lors du développement de méthodes analytique du criblage de la banque, ainsi que lors des conversions de méthodes HPLC analytique/HPLC préparative lors de la purification des MO.

Disponible ici : https://www.thermofisher.com/content/dam/tfs/ATG/CMD/cmd-documents/oper/tech/chrom/ lc/sys/RSLC-Method-Transfer-Tool-2-3.xlsm



Thermo SCIENTIFIC
UHPLC+ focused
RESET

METHOD TRANSFER RECOMMENDATIONS

SELECT LANGUAGE ENGLISH

Best Viewed in 1024 x 768 screen resolution
MACROS MUST BE ENABLED TO USE THE TOOL

VERSION 2.30i
© 2006 - YYYY Thermo Fisher Scientific

Current Column

Length (mm)	150 mm
Diameter (mm)	4,6 mm
Particle Size (µm)	5,0 µm

Planned Column

Length (mm)	50 mm
Diameter (mm)	2,1 mm
Particle Size (µm)	5,0 µm

Peak Details (Critical Pair)

Actual Rs (resolution factor)	
Predicted Rs Change Factor	n.a.
Predicted Rs	n.a.

Current Method Conditions — Consider Gradient Delay Volume (GDV)

Flow (mL/min)	0,500 mL/min	
Injection Volume (µL)	10,0 µL	
Max Pressure	400,0 bar	<< CHANGE PRESSURE UNITS
Number of Samples	700	
Data Collection Rate (Hz)		

Recommended Method Conditions

Boost Factor	2,11 x 0,104 mL/min
Flow (mL/min)	0,220 mL/min — Adjust Flow
Injection Volume (µL)	1,2 µL
Estimated Max Pressure	281,5 bar
Number of Samples	700
Data Collection Rate (Hz)	0 Hz

Gradient Table

Step	Time (min)	%A	%B	%C	%D
1	0,000	20,0	80,0		
2	2,000	20,0	80,0		
3	22,000	60,0	40,0		
4	30,000	60,0	40,0		
5	34,000	20,0	80,0		
6					
7					
8					
9					
10					

Step	Time (min)	%A	%B	%C	%D
1	0,000	20,0	80,0	0,0	0,0
2	0,316	20,0	80,0	0,0	0,0
3	3,474	60,0	40,0	0,0	0,0
4	4,737	60,0	40,0	0,0	0,0
5	5,368	20,0	80,0	0,0	0,0
6					
7					
8					
9					
10					

End Time	34,000 min	End Time	5,368 min	
Recommended Reconditioning Time	19,44 min	Recommended Reconditioning Time	3,07 min	
TOTALS		**TOTALS**		**SAVING**
Eluent Usage	11900,00 mL	Eluent Usage	826,70 mL =	93%
Time	23800,0 min	Time	3757,7 min	Throughput
	396,67 hr		62,63 hr =	84% x6,3
Sample Usage	7000,00 µL	Sample Usage	842,29 µL =	88%

For more information on Rapid Separation LC (RSLC) visit website

Disponible ici : https://www.sigmaaldrich.com/analytical-chromatography/hplc/method-transfer-calculator.html#gradient

SUPELCO Analytical — HPLC Method Transfer Calculator

Isocratic METHOD | Gradient METHOD | Technical SUPPORT

Reset

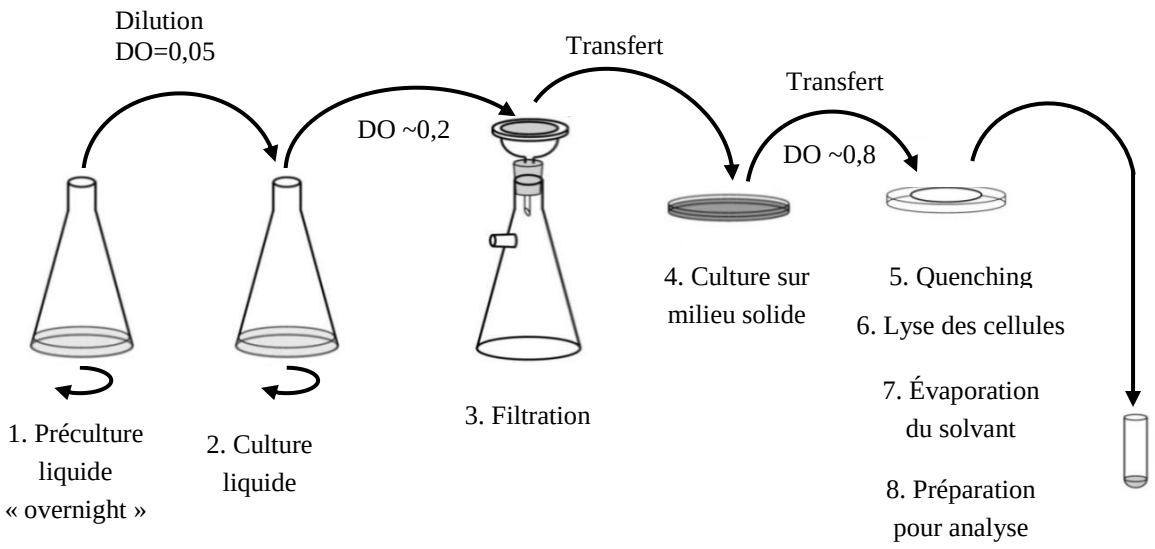
	CURRENT METHOD	NEW METHOD	NEW METHOD ADJUST FLOW RATE
Column Length (cm)	15	5	5
Column I.D. (mm)	4,6	2,1	2.1
Particle Size (µm)	5	5	5
Flow Rate (mL/min)	0,50	0.10	0,20
Injection Volume (µL)	10	0.7	0.7
Pressure (psi/bar)	400	133.3	255.9
Run Time (min)	50,00	16.67	8.68
Equilibration Time (min)	34.88	11.63	6.06
TIME SAVED (MIN)		33.3	41.3
SOLVENT SAVED (ML)		23.3	23.3

GRADIENT CONDITIONS	CURRENT METHOD	NEW METHOD	NEW METHOD ADJUST DWELL VOLUME
Dwell Volume (mL)	0	0,00	0,00
Gradient Delay (min)		0.00	0.00

GRADIENT TABLE

TIME	% A	% B	NEW TIME	NEW TIME ADJUSTED FLOW
0,00	80	20	0.00	0.00
2,00	80	20	0.40	0.46
22,00	80	20	7.30	3.82
30,00	80	20	10.00	5.21
34,00	20	80	11.30	5.91
50	20	80	16.70	8.68

ANNEXE 2: Protocole de préparation d'un métabolome d'ADP1 selon le protocole d'extraction utilisé précédemment par le laboratoire[104] . Issu de [61]



La préparation du métabolome est adaptée de Brauer *et al.* [65] et peut se décomposer en 8 étapes.

1. Préculture liquide 'overnight' : 5 mL de milieu minimum Acinetobacter (MA), contenant la source de carbone souhaitée à 0,02 M, sont ensemencés avec une colonie d'ADP1 dans un tube 50 mL.

2. Culture liquide : la préculture est diluée dans 20 mL du même milieu minimum à une densité optique (DO) à 600 nm de 0,05 dans un tube 50 mL. La culture est incubée à 30°C jusqu'à DO = 0,2 (~2h) sous agitation.

3. Filtration : 5 mL de cette culture sont filtrés sur un filtre en PTFE (0,45 µm) de 47 mm de diamètre (Millipore Omnipore JHWP 4700). Les bactéries restent sur le filtre et le milieu de culture est éliminé.

4. Culture sur milieu solide : Le filtre est positionné sur une gélose d'agarose (1,5%) contenant le même milieu de culture avec les bactéries situées sur le haut du filtre. La gélose est incubée à 30°C jusqu'à obtenir une DO_{600} de l'ordre de 0,8 (phase de croissance exponentielle ~2h). Pour mesurer la DO des cultures, un filtre ayant subi le même protocole jusqu'alors sera immergé dans 5 mL d'une solution de milieu (quinate ou succinate) puis passé aux ultrasons pour détacher les bactéries ; la DO est ensuite mesurée.

5. « Quenching » : le métabolisme est stoppé en immergeant rapidement le filtre (bactéries positionnées vers le bas) dans une cupule en verre préalablement refroidie et remplie avec 5 mL d'un mélange ACN/MeOH 80 : 20 dans un bain à -30°C (glace, carboglace, éthanol). Après 15 minutes à -30°C, la cupule est placée dans un bain liquide à ultrasons à 4°C pendant 5 minutes pour détacher les bactéries du filtre. La phase liquide contenant les bactéries est transférée dans des cryotubes (Simport T3082A).

6. Lyse des cellules : 6 cycles de congélation/décongélation dans de l'azote liquide et dans un bain-marie à 65°C permettent de casser complètement les cellules et d'extraire les métabolites.

7. Évaporation des solvants : 1h30 de 'SpeedVac™' (Eppendorf Concentrator 5301) à 30°C est nécessaire pour évaporer complètement les solvants (fonction 1 adaptée aux solutions alcooliques).

8. Préparation analyse LC/MS : Le culot est repris dans 300 µL d'eau puis soumis aux ultrasons pendant 10 minutes, afin de favoriser l'extraction. Les débris sont éliminés après centrifugation (20000 g, 4°C, 10 min) et le surnageant est filtré (0,22 µm (Millipore Millex-GV 13mm)). 700 µL de phase mobile LC/MS sont utilisés pour rincer le filtre et ajoutés à l'échantillon. L'échantillon final est stocké à 4°C et injecté sous 24h.

ANNEXE 3: Liste des 141 métabolites inconnus dont le fold Q/S est supérieur à 5. En bleu, métabolites ne possédant aucune proposition d'identification dans les bases de données (KEGG et METLIN).

$$Fold\ Q/S\ = \frac{abondance\ moyenne\ du\ métabolite\ dans\ les\ "métabolomes\ quinate"}{abondance\ moyenne\ du\ métabolite\ dans\ les\ "métabolomes\ succinate"}$$

Données issues de [104]

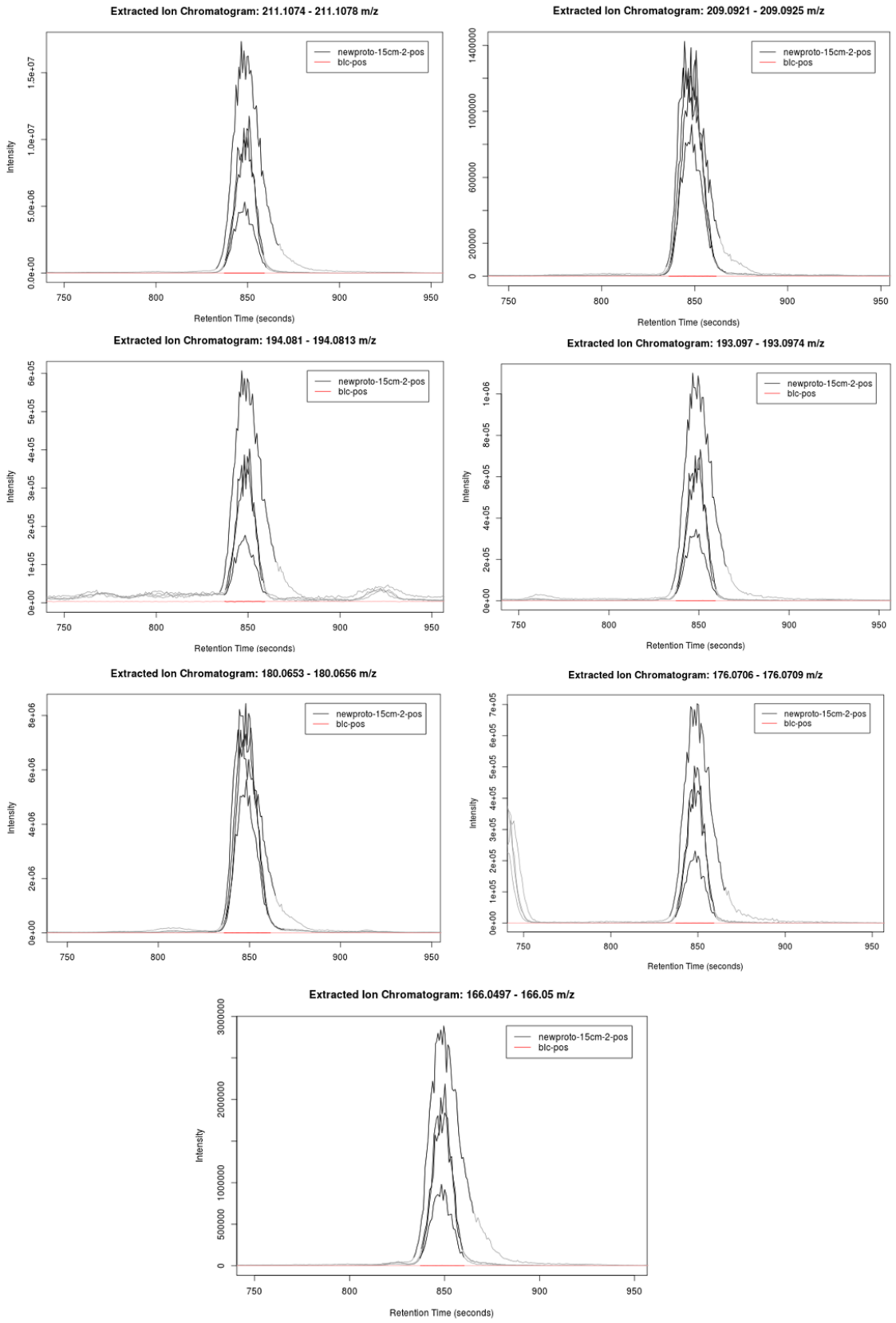
masse détectée	mode d'ionisation de détection préférentiel	masse neutre	temps de rétention sur ZIC-*p*HILIC 150 x 4,6 mm², 5 µm	Fold Q/S
157,13328	+	156,12600	5,89	5,08
256,11714	+	255,10986	5,15	5,09
349,05373	+	348,04645	12,58	5,12
173,00954	-	174,01682	16,70	5,19
198,07555	+	197,06827	8,45	5,28
203,13844	+	202,13116	12,08	5,28
284,06912	+	283,06184	4,86	5,32
110,98575	-	111,99303	11,20	5,38
219,13329	+	218,12601	12,43	5,39
267,03567	-	268,04295	7,56	5,44
233,11265	+	232,10537	10,97	5,53
173,09168	+	172,08440	9,47	5,54
235,12809	+	234,12081	11,75	5,77
231,04492	-	232,05220	3,23	6,26
155,11760	+	154,11032	3,71	6,28
127,05005	+	126,04277	7,60	6,30
85,07598	+	84,06870	26,29	6,87
170,04623	-	171,05351	5,04	6,99
245,04368	-	246,05096	9,22	7,21
873,33170	+	872,32442	11,66	7,29
199,10700	+	198,09972	21,39	7,36
144,98181	+	143,97453	9,45	7,47
851,34869	+	850,34141	11,59	7,67
307,09492	+	306,08764	5,87	7,72
219,13317	+	218,12589	23,17	7,74
452,27634	+	451,26906	2,72	8,43
451,26810	-	452,27538	2,86	8,59
216,06241	+	215,05513	12,12	9,00
263,06877	+	262,06149	9,79	9,15

159,11254	+	158,10526	22,88	9,24
448,12473	-	449,13201	6,00	9,34
130,09729	+	129,09001	7,54	9,53
129,06569	+	128,05841	5,99	9,80
169,09686	+	168,08958	8,69	11,68
201,12271	+	200,11543	23,54	11,73
199,10708	+	198,09980	5,08	12,30
294,04082	-	295,04810	3,28	12,88
316,99266	-	317,99994	15,73	13,40
157,09686	+	156,08958	8,43	13,41
500,97639	-	501,98367	14,53	13,95
217,11748	+	216,11020	22,74	14,30
165,05423	+	164,04695	9,45	14,94
177,02314	-	178,03042	10,55	15,25
311,02818	-	312,03546	5,91	15,63
215,01436	-	216,02164	8,52	16,55
201,12272	+	200,11544	21,65	16,71
165,10198	+	164,09470	15,71	16,95
103,00414	-	104,01142	12,60	17,46
217,01034	-	218,01762	5,93	18,40
175,02527	-	176,03255	15,89	20,52
379,16999	+	378,16271	21,14	23,27
169,09686	+	168,08958	22,44	23,61
381,18591	+	380,17863	14,14	24,83
103,95546	+	102,94818	9,46	25,01
135,09146	+	134,08418	13,50	26,28
202,12600	+	201,11872	23,57	32,11
131,07012	+	130,06284	4,92	33,95
187,02521	-	188,03249	11,69	42,88
178,98371	-	179,99099	15,01	42,99
335,11664	-	336,12392	7,57	43,84
685,26428	+	684,25700	19,70	47,22
251,10184	+	250,09456	5,22	58,79
115,04052	-	116,04780	11,98	59,21
232,12840	+	231,12112	6,00	68,83
262,09132	+	261,08404	14,83	70,46
137,10710	+	136,09982	16,33	75,25
221,09127	+	220,08399	10,13	77,83
850,34897	-	851,35625	12,07	86,08

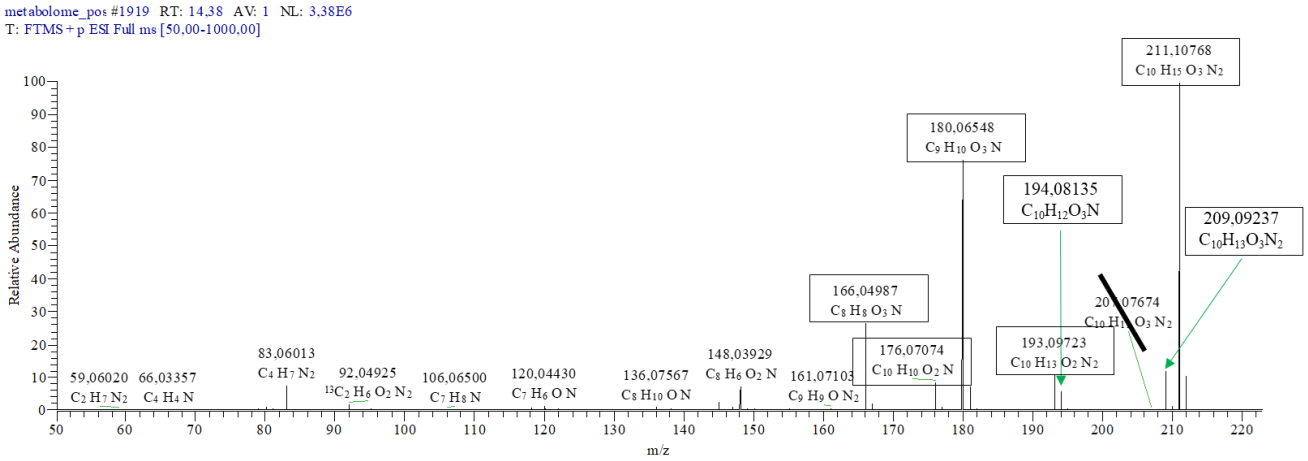
115,04048	-	116,04776	3,86	87,77
829,30642	+	828,29914	18,31	90,24
204,08599	+	203,07871	8,56	91,89
245,11241	+	244,10513	4,98	95,42
208,08359	+	207,07631	5,91	98,16
232,12841	+	231,12113	18,78	100,65
216,98128	-	217,98856	16,50	104,71
427,15006	+	426,14278	12,15	106,13
300,07362	-	301,08090	11,76	108,75
275,21101	+	274,20373	3,14	110,49
213,01736	-	214,02464	8,50	114,19
204,04306	+	203,03578	5,42	136,20
227,10183	+	226,09455	4,03	137,67
523,21236	+	522,20508	18,75	140,46
245,11231	+	244,10503	9,70	144,25
275,10175	+	274,09447	12,22	155,11
251,10173	+	250,09445	9,81	176,80
405,01160	-	406,01888	3,26	183,26
343,08585	-	344,09313	10,75	184,21
211,10702	+	210,09974	3,32	186,69
242,11266	+	241,10538	10,61	248,01
265,11749	+	264,11021	3,96	255,47
331,06194	-	332,06922	11,80	273,15
302,11264	+	301,10536	3,31	276,42
211,10698	+	210,09970	10,40	320,48
283,09159	+	282,08431	14,97	388,66
229,11754	+	228,11026	12,47	457,64
353,13340	+	352,12612	8,32	475,86
265,11739	+	264,11011	8,74	555,17
866,47907	+	865,47179	2,60	677,86
187,02516	-	188,03244	12,54	686,55
166,99898	-	168,00626	14,10	818,74
153,57015	-	154,57743	9,69	911,53
339,15401	+	338,14673	9,65	913,53
265,11742	+	264,11014	8,31	1106,51
216,01528	-	217,02256	10,06	1285,59
189,04079	-	190,04807	5,60	1481,42
313,08109	+	312,07381	2,87	2269,21
167,11762	+	166,11034	5,13	2753,89

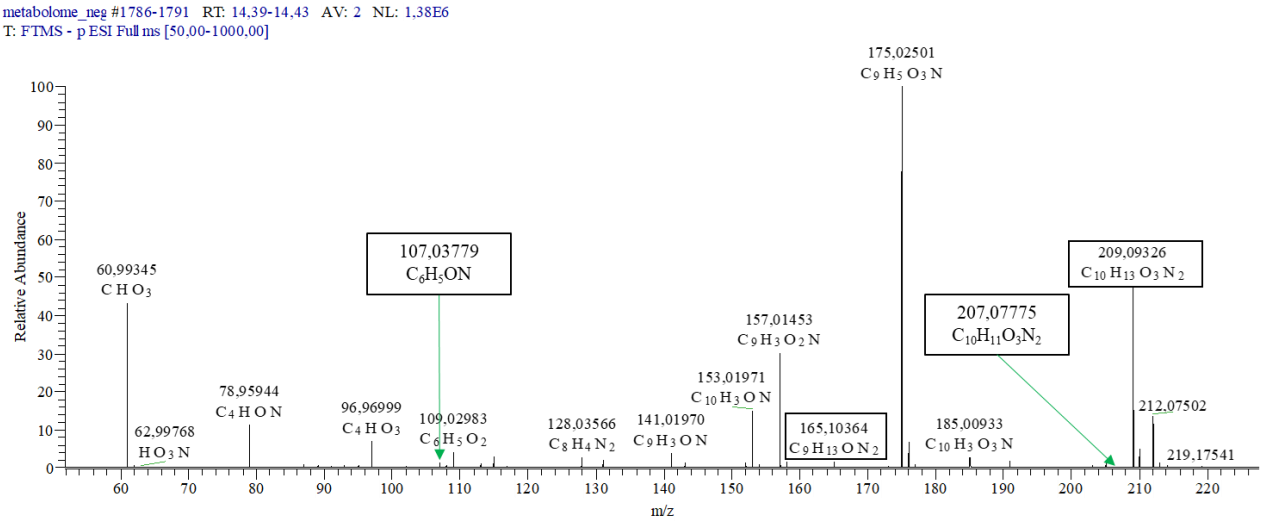
345,08440	-	346,09168	9,61	3540,82
333,07420	+	332,06692	8,54	4074,69
379,19529	-	380,20257	2,60	4107,32
203,02005	-	204,02733	16,88	10217,62
353,13336	+	352,12608	5,34	10769,86
314,03609	-	315,04337	9,71	15437,17
103,01909	-	104,02637	12,68	inf
129,01973	-	130,02701	12,43	inf
142,08608	+	141,07880	12,07	inf
152,03586	-	153,04314	9,46	inf
193,09664	+	192,08936	4,61	inf
203,02009	-	204,02737	13,82	inf
211,10704	+	210,09976	14,48	inf
214,11778	+	213,11050	23,21	inf
219,01481	-	220,02209	17,00	inf
243,01509	-	244,02237	15,00	inf
274,08144	+	273,07416	3,03	inf
299,17047	+	298,16319	5,99	inf
309,01955	-	310,02683	12,74	inf
315,09649	+	314,08921	8,70	inf
317,11233	+	316,10505	7,48	inf
318,16502	+	317,15774	4,80	inf
330,10747	+	329,10019	7,07	inf
333,07416	+	332,06688	9,58	inf
334,12044	-	335,12772	7,61	inf
337,00328	-	338,01056	17,13	inf
345,05906	-	346,06634	5,85	inf
349,00375	-	350,01103	9,71	inf
367,11307	+	366,10579	13,54	inf
403,16026	+	402,15298	14,83	inf
447,02077	-	448,02805	12,67	inf
659,10989	+	658,10261	8,24	inf
716,26067	+	715,25339	17,07	inf
852,53706	+	851,52978	2,08	inf

ANNEXE 4: XIC générés par XCMS des différents m/z semblant provenir du MO211. En noir, injections de métabolomes sur la colonne de 15 cm, en rouge, injections de phase mobile (analyse en mode d'ionisation positif).

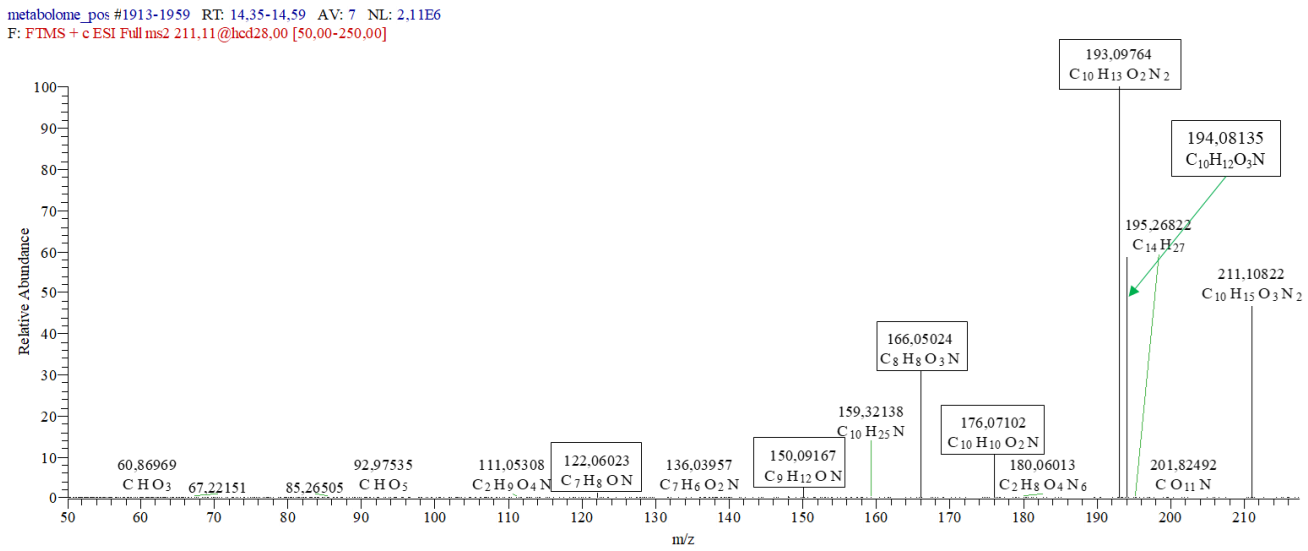


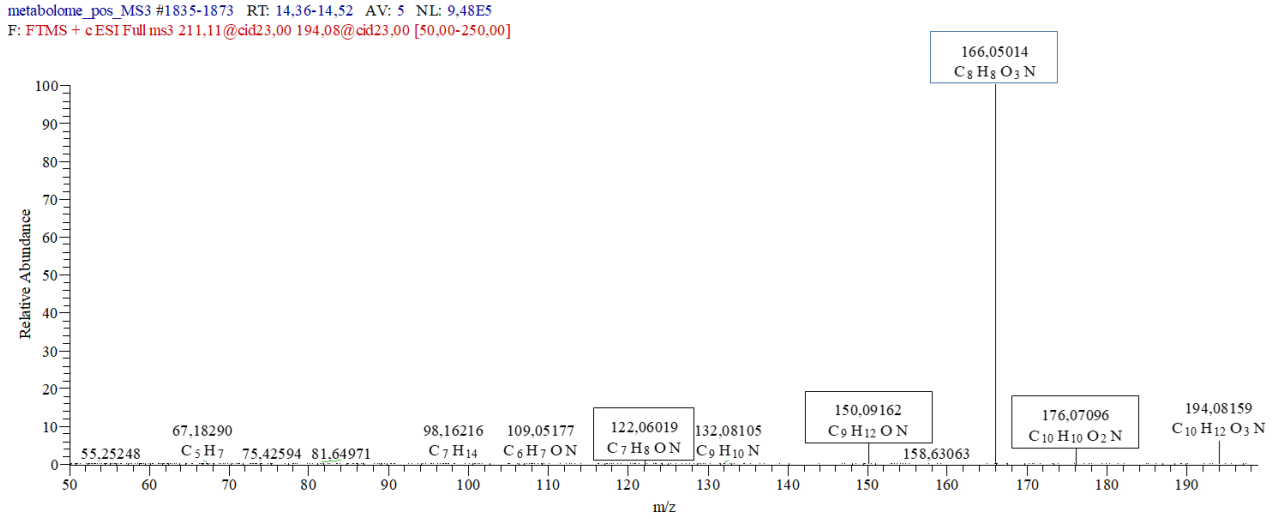
ANNEXE 5 : Spectre de masse MS avec détection FTMS du MO211 en mode d'ionisation positif (haut) et négatif (bas)

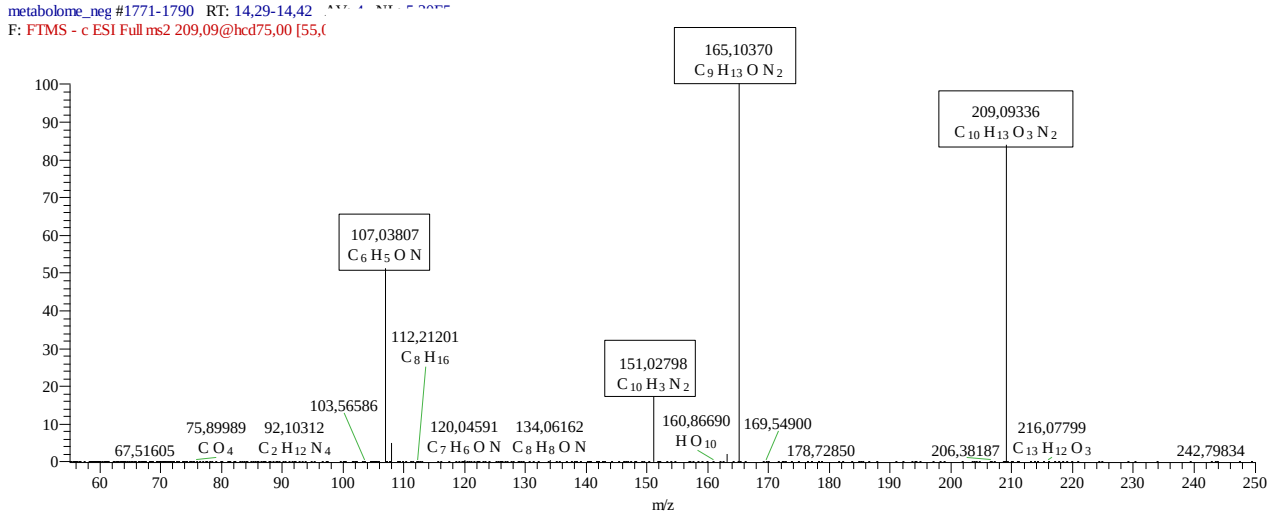




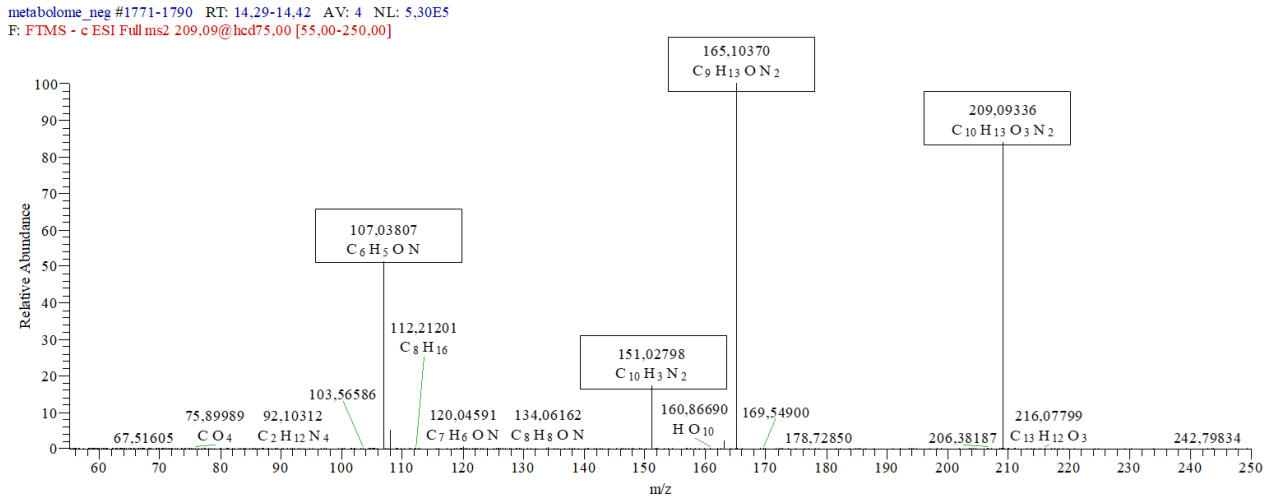
ANNEXE 6 : Spectres MS^n avec détection haute résolution du MO211 en mode d'ionisation positif

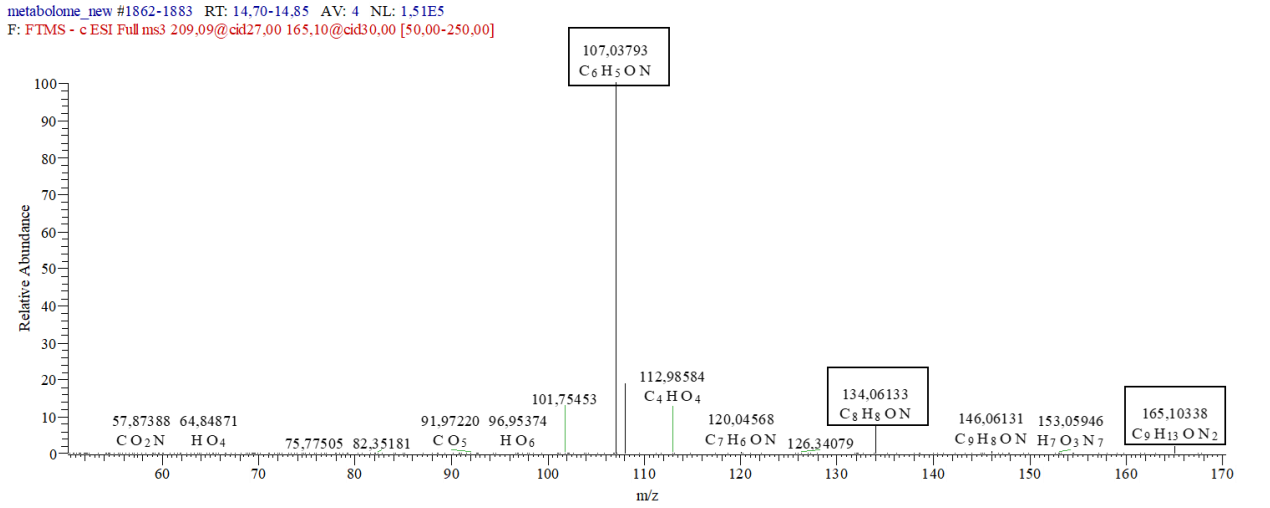




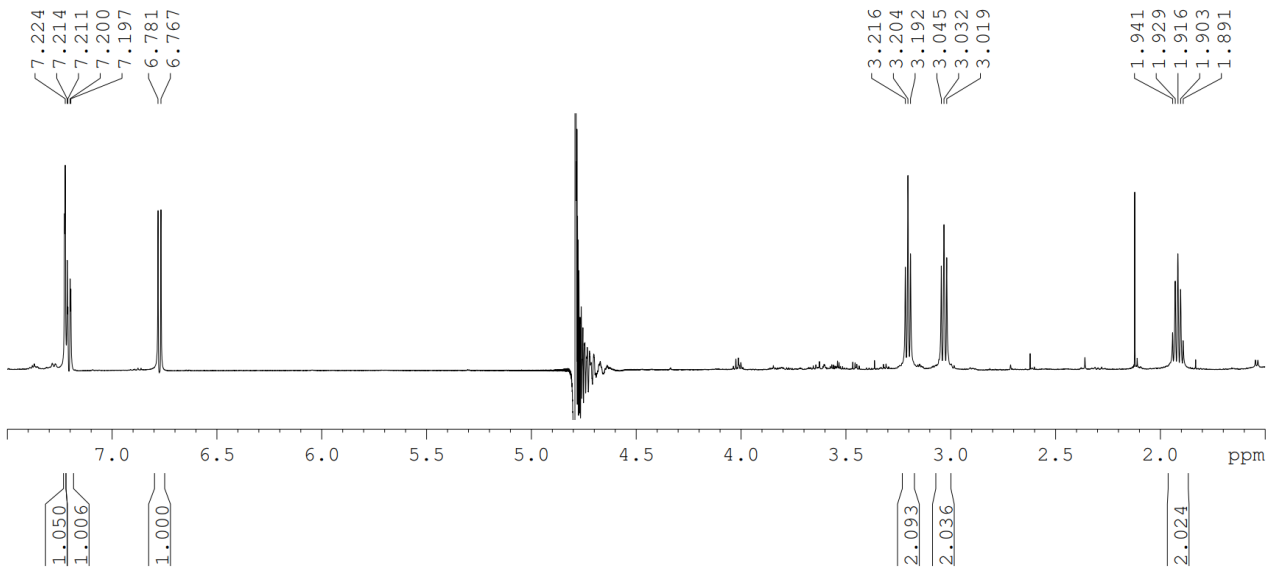


ANNEXE 7: Spectres MS^n avec détection haute résolution de l'ion m/z 209,09336 en mode d'ionisation négatif

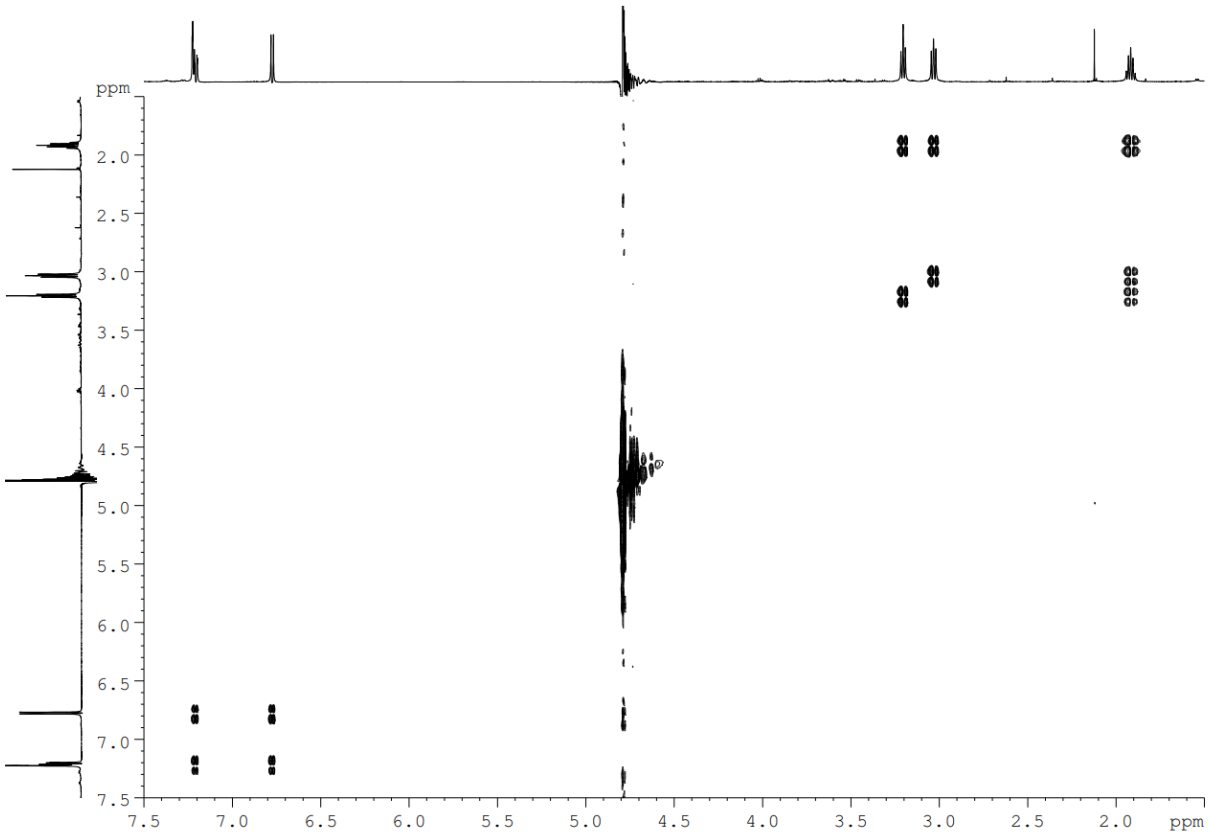


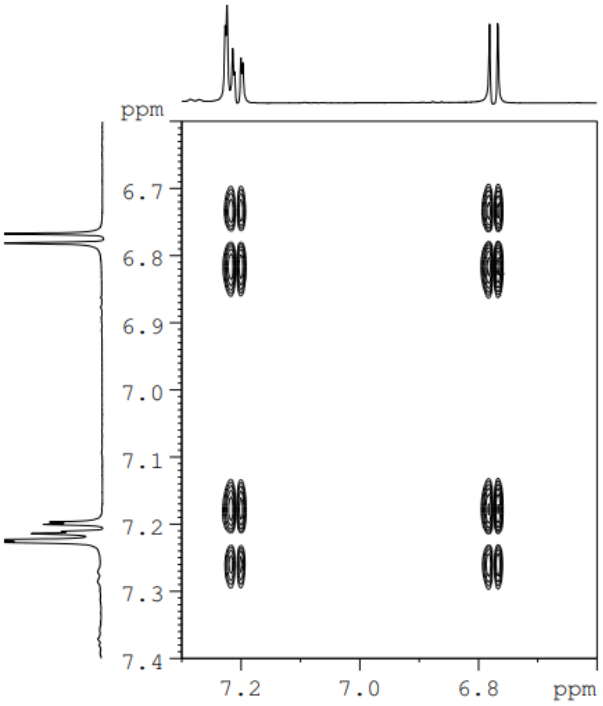


ANNEXE 8: Spectre RMN ^{1}H du MO211 solvant H_2O/D_2O (90:10), 600 MHz).

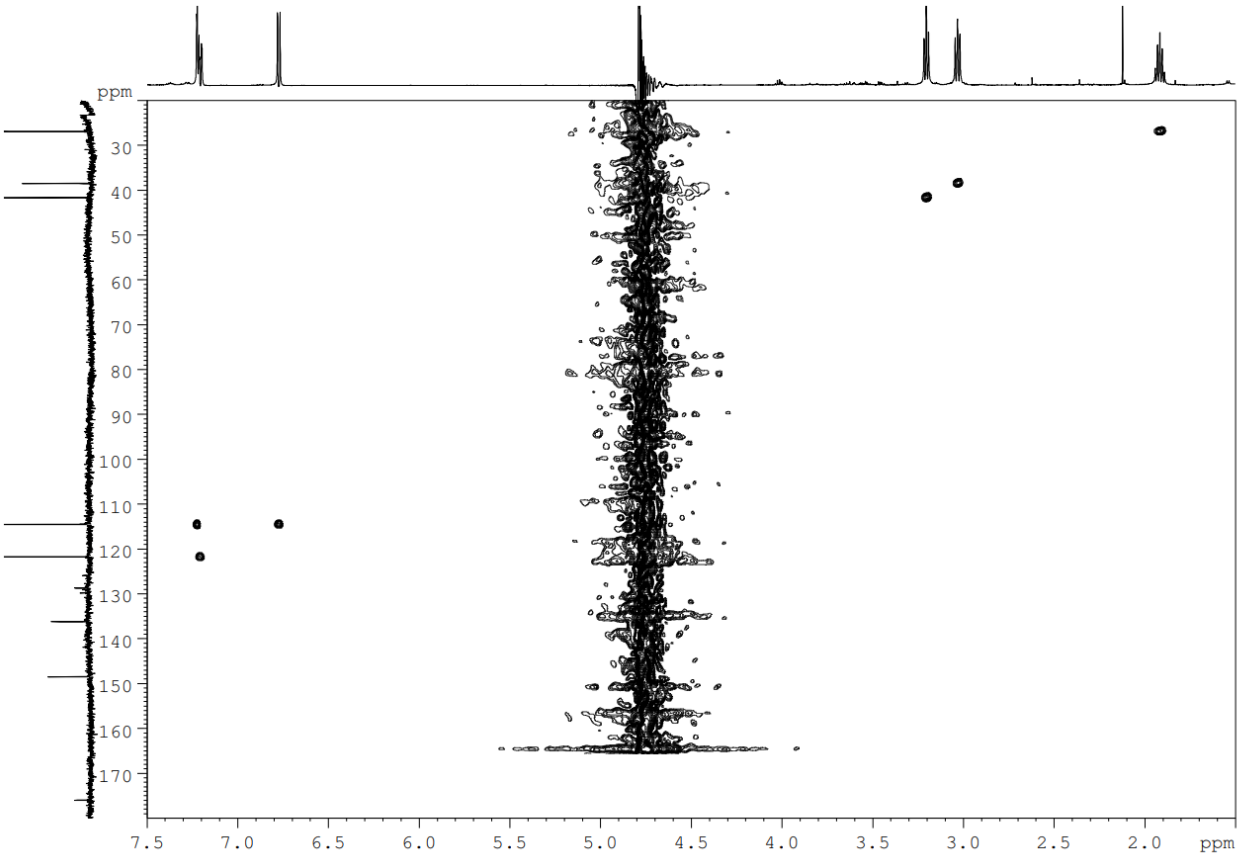


ANNEXE 9: Spectre COSY ^{1}H-^{1}H obtenu du MO211. En haut, spectre complet, en bas, détail de la zone des protons aromatiques (entre 6,6 et 7,4 ppm) ; solvant H_2O/D_2O (90:10)

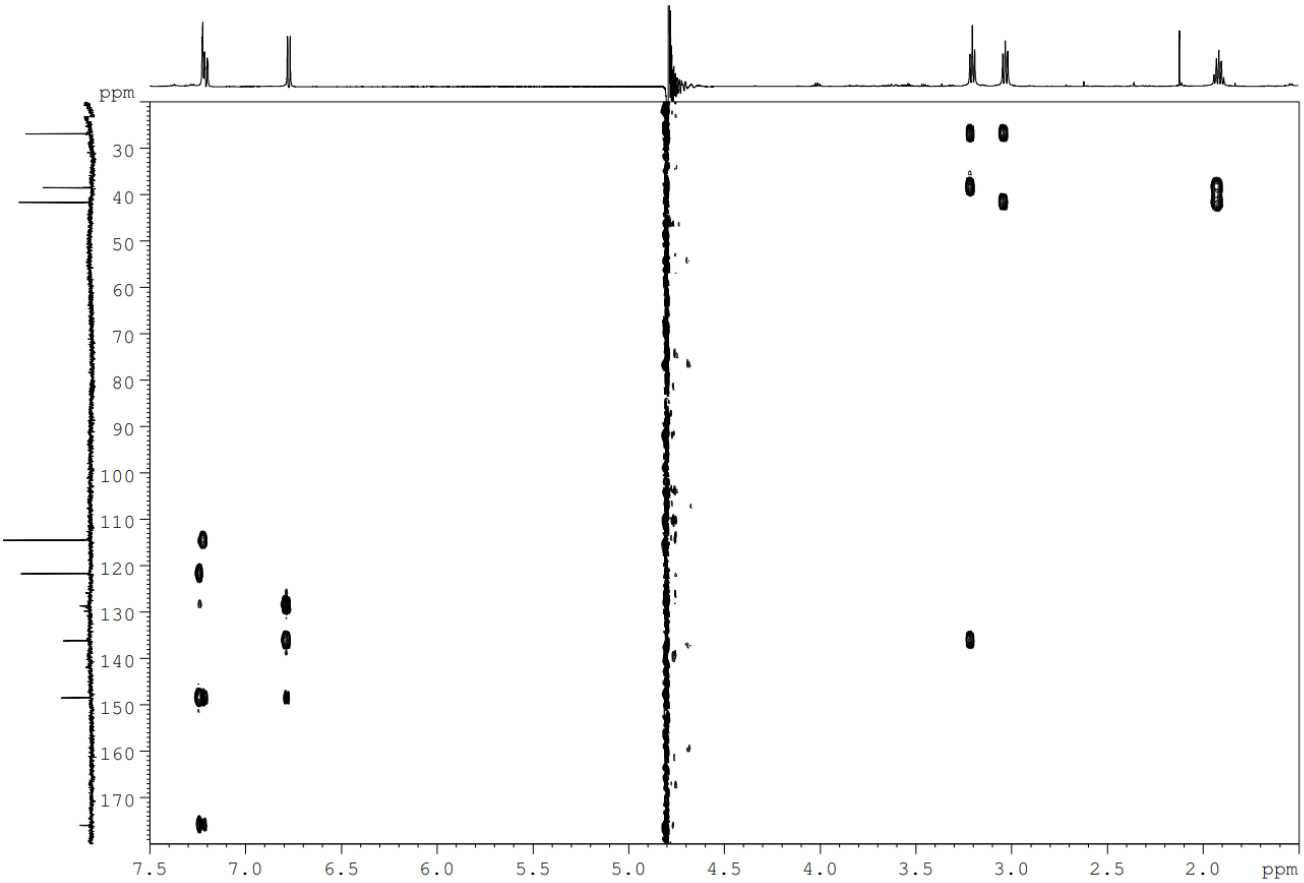




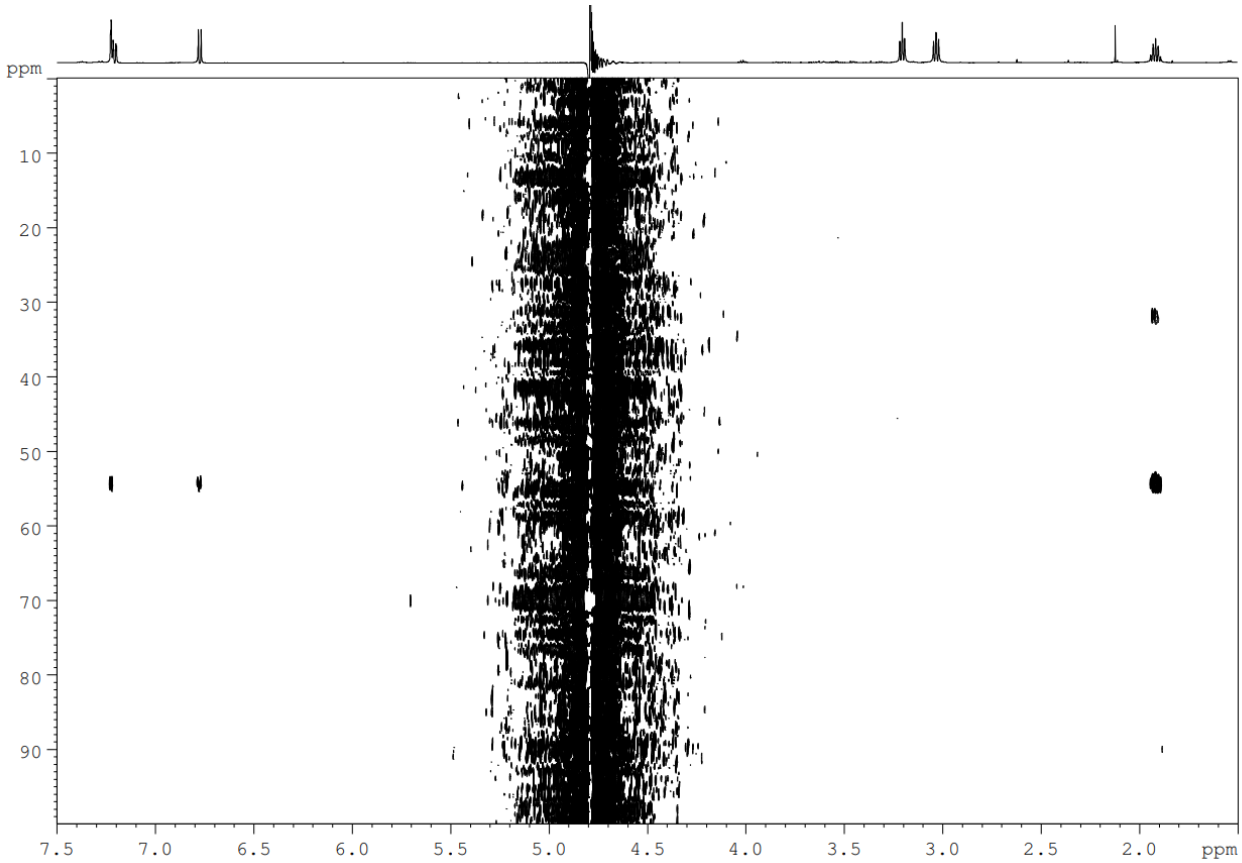
ANNEXE 10 : Spectre RMN HSQC ^{1}H-^{13}C du MO211 ; solvant H_2O/D_2O (90:10)



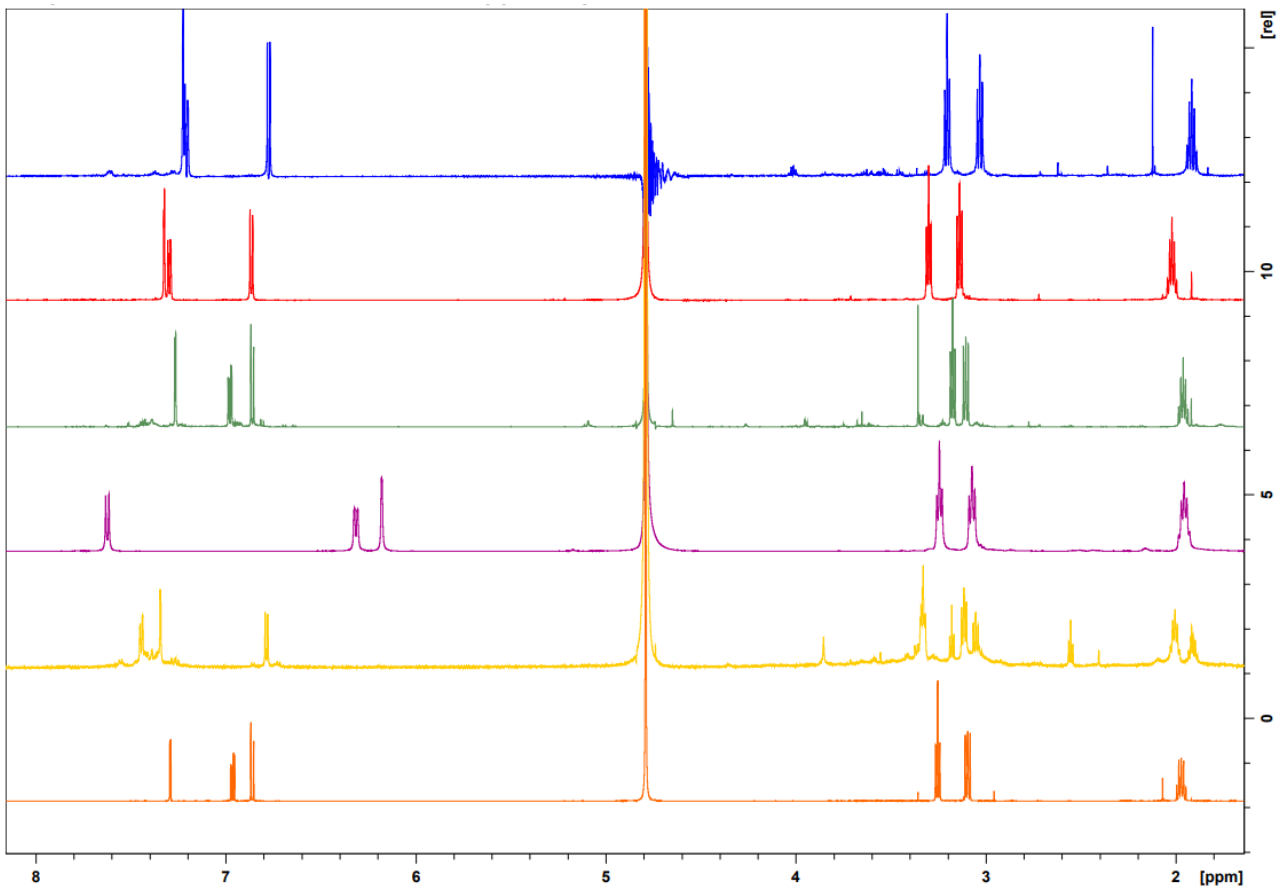
ANNEXE 11 : Spectre RMN HMBC ^{1}H-^{13}C du MO211 ; solvant H_2O/D_2O (90:10)



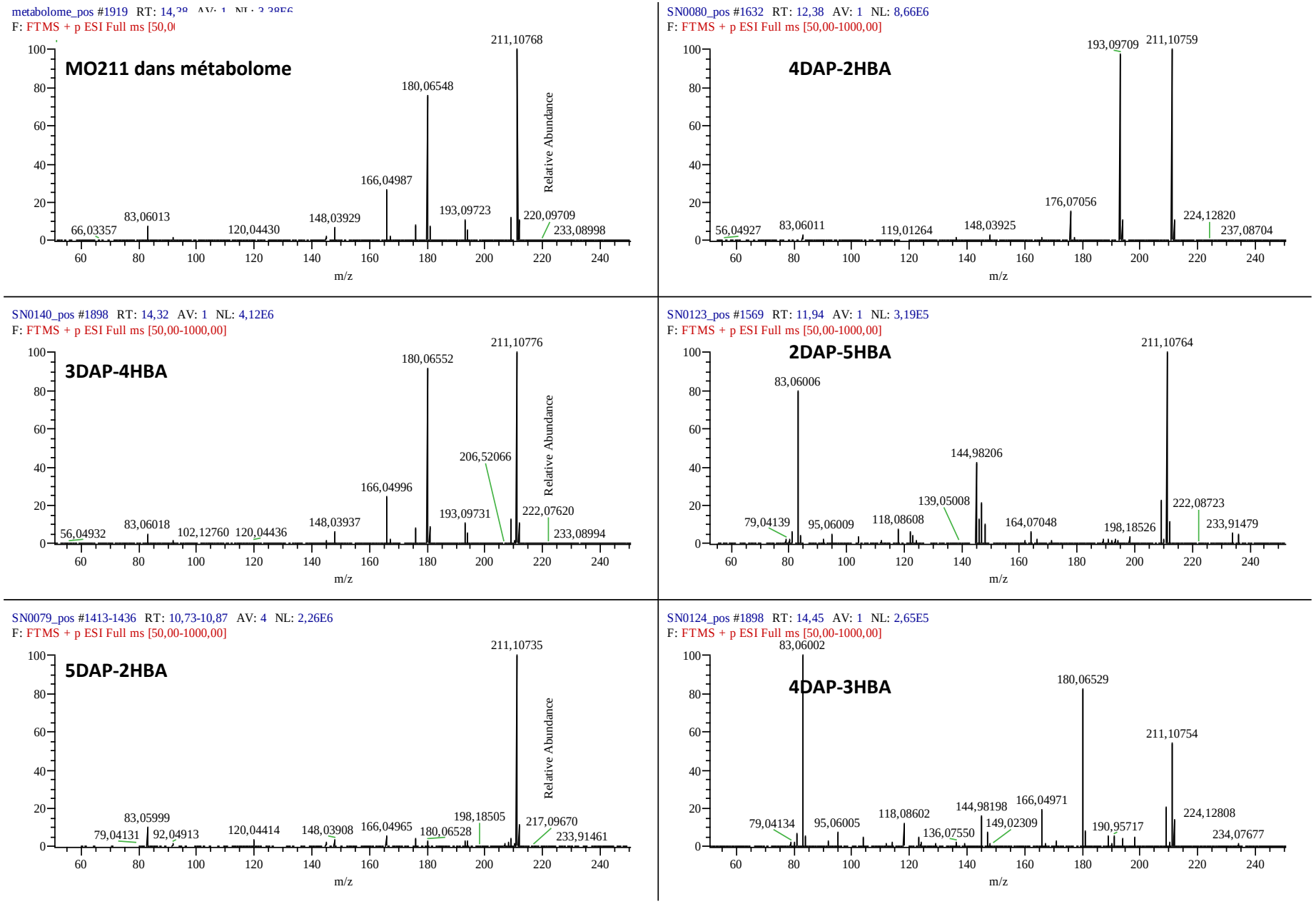
ANNEXE 12: Spectre RMN HMBC ^{1}H-^{15}N du MO211 ; solvant H_2O/D_2O (90:10)



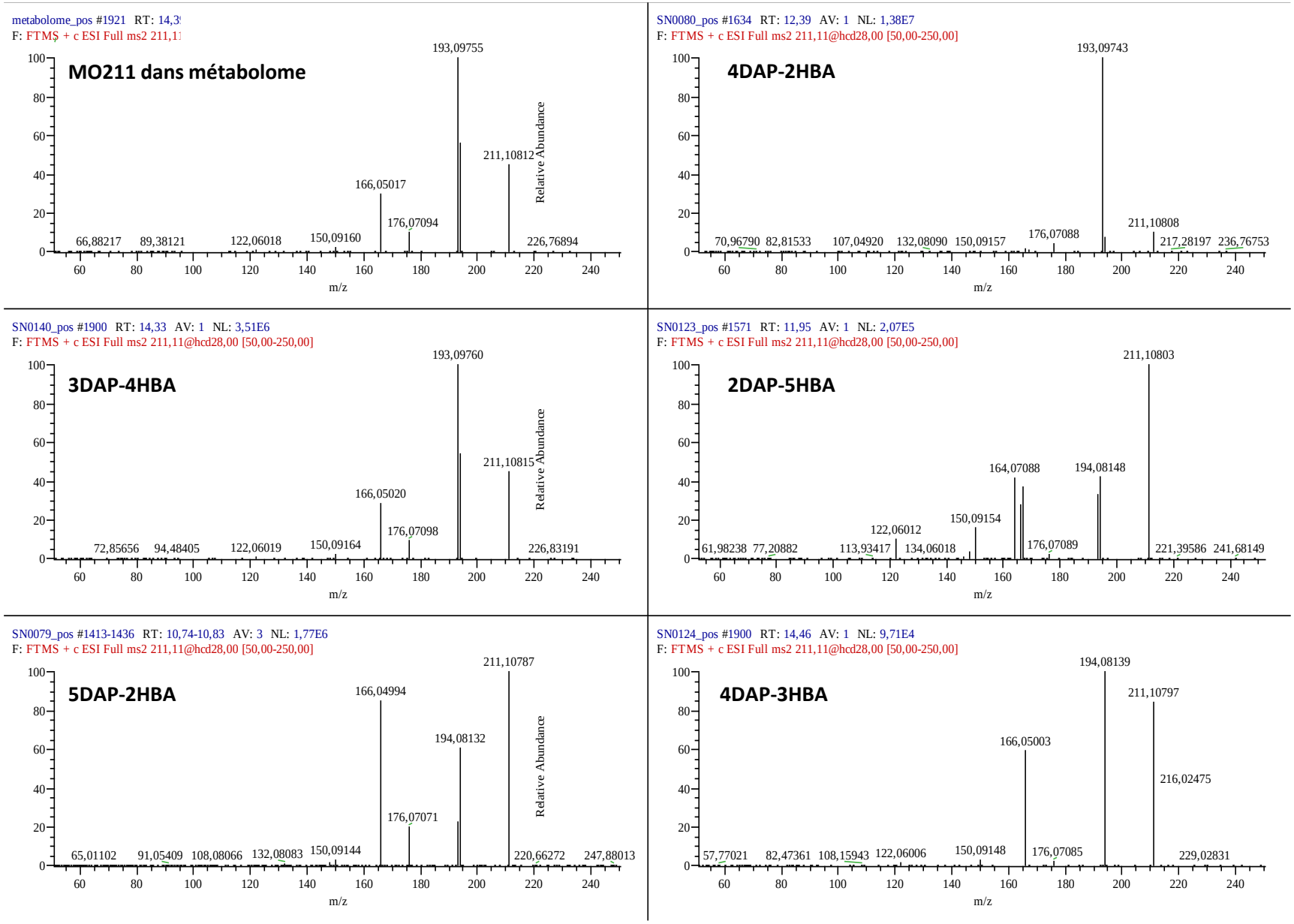
ANNEXE 13: Superposition des spectres RMN du MO211 et des composés synthétisés ; 600MHz, solvant H_2O/D_2O (90:10). (bleu) MO 211, (rouge) 3DAP-4HBA, (vert) 5DAP-2HBA, (violet) 4DAP-2HBA, (jaune) 4DAP-3HBA, (orange) 2DAP-5HBA.



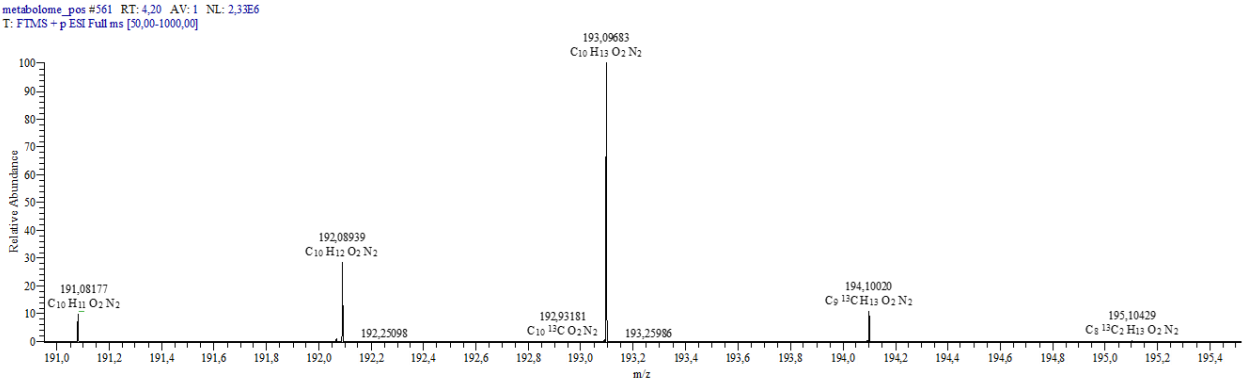
ANNEXE 14 : Spectres de détection en haute résolution du MO211 dans le métabolome et des 5 composés synthétisés, en mode d'ionisation positif

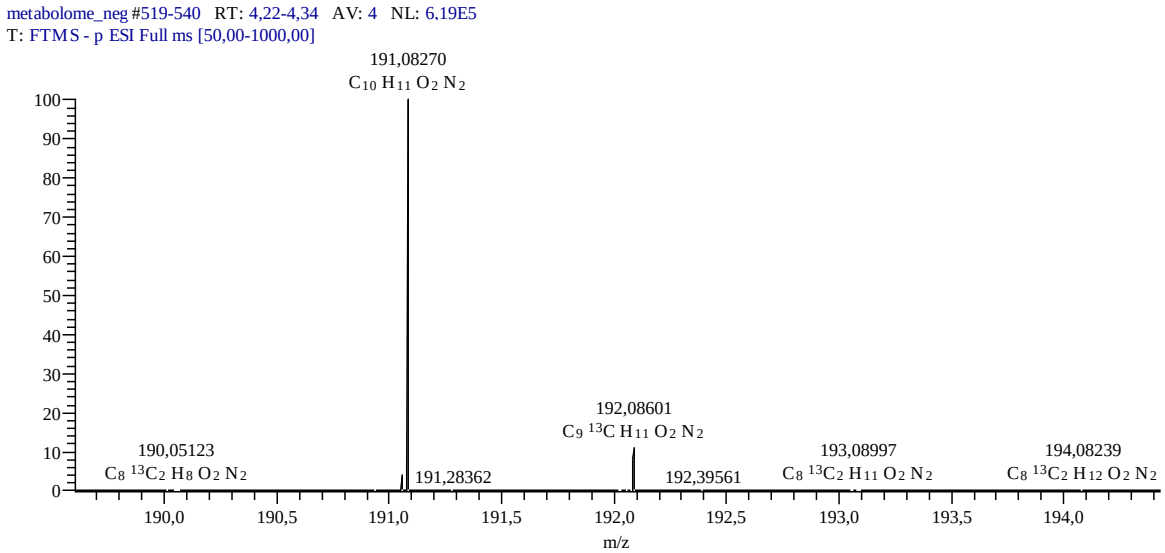


ANNEXE 15: Spectre MS² avec détection haute résolution du MO211 et des 5 composés synthétisés en mode d'ionisation positif

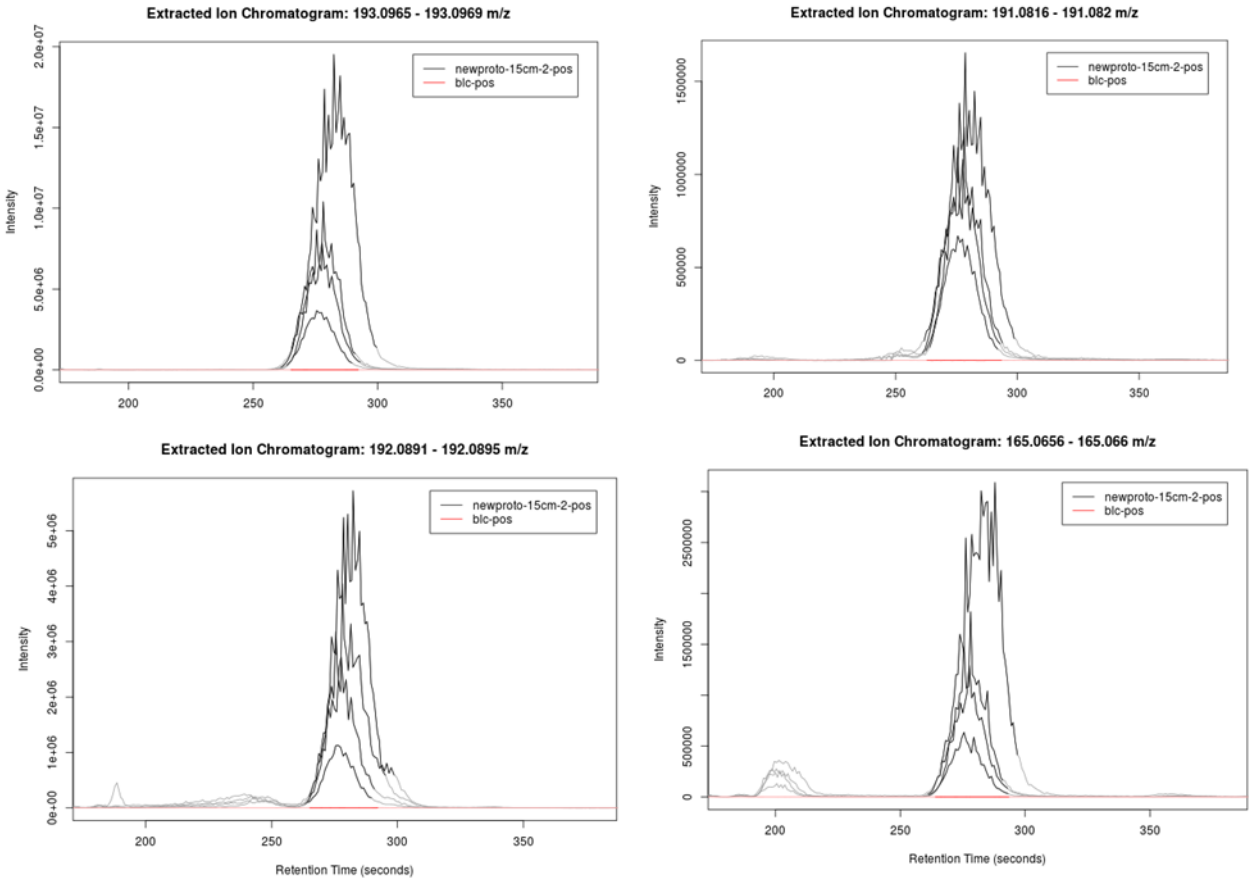


ANNEXE 16: Extrait du spectre de masse MS avec détection en haute résolution du MO193 en mode d'ionisation positif (haut) et négatif (bas).

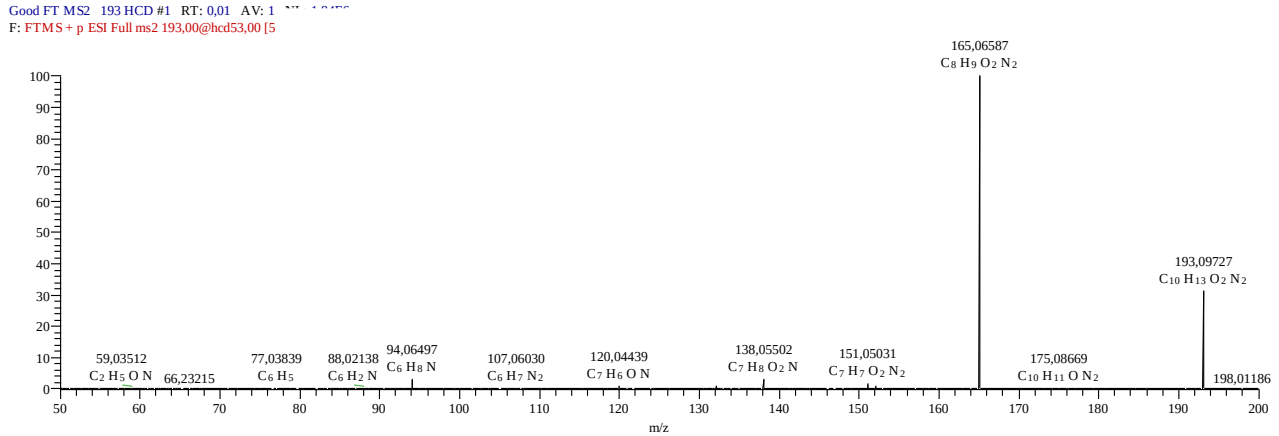


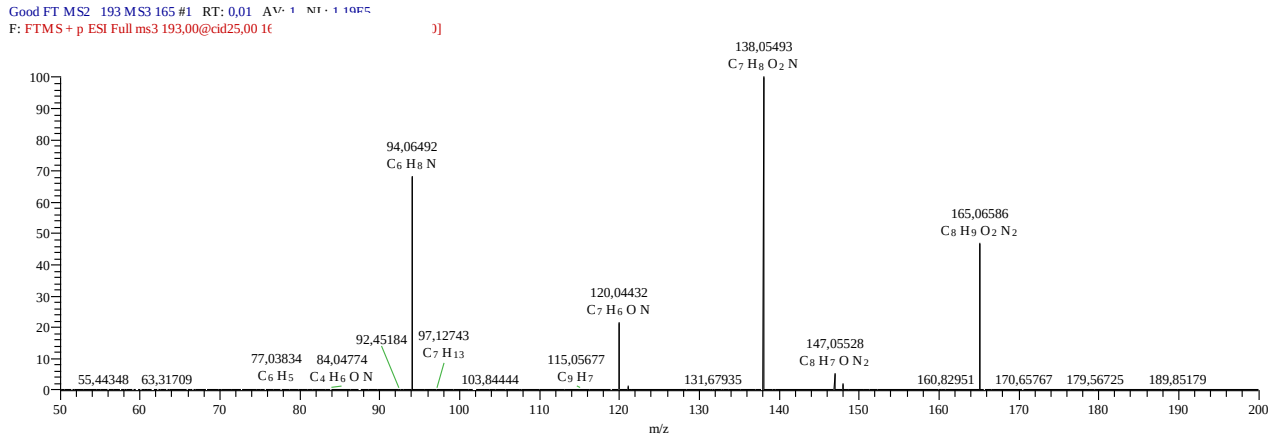


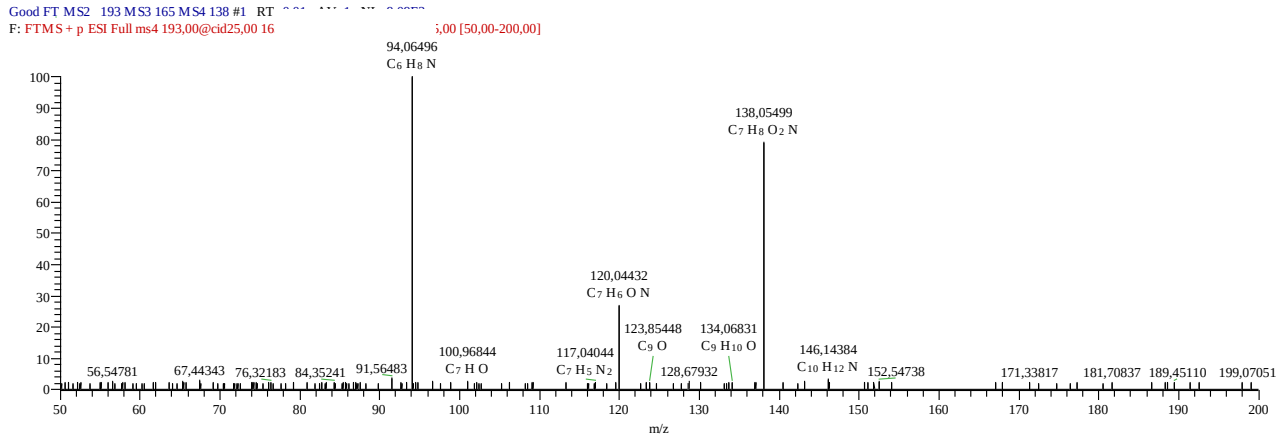
ANNEXE 17: XIC générés par XCMS des différents m/z semblant provenir du MO193. En noir, injections de métabolomes sur la colonne de 15 cm, en rouge, injections de phase mobile (analyse en mode d'ionisation positif).



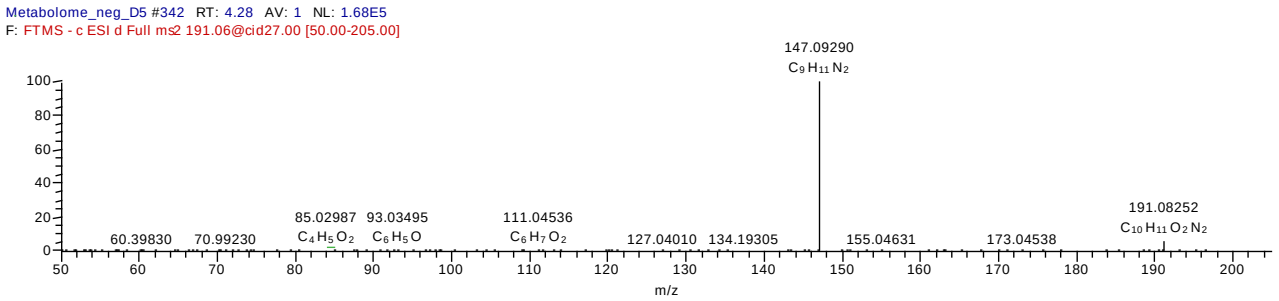
ANNEXE 18 : Spectres de détection et MS^n avec détection haute résolution de l'ion m/z 193,09727 en mode d'ionisation positif







ANNEXE 19: Spectres MS^n avec détection haute résolution de l'ion m/z 191,068252 en mode d'ionisation négatif



Revealing the hidden part of bacterial metabolism in *Acinetobacter baylyi* ADP1

Keywords : Metabolomics, Functional genomics, Mass spectrometry, Nuclear Magnetic Resonance, Structural elucidation, High throughput screening

Our knowledge of metabolism remains incomplete and its understanding is a major goal for both basic and applied research. About 40% of prokaryotic genes have no specific function proposed. Many metabolic pathways are therefore incomplete and others remain to be discovered. These gaps make the modelling and the understanding of cell metabolism very difficult. The laboratory is developing a new strategy to uncover unknown metabolic pathways. This is based on the model bacterium *Acinetobacter baylyi* ADP1 (ADP1). It exploits the complete library of deletion mutants of this organism (2600 mutants) as well as the recent advances of the laboratory in the field of metabolomics by liquid chromatography coupled with very high resolution mass spectrometry (LC / MS). This study is related to the observation that during a change of carbon source (from succinate to quinate), ADP1 produces many metabolites of unknown identity, unrelated to our knowledge of its metabolism. These orphan metabolites (MO) participate in novel metabolic pathways and represent entry points into the dark part of bacterial metabolism.

This project has two objectives. The first is to elucidate of the structure of MOs, and the second is to identify all the genes involved in their synthesis.

As part of this thesis, we established the identity of a first MO: 3- (3-aminopropylamine) - 4-hydroxybenzoic acid. This compound is a new metabolite, never referenced before. Firstly, LC / MS experiments involving sequential fragmentation MS_n and isotopic H / D exchanges allowed us to determine its elementary composition, its number of exchangeable protons and suggested an aromatic structure of the molecule. Then, its identification required its purification on the scale of the hundred µg. Then, a complete NMR analysis (^{1}H, ^{13}C, COSY, ^{1}H-^{13}C HSQC, ^{1}H-^{13}C HMBC, and ^{1}H-^{15}N HMBC) allowed us to resolve the structure of the compound.

Furthermore, to identify genes involved in the synthesis of MOs, we proceeded to the metabolomic analysis of all the strains of the ADP1 mutant collection. For this, we developed high-throughput protocols for the preparation of metabolomes (100 metabolomes per day), for the acquisition of LC / MS data (8 minutes per acquisition) and for their automated analysis. For every mutant, we are looking which MOs are present and which are absent. The absence of an MO in a given mutant indicates that the excised gene is involved in its synthesis. The systematic analysis of the bank should allow to list all the genes involved in the synthesis of each MO. The results of the screening, still in progress, will thus make it possible to begin to reconstitute these new metabolic pathways.

ÉCOLE DOCTORALE
Structure et dynamique des systèmes vivants (SDSV)

Exploration de la face cachée du métabolisme bactérien chez *Acinetobacter baylyi* ADP1

Mots clés : Métabolomique, Génomique fonctionelle, Spectrométrie de masse, Résonance Magnétique Nucléaire, Élucidation structurale, Criblage haut débit

La connaissance du métabolisme demeure incomplète mais sa compréhension reste un but majeur tant pour la recherche fondamentale qu'appliquée. Environ 40% des gènes des organismes procaryotes n'ont pas de fonction précise proposée. En conséquence, de nombreuses voies métaboliques sont incomplètes et d'autres restent à découvrir. Ces lacunes rendent très difficile la modélisation et la compréhension du métabolisme d'une cellule. Le laboratoire développe une nouvelle stratégie pour mettre au jour des voies métaboliques inconnues. Celle-ci est basée sur la bactérie modèle du laboratoire, *Acinetobacter baylyi* ADP1 (ADP1). Elle exploite la banque complète de mutants de délétion de cet organisme (2600 mutants) ainsi que les avancées récentes du laboratoire dans le domaine de la métabolomique par chromatographie liquide couplée à la spectrométrie de masse à très haute résolution (LC/MS).

Cette étude est liée à l'observation que lors d'un changement de source de carbone (succinate *vs* quinate) ADP1 produit de nombreux métabolites d'identité inconnue, non reliés à notre connaissance de son métabolisme. Ces métabolites orphelins (MO) participent à des voies métaboliques nouvelles et représentent des points d'entrée dans la partie obscure du métabolisme bactérien.

Ce projet comporte deux volets. Il s'agit d'une part de procéder à l'élucidation structurale des MO et d'autre part d'identifier l'ensemble des gènes impliqués dans leur synthèse.

Dans le cadre de ce travail de thèse, nous avons établi l'identité d'un premier MO : l'acide 3-(3-aminopropylamine)- 4-hydroxybenzoïque. Ce composé est un bien métabolite nouveau, jamais référencé auparavant. Dans un premier temps, des expériences de LC/MS impliquant notamment des fragmentations séquentielles MS^n et des échanges isotopiques H/D ont permis de déterminer sa composition élémentaire, son nombre d'hydrogènes échangeables et ont suggéré une structure aromatique de la molécule. Ensuite, son identification a nécessité sa purification à l'échelle de la centaine de µg. Puis, une analyse RMN complète (^{1}H, ^{13}C, COSY, ^{1}H-^{13}C HSQC, ^{1}H-^{13}C HMBC, et ^{1}H-^{15}N HMBC) a permis de préciser sa structure.

Par ailleurs, pour identifier les gènes impliqués dans la synthèse des MO, nous avons procédé à l'analyse métabolomique de l'ensemble des mutants de la collection d'ADP1. Pour cela, nous avons développé des protocoles à haut-débit pour la préparation des métabolomes (100 métabolomes par jour), l'acquisition des données LC/MS (8 minutes par acquisition) et enfin leur analyse automatisée. Nous regardons, pour chaque mutant, quels sont les MO présents et absents. L'absence d'un MO chez un mutant donné indique que le gène excisé est impliqué dans sa synthèse. L'analyse systématique de la banque doit permettre *in fine* de dresser la liste de tous les gènes impliqués dans la synthèse de chaque MO. Les résultats du criblage, toujours en cours, permettront ainsi de commencer à reconstituer ces voies métaboliques nouvelles.