

# TABLE DES MATIÈRES

---

|   |           |
|---|-----------|
| <b>Chapitre 1 – Introduction Générale</b>                   | <b>15</b> |
| <b>I.   Produits Naturels</b>                               | <b>17</b> |
| A.   Présentation des produits naturels                     | 17        |
| B.   Exemples de découverte de molécules naturelles actives | 17        |
| C.   Chimie des produits naturels                           | 18        |
| D.   Métabolites secondaires ou spécialisés végétaux        | 20        |
| 1)   Composés phénoliques                                   | 20        |
| 2)   Composés azotés  | 22        |
| 3)   Composés terpéniques                                   | 23        |
| <b>II.   Profilage des Métabolites et Métabolomique</b>     | <b>26</b> |
| A.   Définitions  | 26        |
| 1)   Profilage des métabolites                              | 26        |
| 2)   Métabolomique  | 27        |
| B.   Techniques analytiques                                 | 27        |
| 1)   Chromatographie en Phase-Liquide                       | 28        |
| 2)   Spectrométrie de masse                                 | 31        |
| 3)   Couplage UHPLC-QTOF-HRMS                               | 40        |
| <b>III.   DéréPLICATION et annotation des métabolites</b>   | <b>44</b> |
| A.   Définition et principe                                 | 45        |
| B.   Histoire   | 45        |
| C.   Avancées de la déréPLICATION                           | 46        |
| D.   Annotation des métabolites                             | 48        |
| 1)   Élucidation de structure <i>de novo</i>                | 48        |
| 2)   DéréPLICATION basée sur les bases de données           | 52        |
| <b>IV.   Réseaux Moléculaires</b>                           | <b>55</b> |
| A.   Définition et principe                                 | 55        |

|   |   |            |
|---|---|------------|
| B.  | Génération des réseaux moléculaires et visualisation              | 57         |
| 1)  | Génération des données spectrales                                 | 57         |
| 2)  | Prétraitement et transfert des données MS                         | 58         |
| 3)  | Génération des réseaux moléculaires                               | 59         |
| 4)  | Visualisation des réseaux moléculaires                            | 60         |
| C.  | Interprétation des réseaux moléculaires                           | 60         |
| D.  | Quelques avancées offertes par les réseaux moléculaires           | 61         |
| <b>V.</b>   | <b>Objectifs de la Thèse</b>                                      | <b>63</b>  |
| <b>Chapitre 2 – Développement et optimisation des réseaux moléculaires afin d'étudier la production des lignanes dans les cals cellulaires de <i>Linum usitatissimum L.</i></b> |   | <b>65</b>  |
| <b>I.</b>   | <b>Introduction</b>   | <b>67</b>  |
| <b>II.</b>  | <b>Cals cellulaires de <i>Linum usitatissimum L.</i></b>          | <b>69</b>  |
| A.  | Préparation d'échantillons  | 69         |
| 1)  | Mise en place des cals cellulaires                                | 69         |
| 2)  | Extraction  | 69         |
| B.  | Analyses UHPLC-MS/HRMS  | 70         |
| <b>III.</b>   | <b>Principaux paramètres influençant les réseaux moléculaires</b> | <b>73</b>  |
| A.  | Nombre de précurseurs   | 73         |
| B.  | Énergies de collision   | 75         |
| C.  | Fichiers de transfert des données spectrales                      | 77         |
| D.  | Nombre minimum de correspondances des spectres MS/MS              | 78         |
| E.  | Scores cosine   | 82         |
| <b>IV.</b>  | <b>Réseaux moléculaires de <i>Linum usitatissimum</i></b>         | <b>84</b>  |
| <b>V.</b>   | <b>Comparaison de production de lignanes</b>                      | <b>87</b>  |
| <b>VI.</b>  | <b>Conclusion</b>   | <b>94</b>  |
| <b>Chapitre 3 – Approche métabolomique et réseau moléculaire pour le criblage et la caractérisation de biomarqueurs</b>   |   | <b>97</b>  |
| <b>I.</b>   | <b>Introduction</b>   | <b>99</b>  |
| <b>II.</b>  | <b>Métabolomique Non-Ciblée</b>                                   | <b>101</b> |
| A.  | Préparation des échantillons                                      | 102        |

|  |  |            |
|--|--|------------|
| B.   | Acquisition des données  | 102        |
| C.   | Prétraitement des données                                      | 103        |
| D.   | Analyses Statistiques  | 103        |
| 1)   | Analyses univariées  | 103        |
| 2)   | Analyses multivariées  | 104        |
| E.   | Annotation des biomarqueurs                                    | 104        |
| F.   | Interprétation biologique                                      | 105        |
| <b>III.</b>  | <b>Criblage et identification de biomarqueurs</b>              | <b>106</b> |
| A.   | Préparation des échantillons                                   | 106        |
| 1)   | Culture en mésocosme   | 106        |
| 2)   | Extraction   | 107        |
| B.   | Acquisition des données  | 109        |
| C.   | Prétraitement des données                                      | 110        |
| D.   | Analyses statistiques  | 111        |
| E.   | Caractérisation des composés                                   | 115        |
| 1)   | Réseau moléculaire des extraits bruts                          | 115        |
| 2)   | Caractérisation basée sur l'élucidation de novo                | 119        |
| <b>IV.</b>   | <b>Conclusion</b>  | <b>128</b> |
| <b>Chapitre 4 – Combinaison de réseaux moléculaires et fractionnement CPC pour cibler des molécules d'intérêt cosmétique</b> |  | <b>131</b> |
| <b>I.</b>  | <b>Introduction</b>  | <b>133</b> |
| <b>II.</b>   | <b>Développement méthodologique</b>                            | <b>135</b> |
| A.   | Les plantes  | 135        |
| 1)   | <i>Artemisia annua</i> L.                                      | 135        |
| 2)   | <i>Eclipta alba</i> L.   | 136        |
| B.   | Préparation des échantillons                                   | 138        |
| 1)   | Culture de la matière végétale                                 | 138        |
| 2)   | Extraction   | 139        |
| C.   | Fractionnement par Chromatographie de Partage Centrifuge (CPC) | 140        |

|   |  |            |
|---|--|------------|
| D.  | Évaluation des activités biologiques                       | 142        |
| 1)  | Activité antioxydante                                      | 144        |
| 2)  | Activité anti-enzymatique                                  | 147        |
| 3)  | Activité cellulaire  | 148        |
| E.  | Étude phytochimique et élucidation structurale             | 150        |
| 1)  | Réseaux moléculaires de l'extrait brut                     | 150        |
| 2)  | Identification des composés bioactifs                      | 151        |
| <b>III.</b>   | <b>Criblage et identification des molécules bioactives</b> | <b>152</b> |
| A.  | <i>Artemisia annua</i> L.                                  | 152        |
| 1)  | Réseau moléculaire de l'extrait brut                       | 152        |
| 2)  | Activité antioxydante                                      | 156        |
| 3)  | Identification des molécules bioactives                    | 158        |
| 4)  | Viabilité cellulaire                                       | 162        |
| B.  | <i>Eclipta alba</i> L.                                     | 163        |
| 1)  | Réseau moléculaire de l'extrait brut                       | 163        |
| 2)  | Activités biologiques                                      | 185        |
| 3)  | Identification des molécules bioactives                    | 190        |
| 4)  | Viabilité cellulaire                                       | 198        |
| <b>IV.</b>  | <b>Conclusion</b>  | <b>201</b> |
| <b>Chapitre 5 – Conclusion Générale et Perspectives</b> |  | <b>203</b> |
| <b>Annexes</b>  |  | <b>211</b> |
| <b>Références bibliographiques</b>                      |  | <b>223</b> |

# AVANT-PROPOS

---

Cette thèse doctorale s'inscrit dans le projet **ValBioCosm'** (**V**alorisation de **B**iodiversité en **C**osmétique) financé par la Région Centre-Val de Loire dans le cadre de la deuxième phase du programme ARD 2020 Cosmétosciences. C'est le fruit d'une collaboration de quatre partenaires industriels : **SEDERMA** (Le Perray-en-Yvelines), **CAUDALIE** (Saint Jean de Braye), **Botanicosm'Ethic (BCE)**, Neuville aux Bois et le Comité de Développement Horticole de la Région Centre-Val de Loire (**CDHRC**, Saint-Cyr-en-Val), ainsi que quatre laboratoires académiques : l'Institut de Chimie Organique et Analytique (**ICOA UMR 7311**, Université d'Orléans), le Centre de Biophysique Moléculaire (**CBM UPR 4301**, CNRS Orléans), le laboratoire de Biomolécules et Biotechnologies Végétales (**BBV EA 2106**, Université de Tours) et le Laboratoire de Biologie des Ligneux et des Grandes Cultures (**LBLGC EA 1207**, Université Orléans, antenne de Chartres).

**ValBioCosm'** consiste à rassembler dans un projet inédit les compétences scientifiques et les savoir-faire des acteurs majeurs en Région Centre-Val de Loire (ou proche) de la filière cosmétique regroupant des procédés de *sourcing* végétal, des méthodes d'extraction, de caractérisation phytochimique, des tests d'activités biologiques jusqu'aux étapes de formulation et de commercialisation de nouvelles gammes cosmétiques.

Ce projet allant de la plante à l'actif porte d'une part sur l'innovation des procédés de production de biomasses agricoles et biotechnologiques en Région Centre-Val de Loire afin de démontrer la faisabilité d'implanter une filière de *sourcing* local permettant une meilleure traçabilité de la ressource végétale, des circuits courts d'approvisionnement avec une limitation de l'impact carbone et d'autre part sur le développement des méthodes de criblage de l'extrait et d'activités biologiques pour faciliter l'identification de nouveaux bioactifs propres à la commercialisation.

La Figure 1 est une représentation schématique du projet ValBioCosm' décrivant les différentes tâches mises en œuvre par les différents collaborateurs du projet.

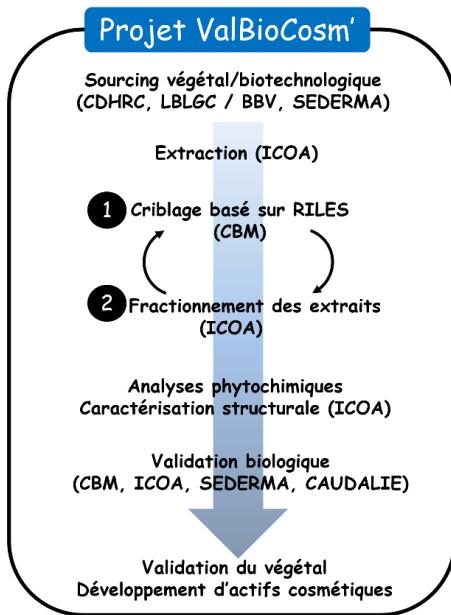


Figure 1 : Représentation schématique du projet ValBioCosm'

L'étape de *sourcing* végétal est mise en œuvre principalement par **CDHRC** et **LBLGC** dans le but de produire des cultures de plantes dans la région et d'améliorer la disponibilité végétale. Le *sourcing* biotechnologique est accompli par **BBV** et **SEDERMA** pour produire des cultures cellulaires de tissus végétaux permettant d'obtenir des matrices végétales indépendamment des conditions climatiques et saisonnières.

Les métabolites spécialisés contenus dans ces matrices végétales sont par la suite extraits à l'**ICOA** et au **LBLGC** pour évaluer les activités biologiques qu'ils pourraient offrir aux extraits bruts. Pour ce faire, des tests chimiques et enzymatiques *in-tubo* sont élaborés à l'**ICOA** et au **LBLGC**, et cellulaires *in-vitro* au **CBM**. De plus, une plateforme de criblage d'activité biologique par des microARN spécifiques (RILES) a été développée au **CBM** dans le cadre du projet **ValBioCosm'** afin d'évaluer l'activité de cicatrisation des extraits de plantes.

Les extraits ayant montré des propriétés biologiques d'intérêt cosmétique sont ensuite fractionnés et analysés en spectrométrie de masse à l'**ICOA** dans le but de cibler et d'identifier les composés responsables de ces activités biologiques. Enfin, l'ensemble des collaborateurs participent à la validation des actifs cosmétiques pour une éventuelle intégration des extraits dans des produits finis.

Dans ce contexte, l'objectif de ce travail de thèse a été de développer et de mettre en œuvre des approches d'identification des compositions moléculaires d'extraits de plantes ou de cultures cellulaires végétales. Ces stratégies doivent être simples, rapides, efficaces et applicables à différents types et nombres d'échantillons. Différentes matrices végétales produites par nos collaborateurs ont été utilisées pour progresser dans le développement de ces outils et dans leur combinaison pour répondre aux différentes problématiques.

Ce mémoire, consacré au criblage de molécules naturelles d'intérêt cosmétique par combinaison d'outils analytiques, bioinformatiques et métabolomiques, est constitué de cinq chapitres.

Le premier chapitre introductif bibliographique, situe l'intérêt des substances naturelles et les différents outils développés et améliorés au cours du temps en chimie des produits naturels dans le but de caractériser au mieux les métabolomes des plantes. Ce chapitre appuie sur l'utilisation des approches de métabolomique, déréplication et réseaux moléculaires en détaillant les techniques analytiques utilisées durant cette thèse.

Le second chapitre présente l'optimisation des paramètres influençant la création des réseaux moléculaires et leur intérêt pour la comparaison de la teneur en lignanes, une famille d'intérêt cosmétique, produits par des cals cellulaires de *Linum usitatissimum* cultivés dans différentes conditions.

Le troisième chapitre est dédié au développement des approches de métabolomique et réseau moléculaire pour mettre en évidence et identifier des biomarqueurs de développement de plants de tomate *Solanum lycopersicum*. En effet, pour déterminer des molécules discriminantes de groupes d'extraits, des approches chimiométriques ont été mise en œuvre et associées aux informations de déréplication obtenues grâce aux réseaux moléculaires pour identifier les biomarqueurs.

Dans le quatrième chapitre, nous nous sommes intéressées au criblage de composés bioactifs d'*Artemisia annua* et d'*Eclipta alba* cultivées en Région Centre-Val de Loire. C'est pourquoi une combinaison de réseaux moléculaires et de chromatographie de partage centrifuge a été élaborée pour identifier les molécules contribuant majoritairement aux activités antioxydante et anti-enzymatique de ces espèces végétales.

Enfin, le cinquième et dernier chapitre présente un bilan ces stratégies d'identification moléculaire appliquées aux problématiques de comparaison d'extraits et de criblage de molécules actives, ainsi que les perspectives qui découlent du travail effectué durant ces trois années de thèse.

# ABRÉVIATIONS ET ACRONYMES

---

| Acronyme | Signification  |
|----------|--|
| ABTS     | 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulphonic acid)            |
| BBV      | Biomolécules et Biotechnologies Végétales                          |
| BPC      | Base Peak Chromatogram   |
| CBM      | Centre de Biophysique Moléculaire                                  |
| CDHRC    | Comité de Développement Horticole de la Région Centre-Val de Loire |
| CID      | Collision-Induced Dissociation                                     |
| Cos      | Cosine score   |
| CPC      | Centrifugal Partition Chromatography                               |
| CUPRAC   | CUpric Reducing Antioxidant Capacity                               |
| DAD      | Diode Array Detection  |
| DDA      | Data Dependent Acquisition   |
| DIA      | Data Independent Acquisition                                       |
| DNP      | Dictionary of Natural Products                                     |
| DPPH     | 2,2-diphenyl-1-picrylhydrazyl                                      |
| Ea       | <i>Eclipta alba</i>  |
| EC       | Electrophoresis Chromatography                                     |
| ELSD     | Evaporative Light Scattering Detector                              |
| ESI      | ElectroSpray Ionization  |
| eV       | Electronvolt   |
| FDA      | Food and Drug Administration                                       |
| FPP      | Fully Porous Particles   |
| FRAP     | Ferric Reducing Antioxidant Power                                  |
| G        | Gravitational constant   |

|                |   |
|----------------|---|
| GC             | Gas Chromatography  |
| Glc            | Glucose   |
| GNPS           | Global Natural Products Social Molecular Networking         |
| HCA            | Hierarchical Clustering Ascendant                           |
| HILIC          | Hydrophilic Interaction LIquid Chromatography               |
| HRMS           | High Resolution Mass Spectrometry                           |
| ICOA           | Institut de Chimie Organique et Analytique                  |
| K <sub>D</sub> | Coefficient de partage                                      |
| kV             | KiloVolt  |
| LBLGC          | Laboratoire de Biologie des Ligneux et des Grandes Cultures |
| LC             | Liquid Chromatography                                       |
| L-DOPA         | L-3,4-dihydroxyphenylalanine                                |
| M              | Molecule  |
| m/z            | Rapport masse sur charge                                    |
| mDa            | milli Dalton  |
| mgf            | Masscot Generic Format                                      |
| min            | Minutes   |
| MS             | Mass Spectrometry   |
| MS/MS          | Tandem mass spectrometry                                    |
| mzXML          | <i>m/z</i> Extensible Markup Language                       |
| NI             | Negative Ionization   |
| NMR            | Nuclear Magnetic Resonance                                  |
| PCA            | Principal Component Analysis                                |
| PI             | Positive Ionization   |
| PLS-Da         | Partial Least Squares Discriminant Analysis                 |
| ppm            | Part Per Million  |
| ROS            | Reactive Oxygen Species                                     |

|             |   |
|-------------|---|
| RP          | Reverse Phase   |
| rpm         | Round Per Minute  |
| Rt          | Retention time  |
| SFC         | Supercritical Fluid Chromatography  |
| SPP         | Superficialy Porous Particles   |
| TLC         | Thin-layer chromatography   |
| TOF         | Time Of Flight analyzers  |
| u           | Arbitrary Unit  |
| UHPLC       | Ultra High-Performance Liquid Chromatography  |
| UV          | Ultraviolet   |
| ValBioCosm' | Valorisation de Biodiversité en Cosmétique  |
| VIP         | Variable Important in Projection  |
| W4M         | Workflow4Metabolomics   |
| XRD         | X-Ray Diffraction   |
| XTT         | Cell Proliferation assay : 2,3-bis-(2-methoxy-4-nitro-5-sulfophenyl)-2H-tetrazolium-5-carboxanilide |

# LISTE DES FIGURES

---

Figure 1 : Représentation schématique du projet ValBioCosm'

Figure 2 : Quelques composés appartenant à la famille des composés phénoliques

Figure 3 : Quelques composés appartenant à la famille des alcaloïdes

Figure 4 : Quelques composés appartenant à la famille des terpènes

Figure 5 : Profilage des métabolites obtenu pour un extrait standardisé de *Ginkgo biloba* présenté par Wolfender *et al.* (2019)<sup>43</sup>. a) Colonne Xbridge BEH C18 (150 x 4,6 mm ; 5 µm). b) Colonne Acquity BEH C18 (150 x 2,1 mm ; 1,7 µm). c) Colonne Acquity BEH C18 (50 x 2,1 mm ; 1,7 µm)

Figure 6 : Les éléments communs des spectromètres de masse

Figure 7 : Profilage UHPLC-HRMS obtenu avec l'ionisation ESI en modes positif (PI) et négatif (NI) d'un extrait brut de feuilles de maïs effectué par G. Marti *et al.* (2012)<sup>49</sup>

Figure 8 : Spectromètre de masse maXis UHR-TOF de Bruker

Figure 9 : Illustration du principe du mode d'acquisition dépendante des données DDA. Figure adaptée de Christina Ludwig *et al.* (2018)

Figure 10 : Illustration du principe du mode d'acquisition indépendante des données DDA. Figure adaptée de Christina Ludwig *et al.* (2018)<sup>57</sup>

Figure 11 : Analyse UHPLC-HRMS d'un extrait de *Ginkgo biloba* réalisée par J-L. Wolfender *et al.* (2019)<sup>43</sup>. (a) Chromatogrammes d'ions totaux (TIC) obtenus en modes d'ionisation positif et négatif. (b) Carte d'ions des données caractéristiques des ions détectés en mode positif avec un seuil d'intensité de 0,01%. (c) Nombre de données caractéristiques en fonction du niveau de seuil d'intensité

Figure 12 : Schéma récapitulatif des outils analytiques et informatiques qui permettent l'identification des métabolites réalisé par J. Bitzer *et al.* (2007)<sup>66</sup>

Figure 13 : Schéma récapitulatif des relations entre les différents espaces impliqués dans la chimie des produits naturels, reporté par A. Bakiri *et al.* (2017)<sup>67</sup>

Figure 14 : Workflow schématique de déréplication/élucidation des structures dans la chimie des produits naturels énoncé par J-L. Wolfender *et al.* (2019)<sup>76</sup>

Figure 15 : Exemple d'annotation d'un pic LC-MS basée sur un profilage haute résolution de *Viola tricolore* réalisé par J-L. Wolfender *et al.* (2015)<sup>84</sup>

Figure 16 : Workflow de déréplication basé sur la détermination de la formule brute et la recherche croisée dans les bases de données pour un flavonoïde glycosylé d'un extrait de *Ginkgo biloba* énoncé par J-L. Wolfender *et al.* (2019)<sup>43</sup>

Figure 17 : Schématisation du processus de création des réseaux moléculaires présenté par R.A. Quinn *et al.* (2017)<sup>113</sup>

Figure 18 : Spectre expérimental en (-) ESI de la catéchine, comparé à ses spectres de référence à différentes énergies de collision tiré de la thèse de Florent Olivon (2018)<sup>294</sup>

Figure 19 : Formule de calcul des scores cosine entre deux spectres MS/MS A et B

Figure 20 : Cals cellulaires de *Linum usitatissimum*. A) culture de cals à la lumière après 30 jours, B) culture de cals à l'obscurité après 30 jours. Images prises par nos collaborateurs du laboratoire Biomolécules et Biotechnologies Végétales à l'université de Tours

Figure 21 : BCP (*Base Peak Chromatogram*) des extraits de cals cellulaires de *Linum usitatissimum* cultivés à la lumière (rose) et à l'obscurité (bleu), obtenus par UHPLC-HRMS en mode (+) ESI

Figure 22 : Réseaux moléculaires de cals cellulaires de *Linum usitatissimum* cultivés à la lumière, obtenus en fonction du nombre d'ions précurseurs sélectionnés. Graphe résumant le nombre total de nœuds, de nœuds singltons et de clusters (> 2 nœuds) des réseaux moléculaires

Figure 23 : Réseaux moléculaires de cals cellulaires de *Linum usitatissimum* cultivées en lumière, obtenus à différentes énergies de collision. Graphe résumant le nombre de nœuds total, de nœuds singltons et de clusters (> 2 nœuds) des réseaux moléculaires

Figure 24 : Graphe présentant les informations des réseaux moléculaires (cals cellulaires de *Linum usitatissimum*) obtenues après transfert de données spectrales sur GNPS par différents formats de fichier

Figure 25 : Correspondances spectrales considérées lors de la comparaison des spectres MS/MS et du calcul des scores cosine par la plateforme GNPS lors de la création des réseaux moléculaires des cals cellulaires de *Linum usitatissimum*

Figure 26 : Réseaux moléculaires de cals cellulaires de *Linum usitatissimum* cultivés à la lumière, obtenus en fonction du nombre de correspondances minimal exigé. Graphe résumant le nombre total de nœuds, de nœuds singltons et de clusters (> 2 nœuds) des réseaux moléculaires

Figure 27 : Réseaux moléculaires de cals cellulaires de *Linum usitatissimum* cultivés à la lumière, obtenus en fonction de différents scores cosine. Graphe résumant le nombre total de nœuds, de nœuds singletons et de clusters (> 2 nœuds) des réseaux moléculaires

Figure 28 : Réseau moléculaire qualitatif (AutoMSn, DataAnalysis) des extraits des cals cellulaires de *Linum usitatissimum* cultivés à la lumière (rose) et à l'obscurité (bleu).

Figure 29 : Réseau moléculaire quantitatif (MetaboScape) des extraits de cals cellulaires de *Linum usitatissimum* cultivés à la lumière (rose) et à l'obscurité (bleu)

Figure 30 : Cluster représentant les lignanes d'intérêt extraits des cals cellulaires de *Linum usitatissimum* cultivés dans la lumière (rose) et l'obscurité (bleu). Le réseau moléculaire a été obtenu par les analyses UHPLC-MS/HRMS en ionisation (+) ESI

Figure 31 : Clusters regroupant les composés majoritairement affectés par les conditions de culture des cals cellulaires de *Linum usitatissimum*, soit lumière (rose) soit obscurité (bleu)

Figure 32 : Schéma général d'une analyse métabolomique par LC-HRMS présenté par Pierre BARBIER (2019)<sup>295</sup>

Figure 33: schéma récapitulatif de la démarche développée pour l'identification de biomarqueurs discriminants des racines de tomate à différents stades de développement

Figure 34 : Culture des plants de tomate en mésocosme

Figure 35 : Les BPCs (*Base Peak Chromatograms*) des analyses UHPLC-HRMS en mode (+) ESI des extraits de racines de tomates obtenus avec différentes proportions (MeOH/H<sub>2</sub>O acidifiée avec 0,1 % Acide formique) comme solvant d'extraction

Figure 36 : Les BPCs (*Base Peak Chromatograms*) d'extraits de racines de tomate jeunes (vert) et matures (rouge) obtenus par UHPLC-HRMS en mode (+) ESI

Figure 37: Analyses multivariées non supervisées A) PCA (*Principal Component Analysis*) des extraits de racines de tomate jeunes (vert) et matures (rouge). B) Dendrogramme HCA (*Hierarchical Clustering Ascendant*). Les paramètres utilisés sont Euclidean distance measure et l'algorithme ward clustering

Figure 38 : A) Analyse multivariée supervisée PLS-DA (*Partial Least Squares Discriminant Analysis*) des deux groupes de racines de tomate jeunes (vert) et matures (rouge). B) Analyse univariée *Volcano Plot*

Figure 39 : A) VIP (*Variable Important in Projection*) scores. B) Box Plots des 15 variables les plus discriminantes entre les groupes des racines de tomate jeunes (vert) et matures (rouge)

Figure 40 : Numérotation des pics chromatographiques correspondants aux biomarqueurs, discriminant les racines jeunes et matures, mis en évidence par les analyses statistiques

Figure 41 : Réseau moléculaire des groupes jeunes (vert) et matures (rouge) racines de tomate, obtenu en utilisant la plateforme GNPS et visualisé sur le logiciel Cytoscape 3.5.1. Les molécules entourées en bleu sont les composés annotés grâce aux bases de données de GNPS, et ceux entourés en mauve sont les biomarqueurs mis en évidence par les analyses statistiques (scores VIP > 1)

Figure 42 : Spectre MS/MS obtenu en mode (+) ESI et interprétation du schéma de fragmentation possible pour le VIP 1, l'acétylagmatine

Figure 43 : Spectre MS/MS obtenu en mode (+) ESI et interprétation du schéma de fragmentation du VIP2, le tryptophane

Figure 44 : Spectre MS/MS obtenu en mode (+) ESI et interprétation du schéma de fragmentation du VIP 3, (S)-5'-Désoxy-5'-(méthylsulfinyl)adénosine

Figure 45 : Spectre MS/MS obtenu en mode (+) ESI et interprétation du schéma de fragmentation du VIP 4, l'adénosine monophosphate

Figure 46 : *Artemisia annua* L.

Figure 47 : Artémisinine, principe actif de médicaments antipaludiques

Figure 48 : *Eclipta alba* L.

Figure 49 : Quelques produits cosmétiques commercialisés à base d'*Eclipta alba*

Figure 50 : La colonne constituée de plusieurs disques sur un rotor (A), ces disques (B) contiennent de cellules, qui sont connectées par un canal mince (C). Images adaptées d'Armen, Gilson

Figure 51 : Élution des composés en CPC selon leurs coefficients de partage  $K_D$  (Kromaton)

Figure 52 : Illustration du stress oxydatif et des facteurs externes aggravant sur la peau

Figure 53 : Mécanisme d'action anti-radicalaire. Exemple de piégeage d'anion superoxyde par la quercétine présenté par Balasaheb Nimse *et al* (2015)<sup>263</sup>

Figure 54 : Mécanismes réactionnels mis en jeu lors des tests chimiques colorimétriques en présence de composés antioxydants. DPPH (2,2-diphénol-1-picrylhydrazyl), ABTS (2,2-azino-bis-(3-éthylbenzothiazoline-6-sulfonic acid)), FRAP (2,3,5-triphenyltetrazolium), CUPRAC (neocuproine copper) et Chélation du fer (ferrozine)

Figure 55 : Illustration des voies d'hydrolyse de collagénase et d'élastase, ainsi que l'implication de la tyrosinase dans la synthèse de mélanine

Figure 56 : Résumé de la démarche développée pour cibler et caractériser les molécules bioactives d'*Artemisia annua* et d'*Eclipta alba*

Figure 57 : Le BPC (Base Peak Chromatogram) de l'extrait éthanolique d'*Artemisia annua* obtenu par UHPLC-HRMS en mode d'ionisation (+) ESI

Figure 58 : Le réseau moléculaire de l'extrait brut *Artemisia annua* obtenu à l'aide de la plateforme GNPS et visualisé avec le logiciel Cytoscape 3.5.1.

Figure 59 : Résultats de l'activité antioxydante de l'extrait brut éthanolique et des fractions CPC. L'absorbance moyenne a été tracée  $\pm$  l'écart-type ( $n = 3$ )

Figure 60 : (A) Focalisation sur le cluster de l'ion d'intérêt contenu dans la Fraction 2. (B) Le chromatogramme UV de la Fraction 2 à  $\lambda = 280$  nm, la séparation a été effectuée sur une colonne Fortis C18 (150 x 2,1 mm ; 1,7  $\mu$ m) à 60°C. (C) Spectre de masse de l'ion d'intérêt. (D) Structure de chrysosplénol D

Figure 61 : Spectre MS/MS de l'ion d'intérêt contenu dans la fraction active (Fraction 2)

Figure 62 : Proposition de fragmentation du chrysosplénol D

Figure 63 : Spectre MS/MS du standard chrysosplénol D présenté par Wang *et al* (2019)<sup>281</sup>

Figure 64 : Flavonoïdes de l'extrait éthanolique d'*Artemisia annua*

Figure 65 : Le BPC (Base Peak Chromatogram) de l'extrait éthanolique d'*Eclipta alba* obtenu en UHPLC-HRMS en mode d'ionisation (+) ESI

Figure 66 : Superposition des BPCs (Base Peak Chromatograms) d'extrait éthanolique d'*Eclipta alba* obtenu par UHPLC-HRMS en utilisant les modes d'ionisation (+) ESI en bleu, et (-) ESI en orange

Figure 67 : A) Réseau moléculaire visualisé sur Cytoscape et B) BPC (Base Peak Chromatogram) de l'extrait éthanolique d'*Eclipta alba* obtenu par UHPLC-HRMS en mode (+) ESI

Figure 68 : A) Réseau moléculaire global de l'extrait éthanolique d'*Eclipta alba* visualisé sur Cytoscape. B) BPC (Base Peak Chromatogram) de l'extrait éthanolique d'*Eclipta alba* obtenu par UHPLC-HRMS en mode (+) ESI. C) Focalisation sur les clusters des composés majoritaires de l'extrait éthanolique d'*Eclipta alba*. Les numéros associés aux nœuds sont les composés annotés sur le réseau moléculaire et détaillés dans le Tableau 8

Figure 69 : A) Spectres MS/MS des composés liés dans le cluster 1 avec une focalisation sur la partie commune  $m/z$  100-550. B) Exemple d'un massif isotopique de l'ion moléculaire du composé 44

Figure 70: Spectres MS/MS des composés 46 et 52

Figure 71 : Interprétation du spectre MS/MS du composé 32 et proposition de voies de fragmentations

Figure 72 : Spectres MS/MS des composés 32, 36 et 37 liés en cluster 3

Figure 73 : Les familles moléculaires regroupant les composés les plus intenses détectés par UHLPLC-HRMS en mode (+) ESI dans l'extrait éthanolique des parties aériennes d'*Eclipta alba*

Figure 74 : Activité antioxydante de l'extrait *Eclipta alba* et ses fractions CPC. L'absorbance moyenne a été tracée  $\pm$  l'écart-type ( $n = 3$ )

Figure 75 : Activité d'inhibition de collagénase par l'extrait *Eclipta alba* et ses fractions CPC. La fluorescence moyenne a été tracée  $\pm$  l'écart-type ( $n = 3$ )

Figure 76 : Activité d'inhibition de l'élastase par l'extrait d'*Eclipta alba* et ses fractions CPC. La fluorescence moyenne a été tracée  $\pm$  l'écart-type ( $n = 3$ )

Figure 77 : Inhibition de la tyrosinase par l'extrait d'*Eclipta alba* et ses fractions CPC. L'absorbance moyenne a été tracée  $\pm$  l'écart-type ( $n = 3$ )

Figure 78 : BPC (Base Peak Chromatogram) de l'extrait d'*Eclipta alba* et des fractions CPC

Figure 79 : Évolution de l'élution des composés polaires à moyennement polaires, présentés en nœuds singletons sur le réseau moléculaire, dans les différentes fractions CPC de l'extrait d'*Eclipta alba*

Figure 80 : Évolution de l'élution des composés du cluster 1 dans les différentes fractions CPC de l'extrait d'*Eclipta alba*

Figure 81 : Évolution de l'élution des composés des clusters 2 3 et 2' dans les différentes fractions CPC de l'extrait d'*Eclipta alba*

Figure 82 : Évaluation de la cytotoxicité XTT des cellules HaCaT à différentes concentrations d'extrait d'*Eclipta alba* après un traitement de 48 heures. Les valeurs sont exprimées en % de viabilité normalisée à 100 % des cellules non traitées. Les valeurs sont la moyenne  $\pm$  l'écart type d'une expérience réalisée au moins 3 fois en quintuplet, \* $p < 0,05$ , \*\* $p < 0,01$ , test T-student utilisant un test bilatéral non apparié

Figure 83 : Protection d'extrait d'*Eclipta alba* contre les espèces réactives de l'oxygène (ROS) dans les cellules HaCaT traitées avec 250  $\mu\text{M}$  de  $\text{H}_2\text{O}_2$ . Les valeurs sont exprimées en % de la production de ROS normalisée à 100 % des cellules non traitées. Les valeurs sont la moyenne  $\pm$  l'écart-type d'une expérience réalisée au moins 3 fois en quintuplet, \* $p < 0,05$ , \*\* $p < 0,01$ , test T-student utilisant un test bilatéral non apparié

# LISTE DES TABLEAUX

---

Tableau 1 : Les contributions et avancées offertes par la chimie des produits naturels présenté par R.A Khan (2018)<sup>11</sup>

Tableau 2 : Principales sources d'ionisation en spectrométrie de masse

Tableau 3 : Quelques exemples de bases de données de métabolites

Tableau 4 : Les outils informatiques utilisés dans le workflow de W4M (*workflow4metabolomics*) pour le prétraitement des données UHPLC-HRMS des racines de tomate

Tableau 5 : Annotation des métabolites impliqués lors du développement des racines de tomates (scores VIP > 1 et p-value < 0,001)

Tableau 6 : Composés annotés sur le réseau moléculaire de l'extrait éthanolique d'*Artemisia annua*

Tableau 7 : Activité antioxydante de l'extrait brut *Artemisia annua* et ses 13 fractions CPC à 0,5 mg/mL, évaluée par les tests DPPH, ABTS, FRAP, CUPRAC et chélation du fer (II)

Tableau 8 : Les composés chimiques contenus dans l'extrait éthanolique des parties aériennes d'*Eclipta alba*

Tableau 9 : Récapitulatif des résultats antioxydants et anti-enzymatiques des fractions CPC d'*Eclipta alba*

Tableau 10 : Gradient utilisé pour l'analyse UHPLC-HRMS des extraits de cals cellulaires de *Linum usitatissimum* (B : acétonitrile avec 0,08% d'acide formique).

Tableau 11 : Gradient utilisé pour l'analyse UHPLC-HRMS des racines de tomate. (B : acétonitrile avec 0,08% d'acide formique).

Tableau 12 : Gradient utilisé pour l'analyse UHPLC-HRMS d'*Artemisia annua*. (B : acétonitrile : 2-propanol à 50:50 avec 0,1% d'acide formique).

Tableau 13: Gradient utilisé pour l'analyse UHPLC-HRMS d'*Eclipta alba*. (B : acétonitrile avec 0,08% d'acide formique)



# **CHAPITRE 1 – INTRODUCTION GÉNÉRALE**

---



## I. PRODUITS NATURELS

### A. Présentation des produits naturels

Les substances ou produits naturels sont des composés chimiques synthétisés par les organismes vivants procaryotes et eucaryotes tels que bactéries, champignons, plantes, organismes marins et animaux. Ils sont divisés en deux grandes classes : les métabolites primaires et les métabolites secondaires ou spécialisés<sup>1,2</sup>.

Les métabolites primaires sont nécessaires à la survie de l'organisme, ce sont des composants des voies métaboliques de base qui sont associés aux fonctions cellulaires essentielles telles que la croissance, le développement, la production d'énergie, etc. Ils comprennent les glucides, les lipides, les acides aminés et les acides nucléiques qui sont des éléments de base de la vie.

Les métabolites secondaires ou spécialisés ne sont pour la plupart pas indispensables à la survie de l'organisme, mais augmentent sa compétitivité dans son environnement<sup>3</sup>. Ils ont un large éventail de fonctions telles que la communication, la défense, la solubilisation et le transport des nutriments, etc. Cependant, certains de ces métabolites tels que la chlorophylle et la lignine, ont des fonctions essentielles pour la croissance de la plante et pourraient faire partie des métabolites primaires, tandis que la fonction de beaucoup d'autres métabolites secondaires n'est pas encore bien connue.

Les métabolites secondaires sont généralement spécifiques aux espèces et sont historiquement davantage corrélés aux plantes, bactéries et champignons, mais on découvre également des métabolites spécifiques à certains groupes d'animaux. Les classes structurales générales des métabolites secondaires comprennent les alcaloïdes, les phénylpropanoïdes et les terpénoïdes. En raison de leur capacité à moduler les voies biochimiques et les voies de transduction du signal, certains métabolites secondaires ont des propriétés biologiques utiles<sup>4</sup>.

### B. Exemples de découverte de molécules naturelles actives

Les civilisations anciennes ont utilisé diverses parties de plantes et d'animaux afin d'en évaluer les effets. Des sources d'effets thérapeutiques et cosmétiques ont été remarquées et ont ainsi été transmises, souvent oralement, de génération en génération dans des pratiques comme la médecine traditionnelle chinoise, l'ayurvédique, etc.

#### Origine animale

Il y a au moins 5000 ans, en Asie et en Afrique, une substance toxique : la cantharidine qui était utilisée comme aphrodisiaque ou pour soigner l'angine de poitrine<sup>5</sup>, était extraite de la mouche espagnole. Le lucane produisait un diurétique qui soignait la goutte, les rhumatismes,

l'otite ou les convulsions<sup>6</sup>. On retrouve également le miel, dont on commence seulement à percevoir l'étendue des applications en médecine comme cicatrisant, antiseptique et antibactérien. Enfin, en Amérique latine, les femmes déposaient les escargots sur leur peau afin de profiter de leurs vertus adoucissantes et réparatrices<sup>7</sup> ; le mucus d'escargot est aujourd'hui intégré dans la composition de certains produits cosmétiques.

## Origine végétale

Les plantes ont aussi été largement utilisées pour leurs propriétés biologiques, un grand nombre de leurs molécules a été utilisé pour diverses applications thérapeutiques ou cosmétiques. On retrouve par exemple, le curare, une substance extraite de certaines lianes d'Amazonie, qui était utilisé par les Amérindiens pour enduire les flèches de chasse afin de provoquer une paralysie des muscles de la proie<sup>8</sup>. Ou encore, l'extrait de belladone qui était instillé dans les yeux des Italiennes élégantes afin de les faire briller, en augmentant la pression oculaire et en dilatant la pupille sous l'action de l'atropine, et ainsi accroître leur attirance. Aujourd'hui, ce principe actif est utilisé par les ophtalmologues pour dilater les pupilles lors d'un fond d'œil<sup>9</sup>.

La compréhension et l'identification des composés responsables des propriétés biologiques diverses et variées des organismes vivants, observées et utilisées par nos ancêtres, est rendue possible grâce à la chimie des produits naturels. En effet, l'étude chimique avec des outils modernes de différents extraits a conduit à l'identification structurale de leurs principes actifs, à la démonstration de leur activité biologique et à la mise au point de nouveaux produits cosmétiques et thérapeutiques<sup>10</sup>.

## C. Chimie des produits naturels

Le concept de la chimie moderne des produits naturels remonte au XVIII<sup>e</sup> siècle. Elle était considérée comme la chimie des substances qui composent les plantes et les animaux, dont les principes avaient été établis en 1789 par le français Antoine Lavoisier. Elle a permis au-delà de l'utilisation des extraits, l'évolution des recherches vers la détermination des composés chimiques actifs issus des plantes. Néanmoins, l'intérêt de la chimie des produits naturels a connu un véritable déclin en raison d'un manque d'outils analytiques adaptés, de programmes universitaires et industriels dédiés, ainsi que la faiblesse des ressources financières de la recherche et du développement de nouvelles découvertes<sup>11</sup>.

Au XIX<sup>e</sup> siècle, un regain d'intérêt a eu lieu pour les produits naturels en raison de leur impact sur divers domaines de la science, des progrès techniques et de l'activité économique. La compréhension des produits naturels et de leur chimie conduit au développement des approches de purification, d'analyse, de détermination de structure et de fonction, de dynamique chimique et métabolique interrelationnelle. Ainsi une meilleure compréhension des

produits plus complexes et de leurs implications dans des utilisations analytiques, de bio-ingénierie et médicales a été atteinte. Le Tableau 1 présente quelques avancées offertes par le développement de la chimie des produits naturels dans divers domaines<sup>11</sup>.

| <b>Domaines</b>                          | <b>Avancées offertes par le développement de la chimie des produits naturels</b>  |
|--|---|
| Thérapeutique                            | <p>Taxol, céphalosporine, pénicilline, tétracycline, camptothécine, etoposide, podophyllotoxine, combretastatine, analogues d'homoharringtonine, bryostatine d'origine marine, ecteinascidine, kahalalide F, etc.</p> <p>Activités : neuro-active, cardiovasculaire, anti-hyperlipidémie, hépato-protecteur, anti-inflammatoire, anticancer, antidiabétique, analgésique, antispasmodique, antiulcéreux.</p> <p>Développement de l'aromathérapie, homéopathie et systèmes Siddha d'usage médicinal.</p> |
| Cosmétique                               | <p>Acide hyaluronique, caféine, urée, glycérine, tocophérol, acide ascorbique, huiles essentielles, etc.</p> <p>Activités : antioxydant, anti-enzymatique, antimicrobien, antibactérien, cicatrisant, anti-inflammatoire, etc.</p>  |
| Contrôle qualité et réglementation       | Développement de la Pharmacopée, traités médicaux et médecines alternatives (par exemple, la médecine unani ou l'ayurveda comme principaux systèmes de guérison à base de plantes).   |
| Chimie de synthèse                       | Approche rétrospective de la synthèse biogénétique, synthèse totale selon un schéma biomimétique, nouvelles voies, nouveaux modèles et disponibilité de structures moléculaires complexes, hémisynthèse.  |
| Agroalimentaire/nutraceutique            | Colorants végétaux, acides gras et lipides, molécules aromatiques, épices, condiments, épaississants et texturants pectines, alginates, cellulose, etc.   |
| Environnement                            | Insectifuge, herbicide, pesticide naturel, larvicide, auxine, phytoalexine, antihelminthique, filaricide.   |
| Produits techniques, fibres et polymères | Fibres de jute, papier, carton, polymères d'origine marine, etc.  |
| Économie                                 | Développement de l'économie mondiale, croissance économique dans les pays riches en ressources végétales,   |

|  |   |
|--|---|
|  | importations mondiales, consommation indigène interne, économies forestières, subsistance durable, etc. |
|--|---|

Tableau 1 : Les contributions et avancées offertes par la chimie des produits naturels présenté par R.A Khan (2018)<sup>11</sup>

Ainsi, c'est grâce aux métabolites secondaires provenant des plantes que la chimie des produits naturels a pris de l'ampleur et a offert des contributions et les progrès importants.

#### D. Métabolites secondaires ou spécialisés végétaux

À ce jour, plus de 250 000 métabolites secondaires ont été répertoriés dans le dictionnaire des produits naturels<sup>12</sup>. Ils sont élaborés à partir des métabolites primaires par des voies métaboliques génétiquement contrôlées, et leur production dépend de nombreux facteurs extrinsèques et intrinsèques. Les métabolites secondaires végétaux sont classés selon leur biogénèse ou biosynthèse ; on retrouve trois grands groupes chez les plantes : les composés phénoliques, les composés azotés et les composés terpénoïdes. L'ensemble des métabolites primaires et de ces grandes classes de métabolites secondaires représente **le métabolome de la plante**<sup>13</sup>.

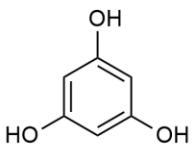
##### 1) Composés phénoliques

Ce sont les métabolites secondaires les plus largement distribués, ils sont omniprésents dans le règne végétal, même si le type de composés présents varie selon l'espèce considérée<sup>14</sup>. C'est un groupe qui est difficile à définir, ces composés contiennent au moins un noyau benzénique porteur d'un groupement hydroxyle libre ou engagé dans une autre liaison (éther, ester, hétéroside, etc.). Ce groupe comprend des familles provenant de deux voies biogénétiques d'aromagénèse : les phénols simples, coumarines, lignanes, stilbènes, chalcones, flavonoïdes, xanthones et tanins provenant de l'acide shikimique ainsi que les quinones et orcinols provenant de la voie de l'acétate<sup>15</sup> (Figure 2).

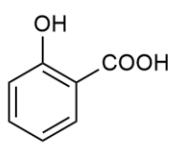
Les composés phénoliques sont considérés comme jouant un rôle clé en tant que composés de défense lorsque des stress environnementaux, tels que la lumière intense, les basses températures, l'infection par des agents pathogènes, les herbivores et la carence en nutriments, peuvent entraîner une production accrue de radicaux libres et d'autres espèces oxydatives dans les plantes. Ils sont principalement utilisés pour leurs effets bénéfiques sur la santé comme antioxydants, et peuvent être ajoutés comme additifs dans les produits des industries agroalimentaires, pharmaceutiques et cosmétiques. Il existe notamment certains produits médicinaux à base de plantes où les principes actifs peuvent être présents sous forme de promédicaments et agir en synergie<sup>16</sup>. Par exemple, les polyphénols ont donné naissance à des médicaments intéressants et la FDA (*Food and Drug Administration*) américaine a approuvé un extrait de polyphénols de thé vert enrichi (Mitsui Norin, Polyphenon

E) en tant que pommade pour le traitement des verrues génitales<sup>17</sup>. On retrouve également le resvératrol, un stilbène extrait de la vigne et utilisé dans des produits cosmétiques pour lutter contre le vieillissement cellulaire (Caudalie, Resvératrol-lift).

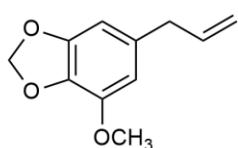
### Phénols simples



Phloroglucinol

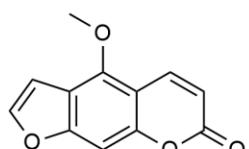


Acide salicylique



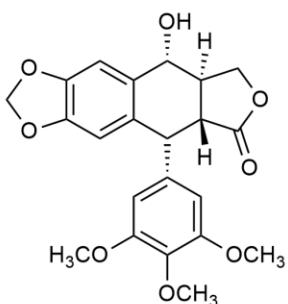
Myristicine

### Coumarines



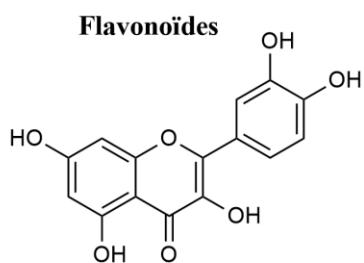
Bergaptène

### Lignanes



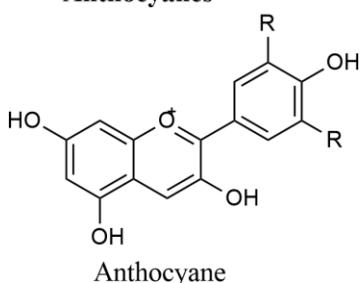
Podophyllotoxine

### Flavonoïdes



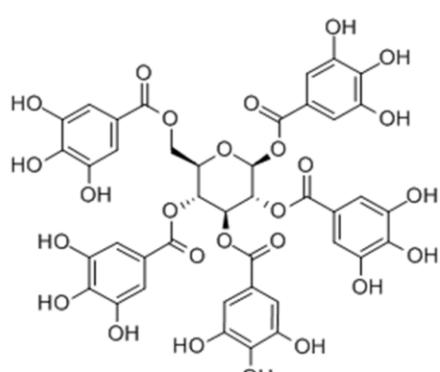
Quercétol

### Anthocyanes



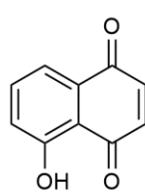
Anthocyane

### Tanins

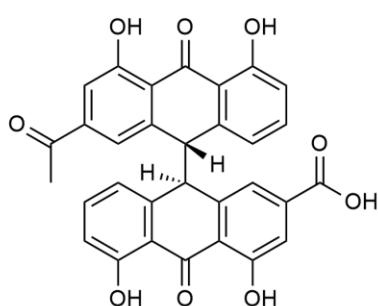


Beta-penta-O-galloyl-glucose

### Quinones

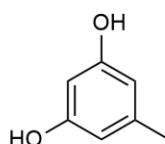


Juglone



Sennidine A

### Orcinols



Orcinol

Figure 2 : Quelques composés appartenant à la famille des composés phénoliques

## 2) Composés azotés

Ce groupe comprend les alcaloïdes, les betalaïne, les hétérosides cyanogènes et les glucosinolates. Néanmoins, les alcaloïdes, principaux composants de ce groupe, sont des substances azotées, basiques et hétérocycliques (Figure 3).

Selon l'acide aminé et métabolite primaire précurseur, on retrouve dans la famille des alcaloïdes des dérivés de l'ornithine, de la lysine, de l'acide nicotinique, de la phénylalanine, de la tyrosine, du tryptophane, de l'histidine, etc. Cette famille contient d'importants métabolites secondaires avec des propriétés pharmacologiques d'intérêt thérapeutique, il en existe environ 12 000 qui seraient répertoriés à ce jour. Parmi les exemples les plus remarquables des produits naturels végétaux présents dans la pharmacopée actuelle et dans la liste des médicaments essentiels de l'organisation mondiale de la santé, figurent les vinca alcaloïdes (vincristine, vinblastine) isolés de la pervenche de Madagascar *catharanthus roseus*.

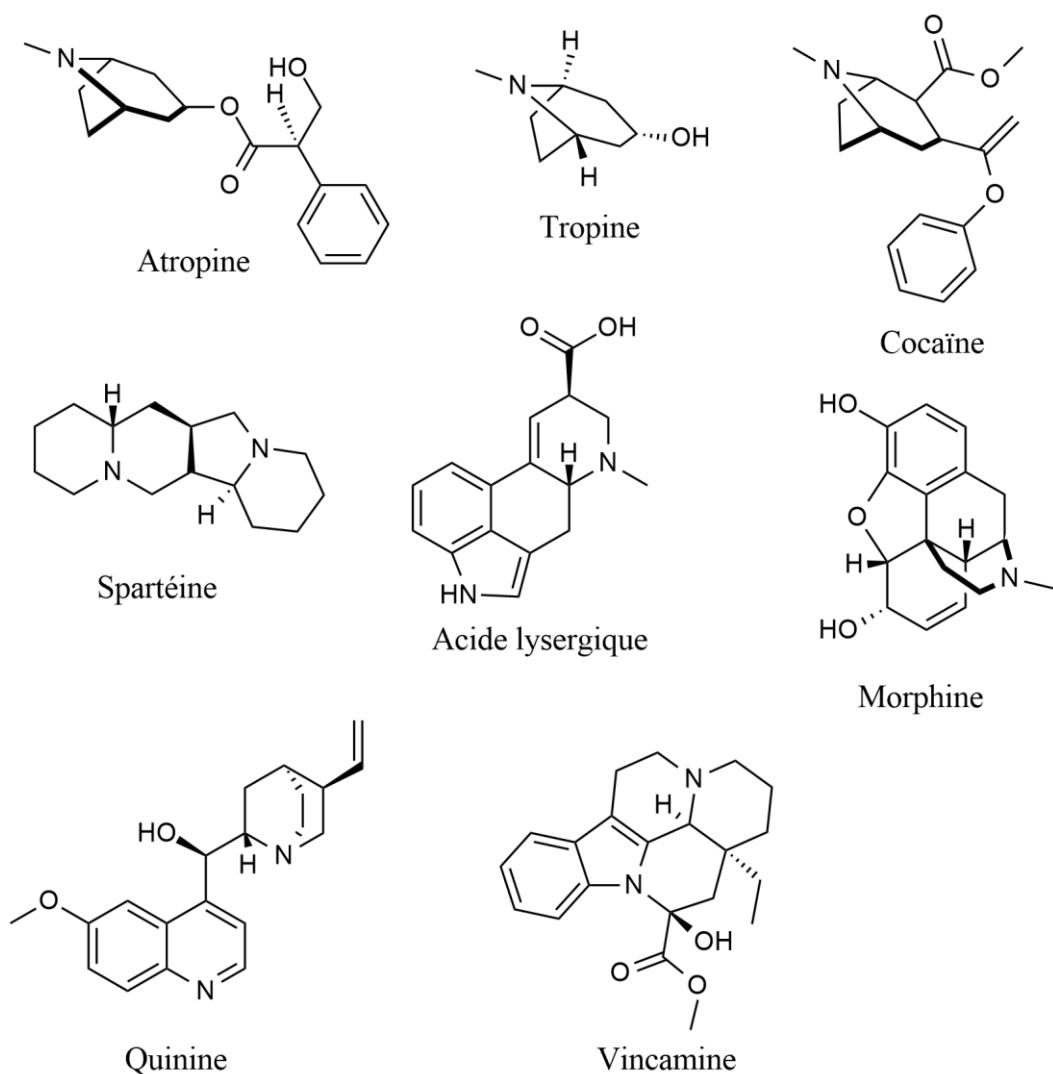


Figure 3 : Quelques composés appartenant à la famille des alcaloïdes

### **3) Composés terpéniques**

C'est le groupe le plus vaste des métabolites secondaires. Les composés sont formés par assemblage d'un nombre entier d'unités d'isoprène. On retrouve dans cette classe les monoterpènes C<sub>10</sub> (2 unités), les sesquiterpènes C<sub>15</sub> (3 unités), les diterpènes C<sub>20</sub> (4 unités), les triterpènes C<sub>30</sub> (6 unités) dont les saponines, les stéroïdes C<sub>27</sub> (6 unités – 3C) et les tétraterpènes C<sub>40</sub> (8 unités) (Figure 4).

Les terpènes volatils sont souvent responsables de l'odeur des végétaux et sont extraits sous forme d'huiles essentielles pour la parfumerie. Certains terpènes ont des rôles biologiques importants (hormones, vitamines). Parmi les molécules bioactives appartenant à cette famille, on retrouve l'artémisinine, une lactone sesquiterpénique contenant un peroxyde. Isolé d'*Artemisia annua*, ce composé est devenu un médicament de référence pour le traitement du paludisme.

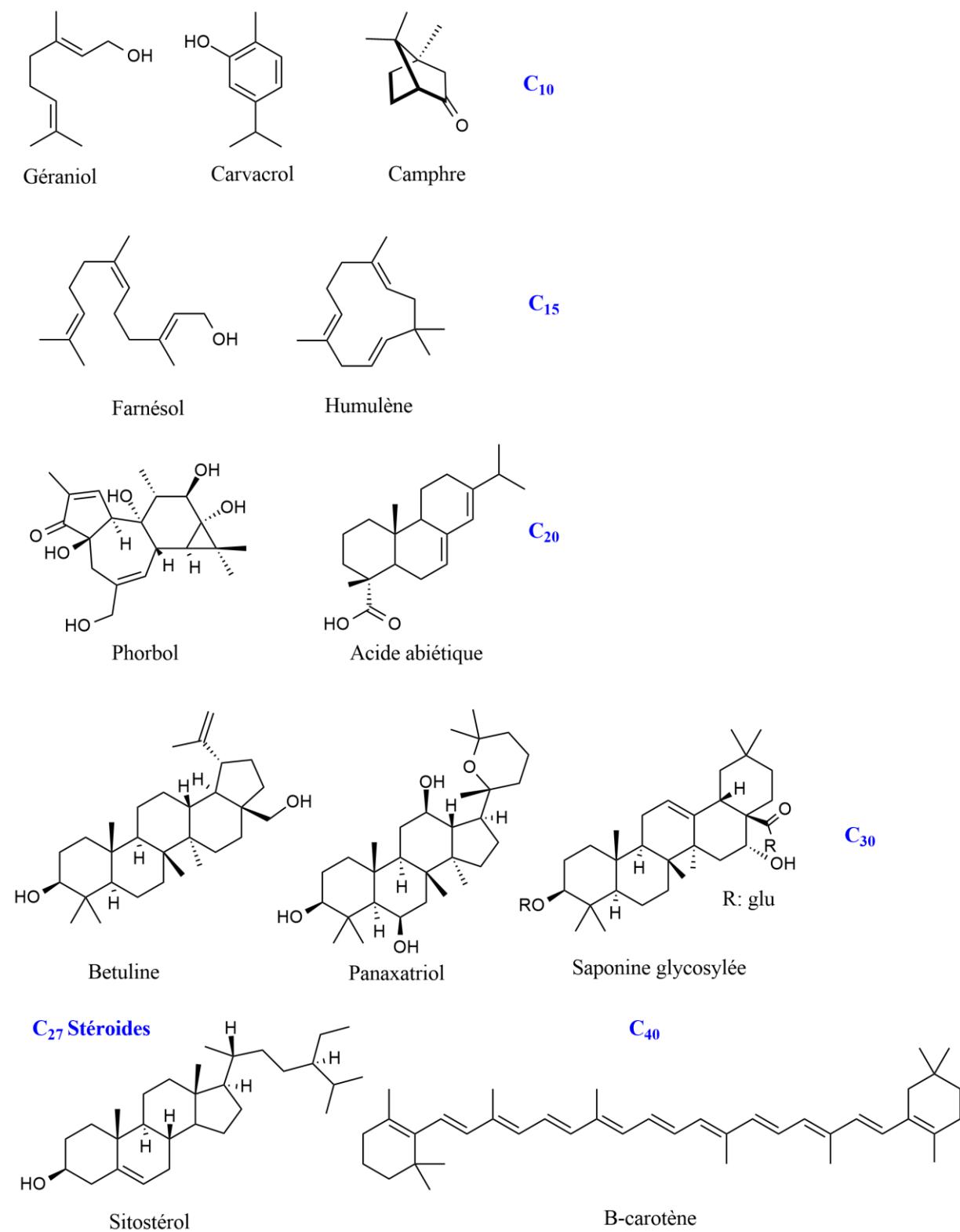


Figure 4 : Quelques composés appartenant à la famille des terpènes

Les métabolites secondaires d'origine végétale représentent une source précieuse et un réservoir de molécules bioactives pour les industries cosmétiques et pharmaceutiques. Aujourd'hui, environ un tiers des médicaments sur le marché sont des structures naturelles d'origine végétale ou des composés dérivés, et les fabricants cosmétiques ont de plus en plus

recours aux ingrédients naturels (extrait végétal obtenu par solvant, huiles essentielles ou composé pur). Cependant, l'étude du métabolome des plantes reste extrêmement difficile compte tenu de la diversité et de la complexité structurale, du grand nombre et du contenu en métabolites secondaires. Ce profil métabolique dépend largement de l'espèce, des conditions de culture et de l'environnement. Par conséquent, le développement de nouvelles techniques pour la caractérisation, la quantification, l'identification et la visualisation d'informations structurales est indispensable afin de continuer à progresser dans la découverte de nouvelles molécules naturelles d'intérêt, et pour suivre l'impact que peut avoir le changement de certaines conditions environnementales sur la production de ces molécules par la plante.

## **II. PROFILAGE DES MÉTABOLITES ET MÉTABOLOMIQUE**

Jusqu'à récemment, la plupart des métabolites végétaux ont été analysés et détectés à l'aide de méthodes ciblées à des fins très spécifiques, notamment l'évaluation de la diversité des métabolites dans certaines espèces végétales, la quantification des métabolites bioactifs des plantes médicinales et l'évaluation des variations chimiques dans certains environnements et conditions.

Au cours des deux dernières décennies, la biologie des systèmes, des sciences "omiques", plus particulièrement de la métabolomique et des méthodes de profilage des métabolites ont considérablement évolué et peuvent fournir une quantité importante d'informations sur la composition chimique des plantes.

### **A. Définitions**

#### **1) Profilage des métabolites**

Le profilage des métabolites des plantes est la mesure des métabolites de faible poids moléculaire (< 1000 Da) et de leurs intermédiaires, qui reflète la réponse dynamique à la modification génétique (par exemple, transgénique ou virale) ainsi qu'aux stimuli physiopathologiques (par exemple, la morbidité de la maladie) et/ou développementaux (par exemple, le vieillissement). Cette approche permet l'identification des composés appartenant à une voie métabolique ou à une famille chimique donnée. La mesure et l'interprétation du profil des métabolites endogènes, à partir d'un échantillon de plante, ont fourni de nombreuses occasions d'étudier les changements induits par des stimuli externes (par exemple, le traitement d'une plante) ou d'améliorer notre connaissance de la variation biologique inhérente au sein des sous-populations (le phénotype métabolique)<sup>18</sup>.

Le concept de profilage des métabolites n'est pas nouveau<sup>19</sup>. En 1971, Horning et Horning ont introduit pour la première fois le concept de profilage métabolique par spectrométrie de masse<sup>20</sup>, et cette même année, Pauling *et al.* ont rapporté des analyses de vapeur d'urine et de souffle par chromatographie en phase gazeuse et ont mis en relation les profils obtenus avec les effets d'un régime alimentaire défini<sup>21</sup>. Le profilage des métabolites est ainsi l'approche la plus ancienne et la plus établie. Depuis lors, de nombreuses méthodes analytiques puissantes, tant chromatographiques que spectroscopiques, ont été développées ou améliorées<sup>22</sup>. Les analyses de métabolites dans divers types de matrices sont devenues nécessaires dans plusieurs domaines de recherche. Cette approche est considérée comme le précurseur de la métabolomique.

## 2) Métabolomique

Le terme métabolomique a fait son apparition avec d'autres technologies "omiques" dans la recherche biologique (génomique, transcriptomique, protéomique et métabolomique)<sup>23</sup> en donnant une vue d'ensemble de l'état biochimique d'un organisme qui peut être utilisé pour surveiller les variations significatives des métabolites. En effet, comme les métabolites sont les produits finaux des processus de régulation cellulaire, leurs niveaux peuvent être considérés comme la réponse ultime des systèmes biologiques aux changements génétiques ou environnementaux. Contrairement au profilage des métabolites qui s'intéresse aux composés d'une seule voie métabolique, cette approche permet une identification et une quantification de tous les métabolites présents dans un échantillon biologique. Deux types d'analyse peuvent alors être envisagés : l'analyse globale ou non ciblée, qui s'intéresse à tous les composés de l'échantillon sans sélection préalable ; et l'analyse ciblée qui se focalise sur un nombre limité de métabolites connus, appartenant à une ou plusieurs familles chimiques, et prédéfinis avant l'analyse afin de déterminer leurs concentrations précises.

Ces deux stratégies (profilage des métabolites et métabolomique), qui peuvent être complémentaires<sup>24</sup>, sont utilisées soit pour une étude complète de la composition chimique d'un extrait de plante d'intérêt, soit en combinaison avec une analyse avancée des données multivariées dans la recherche de biomarqueurs responsables de la discrimination entre des extraits de plantes. Lorsque ces analyses statistiques sont effectuées sur un nombre suffisant de répliques biologiques, elles permettent de discriminer et de classer les échantillons en groupes et de suivre les modifications de la composition du métabolome qui sont liées à un état physiologique donné, à l'influence d'un stress ou d'un stimulus, à la modification génétique, à l'interaction avec d'autres organismes, etc.

## B. Techniques analytiques

Afin de pouvoir d'établir le profil de tous les métabolites des extraits de plantes, la méthode analytique choisie doit être capable de détecter un large éventail de métabolites sur une large gamme dynamique, avec une sensibilité, une résolution et une répétabilité élevées. En outre, une telle méthode devrait fournir des données spectroscopiques qualitatives pour l'identification de biomarqueurs et des informations quantitatives sur les modifications des métabolomes observées. Comme il n'existe actuellement aucune méthode analytique unique pouvant fournir toutes les informations nécessaires, deux approches principales sont utilisées : la Résonance Magnétique Nucléaire (NMR)<sup>25</sup> et la Spectrométrie de Masse (MS)<sup>26</sup>. Étant donné la composition complexe des extraits de plantes et le fait qu'ils contiennent souvent de nombreuses structures isomériques, les métabolites sont souvent séparés par une technique séparative avant leur détection. À cet égard, les méthodes de chromatographie couplées à la

MS, telles que la chromatographie en phase liquide LC-MS<sup>27</sup> et la chromatographie en phase gazeuse GC-MS<sup>28</sup>, sont largement utilisées. D'autres techniques séparatives ont également été explorées comme l'électrophorèse capillaire CE-MS et la chromatographie en fluide supercritique SFC-MS. Néanmoins, le couplage CE-MS présente des limites de sensibilité, de reproductibilité et de robustesse<sup>29</sup>, tandis que les couplages SFC-MS et GC-MS présentent des limites de flexibilité du solvant de dissolution des échantillons pour les composés polaires<sup>30</sup>. Par conséquent, le profilage des métabolites et la métabolomique pour les composés non volatils reposent principalement sur le couplage LC-MS qui est bien adapté à l'analyse des matrices biologiques complexes telles que les extraits de plantes<sup>31</sup>. Le couplage LC-NMR est également possible et permet une détermination de structure sans ambiguïté des composés majoritaires. Toutefois il est moins répandu que le couplage LC-MS car il présente davantage de contraintes instrumentales et une moins bonne sensibilité<sup>32</sup>.

Il existe de nombreuses techniques séparatives et de nombreux types de spectromètres de masse ; seuls les principes de fonctionnement de ceux utilisés au cours de cette thèse seront détaillés dans ce mémoire.

## 1) Chromatographie en Phase-Liquide

La chromatographie de partage en phase liquide est une technique analytique qui permet de séparer des constituants d'un mélange, et dont le principe repose sur l'équilibre de concentrations de composés présents entre deux phases non miscibles en contact : la phase stationnaire (emprisonnée dans la colonne) et la phase mobile, liquide, qui se déplace. La séparation est basée sur l'entrainement des constituants présents dans la colonne, qui se déplacent à des vitesses différentes selon leur affinité pour la phase stationnaire et la phase mobile, dépendant de leurs propriétés intrinsèques (structure, polarité, etc.). À la sortie de la colonne, un détecteur permet de visualiser en continu le passage des solutés selon différentes propriétés physico-chimiques (absorbance, ionisation, etc.). La réponse donnée par le détecteur peut également être dépendante de la quantité de chacun des constituants du mélange<sup>33</sup>.

La chromatographie peut être analytique, en ayant pour objectif de séparer les différentes molécules en vue de les identifier qualitativement et/ou de les quantifier, ou préparative, avec pour objectif de purifier un produit ou de fractionner un mélange de composés. Le produit d'une séparation préparative peut ensuite être utilisé pour effectuer d'autres études telles que des tests d'activités biologiques, de formulation, de synthèse, de caractérisation, etc.

Les grandeurs utilisées en chromatographie pour caractériser une méthode de séparation concernent essentiellement la colonne<sup>34</sup>.

- i) **Le facteur de capacité ou de rétention  $k'$**  : permet de s'affranchir des paramètres géométriques de la colonne pour exprimer le temps de rétention d'un analyte de façon relative par rapport au temps mort de la colonne.
- ii) **La sélectivité  $\alpha$**  : permet de caractériser la distance qui sépare le sommet de deux pics chromatographiques consécutifs.
- iii) **L'efficacité  $N$**  : caractérisée par le nombre de plateaux théoriques. Afin de pouvoir comparer les efficacités de séparation de colonnes de longueurs  $L$  différentes, on définit la hauteur équivalente à un plateau théorique  $H$ .
- iv) **La résolution RS** : mesure la séparation qui existe entre deux pics chromatographiques. Plus la résolution est grande, meilleure est la séparation. Si les pics chromatographiques sont gaussiens, la séparation est pratiquement complète pour une résolution de 1,5.
- v) **La perte de charge  $\Delta P$**  : la loi de Darcy exprime la perte de charge d'une colonne chromatographique, et donc sa résistance à l'écoulement.

La technique HPLC (*High-Performance Liquid Chromatography*) est reconnue depuis le début des années 1980 comme la plus polyvalente pour la séparation efficace et directe des produits naturels à partir de mélanges bruts, sans nécessité de préparation d'échantillon complexe<sup>35</sup>. Depuis lors, les techniques de séparation des produits naturels se sont considérablement améliorées en termes de commodité, de rapidité, de choix des phases stationnaires de la colonne, de sensibilité, d'applicabilité à une variété de matrices d'échantillons et d'utilisation de détecteurs spectroscopiques<sup>36,37</sup>.

Dans le profilage des métabolites, il y a un fort besoin de haute résolution chromatographique afin de déconvoluer un métabolite de la matrice complexe associée. Des colonnes HPLC avec une technologie "Fused-Core" (SPP pour *Superficialy Porous Particles* ou core-shell) sont utilisées afin d'améliorer la résolution. Elles ont été imaginées par Horvàth en 1967 et largement développées par Kirkland qui leur a donné le nom de "Fused-Core"<sup>38</sup>.

Ces colonnes contiennent une couche poreuse soudée à la surface d'un noyau solide résultant en un chemin de diffusion plus court comparé à celui des colonnes remplies de particules totalement poreuses (FPP *Fully Porous Particles*). La dispersion axiale de l'échantillon est ainsi réduite et l'élargissement des pics minimisé. Ces colonnes permettent de meilleures efficacités et de meilleures résolutions que les colonnes HPLC tout en conservant des pressions de travail compatibles avec l'utilisation de chaines HPLC (maximum 400 bar).

La réduction de la taille des particules de la colonne chromatographique permet également une plus grande efficacité de séparation (nombre de plateaux plus élevé) et un temps d'analyse plus court<sup>39</sup>. Par conséquent, l'introduction de systèmes de chromatographie liquide

à ultra-haute pression UHPLC (*Ultra High-Performance Liquid Chromatography*) fonctionnant à des pressions très élevées (600-1500 bar) et utilisant des colonnes de granulométries inférieures à 2 µm a permis une diminution remarquable des temps d'analyse et une augmentation de la capacité de pics, de la sensibilité et de la reproductibilité par rapport à l'HPLC classique<sup>40</sup>. Des exemples d'application montrent qu'il est possible de réduire le temps d'analyse d'un facteur 9 par rapport à l'HPLC en sélectionnant la longueur de colonne appropriée<sup>41</sup>. De telles séparations ultra-rapides ont été démontrées expérimentalement pour des extraits d'origine naturelle, avec des temps d'analyse allant de 1 à 5 min<sup>42</sup>. À titre d'exemple, l'amélioration de la résolution obtenue pour le profilage des métabolites (flavonoïdes) dans un extrait standardisé de *Ginkgo biloba* est présentée sur la Figure 5<sup>43</sup>. Comme illustré, le profilage effectué sur une colonne HPLC avec des particules de 5 µm peut être considérablement amélioré lorsque le même échantillon est analysé sur une colonne UHPLC avec des particules de 1,7 µm. La comparaison de (a) et (b) illustre le gain de résolution (facteur d'environ 3) lorsqu'un temps de gradient similaire est appliqué entre l'HPLC classique et l'UHPLC. En effet, pour un même temps d'analyse, on observe la séparation d'un plus grand nombre de composés qui résulte du gain de résolution, ce qui permet de mettre en évidence des composés minoritaires non détectées sur une colonne HPLC à cause des coélutions. La comparaison entre (a) et (c) illustre le gain de temps pour une résolution similaire ; le temps d'analyse est divisé par 9, soit 5 min au lieu de 45 min.

Cependant, le choix de la nature de la phase stationnaire est fortement dépendant de la polarité et des propriétés physico-chimiques des métabolites qui seront analysés, pour cela, un large panel de phases stationnaires existe avec des diamètres de particules allant de 5 à 1,3 µm.

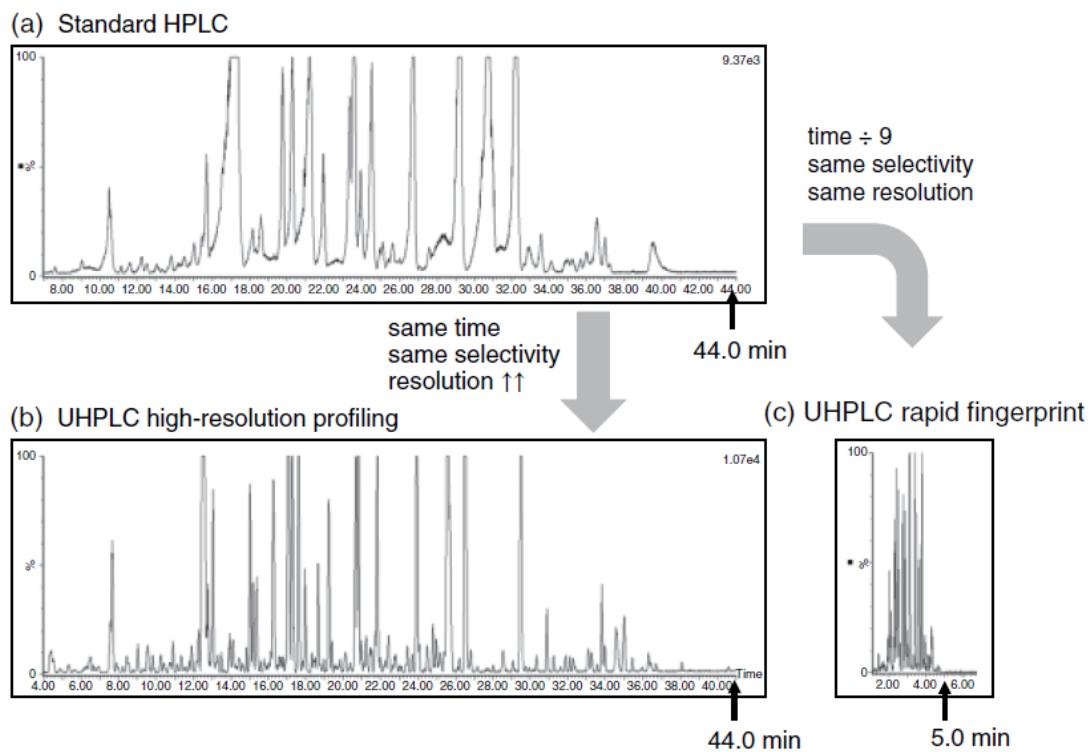


Figure 5 : Profilage des métabolites obtenu pour un extrait standardisé de *Ginkgo biloba* présenté par Wolfender et al. (2019)<sup>43</sup>. a) Colonne Xbridge BEH C18 (150 x 4,6 mm ; 5 µm). b) Colonne Acquity BEH C18 (150 x 2,1 mm ; 1,7 µm). c) Colonne Acquity BEH C18 (50 x 2,1 mm ; 1,7 µm)

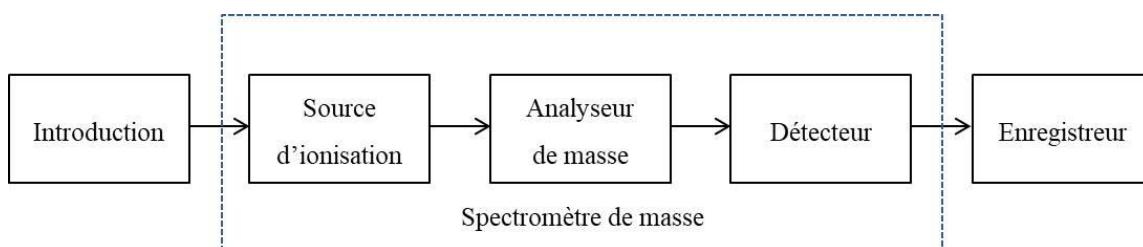
Cette technologie a rapidement été adoptée par la communauté analytique et est largement appliquée à divers domaines de l'analyse des plantes<sup>60</sup>, en particulier la métabolomique<sup>44</sup>.

## 2) Spectrométrie de masse

Un spectromètre de masse permet de caractériser les analytes en mesurant le rapport masse sur charge ( $m/z$ ) de leurs ions et de leurs produits de fragmentation.

Tous les spectromètres de masse possèdent trois éléments en commun : une source, un analyseur et un détecteur comme illustré sur la Figure 6.

Les molécules sont introduites dans la source où sont formés les ions. Ces derniers sont guidés vers l'analyseur où ils sont séparés, puis vers le détecteur où ils sont captés. Les spectres sont enregistrés à l'aide d'un logiciel qui permet également d'assurer le pilotage du spectromètre<sup>45</sup>.



*Figure 6 : Les éléments communs des spectromètres de masse*

La MS a joué un rôle très important dans la progression de la métabolomique car elle permet la détection sensible de la plupart des métabolites de plantes<sup>46</sup>. Dans les applications métabolomiques, deux étapes sont essentielles : l'ionisation des métabolites (la méthode d'ionisation doit être choisie pour observer un maximum de molécules avec des *m/z* caractéristiques dans un échantillon donné) et la séparation des ions par l'analyseur de masse (des instruments MS à basse ou haute résolution peuvent être utilisés et influenceront fortement le nombre de structures détectées). Contrairement à la NMR, la détection par MS n'est pas universelle, et l'ionisation est fortement dépendante d'une part du composé et de la matrice dans laquelle il est analysé et d'autre part de la source et du mode d'ionisation choisis. Ainsi, les intensités des ions observés ne peuvent pas être directement reliées à la concentration des analytes et nécessitent l'utilisation d'étalons pour obtenir des informations quantitatives.

**a) Ionisation**

L'étape d'ionisation permet d'obtenir des ions à l'état gazeux à partir des molécules introduites dans la source, partie du spectromètre de masse qui a pour rôle de volatiliser et d'ioniser. La nature de la source utilisée dépend des caractéristiques physico-chimiques des molécules à analyser.

Il existe des sources d'ionisation douces ou dures. Les conditions douces engendrent des ions moléculaires relativement stables, avec des durées de vie suffisantes pour traverser l'analyseur et parvenir au détecteur. Ce type d'ionisation produit peu ou pas de fragments, tandis que les conditions dures engendrent souvent des ions moléculaires qui se fragmentent beaucoup, voire totalement, avant la sortie de la source. Dans ce cas, les fragments peuvent être analysés et donnernt des informations structurales<sup>47</sup>.

Les principales sources d'ionisation sont présentées dans le Tableau 2 ci-dessous. Cette liste permet d'avoir un aperçu de la grande diversité des méthodes d'ionisation qui ont été développées pour satisfaire à tous types d'analyses<sup>48</sup>.

| Source d'ionisation  | Douces | Assez douces | Dures |
|--|--------|--------------|-------|
| ICP ( <i>Inductively Coupled Plasma</i> )                    |        |              | X     |
| EI ( <i>Electron Ionization</i> )                            |        |              | X     |
| CI ( <i>Chemical Ionization</i> )                            |        | X            |       |
| ESI ( <i>ElectroSpray Ionization</i> )                       | X      |              |       |
| APPI ( <i>Atmospheric Pressure Photolionization</i> )        |        | X            |       |
| APCI ( <i>Atmospheric Pressure Chemical Ionisation</i> )     |        | X            |       |
| SIMS ( <i>Secondary Ions Mass Spectrometry</i> )             |        | X            |       |
| MALDI ( <i>Matrix Assisted Laser Desorption Ionization</i> ) | X      |              |       |
| LDI ( <i>Laser Desorption Ionization</i> )                   |        | X            |       |
| LSI ( <i>Liquid Secondary Ionization</i> )                   |        | X            |       |

Tableau 2 : Principales sources d'ionisation en spectrométrie de masse

La majorité des études métabolomiques utilisant la MS est basée sur l'ionisation par électrospray (ESI) car c'est une source à pression atmosphérique qui permet l'ionisation d'un large éventail d'analytes et est compatible avec une introduction de l'échantillon par chromatographie en phase liquide. En ESI, la protonation/déprotonation des molécules est la principale source de charge. Différents types d'adduits avec des sels ( $\text{Na}^+$ ,  $\text{K}^+$ ) ou des molécules de solvant (MeOH, HCOOH) peuvent également être générés selon la nature des analytes. Il est important de noter que l'ionisation de l'analyte dépend largement de sa structure. En bonne approximation, les molécules acides (par exemple, les acides carboxyliques ou phénoliques) produiront principalement des ions  $[\text{M}-\text{H}]^-$  en ionisation négative (NI), tandis que les bases (par exemple, les alcaloïdes, les amines) produiront facilement des ions  $[\text{M}+\text{H}]^+$  en ionisation positive (PI). Ainsi, pour étudier un nombre maximal de métabolites, il est souvent nécessaire d'utiliser les modes d'ionisation PI et NI<sup>27</sup>.

La Figure 7, présente un profil typique obtenu en ESI dans les modes PI et NI pour un extrait brut de feuilles de maïs<sup>49</sup>. La comparaison des deux chromatogrammes d'ions totaux (TIC) montre la complémentarité de la détection des deux modes d'ionisation. Les panneaux de droite montrent les spectres de masse PI et NI extraits du pic à  $t_R = 8$  min qui mettent en évidence plusieurs adduits rencontrés pour le DIM2BOA-Glc, un benzoxazinoïde connu.

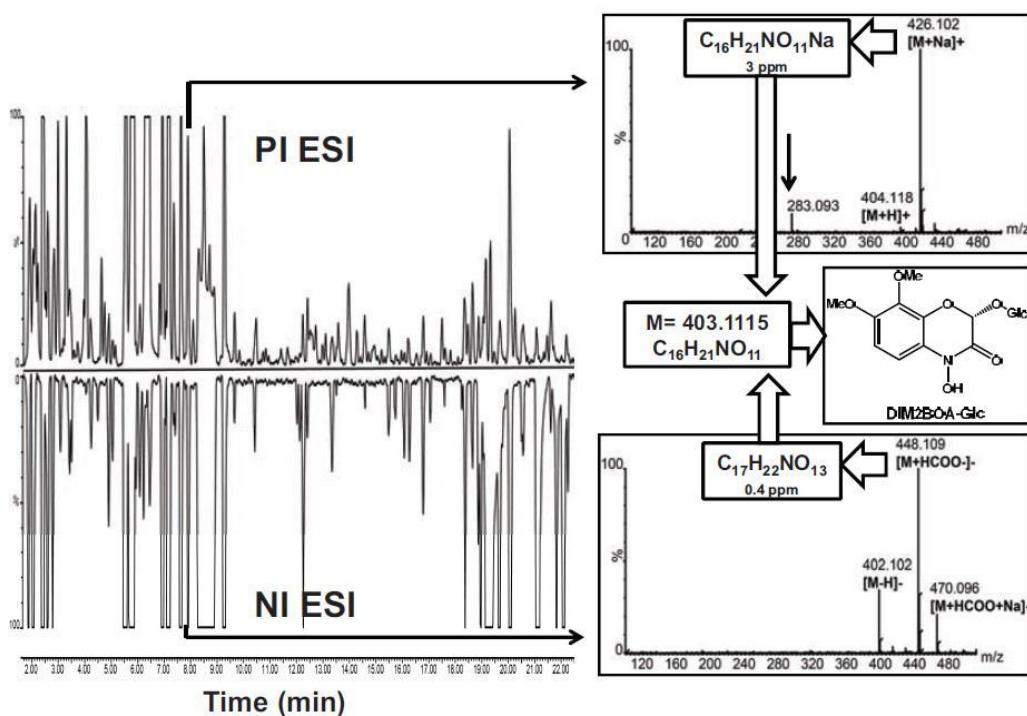


Figure 7 : Profilage UHPLC-HRMS obtenu avec l'ionisation ESI en modes positif (PI) et négatif (NI) d'un extrait brut de feuilles de maïs effectué par G. Marti et al. (2012)<sup>49</sup>

Même si la plupart des composés ont été ionisés à la fois en mode PI et NI, certains n'ont été détectés que dans une seule polarité. Les spectres de masse du composé DIM2BOA-Glc montrent la diversité des espèces d'ions qui peuvent être enregistrées (ion de molécule protonée ou déprotonnée, adduit, fragment, multichargé). Ainsi, une combinaison des informations qualitatives obtenues dans les deux modes est très utile pour déterminer la masse moléculaire [M] d'un analyte inconnu détecté.

Néanmoins, cette méthode d'ionisation subit d'importants effets de suppression ionique qui peuvent limiter la reproductibilité, l'exactitude et la sensibilité des études. Afin de diminuer l'impact de la suppression des ions sur les données acquises, il est important de développer la méthode avec soin, choisir une méthode appropriée de préparation des échantillons, choisir les bons solvants et additifs, utiliser des séparations chromatographiques à haute résolution et optimiser les paramètres de la source (température, tension, débit de gaz, etc.) pour optimiser l'ionisation.

### b) Séparation - Résolution

L'analyseur est la partie du spectromètre qui sépare les ions en fonction de leur rapport  $m/z$ . Il y règne un vide tel que le libre parcours moyen des ions soit supérieur à la distance à parcourir dans l'appareil pour atteindre le détecteur. La source est généralement reliée à l'analyseur par l'intermédiaire d'une interface. Cette interface est composée de champs

électrostatiques (lentilles et multipôles) qui permettent d'extraire les ions de la source et de les focaliser vers l'entrée de l'analyseur. Les champs électrostatiques appliqués peuvent être de l'ordre de quelques volts (quadripôles, pièges à ions), ou de plusieurs dizaines de kilovolts (secteurs magnétiques, temps de vol). Le rôle des lentilles est également de créer un gradient de vide entre les différentes parties du spectromètre.

Dans les approches de profilage de métabolites basées sur la spectrométrie de masse décrites, de nombreux types d'analyseurs peuvent être utilisés. Quelques-unes des principales caractéristiques de l'analyseur qui influencent la qualité des données obtenues sont :

- i) **La gamme de masse** : représente l'étendue de la gamme de valeurs  $m/z$  mesurable par le spectromètre.
- ii) **La vitesse d'acquisition** : détermine la vitesse à laquelle l'analyseur mesure une gamme de masse donnée.
- iii) **La transmission** : correspond au ratio entre le nombre d'ions atteignant le détecteur sur le nombre d'ions entrants réellement dans le spectromètre. Cette caractéristique prend en compte par exemple les ions perdus dans d'autres sections du spectromètre de masse telles que les lentilles. La transmission est associée à la sensibilité de l'instrument.
- iv) **La résolution** : correspond à la faculté d'un appareil à différencier des signaux distincts pour des ions possédant des valeurs  $m/z$  proches. Plus un pic est résolu, plus son centre peut être déterminé de manière fiable. Dans l'absolu, un pic large peut être correctement centré à la bonne valeur  $m/z$ , mais ce résultat sera forcément faussé si le pic observé contient un mélange de composés de valeurs  $m/z$  très proches. La résolution est donc un facteur essentiel pour évaluer la faculté d'un instrument à mesurer précisément une valeur  $m/z$ . Un instrument permettant une mesure en haute ou basse résolution doit être calibré à l'aide de composés de référence pour être centré à la bonne valeur  $m/z$  et donc fournir une précision de masse satisfaisante.
- v) **La précision de la masse** : indique l'exactitude de la mesure du rapport  $m/z$  que l'appareil fournit. Elle correspond à la déviation entre la valeur  $m/z$  théorique et celle qui est mesurée pour un ion donné. Cette précision peut être exprimée en valeur absolue (mDa) ou bien en valeur relative (partie par million - ppm) qui permet de prendre en considération la masse de l'ion dans le calcul. La précision de mesure de masse d'un composé issu d'un mélange dépend de la détermination correcte de la médiane du pic qui peut donc être largement influencée par la résolution du spectromètre. Une haute précision de masse permet différentes applications d'intérêt, notamment la détermination de formule brute.

- vi) **La précision spectrale :** indique la reproductibilité du modèle isotopique d'un ion donné. Elle permet notamment de comparer un massif isotopique théorique d'une formule brute (ion moléculaire ou adduit) avec un profil isotopique expérimental afin de vérifier la concordance.

En conséquence, deux principaux types d'analyseurs de masse peuvent être considérés : i) les spectromètres de masse à faible résolution (LR), tels que les quadripôles (Q) ou les piège à ions (IT)<sup>45</sup> et ii) les spectromètres de masse à haute résolution (HR) tels que les instruments à temps de vol (TOF)<sup>45</sup>, les Orbitraps<sup>50</sup> ou les spectromètres à résonance cyclotronique ionique (ICR)<sup>51</sup>.

Afin de détecter le plus grand nombre possible de structures et obtenir une image détaillée du métabolome de la plante, une grande sensibilité, une grande gamme dynamique et une haute résolution sont nécessaires. Dans le domaine de la métabolomique, des instruments HRMS plus coûteux et plus efficaces<sup>52</sup> sont utilisés pour augmenter la résolution des analytes dans la dimension MS.

Les études présentées dans ce mémoire ont été réalisées à l'aide du maXis Q-TOF illustré en Figure 8. Il a une résolution allant jusqu'à 40 000 à  $m/z$  1222 et peut être utilisé avec une fréquence d'acquisition élevée sur une large gamme de masse, compatible avec des méthodes chromatographiques rapides, sans compromettre la sensibilité<sup>53</sup>.

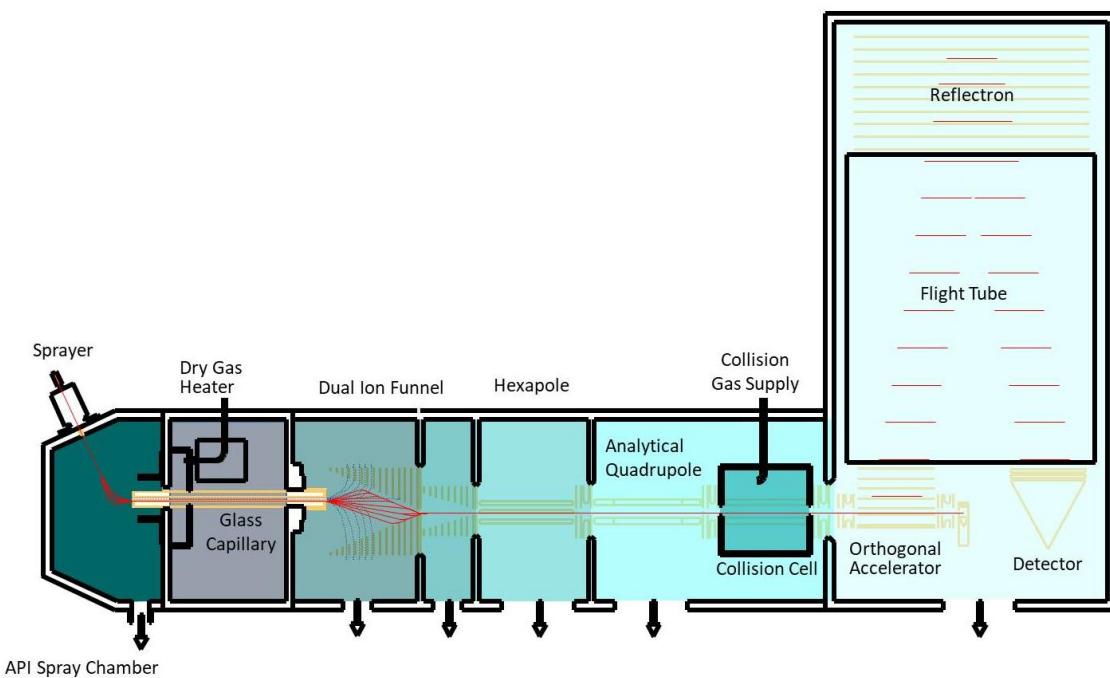


Figure 8 : Spectromètre de masse maXis UHR-TOF de Bruker

Lorsque deux analyseurs sont couplés pour agir de façon séquentielle, on parle de spectrométrie de masse à deux dimensions : MS/MS, ou spectrométrie de masse en tandem.

Un protocole de MS/MS comporte trois étapes : i) isolation d'un ion caractéristique de l'analyte dit ion précurseur, ii) fragmentation de cet ion par collisions avec un gaz inert et iii) détection des ions fragments obtenus.

### c) Spectrométrie de masse en tandem

Dans l'analyse non ciblée des métabolites, l'acquisition de spectres MS/MS pour tous les métabolites détectés est une condition préalable à la génération d'informations structurales complémentaires, nécessaires à l'établissement des formules développées<sup>54</sup>. La spectrométrie de masse en tandem (MS/MS) est en effet un outil expérimental inestimable pour fournir des données analytiques permettant de soutenir l'identification de petites molécules<sup>55</sup>. Les principales méthodes utilisées pour réaliser cette tâche sont l'acquisition dépendante des données (DDA) et l'acquisition indépendante des données (DIA)<sup>56</sup>.

Le mode d'acquisition DDA est utilisé pour effectuer la recherche de métabolites dont on ne connaît pas la structure. Il se déroule en deux étapes successives. Dans un premier temps le spectromètre acquiert un spectre en mode full scan MS. Dans un second temps il sélectionne automatiquement les n ions les plus intenses de chaque spectre et les fragmente par CID (*Collision-Induced Dissociation*) afin d'obtenir des spectres full scan MS/MS. De nombreux paramètres (tels que le seuil d'intensité pour la sélection de l'ion précurseur, son état de charge, le temps d'exclusion pendant lequel il ne sera pas sélectionné pour être de nouveau fragmenté, etc.) permettent d'ajuster la sélection et les conditions de fragmentation des ions à identifier. Cependant, ce mode d'acquisition est limité par le temps nécessaire à l'acquisition des spectres MS/MS par rapport à la largeur des pics chromatographiques. En fonction de la fréquence d'acquisition de l'instrument, le nombre de spectres MS/MS par cycle variera, et les ions de faible abondance dans les spectres MS ne seront pas sélectionnés comme précurseurs pour l'enregistrement de leurs spectres MS/MS, limitant ainsi la couverture de l'annotation des métabolites pour un échantillon donné. Toutefois, la plupart des annotations non ciblées sur les métabolomes sont actuellement réalisées à partir d'ensembles de données acquises par DDA, qui fournissent des spectres de bonne qualité avec un nombre d'ions fragmentés plus limité mais une meilleure association ion parent/ion fragment (Figure 9). Pour pallier à cette lacune de couverture d'échantillon, le mode DIA a été développé.

## DDA – data dependent acquisition

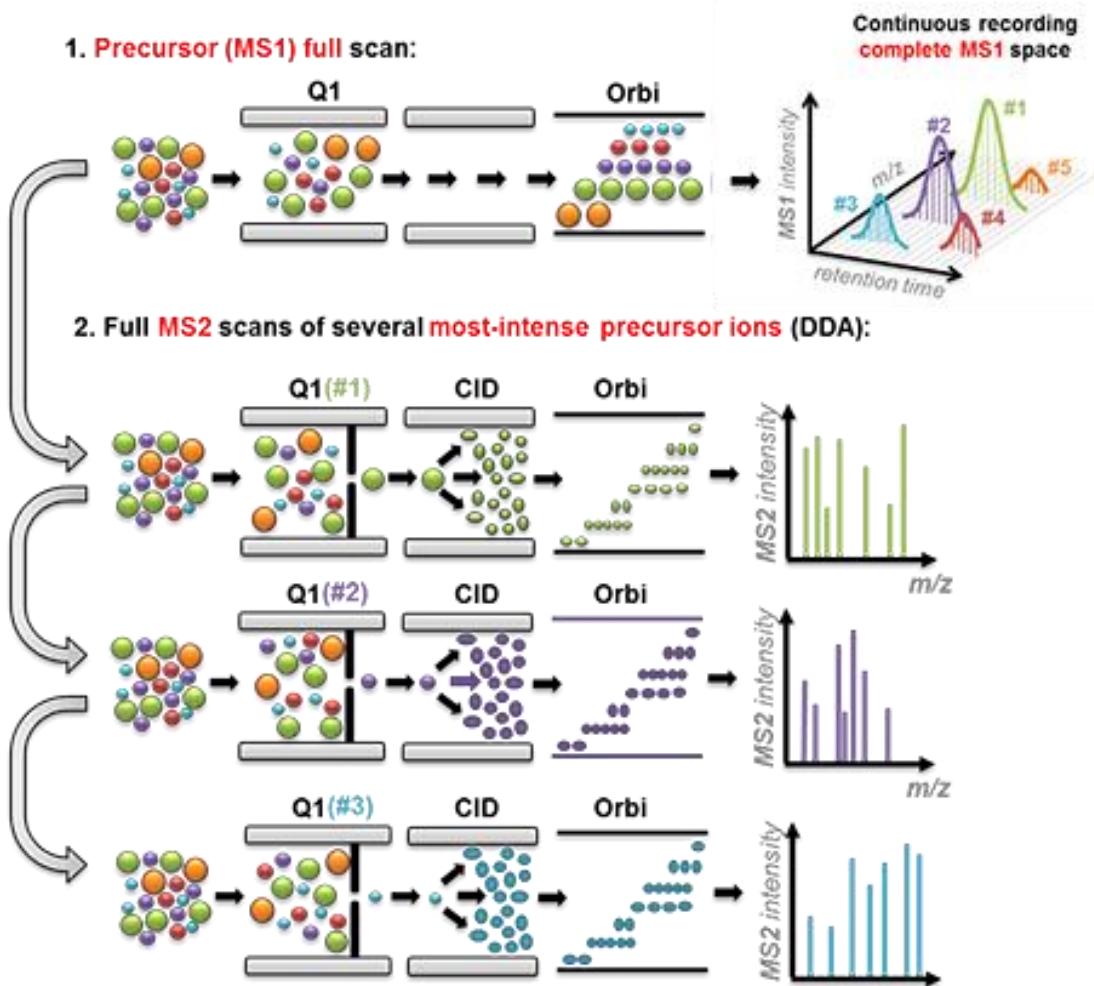


Figure 9 : Illustration du principe du mode d'acquisition dépendante des données DDA. Figure adaptée de Christina Ludwig et al. (2018)<sup>57</sup>

Le mode d'acquisition DIA permet la fragmentation de tous les ions et leurs fragments sont détectés en mélange dans le spectre MS/MS. Ce mode permet un meilleur recouvrement et une meilleure sensibilité des métabolites présents en mélange en dépit de la perte du lien entre l'ion parent et l'ion fragment (Figure 10). Ce type d'acquisition est peu utilisé en métabolomique en comparaison avec la protéomique car il engendre un problème de déconvolution et nécessite l'utilisation des bases de données. Néanmoins, le développement de nouvelles approches informatiques adaptées au traitement spécifique des analyses en DIA pourrait populariser ce mode dans les prochaines années<sup>58</sup>.

## DIA – data independent acquisition

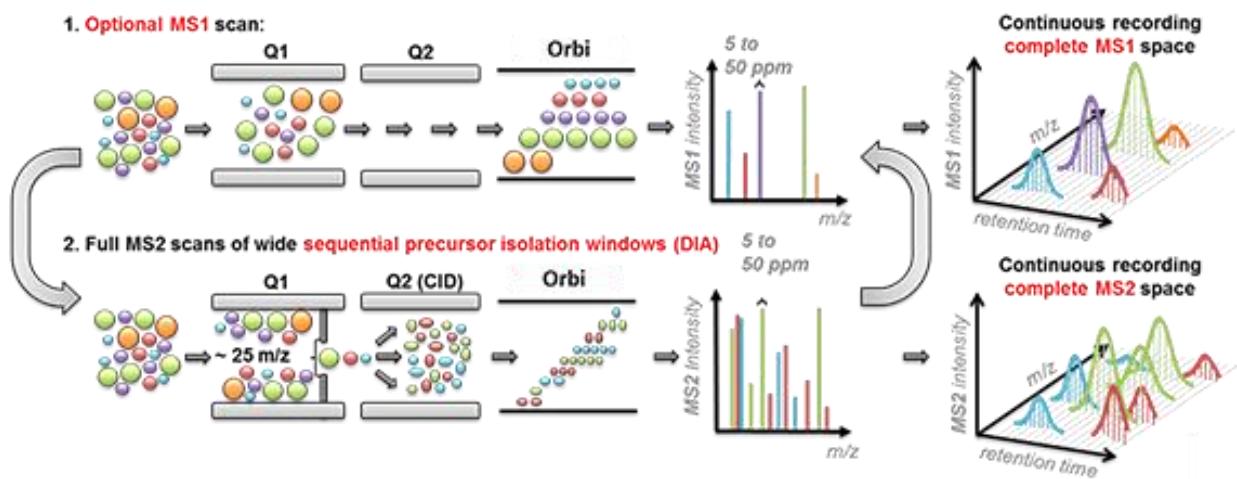


Figure 10 : Illustration du principe du mode d'acquisition indépendante des données DDA. Figure adaptée de Christina Ludwig et al. (2018)<sup>57</sup>

Une expérience de MS/MS dans une cellule de collision se déroule de la façon suivante : les ions précurseurs sont sélectionnés au niveau du quadripôle, puis accélérés afin d'accroître leur énergie cinétique dans la cellule de collision où ils vont entrer en collision avec des molécules de gaz inerte ( $N_2$ , Ar, etc.), ce qui produit des fragments chargés qui seront détectés et enregistrés dans les spectres MS/MS, ainsi que des pertes de neutres qui ne seront pas détectées. La distribution, en particulier l'intensité des différents fragments, varie en fonction du spectromètre de masse (géométrie, Figure 8) et des réglages utilisés.

Un point important commun à tous les spectres MS/MS, est l'influence de l'énergie donnée aux ions sur les modèles de fragmentation dans les expériences MS/MS induites par collision CID. En effet, la normalisation des énergies dans la CID est une tâche très difficile et le mode de fragmentation a un impact sur la façon dont un spectre MS/MS donné change avec l'augmentation des énergies. L'optimisation de ces énergies a fait l'objet de recherches approfondies, notamment en protéomique, où des règles de fragmentation claires ont été établies pour la déduction des séquences peptidiques<sup>59</sup>. Cependant, en protéomique c'est la liaison peptidique qui subit une rupture, tandis que sur les petites molécules divers sites de rupture existent.

En conclusion, la possibilité d'acquérir des données MS/HRMS à haute fréquence est un point clé pour l'enregistrement des empreintes de métabolites. Comme une méthode à haut débit est généralement nécessaire pour traiter le grand nombre d'échantillons et de répliques qui sont

analysés, des méthodes chromatographiques rapides telles que l'UHPLC sont utilisées en amont de la détection par MS<sup>60</sup>.

### 3) Couplage UHPLC-QTOF-HRMS

#### 1) Obtention des empreintes métaboliques

Le couplage UHPLC-QTOF-HRMS est reconnu comme étant très efficace pour l'étude des métabolomes des plantes<sup>61,62</sup>. Il peut également fournir un profilage très détaillé des métabolites. Ce type d'analyse peut être nécessaire pour la séparation d'isomères étroitement liés qui doivent être résolus pour des études spécifiques. La haute résolution apportée par l'UHPLC permet une localisation précise des biomarqueurs d'intérêt dans le cadre de leur isolement ou de l'enregistrement de leurs spectres MS et MS/MS<sup>63</sup>.

Afin de détecter un nombre maximum de métabolites ayant des polarités différentes, la séparation UHPLC-MS est effectuée en phase inverse (RP) en utilisant de larges gradients, aqueux-organique comme des systèmes eau-méthanol ou eau-acétonitrile avec des modificateurs volatils, principalement acides tels que l'acide formique ou l'acide acétique afin de faciliter l'ionisation et/ou la séparation<sup>64</sup>. D'autres modificateurs basiques (formate ou acétate d'ammonium) peuvent être employés en fonction du type de composés analysés. Néanmoins, pour analyser des constituants très polaires, d'autres modes chromatographiques peuvent être envisagés, comme la HILIC (*Hydrophilic Interaction Liquid Chromatography*).

Lorsque l'UHPLC est couplée à l'analyseur QTOF-HRMS, les données obtenues donnent des cartes d'ions ( $m/z$  vs temps de rétention  $t_R$ ) qui peuvent être utilisées pour l'identification de nombreuses structures. En outre, les problèmes de suppression d'ion liés à la coélution des métabolites sont moins fréquents avec l'UHPLC qu'avec l'HPLC<sup>42</sup>.

Lorsque les profils des métabolites ont été acquis au plus haut niveau de résolution et de sensibilité possible, avec un degré élevé de séparation des composants chimiques d'un extrait, l'identification des différents composés devient le prochain défi. En effet, l'identification sans ambiguïté des métabolites demeure un difficile challenge<sup>65</sup>.

#### 2) Identification des métabolites

Le type d'information délivré par le profilage UHPLC-HRMS des métabolites est illustré dans la Figure 11 qui présente une empreinte d'un extrait de plante<sup>43</sup>. Comme on peut le voir, un grand nombre de pics peut être détecté par UHPLC-HRMS (a), chaque pic correspondant à, au moins, un métabolite spécifique. Tous les points rouges de la carte des ions (b) représentent les caractéristiques moléculaires des métabolites détectés, elles sont appelées données caractéristiques ou *features* (ion moléculaire, adduit, fragment). Un métabolite peut être représenté par plusieurs données caractéristiques, par exemple, le  $m/z$  de l'ion

moléculaire  $[M+H]^+$  et le  $m/z$  de son fragment sont deux données caractéristiques qui correspondent au même métabolite.

Pour cet extrait de plante, la haute sensibilité de l'HRMS permet de détecter plus de 3000 données caractéristiques ( $m/z$ ) pour lesquelles les spectres MS/MS correspondants ont pu être enregistrés. Le nombre de données caractéristiques détectées change en fonction du seuil utilisé pour le *peakpicking* (extraction des informations de spectrométrie de masse à partir de pics chromatographiques). Plus le seuil de *peakpicking* est faible, plus on va détecter de données caractéristiques et donc de métabolites (c).

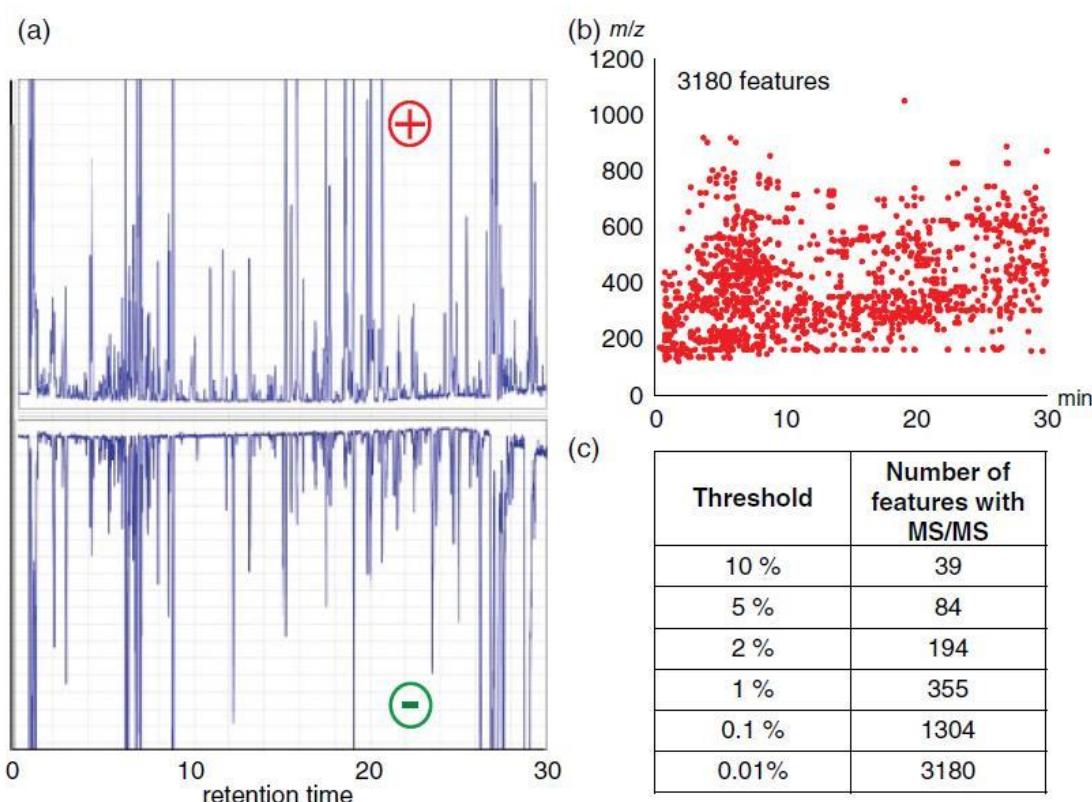


Figure 11 : Analyse UHPLC-HRMS d'un extrait de *Ginkgo biloba* réalisée par J-L. Wolfender et al. (2019)<sup>43</sup>. (a) Chromatogrammes d'ions totaux (TIC) obtenus en modes d'ionisation positif et négatif. (b) Carte d'ions des données caractéristiques des ions détectés en mode positif avec un seuil d'intensité de 0,01%. (c) Nombre de données caractéristiques en fonction du niveau de seuil d'intensité

Souvent, des détecteurs tels que le DAD (*Diode Array Detector*) ou ELSD (*Evaporative Light Scattering Detector*) sont utilisés afin d'appuyer l'identification des molécules car ils fournissent des informations complémentaires pour cibler les familles moléculaires. Contrairement au détecteur ELSD, le DAD ne détruit pas l'échantillon, il peut donc être directement intégré dans le couplage (UHPLC-DAD-HRMS).

Par conséquent, l'identification de tous les composés d'un extrait de plante consiste en l'interprétation i) des données chromatographiques (temps de rétention) ii) des données DAD et ELSD pour une identification de premier niveau iii) des spectres MS et MS/MS par l'annotation des pics de chacun des métabolites (qui sera détaillée ultérieurement). Une annotation manuelle est tout à fait possible, mais c'est un processus long. À cet égard, des approches bioinformatiques ont été développées afin d'accélérer l'identification des composés chimiques comme cela est illustré dans la Figure 12<sup>66</sup>.

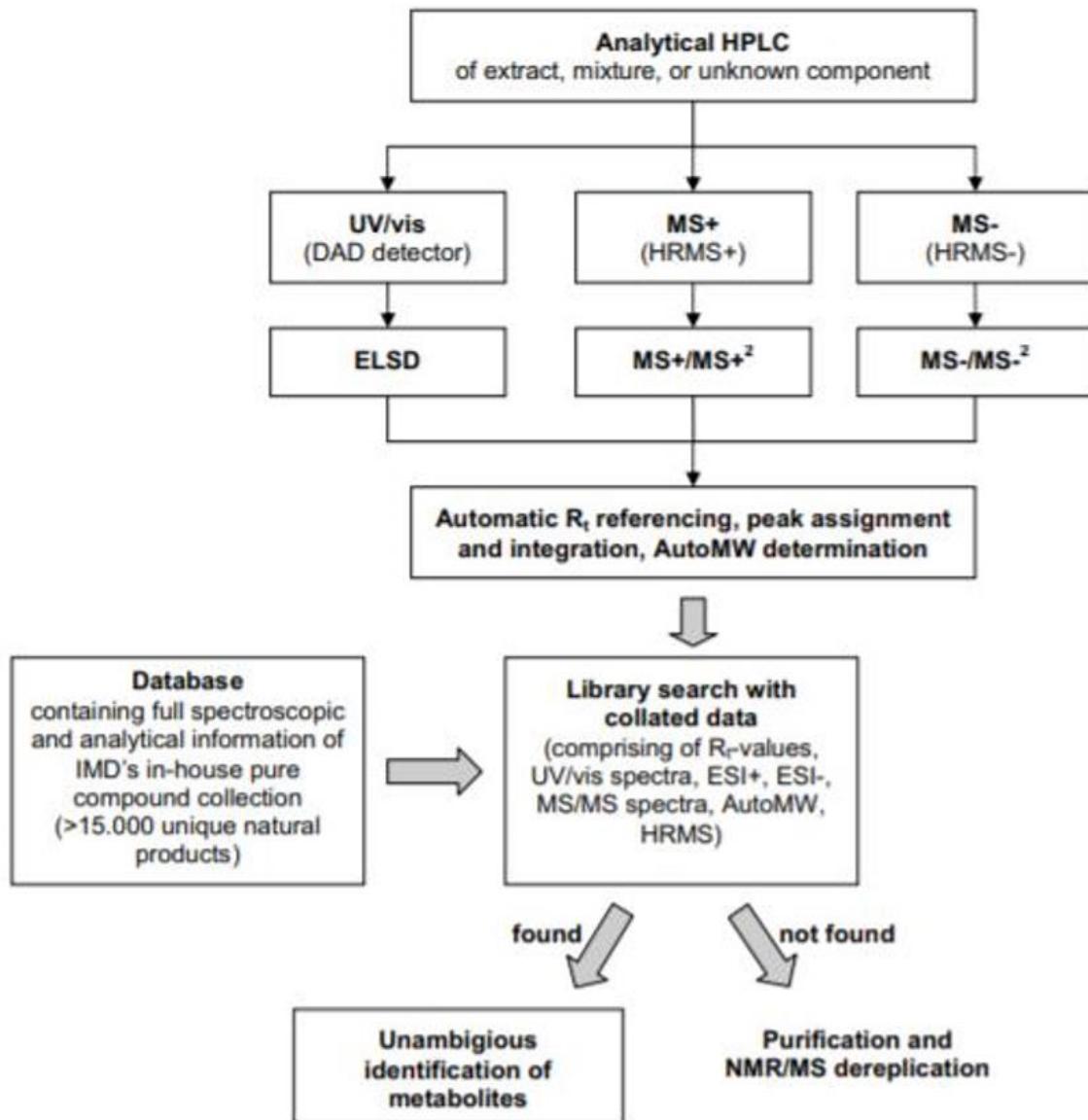


Figure 12 : Schéma récapitulatif des outils analytiques et informatiques qui permettent l'identification des métabolites réalisé par J. Bitzer et al. (2007)<sup>66</sup>

En conclusion, l'identification sans ambiguïté des produits naturels est une étape clé. Toutefois, c'est une étape délicate et longue pour le profilage des extraits naturels en raison de la diversité des outils utilisés et des réponses différentes qu'ils peuvent produire pour une

même molécule. De nombreuses bases de données accessibles en ligne sont utilisables pour soumettre des valeurs expérimentales  $m/z$  et obtenir des propositions de formules brutes et/ou développées mais aucune ne peut être complète. C'est pourquoi des stratégies d'aide à l'identification des produits naturels se développent afin d'accélérer et d'améliorer l'identification des composés d'intérêt. On retrouve notamment les approches de **Déréplication** ou encore des **Réseaux Moléculaires** (*Molecular Networking*) qui sont présentées dans la suite de ce chapitre.

### III. DÉRÉPLICATION ET ANNOTATION DES MÉTABOLITES

L'étude du contenu chimique d'un extrait de plante commence généralement par l'identification détermination exacte de l'espèce étudiée et par un travail bibliographique relatif à ce qui a déjà été publié sur la même espèce, le même genre ou encore la même famille, selon le nombre d'espèces apparentées. L'exploitation de ces études bibliographiques permet de définir une liste de composés qui pourraient potentiellement être présents dans l'extrait étudié, car une proximité taxonomique est généralement corrélée à une proximité chimique.

D'une manière formelle, la bibliographie a établi une relation entre l'espace des espèces et l'espace de la structure moléculaire. La Figure 13 présente les relations entre les différents espaces impliqués dans la chimie des produits naturels<sup>67</sup>. Les trois cases délimitées par la ligne de contour pointillée correspondent à des objets physiques et les autres à des objets intellectuels. La relation A relie les structures des molécules aux espèces. La relation B relie les espèces aux données biologiques<sup>67</sup>. En effet, une espèce donnée est liée à un ensemble de structures moléculaires (pharmacognosie) et une structure moléculaire donnée est liée à un ensemble d'espèces (pharmacognosie inverse)<sup>68</sup>.

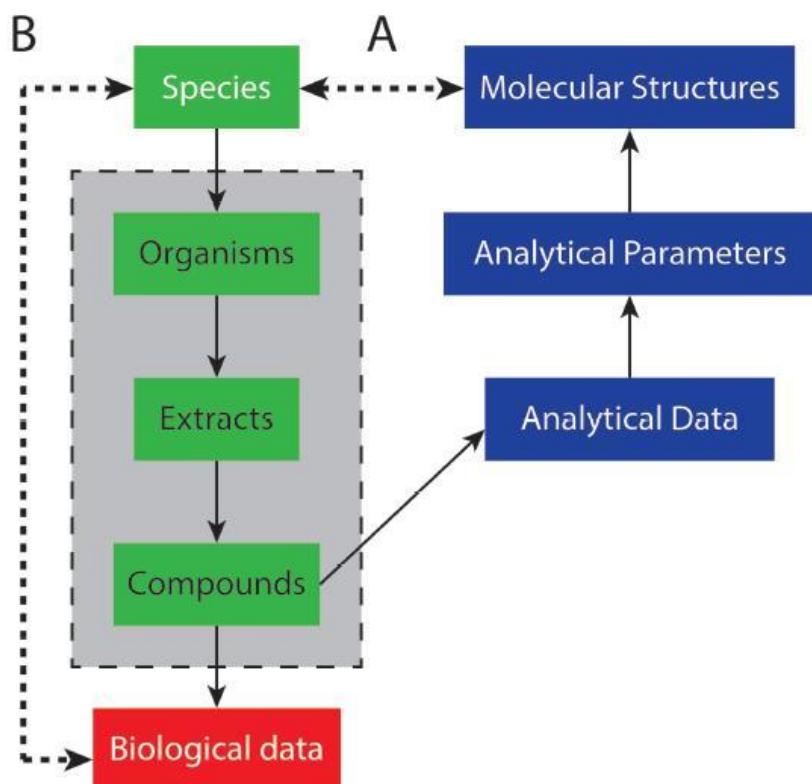


Figure 13 : Schéma récapitulatif des relations entre les différents espaces impliqués dans la chimie des produits naturels, reporté par A. Bakiri et al. (2017)<sup>67</sup>

Le workflow communément adopté dans la chimie des produits naturels implique l'isolement des composés à partir d'extraits d'organismes. L'extraction peut être plus ou moins sélective et/ou effectuée sur des parties particulières des organismes (telles que les écorces ou les fleurs des plantes), lorsque cela est possible, afin de simplifier l'isolement des composés et éventuellement de mettre en relation les organes et leur fonction. Un composé est généralement purifié par un chimiste spécialiste des produits naturels en vue d'étudier son activité biologique. De la même manière qu'un échantillon de plante doit être associé au nom d'une espèce vivante, un composé isolé doit être associé à une structure moléculaire afin de donner une valeur au test biologique. Un extrait étant riche en molécules, il peut être difficile de sélectionner les composés d'intérêt à isoler, et les chimistes spécialisés dans les produits naturels ont souvent la désagréable sensation de redécouvrir des composés déjà connus. C'est pourquoi des approches telles que la déréplication permettant de cibler de façon pertinente les composés d'intérêt à étudier de façon plus approfondie se développent.

### A. Définition et principe

La déréplication consiste à identifier les molécules déjà connues dans un mélange complexe. Dans ce sens, la déréplication est un moyen d'attribuer une structure chimique à un composé en utilisant les connaissances accumulées pendant des décennies de recherche sur les produits naturels. Par l'identification rapide des composés déjà décrits, la déréplication permet aux chercheurs de se focaliser sur les composés inconnus.

L'analyse physico-chimique est un moyen d'y parvenir en reliant l'espace des échantillons physiques à celui des données analytiques. Une relation entre ces dernières et l'espace des structures moléculaires est nécessaire afin de relier les composés isolés à leur structure moléculaire comme présenté dans la Figure 13. La déréplication consiste donc en l'acquisition de données analytiques et leur comparaison à celles déjà publiées. Cette tâche peut sembler facile et instantanément réalisable à travers le temps et les continents à l'ère d'Internet<sup>69</sup>.

### B. Histoire

De nombreuses stratégies de déréplication à base de produits chimiques ont été publiées au cours des décennies. L'adoption de ces stratégies de "dépistage chimique" a été renforcée lorsque des stratégies simples et des techniques chromatographiques stables ont été développées<sup>70</sup>. La principale technique chromatographique utilisée était la chromatographie sur couche mince (TLC). Une deuxième évolution dans les stratégies basées sur la chimie est intervenue au début des années 1970, lorsque des ordinateurs capables de traiter de grandes quantités de données interrogables ont été introduits. À l'origine, les chercheurs souhaitaient disposer d'une méthode pour obtenir des informations sur les classes chimiques des

composés d'un extrait de plante, une façon d'obtenir ces informations était d'utiliser des tests de réaction chimique.

Le premier type de procédure de criblage a été l'approche initiée par les chercheurs dans la dernière partie du XIX<sup>e</sup> siècle et le début du XX<sup>e</sup> siècle, appelée criblage phytochimique. Le criblage phytochimique des plantes est le processus par lequel des informations sur les classes structurales des produits naturels présents dans un extrait de plante sont obtenues, généralement par des réactions colorimétriques spécifiques avec des réactifs chimiques sélectionnés. Des études sur les types de réactifs chimiques utilisés pour différentes classes de composés ont été publiées<sup>71,72</sup>. Parmi les classes chimiques qui pouvaient être examinées, on trouve les alcaloïdes, les polyphénols, les glycosides et les saponines.

Le développement de la chromatographie sur couche mince a fourni aux chercheurs une technique de séparation des composés et de visualisation rapide pour l'identification des classes chimiques des produits naturels. De nombreux réactifs chromogènes ont été développés afin de mettre en évidence la présence de différentes classes de métabolites secondaires dans les plantes et de nombreuses études sur la spécificité et la sensibilité de ces réactifs ont été publiées<sup>73</sup>. Les chercheurs s'intéressaient d'abord au criblage des familles chimiques avant de se focaliser sur l'isolement et l'identification du composé actif de l'extrait de plante. Cependant, cette stratégie était d'une utilité limitée si l'objectif était de trouver de nouveaux composés biologiquement actifs indépendamment de la classe chimique. Le terme de déréplication a alors été défini en 1990 par Butler *et al.*<sup>74</sup> comme un processus d'identification rapide des chémotypes connus.

### C. Avancées de la déréplication

Depuis 30 ans, la déréplication a considérablement évolué. Elle a été utilisée dans différents procédés allant de l'identification de composés majeurs et de l'accélération du fractionnement bioguidé jusqu'au profilage chimique de collections d'extraits<sup>75</sup>. Ces développements ont été en partie dus aux progrès réalisés dans les techniques d'acquisition pour le profilage des métabolites dans des matrices biologiques complexes dans les domaines de la MS et de la NMR en termes de sensibilité, de résolution et de débit, mais aussi, plus récemment, à l'introduction de méthodes *in silico* et chimiométriques associées<sup>76</sup>. Il a été reconnu que les méthodes d'analyse traditionnelles ne plongeaient que légèrement dans le réservoir complet de molécules présentes dans les mélanges complexes, en y laissant une grande quantité de "matière noire"<sup>77,78</sup>.

Le workflow schématique de déréplication et d'élucidation des structures dans la chimie des produits naturels est présenté dans la Figure 14<sup>79</sup>. La tâche principale consiste à relier les échantillons (extraits, fractions chimiquement simplifiées ou composés isolés) et les structures

moléculaires qu'ils contiennent. Les extraits sont obtenus par différents procédés d'extraction qui conduisent à des mélanges complexes de composés ayant des propriétés physico-chimiques données selon la nature du solvant et de la technique utilisée. Les fractions et les composés purs sont obtenus après une ou plusieurs étapes chromatographiques préparatoires. L'élucidation des structures est réalisée par une combinaison de méthodes spectrométriques et spectroscopiques, principalement la MS et la NMR. Des méthodes supplémentaires peuvent également être utilisées, telles que, la diffraction des rayons X (XRD) pour les composés purs, la LC-DAD pour les fractions ou les extraits, etc.

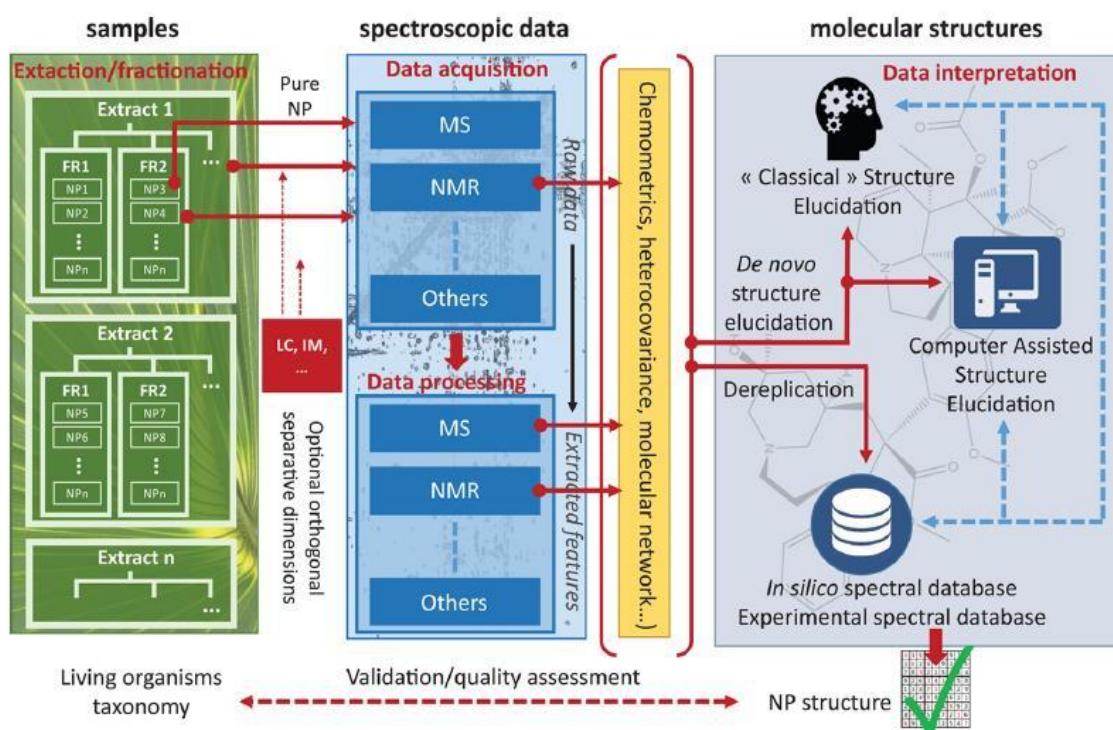


Figure 14 : Workflow schématique de déréplication/élucidation des structures dans la chimie des produits naturels énoncé par J.-L. Wolfender et al. (2019)<sup>76</sup>

Les données spectroscopiques sont divisées en deux sous-espaces : les "données brutes" acquises et les "données traitées" (par exemple, l'extraction des données caractéristiques MS et la combinaison des spectres MS et MS/MS d'une analyse LC-MS). La détermination de la structure moléculaire résulte de l'interprétation des données par deux stratégies différentes : l'élucidation *de novo* de la structure et la déréplication. L'approche *de novo* a principalement recourt à des stratégies manuelles, tandis que la déréplication est généralement assistée par ordinateur. L'ensemble de ces deux stratégies permettra l'annotation des métabolites.

## D. Annotation des métabolites

### 1) Élucidation de structure *de novo*

Comme mentionné précédemment, l'étude préliminaire de la composition des extraits naturels (profilage des métabolites) est principalement réalisée par LC-HRMS<sup>60</sup>. Dans ce processus, la séparation efficace des métabolites dans des mélanges complexes avant la détection MS est importante afin de générer des données de haute qualité pour la déréplication.

Dans la plupart des applications LC-HRMS, les métabolites sont ionisés à l'aide d'une source ESI, une technique d'ionisation douce qui génère peu de fragments. L'ESI a le grand avantage de produire des espèces d'ions essentiellement moléculaires qui apparaissent sous forme d'adduits, tels que  $[M+H]^+$ ,  $[M+Na]^+$  et  $[M+NH_4]^+$  en mode positif, ou  $[M-H]^-$ ,  $[M+HCOO]^-$  et  $[M+Cl]^-$  en mode négatif. En outre, des adduits ion-molécule (dimères), qui compliquent l'interprétation, peuvent également être formés dans la source d'ions. Par conséquent, avant l'identification, le poids moléculaire correct doit être défini par la reconnaissance des adduits<sup>80</sup>. Une fois que l'ion moléculaire est soigneusement sélectionné, manuellement par le chimiste, des hypothèses de formules brutes (MF) sont déterminées en fonction de l'exactitude de la masse et du spectre<sup>81</sup>. Cependant, la détermination définitive de la formule brute demeure une tâche difficile, même lorsque l'on travaille avec une exactitude inférieure à 2 ppm<sup>82</sup>. Ce fait est particulièrement vrai pour les composés ayant un poids moléculaire élevé ( $> 500$  Da). Afin de réduire le nombre de formules brutes possibles, différents filtres heuristiques peuvent être appliqués<sup>83</sup>, notamment des restrictions sur le nombre d'éléments, des modèles isotopiques, des rapports hydrogène/carbone ou des rapports élémentaires de l'azote, de l'oxygène, du phosphore et du soufre par rapport au carbone.

La Figure 15 illustre un exemple d'annotation de pic LC-MS, basé sur un profilage haute résolution de *Viola tricolore*<sup>84</sup>. (A) Chromatogrammes UHPLC-HRMS obtenus en ionisation (+) ESI et (-) ESI et le chromatogramme UV (366 nm). (B) Assignation de formules brutes basées sur une exactitude de 15 ppm et le profil isotopique obtenu à partir du spectre (-) ESI du pic à  $t_R = 11,33$  min. L'application des filtres heuristiques permet de déterminer la formule brute. (C) Le pic à  $t_R = 11,33$  min est annoté sur la base de l'attribution des formules brutes en PI, NI et du spectre UV. L'affectation structurale finale est basée sur une recherche croisée avec les informations chimiotaxonomiques, ce qui réduit considérablement le nombre de possibilités.

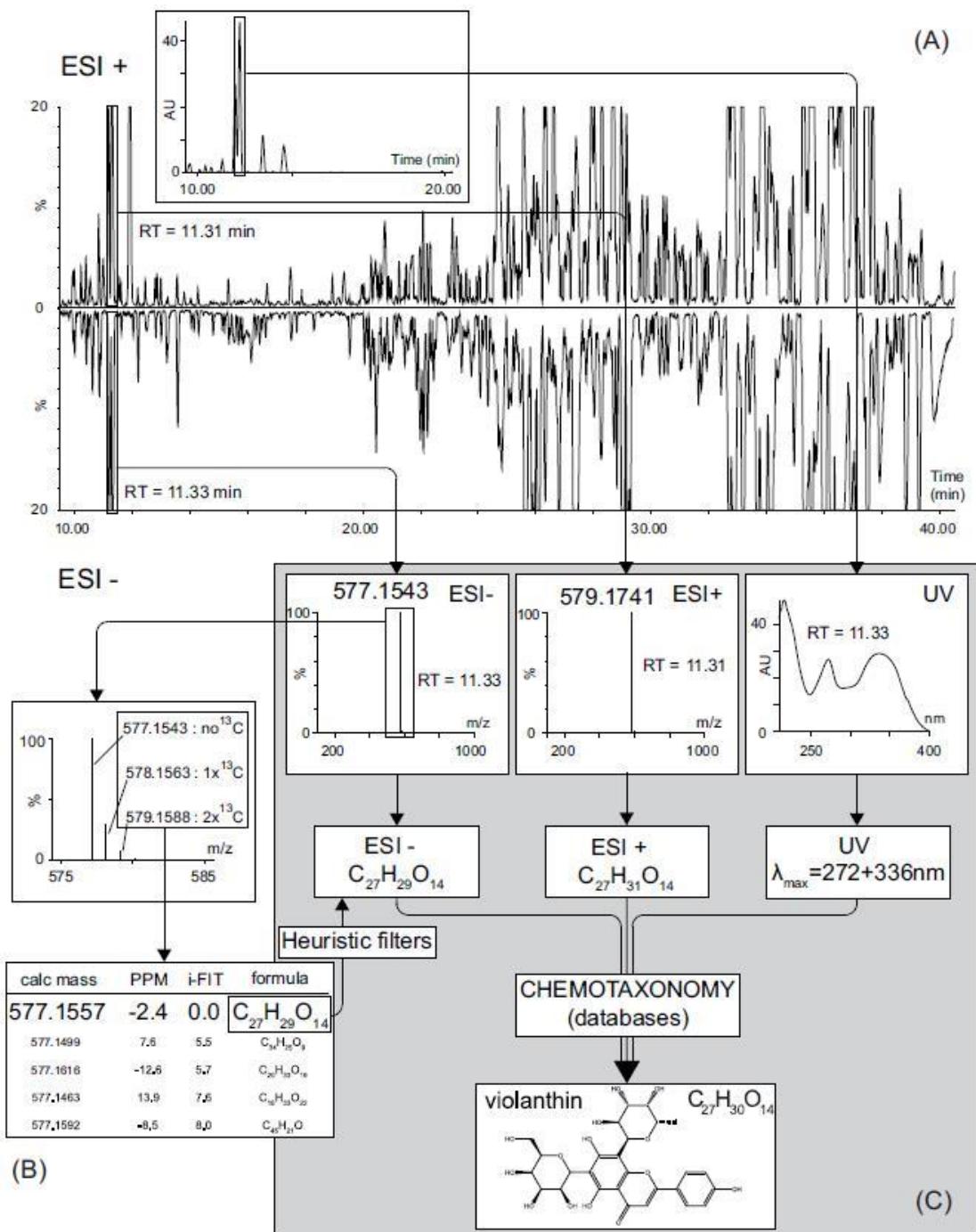


Figure 15 : Exemple d'annotation d'un pic LC-MS basée sur un profilage haute résolution de *Viola tricolore* réalisé par J-L. Wolfender et al. (2015)<sup>84</sup>

Sur la base des formules brutes présumées, l'identification des métabolites peut être réalisée en utilisant diverses bases de données (SciFinder, PubChem, ChEBI, DNP, etc.). Cependant, cette stratégie conduit généralement à de multiples identités présumées pour chaque composé détecté. C'est pourquoi diverses stratégies ont été utilisées pour réduire, autant que possible, le nombre de structures possibles qui correspondent à une formule brute donnée afin d'améliorer l'annotation des pics et l'identification des composés. Par exemple, les données

spectrales UV-visible peuvent être très instructives, en fonction des chromophores caractéristiques d'une classe de constituants, ces données sont souvent utilisées pour améliorer la déréplication<sup>85,86,87</sup>. En outre, les informations phylogénétiques de l'organisme analysé (famille, genre, espèce) réduisent également le nombre de structures possibles et améliorent la précision de l'annotation des pics<sup>88,89</sup>. Comme illustré dans la Figure 16<sup>43</sup>, (a) Spectre ESI-HRMS enregistré en mode positif d'une donnée caractéristique, (b) Détermination de la formule brute après filtrage heuristique et classement selon la similarité des profils isotopiques, (c) Recherche croisée de la formule brute obtenue avec les données chimiotaخonomiques, (d) Éventuels résultats structuraux, deux diglycosides isomères de flavone signalés chez *Ginkgo biloba*. Ce workflow ne permet pas de résoudre les structures isomériques.

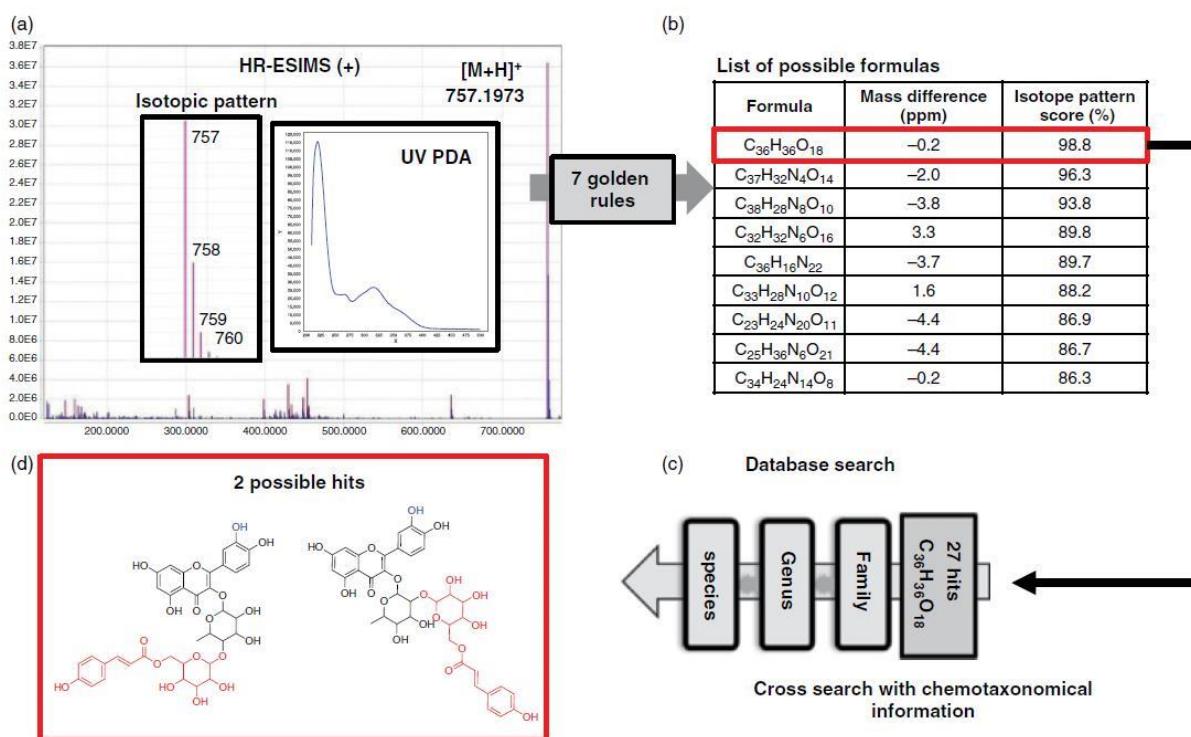


Figure 16 : Workflow de déréplication basé sur la détermination de la formule brute et la recherche croisée dans les bases de données pour un flavonoïde glycosylé d'un extrait de *Ginkgo biloba* énoncé par J-L. Wolfender et al. (2019)<sup>43</sup>

En plus de la formule brute, qui peut être utilisée comme paramètre générique pour les recherches dans les bases de données, le temps de rétention ( $t_R$ ) des analytes est l'autre information accessible dans le profilage des métabolites, chaque métabolite étant caractérisé par ses  $t_R$ ,  $m/z$  et intensité dans l'ensemble des données LC-MS. Le temps de rétention apporte des informations sur la polarité des molécules et peut permettre de discriminer certaines propositions de structures. Enfin, l'interprétation manuelle des spectres MS/MS d'un

composé et la compréhension de la fragmentation que subit la molécule est l'étape clé qui permet d'avoir un haut degré de certitude.

Cette étape *de novo* a été informatisée par la suite pour permettre une annotation plus exhaustive des composés avec génération d'hypothèses structurales pour les inconnus. Elle consiste en une annotation à l'aide d'un logiciel tel que MZmine<sup>90</sup> ou MetaboScape de Bruker et permet d'obtenir plus d'information, même si des annotations ou des identifications précises ne peuvent pas toujours être réalisées. Le workflow général comprend 5 étapes pour chaque pic d'un chromatogramme avec ses spectres MS et MS/MS associés :

- i) L'interprétation des spectres MS pour rechercher les adduits, les isotopes et les pertes de neutres.
- ii) La détermination de formule brute de l'ion détecté basée sur les informations de la MS ; sa masse et sa précision spectrale, les filtres heuristiques et la cohérence du modèle isotopique.
- iii) La recherche de la formule brute dans les bases de données disponibles Tableau 3, pour obtenir des annotations putatives (liste des structures possibles). En raison du grand nombre de bases de données disponibles, cette étape est l'une des plus longues de tout le processus.
- iv) La réduction du nombre d'annotations putatives basées sur des informations taxonomiques<sup>91</sup>.
- v) Utiliser les spectres MS/MS acquis pour déterminer la structure la plus probable parmi celles générées après les étapes iii et iv. Cette étape demeure la plus complexe.

Afin d'évaluer le niveau de confiance de l'annotation des métabolites, différentes normes de déclaration pour l'identification ont été définies par les chercheurs en métabolomique. Il en résulte un système à quatres niveaux<sup>65</sup>.

**Le niveau 0** ajouté depuis 2018 à la liste des niveaux de confiance de l'annotation des métabolites, correspond aux composés identifiés dont la stéréochimie a été déterminée.

**Le niveau 1** correspond aux composés identifiés. Pour qu'un composé puisse être considéré comme identifié, il faut qu'il soit validé par à minima deux critères indépendants et orthogonaux par comparaison à un composé de référence analysé dans les mêmes conditions expérimentales. Il peut par exemple s'agir du temps de rétention et de la masse précise ou de la masse précise et du spectre de fragmentation. L'utilisation de valeurs rapportées dans la littérature à partir de standards analysés dans d'autres laboratoires est insuffisante pour considérer un composé comme identifié. En général, c'est la comparaison de spectres de fragmentation.

**Le niveau 2** correspond aux composés putativement annotés. Il s'agit en général de composés pour lesquels une annotation est très probable, mais pour laquelle le standard n'est pas disponible ou de composés qui sont annotés à partir de bases de données externes au laboratoire menant l'étude.

**Le niveau 3** correspond aux composés dont l'annotation n'est pas complète, et/ou pour lesquels plusieurs annotations sont proposées, mais pour lesquels les informations disponibles permettent de les classer au sein d'une famille chimique.

**Le niveau 4** correspond aux composés inconnus. Il s'agit des variables détectées pour lesquels les informations disponibles ne permettent pas de tirer de conclusions claires quant à la nature du composé.

## 2) DéréPLICATION basée sur les bases de données

La première et la plus rapide des stratégies de déréPLICATION permettant d'identifier précisément les pics observés en LC-HRMS, est basée sur l'interrogation de bases de données internes aux laboratoires. Cette stratégie consiste à comparer les informations de  $t_R$  et MS obtenues avec celles de composés précédemment isolés qui ont été analysés dans les mêmes conditions. Cette stratégie est assez simple et présente l'avantage d'identifier diverses molécules avec un très haut degré de certitude.

Contrairement à la démarche d'élucidation de structure *de novo*, la déréPLICATION est principalement réalisée par des moyens informatiques, et non manuellement par le chimiste. Toutefois, cette stratégie se limite à l'analyse de composés connus et déjà étudiés. Cela a cependant été fait à une échelle assez large, comme le montre la bibliothèque de NP WEIZMASS<sup>92</sup>. Lorsqu'un métabolite est nouvellement observé dans un laboratoire, dans un profil donné, l'ensemble des données spectroscopiques acquises ainsi que les données structurales, devraient être conservées à des fins de comparaison future. Des stratégies d'annotation avancée, basées sur la LC-MS, ont été développées<sup>79</sup> afin de permettre l'identification de tous les produits naturels détectés dans un mélange.

### Annotation automatisée par rapport aux composés connus :

Cette stratégie a été développée en tenant compte de la nature difficile et longue de la création de bases de données internes. Celle-ci est très ciblée et consiste en la comparaison directe des spectres MS/MS acquis au laboratoire avec ceux déjà enregistrés dans des bases de données publiques.

De plus, la précision des rapports  $m/z$  des ions précurseurs permet de se concentrer rapidement sur des spectres précis. Cette recherche doit cependant être réalisée dans les

différentes bases de données MS/MS expérimentales qui sont généralement disponibles en ligne (Tableau 3), comme c'est le cas, par exemple, avec ReSpect DB<sup>93</sup> ou GNPS<sup>94</sup>.

### Analyse des suspects à l'aide de caractéristiques spectrales connues :

Cette stratégie a été introduite pour réduire l'espace de recherche dans les vastes ensembles de données LC-MS/HRMS à des molécules d'intérêt, comme par exemple, les dérivés de molécules dans les eaux usées<sup>95</sup>. Dans les extraits naturels, cela peut être fait en limitant la recherche des structures de produits naturels à une espèce ou un genre donné (recherche dans la littérature/les bases de données) et en générant *in silico* (informatiquement) un ensemble restreint de fragments MS/MS correspondants et une liste correspondante de formules brutes<sup>96</sup>. En utilisant des caractéristiques spectrales connues, telles que les masses ou les fragments de masse des molécules attendues (suspects), les bases de données sont interrogées pour trouver les données spectrales qui contiennent ces caractéristiques.

| Nom                      | Nombre de composés    | Accessibilité au public | Spectres MS/MS | Spectres UV | Spectres NMR    |
|--------------------------|-----------------------|-------------------------|----------------|-------------|-----------------|
| AntiBase <sup>97</sup>   | 40 000                | non                     | oui            | oui         | <sup>13</sup> C |
| ChEBI <sup>98</sup>      | 30 500                | oui                     | non            | non         | oui             |
| DNP <sup>12</sup>        | 275 000               | non                     | non            | oui         | non             |
| GNPS <sup>94</sup>       | Σ de bases de données | oui                     | oui            | non         | non             |
| KNApSAck <sup>99</sup>   | 51 000                | oui                     | non            | non         | non             |
| MarinLit <sup>100</sup>  | 50 000                | non                     | non            | oui         | <sup>1</sup> H  |
| UNPD-ISBD <sup>101</sup> | 170 000               | oui                     | oui            | non         | non             |

Tableau 3 : Quelques exemples de bases de données de métabolites

Aujourd'hui, les chercheurs contribuent au développement et à l'amélioration des bases de données afin de faciliter la déréplication. Idéalement, chaque composé nouvellement déclaré devrait avoir, au moins, sa structure moléculaire bien définie et de préférence lisible par ordinateur, ses données spectrales et ses paramètres spectraux préservés dans une base de données accessible au public afin que la déréplication puisse être effectuée avec un minimum d'efforts.

En conclusion, la caractérisation complète permet l'identification univoque de haute qualité des métabolites, mais elle nécessite beaucoup de temps et n'est souvent pas indispensable, surtout lorsque des composés connus sont caractérisés de manière redondante et que leurs

données spectroscopiques ont déjà été publiées plusieurs fois. De plus, l'isolement et l'identification complète de composés mineurs nécessitent de grandes quantités de matériel biologique. Afin de pallier à cela, des stratégies d'aide à l'identification et à la caractérisation de composés sans purification ont été développées telle que la déréplication.

Actuellement, le regroupement des composés en familles moléculaires est rendu possible par la comparaison des spectres de masse et la détermination de similarités spectrales ainsi que le développement des réseaux moléculaires (Molecular Networking).

## IV. RÉSEAUX MOLÉCULAIRES

Le profilage des métabolites par LC-MS/HRMS génère des fichiers de données de spectrométrie de masse riches en informations dans lesquels, pour un extrait complexe, plusieurs milliers de données caractéristiques LC-MS et MS/MS sont enregistrées.

Afin d'identifier des composés chimiques connus et nouveaux, la comparaison des spectres MS/MS est la clé de voûte de l'analyse par spectrométrie de masse. La raison principale en est que les similitudes spectrales sont souvent représentatives de similitudes structurales<sup>102,103</sup>. En effet, des modifications mineures apportées à une molécule, par exemple l'ajout d'un groupement méthyle, laissent la structure globale intacte. Comme cela n'a généralement pas d'incidence sur les principales voies de fragmentation qui, ensemble, aboutissent à un spectre de fragmentation. On peut donc déduire des similitudes structurales sur la base des similitudes spectrales. Les paragraphes suivants expliquent comment ce concept a été adapté dans la recherche sur les produits naturels et comment des outils de calcul ont été développés pour permettre son utilisation à grande échelle.

Il est reconnu depuis longtemps que les produits naturels partagent des structures qui donnent naissance à un groupe de composés structuralement similaires, c'est-à-dire une famille moléculaire de composés. Historiquement de tels squelettes ont été mis en évidence par la recherche manuelle d'ions fragments communs à travers différents spectres de MS/MS<sup>104,105</sup>. De plus, non seulement des molécules entières ou des squelettes plus grands, mais aussi des parties plus petites, c'est-à-dire des sous-structures de ces squelettes, donnent souvent lieu à des schémas de fragmentation similaires, même lorsqu'ils sont acquis à partir de molécules différentes. Toutefois, jusqu'à récemment, il n'était pas possible d'exploiter ce concept à une plus grande échelle.

Les spectres de fragmentation peuvent maintenant être comparés au sein d'un profil complet LC-MS/HRMS et entre différents profils. Diverses stratégies ont été utilisées pour comparer les spectres de fragmentation à grande échelle en comparant les ions fragments<sup>106,107</sup>, les pertes de neutres<sup>108,109</sup> ou les deux<sup>110,111</sup>. Ainsi, ces approches offrent aux chercheurs de nouvelles façons de représenter et de visualiser les données LC-MS/HRMS en exploitant différents aspects des similitudes entre les spectres de fragmentation.

### A. Définition et principe

Le concept de réseau moléculaire (MN) a été développé par les groupes de recherche de Pieter C. Dorrestein et Nuno Bandeiras de l'Université de Californie à San Diego. Ils ont mis en ligne en janvier 2014 la plateforme internet d'analyse et de partage de données MS/MS "Global Natural Products Social Molecular Networking – GNPS, Spectrometry Interactive

"Virtual Environnement MassIVES" (<http://gnps.ucsd.edu>) réunissant déjà une communauté de 10 000 utilisateurs provenant d'une centaine de pays en novembre 2015<sup>112</sup>.

Le réseau moléculaire représente l'outil le plus largement utilisé pour regrouper les molécules en familles moléculaires sur la base de leurs similitudes spectrales MS/MS, c'est-à-dire que plus deux spectres MS/MS partagent d'ions ou de pertes neutres, plus ils sont similaires. Il est basé sur la "fonction cosinus modifiée" qui i) examine les pics partagés entre deux spectres à l'intérieur d'un seuil défini par l'utilisateur, ii) considère les intensités des ions fragments et iii) considère la différence entre les deux masses parentales en déplaçant les ions fragments à l'intérieur de cette différence, et en vérifiant les correspondances entre les ions fragments (Figure 17)<sup>113</sup>. Après avoir vectorisés tous les spectres MS/MS, ils sont comparés les uns aux autres, chaque combinaison spectrale reçoit un score cosine allant de 0 (spectres complètement différents) à 1 (spectres identiques).

Ensuite, une "représentation du réseau" peut être construite comme suit : les molécules fragmentées servent de "nœuds", et elles ont une connectivité "arête" avec d'autres nœuds si elles partagent une similarité spectrale suffisante. Après avoir rassemblé tous les nœuds et les arêtes, on peut dessiner un réseau (Figure 17).

Des nœuds présentants des similitudes spectrales se lient et forment un cluster. Si un autre nœud ne présente pas de similarité spectrale avec ce cluster, soit il reste solitaire "singleton", soit il forme à son tour un cluster avec d'autres nœuds similaires. Le réseau moléculaire résultant contiendra alors des clusters qui vont correspondre à des "familles moléculaires", c'est-à-dire des clusters de molécules fragmentées qui sont interconnectées, ainsi que des "singletons", c'est-à-dire des molécules sans voisins connectés.

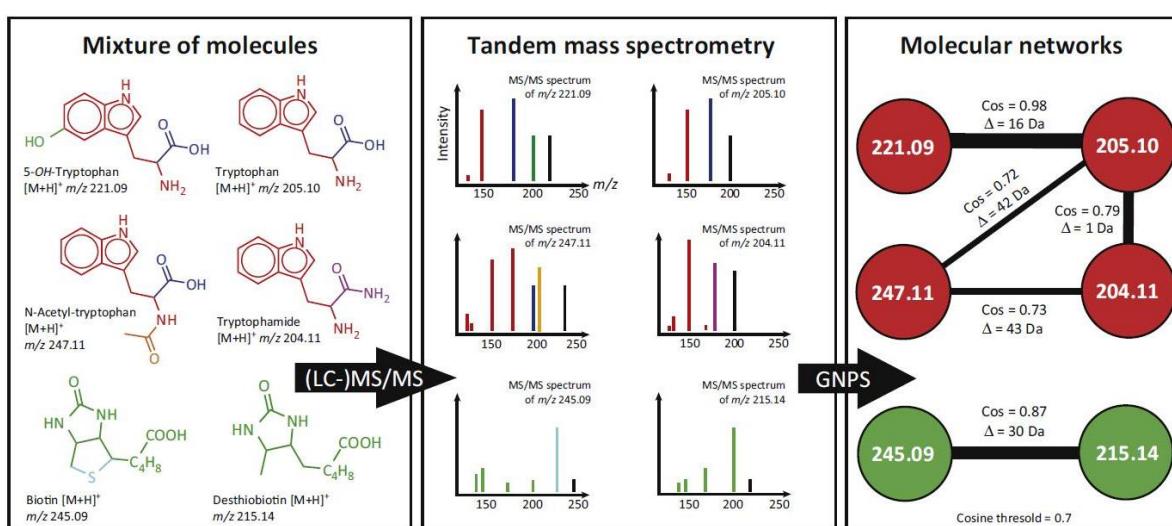


Figure 17 : Schématisation du processus de création des réseaux moléculaires présenté par R.A. Quinn et al. (2017)<sup>113</sup>

## B. Génération des réseaux moléculaires et visualisation

### 1) Génération des données spectrales

Le mode d'acquisition DDA principalement utilisé pour effectuer la recherche de métabolites dont on ne connaît pas la structure, est appliqué dans la plupart des études de produits naturels où l'annotation est basée sur les réseaux moléculaires<sup>113,114</sup> car il fournit des spectres MS/MS de bonne qualité et associe les ions fils à leur ions parents.

Comme les spectres MS/MS diffèrent d'un spectromètre à un autre, l'efficacité de la reconnaissance de spectres MS/MS expérimentaux par rapport à des spectres MS/MS enregistrés dans des bases de données avec des énergies différentes est limitée. Souvent, les intensités relatives diffèrent, et certains fragments spécifiques n'apparaissent que lorsqu'une énergie donnée est atteinte, alors qu'à des énergies trop élevées, d'autres ions fragments peuvent disparaître. Ceci est illustré dans la Figure 18 qui montre un spectre expérimental en (-) ESI de la catéchine, comparé à ses spectres de référence collectés à des énergies de collision de 10, 20 et 40 eV. Si la base de données ne contenait pas le spectre à 20 eV, le composé n'aurait probablement pas été annoté.

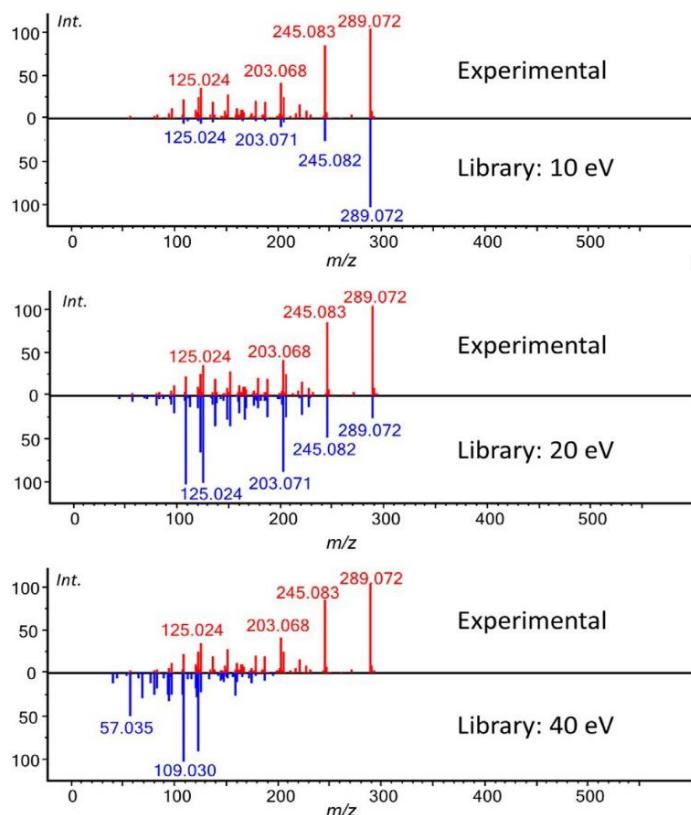


Figure 18 : Spectre expérimental en (-) ESI de la catéchine, comparé à ses spectres de référence à différentes énergies de collision tiré de la thèse de Florent Olivon (2018)<sup>294</sup>

Néanmoins, la plupart des spectromètres de masse modernes proposent une fonctionnalité "d'énergie de collision par palier" où, pendant l'acquisition, plusieurs spectres MS/MS de différentes énergies de collision sont fusionnés pour former un spectre moyen<sup>111</sup>. Cela permet d'obtenir des spectres riches en information, mais le choix des bonnes énergies à combiner demeure un défi car toutes les molécules ne se comportent pas de manière optimale à la même énergie. Il est donc important d'optimiser les énergies de collision en fonction des molécules à analyser et de faire des compromis afin d'obtenir un maximum de spectres MS/MS de bonne qualité.

## 2) Prétraitement et transfert des données MS

Les données MS/MS peuvent être directement intégrées dans la plateforme GNPS en utilisant les formats de fichier .mzML ou .mzXML sans prétraitement préalable. Le réseau est alors généré en activant l'algorithme MSCLustering de GNPS. En effet, avant de procéder à l'étape de comparaison entre les spectres MS/MS par le calcul des scores cosine, l'algorithme MSCLustering de GNPS fusionne les scans possédant le même ratio  $m/z$  d'ion précurseur ( $\pm$  la tolérance en Da entrée pour le paramètre), et possédant des spectres MS/MS quasi-identiques en un seul spectre consensus (spectre moyen). Cette étape permet d'améliorer la qualité des spectres des ions de faible intensité, mais a pour défaut de fusionner les différents isomères possédant des voies de fragmentation similaires car l'outil MSCLustering ne prend pas en compte la dimension de temps de rétention pour le clustering des données. Le nombre de spectres fusionnés par échantillon pour chaque spectre consensus est conservé et servira de seul moyen de quantification relative des ions (nombre de scan). Ce moyen de quantification est appelé *spectral counting*.

Néanmoins, si les données sont prétraitées au préalable, la taille de fichier est bien plus faible et il est possible d'obtenir des réseaux plus fiables avec moins de redondance d'information liée par exemple aux ions adduits d'un composé, une discrimination des isomères et un système de quantification plus précis qui va se baser sur les aires ou les intensités des pics et non sur le nombre de scan clusterisés. De plus, les réseaux peuvent être annotés avec les formules brutes calculées, et les nœuds peuvent avoir une taille proportionnelle à l'intensité des ions qu'ils représentent dans l'échantillon. L'un des logiciels les plus utilisés pour réaliser cette étape est MZmine dont l'accès libre.

Durant cette thèse, deux méthodes de prétraitement des données MS acquises avec le spectromètre maXis de Bruker ont été testées. La première méthode est basée sur l'utilisation de l'algorithme AutoMSMS du logiciel DataAnalysis 4.4 (Bruker) et la deuxième méthode est basée sur l'utilisation du logiciel MetaboScape 4.0 (Bruker). Ces deux outils permettent

d'obtenir des fichiers .mgf (*mascot generic format*) qui sont téléversés dans la plateforme GNPS avec des résultats fiables et parfaitement représentatifs des données brutes de départ.

### 3) Génération des réseaux moléculaires

La mise en réseau moléculaire des données MS/MS est réalisée sur la plateforme en ligne GNPS (<http://gnps.ucsd.edu>). C'est une plateforme web qui propose à la fois un dépôt de stockage des données et des outils d'analyse métabolomique et sert d'interface utilisateur graphique (GUI) pour un certain nombre d'outils.

La génération des réseaux moléculaires est un workflow qui vise à organiser de grands ensembles de données MS en exploitant les similitudes spectrales entre les modèles de fragmentation MS/MS d'ions précurseurs différents, mais structuralement liés.

Premièrement, les données MS/MS sont simplifiées pour réduire la charge de calcul en aval et pour améliorer l'efficacité de l'algorithme de similarité spectrale<sup>115,116</sup>. En particulier, les ions fragments de faible intensité et l'ion précurseur sont éliminés des spectres MS/MS. Ensuite, si l'utilisateur le souhaite, les spectres ayant le même ion précurseur et des spectres MS/MS similaires sont fusionnés en un seul spectre, il est appelé "spectre consensus". Ces spectres consensus MS/MS sont ensuite convertis en vecteurs dans un espace multidimensionnel normalisé où chaque dimension correspond à une valeur *m/z* et à son intensité respective. Ces vecteurs sont par la suite utilisés pour calculer un score cosine (Figure 19) entre chaque paire possible de spectres MS/MS, ce qui permet de déterminer le degré de similarité spectrale entre eux<sup>115</sup>.

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|_2 \|\mathbf{B}\|_2} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

Somme des « match scores »  
Racine de la somme des intensités au carré du spectre B  
Racine de la somme des intensités au carré du spectre A

Figure 19 : Formule de calcul des scores cosine entre deux spectres MS/MS A et B<sup>117</sup>

Pour ce faire, plusieurs paramètres peuvent être modifiés en fonction du spectromètre de masse utilisé, tels que la tolérance de masse des ions précurseurs pour le spectre consensus et la tolérance de masse des ions fragments (généralement 0,02 Da pour les instruments à haute résolution, dans la gamme *m/z* 100-1500). En outre, le nombre minimal d'ions fragmentés en commun peut être ajusté pour répondre à la spécificité du comportement de fragmentation des molécules analysées (généralement entre 3 et 7 en fonction du type de

molécules de l'extrait). Les résultats de ces comparaisons vectorielles peuvent ensuite être visualisés sous forme de graphiques de similarité spectrale appelés réseaux moléculaires, où chaque nœud est un spectre MS/MS et où les arêtes entre les nœuds indiquent le degré de similarité entre les spectres MS/MS (au-dessus du seuil des scores cosine de similarité défini par l'utilisateur, généralement entre 0,55 et 0,75).

Le choix de paramètres corrects est important lors de la réalisation de réseaux moléculaires pour garantir que i) les bons pics sont corrélés entre eux, ii) la plus grande partie du bruit est éliminée sans perte du signal, car cela affecte le score cosine, et iii) le seuil du score cosine est fixé de manière appropriée pour associer au sein du même cluster les composés de structure similaire.

#### **4) Visualisation des réseaux moléculaires**

Les premières études sur les réseaux moléculaires utilisaient des scripts MATLAB (The MathWorks, Inc.) installés sur un ordinateur de bureau pour le calcul des scores de similarité et la visualisation était réalisée à l'aide du logiciel Cytoscape<sup>118</sup>. Avec l'introduction de la plateforme web GNPS, les utilisateurs peuvent téléverser et stocker des données MS/MS en ligne, générer des réseaux moléculaires, saisir des connaissances sur les réseaux en tant qu'individu ou en tant que communauté, et ajouter des informations d'échantillon et d'autres métadonnées pour comprendre le réseau<sup>94</sup>.

Les clusters ou les sous-groupes d'un réseau moléculaire représentent des molécules structuralement liées et sont appelés familles moléculaires<sup>119</sup>. La visualisation des données des familles moléculaires peut être effectuée directement en ligne sur GNPS, alors que la visualisation d'un réseau moléculaire entier et de ses familles moléculaires constitutives se fait de préférence hors ligne avec Cytoscape ou un autre outil de visualisation de réseau. Dans Cytoscape, les propriétés des réseaux moléculaires peuvent être réglées pour faciliter l'interprétation des données, comme l'épaisseur des arêtes, qui peut être proportionnelle au score cosine, la taille des nœuds, qui peut être corrélée à l'intensité des ions précurseurs, ou la couleur des nœuds, qui peut être utilisée pour cartographier diverses métadonnées associées aux échantillons (cartographie de groupe).

#### **C. Interprétation des réseaux moléculaires**

Partant de l'hypothèse que des molécules apparentées produisent des schémas de fragmentation similaires en MS/MS, la mise en réseau moléculaire produit une carte de similarités spectrales MS/MS qui permet de visualiser le rassemblement de molécules structuralement apparentées au sein d'un même cluster et de molécules de structures différentes réparties dans des clusters différents. Le principal avantage de cette approche est qu'elle peut être utilisée pour l'exploration de nombreux spectres MS/MS sans aucune

connaissance préalable de la composition chimique des échantillons. De plus, GNPS peut automatiquement effectuer une recherche dans une base de données spectrale pour annoter les molécules si leurs spectres MS/MS sont répertoriés dans des bibliothèques spectrales publiques (MassBank<sup>120</sup>, Metlin<sup>121</sup>, ReSpect<sup>122</sup>, NIST<sup>123</sup> et GNPS<sup>94</sup>) qui hébergent les données MS/MS de plus de 22 644 composés individuels<sup>124</sup>. Ce processus de déréPLICATION appelé identification des "inconnus connus" est essentiel non seulement pour l'annotation moléculaire des spectres MS/MS, mais également pour la propagation de ces annotations à travers les réseaux, permettant la détection d'analogues et la découverte de nouveaux produits chimiques<sup>125</sup>. De plus, des molécules étalon peuvent être analysées afin d'annoter le réseau moléculaire sans ambiguïté. Enfin, l'annotation des composés déjà cités dans la bibliographie de l'espèce analysée, l'élucidation de structure *de novo* et l'annotation par les stratégies de déréPLICATION (présentées précédemment) sont les moyens d'annotation les plus utilisés pour interpréter les réseaux moléculaires.

Ainsi, l'interprétation des réseaux moléculaires peut être élaborée dans deux concepts complémentaires i) un concept initial dans lequel les données sont organisées sans connaissance préalable de leur composition chimique et ii) un concept avancé dans lequel une annotation moléculaire sûre est propagée à travers les réseaux moléculaires. Les connexions découvertes peuvent ensuite être exploitées, par exemple, lorsqu'une molécule est connue, on peut déduire que sa structure (ou une partie de sa structure) est connectée à d'autres molécules du réseau, étendant ainsi l'identification et l'annotation des métabolites "inconnus".

#### **D. Quelques avancées offertes par les réseaux moléculaires**

Le regroupement des molécules en familles moléculaires sur la base de leurs similitudes spectrales MS/MS a permis de nombreuses découvertes<sup>126,127,128,129</sup>. Ces quelques exemples montrent comment de grands ensembles de données complexes peuvent être utilisés pour permettre aux chercheurs de formuler rapidement des hypothèses sur les molécules qui sont structuralement liées et sur la façon dont elles sont réparties dans les échantillons (c'est-à-dire les souches, les espèces ou les lieux).

Par exemple, l'indexation du métabolome spécialisé de l'écorce séchée et des tiges de *Geissospermum laeve* a permis à F. Ramos *et al.* (2017)<sup>124</sup> l'élucidation structurale de trois nouveaux alcaloïdes et l'évaluation de leurs activités antiparasitaires et cytotoxiques. Leur processus d'isolement a été guidé par une stratégie de déréPLICATION basée sur un réseau moléculaire utilisant une base de données interne d'alcaloïdes-indole-monoterpènes.

De plus, les comparaisons de 300 espèces d'euphorbiacées ont révélé une tendance commune dans leur composition et ont permis l'isolement ciblé de nouveaux produits naturels bioactifs<sup>114</sup>. En effet, l'étude de F. Olivon *et al.* (2017)<sup>114</sup> montre que l'outil des réseaux

moléculaires peut être particulièrement adapté à la recherche de produits bioactifs au sein de collections de mélanges complexes. Dans un premier temps l'application d'un système de filtres de couleurs (activité ou taxonomie) et l'observation des réseaux leur a permis de prioriser la sélection des extraits d'intérêt par la comparaison de leurs profils métabolomiques, et de mettre en place des étapes d'annotation et de déréplication. Cette stratégie leur a ensuite permis d'isoler directement et de cibler les substances d'intérêt biologique tout en s'affranchissant d'effectuer des tests biologiques lors des étapes intermédiaires de purification.

En outre, l'étude de F. Di Ottavio *et al.* (2020)<sup>130</sup> qui porte sur 21 types d'échantillons d'aliments étiquetés comme superaliments ont été analysés et comparés à un grand nombre de fruits et légumes crus en utilisant l'UHPLC-HRMS. La mise en réseau moléculaire de l'ensemble des données a révélé la présence de plusieurs familles moléculaires constituées de molécules spécifiques non partagées avec le reste des échantillons d'aliments d'origine végétale. Certaines de ces familles moléculaires contiennent des molécules ayant des activités anti-inflammatoires, anticancéreuses, antibactériennes ou antioxydantes, soutenant les allégations de santé des superaliments correspondants.

Dans un autre contexte, une étude a montré comment le dépistage des drogues dans les échantillons d'urine peut être enrichi grâce à une approche non ciblée basée sur les réseaux moléculaires : des médicaments antihypertenseurs et des drogues métabolisées ont été trouvés au sein d'extraits d'urine complexes<sup>111</sup>. Enfin, le regroupement à grande échelle de molécules sur la base de leurs données spectrales a inspiré plusieurs outils d'annotation qui exploitent la topologie du réseau<sup>101,131,132</sup>.

En conclusion, les chimistes spécialisés dans les produits naturels ont été et sont toujours impliqués dans l'identification *de novo* complète et sans ambiguïté de composés inconnus provenant de matrices biologiques naturelles complexes. Leur implication dans la métabolomique a aidé au développement de méthodes pour améliorer l'identification des métabolites, un point clé qui représente un obstacle majeur dans le domaine de la métabolomique. Grâce au développement des plateformes analytiques et des outils bioinformatiques tels que la Déréplication, le Molecular-Networking (réseaux moléculaires) et la métabolomique, l'identification des produits naturels évolue rapidement vers des annotations de plus en plus complètes du métabolome des organismes pour continuer à découvrir de nouvelles molécules.

## V. OBJECTIFS DE LA THÈSE

Dans les parties précédentes il a été montré que les plantes sont une source de métabolites secondaires procédant de réelles propriétés biologiques, notamment, cosmétiques. La compréhension de ces activités biologiques repose sur le criblage et l'identification des molécules bioactives à partir des extraits de plantes très riches en composés. Néanmoins, cette procédure pourrait être longue et fastidieuse si des étapes de purification, de test et de caractérisation de chaque molécule sont envisagées. Toutefois, il est possible de combiner des outils analytiques, bioinformatiques et métabolomiques pour mettre en place des stratégies plus simples et efficaces permettant de cibler et caractériser des molécules bioactives ou des biomarqueurs d'extraits de plantes. Tel a été l'objectif de cette thèse.

Les stratégies envisagées et qui seront expliquées dans la partie résultats, reposent sur l'analyse d'extrait de plantes par réseaux moléculaires multi-informatifs obtenus par UHPLC-MS/HRMS combinée ou non à une approche de métabolomique ou de fractionnement et évaluation des activités biologiques.

Dans ce contexte, l'approche de réseaux moléculaires a été utilisée dans un premier temps pour la comparaison de la teneur en lignanes produits par des cals cellulaires de *Linum usitatissimum* soumis à deux conditions de culture différentes (lumière et obscurité). Ensuite, elle a été combinée à une étude métabolomique dans le but de déterminer et d'identifier des biomarqueurs discriminant des groupes de plusieurs extraits de racines de tomate *Solanum lycopersicum*. Enfin, elle a été associée au fractionnement par chromatographie de partage centrifuge afin de mettre en évidence et caractériser les molécules bioactives d'intérêt cosmétique (antioxydantes et anti-enzymatiques) de deux extraits de plantes ; *Artemisia annua* et *Eclipta alba* cultivées en Région Centre-Val de Loire.

Les stratégies développées doivent être simples, rapides, efficaces et applicables à différents types et nombres d'échantillons pour cibler des composés d'intérêt biologique, en particulier cosmétique, et pour répondre aux différentes problématiques.



**CHAPITRE 2 – DÉVELOPPEMENT ET OPTIMISATION DES  
RÉSEAUX MOLÉCULAIRES AFIN D'ÉTUDIER LA  
PRODUCTION DES LIGNANES DANS LES CALS  
CELLULAIRES DE *LINUM USITATISSIMUM L.***

---



## I. INTRODUCTION

L'espèce *Linum usitatissimum* L., souvent appelée lin et appartenant à la famille des Linacées, est une plante présentant un intérêt commercial. La littérature du lin remonte à 5000 avant J-C., lorsqu'il était principalement cultivé pour la fibre et l'huile en Europe occidentale, en Afrique du Nord et en Asie du Sud-Ouest<sup>133,134</sup>.

Des études récentes sur *Linum usitatissimum* ont mis en évidence plusieurs propriétés biologiques intéressantes de la plante, notamment anticancéreuse<sup>135</sup>, antidiarrhéique<sup>136</sup>, antimicrobienne<sup>137</sup>, anti-inflammatoire<sup>134</sup>, antioxydante et préventive contre les maladies cardiovasculaires<sup>138</sup>. Une partie de ces avantages pour la santé a été associée à la présence de lignanes. Après leur consommation, les lignanes végétaux sont convertis en entéro-lignanes (entérolactone et entérodiol) par les microorganismes présents dans l'intestin qui, selon la littérature, réduisent l'incidence de différents cancers<sup>139,138</sup>.

De plus, le lin est considéré comme un ingrédient cosmétique potentiel dans le monde entier, y compris en Chine<sup>140</sup>. Son utilisation en culture cellulaire par rapport à la plante entière apporte l'avantage de son obtention indépendamment des considérations climatiques ou saisonnières garantissant ainsi une production plus reproductible d'extraits de lin d'intérêt cosmétique.

Aujourd'hui, il existe de nombreux rapports décrivant des systèmes de culture tissulaire *in vitro* utilisés pour la propagation de *Linum usitatissimum* à partir d'explants d'hypocotyles<sup>141,142,143</sup> et d'anthères<sup>144,145</sup>. En outre, des cals<sup>146,147</sup> et des cultures en suspension cellulaire<sup>148,149</sup> produisant des quantités plus importantes de lignanes par rapport à la plante ont également été décrits.

De plus, l'utilisation de cultures de cellules végétales au lieu de plantes cultivées pour la production de biomolécules actives peut aider à surmonter les limites de qualité inégale dues aux changements saisonniers, aux méthodes de culture et aux variations géographiques. Les incohérences d'un lot à l'autre peuvent également être évitées en créant un environnement exempt d'agents pathogènes et de contamination<sup>150</sup>. En outre, en raison des conditions contrôlées associées à cette technologie, elle peut également être utilisée avec succès pour augmenter la quantité de production de biomolécules actives en optimisant les techniques de biotransformation et/ou en provoquant des conditions de stress (biotiques et abiotiques) incitant les cellules à se défendre en synthétisant les molécules adaptées.

Dans cette étude, cette technologie de culture cellulaire de tissus végétaux (cals cellulaires) a été appliquée à l'espèce *Linum usitatissimum* L. Ainsi, des lignées de cals cellulaires de *Linum usitatissimum* obtenues dans deux conditions de culture différentes (lumière et obscurité), produisant des métabolites secondaires notamment des lignanes précieux pour la cosmétique

ont été établies par nos collaborateurs du laboratoire Biomolécules et Biotechnologies Végétales de l'Université de Tours dans le cadre du projet ValBioCosm'.

Notre objectif était l'extraction et l'analyse des lignanes des cals cellulaires, et la comparaison de leur production dans les deux conditions de culture obscurité et lumière dans le but de déterminer si une condition de culture permettait d'accroître la production de lignanes. Les analyses d'extraits de cals cellulaires de lin ont d'abord été utilisées pour étudier les paramètres d'acquisition et de traitement qui influencent directement la construction des réseaux moléculaires. Ensuite, les réseaux moléculaires des extraits obtenus à partir des cals cellulaires de lin dans les deux conditions de culture (obscurité et lumière) ont été comparés afin d'évaluer les différences de production de métabolites d'intérêt.

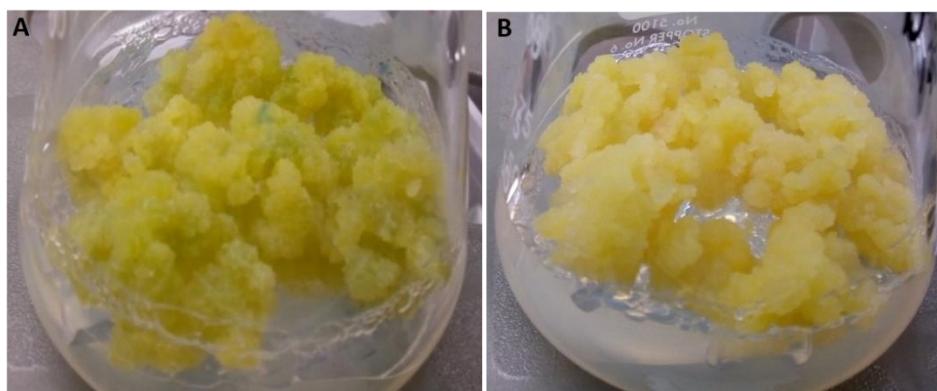
## II. CALS CELLULAIRES DE *LINUM USITATISSIMUM L.*

### A. Préparation d'échantillons

#### 1) Mise en place des cals cellulaires

Cette étape a été réalisée dans le laboratoire Biomolécules et Biotechnologies Végétales à l'Université de Tours. En effet, la sélection des explants (hypocotyles, cotylédons et racines) et la formation des cals de *Linum usitatissimum* ont été réalisées selon le protocole décrit par Hano *et al.* (2006)<sup>148</sup>, avec de légères modifications. Les explants hypocotyles ont été choisis et les meilleurs cals en croissance ont été obtenus dans les milieux de culture Murashige et Skoog (1962) contenant 2 mg/L de 6-benzylaminopurine et 0,5 mg/L d'acide acétique 1-naphthalene. Les cultures ont été maintenues dans deux conditions d'éclairage : l'une avec 12 h de lumière avec une intensité lumineuse de 25 µE m<sup>-2</sup>s<sup>-1</sup>/12 h d'obscurité, tandis que l'autre dans l'obscurité totale (24 h par jour).

L'intensité lumineuse a été mesurée en utilisant un luxmètre juste sous la source de lumière. L'éclairage était assuré par des LED au néon blanc (20 W, 380-780 nm, Philips). La température de la chambre de croissance a été maintenue à 24°C pour les deux conditions. Les cals ont été repiqués tous les 30 jours pour les deux conditions (Figure 20).



*Figure 20 : Cals cellulaires de *Linum usitatissimum*. A) culture de cals à la lumière après 30 jours, B) culture de cals à l'obscurité après 30 jours. Images prises par nos collaborateurs du laboratoire Biomolécules et Biotechnologies Végétales à l'université de Tours*

#### 2) Extraction

Les cellules lyophilisées de *Linum usitatissimum* ont été broyées en une fine poudre à l'aide d'un mortier et d'un pilon. 20 mg de poudre ont été extraits dans 1 mL de mélange éthanol/eau (70/30) dans un bain ultrasons pendant 1 h. Les extraits ont ensuite été centrifugés à 10 000 g pendant 7 min. Le surnageant a été collecté et analysé en UHPLC-HRMS. Cette méthode d'extraction a permis d'extraire les lignanes produits par les cals cellulaires de lin.

## B. Analyses UHPLC-MS/HRMS

Les analyses chromatographiques des extraits des cals cellulaires de *Linum usitatissimum* ont été effectuées à l'aide d'un système Ultimate 3000 RSLC en suivant les conditions décrites en Annexe 1. Les analytes ont été séparés sur une colonne C18 (Luna, Phenomenex) compte tenu de leur gamme de polarité. L'eau et l'acétonitrile ont été choisis pour éluer les lignanes d'intérêt moyennement polaires.

Les expériences de MS/MS ont été réalisées sur un spectromètre de masse UHR-Q-TOF maXis (Bruker, Brême, Allemagne) en mode d'ionisation électrospray (+) ESI avec une tension capillaire fixée à 4,5 kV. Les débits de gaz de nébulisation et de séchage (azote) ont été respectivement fixés à 2 bar et 9 L/min et le gaz de séchage a été chauffé à 200°C. Les analyses ont été effectuées avec une fréquence d'acquisition de 0,6 Hz pour les MS et les MS/MS et la plage de balayage de la masse a été fixée de  $m/z$  50 à 1550.

Les expériences MS/MS ont été réalisées en utilisant le mode d'acquisition dépendant des données (DDA). Plusieurs nombres de précurseurs sélectionnés pour chaque spectre MS ont été testés (2, 3, 4 et 5 précurseurs). Les ions précurseurs ont été fragmentés en utilisant une combinaison de deux énergies de collision et les spectres MS/MS ont été moyennés pour obtenir un maximum d'information. Pour ce faire, plusieurs combinaisons ont été testées, 15-35 eV, 20-45 eV et 15-65 eV pour étudier leur influence lors de la création des réseaux moléculaires.

Les données ont été traitées à l'aide du logiciel DataAnalysis 4.4. Les formules moléculaires ont été calculées en utilisant une composition élémentaire C, H, O, N<sub>0-5</sub> et une précision de masse ≤ 3 ppm pour la MS et ≤ 7 ppm pour les ions de la MS/MS.

Les chromatogrammes en (+) ESI des extraits des cals cellulaires de *Linum usitatissimum* cultivés à la lumière et à l'obscurité sont présentés dans la Figure 21. Des composés ayant une large gamme de polarité sont séparés en cours de l'analyse chromatographique allant de 3% de phase organique (acétonitrile) jusqu'à 90%. Des profils chromatographiques similaires sont observés pour les deux extraits de lin avec principalement des variations d'intensité pour certains composés, en particulier les composés polaires élus au début du gradient qui sont plus intenses dans les cals cellulaires cultivés à la lumière.

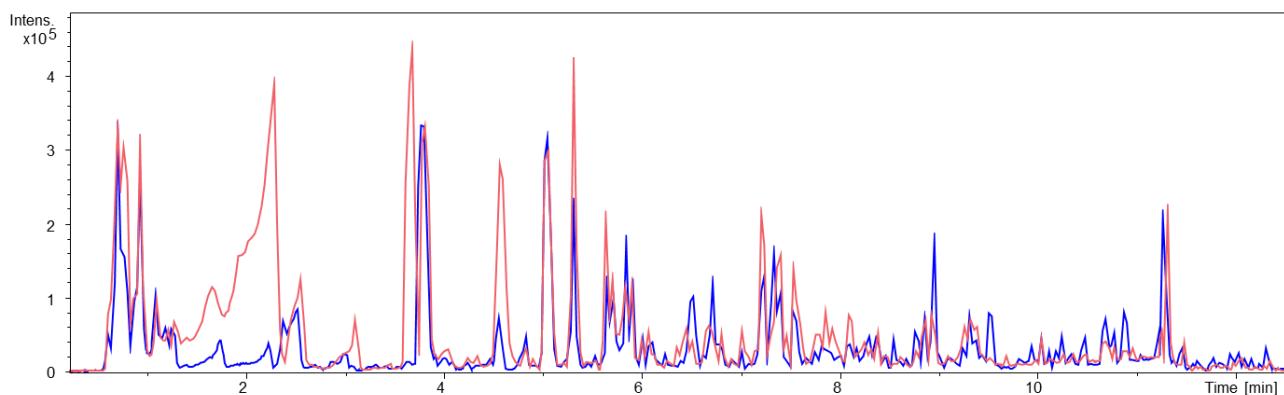


Figure 21 : BCP (Base Peak Chromatogram) des extraits de cals cellulaires de *Linum usitatissimum* cultivés à la lumière (rose) et à l'obscurité (bleu), obtenus par UHPLC-HRMS en mode (+) ESI

Les analyses des extraits ont ensuite été traitées afin de les exporter dans la plateforme GNPS pour la création des réseaux moléculaires. Pour ce faire, deux méthodes ont été utilisées selon le type de réseau moléculaire recherché, l'algorithme AutoMSMS du logiciel DataAnalysis, et le logiciel MetaboScape 4.0.

L'algorithme AutoMSMS de DataAnalysis 4.4 est rapide, simple et facile à mettre en œuvre ; toutes les étapes nécessaires étant regroupées dans un seul module avec des paramètres à optimiser en fonction des analyses LC-MS. Il permet d'exporter des spectres MS/MS moyennés correspondant à la fusion des spectres acquis à plusieurs énergies de collision avec une bonne sensibilité spectrale. Les informations exportées pour chaque ion précurseur fragmenté sont : temps de rétention, numéro de scan, état de charge ; les ions fragments avec leurs intensités et états de charge charges. Cependant, cet outil a trois défauts majeurs ; il est applicable uniquement sur les données acquises par les instruments Bruker (les autres constructeurs ont des algorithmes équivalents), il ne présente pas d'étape d'alignement des temps de rétention, ce qui rend l'analyse comparative de plusieurs échantillons difficile, et ne permet de générer que des réseaux qualitatifs et non quantitatifs car il n'exporte pas les intensités ou les aires de pics des composés.

MetaboScape 4.0 acquis au laboratoire en fin 2018 qui utilise un algorithme puissant pour l'extraction des caractéristiques spectrales (*peakpicking*), appelé T-ReX (*Time aligned Region complete eXtraction*). Cet algorithme 3D avancé permet d'extraire et de combiner toutes les informations pertinentes. Il permet de réaliser le même traitement de données avec plusieurs autres fonctionnalités additionnelles (alignement de données, identifications, analyses statistiques, visualisation, etc.). En effet, celui-ci offre la possibilité d'aligner plusieurs analyses pour obtenir un fichier .mgf unique comportant les intensités ou les aires des pics des composés dans toutes les analyses effectuées afin d'aller plus loin dans l'analyse comparative des échantillons. Il permet de générer des formules brutes et d'effectuer la recherche de

composés dans des bases de données avant le transfert vers la plateforme GNPS, et il permet de regrouper les adduits de l'ion moléculaire voire certains fragments (perte d'eau, de sucre) en une seule caractéristique MS. Enfin, MetaboScape permet d'exporter les données directement dans GNPS "export for GNPS" facilitant ainsi le transfert des données. L'inconvénient de MetaboScape est sa spécificité aux analyses des instruments Bruker et qu'il ne s'agit pas d'un logiciel libre contrairement par exemple à MZmine. Cet inconvénient représente également la force de ce logiciel lorsque les analyses sont effectuées sur les instruments de la marque, puisqu'il n'induit aucun traitement supplémentaire des données brutes.

Dans un premier temps, l'algorithme AutoMSMS du logiciel DataAnalysis a été utilisé pour traiter et exporter les données dans le but d'évaluer les facteurs qui influencent directement la création des réseaux moléculaires qualitatifs. Alors que le logiciel MetaboScape de Bruker a été utilisé dans un deuxième temps pour créer des réseaux moléculaires quantitatifs afin de comparer la production de lignanes par les cals cellulaires de *Linum usitatissimum* cultivés à la lumière et à l'obscurité.

### **III. PRINCIPAUX PARAMÈTRES INFLUENÇANT LES RÉSEAUX MOLÉCULAIRES**

#### **A. Nombre de précurseurs**

Le nombre de précurseurs d'un spectre MS qui sont sélectionnés pour être fragmentés en MS/MS influence directement les informations qui peuvent être observées sur les réseaux moléculaires. En effet, sur un spectre MS, les ions peuvent provenir soit d'une fragmentation en source de l'ion moléculaire d'une molécule X ou alors d'une coélution, c'est-à-dire un ion moléculaire, fragment ou adduit d'une molécule Y éluée au même temps de rétention que la molécule X. Parmi l'ensemble de ces ions MS, certains ions appelés précurseurs seront sélectionnés et fragmentés en MS/MS s'ils remplissent les critères sélectionnés pour l'acquisition DDA. Chaque ion précurseur qui rentre dans la cellule de collision pour être fragmenté, est représenté par un nœud sur le réseau moléculaire, qu'il vienne de la molécule X ou de la molécule Y. Lorsque les ions moléculaires se fragmentent en source en donnant plusieurs ions fragments aptes à rentrer dans la cellule de collision, les molécules sont représentées par plusieurs nœuds générant ainsi de la redondance d'information. Plus le nombre de précurseurs augmente plus le taux d'information est élevé. Il est ainsi possible d'obtenir les nœuds correspondant aux molécules X et Y, en revanche, le nombre de redondance d'information peut croître si les ions fragmentés proviennent tous de la même molécule X ou Y. À l'inverse, si le nombre de précurseurs est trop faible, une des molécules X ou Y, peut ne pas être fragmentée en MS/MS ce qui implique une perte d'information.

La Figure 22 présente les réseaux moléculaires de cals cellulaires de *Linum usitatissimum*, cultivés à la lumière, obtenus en choisissant différents nombres d'ions précurseurs. Le nombre total de nœuds, de nœuds singulaires et de clusters ( $> 2$  nœuds) sont résumés dans le graphe de droite. On peut observer que le nombre total de nœuds et de nœuds singulaires est considérablement augmenté avec le nombre d'ions précurseurs comme expliqué précédemment. À l'inverse, le nombre de clusters est réduit lorsqu'on passe de 2 à 5 précurseurs.

Ceci peut s'expliquer par le fait que lorsqu'on ne sélectionne que 2 précurseurs, l'instrument passe plus de temps sur chaque masse permettant ainsi d'obtenir des spectres MS/MS de meilleure qualité et un plus grand nombre de clusters. Tandis que lorsqu'on sélectionne 5 ions précurseurs, le taux d'information augmente mais la qualité des spectres MS/MS est réduite car l'instrument passe moins de temps sur la fragmentation de chaque masse d'où le faible nombre de clusters. L'idéal pour cet extrait est donc la sélection de 4, voire 3 ions précurseurs pour avoir le plus grand nombre de nœuds uniques et de clusters et le moins de nœuds redondants.

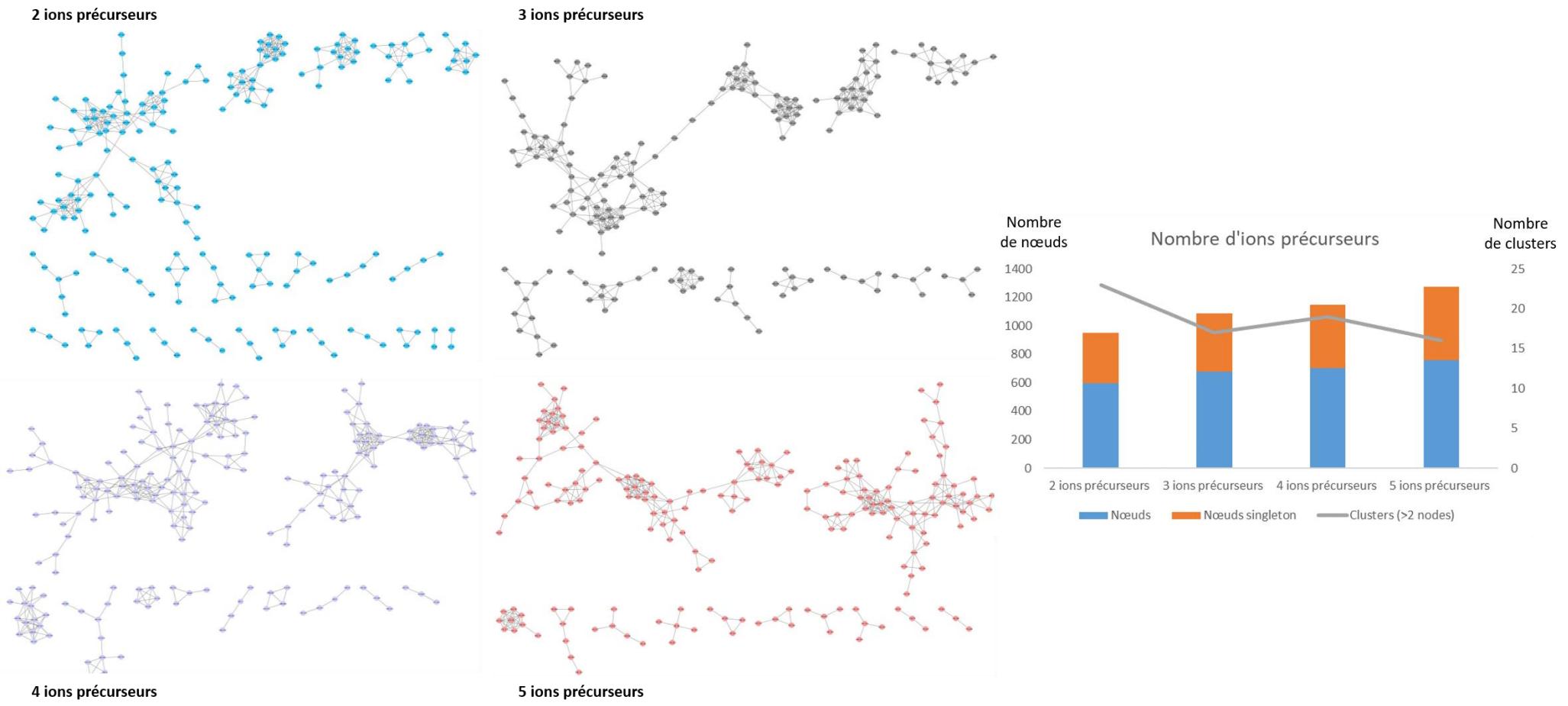


Figure 22 : Réseaux moléculaires de cals cellulaires de *Linum usitatissimum* cultivés à la lumière, obtenus en fonction du nombre d'ions précurseurs sélectionnés. Graphe résumant le nombre total de nœuds, de nœuds singlets et de clusters (>2 nœuds) des réseaux moléculaires

## B. Énergies de collision

L'optimisation des énergies de collision est très importante lors de la création des réseaux moléculaires puisque toutes les molécules ne se comportent pas de la même manière face aux différentes énergies de collision. Le graphe présenté en Figure 23 montre l'impact de différentes énergies de collisions sur les réseaux moléculaires présentés à gauche. Le nombre total de nœuds, de nœuds singlons et de clusters sont présentés par des histogrammes et courbe respectivement et peuvent varier en fonction de l'énergie de collision utilisée.

En effet, si les molécules sont trop fragmentées, leurs spectres MS/MS vont contenir très peu voire pas d'information, elles sont donc représentées par des nœuds singlons qui ne vont pas se lier à des clusters même s'ils comportent des similarités structurales. De la même façon, si les molécules sont trop rigides, elles ne se fragmentent pas à des énergies de collisions faibles, et seront présentées par des nœuds singleton puisque leurs spectres MS/MS ne présenteront pas assez d'information lors du calcul du score cosine pour être liés avec d'autres nœuds. Le graphe ci-dessous montre que le nombre de nœuds singlons est plus faible lorsqu'une combinaison de 20 et 45 eV est utilisée. Ce qui signifie que les spectres MS/MS obtenus avec cette combinaison sont de meilleure qualité et ont offert plus d'information pour la comparaison des spectres et le calcul des scores cosine augmentant ainsi le nombre de clusters. Tandis que le nombre total de nœuds qui représentent les analytes varie très légèrement ce qui peut être dû au caractère aléatoire du mode d'acquisition DDA.

En conclusion, la combinaison d'énergies de collision 20 et 45 eV offre le meilleur compromis, avec un nombre élevé de clusters et le plus faible nombre de nœuds singlons.

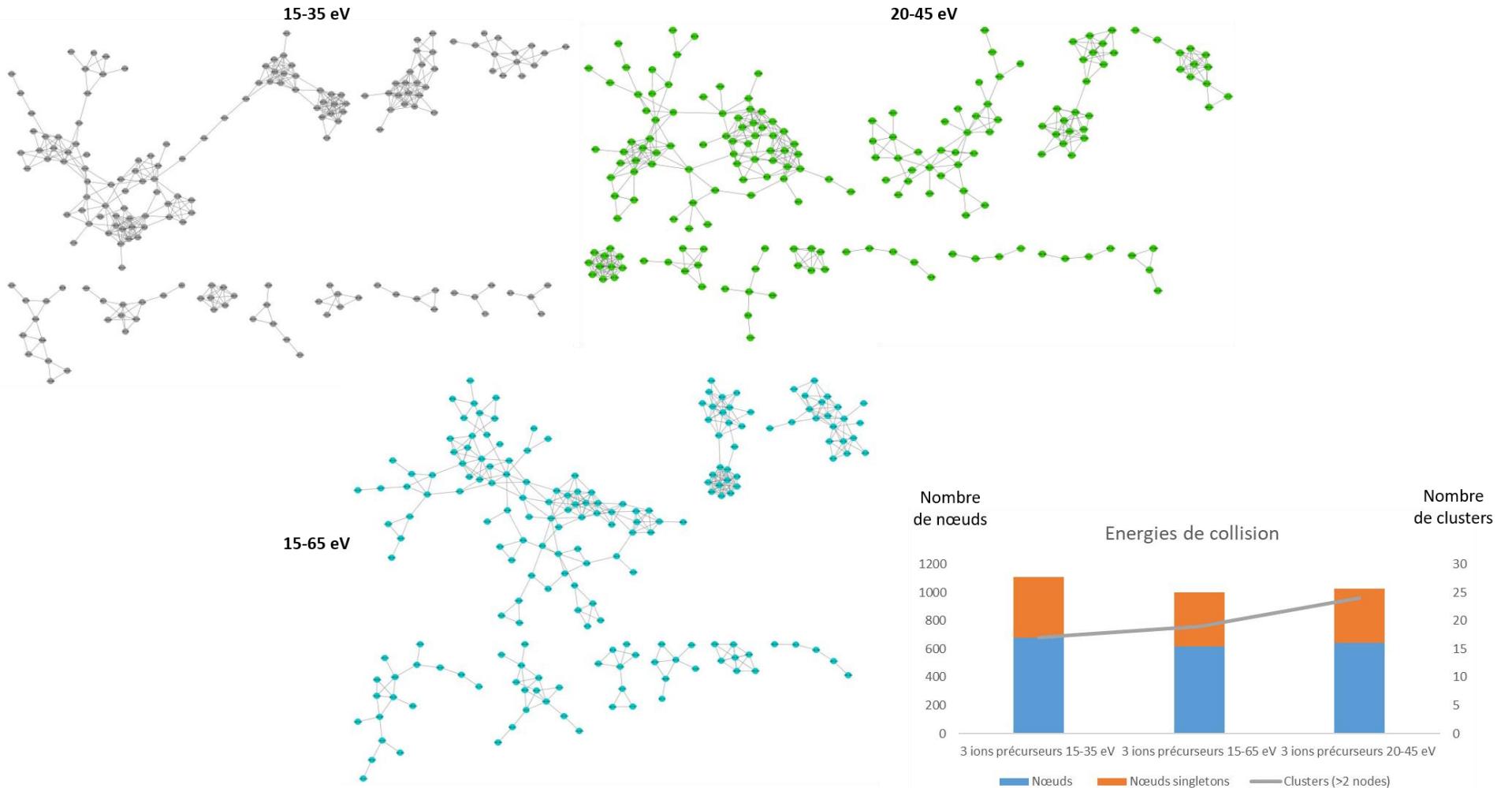


Figure 23 : Réseaux moléculaires de cals cellulaires de *Linum usitatissimum* cultivées en lumière, obtenus à différentes énergies de collision. Graphe résumant le nombre de nœuds total, de nœuds singlons et de clusters (> 2 nœuds) des réseaux moléculaires

### C. Fichiers de transfert des données spectrales

La plateforme GNPS supporte trois types de fichier ; mzXML, mzML et mgf. Il est donc nécessaire de convertir les données brutes avant de les exporter sur GNPS afin de créer les réseaux moléculaires. Pour ce faire, il existe différents outils de conversion de données brutes en mzXML et mzML : serveur de GNPS (mzML), vendor conversion (mzXML), MSConvertGui (mzXML) et l'outil CompassXport de DataAnalysis. Tandis que le format mgf est lui généré après traitement de données par l'algorithme AutoMSn DataAnalysis dans cette partie.

Les formats de fichiers les plus utilisés mzXML et mgf. ont donc été générés en utilisant toutes les plateformes de conversion citées plus haut, et exportés dans la plateforme GNPS pour créer les réseaux moléculaires dans le but d'étudier si le fichier exporté peut avoir un impact sur les résultats obtenus.

Les résultats obtenus ont permis d'observer que tous les fichiers transformés en mzXML donnent la même réponse que cela soit en nombre total de nœuds, de nœuds singleton et de clusters. Les réseaux moléculaires obtenus sont identiques peu importe la plateforme utilisée pour la conversion des fichiers. Tandis que le fichier mgf, obtenu après traitement de données diffère en tout point des autres fichiers (Figure 24). En effet, d'un côté le fichier mzXML permet d'exporter la totalité des informations sans tri, ensuite l'outil MSClustering est activé sur la plateforme GNPS pour filtrer un peu les données mais cet outil présente des inconvénients comme mentionné dans le chapitre précédent (introduction générale). De l'autre côté, le fichier mgf exporte les données avec moins d'information puisque lors de la formation du fichier mgf, on fixe un seuil d'intensité afin de réduire le bruit de fond et les signaux très faibles. De plus, les ions précurseurs avec des *m/z* identiques élués dans un intervalle de temps de 0,4 min sont fusionnés en une seule donnée. L'ensemble de ces traitements réduisant considérablement le taux d'information exportée et donc le nombre total de nœuds. En revanche, le nombre de clusters augmente puisque les spectres MS/MS des fichiers mgf sont de meilleure qualité.

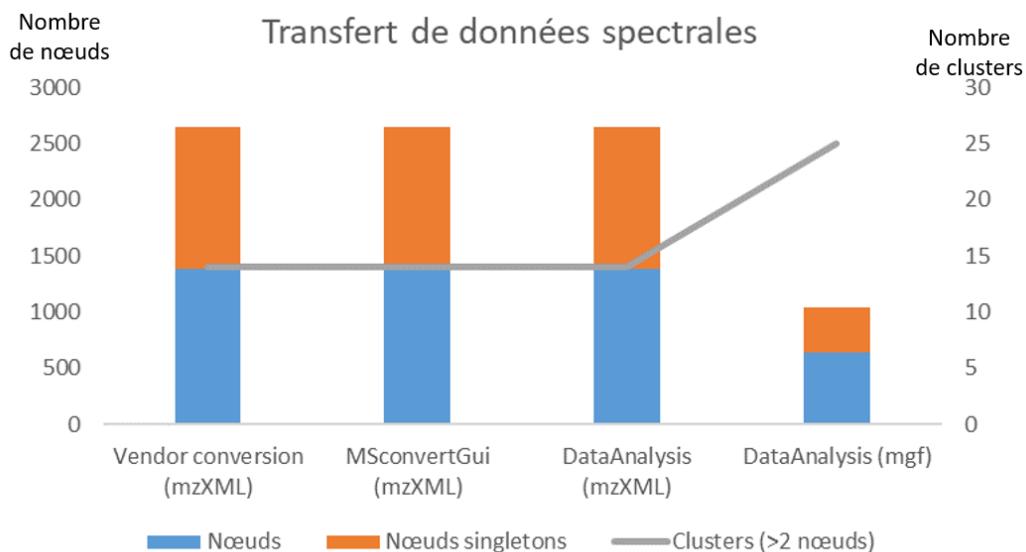


Figure 24 : Graphe présentant les informations des réseaux moléculaires (cals cellulaires de *Linum usitatissimum*) obtenues après transfert de données spectrales sur GNPS par différents formats de fichier

#### D. Nombre minimum de correspondances des spectres MS/MS

En plus des paramètres d’acquisition de données MS/MS, les consignes sélectionnées sur GNPS influencent la création des réseaux moléculaires et doivent être optimisés. Parmi elles, le nombre minimum de correspondances exigées pour que deux nœuds soient liés en un cluster. Les correspondances peuvent être des ions identiques et/ou des pertes communes en tenant compte des proportions des ions, et plus on retrouve de correspondances, plus le score cosine est élevé et donc la similarité structurale des composés. La Figure 25 présente les spectres MS/MS de 5 composés élués à différents temps de rétention. La comparaison de ces spectres montre des correspondances spectrales entre les composés 1, 2, 3 et 4 qui ont l’ion commun à  $m/z$  185,0434 et les pertes communes de 144 et 162. Quant au composé 5, il présente le même spectre ( $m/z$ , ions fragments et leurs proportions) que le composé 4, son isomère élué 0,36 min avant.

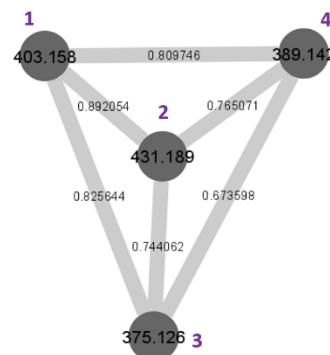
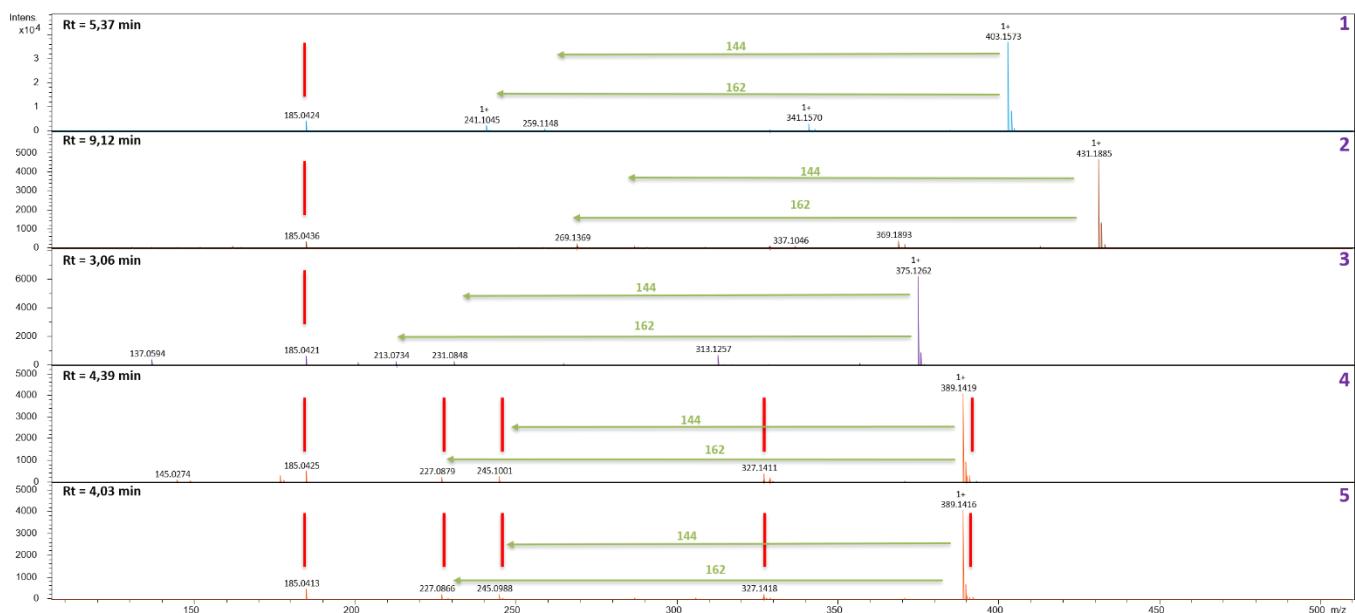


Figure 25 : Correspondances spectrales considérées lors de la comparaison des spectres MS/MS et du calcul des scores cosine par la plateforme GNPS lors de la création des réseaux moléculaires des cals cellulaires de *Linum usitatissimum*

Lors de l'export de ces données spectrales sur GNPS, les composés 1, 2, 3 et 4 ont été représentés par des nœuds liés en un cluster (Figure 25) s'ils présentent au moins 3 correspondances. Les scores cosine sont mentionnés sur les arrêtes qui lient les nœuds dans le cluster. Les composés 1 et 2 ayant le score cosine le plus élevé (0,89) présentent les deux spectres les plus similaires. Tandis que les composés 3 et 4 avec un score cosine de 0,67 ont les spectres les moins similaires parmi les spectres présentés.

Lors de l'export des données, les spectres MS/MS des ions précurseurs ayant le même *m/z* élués dans un intervalle de 0,4 min ont été fusionnés, c'est la raison pour laquelle les composés isomères 4 et 5 ont été fusionnés et un seul nœud est observé. Par conséquent ce temps de fusion a été réduit à 0,3 min pour séparer ces isomères, d'où l'importance

d'optimiser les paramètres de prétraitement de données par rapport aux données chromatographiques et spectrales.

De la même manière que les paramètres présentés précédemment, le réseau moléculaire varie en fonction du nombre de correspondances minimal exigé pour que les nœuds soient liés comme présenté dans la Figure 26. En effet, on peut remarquer que le nombre total de nœuds ne varie pas puisque les données spectrales sont les mêmes, néanmoins, le nombre de nœuds singltons augmente lorsque le nombre de correspondances minimal exigé est élevé. A l'inverse, le nombre de clusters réduit considérablement puisque les spectres MS/MS doivent au moins avoir 5 correspondances pour être liés en cluster.

En conclusion, le paramètre optimum pour l'extrait des cals de *Linum usitatissimum* est 3 voire 4 correspondances minimales exigées afin d'obtenir un bon compromis entre le nombre de nœuds singltons et de clusters et permet d'avoir plus de confiance que si 2 correspondances seulement sont exigées.

La comparaison des spectres MS/MS expérimentaux et ceux enregistrés dans les bases données spectrales de GNPS pour l'information de déréPLICATION se fait de la même manière que la comparaison entre les spectres MS/MS d'un même échantillon. Pour ce faire, il est préférable d'augmenter le nombre minimum de correspondances exigés à 4 voire 5 pour une identification plus fiable.

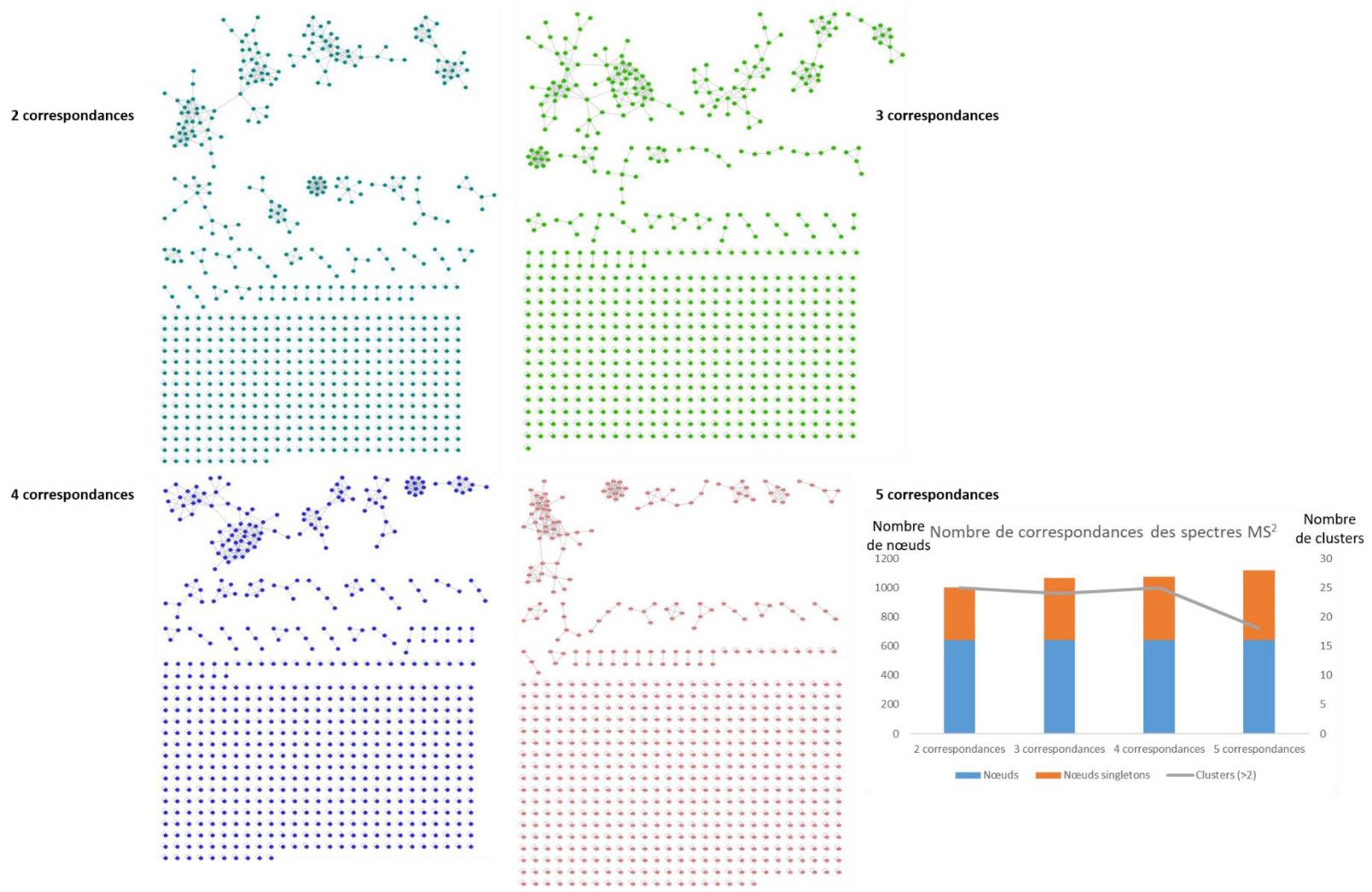


Figure 26 : Réseaux moléculaires de cals cellulaires de *Linum usitatissimum* cultivés à la lumière, obtenus en fonction du nombre de correspondances minimal exigé. Graphe résumant le nombre total de nœuds, de nœuds singltons et de clusters (> 2 nœuds) des réseaux moléculaires

## E. Scores cosine

Sur la plateforme GNPS, il est recommandé d'exiger un score cosine minimum compris entre 0,55 et 0,75 pour que 2 nœuds soient liés par une arête. En effet, si deux spectres MS/MS comparés présentent un score cosine inférieur à 0,55 cela signifie qu'ils ne présentent pas voire peu de similarité spectrale et doivent être séparés dans le réseau moléculaire. Tandis qu'un score cosine supérieur à 0,75 résulte d'une très grande similarité spectrale et il est difficile à atteindre lorsque deux molécules ne présentent qu'une partie de structure commune.

La Figure 27 présente des réseaux moléculaires d'un même échantillon obtenus avec différents scores cosine. On peut remarquer que le nombre de nœuds total ne varie pas puisque le fichier d'analyse est le même. Alors que le nombre de singletons et de clusters varie en fonction du score cosine. En effet, le nombre de nœuds singletons augmente avec le score cosine et certains spectres MS/MS similaires mais avec un score cosine de 0,67 sont séparés lorsqu'un score cosine de 0,75 est exigé d'où la diminution du nombre de clusters. A l'inverse, le score cosine à 0,55 offre moins de nœuds singleton mais certains spectres MS/MS avec peu de similarité ont été liés. Par conséquent, le score cosine de 0,65 est le paramètre optimum permettant de lier des spectres similaires et d'avoir un bon compromis entre le nombre de nœuds singletons et de clusters.

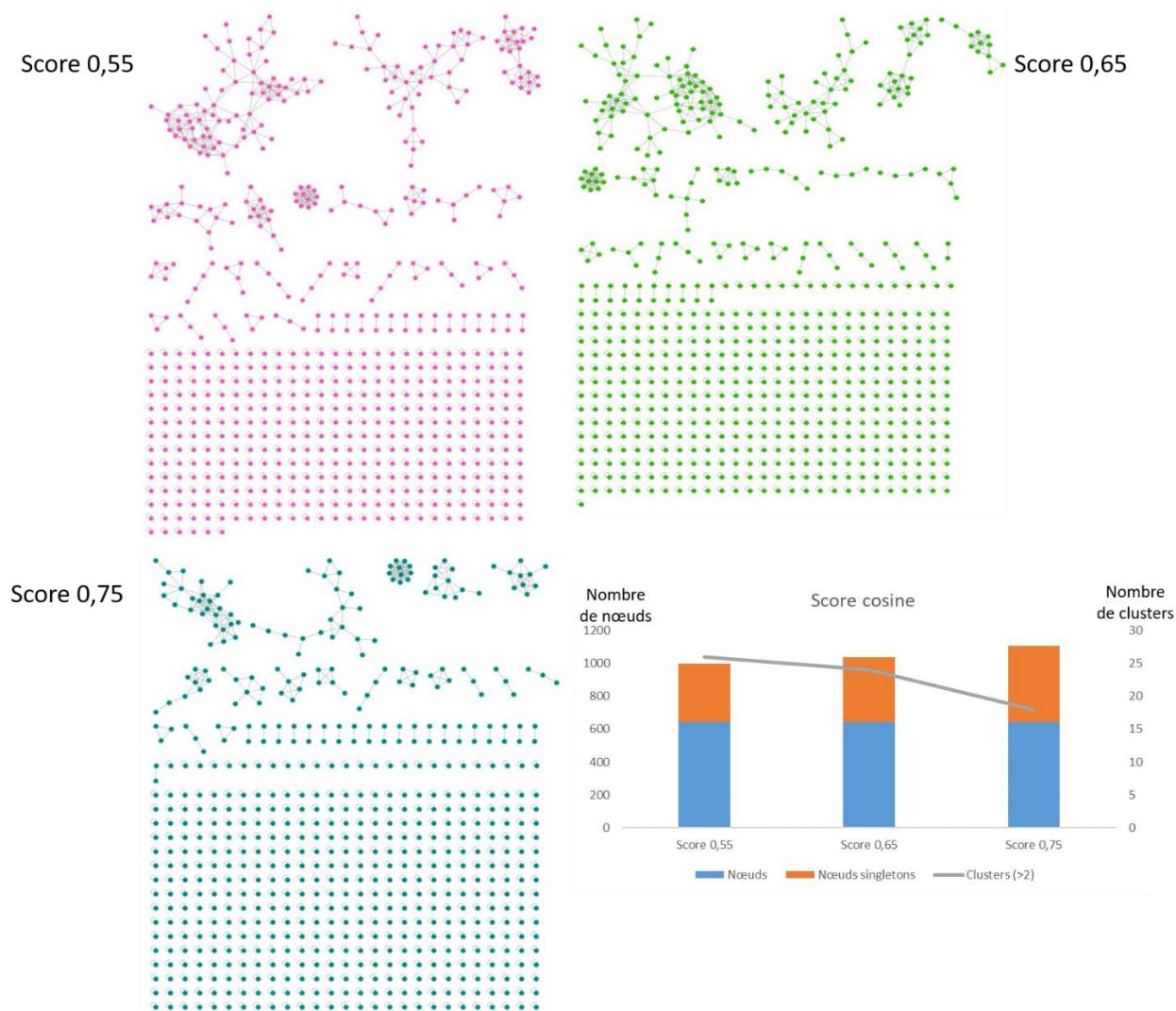


Figure 27 : Réseaux moléculaires de cals cellulaires de *Linum usitatissimum* cultivés à la lumière, obtenus en fonction de différents scores cosine. Graphe résumant le nombre total de nœuds, de nœuds singlons et de clusters (> 2 nœuds) des réseaux moléculaires

#### **IV. RÉSEAUX MOLÉCULAIRES DE *LINUM USITATISSIMUM***

Après l'étude et l'optimisation des paramètres d'acquisition et de traitement de données (sélection de 3 ions précurseurs à fragmenter avec une combinaison de 20-45 eV pour chaque *m/z*, 3 correspondances spectrales et un score cosine de 0,65 pour que deux spectres MS/MS soient liés en cluster), le réseau moléculaire qualitatif (traitement AutoMSn de DataAnalysis) des extraits de cals cellulaires de *Linum usitatissimum* cultivés à la lumière et à l'obscurité a été généré et présenté dans la Figure 28. Après l'annotation sur le réseau moléculaire de certains composés décrits dans la littérature comme étant présents dans le lin, trois familles majoritaires ont été mises en évidence : des lignanes, des dérivés glycosylés et des acides aminés.

La famille de lignanes est sous forme d'un seul cluster, la famille des molécules glycosylés est divisée en trois clusters car les motifs liés au sucre sont spécifiques à chaque cluster, et de la même manière, les acides aminés sont représentés par deux clusters avec deux schémas de fragmentation ; on retrouve le tryptophane dans un cluster et la phénylalanine dans l'autre cluster.

En outre, l'utilisation de l'algorithme AutoMSn (DataAnalysis), ne permettant pas de regrouper les adduits et les fragments générés en source (par exemple les pertes d'eau) avec les ions moléculaires des composés, on retrouve une forte redondance d'information dans les clusters du réseau moléculaire. Cependant, cela permet de lier des fragments générés en source avec leurs ions moléculaires puisqu'ils sont directement liés dans le réseau (spectres MS/MS similaires).

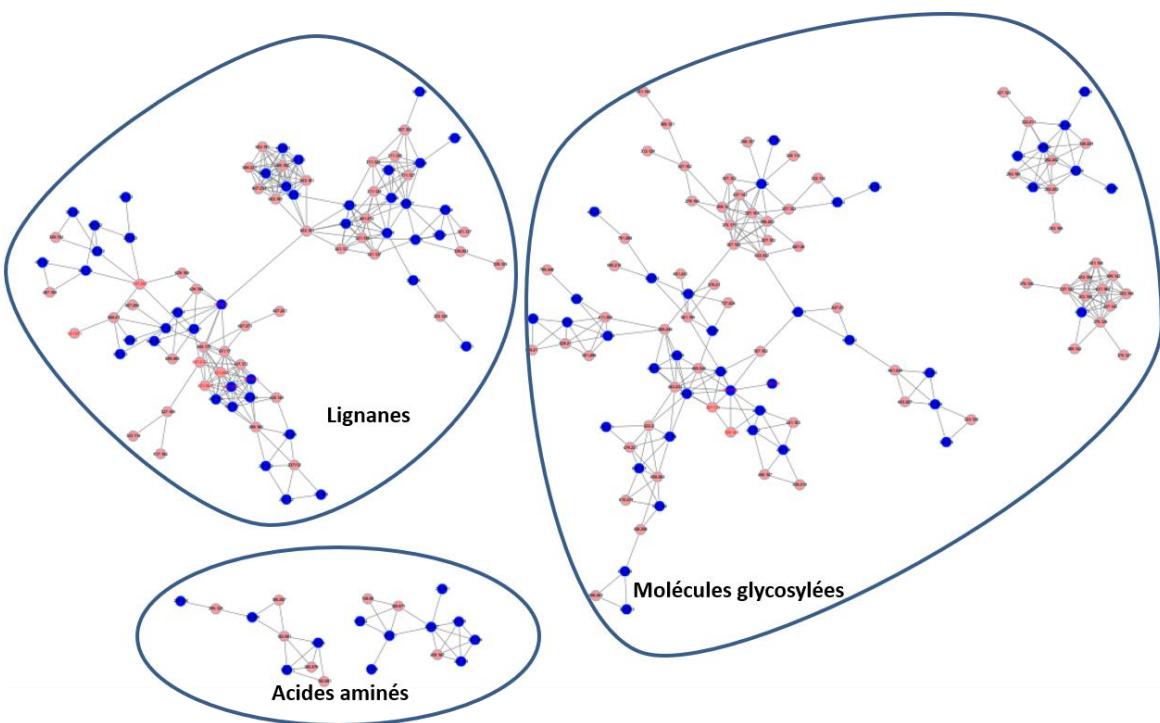


Figure 28 : Réseau moléculaire qualitatif (AutoMSn, DataAnalysis) des extraits des cals cellulaires de *Linum usitatissimum* cultivés à la lumière (rose) et à l'obscurité (bleu).

Toutefois, pour éviter cette redondance d'information, le réseau moléculaire a été généré en utilisant un traitement de données par MetaboScape pour regrouper avec les ions moléculaires, les adduits et les pertes d'eau créées au niveau de la source d'ionisation.

En plus des réseaux moléculaires qualitatifs, qui permettent d'avoir une vue globale de la composition chimique d'un extrait de plante, il est également possible de créer des réseaux moléculaires quantitatifs qui mettent en évidence l'abondance des composés dans un ou plusieurs extraits. En effet, lorsque deux chromatogrammes d'extraits sont superposés, il est difficile d'observer toutes les variations de concentration des composés chimiques notamment dans le cas des composés coélus. Les réseaux quantitatifs permettent de représenter chaque composé accompagné par son aire de pic, obtenue par extraction de l'ion correspondant permettant ainsi de s'affranchir des coélutions. Ceci est d'autant plus intéressant lorsque plusieurs répliques d'échantillons ne sont pas disponibles pour faire une étude métabolomique et des analyses statistiques afin de comparer les teneurs des molécules dans différents extraits.

Les paramètres ont été optimisés sur les analyses UHPLC-HRMS des extraits de cals cellulaires de *Linum usitatissimum* afin de créer une liste (appelée *Bucket*) de caractéristiques *m/z* associées au temps de rétention, à l'aire du pic pour la quantification et à des formules brutes proposées par l'outil SmartFormula de MetaboScape.

Tous les pics chromatographiques ayant une intensité supérieure à 2000 ua avec un nombre minimum de 3 spectres MS ont été sélectionnés pour la création de la *peaklist* afin de récupérer le maximum d'information tout en limitant le bruit de fond.

Les spectres MS/MS d'un seul ion précurseur générés avec les deux énergies de collision sont moyennés pour générer un seul spectre. De plus, pendant l'acquisition, si des spectres MS/MS avec des énergies de collision multiples ont été stockés séparément, ils sont additionnés au cours de l'importation des spectres et le meilleur est importé.

Les caractéristiques spectrales ont été extraites dans un intervalle de temps de 0 à 13 min et de  $m/z$  50 à 1200. Ensuite, les caractéristiques  $[M+H]^+$ ,  $[M+Na]^+$ ,  $[M+K]^+$ ,  $[M+H-H_2O]^+$  d'un seul composé sont groupées en une seule donnée permettant de réduire la redondance d'information. Enfin, la table créée est directement exportée vers GNPS (via la plateforme de transfert de données WinSCP).

Une fois les données exportées sur GNPS, le réseau moléculaire est généré en liant deux spectres lorsqu'ils présentent un score cosiné  $\geq 0,65$ , 3 correspondances spectrales avec une tolérance de 0,01 Da pour les valeurs  $m/z$ . Le nombre maximum de voisins d'un composé est de 99 en désactivant l'outil MSClustering de GNPS puisque les données ont été retraitées avec MetaboScape avant l'export vers GNPS. La recherche dans les bases de données GNPS a été activée, avec des consignes plus restreintes soit un score cosiné  $\geq 0,75$  et un nombre minimum de 4 ions qui concordent.

Le réseau moléculaire présenté dans la Figure 29, présente donc une plus faible redondance d'information que celui de la Figure 28, ainsi que les proportions des composés dans les deux extraits de *Linum usitatissimum* superposés (lumière vs obscurité).

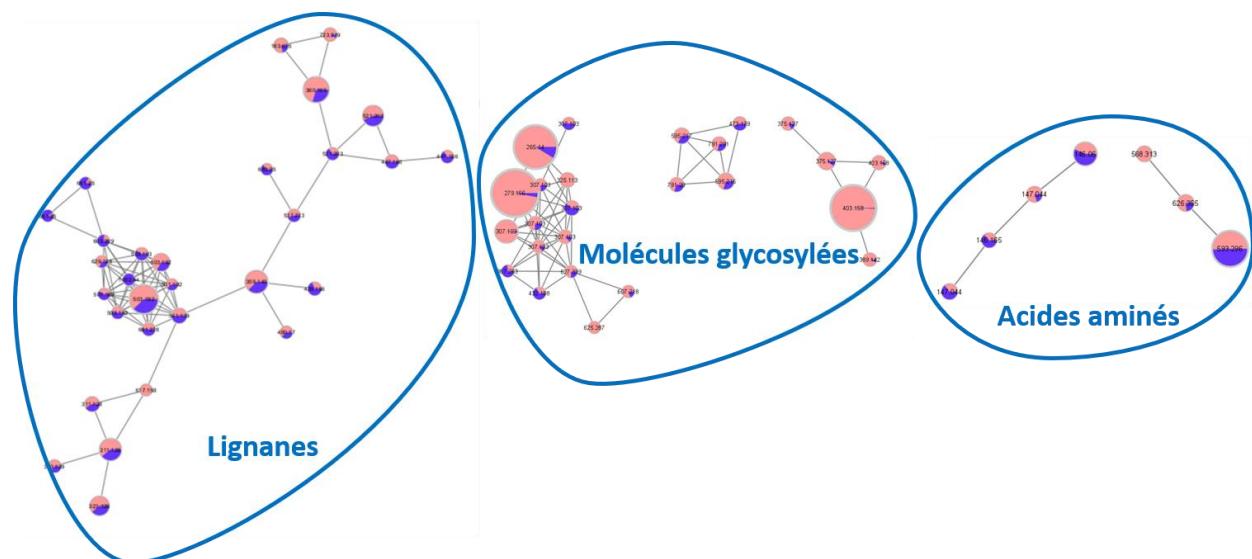


Figure 29 : Réseau moléculaire quantitatif (MetaboScape) des extraits de cals cellulaires de *Linum usitatissimum* cultivés à la lumière (rose) et à l'obscurité (bleu)

## V. COMPARAISON DE PRODUCTION DE LIGNANES

L'analyse du réseau moléculaire quantitatif de la Figure 29 permet de comparer les teneurs en lignanes, famille moléculaire d'intérêt cosmétique, synthétisée par les cals cellulaires de *Linum usitatissimum* dans les différentes conditions de cultures (lumière vs obscurité). Ces molécules sont regroupées dans un seul cluster, ce qui facilite leur identification et la comparaison de leur production dans les deux conditions de culture.

Le cluster représentant les lignanes biosynthétisés par les cals cellulaires de *Linum usitatissimum* cultivés dans à la lumière et à l'obscurité est présenté dans la Figure 30. Leur structure chimique se caractérise par le couplage de deux unités dérivées du 1-phénylpropane, provenant elles-mêmes du métabolisme de la phénylalanine et de précurseurs appelés monolignols<sup>151</sup>. Jusqu'à ce jour, le lin est la matière première végétale contenant le plus de lignanes. Certains d'entre eux ont été identifiés et rapportés sur le cluster de la Figure 30.

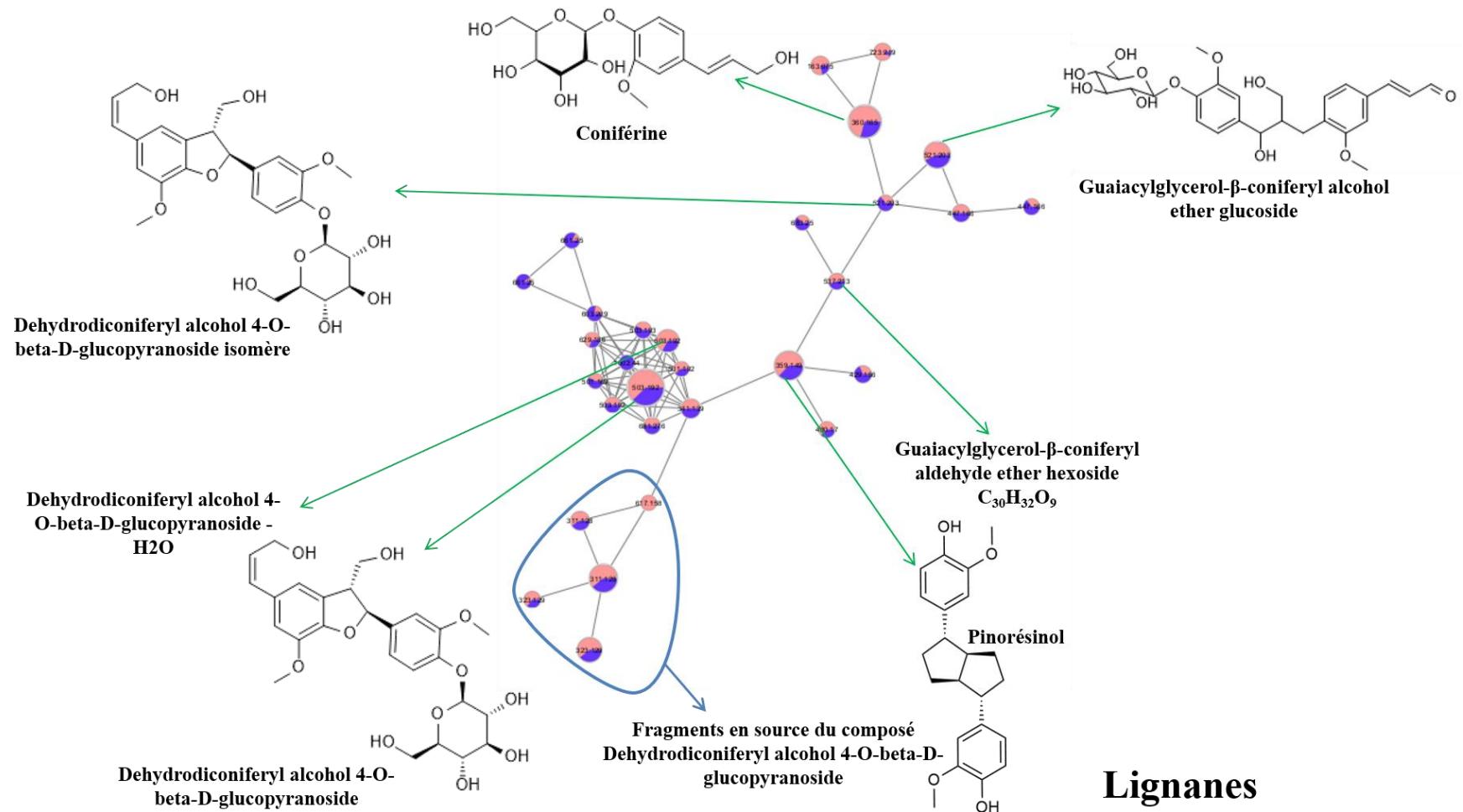


Figure 30 : Cluster représentant les lignanes d'intérêt extraits des cals cellulaires de *Linum usitatissimum* cultivés dans la lumière (rose) et l'obscurité (bleu). Le réseau moléculaire a été obtenu par les analyses UHPLC-MS/HRMS en ionisation (+) ESI

En effet, en mode (+) ESI, les ions suivants sont produits :  $[M+Na]^+$  à  $m/z$  561,1946,  $[M+NH_4]^+$  à  $m/z$  556,2394,  $[M+H-H_2O]^+$  à  $m/z$  521,2022 qui donne après une perte de glucose le  $m/z$  359,1516,  $[M+H-gluc-2H_2O]^+$  à  $m/z$  341,1388. Ces caractéristiques spectrales correspondaient aux deux isomères précédemment décrits dans la littérature comme présents dans les extraits de cellules de lin (Beejmohun *et al.*, 2007) ; les deux isomères *erythro* et *threo* gaïacylglycérol- $\beta$ -coniferyl alcohol ether glucosides  $C_{26}H_{34}O_{12}$  (GGCG). Dans le cluster des lignanes, ils sont représentés par un seul nœud (521) puisqu'ils sortent à 0,13 min d'écart et présentent le même spectre MS/MS.

De plus, on retrouve le  $[M+Na]^+$  à  $m/z$  543,1849 et  $[2M+Na]^+$  à  $m/z$  1063,3822,  $[M+H-H_2O]^+$  à  $m/z$  503,1849 qui perd un glucose en source et donne le  $m/z$  341,1386. Ces caractéristiques chimiques correspondent aux deux isomères déhydrodiconiféryl-alcool-4- $\beta$ -D-glucoside  $C_{26}H_{32}O_{11}$  (DCG), en accord avec la littérature (Beejmohun *et al.* 2007). Nos analyses ont permis de détecter deux isomères attribués ( $t_R=5,71$  min et  $t_R=5,84$  min) au déhydrodicononiféryl-alcool-4- $\beta$ -D-glucoside (DCG), alors que les études précédentes n'ont reporté qu'un seul isomère (Attoumbre *et al.*; Beejmohun *et al.*, 2007).

En outre, on retrouve l'ion moléculaire  $[M+H]^+$  à  $m/z$  537,2128 et son adduit  $[M+Na]^+$  à  $m/z$  559,1949. L'ion  $[M+H-H_2O]^+$  à  $m/z$  519,2027, un ion  $[M+H-H_2O-hexoside]^+$  à  $m/z$  357,1343. Ces caractéristiques MS correspondraient à la forme hexoside de l'éther de gaïacylglycérol- $\beta$ -coniféryl aldéhyde  $C_{30}H_{32}O_9$  décrite par Yao *et al.* (2018). À notre connaissance, nous sommes les premiers à signaler sa présence dans les cellules de lin.

Les adduits moléculaires  $[M+Na]^+$  à  $m/z$  365,1210 et  $[M+NH_4]^+$  à  $m/z$  360,1656 ainsi que l'ion  $[M+H-glucose]$  à  $m/z$  163,0755 ont été observés. Ces caractéristiques correspondraient à la coniférine  $C_{16}H_{22}O_8$ . Cette molécule de type monolignol subit une dimérisation pour former les lignanes dans les cals et suspensions cellulaires<sup>152</sup>.

Enfin, le pinorésinol  $C_{20}H_{22}O_6$ , un lignane de type furofurane a été identifié sans ambiguïté par comparaison avec l'échantillon pur. Son ion moléculaire  $[M+H]^+$  à  $m/z$  359,1496, l'ion  $[M+H-H_2O]^+$  à  $m/z$  341,1386,  $[M+H-2H_2O]^+$  à  $m/z$  323,1267, et  $[M+H-H_2O-2CH_3]^+$  à  $m/z$  311,1275.

Les nœuds sous forme de camembert représentent la proportion des aires des composés dans les deux extraits de cals cellulaires de lin, on peut remarquer que tous les composés sont retrouvés dans les deux extraits de cals cellulaires mais qu'une majorité des lignanes légèrement plus abondante dans les cals cellulaires cultivés à la lumière (rose).

De plus, la taille de chaque nœud est proportionnelle à l'aire du pic chromatographique du composé qu'il représente. Par conséquent, la Figure 30 montre que les lignanes les plus abondants et accumulés dans les cals cellulaires sont un des deux isomères de dehydrodiconiferyl alcohol-4- $\beta$ -D-glucoside, le pinorésinol, le guaiacylglycerol- $\beta$ -coniferyl

alcool ether et la coniférine. Toutefois ces composés ne sont que légèrement surexprimés dans les cals cultivés à la lumière par rapport aux cals cultivés à l'obscurité.

Le réseau moléculaire quantitatif a donc permis de mettre en évidence la présence de lignanes, famille moléculaire ciblée, dans les deux extraits de cals cellulaires cultivés à la lumière et à l'obscurité avec une légère surexpression de certains d'entre eux dans les cals cultivés à la lumière.

Toutefois l'analyse globale du réseau moléculaire a permis de mettre en évidence en plus des lignanes, d'autres familles moléculaires présentes dans les extraits des cals cellulaires du lin et de repérer rapidement les composés affectés par les conditions de culture, soit lumière et obscurité. Ainsi, deux clusters regroupant des composés présents en très faibles intensités dans cals cellulaires cultivés à l'obscurité et majoritairement accumulés dans les cals cellulaires cultivés à la lumière (Figure 31) apparaissent. Ces composés sont des dérivés glycosylés élus au début du chromatogramme.

Parmi eux, on retrouve la linamarine  $C_{10}H_{17}NO_6$  et ses isomères qui forment un pic large à 2 min, c'est un glycoside cyanogène produit par le lin et représenté sur le réseau moléculaire par son adduit  $[M+NH_4]^+$  à  $m/z$  265,1401. On observe également l'heterodendrine  $C_{10}H_{17}NO_6$  élue à 3,63 min, de structure très proche de la linamarine représentée par son adduit  $[M+NH_4]^+$  à  $m/z$  279,1552. Les autres nœuds du même cluster n'ont à ce stade pas pu être précisément identifiés. En effet, on retrouve plusieurs nœuds représentant le  $m/z$  307,1029 avec la formule brute de  $C_{12}H_{18}O_9$ , c'est un fragment commun de plusieurs composés glycosylés qui est généré en source.

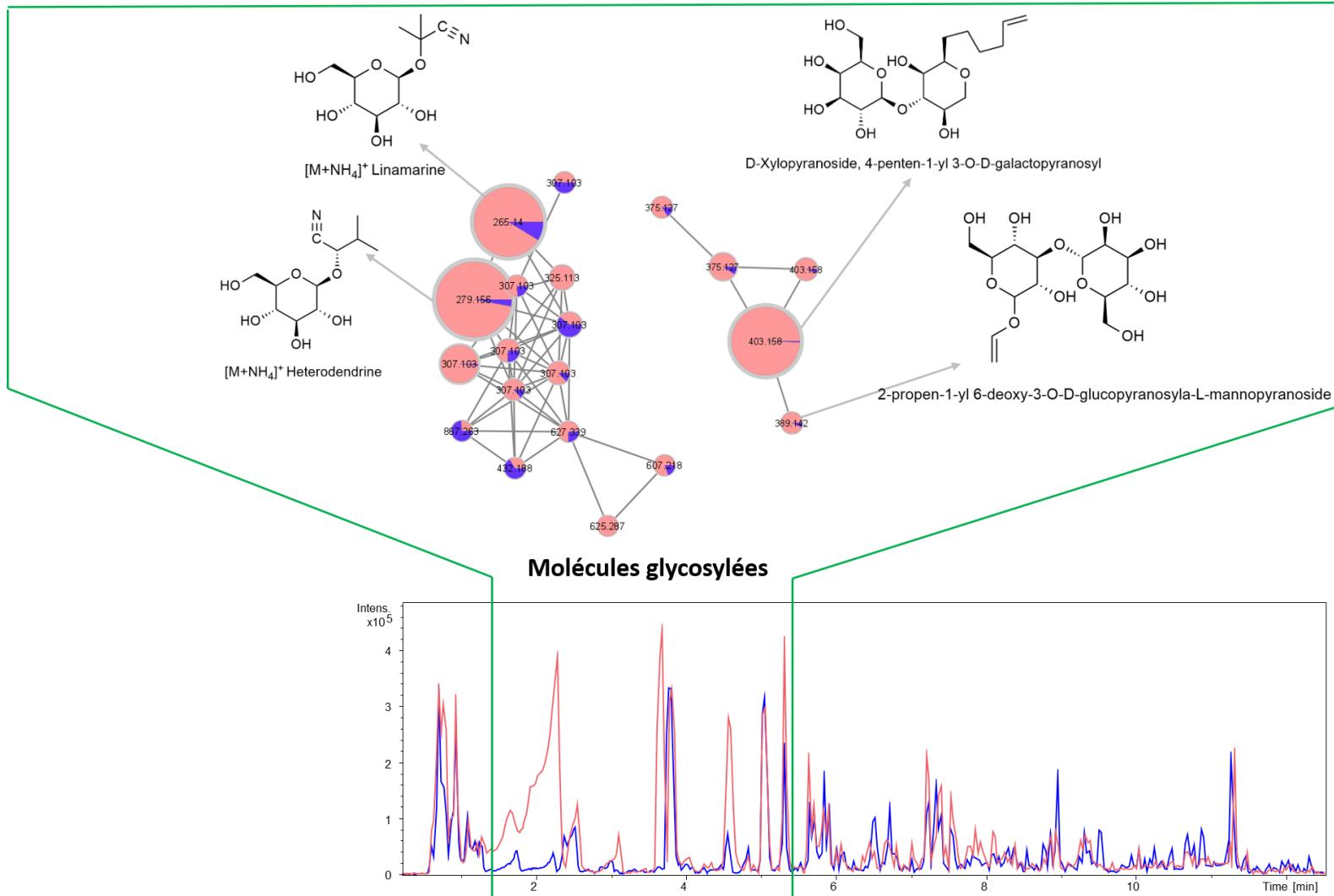


Figure 31 : Clusters regroupant les composés majoritairement affectés par les conditions de culture des cals cellulaires de *Linum usitatissimum*, soit lumière (rose) soit obscurité (bleu)

L'autre cluster de droite regroupe également des composés glycosylés avec des fragments et perte communes. Ces composés n'ont pas été décrits dans la littérature et ont les formules brutes suivantes :

$[M+Na]^+$  à  $m/z$  403,1582 et  $[2M+Na]^+$  à  $m/z$  783,3274 correspondant à la formule brute  $C_{16}H_{28}O_{10}$ .

$[M+Na]^+$  à  $m/z$  389,1426 et  $[2M+Na]^+$  à  $m/z$  755,2980 correspondant à la formule brute  $C_{15}H_{26}O_{10}$ .

$[M+Na]^+$  à  $m/z$  375,1267 et  $[2M+Na]^+$  à  $m/z$  727,2657 correspondant à la formule brute  $C_{14}H_{24}O_{10}$ .

Ces formules brutes ont des écarts de  $CH_2$  et n'ont pas été décrites dans la littérature du lin, et peu de formules développées ont été proposées par les bases de données publiques (SciFinder, DNP). Parmi les structures proposées, celles qui concordent avec le schéma de fragmentation sur les spectres MS/MS sont  $\alpha$ -D-Xylopyranoside, 4-penten-1-yl 3-O- $\beta$ -D-galactopyranosyl pour  $C_{16}H_{28}O_{10}$  ( $m/z$  241,1045 correspondant à  $C_{10}H_{18}O_5^+$  après une perte d'une unité de sucre 162u détectée à  $m/z$  185,0436  $C_6H_{10}O_5^+$ ) et  $\alpha$ -L-Mannopyranoside, 2-propen-1-yl 6-deoxy-3-O- $\alpha$ -D-glucopyranosyl pour  $C_{15}H_{26}O_{10}$  ( $m/z$  371,1312  $C_{15}H_{24}O_9$  correspondant à une perte d'eau  $[M-H_2O+Na]^+$ ,  $[M-162+Na]^+$  à  $m/z$  227,0886  $C_9H_{16}O_5$  après une perte d'une unité de sucre détectée à  $m/z$  185,0436  $C_6H_{10}O_5^+$ ). Ces molécules sont présentées sur la Figure 31. Un hétéroside a été proposé pour la formule brute  $C_{14}H_{24}O_{10}$ .

Une prédiction de famille moléculaire pour ce cluster a été lancée sur la plateforme GNPS-MS/MSLDA (*Masse2Motifs-Latent Dirichlet Allocation*) et la famille de "saccharolipide" a été proposée. De plus, une prédiction de structure *in-silico* a également été recherchée sur GNPS-NAP (*Network Annotation Propagation*) en utilisant les bases de données GNPS, néanmoins, les prédictions de structures n'étaient pas cohérentes avec les spectres MS/MS expérimentaux.

L'ensemble de ces composés glycosylés annotés sont donc largement impactés par les différentes conditions de culture et sont très accumulés lors de la culture à la lumière par rapport à l'obscurité.

En conclusion, les réseaux moléculaires qualitatifs ont permis de mettre en évidence les familles majoritaires contenues dans les extraits de cals cellulaires de lin et de propager l'annotation des composés dans le réseau. Les réseaux moléculaires quantitatifs ont permis de comparer l'accumulation des métabolites dans deux conditions de culture. Nous avons observé une accumulation de la majorité des composés dans les cals cultivés en lumière.

En effet, les lignanes, principaux métabolites recherchés, sont légèrement accumulés dans les cals cellulaires cultivés à la lumière. De plus, les réseaux moléculaires ont mis en évidence rapidement un cluster regroupant des composés glycosylés très intenses dans les cals cellulaires cultivés à la lumière par rapport à l'obscurité. Toutefois, ce groupe de molécules glycosylées comprend des composés cyanogènes qui peuvent se dégrader et libérer des groupements toxiques, il serait donc intéressant d'étudier la stabilité et la toxicité de cette famille afin de conclure quant à la méthode de culture adéquate.

## VI. CONCLUSION

L'extrait de lin (*Linum usitatissimum* L.) est considéré comme un ingrédient cosmétique potentiel dans le monde entier notamment pour sa richesse en lignanes. Il est donc intéressant de stimuler la production de cette classe moléculaire en évitant la surexploitation végétale.

La technologie de cals cellulaires végétaux a été proposée comme un système utile pour la production de biomasse végétale capable de produire et d'accumuler des composés bioactifs. Par conséquent, les collaborateurs du projet ValBioCosm' se sont intéressés à la culture de cals cellulaires de lin afin de garantir la production de ses métabolites secondaires indépendamment des conditions climatiques et saisonnières et afin d'améliorer la production de lignanes ayant un intérêt économique.

Pour ce faire, des cals cellulaires de *Linum usitatissimum* ont été cultivés dans deux conditions d'éclairage puisque la lumière a été mentionnée comme un inducteur de stress chez plusieurs espèces végétales, qui déclenche la biosynthèse de métabolites secondaires<sup>153,154</sup>. Des cals cellulaires ont donc été cultivés à la lumière et à l'obscurité afin d'estimer l'effet des paramètres de culture sur l'accumulation des composés et afin de repérer la condition de culture qui permet une meilleure accumulation des métabolites secondaires de lin, notamment des lignanes.

Des méthodes d'extraction, de séparation et d'analyse ont donc été développées afin d'extraire un large éventail de métabolites produits dans les cals cellulaires de lin. Les analyses UHPLC-HRMS ont ensuite été utilisées pour créer les réseaux moléculaires des extraits de cals cellulaires de lin.

D'une part, une étude qualitative a été effectuée afin d'étudier l'influence des paramètres d'acquisition et de traitement sur la construction des réseaux moléculaires. Parmi ces paramètres, on retrouve le nombre d'ions précurseurs, l'énergie de collision, le traitement des données et les paramètres de la plateforme GNPS (score cosine, nombre de correspondance spectrales) et pour mettre en évidence les familles majoritaires contenues dans les extraits de lin. D'autre part, une étude quantitative a été réalisée afin de comparer l'accumulation des métabolites secondaires dans les deux conditions de culture des cals cellulaires de lin (lumière vs obscurité). Cette étude a montré une influence bénéfique de la lumière sur l'accumulation relativement élevée des métabolites secondaires dans les cals cellulaires de lin par rapport à l'obscurité notamment sur les dérivés glycosylés et non sur les lignanes à priori envisagés.

Par conséquent, dans cette étude, en utilisant une méthode de chromatographie à haute résolution et l'approche de réseaux moléculaires, nous avons montré pour la première fois que les cals de lin cultivés dans la lumière sont capables d'accumuler au moins 11 métabolites secondaires (6 lignanes et 5 dérivés glycosylés). Cela montre la polyvalence de ce système

cellulaire et l'utilité de développer de nouvelles lignées cellulaires de *Linum usitatissimum* pour améliorer le développement métabolique spécifique en fonction des applications ciblées.

Néanmoins, cette étude a également permis d'appréhender rapidement les limites de l'approche des réseaux moléculaires et de la plateforme GNPS. En effet, le réseau moléculaire est un bon moyen de première approche pour lier les composés ayant des similarités spectrales et permet d'associer les molécules appartenant aux mêmes familles moléculaires mais nécessite l'utilisation d'informations complémentaires et de recherche manuelle pour compléter l'élucidation structurale des composés d'un échantillon. De plus, cette approche de réseaux moléculaires est très dépendante des paramètres d'acquisition, il est donc très facile d'obtenir des réseaux aberrants. Par exemple si une famille se fragmente en source, elle sera représentée par un cluster riche en nœuds contrairement à une autre famille qui se fragmente seulement en cellule de collision.

En outre, la plateforme GNPS comporte un système de prédiction de famille moléculaire et de fragmentation *in-silico* nécessitant des améliorations qui seront apportées au fur et à mesure de son utilisation. De plus, les bases de données spectrales permettant d'obtenir les informations de déréplication sont également limitées par le fait que leurs spectres MS/MS sont obtenus par différents spectromètres de masses et différentes méthodes.

L'approche des réseaux moléculaires et la plateforme GNPS sont en plein émergence et leurs limites sont continuellement repoussées.



# **CHAPITRE 3 – APPROCHE MÉTABOLOMIQUE ET RÉSEAU MOLÉCULAIRE POUR LE CRIBLAGE ET LA CARACTÉRISATION DE BIOMARQUEURS**

---



## I. INTRODUCTION

La tomate, *Solanum lycopersicum* L. (*Solanaceae*) est l'un des légumes les plus consommés au monde et joue un rôle important dans l'alimentation humaine, qu'elle soit consommée fraîche ou sous forme de produits cuits transformés. Malgré une culture tropicale (originaire de la région Pérou-Equateur), la tomate est cultivée dans presque tous les coins du monde et elle est plantée non seulement en plein champ mais aussi sous serre<sup>155</sup>. Cependant, plusieurs agents pathogènes peuvent attaquer les plants de tomates et provoquer des dégâts et pertes de rendement énormes.

Les plantes produisent une vaste gamme de métabolites secondaires qui peuvent être utilisés comme substances toxiques, répulsives ou anti-nutritives et qui ont un impact direct sur le comportement, le développement, la fécondité ou la survie des organismes nuisibles<sup>156,157</sup>.

Afin de protéger les plants de tomate contre l'attaque des ravageurs tout en limitant l'utilisation de produits phytosanitaires il serait important de comprendre leur système de défense afin de mettre en évidence les molécules biosynthétisées pour se protéger. Dans ce sens, plusieurs études métabolomiques ont été développées et la plupart portent sur les parties aériennes (principalement les feuilles et les fruits). Ces études ont montré que les cultures de tomates sont attaquées par divers types de parasites, ce qui implique des stratégies de défense chimique diverses et complexes<sup>158,159</sup>. Parmi ces molécules de défense, les alcaloïdes stéroïdiens et leurs formes glycosylées (glycoalcaloïdes) sont les principaux métabolites spécialisés produits par le genre *Solanum* pour se protéger contre les parasites<sup>160,161,162</sup>. Les esters d'hydroxycinnamoyle, les flavonoïdes et les phénolamides sont d'autres métabolites secondaires potentiellement impliqués dans les défenses de la tomate<sup>163</sup>. Ces études montrent également que certains composés, comme les alcaloïdes, sont constitutifs (toujours présents dans le tissu végétal comme défense préventive) alors que d'autres, comme les phénolamides, sont induits par une attaque de parasites (défense active).

Néanmoins, on sait peu de choses sur la composition chimique des racines des plants de tomate malgré leur importance dans le développement de la plante et de ses défenses. En effet, des métabolites spécialisés qui protègent les feuilles peuvent être synthétisés dans les tissus racinaires<sup>164</sup>, ce qui implique une connexion entre les parties aériennes et souterraines.

Toutefois, la composition chimique des racines peut également varier avec le stade de développement des parties souterraines impliquant des variations chimiques indépendantes de l'attaque d'agents pathogènes. Il est donc important de caractériser le métabolome des racines et de comprendre sa variation au court du temps afin de mieux étudier l'impact de différents ravageurs sur les parties racinaires et aériennes des plants de tomates.

C'est pourquoi, une approche métabolomique non ciblée a été développée et combinée aux réseaux moléculaires dans le but d'étudier le métabolome des racines de plants de tomate et afin de cibler et identifier les composés chimiques qui varient dans les racines au cours du développement des plants. Cette étude constitue la base d'une étude plus large concernant l'identification des molécules de défenses biosynthétisées par les parties racinaires et aériennes lorsque les plants de tomates sont soumis à l'attaque de quatre types de ravageurs *Macrosiphum euphorbiae*, *Helicoverpa armigera*, *Oidiumneo lycopersici*, *Meloidogyne incognita*, seuls ou en combinaisons, coordonnée par nos collègues de l'**INRAE** et l'**ICN** à Nice.

## II. MÉTABOLOMIQUE NON-CIBLÉE

La Figure 32 présente le schéma général des différentes étapes qui composent généralement une analyse métabolomique par LC-HRMS et que nous avons mis en œuvre pour l'étude des mécanismes de défense de *Solanum lycopersicum* L.

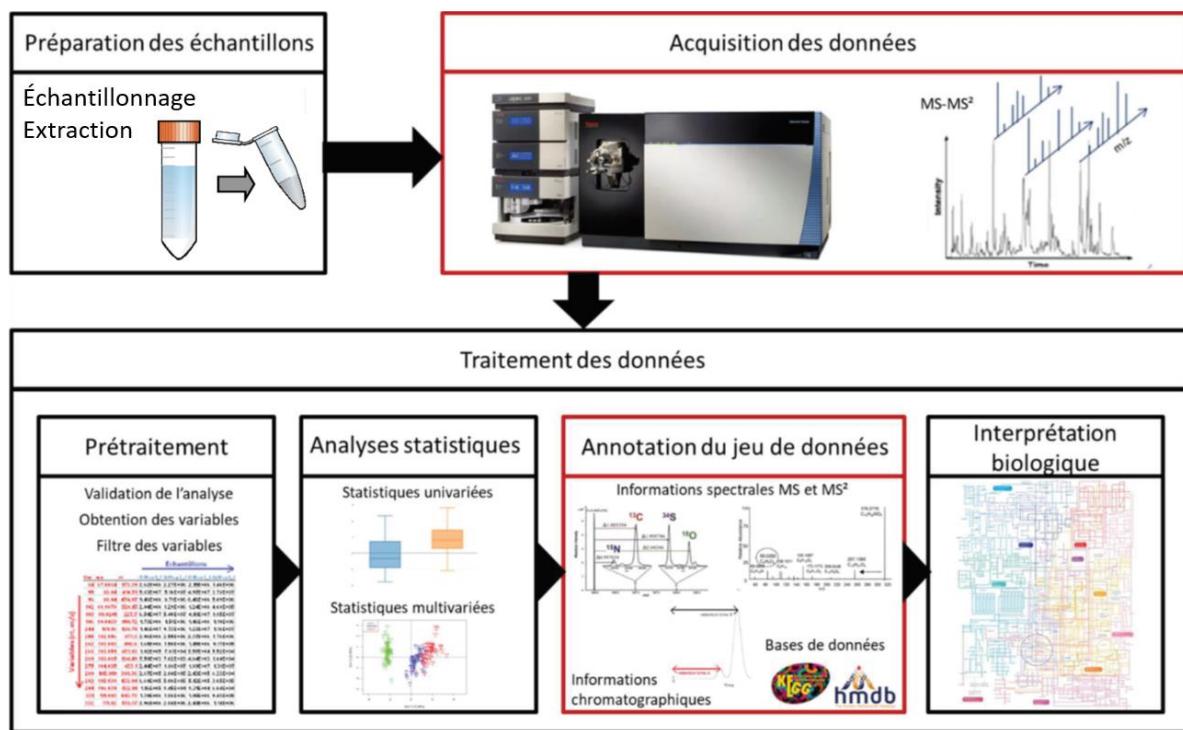


Figure 32 : Schéma général d'une analyse métabolomique par LC-HRMS présenté par Pierre BARBIER (2019)<sup>295</sup>

Après la récolte des échantillons végétaux, qui doit être rigoureusement effectuée car elle conditionne en partie la pertinence de l'étude et des résultats obtenus, la première étape est la préparation des échantillons. Son objectif est d'extraire de façon la plus complète possible les métabolites présents dans un échantillon. Cette préparation est dépendante du type de matrice étudiée. La seconde étape est l'acquisition des données fournies par la séparation des analytes au niveau du système chromatographique et l'acquisition des spectres de masse par le spectromètre. À ce niveau, différentes approches sont étudiées pour améliorer la qualité et la quantité d'informations acquises, notamment grâce à l'utilisation de spectromètre de masse à haute résolution. Après leur acquisition, les données sont traitées. Cette étape est cruciale pour extraire les informations pertinentes et utilisables à partir des données acquises. L'étape d'analyse statistique permet d'identifier des variables discriminantes entre les groupes d'individus étudiés. Cette étape peut mettre en évidence de potentiels biomarqueurs et l'ensemble de ces biomarqueurs constitue la signature métabolique liée à la perturbation étudiée. L'étape finale consiste en l'interprétation biologique de la signature métabolique obtenue. Cette étape est à l'heure actuelle la plus complexe.

## A. Préparation des échantillons

La préparation des échantillons comporte quatre étapes essentielles : la collecte d'échantillons, l'arrêt du métabolisme dans le but d'éviter toute modification du métabolome après la collecte, la conservation de l'échantillon avant analyse et l'extraction des métabolites. Toutes ces étapes peuvent être sources de biais et sont donc déterminantes dans la qualité globale d'une étude métabolomique.

Une préparation d'échantillons doit être non sélective, simple, rapide, reproductible, et doit comprendre un nombre limité d'étapes<sup>165</sup>. La non-sélectivité de la méthode d'extraction peut être caractérisée par sa capacité à préserver une large gamme de métabolites avec des taux élevés de recouvrement pour les différentes classes de composés afin d'obtenir une sensibilité maximale. La simplicité et la rapidité de la méthode permet de minimiser la dégradation des métabolites ainsi que leur adsorption sur différentes matrices tout en évitant l'introduction de composés exogènes.

La reproductibilité est également un facteur essentiel. En effet, les analyses métabolomiques impliquent une comparaison de données entre individus. Il est donc essentiel que la variabilité générée par l'échantillonnage et l'extraction ainsi que le stockage soit la plus faible possible afin de ne pas masquer les variations (faux négatifs) et ne pas créer des variations (faux positifs) entre les différents échantillons. Enfin l'arrêt du métabolisme est important pour éviter des dégradations ou des surexpressions chimiques (oxydations) ou biologiques (actions enzymatiques) des métabolites après l'instant du prélèvement.

## B. Acquisition des données

Une fois les métabolites extraits, ils sont analysés en LC-HRMS. De la même manière que pour la préparation des échantillons, cette étape ne doit pas être une source de biais. En effet, l'ensemble des échantillons doit être analysé dans une même séquence, le même jour en suivant un ordre aléatoire afin de minimiser au maximum les variations expérimentales. Il est également important d'injecter des échantillons contrôles QC (*Quality Control*) au début, à intervalles réguliers et à la fin de la séquence d'analyse, ils servent à suivre la stabilité temporelle de l'ensemble des performances de l'analyse LC-MS et à détecter d'éventuelles dérives analytiques<sup>166</sup>. Il est également important d'intégrer des étalons afin de suivre la reproductibilité des analyses et de vérifier leur bon déroulement. Les étalons peuvent être des composés injectés indépendants, ou ajoutés directement dans le solvant d'extraction ou encore dans l'échantillon avant l'analyse. Enfin, l'utilisation d'échantillons "blancs" est essentielle à la validation de l'analyse. Dans l'idéal, un blanc correspond à une matrice sans analytes d'intérêt. La préparation de cet échantillon doit être comparable à celle des échantillons réels en termes d'extraction, de stockage et d'analyse<sup>167</sup>.

## C. Prétraitement des données

Une fois l'analyse validée, la première étape du traitement de données est la détection automatique des informations présentes dans les fichiers d'acquisition. Il s'agit à la fois d'informations spectrales avec la détection des différents signaux de masse mais aussi d'informations chromatographiques avec la détection des pics chromatographiques correspondant aux différents ions détectés au cours du temps. L'objectif de cette étape est de transformer les signaux bruts en un format matriciel permettant de traiter facilement et de manière groupée les données de l'ensemble des échantillons.

Différents logiciels permettent de réaliser cette étape. On distingue les logiciels commerciaux spécifiques d'un constructeur donné (MarkerLynx™ pour Waters, MarkerView™ pour AB SCIEX, MetaboScape pour Bruker, etc.) qui ne sont utilisables qu'avec des données acquises sur les appareils provenant du même fournisseur, et des logiciels libres et gratuits (XCMS, MZmine). Les fournisseurs d'instruments utilisant des formats de données propriétaires, il est souvent nécessaire de convertir les données brutes en fichiers de formats universels afin de pouvoir utiliser ces logiciels gratuits. Deux types de formats universels ont été développés pour les données de spectrométrie de masse : le netCDF (*Network Common Data Form*) et le mzXML<sup>168,169</sup>.

Le traitement des données débute par une étape de soustraction du bruit de fond spectral. Les signaux sont ensuite extraits, alignés et restitués sous la forme d'une matrice contenant des variables, identifiées par un rapport  $m/z$  et un temps de rétention, et des numéros d'échantillons. Les valeurs contenues dans la matrice correspondent aux aires ou aux abondances de chaque variable pour chacun des échantillons.

## D. Analyses Statistiques

L'étape suivante au cours d'une analyse métabolomique est l'utilisation des statistiques pour faire ressortir les métabolites pouvant être marqueurs d'un statut physiologique. Deux grandes catégories d'analyses statistiques peuvent être utilisées et sont présentées ci-après.

### 1) Analyses univariées

Les tests statistiques univariés permettent de comparer les différents échantillons métabolites par métabolites. Il s'agit en général de test de comparaison des moyennes. Il existe plusieurs tests paramétriques (T-test, Welch-Test, ANOVA, etc) ou non paramétriques (Mann-Withney-Wilcoxon, Kruskal-Wallis, etc)<sup>170</sup>. Ces approches permettent de décrire la variable, de représenter sa dispersion lors de la comparaison d'échantillons ou lors de la validation de méthodes.

Par exemple, le test *Volcano Plot* est un type de diagramme de dispersion qui montre la signification statistique (P-value) en fonction de l'ampleur du changement (Fold change). Il permet une identification visuelle rapide des variables présentant des variations importantes qui sont également statistiquement significatives<sup>171</sup>.

## 2) Analyses multivariées

Ces outils vont traiter l'ensemble du jeu de données comme un seul élément. Cela permet une visualisation plus immédiate de la dispersion des échantillons. Elles sont basées sur l'utilisation de variables latentes qui sont calculées comme des combinaisons linéaires des variables initiales et permettent de mettre en avant des interactions existantes entre ces différentes variables. Il existe des analyses multivariées supervisées et non supervisées. Les analyses statistiques multivariées non supervisées vont chercher à faire ressortir la variance qui existe dans le jeu de données sans *a priori* sur la nature de chaque échantillon. Les méthodes d'analyses statistiques multivariées supervisées vont chercher à mettre en évidence une hiérarchisation du jeu de données à partir de groupes d'appartenance prédéfinis pour chaque échantillon<sup>172</sup>.

La PCA (*Principal Component Analysis*) est l'analyse descriptive la plus simple et la plus connue des techniques d'analyse de données multivariées non supervisées. C'est souvent la première méthode employée pour l'analyse des données afin d'étudier leur organisation et observer si les variations liées à l'appartenance à un groupe font partie des causes majeures de variance ou si cette dernière est plutôt reliée à un effet expérimental. Elle est suivie d'une analyse dite explicative permettant l'interprétation des données plus en profondeur afin de renforcer la séparation des groupes et de déterminer quelles sont les variables initiales du jeu de données qui contribuent à l'obtention de cette séparation.

La PLS (*Partial Least Square*) ainsi que la PLS-DA (*Partial Least Squares Discriminant Analyses*) sont des analyses multivariées supervisées qui permettent la construction d'un modèle explicatif et discriminant respectivement. Ce modèle permet ainsi d'isoler les métabolites dont l'intensité est caractéristique d'un état biologique donné et qui contribuent à la séparation des différents groupes que forment les échantillons. Ce modèle construit à partir d'un jeu de données initiales peut également être utilisé pour expliquer un nouveau jeu de données. On cherche ainsi à prédire la classe d'appartenance des individus du nouveau jeu de données.

### E. Annotation des biomarqueurs

Une fois la liste de variables pertinentes permettant de discriminer les groupes étudiés sont mis en évidence, ces variables doivent être annotées pour pouvoir conduire ensuite à une interprétation biologique des données. Cette étape consiste à attribuer un nom de métabolite

aux variables détectées. Elle est cruciale au cours du processus métabolomique et reste aujourd’hui une des principales étapes limitantes lors d’une analyse par LC-HRMS<sup>173</sup>. Ces limites sont dues à notre méconnaissance de l’intégralité du métabolome<sup>174</sup>.

Des travaux ont cherché à raffiner le système de score pour les composés putativement annotés se basant sur des éléments chromatographiques et des éléments de spectrométrie de masse tels que les spectres de fragmentation ou le massif isotopique<sup>175</sup>. Cependant, ces approches sont compliquées et non consensuelles et les niveaux d’identification expliqué dans le chapitre 1 restent pour l’instant l’outil de référence.

Une fois identifiés, les métabolites permettant de discriminer les différents groupes vont pouvoir être interprétés biologiquement afin de chercher une cohérence entre ces biomarqueurs et le facteur étudié.

## F. Interprétation biologique

L’interprétation biologique des résultats se base principalement sur la visualisation des différentes voies métaboliques connues et sur le positionnement des métabolites identifiés sur ces voies. Ces voies sont en général focalisées sur un nombre restreint de substrats initiaux ou produits finaux. Chaque voie va alors représenter les différentes étapes de biosynthèse ou de biodégradation de ce petit nombre de métabolites en les connectant par l’intermédiaire des enzymes capables de catalyser les réactions. Cette étape consiste donc à expliquer la cohérence biologique des métabolites discriminants.

### III. CRIBLAGE ET IDENTIFICATION DE BIOMARQUEURS

Dans ce travail, une approche métabolomique non ciblée a été développée pour caractériser la diversité phytochimique des racines de tomate et pour mettre en évidence son évolution au cours du développement des plants. Comme décrit précédemment, la mise en œuvre d'une approche métabolomique, soit la collecte des échantillons, l'extraction des métabolites, les analyses chromatographiques et spectrométriques, le traitement des données comprenant les statistiques et l'identification des métabolites doit être soigneusement optimisée afin d'obtenir le plus grand nombre de caractéristiques pertinentes. La Figure 33 montre les différentes étapes qui ont été étudiées et optimisées pour l'étude des variations chimiques des racines de tomates à différents stades de développement en utilisant une approche métabolomique et les réseaux moléculaires.

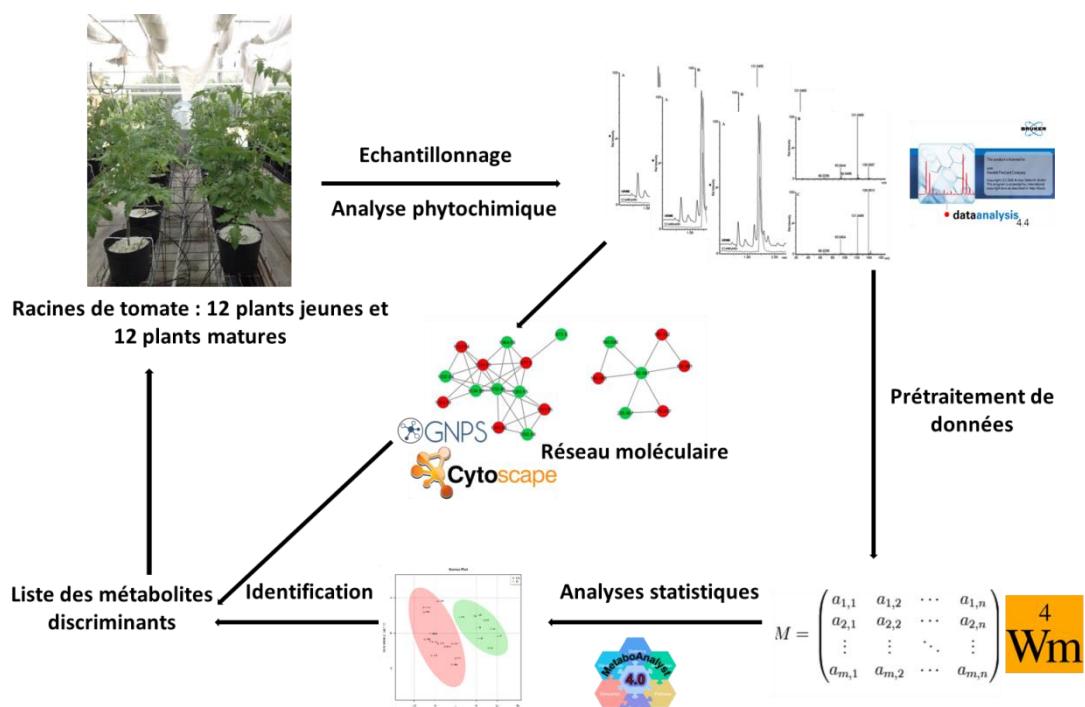


Figure 33: schéma récapitulatif de la démarche développée pour l'identification de biomarqueurs discriminants des racines de tomate à différents stades de développement

#### A. Préparation des échantillons

##### 1) Culture en mésocosme

Des plants de tomates, *Solanum lycopersicum* appartenant à la famille des Solanacées, ont été cultivés en mésocosme à l'INRAE de Nice (Figure 34). Des jeunes plants d'une semaine ont été séparés dans des pots et plantés dans un mélange de terre et de perlite sous des conditions de culture contrôlées et détaillées en Annexe 4. Le 34<sup>ème</sup> jour après le semis (plants

de tomates présentant 3 feuilles complètes), les racines de tomates ont été échantillonnées et stockées à -80°C en tant qu'échantillons au stade jeune ( $n = 12$  individus). De la même manière, des racines ont été prélevées sur 12 autres plants en tant qu'échantillons de plants matures (plants en fleur avec 9 feuilles complètes) le 62<sup>ème</sup> jour, également stockés à -80°C. Tous les échantillons ont été lyophilisés pendant 48 heures avant d'être broyés avec un broyeur à boulets (Retsch MM301) par agitation (30 Hz) formant ainsi une poudre de racine fine et homogène.



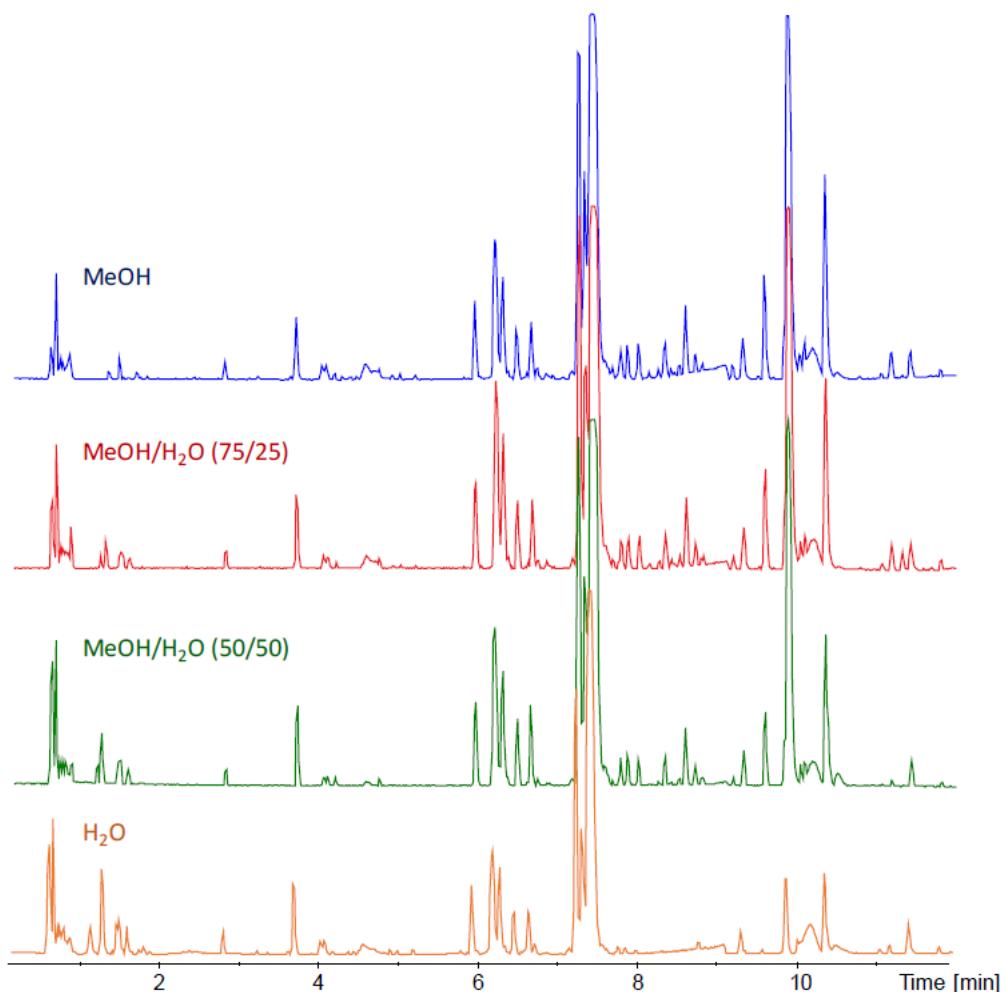
Figure 34 : Culture des plants de tomate en mésocosme

## 2) Extraction

Afin d'extraire un large éventail de composés présentant de propriétés physico-chimiques et pouvant se trouver à des concentrations variables un mélange de solvant (tels que l'éthanol, le méthanol et l'eau) est souvent préconisé<sup>176,177</sup>. En outre, les travaux métabolomiques précédents sur la tomate ont utilisé du méthanol pur ou à 75% pour l'extraction non sélective des composés<sup>178,179</sup> et l'ajout d'acide formique dans le solvant d'extraction contribue à l'extraction des glycoalcaloïdes<sup>180</sup>.

Dans ce contexte, différentes proportions du mélange méthanol : eau acidifié avec 0,1% d'acide formique (100:0), (75:25), (50:50), (0:100) ont été testées afin de maximiser le nombre de composés extraits et leurs intensités à partir des racines de tomate. Les différents extraits analysés suivant la méthode présentée en Annexe 1 et leurs profils chromatographiques ont été comparés (Figure 35) et ont permis de choisir le rapport (75:25) avec 0,1% d'acide formique puisque l'intensité des glycoalcaloïdes est plus importante que lorsque l'acide

formique n'a pas été ajouté au solvant d'extraction. Ce solvant a donc permis d'extraire des composés polaires et moyennement polaires.



*Figure 35 : Les BPCs (Base Peak Chromatograms) des analyses UHPLC-HRMS en mode (+) ESI des extraits de racines de tomates obtenus avec différentes proportions (MeOH/H<sub>2</sub>O acidifiée avec 0,1 % Acide formique) comme solvant d'extraction*

Enfin, deux étalons internes ont été ajoutés au solvant d'extraction : l'acide gallique et la digitonine. Tous les deux sont absents des extraits de racines et n'induisent pas de coélution. Le premier a été sélectionné par rapport à la présence de composés phénoliques tandis que le second, une molécule de type saponine, a été choisi pour imiter le comportement des glycoalcaloïdes.

Le protocole d'extraction optimisé est le suivant : une quantité de 5 mg des racines séchées et broyées ont été introduites dans 400 µL de mélange méthanol/eau (75/25) acidifié avec de l'acide formique à 0,1%. La digitonine (25 ng/µL) et l'acide gallique (50 ng/µL) ont été introduits comme étalons internes dans le solvant d'extraction afin qu'ils suivent l'ensemble du protocole de préparation des échantillons. L'extraction a été assistée par les ultrasons (50 Hz) à

température ambiante pendant 20 min. Ensuite, les 24 échantillons (12 échantillons de racines jeunes et 12 matures) ont été centrifugés pendant 10 min à 10 000 g et les surnageants ont été transférés dans des flacons pour analyse. Deux échantillons préparés à partir d'un mélange des échantillons de racines jeunes ou matures (100 µL de chaque échantillon) ont été utilisés comme échantillons contrôle au cours des séquences d'analyse.

## B. Acquisition des données

Les composés chimiques des racines de tomates ont été séparés sur une colonne de phase inverse. Compte tenu de la polarité des métabolites extraits, l'eau et l'acétonitrile ont été choisis pour séparer les composés, 0,1% d'acide formique est ajouté dans chacune des voies de la phase mobile car plusieurs composés restent accrochés à la colonne chromatographique lorsque des phases mobiles non acidifiées ont été testées. Les conditions chromatographiques sélectionnées représentaient un compromis entre la résolution du pic et le temps d'analyse, ce qui a permis de séparer différentes classes de molécules. En outre, les pics chromatographiques des étalons internes ont été bien séparés des autres molécules des extraits afin d'éviter les coélutions et permettre une intégration répétable. Les détails des conditions d'analyse sont décrits en Annexe 1.

Les expériences MS ont été réalisées en mode (+) ESI et (-) ESI pour détecter le plus grand nombre de composés. Le mode (+) ESI a été choisi pour le traitement des données car il a permis de détecter certains composés ne répondant pas en mode négatif. De plus, les spectres MS de glycoalcaloïdes sont plus faciles à interpréter en mode (+) ESI, avec des ions et des adduits moléculaires. Les spectres MS obtenus en mode négatif ont aidé à l'identification des composés.

Différentes fréquences d'acquisition ont été testées de 1 à 5 Hz pour améliorer la résolution des pics et obtenir suffisamment de points de données qui définissent un pic afin de simplifier la déconvolution par les algorithmes de *peakpicking*. La fréquence d'acquisition de 2,5 Hz a été choisie car elle a offert une meilleure résolution et un rapport signal/bruit optimal. La gamme de masse a été de *m/z* 50 à 1550. La tension du capillaire a été fixée à 4,5 kV en mode positif et à 4,0 kV en mode négatif. Les débits de gaz de nébulisation et de séchage ( $N_2$ ) ont été fixés à 2 bar et 9 L/min respectivement et le gaz de séchage a été chauffé à 200°C.

Les expériences MS/MS ont été réalisées en utilisant le mode d'acquisition DDA. Plusieurs énergies de collision ont été testées pour les expériences de MS/MS afin de trouver celles qui sont optimales et qui permettent de fragmenter les structures rigides tout en ayant des spectres exploitables pour les molécules qui se fragmentent facilement. Une énergie de 12 eV a donc été choisie pour fragmenter la plupart des composés, de 50 eV pour fragmenter les glycoalcaloïdes (*m/z* 800-1550) et de 70 eV pour induire la fragmentation des alcaloïdes

stéroïdiens ( $m/z$  400-450). Le logiciel DataAnalysis 4.4 a été utilisé pour traiter les données, analyser et identifier les composés à l'aide des spectres MS et MS/MS.

Les chromatogrammes (BPC) obtenus par UHPLC-HRMS en mode positif de deux extraits de racines jeunes et matures sont présentés dans la Figure 36. Des composés de différentes polarités peuvent être observés sur le chromatogramme allant des plus polaires élués au début du gradient, moyennement polaires au milieu du chromatogramme et certains composés lipophiles vers la fin de l'analyse. Ces résultats montrent que la plupart des composés sont présents dans les deux types d'échantillons, les différences se manifestant principalement par des changements d'intensité de certains composés plus ou moins intenses en fonction des stades de développement du plant.

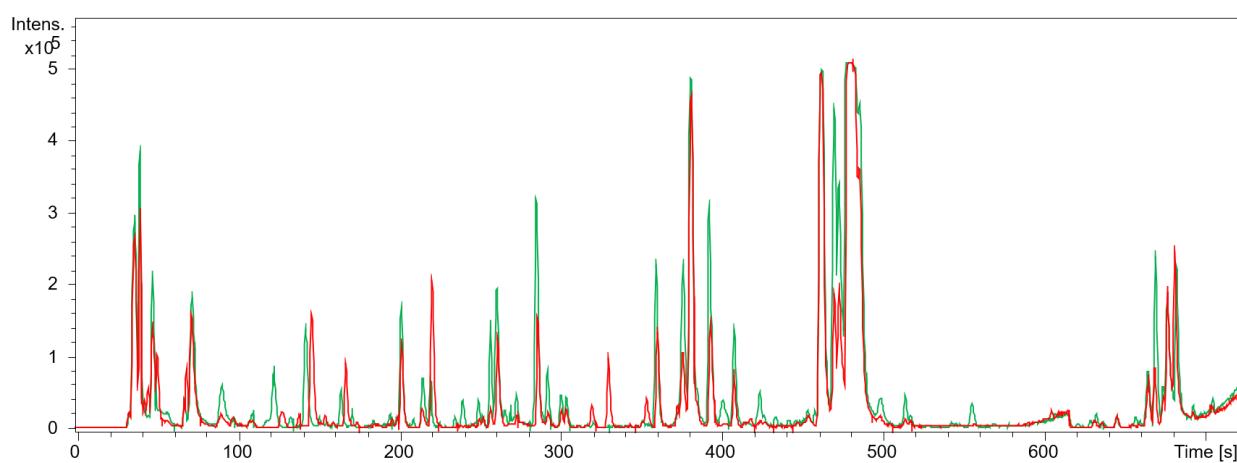


Figure 36 : Les BPCs (Base Peak Chromatograms) d'extraits de racines de tomate jeunes (vert) et matures (rouge) obtenus par UHPLC-HRMS en mode (+) ESI

### C. Prétraitement des données

Les données acquises en mode (+) ESI ont été converties au format .mzXML à l'aide du logiciel CompassXport 3.0.13 (Bruker). Le prétraitement des données a été effectué sur la plateforme Workflow4Metabolomics (W4M), disponible en ligne gratuitement <http://workflow4metabolomics.org> et développée depuis 2013 pour le prétraitement, l'analyse statistique et l'interprétation de données métabolomiques complètes<sup>181</sup>. W4M est une plateforme de recherche virtuelle entièrement construite sur Galaxy<sup>182</sup> pour les développeurs en bioinformatique et les utilisateurs de métabolomique, avec une mise en place des algorithmes dans les modules, en plus de fonctionnalités simples d'utilisation pour la gestion des workflows<sup>183</sup>. Ce processus de prétraitement est réalisé par l'algorithme XCMS<sup>184</sup> installé sur W4M et comprend plusieurs étapes telles que le prélèvement des pics, l'alignement et l'intégration afin d'obtenir une matrice prête pour les analyses statistiques ou le "post-traitement".

Dans cette étude, W4M a été utilisée afin d'obtenir une matrice à partir des données brutes de départ, avec les échantillons en lignes (deux groupes de 12 individus), les composés en colonnes (les variables) et leurs aires de pics dans les différents échantillons (la réponse des variables). Pour ce faire, un *workflow* a été créé et optimisé sur les données UHPLC-HRMS des racines de tomate et comporte 4 outils résumés dans le Tableau 4.

| Outils         | Paramètres                                    |
|----------------|---|
| xcms.xcmsSet   | Scanrange 50-708<br>S/N 20<br>centWave method |
| xcms.group     | Bandwidth 5                                   |
| xcms.retcor    | Obiwarp method                                |
| xcms.fillpeaks | Filling method chrom                          |

Tableau 4 : Les outils informatiques utilisés dans le workflow de W4M (workflow4metabolomics) pour le prétraitement des données UHPLC-HRMS des racines de tomate

Le premier outil est utilisé pour filtrer et identifier les pics en extrayant des chromatogrammes d'ions, en appliquant un modèle de filtre adapté aux pics, puis en identifiant et en intégrant les aires de pics. Le second outil est fait pour regrouper les caractéristiques qui représentent le même analyte après identification des pics. Le troisième est appliqué à la correction du temps de rétention, en identifiant les groupes à utiliser comme étalons, en calculant les écarts de temps de rétention et en alignant les pics. La dernière étape du prétraitement des données est appliquée pour intégrer les aires des pics de toutes les analyses et créer une matrice contenant les variables avec leurs aires dans chaque échantillon<sup>184</sup>.

Ce workflow a été appliqué à un fichier compressé avec 7-zip contenant les 24 analyses mzXML et introduit dans W4M à l'aide de Cyberduck. En plus des analyses, un fichier de métadonnées (.txt) a été inséré pour avoir des informations sur l'échantillon (nom de l'échantillon, classe, type d'échantillon, groupe, ordre d'injection).

Une fois générée, la matrice a été comparée aux analyses brutes pour confirmer qu'elle représente bien les données chromatographiques et spectrales. La redondance d'information a ensuite été réduite en éliminant les caractéristiques représentant le même composé et les caractéristiques présentes dans le blanc (contaminants) afin de réduire le taux d'information.

#### D. Analyses statistiques

Un total de 855 caractéristiques a été généré à partir des analyses UHPLC-HRMS par W4M, ce grand nombre de variables pourrait entraîner une instabilité du modèle statistique, ce qui

implique des résultats statistiques non pertinents. Pour résoudre ce problème, des tests statistiques ont été combinés pour obtenir un jeu de données plus petit à partir de la matrice initiale. Pour ce faire, la matrice de départ a été introduite dans le logiciel en ligne MetaboAnalyst 4.0 ; les variables ont été centrées et réduites.

L'analyse discriminante des moindres carrés partiels PLS-DA (*Partial Least Squares Discriminant Analyses*) a été appliquée pour obtenir les scores VIP (*Variable Important in Projection*) des variables, permettant de choisir les caractéristiques et d'éliminer toutes celles qui ne sont pas discriminantes entre les deux groupes de racines de tomate. Seules 72 variables (nombre de variables = 3 x nombre d'individus) ayant les scores VIP les plus élevés selon la composante 1 ont été choisies pour créer une sous-matrice contenant les variables les plus pertinentes. Le jeu de données de 72 variables et 24 individus a donc été extrait de la matrice de départ et utilisé pour l'analyse statistique afin d'effectuer l'étude comparative et déterminer les biomarqueurs discriminants les jeunes racines de tomates par rapport aux racines matures.

La matrice réduite a été introduite dans MetaboAnalyst 4.0, les données ont été centrées réduites pour normaliser le jeu de données. Les analyses univariée (*Volcano Plot*), multivariées non supervisée, notamment l'analyse en composantes principales PCA (*Principal Component Analysis*) et analyse de regroupement hiérarchique HAC (*Hierarchical Clustering Analysis*), ainsi que supervisées PLS-DA, ont été utilisées pour identifier les différences entre les deux groupes de racines.

Des approches d'analyse en composantes principales (PCA) et d'analyse de regroupement hiérarchique ascendant (HCA) non supervisées ont d'abord été utilisées pour visualiser les groupes, les tendances et les valeurs aberrantes parmi les observations. Sur la Figure 37 A, le résultat de PCA a montré une séparation nette entre les deux groupes, suggérant des différences de profil métabolique lorsque les plants de tomates sont collectés avec un décalage de 22 jours. Les deux premières composantes représentent 85% de la variation totale. Les groupes jeunes et matures sont bien différenciés selon la composante 1 (79%) sans observation de valeurs aberrantes.

Sur la Figure 37 B, le dendrogramme représente le regroupement hiérarchique des 24 échantillons de racines de tomate. L'axe horizontal représente la distance ou la dissimilitude entre les groupes et l'axe vertical les objets et les groupes. Cette présentation montre deux grappes, car deux branches se trouvent à peu près à la même distance horizontale : les groupes jeunes (vert) et matures (rouge). La dispersion intra-groupe des jeunes racines observée sur la PCA serait due à aux individus jeunes Y1 et Y2 qui ont une distance plus grande sur le dendrogramme par rapport aux autres individus.

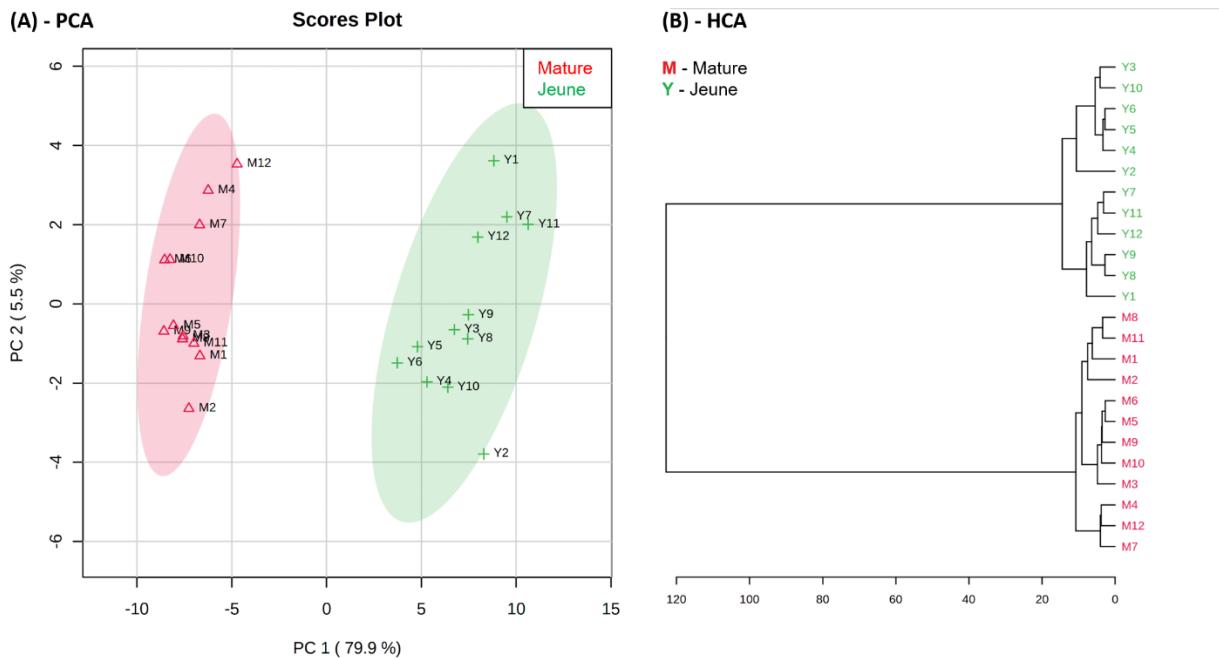
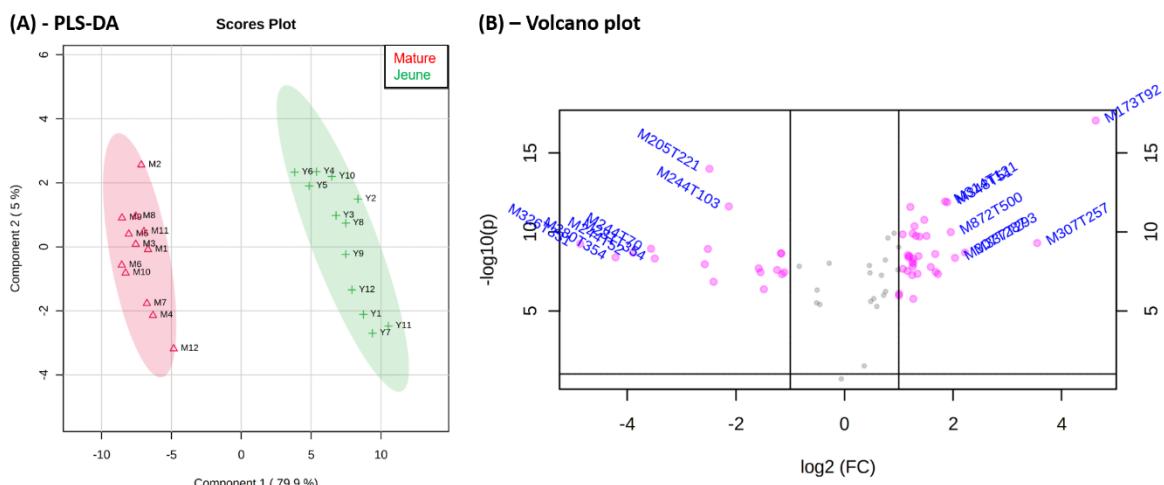


Figure 37: Analyses multivariées non supervisées A) PCA (Principal Component Analysis) des extraits de racines de tomate jeunes (vert) et matures (rouge). B) Dendrogramme HCA (Hierarchical Clustering Ascendant). Les paramètres utilisés sont Euclidean distance measure et l'algorithme ward clustering

Par la suite, une analyse multivariée supervisée a été appliquée pour déterminer les variables discriminantes, correspondant aux biomarqueurs impliqués dans le développement des racines de tomate. L'analyse PLS-DA présentée sur la Figure 38 A décrit 80% de la variance en X ( $R^2X$ ) et 5% en Y ( $R^2Y$ ) avec une capacité prédictive  $Q_2$  de 0,95. Ces paramètres indiquent que le modèle statistique est prédictif avec une bonne finesse.



Ensuite, la distribution des composés a été examinée à l'aide de l'analyse univariée *Volcano Plot* présentée sur la Figure 38 B. Le test *Volcano Plot* montre les variables non discriminantes au centre, et les variables discriminantes en dehors du rectangle central, à droite celles qui sont plus abondantes dans le groupe des jeunes racines, et à gauche celles qui sont abondantes dans le groupe des racines matures. Ce test donne une vision globale de la distribution des composés. Il permet de voir les composés qui ne varient pas entre les groupes, et ceux qui sont plus ou moins intenses dans l'un des groupes. Afin d'évaluer la variation des variables indépendamment les unes des autres, des *Box Plots* des VIP ont été examinés.

Enfin, la corrélation entre les variables est analysée pour éliminer celles qui représentent le même composé et pour ne garder que les biomarqueurs. Les métabolites les plus discriminants (c'est-à-dire ceux qui ont le scores VIP les plus élevés) sont présentés à la Figure 39 A. Ces composés ont été sélectionnés car ils ont des scores  $VIP > 1$ . Les diagrammes en boîtes (*Box Plots*) des biomarqueurs montrent qu'une majorité des composés sont régulés à la hausse au début du développement des racines et leurs quantités sont fortement réduites à maturité (Figure 39 B).

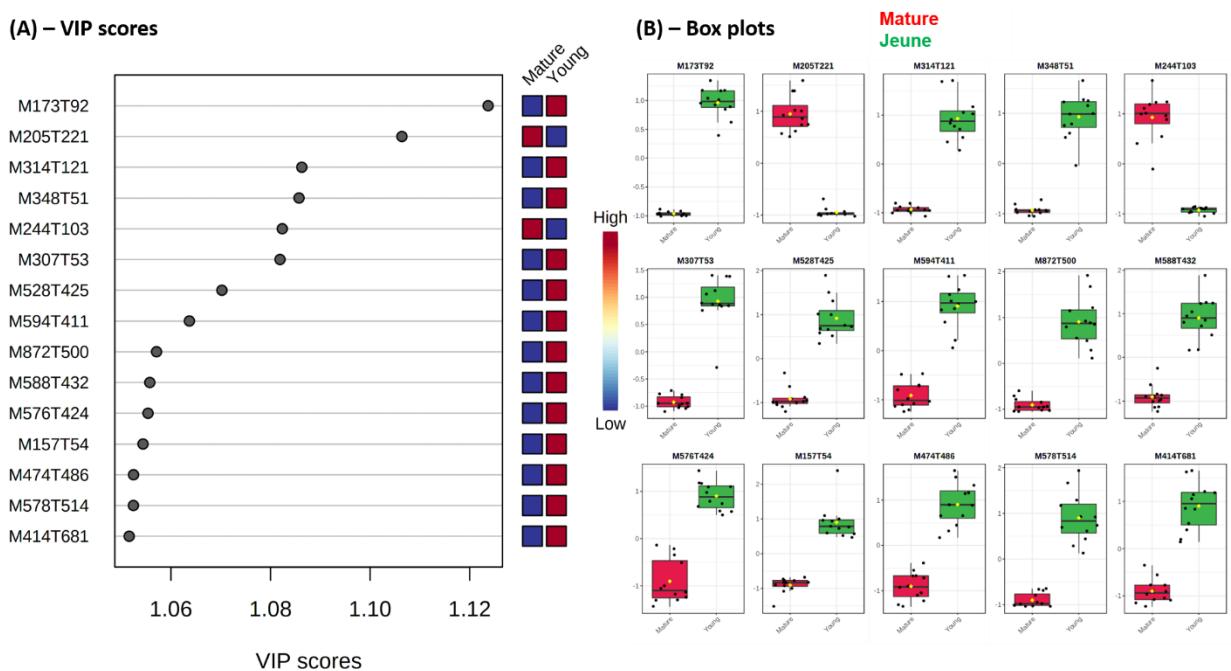


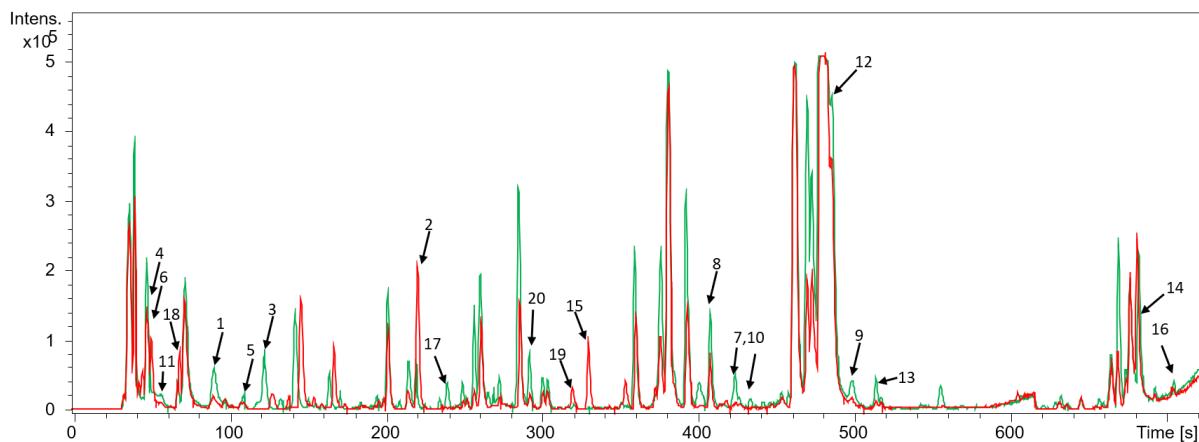
Figure 39 : A) VIP (Variable Important in Projection) scores. B) Box Plots des 15 variables les plus discriminantes entre les groupes des racines de tomate jeunes (vert) et matures (rouge)

En conclusion, les groupes de racines de tomates ont été comparés et analysés et les variables discriminant les deux groupes jeunes et matures ont été mises en évidence par l'ensemble des tests statistiques univariés et multivariés non supervisés comme supervisés.

Les données chromatographiques et spectrales ont ensuite été étudiées dans le but d'identifier ces biomarqueurs.

### E. Caractérisation des composés

La distribution chromatographique (Figure 40) révèle que les biomarqueurs ( $VIP > 1$ ) sont répartis sur l'ensemble du chromatogramme, ce qui suggère des différences en termes de structure et de polarité.



*Figure 40 : Numérotation des pics chromatographiques correspondants aux biomarqueurs, discriminant les racines jeunes et matures, mis en évidence par les analyses statistiques*

L'identification des composés discriminants le métabolome des racines de tomates jeunes et matures a été réalisée en utilisant d'une part, les informations de déréplication obtenues par le réseau moléculaire qui permet l'identification rapide des substances connues présentes dans les bases de données spectrales GNPS ; et d'autre part en utilisant l'élucidation structurale *de novo* appuyée par le logiciel DataAnalysis pour générer les formules chimiques possibles des ions de spectres MS, avec une faible incertitude entre le *m/z* théorique et expérimental, et en comparant le modèle isotopique expérimental au modèle théorique.

#### 1) Réseau moléculaire des extraits bruts

Le réseau moléculaire a été généré sur la plateforme GNPS. Tout d'abord, l'algorithme AutoMS(n) du logiciel DataAnalysis 4.4 a été appliqué pour générer les *peaklists* des composés de racines de tomates jeunes et matures. La fenêtre de temps de rétention est de 0,3 min, le seuil d'intensité TIC (AllMSn) a été fixé à 50 pour le mode positif (+) ESI et le nombre maximum de composés à 7500, les 10 ions non déconvoltés les plus abondants avec un seuil d'intensité de 100 ont été extraits. Les listes ont ensuite été exportées au format .mgf.

Les données ont ensuite été traitées sur GNPS avec une tolérance de 0,02 Da pour la masse des ions parents fragments en désactivant l'outil MSCLustering. Des arêtes ont été formées

pour les similarités spectrales avec des scores cosine  $\geq 0,65$  avec un nombre de correspondances (ions communs ou pertes de neutres)  $\geq 4$ , en outre, les liaisons entre deux nœuds ont été conservées dans le réseau si et seulement si chacun des nœuds apparaissait dans le top 10 des nœuds les plus similaires.

De plus, pour l'identification des composés, les spectres présents dans le réseau ont été comparés aux spectres enregistrés dans les bases de données spectrales de GNPS de la même manière que les données d'entrée. Les paramètres utilisés pour la recherche dans les bibliothèques sont les suivants : seuil de score cosine de 0,65, minimum de correspondances exigé à 6. Enfin, le réseau moléculaire a été visualisé à l'aide du logiciel Cytoscape 3.5.1.

Le réseau moléculaire obtenu à partir des analyses des racines de tomates est présenté sur la Figure 41. À droite, le réseau moléculaire global contenant des nœuds liés en clusters avec au moins quatre ions communs et/ou des pertes neutres, et des nœuds solitaires qui ne sont pas liés aux clusters parce qu'ils ont des structures chimiques différentes ou ont moins de quatre ions communs avec d'autres nœuds. À gauche, un zoom sur quatre clusters identifiés représentant les principales familles moléculaires contenues dans les racines de tomate. Pour identifier ces familles moléculaires le réseau a été annoté par comparaison des spectres MS/MS expérimentaux avec ceux de la base de données GNPS. Ainsi 5 composés ont été identifiés (encadré en bleu dans la Figure 41) et ont permis d'attribuer les clusters aux familles décrites dans la littérature. Les dérivés d'alcaloïdes ont été séparés en trois groupes : alcaloïdes stéroïdiens, alcaloïdes de type tomatidine (génine A) et alcaloïdes de type solanoside (génine B). Les acides aminés ont été rassemblés en un cluster plus petit. Chacune de ces familles moléculaires a été détectée dans les deux types d'échantillons. Des groupes d'esters d'hydroxycinnamoyle, de flavonoïdes et de phénolamides n'ont pas été observés sous forme de clusters malgré leur présence dans les racines de tomates<sup>185</sup>. Cela est probablement dû au faible nombre d'ions fragments communs ( $< 4$ ) dans leurs spectres MS/MS. En effet, on les trouve majoritairement sous forme de nœuds singulaires dans le réseau moléculaire.

Parmi les 20 VIP les plus élevés, soit les composés discriminants les groupes de racines, 5 métabolites ont été pu être identifiés grâce aux bases de données GNPS, dont le tryptophane (2), la déhydrotomatine (7),  $\beta$ 2-tomatine (9), l'acétoxytomatine (12) et le tomatidénol (14). Les 15 autres VIP ont été annotés par interprétation de leurs spectres MS/MS et de leurs positions dans le réseau moléculaire lorsqu'ils sont liés en cluster puisqu'ils présentent des similitudes spectrales avec les nœuds voisins. Toutefois, lorsque les composés sont présentés par des nœuds singulaires, on peut supposer qu'ils font partie de la classe des composés phénoliques présents dans les racines de tomates mais qui ne sont pas liés en cluster.

Le réseau moléculaire a donc fourni une vue globale de la composition chimique des racines de tomate, en mettant en évidence les principales familles moléculaires. Cependant, de nombreux métabolites sont restés non identifiés.

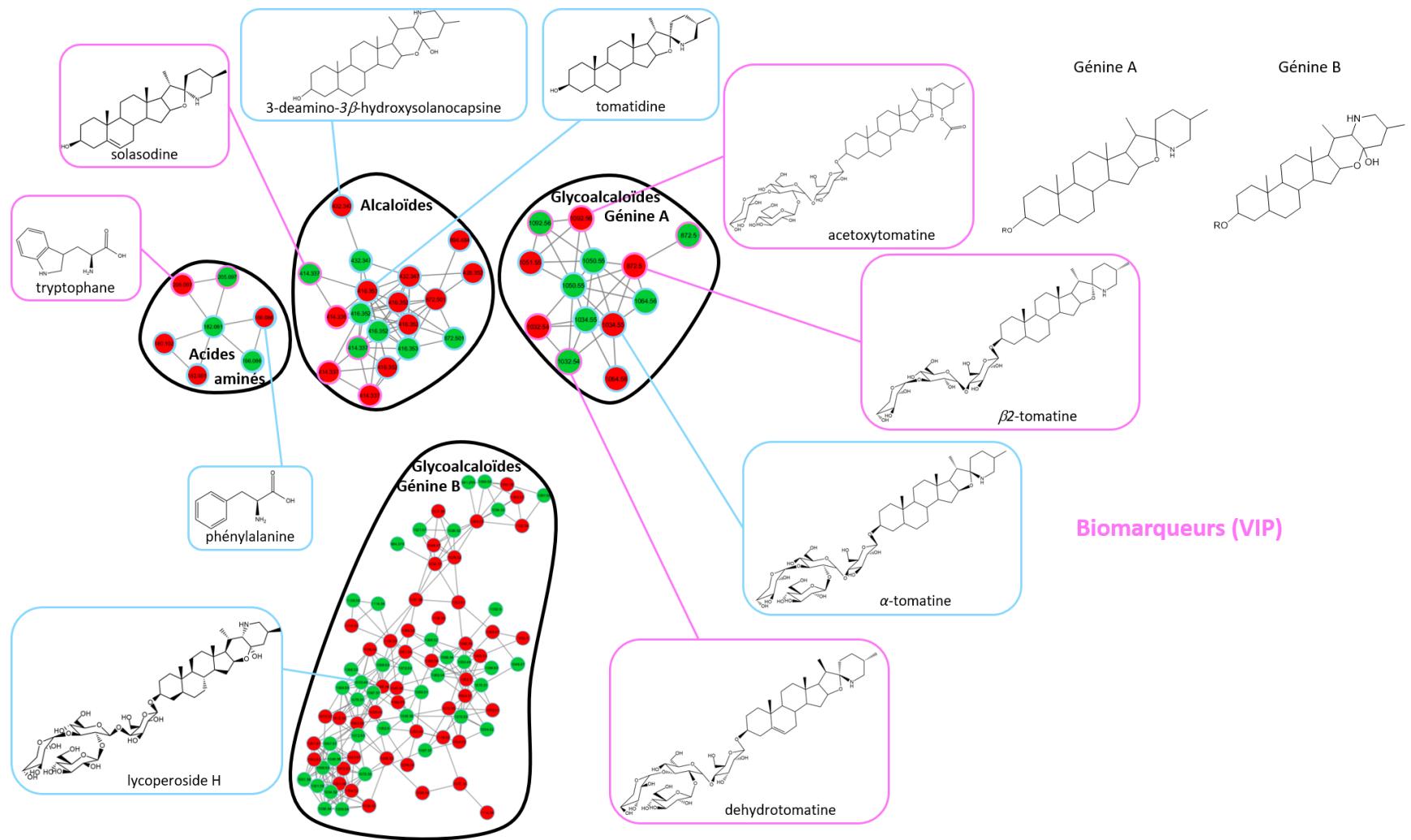


Figure 41 : Réseau moléculaire des groupes jeunes (vert) et matures (rouge) racines de tomate, obtenu en utilisant la plateforme GNPS et visualisé sur le logiciel Cytoscape 3.5.1. Les molécules entourées en bleu sont les composés annotés grâce aux bases de données de GNPS, et ceux entourés en mauve sont les biomarqueurs mis en évidence par les analyses statistiques (scores VIP > 1)

## 2) Caractérisation basée sur l'élucidation de novo

Pour identifier les biomarqueurs de racines de tomate, la tolérance de masse a été fixée à 4 ppm pour la MS et à 7 ppm pour la MS/MS ou les petites valeurs de *m/z*. Les éléments C, H, O, N, P et S ont été sélectionnés pour le calcul de la composition élémentaire et seules les formules chimiques cohérentes ont été prises en compte. En outre, les caractéristiques chromatographiques et les données HRMS/MS ainsi que les informations bibliographiques ont été utilisées pour l'identification. L'ion adduit  $[M+Na]^+$ , spécialement observé dans le cas des glycoalcaloïdes, a confirmé la formule proposée pour l'ion  $[M+H]^+$ . Les formules moléculaires possibles pour chaque pic sélectionné ont été recherchées dans les bases de données publiques, à savoir KNapSAcK, Dictionary of Natural Products, HMBD, Super natural II et PubChem afin de caractériser les produits naturels connus. De plus, certains composés ont été identifiés sans ambiguïté en comparant leur temps de rétention et leurs données spectrales avec ceux des étalons de référence. Enfin, la connaissance de la famille moléculaire des biomarqueurs par le réseau moléculaire permet de choisir et de confirmer des formules chimiques lorsque plusieurs sont cohérentes.

Le Tableau 5 regroupe les 20 composés discriminant les groupes de racines avec des VIP scores > 1 classés par ordre décroissant de discrimination. Le nom des variables correspond au *m/z* et au temps de rétention (en seconde) des composés. Par exemple, la variable M173T92 est la plus discriminante avec un VIP score de 1,1237, sa caractéristique est  $[M+H]^+$  à *m/z* 173,1399 éluée à 92 secondes (1,49 min).

L'erreur en ppm est la différence entre le *m/z* expérimental de l'ion moléculaire  $[M+H]^+$  ou  $[M+H+Na]^{2+}$  et le *m/z* théorique de l'ion moléculaire de la formule brute sectionnée.

En outre, certains glycoalcaloïdes ont été fragmentés en source mais leurs ions fragments n'apparaissent pas dans les spectres MS/MS car l'énergie de collision est assez forte. Toutefois, ces ions fragments générés en source ont été utilisés pour l'identification et sont des fragments typiques de la famille des glycoalcaloïdes. Par conséquent, ils ont été présentés dans le tableau.

Enfin, les propositions de structures dans le Tableau 5 sont présentées pour 9 métabolites avec un niveau 2 et pour 2 métabolites avec un niveau 3 d'identification. Alors que les composés qui n'ont pas été annotés ont un niveau 4 d'identification.

| N° | Variables VIP > 1 | Score VIP | t <sub>R</sub> (min) | m/z mesuré [M-H] <sup>-</sup> | Erreur (ppm) | m/z mesuré [M+H] <sup>+</sup> , [M+H+Na] <sup>2+</sup> | Erreur (ppm) | Formule brute [M]   | MS/MS (intensité)   | Ions issus de la fragmentation en source | Propositions de structure                 |
|----|-------------------|-----------|----------------------|-------------------------------|--------------|--|--------------|---|---|--|---|
| 1  | M173T92           | 1,1237    | 1,49                 | -                             | -            | 173,1399   | 6,8          | C <sub>7</sub> H <sub>16</sub> N <sub>4</sub> O                 | 173,1396 (4761),<br>156,1133<br>(156,131),<br>131,1170 (359),<br>114,0914 (2281),               | -  | Acétylagmatine                            |
| 2  | M205T221          | 1,1064    | 3,65                 | 203,0825                      | 0,4          | 205,0970   | 0,7          | C <sub>11</sub> H <sub>12</sub> N <sub>2</sub> O <sub>2</sub>   | 205,0972 (862),<br>188,0704 (21236),<br>159,0919 (1420),<br>146,0599 (5909),<br>118,0641 (1108) | -  | L-Tryptophane                             |
| 3  | M314T121          | 1,0863    | 2,07                 | 312,0765                      | 4,3          | 314,0915   | 5,3          | C <sub>11</sub> H <sub>15</sub> N <sub>5</sub> O <sub>4</sub> S | 314,0918 (2220),<br>296,0812 (337),<br>179,0380 (166),<br>136,0614 (1728),<br>97,0286 (340)     | -  | (S)-5'-Deoxy-5'-(methylsulfinyl)adénosine |
| 4  | M348T53           | 1,0857    | 0,80                 | 346,0553                      | 4,5          | 348,0704   | 3,0          | C <sub>10</sub> H <sub>14</sub> N <sub>5</sub> O <sub>7</sub> P | 348,0702 (6391),<br>215,0525 (330),<br>157,0106 (253),<br>136,0621(4886),                       | -  | Adénosine monophosphate                   |
| 5  | M244T103          | 1,0823    | 1,67                 | 242,1366                      | 13,1         | 244,1549   | -2,5         | C <sub>12</sub> H <sub>21</sub> NO <sub>4</sub>                 | 244,1555 (626),<br>186,1128 (652)   | -  | -   |
| 6  | M307T53           | 1,0819    | 0,76                 | -                             | -            | 307,0421   | -4,1         | C <sub>7</sub> H <sub>16</sub> O <sub>11</sub> P                | 307,1027 (203),   | -  | -   |

|    |          |        |      |         |     |                         |      |  |  |  |                            |
|----|----------|--------|------|---------|-----|-------------------------|------|--|--|--|----------------------------|
|    |          |        |      |         |     |                         |      |  | 292,1017 (688),<br>273,1062 (216),<br>211,0600 (112),<br>145,0495 (338),<br>127,0388(499),   |  |                            |
| 7  | M528T425 | 1,0702 | 7,15 | -       | -   | 1032,5388<br>(527,7656) | 2,8  | C <sub>50</sub> H <sub>81</sub> NO <sub>21</sub> | -  | 918,5020,<br>738,4464,<br>576,3917,<br>457,1541,<br>432,3488,<br>414,3356, | Dehydrotomatine<br>isomère |
| 8  | M594T411 | 1,0637 | 6,41 | -       | -   | 1050,5482<br>(536,7661) | -0,3 | C <sub>50</sub> H <sub>83</sub> NO <sub>22</sub> | 1050,5481 (39),<br>1032,5357 (845),<br>414,3358 (116),<br>255,2108 (276),<br>145,0496 (259),   | 888,4968,<br>726,4439,<br>594,3973,<br>457,1532,<br>432,3488,<br>414,3347, | Lycoperoside H             |
| 9  | M872T500 | 1,0571 | 8,40 | 870,485 | 0,8 | 872,4995<br>(447,7431)  | 0,8  | C <sub>44</sub> H <sub>73</sub> NO <sub>16</sub> | 872,5009 (6494),<br>854,4893 (7050),<br>398,3417 (411),<br>255,2110 (3300),<br>161,1327 (2600),<br>126,1265 (1663),<br>97,0276 (1032), | 740,4595,<br>578,4056,<br>416,3510,  | β2-Tomatine                |
| 10 | M588T432 | 1,0557 | 7,15 | -       | -   | 1044,5340               | 3,2  | C <sub>51</sub> H <sub>81</sub> NO <sub>21</sub> | 1044,5405 (82),  | 588,3891,  | Derivé de                  |

|    |          |        |       |          |      |                         |      |  |  |                        |   |
|----|----------|--------|-------|----------|------|-------------------------|------|--|--|------------------------|---|
|    |          |        |       |          |      | (533,7657)              |      |  | 1026,5250 (217),<br>136,1116 (238),  | 457,1537,<br>426,3386, | glycoalcaloïde                            |
| 11 | M157T54  | 1,0554 | 0,91  | 291,0193 | -    | -                       | 2,5  | C <sub>6</sub> H <sub>10</sub> O <sub>3</sub>    | -  | -                      | -   |
| 12 | M474T486 | 1,0525 | 7,89  | 1090,545 | -1,4 | 1092,5577<br>(557,7708) | 0,7  | C <sub>52</sub> H <sub>85</sub> NO <sub>23</sub> | 1092,5573 (570),<br>1074,5435 (1327),<br>1032,5373 (6444),<br>1014,5249 (656),<br>576,3859 (457),<br>414,3362 (853),<br>273,2226 (2336),<br>255,2104 (4306),<br>145,0494 (3425),<br>97,0276 (1191),<br>85,0283 (488) | 634,3965,<br>472,3419, | Acétoxytomatine                           |
| 13 | M578T514 | 1,0524 | 8,40  | -        | -    | 1004,5417<br>(513,7666) | 0,8  | C <sub>49</sub> H <sub>81</sub> NO <sub>20</sub> | 1004,5418 (826),<br>986,5362 (242),<br>416,3438 (47),<br>255,2103 (191),<br>126,1286 (91),<br>97,0292 (73),  | 578,4048,<br>416,3508, | Glycoalcaloïde à<br>base de<br>tomatidine |
| 14 | M414T681 | 1,0516 | 11,41 | -        | -    | 414,3368                | -0,4 | C <sub>27</sub> H <sub>43</sub> NO <sub>2</sub>  | 414,3368 (117),<br>396,3263 (184),<br>271,2052 (175),<br>213,1644 (106),<br>161,1321 (2714),   | -                      | Solasodine                                |

|    |          |        |       |          |      |          |      |   |   |   |   |
|----|----------|--------|-------|----------|------|----------|------|---|---|---|---|
|    |          |        |       |          |      |          |      |   | 124,1109 (273),<br>105,0696 (500)   |   |   |
| 15 | M326T331 | 1,042  | 5,48  | 324,1813 | 0,9  | 326,1961 | 0,2  | C <sub>17</sub> H <sub>27</sub> NO <sub>5</sub> | 326,1962 (33490),<br>226,1440 (1483)                                      | - | - |
| 16 | M318T714 | 1,0499 | 11,93 | -        | -    | 318,3003 | -0,2 | C <sub>18</sub> H <sub>39</sub> NO <sub>3</sub> | 318,3004 (22529),<br>300,2895 (3524),<br>282,2792 (2148)                  | - | - |
| 17 | M307T257 | 1,0432 | 4,24  | 305,1615 | 1,5  | 307,1766 | -4,8 | C <sub>14</sub> H <sub>26</sub> O <sub>7</sub>  | 307,1765 (28481),<br>290,1474 (1199),<br>177,0541 (1024)                  | - | - |
| 18 | M244T70  | 1,035  | 1,16  | 242,1399 | -0,6 | 244,1546 | -1,1 | C <sub>12</sub> H <sub>21</sub> NO <sub>4</sub> | -   | - | - |
| 19 | M110T320 | 1,0347 | 5,3   | -        | -    | 210,1489 | -0,4 | C <sub>12</sub> H <sub>19</sub> NO <sub>2</sub> | 210,1491 (7525),<br>110,0964 (11041)                                      | - | - |
| 20 | M337T293 | 1,0294 | 4,88  | 335,1722 | 0,9  | 337,1872 | -4,3 | C <sub>15</sub> H <sub>28</sub> O <sub>8</sub>  | 337,1869 (16791),<br>320,1586 (368),<br>299,2961 (450),<br>207,0653 (744) | - | - |

Tableau 5 : Annotation des métabolites impliqués lors du développement des racines de tomates (scores VIP > 1 et p-value < 0,001)

L'identification des 4 premières variables a été faite par l'interprétation de leurs schémas de fragmentation en MS/MS. En effet, le VIP principal (1) a été détecté à  $m/z$  173,1401. Son schéma de fragmentation présente trois principaux ions produits à  $m/z$  156,1131, 131,1187 et 114,0918 correspondant aux pertes des groupes acétyle (-42 u) et amine (-17 u) (Figure 42). La formule ionique donne  $C_7H_{17}N_4O^+$  (erreur = 2,4 ppm) qui pourrait correspondre à l'acétylagmatine, une guanidine dérivée du métabolisme primaire.

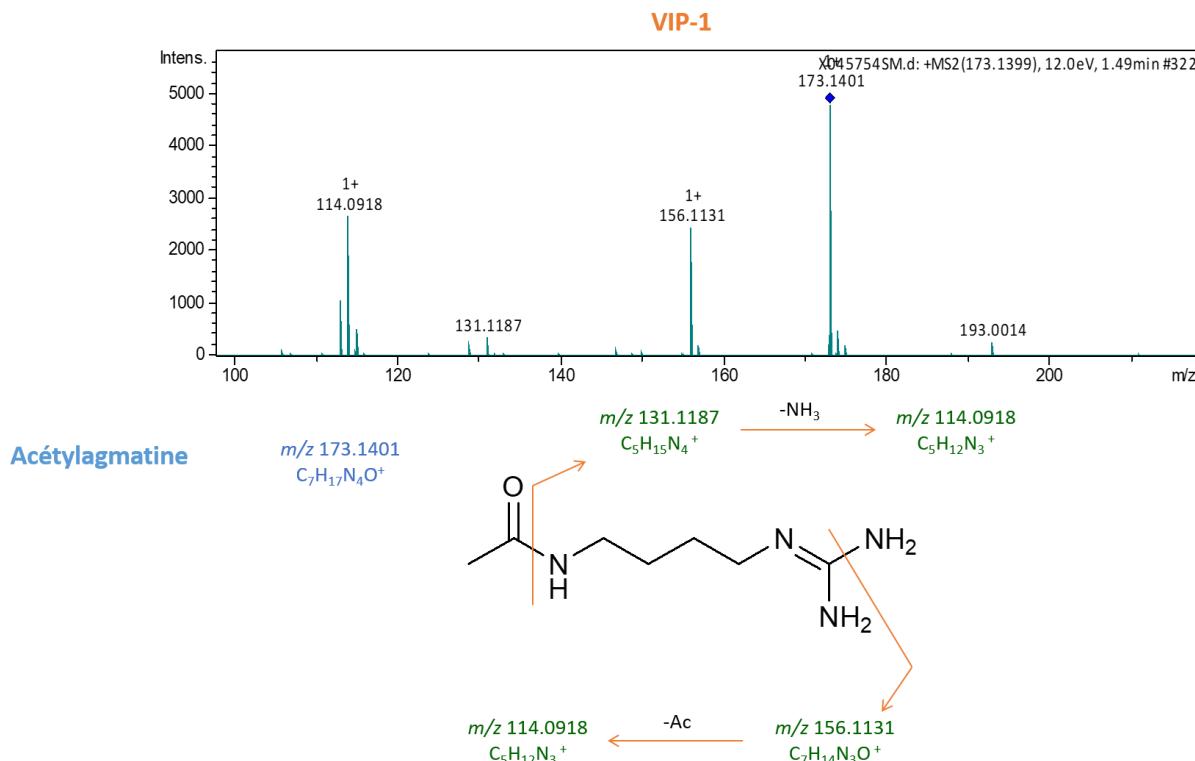


Figure 42 : Spectre MS/MS obtenu en mode (+) ESI et interprétation du schéma de fragmentation possible pour le VIP 1, l'acétylagmatine

Le composé ayant le deuxième score VIP le plus élevé (2) a été identifié comme étant le tryptophane :  $C_{11}H_{12}N_2O_2^+$  (0,2 ppm). Le schéma de fragmentation de cet acide aminé aromatique est en accord avec la littérature<sup>186</sup>. Son ion moléculaire à  $m/z$  205,0972 est principalement fragmenté pour former l'ion fragment à  $m/z$  188,0704 par perte du  $\text{NH}_3$ . D'autres pertes de  $\text{CH}_5\text{NO}_2$  et  $\text{CH}_2\text{O}_2$  donnent lieu à des fragments mineurs à  $m/z$  159,0919 et 146,0599, respectivement (Figure 43).

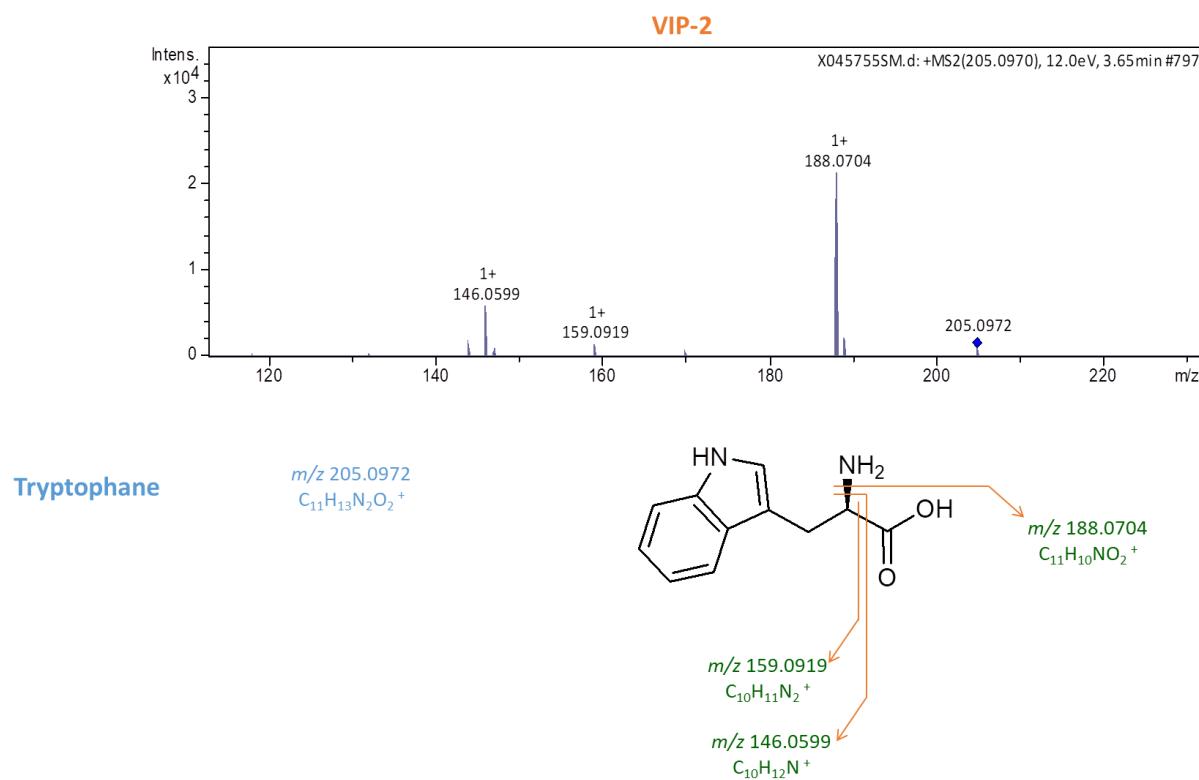


Figure 43 : Spectre MS/MS obtenu en mode (+) ESI et interprétation du schéma de fragmentation du VIP2, le tryptophane

Les composés 3 et 4 ont tous les deux été identifiés comme des dérivés de l'adénosine. L'ion  $[M+H]^+$  du composé 3 à  $m/z$  314,0918 :  $C_{11}H_{16}N_5O_4S^+$  (0,2 ppm) est fragmenté en deux principaux ions fils respectivement observés à  $m/z$  179,0380 et 136,0614, correspondant au clivage entre les groupements adénosine et désoxyméthylsufynil. Le spectre de fragmentation a révélé deux autres ions à  $m/z$  296,0812 (perte d'eau) et  $m/z$  97,0286 ( $C_5H_4O_2^+$ ). Le composé 3 a donc été désigné comme (S)-5'-Désoxy-5'-(méthylsulfynil)adénosine (Figure 44).

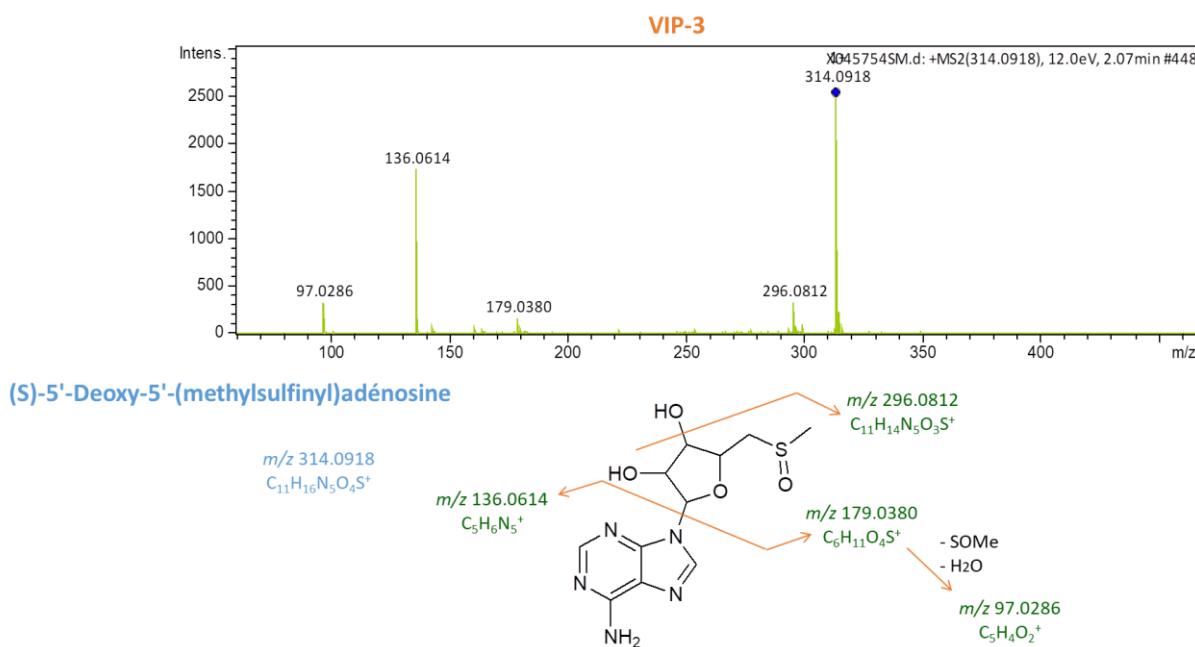


Figure 44 : Spectre MS/MS obtenu en mode (+) ESI et interprétation du schéma de fragmentation du VIP 3, (S)-5'-Désoxy-5'-(méthylsulfinyl)adénosine

Le même type de clivage a également été observé pour le composé 4. L'ion protoné  $m/z$  348,0702  $C_{10}H_{15}N_5O_7P^+$  (0,5 ppm) se fragmente directement en ion fragment majeur à  $m/z$  136,0621, ce qui suggère la présence d'adénosine monophosphate (Figure 45).

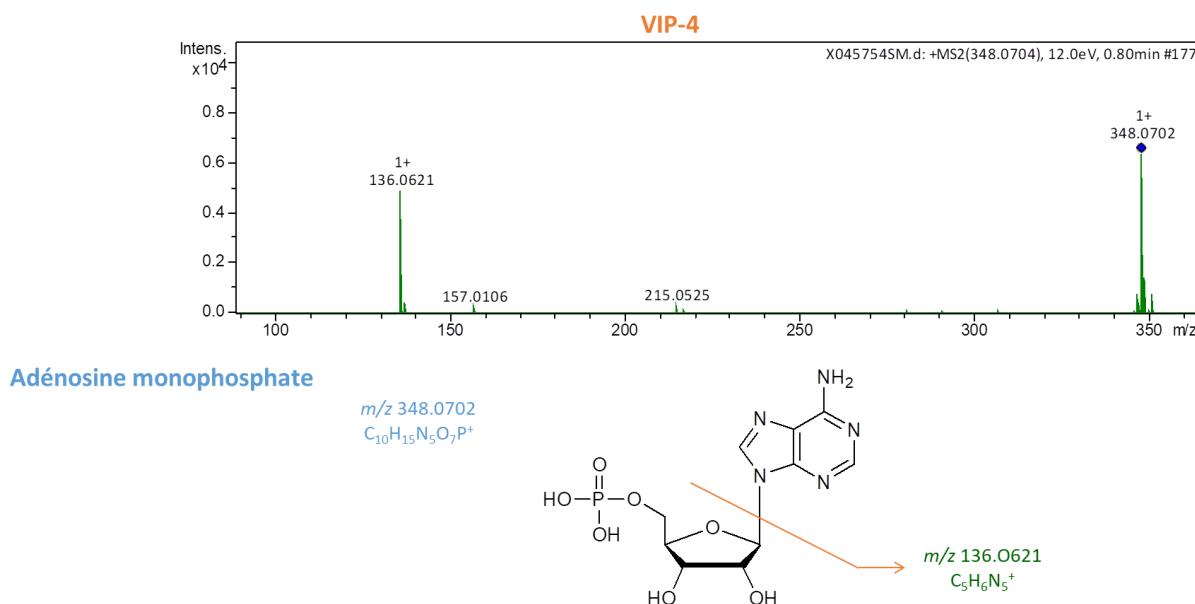


Figure 45 : Spectre MS/MS obtenu en mode (+) ESI et interprétation du schéma de fragmentation du VIP 4, l'adénosine monophosphate

Les glycoalcaloïdes stéroïdiens ont été facilement identifiés sur la base de leur comportement typique en mode positif<sup>187</sup>. Leur spectre MS est caractérisé par la présence de l'ion moléculaire  $[M+H]^+$  et de l'ion doublement chargé  $[M+H+Na]^{2+}$ . Les pertes successives des unités sucre

donnent des informations supplémentaires sur la structure des glycoalcaloïdes. Par exemple, le composé 7 ( $C_{50}H_{81}NO_{21}$ ) a été attribué à un isomère de la déhydrotomatine. L'ion moléculaire  $[M+H]^+$  a été observé à  $m/z$  1032,5374 avec son ion doublement chargé à  $m/z$  527,7635. Des ions fragments correspondant aux pertes d'unité xyl-glu (associée à l'eau) à  $m/z$  756,4525, d'unité xyl-glu-glu à  $m/z$  576,3883 et de groupement lycotétraose à  $m/z$  414,3362 ont également été observés. Sa forme aglycone tomatidénol  $[M+H]^+$  (composé 14) a également été détectée à  $m/z$  414,3368. La fragmentation du tomatidénol entraîne une perte de molécule d'eau neutre ( $m/z$  396,3263), ainsi qu'un ion à  $m/z$  271,2052 (perte de 143 unités de masse) obtenu par clivage typique des alcaloïdes stéroïdiens de type tomatidine<sup>187</sup>.

D'autres glycoalcaloïdes (composés 8, 9, 10, 12 et 13) ont été identifiés sur la base de leurs schémas de fragmentation typiques. Il est fortement probable que le composé 8 ( $m/z$  1050,5482) corresponde au lycopéroside H ( $C_{50}H_{83}NO_{22}$ ). La fragmentation en source de ce métabolite montre des pertes d'unités de xyl ( $m/z$  918,5088), de glu-glu ( $m/z$  726,4439), de glu-glu-xyl ( $m/z$  594,3973) et de lycotétraose ( $m/z$  414,3347). Les spectres MS/MS indiquent la perte d'eau ( $m/z$  1032,5357) et la fragmentation typique du groupement aglycone. L'adduit doublement chargé  $[M+H+Na]^{2+}$  a été observé à  $m/z$  536,7661. Le composé 9 à  $m/z$  872,5001 suit le même schéma de fragmentation. Il a pu être identifié comme étant la  $\beta$ 2-tomatine ( $C_{44}H_{73}NO_{16}$ ). Les ions fragments à  $m/z$  740,4586, 578,4054 et 416,4145 ont été obtenus par perte d'unité de xyl (-132 u), de xyl-glu (-294 u) et de xyl-glu-glu (-456 u), respectivement. Le  $[M+H+Na]^{2+}$  dichargé a été observé à  $m/z$  447,7449. De la même manière, le composé 12 a été identifié comme étant l'acétoxytomatine ( $C_{52}H_{85}NO_{23}$ ), tandis que les composés 10 ( $C_{51}H_{81}NO_{21}$ ) et 13 ( $C_{49}H_{81}NO_{20}$ ) sont probablement des glycoalcaloïdes à base de tomatidine. Enfin, les métabolites identifiés sont globalement des dérivés d'alcaloïdes et d'acides aminés.

#### **IV. CONCLUSION**

Il est connu que les plantes adaptent leur métabolisme primaire et secondaire dans les organes de surface pour résister et tolérer les facteurs environnementaux, tandis que les réponses métaboliques spécifiques sur les racines ne sont pas bien définies. Cela est surprenant étant donné que les racines sont considérées comme un système d'organes vitaux des plantes en raison de leur implication dans l'acquisition, l'ancrage, la propagation, le stockage, la protection contre la prédatation, la sécheresse ou les agents pathogènes de l'eau et des nutriments. Elles constituent également un site d'interaction majeur avec les mycorhizes ou les organismes fixateurs d'azote<sup>188</sup>. De plus, chez plusieurs espèces de solanacées, les alcaloïdes sont synthétisés dans les racines et sont transportés dans les parties aériennes des plantes. Par conséquent, le niveau d'alcaloïdes trouvés dans les feuilles est déterminé par le métabolisme des racines<sup>189</sup>. L'élucidation de la nature exacte des changements biochimiques au cours des premiers stades du développement des racines est d'une importance agronomique. C'est pourquoi une approche métabolomique UHPLC-HRMS non ciblée et efficace a été développée pour étudier les variations chimiques des racines de tomates durant les premiers stades de leur développement.

L'approche métabolomique développée comprend l'optimisation des méthodes d'extraction, de séparation et d'analyse. Une analyse chimiométrique utilisant des analyses statistiques supervisées et non supervisées appropriées a ensuite été établie afin mettre en évidence les variations chimiques et déterminer les composés discriminants les groupes de racines de tomate. Enfin la stratégie de réseau moléculaire combinée à l'élucidation structurale *de-novo* a conduit à la caractérisation de 20 composés d'intérêt, métabolites primaires et secondaires, qui sont régulés entre le 34<sup>ème</sup> et le 62<sup>ème</sup> jour après le semis. Parmi ces composés discriminants, 9 molécules ont été identifiées à un niveau 2 d'identification, 2 autres molécules à un niveau 3 d'identification et 9 composés n'ont pas été identifiés (niveau 4).

Les glycoalcaloïdes ont été identifiés comme l'une des principales classes de métabolites secondaires dans les racines de tomate et la principale famille qui évolue au cours du développement racinaire. Des réarrangements métaboliques spécifiques à des organes dans les plantules au cours du développement de la tomate (entre 3 et 9 jours après le semis) ont également révélé que les alcaloïdes stéroïdiens, y compris les dérivés de la tomatine, étaient les principaux composés observés dans les racines de tomate<sup>190</sup>.

En outre, la plupart des métabolites ont été régulés à la hausse dans les jeunes racines de tomates (34 jours), tandis que seuls les composés 2 (tryptophane) et 5 (inconnu) ont été régulés à la hausse dans les racines plus âgées (62 jours). Il est bien connu que les métabolites primaires et secondaires sont les produits intermédiaires ou ultimes de réseaux

complexes de voies biochimiques impliquées dans le métabolisme des plantes<sup>191</sup>. On peut donc supposer que les deux types de métabolites sont régulés à la hausse dès les premiers jours pour assurer la croissance et la défense précoce des jeunes racines de tomates, qui sont des tissus précieux à protéger<sup>192</sup>. De plus, il n'est pas surprenant de trouver du tryptophane en plus forte concentration dans les racines de 62 jours car c'est un précurseur de l'acide indole-3-acétique, la principale auxine naturelle. Cette phytohormone est essentielle pour presque tous les aspects de la croissance et du développement des plantes, en particulier la croissance des semis et des racines<sup>193</sup>. L'auxine est notamment responsable de divers stades de formation des racines, tels que la racine primaire, la racine latérale et le développement des poils de la racine<sup>194,188,195</sup> et tous les types de racines se développent continuellement pendant toute la vie du plant de tomate.

En conclusion, les changements métaboliques et les métabolites identifiés dans ce travail contribuent à une meilleure compréhension du développement des racines et peuvent nous aider à comprendre les processus complexes impliqués dans l'arsenal de défense des racines de tomates.

Cette méthode a également été appliquée sur des groupes de racines matures de plants de tomates, pour éviter les variations observées dans les racines jeunes, attaquées par différents ravageurs ; *Macrosiphum euphorbiae*, *Helicoverpa armigera*, *Oidiumneo lycopersici*, *Meloidogyne incognita*.

Les premiers résultats ont mis en évidence un biomarqueur principal ; Feruloyltyramine ( $C_{18}H_{19}NO_4$ ), régulé à la hausse dans les racines infestées par rapport aux racines saines. Ce composé appartenant à la famille des phénolamides a déjà été décrit dans la littérature comme étant fortement induit dans les feuilles de tomate infectées par la bactérie *Pseudomonas syringae* alors qu'il est présent à de faible concentration dans la plante saine<sup>163</sup>.

Sur la base de ces résultats, l'approche métabolomique non ciblée accompagnée de réseaux moléculaire développée devrait être considérée comme une stratégie pertinente et efficace pour évaluer les variations chimiques des métabolomes des plantes. Dans cette étude, la méthode a été développée pour un intérêt écologique, mais pourrait très bien être appliquée sur des extraits de plantes d'intérêt cosmétique.



**CHAPITRE 4 – COMBINAISON DE RÉSEAUX  
MOLÉCULAIRES ET FRACTIONNEMENT CPC POUR  
CIBLER DES MOLÉCULES D'INTÉRÊT COSMÉTIQUE**

---



## I. INTRODUCTION

L'industrie cosmétique est un secteur dynamique en progression croissante en France et dans le monde<sup>196</sup>. La France compte 79 000 établissements liés au monde des produits cosmétiques, leur vente a franchi la barre des 15 milliards d'euros en 2019 et affiche 14,5 milliards d'euros d'exportation confirmant ainsi son rôle de Leader mondial devant les Etats-Unis et l'Allemagne<sup>197</sup>. Les consommateurs quant à eux, s'intéressent de plus en plus à la composition des produits qu'ils achètent. Ils souhaitent des ingrédients cosmétiques naturels dont la vente a augmenté de 66% en 2019. Ceci est dû d'une part, à la remise en cause de l'innocuité de certains produits d'origine synthétique par des études scientifiques et par les associations de consommateurs depuis les années 2000 et d'autre part, à l'urgence au niveau mondial de renforcer la protection environnementale, le développement durable et la protection de la biodiversité. Par conséquent, la recherche d'ingrédients naturels engage les fabricants cosmétiques dans une politique de responsabilité sociétale, de préservation de la biodiversité et d'utilisation durable de ressources renouvelables en utilisant des procédés écoresponsables.

De manière générale, les ingrédients végétaux actifs sont intégrés dans une formule cosmétique directement sous forme d'un mélange de composés (complexe si c'est un extrait de plante ou plus simplifié si c'est une fraction). La propriété biologique du mélange de composés peut soit provenir d'une seule molécule bioactive ou bien d'un ensemble de composés agissant potentiellement en synergie pour assurer la propriété biologique du produit cosmétique.

La compréhension de l'activité biologique d'un extrait de plante repose donc sur le criblage et la caractérisation des molécules qui lui confèrent ses propriétés biologiques qui sont ensuite utilisées comme marqueurs pour contrôler la qualité des extraits au cours du temps. De plus, l'expression des molécules bioactives par la plante peut être régulée dans l'objectif de favoriser la biosynthèse de ces molécules d'intérêt cosmétique en adaptant les conditions de culture.

Jusqu'à présent, cette étape de criblage et d'identification des composés bioactifs d'un extrait de plante était souvent longue et fastidieuse, puisqu'elle consistait à purifier, tester et identifier individuellement différents composés d'un extrait riche en molécules. C'est pourquoi de nouvelles approches plus simples et rapides se développent. Ainsi, afin de mettre en évidence les composés qui sont impliqués dans les activités biologiques d'intérêt cosmétique de deux extraits de plantes, nous avons combiné i) les informations de déréplication et de similarités structurales obtenues par les réseaux moléculaires et ii) le fractionnement de l'extrait brut par chromatographie de partage centrifuge CPC (*Centrifugal Partition Chromatography*), une

technique de chromatographie préparative liquide-liquide qui présente de nombreux avantages et qui permet d'obtenir différentes fractions simplifiées à partir d'un extrait brut complexe. En utilisant cette combinaison de techniques, une méthode a été développée pour produire et tester des fractions simplifiées et faciliter le criblage et l'identification des métabolites actifs. Cette méthodologie a été appliquée à deux extraits de plantes cultivées en région Centre-val de Loire pour la recherche des composés d'intérêt cosmétique. Les plantes sélectionnées sont *Artemisia annua* L. et *Eclipta alba* L., qui ont fait l'objet de plusieurs études pour leurs propriétés biologiques et leurs richesses moléculaires.

## II. DÉVELOPPEMENT MÉTHODOLOGIQUE

### A. Les plantes

#### 1) *Artemisia annua* L.

Le genre *Artemisia* regroupe un grand nombre d'espèces dont la culture a permis notamment la production d'épices (*Artemisia dracunculus*), de liqueur comme l'absinthe (*Artemisia absinthium*) et de répulsifs (*Artemisia vulgaris*).

*Artemisia annua* ou armoise annuelle, est l'une des 500 espèces du genre *Artemisia* (Figure 46)<sup>198</sup> appartenant à la famille des astéracées. Il s'agit d'une plante herbacée annuelle, glabre et très odorante, de 30 à 150 cm de haut, pouvant atteindre parfois 250 cm dans les conditions optimales de culture. Elle est traditionnellement utilisée sous forme de thé ou de jus pressé pour le traitement du paludisme en Asie et en Afrique<sup>199</sup>.



Figure 46 : *Artemisia annua* L.<sup>200</sup>

Bien qu'*Artemisia annua* soit originaire des régions tempérées de Chine où elle pousse dans une végétation de steppe à 1000-1500 m d'altitude, son habitat est bien plus large. À partir de la chine, elle s'est propagée à l'état sauvage dans de nombreux pays<sup>201</sup>. En France, elle pousse d'août à octobre dans le sud (Var, Hérault, Gard, Isère) et en région parisienne dans les terrains vagues, les chemins et les bords de route.

Cette espèce est la principale source d'artémisinine, un endoperoxyde sesquiterpene lactone<sup>202</sup> responsable de l'activité antipaludique<sup>203</sup> utilisé comme principe actif dans les médicaments antipaludiques (Figure 47). Cette molécule a été découverte par la chercheuse Tu Youyou qui a reçu le prix Nobel en 2015 pour ses travaux sur l'*Artemisia annua* et le paludisme (malaria).

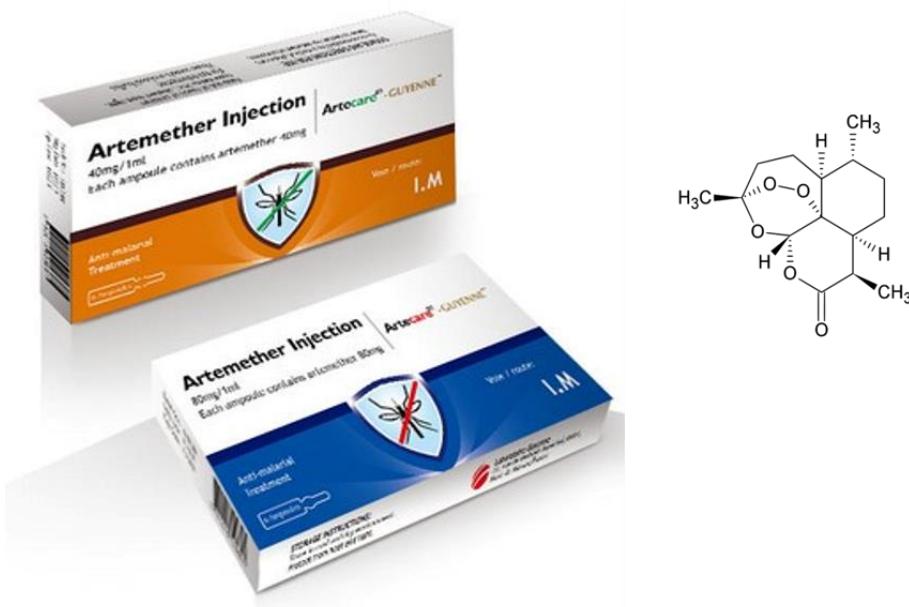


Figure 47 : Artémisinine, principe actif de médicaments antipaludiques

En plus de l'artémisinine, *Artemisia annua* présente un profil nutritionnel abondant, avec des acides aminés, des vitamines et des éléments minéraux<sup>204</sup>. Au fil des ans, plus de 600 métabolites secondaires ont été identifiés dans la plante entière<sup>205</sup>, dont plusieurs sesquiterpénoïdes, triterpénoïdes, monoterpénoïdes, stéroïdes, flavonoïdes, coumarines, alcaloïdes et benzénoïdes<sup>206</sup>.

Cette grande variété de composés et la richesse des extraits confèrent à *l'Artemisia annua* un certain nombre de propriétés biologiques qui ont suscité l'intérêt des chercheurs, on retrouve par exemple des propriétés hépatoprotectrices<sup>207</sup>, antifongiques<sup>208</sup>, antitumorales<sup>209</sup> et antioxydantes<sup>210</sup>. *Artemisia annua* est donc une plante essentiellement à usage thérapeutique et aucune utilisation cosmétique n'a été recensée.

Cependant, l'activité antioxydante est d'intérêt pour la cosmétique et facilement évaluable avec les tests colorimétriques tels que les tests DPPH, FRAP, ABTS, etc. Bien que plusieurs études aient démontré la capacité antioxydante d'*Artemisia annua*<sup>211,212</sup>, l'activité n'a jamais été corrélée à l'identification des molécules actives. De plus, la richesse en familles moléculaires d'*Artemisia Annua* en fait une plante modèle bien adaptée au développement méthodologique du criblage de molécules actives.

## 2) *Eclipta alba* L.

*Eclipta alba* L. ou *Eclipta prostrata* L., communément appelée le Mahakanni en français, *False dasy* en anglais, *Bhringraj* en ayurveda, et *Ecliptae herba* en chinois, est une plante annuelle herbacée de la famille des astéracées (Figure 48). Elle fait partie du genre *Eclipta* qui compte

selon les considérations taxonomiques de 1 à 6 espèces<sup>213</sup>. Elle mesure en moyenne entre 10 et 50 cm et peut atteindre jusqu'à 70 cm de haut dans les conditions de culture optimales.



Figure 48 : *Eclipta alba* L.<sup>214</sup>

Plusieurs sources bibliographiques indiquent une origine du sud de l'Asie tandis que d'autres mentionnent un indigénat en Amérique<sup>215</sup>. En effet, *Eclipta alba* est largement répartie dans les régions tropicales, subtropicales, et zones tempérées chaudes du globe. Elle est présente dans la plupart des collectivités françaises d'outre-mer et également en métropole, où elle aurait été introduite de façon accidentelle en tant qu'adventice dans des lots de semences dans les rizières de Camargue. La région Provence-Alpes-Côte-D'azur constitue la zone principale de sa présence.

De nombreux types de composés chimiques ont été isolés et identifiés à partir de différents extraits d'*Eclipta alba*, notamment des alcaloïdes, alcénones, glycosides cardiaques, flavonoïdes, coumestans, lipides, polyacétylène, stéroïdes, saponines, alcaloïdes stéroïdiens, phytostérols et des triterpènes<sup>216,217,218,219,220</sup>. Il est important de mentionner que les facteurs environnementaux comme les sources géographiques, le temps de récolte et le temps de stockage, sont souvent à l'origine de la variabilité des composants chimiques contenus dans *Eclipta alba*<sup>216</sup>. Par conséquent, cette plante a une composition chimique très riche et contient des familles moléculaires qui lui confèrent des activités biologiques diverses et variées.

De plus, *Eclipta alba* est inscrite sur la liste des plantes de la pharmacopée française<sup>221</sup> et est largement utilisée pour ses propriétés médicinales et cosmétiques. En effet, cette espèce est connue pour ses propriétés curatives et a été utilisée comme anti-hépatotoxique<sup>222,223</sup>, antifongique<sup>224</sup>, anticancéreuse<sup>225,226</sup>, anti-hyperlipidémie<sup>227</sup>, anti-adipogène<sup>228,229</sup>, antibactérienne<sup>230,231</sup>, antipaludéenne<sup>232</sup>, anti-hyper glycémique<sup>233</sup>, anti-inflammatoire<sup>234</sup>, anti-protéolytique et antihémorragique<sup>235</sup>. De plus, des études ont montré l'activité analgésique d'*Eclipta alba*<sup>236</sup>, ses effets neuropharmacologiques<sup>237,238</sup> et son activité immunomodulatrice<sup>239</sup>.

Cette espèce s'est également avérée particulièrement efficace en agriculture, une augmentation des rendements de sorgho a été observée en traitant les semences avec un extrait aqueux d'*Eclipta alba* qui a impliqué un double mécanisme d'hydro-amorçage et de suppression des pathogènes fongiques<sup>240</sup>.

En ce qui concerne l'activité cosmétique, *Eclipta alba* entre dans la composition de certains laits autobronzants, soins capillaires et soins de la peau, elle est commercialisée sous forme de poudre (Figure 49), pour lutter contre la chute de cheveux et les cheveux blanc précoce<sup>241,242</sup>. En outre, quelques études ont démontré son activité antioxydante en utilisant des tests chimiques : capacité de piégeage des radicaux libres de l'oxyde nitrique et de DPPH, la capacité antioxydante totale, l'activité d'inhibition de la peroxydation des lipides, mais sans identifier les molécules qui pourraient être responsables de cette activité antioxydante<sup>243,244,245</sup>. Enfin, aucune évaluation de l'activité anti-enzymatique n'a été étudiée sur *Eclipta alba* pour des intérêts cosmétiques.



Figure 49 : Quelques produits cosmétiques commercialisés à base d'*Eclipta alba*<sup>246</sup>

Par conséquent, ces deux plantes ; *Artemisia annua* et *Eclipta alba*, possédant des propriétés biologiques d'intérêt et ayant une origine lointaine ont suscité notre intérêt afin d'évaluer la possibilité de les cultiver localement tout en conservant de bons rendements de biomasse et de contenu métabolique d'intérêt et ainsi, développer une voie d'approvisionnement en circuit court.

Une démarche a donc été développée pour le criblage de molécules à activité antioxydante dans les extraits éthanoliques d'*Artemisia annua* et d'*Eclipta alba*, ainsi que des inhibiteurs potentiels de la collagénase, d'élastase et de tyrosinase dans l'extrait éthanolique d'*Eclipta alba*.

## B. Préparation des échantillons

### 1) Culture de la matière végétale

Le Comité de Développement Horticole de la Région Centre-Val de Loire (**CDHRC**, Saint Cyr en Val) est une association née en 1981 de la volonté d'horticulteurs et de pépiniéristes de la région de créer une structure commune d'expérimentation, de formations et de conseils. Le **CDHRC** accompagne ainsi depuis plus de 30 ans les professionnels horticulteurs et pépiniéristes adhérents à travers l'évolution des métiers grâce à des expérimentations en cultures ornementales, des formations et du conseil. Cette structure mène différentes expérimentations afin d'optimiser les protocoles de culture dans le but de développer des méthodes alternatives pour la maîtrise des ravageurs et d'éventuelles maladies ainsi que pour l'amélioration de la qualité de production. Le **CDHRC** joue le rôle d'interface entre les différents acteurs du végétal et s'implique également dans des projets innovants, de leur conception jusqu'à leur transfert en entreprise. C'est ainsi que ces dernières années le **CDHRC** a souhaité élargir ses compétences dans la culture de plantes d'intérêt alimentaire, cosmétique, thérapeutique qui ne sont pas originaires de la Région Centre-Val de Loire afin de pouvoir assurer une production locale de ces végétaux. En fonction des plantes, des organes et des molécules d'intérêt, des systèmes de culture biologique, de culture par hydroponie et aéroponie peuvent être mis en œuvre ainsi que des systèmes d'élicitation lumineuse qui peut favoriser la biosynthèse des molécules d'intérêt. Ce développement d'activité permet au **CDHRC** d'apporter de nouvelles voies de valorisation des cultures à ses adhérents et d'optimiser l'utilisation de ses serres et de son matériel.

C'est dans ce contexte que les plantes *Artemisia annua* et *Eclipta alba* issues respectivement de la pharmacopée chinoise et de la médecine ayurvédique ont été cultivées au sein du **CDHRC**, sous serres en conditions contrôlées. Tous les paramètres influençant la culture de la plante (luminosité, température, hydrométrie, etc.) ont été étudiés et optimisés par le **CDHRC** pendant environ 2 ans afin que le rendement en biomasse soit au moins équivalent à celui obtenu en conditions naturelles d'origine.

Les parties aériennes ont été collectées (en 2017), séchées à l'air libre sous serre pendant environ 2 semaines et broyées à l'aide d'un broyeur par couteaux POLYMIX PX-MFC 90 D KINEMATICA (granulométrie < 40 µm) afin d'obtenir une poudre fine et homogène.

## 2) Extraction

L'extraction assistée par les ultrasons a été utilisée pour cette étude permettant un temps d'extraction réduit par rapport à une macération traditionnelle, une réduction de la consommation de solvants organiques, d'énergie et de coût par rapport à des techniques d'extraction plus sophistiquées (Microonde, ASE)<sup>247</sup>. De plus cette technique pouvait être mise en œuvre par l'ensemble des partenaires académiques du projet ValBioCosm' pour effectuer des extractions dans des conditions similaires. L'éthanol a été utilisé comme solvant

d'extraction afin de solubiliser des composés d'un large éventail de polarité. Pour chacune des deux plantes modèles, 5 g des parties aériennes séchées et broyées ont été introduites dans 50 mL d'éthanol pour une extraction d'une heure aux ultrasons. Une centrifugation est ensuite effectuée pour séparer l'extrait du marc (10 000 g pendant 15 min à température ambiante). Le surnageant est récupéré et séché sous un flux d'azote afin d'obtenir les extraits secs d'*Artemisia annua* et d'*Eclipta alba*. Le rendement d'extraction défini comme la masse de l'extrait brut sec sur la masse de la matière première sèche engagée était de 2,3% pour *Artemisia annua* et 4% pour *Eclipta alba*.

### C. Fractionnement par Chromatographie de Partage Centrifuge (CPC)

La richesse et la complexité de la composition chimique d'un extrait de plante sont les principales difficultés des études phytochimiques et de l'identification des molécules présentant l'activité biologique ciblée. Afin de remédier à cela, une simplification des extraits bruts en différentes fractions a été effectuée par CPC. L'activité biologique des différentes fractions sera évaluée avec celle de l'extrait brut afin d'identifier les molécules majoritaires participant à l'activité de l'extrait.

La CPC est une technique de chromatographie liquide-liquide préparative qui utilise un système de solvants biphasique sans phase stationnaire solide. La séparation des molécules est basée sur leur différence de partage entre les deux phases liquides non miscibles. Une colonne CPC est constituée d'une superposition de plusieurs disques en acier inoxydable et sur lesquels sont gravés des cellules de partage reliées entre elles par des capillaires (Figure 50), qui peuvent être assimilées à une succession d'ampoules à décanter.

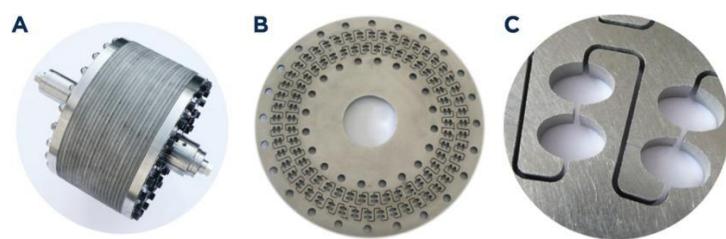


Figure 50 : La colonne constituée de plusieurs disques sur un rotor (A), ces disques (B) contiennent de cellules, qui sont connectées par un canal mince (C). Images adaptées d'Armen, Gilson<sup>248</sup>

Après la sélection du système biphasique, qui dépend fortement de l'échantillon à fractionner et des familles moléculaires, une des deux phases est pompée dans la colonne. Lorsque cette dernière est mise en rotation, le champ de force centrifuge maintient la phase pompée stationnaire au sein de la colonne. L'autre phase, dite mobile, est pompée au travers de la phase stationnaire éluant ainsi les composés injectés qui seront séparés selon leurs coefficients de partage  $K_D$ .

Le  $K_D$  est défini par la concentration d'un composé A dans la phase stationnaire sur sa concentration dans la phase mobile comme présenté sur la Figure 51.

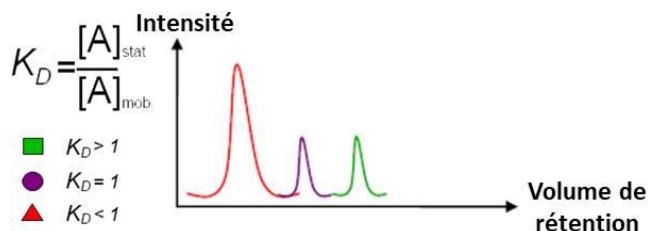


Figure 51 : Élution des composés en CPC selon leurs coefficients de partage  $K_D$  (Kromaton)

L'un des principaux avantages de la CPC est la large variété de systèmes biphasiques qui peuvent être appliqués lorsqu'on travaille sur des échantillons naturels et qui peuvent couvrir toute la gamme de polarité d'un extrait. En effet, il existe différents systèmes biphasiques allant de deux à cinq solvants qui sont largement utilisés pour le fractionnement des extraits végétaux. On retrouve par exemple le système HEMWat<sup>249</sup> constitué du mélange (hexane : acétate d'éthyle : méthanol : eau) ou le système Arizona<sup>250</sup> (heptane : acétate d'éthyle : méthanol : eau) ou encore un mélange de cinq solvants (hexane : acétate d'éthyle : butanol : méthanol : eau), etc. Lorsque la phase mobile est plus dense dans le système biphasique, elle est pompée en mode descendant et si elle est plus légère, elle est pompée en mode ascendant pour pouvoir dans les deux cas, traverser la phase stationnaire et éluer les composés.

L'autre avantage majeur de la CPC est l'utilisation d'une phase stationnaire liquide, puisqu'elle permet de récupérer l'intégralité des composés injectés qu'ils soient élusés dans la phase mobile ou retenus dans la phase stationnaire avec peu ou pas de dégradation grâce à l'absence d'adsorption irréversible. En effet, la CPC permet de travailler avec différents modes de développement dont l'un des plus utilisés est le mode d'élution-extrusion qui consiste, après une période d'élution classique, à pousser avec de la phase stationnaire fraîche la phase mobile puis la phase stationnaire initiales hors de la colonne, et ce sans intervertir le sens de déplacement des solvants. Cette méthode permet de récupérer les deux phases liquides initialement présentes dans la colonne et contenant potentiellement encore des molécules d'intérêt après la phase d'élution et ainsi de collecter l'ensemble des constituants de l'échantillon, tout en diminuant le volume de solvant normalement nécessaire à l'élution des molécules ayant une forte affinité pour la phase stationnaire<sup>251,252</sup>.

D'autres modes de séparations peuvent également être mis en œuvre afin d'améliorer la séparation/purification des molécules tels que l'élution par *Dual mode*<sup>253</sup> (inversion du sens d'élution des composés en cours de séparation), le *déplacement*<sup>254</sup> (échange d'ions ou pH-zone refining utilisés pour les composés ionisés ou ionisables) et *Capture ou Centrifugal*

*Partition Extraction CPC*<sup>255</sup> utilisé pour l'extraction des composés d'intérêt. La CPC est donc une technique versatile avec plusieurs modes de développement permettant son utilisation dans divers domaines pour différents types d'échantillons.

Enfin, cette technique de chromatographie a une forte capacité de charge permettant de purifier de grande quantité d'extrait par rapport au volume de solvant utilisé. Néanmoins, son principal inconvénient reste sa plus faible efficacité (nombre de plateaux théoriques) par rapport à l'HPLC préparative qui peut induire des résolutions de pics insatisfaisantes lorsqu'elle n'est pas compensée par la sélectivité du système solvant biphasique utilisé.

Durant cette thèse, la CPC a été utilisée pour fractionner et simplifier les extraits d'*Artemisia annua* et d'*Eclipta alba*. Pour ce faire, différents systèmes biphasiques ont été testés (1 mg d'extrait dans 4 mL du mélange biphasique) afin de trouver le système optimal qui permet une bonne solubilité de l'extrait et une répartition différente des composés entre les deux phases, ce qui conduit à leur séparation en différentes fractions.

Six systèmes de solvants différents de la table Arizona (heptane : acétate d'éthyle : méthanol : eau) ont été testés : D (1:6:1:6), H (1:3:1:3), M (5:6:5:6), P (6:5:6:5), R (2:1:2:1) et U (4:1:4:1). Ces systèmes couvrent un large éventail de polarités liées aux familles moléculaires, ce qui augmente les chances de cibler le système biphasique optimal. Les systèmes D et H ont entraîné des temps de stabilisation très longs, les systèmes R et U ont un bon temps de stabilisation mais la majorité des composés est restée en phase organique. Enfin, les systèmes M et P, avec des temps de stabilisation du système biphasique courts d'environ 10 secondes, ont permis une répartition satisfaisante des composés entre les deux phases.

Les phases organiques et aqueuses de ces deux systèmes (M et P) ont été analysées par HPLC et le système M a été sélectionné car il offrait une meilleure répartition des composés. Les valeurs de  $K_D$  des composés majoritaires étaient supérieures à 1 pour éviter leur élution dans le volume mort lors du fractionnement CPC et suffisamment différentes les unes des autres pour qu'ils soient séparés dans des fractions différentes. Les composés ayant une valeur de  $K_D$  intermédiaire peuvent sortir de la colonne pendant l'étape d'élution, tandis que certains composés non polaires ayant une valeur de  $K_D$  élevée peuvent être récupérés dans la phase organique pendant l'étape d'extrusion. Les modes opératoires utilisés pour chaque extract sont présentés en Annexe 2.

#### **D. Évaluation des activités biologiques**

Les activités biologiques des extraits bruts ont d'abord été évaluées. Les extraits d'*Artemisia annua* et d'*Eclipta alba* présentant des activités intéressantes ont été fractionnés par CPC et l'activité biologique de chaque fraction a été évaluée dans le but de cibler les fractions actives.

Lors de cette thèse, des tests chimiques, enzymatiques et cellulaires ont été utilisés pour rechercher des propriétés biologiques d'intérêt cosmétique.

## 1) Activité antioxydante

L'activité antioxydante est importante pour protéger le système biologique, car les espèces réactives de l'oxygène (ROS) jouent un rôle important dans l'oxydation des protéines, des lipides ou de l'ADN dans les tissus, entraînant des maladies dégénératives/chroniques, y compris le cancer, le diabète ou les maladies cardiovasculaires<sup>256</sup>. Les ROS sont des espèces très toxiques mais essentielles pour la signalisation et la régulation cellulaire. Leur concentration est régulée par le maintien d'un équilibre entre leur production, lors de la réduction de l'oxygène dans le métabolisme normal, et leur élimination par les systèmes antioxydants intrinsèques des cellules d'un organisme<sup>257</sup>.

Le stress oxydatif est un des mécanismes majeurs impliqué dans le vieillissement cutané et les affections dermatologiques<sup>258</sup>. Une exposition à des facteurs environnementaux tels que les rayons UV, entraîne l'altération des tissus conjonctifs par la formation excessive de ROS, et impacte l'activité de certaines enzymes, induisant ainsi des problèmes cutanés<sup>259</sup> (Figure 52).

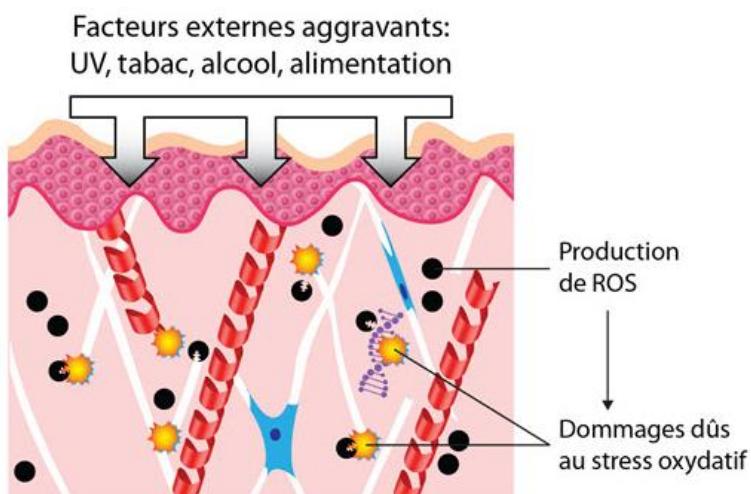


Figure 52 : Illustration du stress oxydatif et des facteurs externes aggravant sur la peau<sup>260</sup>

Le rôle principal des composés antioxydants est de retarder ou d'inhiber l'oxydation des lipides ou d'autres molécules en empêchant le déclenchement ou l'aggravation des réactions d'oxydation en chaîne<sup>261</sup>. Ils présentent un grand intérêt pour les applications cosmétiques car ils préviennent les dommages cutanés et le vieillissement précoce.

De nombreux composés naturels sont des antioxydants efficaces pour lutter contre ces ROS (tanins, acides phénoliques, acides cinnamiques, lignanes, stilbènes, flavonoïdes, etc.)<sup>262,257</sup>. Ils peuvent prévenir les effets de ces radicaux libres en les neutralisant ou en favorisant leur décomposition, et interviennent, selon leur nature, à différents stades des chaînes de réactions oxydatives (Figure 53)<sup>259,263</sup>.

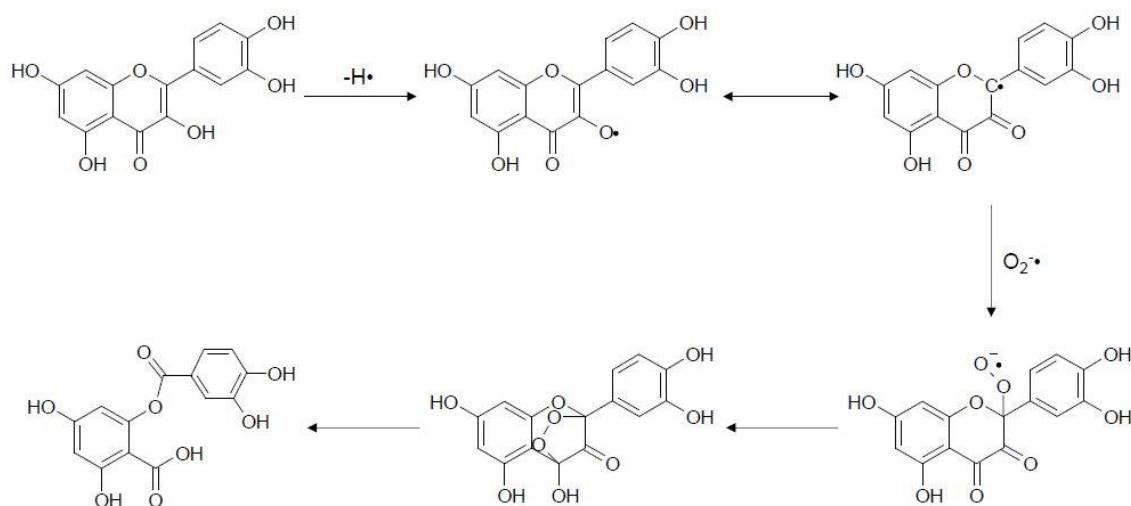


Figure 53 : Mécanisme d'action anti-radicalaire. Exemple de piégeage d'anion superoxyde par la quercétine présenté par Balasaheb Nimse et al (2015)<sup>263</sup>

Un certain nombre d'études décrivent des techniques colorimétriques simples, faciles à mettre en œuvre et à interpréter et adaptées pour l'estimation du pouvoir antioxydant. Parmi les méthodes les plus couramment utilisées, on retrouve notamment les méthodes DPPH (2,2-diphényl-1-picrylhydrazyl), ABTS (2,2-azino-bis-(3-éthylbenzothiazoline-6-sulfonic acid)), FRAP (*ferric reducing antioxidant power*), CUPRAC (*cupric reducing antioxidant capacity*), et chélation du fer<sup>264,265</sup>.

Ces essais reposent sur les réactions de transfert d'un seul électron ou d'un atome d'hydrogène et présentent soit une capacité de réduction (FRAP, CUPRAC), soit une inhibition directe des radicaux libres (DPPH, ABTS)<sup>266</sup> (Figure 54).

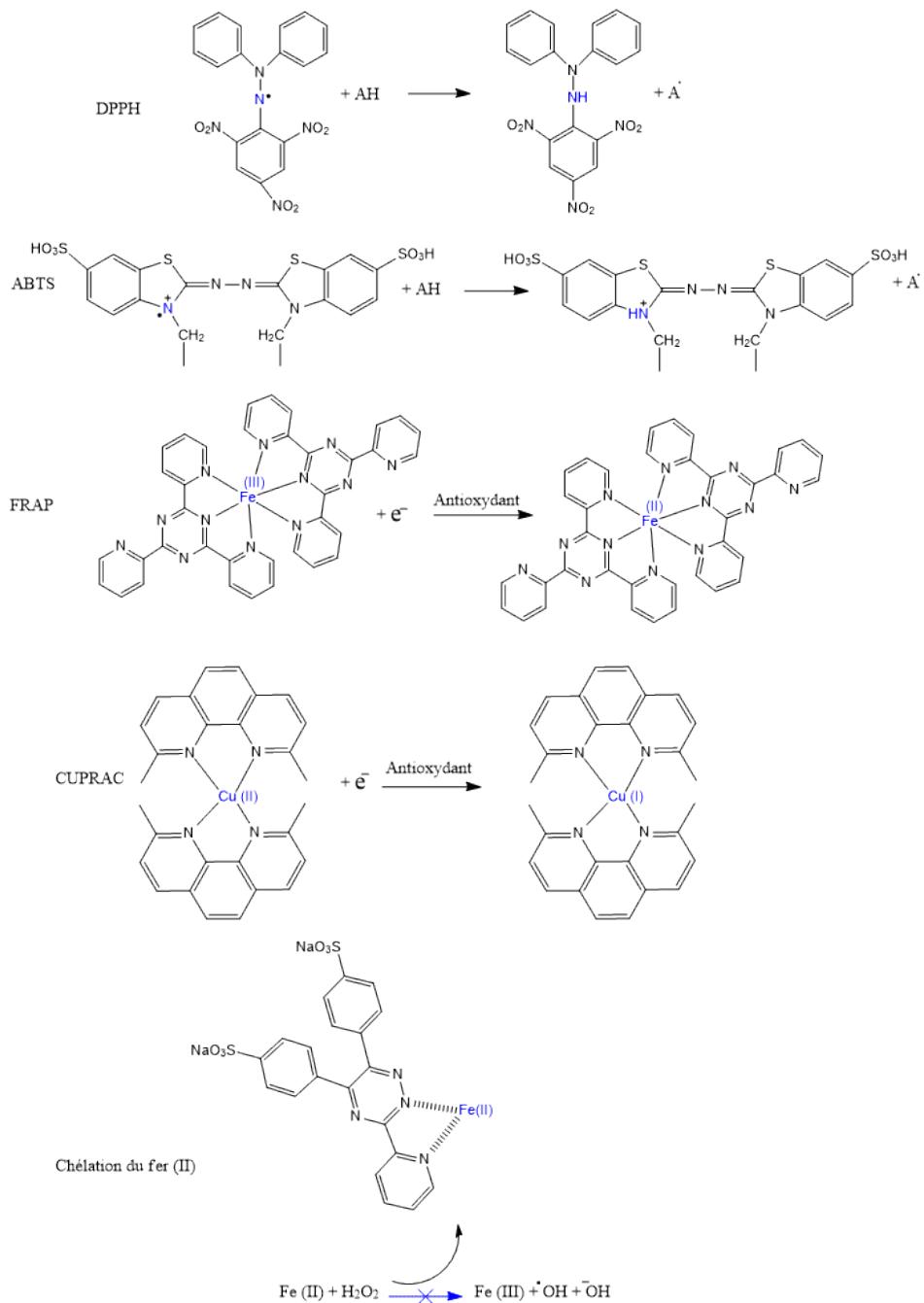


Figure 54 : Mécanismes réactionnels mis en jeux lors des tests chimiques colorimétriques en présence de composés antioxydants. DPPH (2,2-diphényl-1-picrylhydrazyl), ABTS (2,2-azino-bis-(3-éthylbenzothiazoline-6-sulfonic acid)), FRAP (2,3,5-triphenyltetrazolium), CUPRAC (neocuproine copper) et Chélation du fer (ferrozine)

Les résultats obtenus avec les tests chimiques *in vitro* ne peuvent pas être directement transférés à un effet antioxydant *in vivo*, néanmoins, ce sont des tests de première intention avant une étude plus approfondie de l'activité des extraits de plantes et le développement d'un produit cosmétique.

Par conséquent, la capacité antioxydante des extraits bruts et fractions CPC obtenues pour *Artemisia Annua* et *Eclipta alba* a été évaluée à l'aide des tests de DPPH, ABTS, FRAP, CUPRAC et de chélation du fer (II). Elle a été déterminée en mesurant l'absorbance du réactif en présence des différents échantillons (solvant pour le blanc, trolox pour le contrôle positif, extrait brut ou fractions) puisque les réactions d'oxydation/réduction et de complexation sont mises en évidence par un changement de couleur de la solution.

L'évaluation de l'activité antioxydante a été réalisée dans des plaques de 96 puits ; chaque échantillon a été déposé à 3 concentrations en cascade. L'éthanol a été utilisé pour diluer les échantillons et comme contrôle négatif. Le trolox, un puissant antioxydant, a été utilisé comme témoin positif et évalué à différentes concentrations. Enfin, tous les essais ont été effectués en trois répétas ( $n = 3$ ) en suivant le mode opératoire présenté en Annexe 3.

## 2) Activité anti-enzymatique

Outre les tests colorimétriques pour évaluer l'activité antioxydante, il existe les tests d'inhibition enzymatique d'intérêt pour des applications cosmétiques. Les principaux sont les tests d'inhibition de la collagénase, de l'élastase et de la tyrosinase, qui sont des enzymes très intéressantes en raison de leur implication directe dans le vieillissement de la peau, elles sont considérées comme des marqueurs enzymatiques majeurs pour la santé de la peau<sup>267,268,269</sup>.

En effet, les collagénases sont responsables de la dégradation du collagène, une protéine fibreuse insoluble majeure qui a pour fonction première de maintenir la fermeté de la peau<sup>270</sup> (Figure 55 A). Quant à l'élasticité de la peau, elle est considérablement réduite par l'activité de l'élastase qui dégrade les fibres d'élastine, une protéine vitale pour les propriétés élastiques, provoquant le relâchement de la peau<sup>271</sup> (Figure 55 B). Enfin, la tyrosinase est une catéchol oxydase multifonctionnelle contenant du cuivre, qui joue un rôle essentiel dans la biosynthèse de la mélanine responsable de la couleur de la peau, des cheveux et des yeux chez l'Homme (Figure 55 C). Cependant, la surproduction de mélanine dans les couches épidermiques peut provoquer de graves problèmes dermatologiques, comme le lentigo, le mélasme et l'hyperpigmentation dus au vieillissement de la peau. C'est pourquoi les inhibiteurs de la tyrosinase sont considérés comme des agents potentiels de blanchiment de la peau.

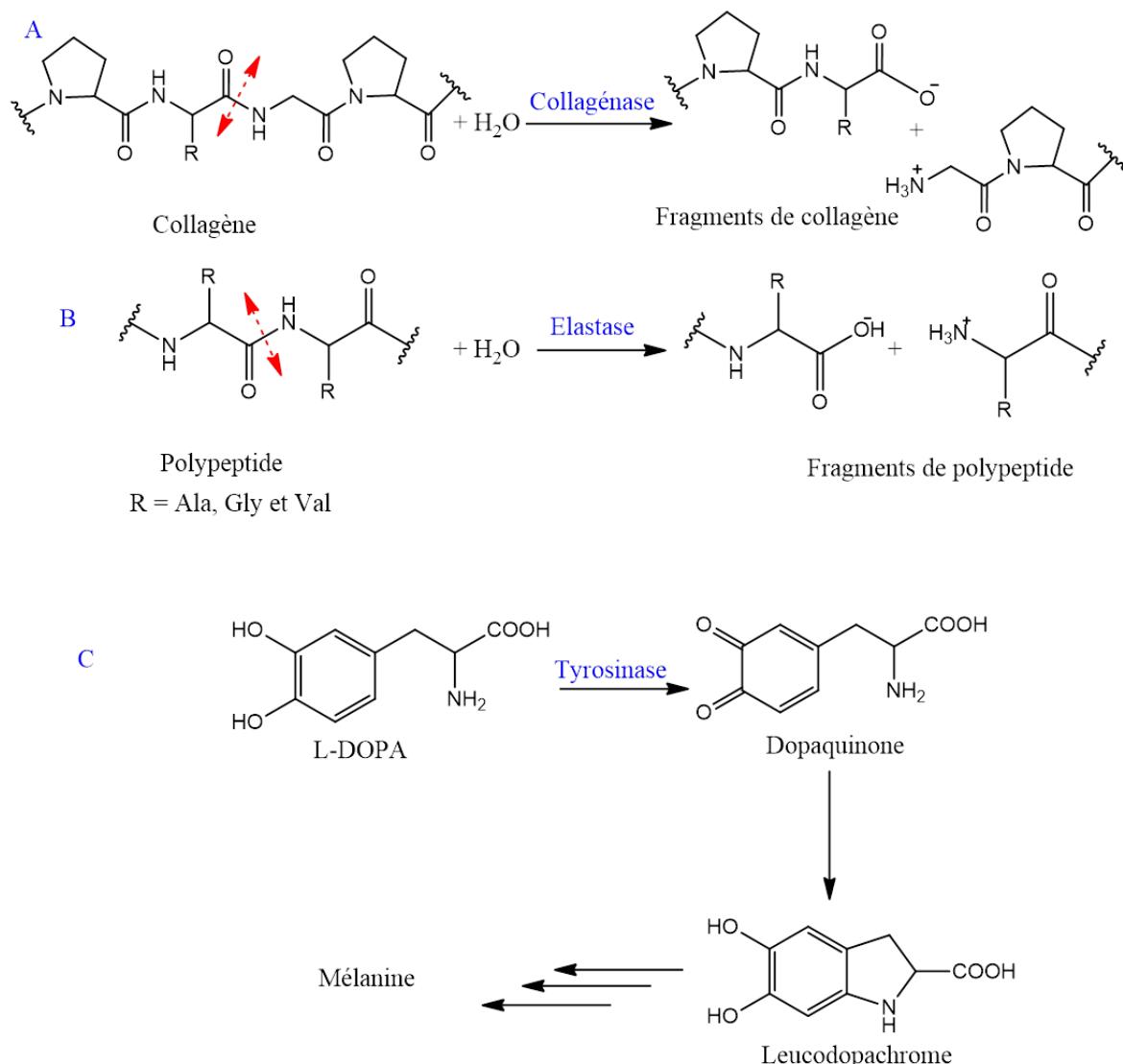


Figure 55 : Illustration des voies d'hydrolyse de collagénase et d'élastase, ainsi que l'implication de la tyrosinase dans la synthèse de mélanine

Les composés naturels sont recherchés car les produits synthétiques bien connus pour inhiber la tyrosinase, par exemple l'hydroquinone, ont une toxicité et des effets secondaires importants<sup>272</sup>. Par conséquent, nous avons étudié l'existence d'inhibiteurs potentiels de collagénase, elastase et tyrosinase dans l'extrait d'*Eclipta alba*.

Tous les tests enzymatiques ont été effectués dans des plaques de 96 puits où chaque essai a été répété 3 fois ( $n = 3$ ). Le mode opératoire des tests enzymatiques est détaillé en Annexe 3.

### 3) Activité cellulaire

Cette étape a été réalisée par nos collaborateurs biologistes du laboratoire Centre de Biophysique Moléculaire (**CMB**), CNRS Orléans.

Dans un premier temps, la toxicité cellulaire des extraits d'*Eclipta alba* et d'*Artemisia annua* a été mesurée sur des cellules de kératinocytes humains immortalisés HaCaT. Le test d'innocuité cellulaire (XTT) est effectué en présence de différentes concentrations d'extraits afin de déterminer la concentration maximale qui peut être utilisée sans induire de toxicité (maximum de viabilité cellulaire) pour les cellules. La viabilité cellulaire est proportionnelle à l'activité métabolique. En effet, le réactif, sel de tétrazolium (XTT), est réduit par la déshydrogénase mitochondriale en formazan (produit coloré) et la mesure de la densité optique permet de calculer le pourcentage de viabilité cellulaire. On considère qu'un actif est non toxique si la viabilité cellulaire est supérieure ou égale à 80%.<sup>273</sup>

Ensuite, l'activité antioxydante de l'extrait d'*Eclipta alba* a été évaluée sur les cellules kératinocytes humains immortalisés HaCaT exposés au H<sub>2</sub>O<sub>2</sub> comme source de production de ROS dans les cellules. L'objectif étant de valider au niveau cellulaire l'activité antioxydants observée par les tests chimiques.

## E. Étude phytochimique et élucidation structurale

Afin d'évaluer la composition phytochimique des extraits et d'identifier les molécules actives nous avons dans un premier temps générés les réseaux moléculaires des extraits bruts d'*Artemisia annua* et d'*Eclipta alba* pour obtenir les informations de déréPLICATION et visualiser les différentes familles moléculaires, les relations structurales entre les composés appartenant à une même famille. Puis les différentes fractions générées par CPC ont été analysées afin d'identifier les composés responsables des activités observées.

### 1) Réseaux moléculaires de l'extrait brut

Les réseaux moléculaires des extraits bruts d'*Artemisia annua* et d'*Eclipta alba* ont été générés à partir des analyses UHPLC-MS/HRMS. Les analyses MS et MS/MS ont été acquises en utilisant le mode d'acquisition DDA avec l'ionisation ESI qui a permis de détecter un grand nombre de composés sur une large gamme de polarité. Le mode positif a été appliqué car il engendre de meilleures ionisation et fragmentation des composés et permet d'obtenir des spectres MS/MS plus riches en information que les spectres obtenus en mode négatif, ce qui augmente le nombre de pics à comparer lors de la construction des réseaux moléculaires. Néanmoins, le mode négatif a été acquis et a aidé à l'identification des composés notamment pour *Eclipta alba*.

L'azote a été utilisé comme gaz de séchage à un débit de 9 L/min chauffé à 200°C et comme gaz de nébulisation à une pression de 2 bar, la tension du capillaire a été fixée à 4,5 kV. Les spectres de masse MS et MS/MS ont été enregistrés dans la gamme  $m/z$  50-1650.

Compte tenu des spectres MS obtenus, trois ions précurseurs ont été sélectionnés par cycle et fragmentés, à deux énergies de collision, dont la moyenne a été calculée dans un spectre MS/MS. De plus, un temps d'exclusion dynamique a été défini pour que chaque ion parent déjà fragmenté soit exclu d'une deuxième fragmentation pendant 0,03 min sauf si son intensité augmente d'un facteur de 3 après une première fragmentation. Le but de cette exclusion est de permettre à l'appareil de se focaliser sur l'analyse des ions les moins intenses après fragmentation des ions majoritaires.

Les énergies de collision ont également été optimisées sur chacun des extraits puisqu'elles dépendent fortement des familles moléculaires et des composés analysés. L'objectif était d'arriver à fragmenter les composés rigides tout en appliquant une énergie de collision plus faible pour les molécules plus fragiles afin d'obtenir des spectres de masse propres et exploitables. Enfin, toutes les données ont été traitées à l'aide de DataAnalysis 4.4 (Bruker), les formules moléculaires ont été générées en utilisant la composition élémentaire (C, H, O) sans aucune restriction, Na ≤ 1 et N ≤ 5 avec une précision de masse ≤ 4 ppm pour la MS et ≤ 7 ppm pour la MS/MS et les petites valeurs  $m/z$ .

Une fois les données spectrales générées, elles ont été transférées sur la plateforme GNPS sous forme de *peaklist* (.mgf) après l'étape de prétraitement de données. Deux méthodes différentes de prétraitement ont été utilisées pour les deux extraits et seront détaillées plus tard. Enfin, pour générer les réseaux moléculaires des deux échantillons sur la plateforme GNPS, les paramètres ont été optimisés pour chacun des extraits de plantes indépendamment et seront détaillés plus tard.

## 2) Identification des composés bioactifs

Après évaluation de leurs activités biologiques, les fractions CPC des deux extraits de plantes ont été analysées en spectrométrie de masse dans le but d'étudier leurs compositions chimiques et identifier les molécules participant majoritairement à l'activité de l'extrait brut. Pour ce faire, les méthodes d'élucidation de structure *de novo*, de déréPLICATION et de réseaux moléculaires ont été appliquées.

En effet, après l'analyse des fractions actives par HRMS, les spectres MS et MS/MS ont été travaillés manuellement pour cibler les ions moléculaires, générer des formules brutes, les soumettre à des bases de données pour trouver les formules développées possibles, interpréter par la suite les spectres MS/MS et les voies de fragmentation, prendre en considération le temps de rétention et les données UV pour réduire au maximum les possibilités de formules développées.

De plus, les spectres MS/MS ont été comparés aux bases de données spectrales de GNPS lors de la génération des réseaux moléculaires. Les propositions de composés ont été vérifiées puisque les bases de données sont externes, il est donc important de tenir compte des variations instrumentales, utilisateurs, etc. Enfin, l'analyse des clusters des réseaux moléculaires créés qui associent les composés ayant des voies de fragmentation similaires permet d'identifier les nœuds voisins d'un composé identifié. En outre, l'injection des molécules standards ainsi que l'annotation des molécules déjà décrites dans la littérature des plantes étudiées a permis d'accélérer l'annotation des réseaux moléculaires et la caractérisation des molécules.

Il est important de noter que c'est l'ensemble des informations obtenues par les approches mentionnées ci-dessus qui va permettre de proposer une identification et d'augmenter le degré de confiance de l'élucidation structurale. Ces approches sont complémentaires et donnent des indices convergents qui confortent les identifications proposées.

### III. CRIBLAGE ET IDENTIFICATION DES MOLÉCULES BIOACTIVES

*Artemisia annua* et *Eclipta alba* ont fait l'objet de plusieurs études pour leur richesse moléculaire et leurs propriétés biologiques. Leurs propriétés cosmétiques ont été évaluées durant cette thèse en utilisant la combinaison de techniques de fractionnement par CPC et de réseaux moléculaires pour l'identification des molécules actives comme résumé dans la Figure 56.

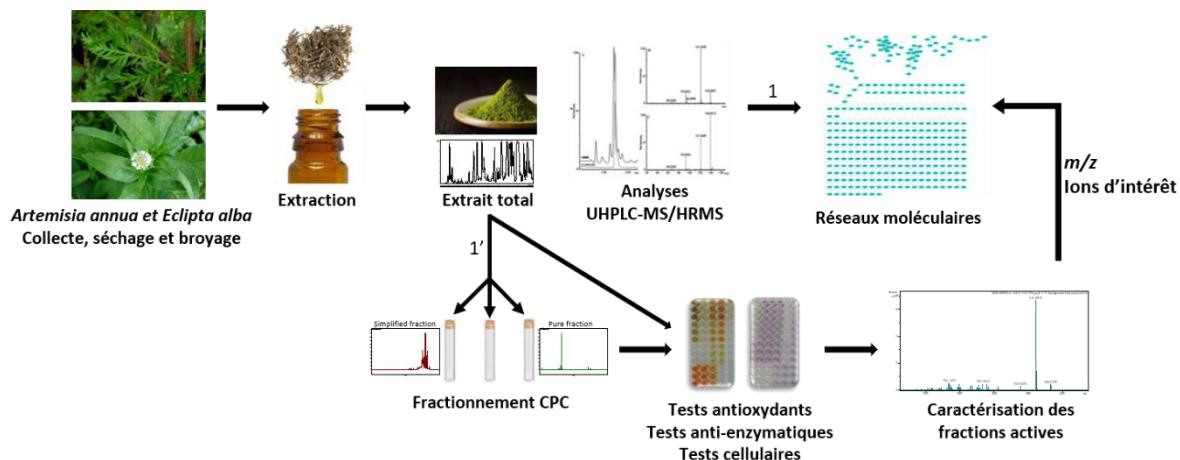


Figure 56 : Résumé de la démarche développée pour cibler et caractériser les molécules bioactives d'*Artemisia annua* et d'*Eclipta alba*

#### A. *Artemisia annua* L.

##### 1) Réseau moléculaire de l'extrait brut

Le BPC (Base Peak Chromatogram) de l'extrait éthanolique d'*Artemisia annua* obtenu par UHPLC-HRMS en mode d'ionisation (+) ESI est présenté dans la Figure 57.

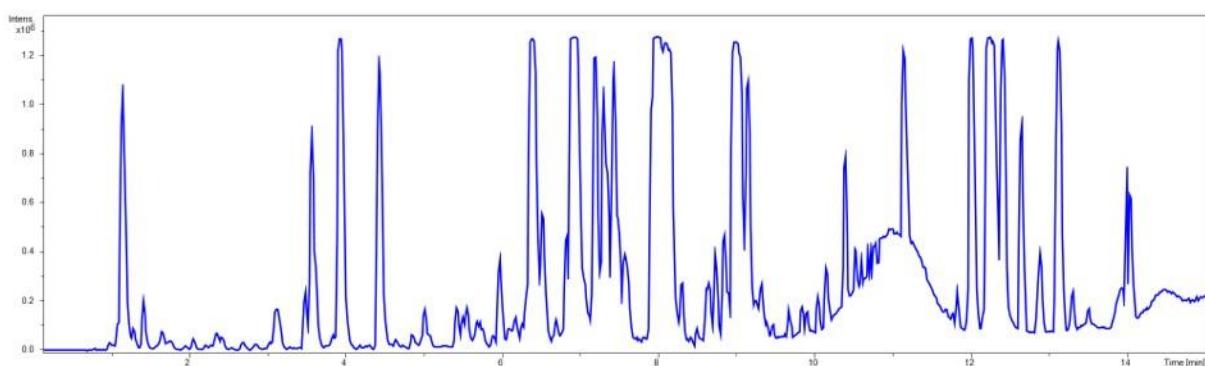


Figure 57 : Le BPC (Base Peak Chromatogram) de l'extrait éthanolique d'*Artemisia annua* obtenu par UHPLC-HRMS en mode d'ionisation (+) ESI

Le chromatogramme acquis en mode d'ionisation positif offre une première empreinte moléculaire de l'échantillon et met en évidence sa complexité. Cette analyse permet la détection d'une centaine de pics chromatographiques avec une large gamme de polarité, puisque les composés sont élusés entre 10 et 100% de solvant organique. Deux énergies de collision ont été appliquées à chaque valeur  $m/z$  allant 20 et 45 eV à  $m/z$  140 jusqu'à 35 et 78 eV à  $m/z$  1000 (énergies de collision proportionnelles au rapport  $m/z$ ). Un spectre MS/MS moyen des deux énergies pour chaque ion parent est calculé permettant ainsi d'obtenir un maximum d'information, c'est-à-dire, une bonne fragmentation des structures rigides et des spectres exploitables pour les structures fragiles.

Avant d'aller plus loin dans l'interprétation des résultats, on peut émettre l'hypothèse au vu des familles moléculaires décrites dans la littérature, que les molécules antioxydantes pourraient être des flavonoïdes et/ou des coumarines<sup>274,275</sup>. Il serait donc possible de rechercher dans l'extrait les composés phénoliques décrits dans la littérature et qui pourraient être présumés responsables de l'activité antioxydante en identifiant chaque pic du chromatogramme de la Figure 57, puisque pour chacun, les spectres UV, MS et MS/MS ont été acquis et pourraient permettre une proposition de structure. Toutefois, pour éviter cette identification systématique, l'approche par réseau moléculaire a été utilisée pour regrouper les molécules en fonction de leurs similitudes spectrales.

Le prétraitement des données spectrales d'*Artemisia annua* effectué avant d'exporter les données spectrales sur la plateforme GNPS a été réalisé avec l'algorithme AutoMS(n) du logiciel DataAnalysis 4.4. La fenêtre de temps de rétention était de 0,3 min, le seuil d'intensité TIC (AllMSn) a été fixé à 50 pour le mode positif (+) ESI et le nombre maximum de composés limité à 7500. Pour chaque spectre, les 10 ions non déconvolus les plus abondants avec un seuil d'intensité de 100 ont été exportés. La *peaklist* a été exportée au format .mgf qui contient toutes les informations sur l'analyse MS/MS: les ions précurseurs ( $m/z$ ) avec leur temps de rétention et les ions fragments ( $m/z$ , intensités, état de charge).

Après l'export des données, le réseau moléculaire a été généré sur GNPS, deux spectres sont liés lorsqu'ils présentent un score cosine  $\geq 0.65$  avec 3 ions et/ou pertes de neutres identiques avec une tolérance de 0,02 Da pour les valeurs  $m/z$ . Le nombre maximum de voisin d'un composé est de 99 (chaque cluster peut avoir une taille maximale de 100 données spectrales) et enfin, l'option MSClustering de GNPS a été désactivée puisque les données ont été retraitées avant leur exportations vers GNPS. Pour l'identification de composés, la recherche dans les bases de données GNPS a été activée, avec des consignes plus restreintes soit un score cosine  $\geq 0.75$  et un nombre minimum de 4 ions (ion fragment et/ou perte neutre) correspondants entre les spectres expérimentaux et ceux des bases de données pour une confiance plus élevée dans les structures proposées.

Le réseau moléculaire de l'extrait brut *Artemisia annua* visualisé à l'aide du logiciel Cytoscape 3.6.1 est présenté sur la Figure 58. Ce réseau moléculaire contient 2 grands et 8 petits clusters qui peuvent être reliés à des familles ou des groupes moléculaires en annotant l'étalon (artémisinine) et certains composés représentatifs des différentes familles moléculaires déjà décrites dans la littérature pour des extraits éthanologiques des parties aériennes d'*Artemisia annua*<sup>276</sup>. Les structures des composés annotés présentés dans le Tableau 6 ont également été reconnues et confirmées par les bibliothèques spectrales de GNPS (par exemple. Massbank : FIO00415 Artémisinine).

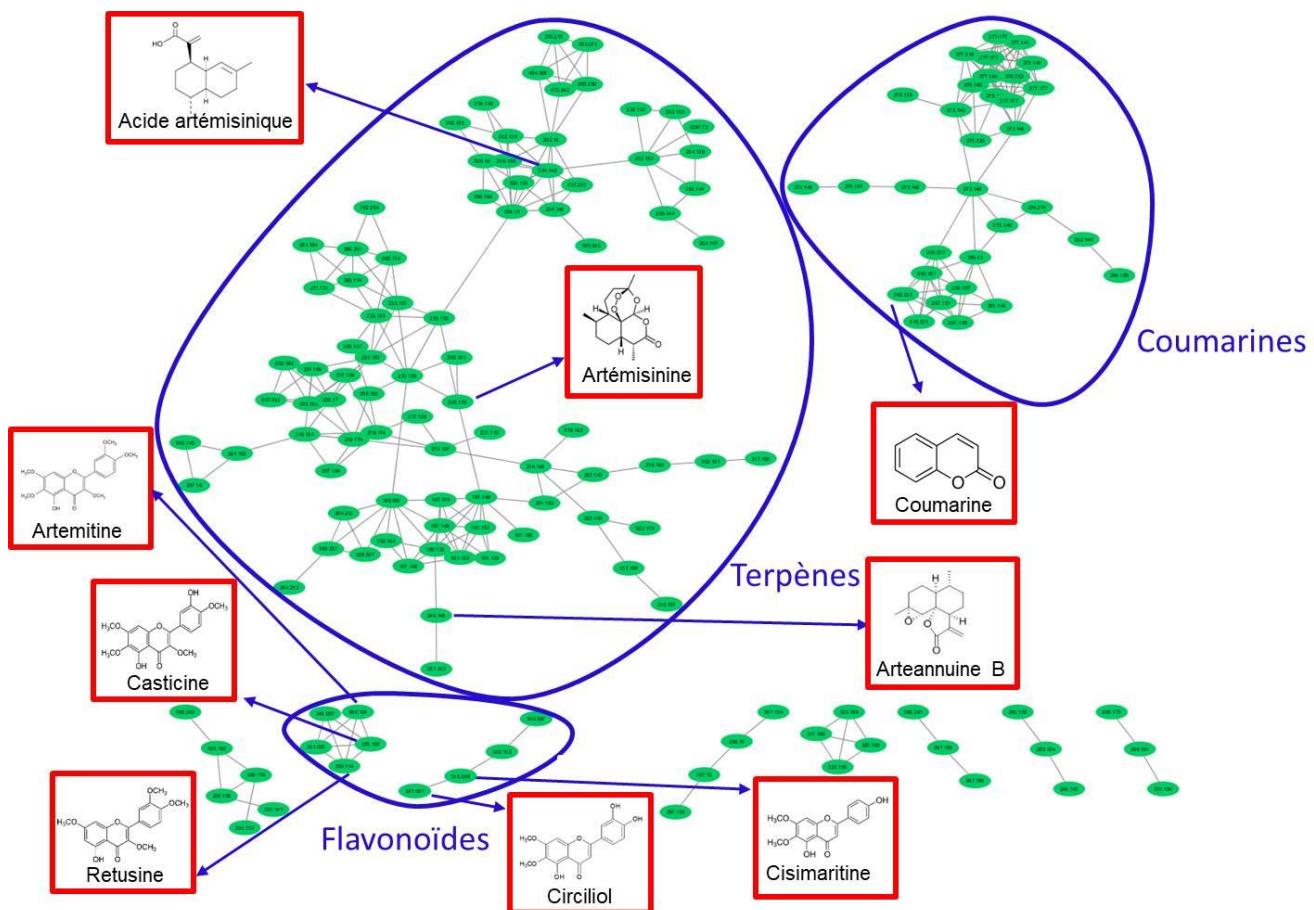


Figure 58 : Le réseau moléculaire de l'extrait brut *Artemisia annua* obtenu à l'aide de la plateforme GNPS et visualisé avec le logiciel Cytoscape 3.5.1.

| <b><i>m/z</i> [M+H]<sup>+</sup><br/>mesuré</b> | <b>Formule<br/>brute [M]</b>                   | <b><i>m/z</i> [M+H]<sup>+</sup><br/>calculé</b> | <b>Erreur<br/>(ppm)</b> | <b>Annotation</b>      | <b>Références</b> | <b>Score<br/>cosine</b> |
|--|--|---|-------------------------|------------------------|-------------------|-------------------------|
| 283,1540                                       | C <sub>15</sub> H <sub>22</sub> O <sub>5</sub> | 283,1540  | 0,1                     | Artémisinine           | Étalon            | 0,73                    |
| 147,0439                                       | C <sub>9</sub> H <sub>6</sub> O <sub>2</sub>   | 147,0441  | 1,2                     | Coumarine              | [206]             | -                       |
| 331,0813                                       | C <sub>17</sub> H <sub>14</sub> O <sub>7</sub> | 331,0812  | 0,4                     | Circiliol              | [206]             | -                       |
| 249,1485                                       | C <sub>15</sub> H <sub>21</sub> O <sub>3</sub> | 249,1485  | 0,0                     | Arteannuine B          | [206]             | -                       |
| 375,1074                                       | C <sub>19</sub> H <sub>18</sub> O <sub>8</sub> | 375,1074  | 0,1                     | Casticin               | [206]             | 0,80                    |
| 315,0865                                       | C <sub>17</sub> H <sub>14</sub> O <sub>6</sub> | 315,0863  | 0,5                     | Cisimaritine           | [206]             | -                       |
| 235,1699                                       | C <sub>15</sub> H <sub>22</sub> O <sub>2</sub> | 235,1693  | 2,7                     | Acide<br>artémisinique | [206]             | 0,77                    |
| 389,1236                                       | C <sub>20</sub> H <sub>21</sub> O <sub>8</sub> | 389,1231  | 0,5                     | Artemetine             | [206]             | -                       |
| 359,1126                                       | C <sub>19</sub> H <sub>18</sub> O <sub>7</sub> | 359,1125  | 0,3                     | Retusine               | [206]             | -                       |
| 361,0915                                       | C <sub>18</sub> H <sub>16</sub> O <sub>8</sub> | 361,0915  | 0,1                     | Chrysosplenol D        | [289]             |                         |

Tableau 6 : Composés annotés sur le réseau moléculaire de l'extrait éthanolique d'*Artemisia annua*

Ces résultats mettent en évidence la présence de 3 grandes familles moléculaires dans l'extrait d'*Artemisia annua* : les terpènes, les coumarines et les flavonoïdes. De petits groupes ne contenant que quelques nœuds n'ont pas été identifiés. Conformément à la richesse bien connue d'*Artemisia annua*, les terpènes semblent être la plus grande famille avec de nombreux nœuds associés. Les coumarines sont également bien représentées dans cet extrait éthanolique avec un groupe assez important, tandis que les flavonoïdes sont moins nombreux.

Néanmoins, plusieurs composés, notamment les terpènes, ont été fragmentés en source, ces ions fragments ont été fragmentés par la suite en MS/MS ce qui a généré de la redondance d'information. Puisque les ions fragments générés en source ont des spectres MS/MS très similaires au spectre MS/MS de l'ion parent, par conséquent, ils ont été liés en cluster dans le réseau moléculaire avec un haut degré de similitude.

Les flavonoïdes quant à eux, n'ont pas été fragmentés en source, c'est la raison pour laquelle les clusters de flavonoïdes contiennent des nœuds qui représentent uniquement les ions moléculaires sans redondance d'information. De plus, les flavonoïdes ont été séparés en deux clusters sur le réseau moléculaire le premier regroupant les structures de type flavonol et le second regroupe les structures de type flavone.

Les réseaux moléculaires ont permis l'identification rapide des principales familles moléculaires et de leurs molécules contenues dans l'extrait d'*Artemisia annua* en regroupant les composés selon leurs structures sans interpréter tous les spectres MS et sans annoter l'ensemble du réseau mais seulement les principaux composés déjà décrits dans la littérature.

Toutefois, afin de cibler les composés antioxydants parmi toutes ces molécules, l'extrait a été simplifié en plusieurs fractions à l'aide de la CPC.

## 2) Activité antioxydante

Une évaluation complète de l'activité antioxydante de l'extrait brut éthanolique et des 13 fractions CPC évaporées et solubilisées dans l'éthanol a été réalisée et les résultats sont présentés dans la Figure 59. L'absorbance mesurée à la longueur d'onde adaptée à chaque test a été tracée en fonction de la concentration de l'échantillon (trolox, extrait brut, fraction).

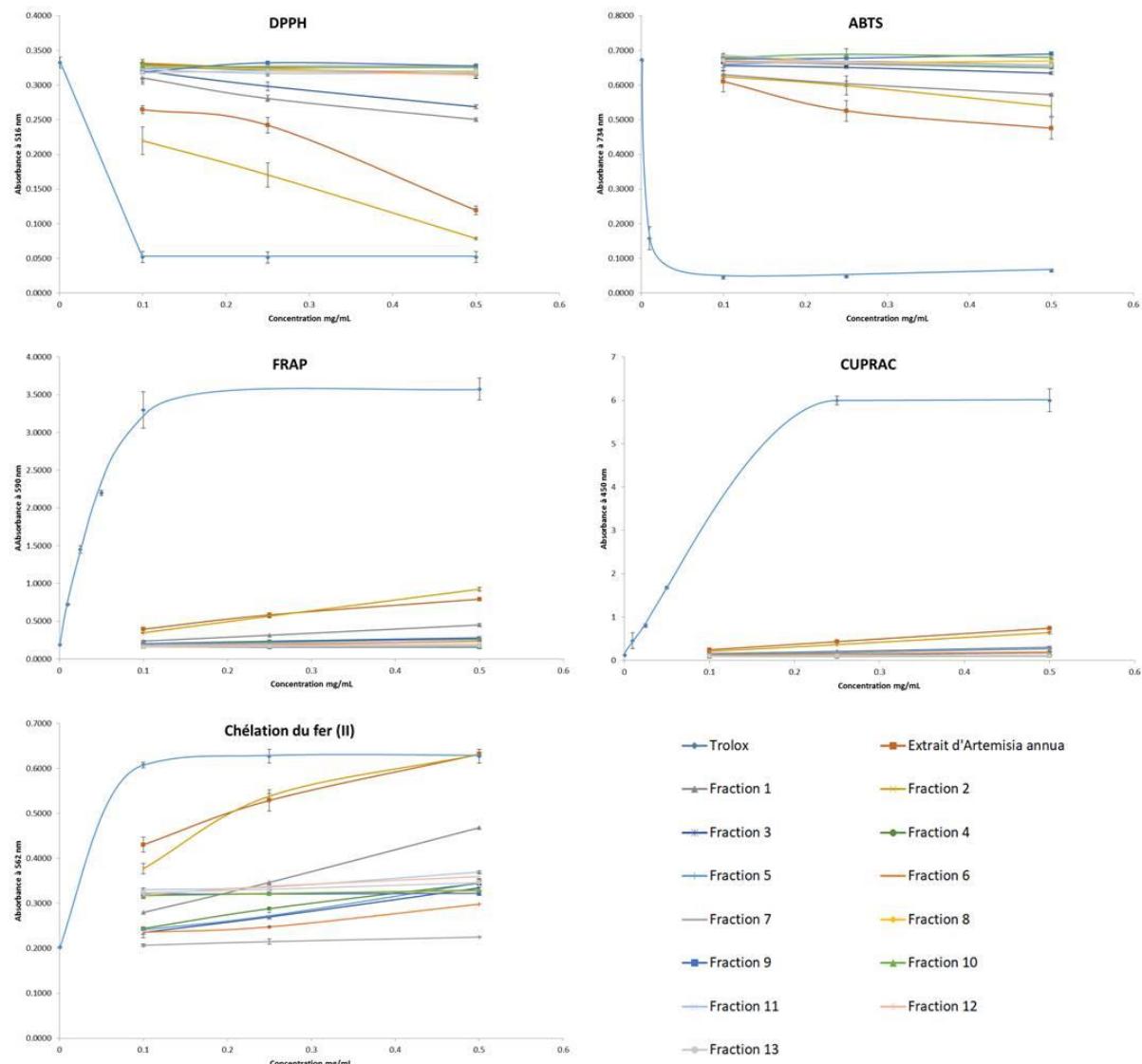


Figure 59 : Résultats de l'activité antioxydante de l'extrait brut éthanolique et des fractions CPC. L'absorbance moyenne a été tracée  $\pm$  l'écart-type ( $n = 3$ )

La capacité antioxydante de toutes les fractions CPC d'*Artemisia annua* a été comparée au témoin positif et à l'extrait brut pour cibler les fractions actives. Les 5 graphiques de la Figure 59 montrent le même ordre d'activité, ce qui indique que les 5 tests classent les

fractions dans le même ordre. L'extrait brut et la Fraction 2 ont invariablement donné les meilleurs résultats, avec l'activité antioxydante la plus élevée. La Fraction 1 a montré une légère activité mais beaucoup plus faible que celle de la Fraction 2, tandis qu'aucune activité n'a été observée pour les autres fractions. Les tests de DPPH et de chélation du fer (II) ont donné les meilleurs résultats avec des valeurs d'absorbance proches de celles du trolox pour la Fraction 2 à 0,5 mg/mL. Les activités mesurées avec les tests ABTS, FRAP et CUPRAC sont plus faibles, avec des valeurs éloignées du trolox mais toujours assez proches entre extrait brut et Fraction 2.

Des différences de niveau d'activité entre les tests chimiques basés sur le mécanisme de transfert des atomes d'hydrogène tels que le DPPH et l'ABTS ont déjà été observées<sup>265,266</sup>. En effet, les différentes structures stéréochimiques des radicaux ABTS<sup>•+</sup> et DPPH<sup>•</sup> ainsi que la solubilité de l'extrait, la polarité et la capacité de chélation des métaux des composés antioxydants peuvent engendrer des divergences entre les deux tests<sup>265,266</sup>. De plus, le DPPH<sup>•</sup> est un radical libre qui accepte les électrons ou les radicaux d'hydrogène des composés donneurs<sup>277</sup>, tandis que le test ABTS détermine la capacité antioxydante en réagissant avec un agent oxydant puissant (persulfate de potassium) en présence d'un cation radicalaire ABTS<sup>•+</sup><sup>278,279</sup>.

La capacité de réduction de l'extrait d'*Artemisia annua* et des fractions est faible par rapport à celle de l'étoile positif (trolox), que ce soit pour la réduction du cuivre ou du fer ce qui indique que les composés n'agissent pas ou peu par transfert d'électron. Les valeurs élevées obtenues pour l'extrait brut et la Fraction 2 pour le test de chélation du fer (II) suggèrent la capacité des composés à former facilement un complexe métallique qui pourrait être dû à une structure spécifique comme celle des flavonoïdes<sup>280</sup>.

Le Tableau 7 présente le pourcentage d'activité calculé pour la concentration testée la plus élevée 0,5 mg/mL, pour l'extrait brut et les 13 fractions. Une analyse ANOVA a été effectuée sur ces résultats et a montré que pour tous les tests, l'extrait brut, la Fraction 2 et la Fraction 1 étaient significativement différents des autres fractions, confirmant la tendance observée dans la Figure 59.

| Échantillons                  | DPPH<br>% inhibition | ABTS<br>% inhibition | FRAP<br>% inhibition | CUPRAC<br>% inhibition | Chélation du fer<br>% inhibition |
|-------------------------------|----------------------|----------------------|----------------------|------------------------|----------------------------------|
| <b><i>Artemisia annua</i></b> | 64,13±0,8*           | 29,9±0,6*            | 22,2±0,7*            | 12,35±0,2*             | 100,7±0,8**                      |
| <b>Fraction 1</b>             | 24,8±0,5*            | 15,6±0,3*            | 12,6±0,5*            | 5,0±0,3*               | 74,7±0,6*                        |
| <b>Fraction 2</b>             | 76,3±0,8*            | 20,5±0,4*            | 25,9±0,7*            | 10,6±0,3*              | 100,7±0,8**                      |
| <b>Fraction 3</b>             | 19,3±0,3*            | 6,5±0,1              | 7,8±0,3              | 2,9±0,1                | 53,3±0,6                         |
| <b>Fraction 4</b>             | 5,5±0,1              | 4,2±0,4              | 7,2±0,1              | 3,2±0,1                | 55,0±0,3                         |
| <b>Fraction 5</b>             | 5,3±0,3              | 3,0±0,6              | 7,5±0,1              | 4,4±0,2                | 55,1±0,5                         |
| <b>Fraction 6</b>             | 5,3±0,1              | 3,1±0,4              | 6,5±0,1              | 2,4±0,1                | 47,5±0,2                         |
| <b>Fraction 7</b>             | 1,7±0,1              | 3,3±0,2              | 5,0±0,1              | 1,6±0,05               | 35,9±0,2                         |
| <b>Fraction 8</b>             | 4,2±0,2              | 1,4±0,3              | 4,6±0,2              | 1,6±0,05               | 52,2±0,8                         |
| <b>Fraction 9</b>             | 1,6±0,1              | 1,6±0,3              | 4,4±0,1              | 1,6±0,05               | 51,4±0,8                         |
| <b>Fraction 10</b>            | 2,3±0,1              | 0,2±0,05             | 4,6±0,1              | 1,6±0,1                | 52,52±0,4                        |
| <b>Fraction 11</b>            | 4,6±0,1              | 3,9±0,1              | 5,4±0,2              | 2,0±0,05               | 58,92±0,8                        |
| <b>Fraction 12</b>            | 5,5±0,2              | 3,0±0,1              | 5,2±0,1              | 1,7±0,05               | 57,3±0,6                         |
| <b>Fraction 13</b>            | 3,8±0,3              | 1,8±0,1              | 4,9±0,1              | 1,6±0,05               | 55,2±0,7                         |

\* Indique une différence significative (ANOVA p-valeur < 0,05).

\*\* Indique une différence significative entre les deux mais pas entre eux.

*Tableau 7 : Activité antioxydante de l'extrait brut *Artemisia annua* et ses 13 fractions CPC à 0,5 mg/mL, évaluée par les tests DPPH, ABTS, FRAP, CUPRAC et chélation du fer (II)*

La combinaison de ces résultats indique que les composés antioxydants de l'extrait *Artemisia annua* sont principalement élusés dans la Fraction 2 pendant le fractionnement CPC. Cette méthode a permis de mettre en évidence la fraction la plus active parmi les 13 de départ. Afin d'identifier les composés antioxydants d'*Artemisia annua*, les fractions ont été analysées par spectrométrie de masse.

### 3) Identification des molécules bioactives

Les fractions CPC ont été rapidement analysées par FIA-HRMS (injection de 0,2 µL de l'échantillon directement dans le spectromètre sans séparation sur la colonne). Les principaux ions obtenus sur le spectre de masse ont pu être reportés sur le réseau moléculaire de l'extrait brut afin d'identifier les molécules contenues dans la fraction F2 active.

L'analyse de la Fraction 2 a montré un ion principal à *m/z* 361,0915 et son adduit sodium [M+Na]<sup>+</sup> à *m/z* 383,0728 (Figure 60 A et C). Bien que l'objectif initial de la CPC n'était pas de purifier les molécules, il s'est avéré que cette fraction semblait contenir un composé principal. L'analyse UHPLC-DAD de la Fraction 2 (Figure 60 B) a confirmé le résultat en détectant un pic principal à 6,3 min. La formule brute de l'ion d'intérêt *m/z* 361,0915 a ensuite été calculée à l'aide de l'algorithme SmartFormula du logiciel d'analyse de données (DataAnalysis) et a

donné deux propositions :  $C_{18}H_{17}O_8^+$  (0,1 ppm) et  $C_{19}H_{13}N_4O_4^+$  (3 ppm). La seconde formule a été rejetée car son profil isotopique théorique est très différent de celui de l'expérience. La formule brute  $C_{18}H_{16}O_8$  a été soumise à des bases de données publiques pour la recherche de structure PubChem, ChemSpider, ChEBI, DNP et 179 composés ont été proposés comme structures potentielles correspondant à cette formule brute.

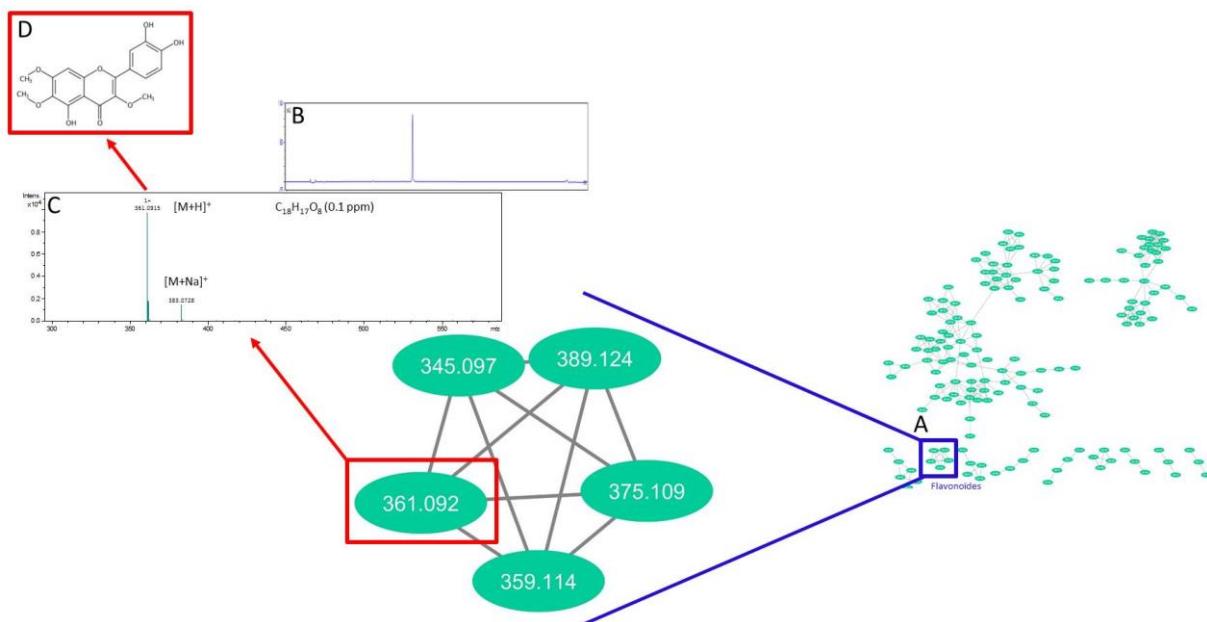


Figure 60 : (A) Focalisation sur le cluster de l'ion d'intérêt contenu dans la Fraction 2. (B) Le chromatogramme UV de la Fraction 2 à  $\lambda = 280\text{ nm}$ , la séparation a été effectuée sur une colonne Fortis C18 ( $150 \times 2,1\text{ mm} ; 1,7\text{ }\mu\text{m}$ ) à  $60^\circ\text{C}$ . (C) Spectre de masse de l'ion d'intérêt. (D) Structure de chrysosplénol D

Afin d'affiner l'identification du composé et la sélection de la structure parmi toutes les propositions, l'ion d'intérêt a été localisé sur le réseau moléculaire. Le composé actif a été trouvé dans le cluster des flavonoïdes comme le montre la Figure 60. Son spectre MS/MS en Figure 61 a montré des ions fragments caractéristiques des flavonoïdes méthylés ( $m/z$  346,0674  $C_{17}H_{14}O_8^+$  due à la perte de  $CH_3$ ,  $m/z$  328,0577  $C_{17}H_{12}O_7^+$  due à la perte de  $CH_3$  et de  $H_2O$ ,  $m/z$  303,0497  $C_{15}H_{11}O_7^+$  due à la perte de  $CH_3$  et de  $C_2H_3O^+$ ,  $m/z$  285,0391  $C_{15}H_9O_6^+$  due à la perte de  $H_2O$  à partir de  $m/z$  303,0497,  $m/z$  257,0443  $C_{14}H_9O_5^+$  due à la perte de CO à partir du  $m/z$  285,0391,  $m/z$  161,0233  $C_9H_5O_3^+ ({}^{0,2}A^+)$  et  $137,0231 C_7H_5O_3^+ ({}^{0,2}B^+)$  obtenus par le clivage rétro Diels-Alder (Figure 62).

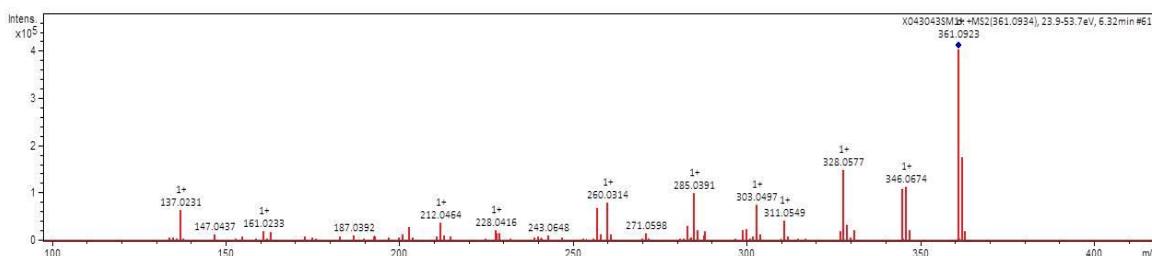


Figure 61 : Spectre MS/MS de l'ion d'intérêt contenu dans la fraction active (Fraction 2)

Sur le spectre MS/MS de la Figure 61, on observe des ions (impair) et des ions radicaux (pairs) générés par les ruptures hétérolytique et homolytique respectivement.

Ce spectre MS/MS a également été soumis aux bibliothèques du GNPS et le 2-(3,4-dihydroxyphényle)-5-hydroxy-3,6,7-triméthoxy-4H-1-benzopyran-4-one

(CCMSLIB00000848806) également appelé chrysosplénol D a été trouvé avec le numéro de registre CAS 14965-20-9 dans SciFinder. La Figure 62 présente une proposition de fragmentation du chrysosplénol D.

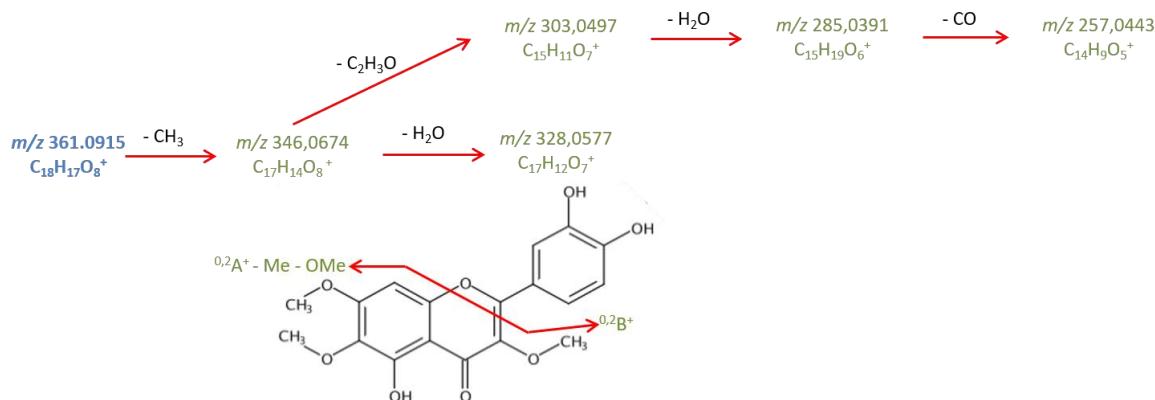


Figure 62 : Proposition de fragmentation du chrysosplénol D

Le niveau de confiance d'identification est fixé à 2 "Composés supposés annotés (par exemple, sans étalons de référence chimiques, basés sur des propriétés physico-chimiques et/ou une similarité spectrale avec les bibliothèques spectrales publiques/commerciales". Néanmoins, Wang *et al.* ont récemment étudié les principaux flavonoïdes d'*Artemisia Annua* et ont rapporté le spectre MS/MS du standard chrysosplénol D, qui est conforme à nos résultats (Figure 63)<sup>281</sup>. Ce qui donne un argument supplémentaire pour confirmer notre hypothèse d'identification.

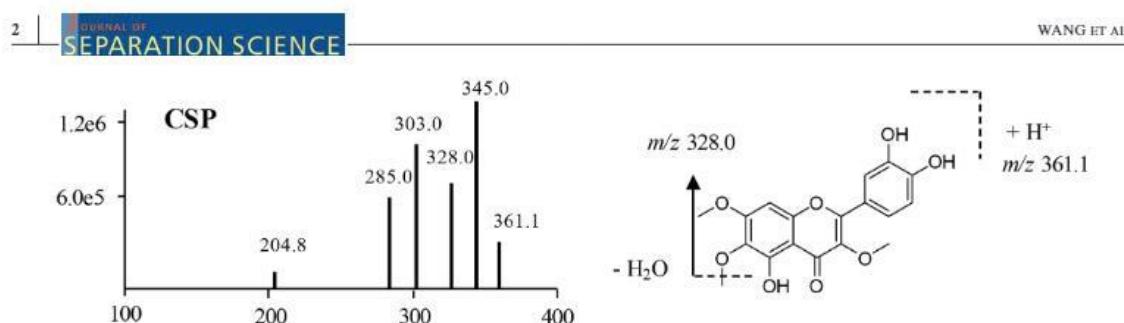


Figure 63 : Spectre MS/MS du standard chrysosplénol D présenté par Wang et al (2019)<sup>281</sup>

L'ion d'intérêt a été rapidement obtenu par analyse FIA-HRMS de la fraction active. Grâce à sa position dans le réseau moléculaire, la famille du composé a été déterminée et son identification facilitée. Cette molécule a déjà été décrite dans la littérature comme étant présente dans *Artemisia annua*<sup>76, 281</sup>, mais elle n'a jamais été corrélée à la capacité antioxydante de cette plante. Ce composé a également été détecté dans la Fraction 3, peu active, mais à une concentration bien inférieure à celle de la Fraction 2.

En se focalisant sur ce cluster (Figure 60 A), on peut remarquer que quatre autres molécules de type flavonoïde méthylé sont présentes dans l'extrait et pourraient avoir une activité antioxydante étant donné leur similarité de structure. Ainsi, ces composés ont également été recherchés dans les fractions CPC (Figure 64). La molécule présentant un  $[M+H]^+$  avec  $m/z$  345,0897, pourrait correspondre à la penduletine, elle a été observée dans la Fraction 1, qui est modérément active. Cette molécule ne possède qu'un seul groupe hydroxyle libre sur le cycle B, ce qui pourrait expliquer sa plus faible activité antioxydante. Les autres composés du cluster, correspondant probablement à la casticine  $m/z$  375,1074, la rétusine  $m/z$  359,1125 et l'artémisitine  $m/z$  389,1231, ont été trouvés dans d'autres fractions qui ne présentaient pas d'activité antioxydante.

Cela peut s'expliquer d'une part par leur présence en faible concentration dans les fractions, et d'autre part par leurs structures. En effet, malgré leur similitude avec le composé actif, la rétusine et l'artémisitine ne possèdent que des groupes OMe sur le cycle B, par conséquent, elles possèdent moins de H labiles nécessaires à l'activité antioxydante.

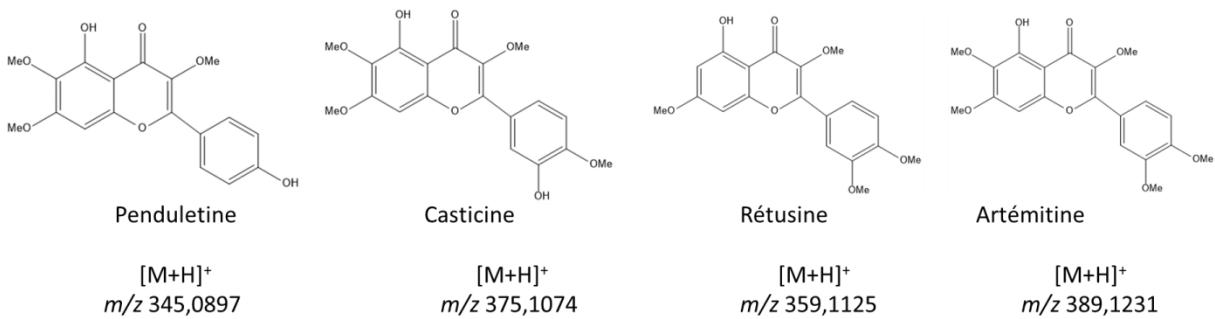


Figure 64 : Flavonoïdes de l'extrait éthanolique d'*Artemisia annua*

#### 4) Viabilité cellulaire

Le test d'innocuité cellulaire XTT effectué à différentes concentrations (de 31,25 à 500 µg/mL dans les puits de mesure) d'extrait éthanolique d'*Artemisia annua* a montré que la concentration maximale n'induisant pas de toxicité cellulaire est de 62,5 µg/mL d'extrait. Par conséquent, toutes les concentrations évaluées en tests d'activité antioxydante sont bien tolérées par les cellules (concentration maximale de 25 µg/mL dans les puits pour une concentration mère de 0,5 mg/mL).

En conclusion, grâce à la combinaison des approches de réseau moléculaire et fractionnement CPC, le flavonoïde chrysosplénol D de formule brute  $C_{18}H_{16}O_8$  a été ciblé et identifié, parmi les centaines de composés de l'extrait, et identifié comme la principale molécule contribuant à l'activité antioxydante de l'extrait d'*Artemisia annua*. Cet extrait peut être incorporé dans une formule cosmétique à une concentration de 62,5 µg/mL pour un effet antioxydant.

## B. *Eclipta alba* L.

### 1) Réseau moléculaire de l'extrait brut

Le BPC (*Base Peak Chromatogram*) de l'extrait éthanolique d'*Eclipta alba* obtenu en UHPLC-HRMS est présenté en Figure 65. Celui-ci représente une première empreinte moléculaire de l'extrait brut et montre une cinquantaine de pics majoritaires des composés détectés en (+) ESI.

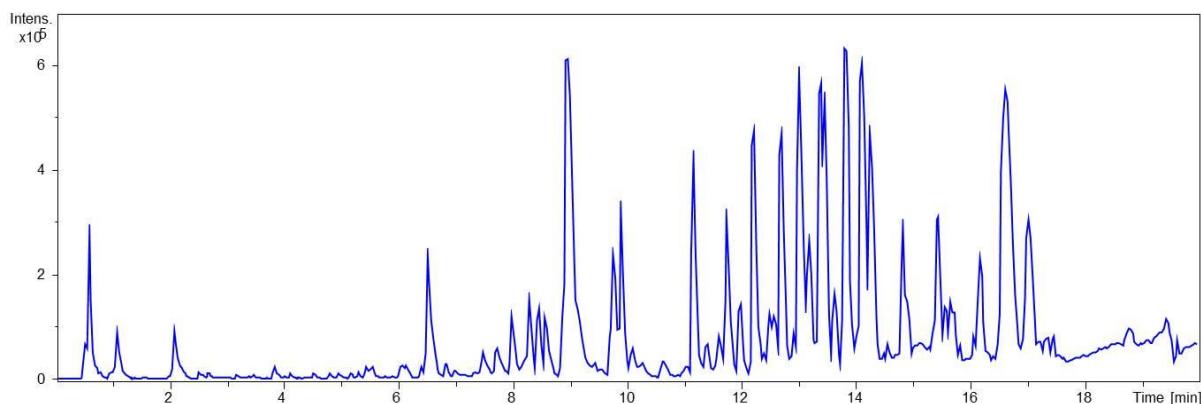


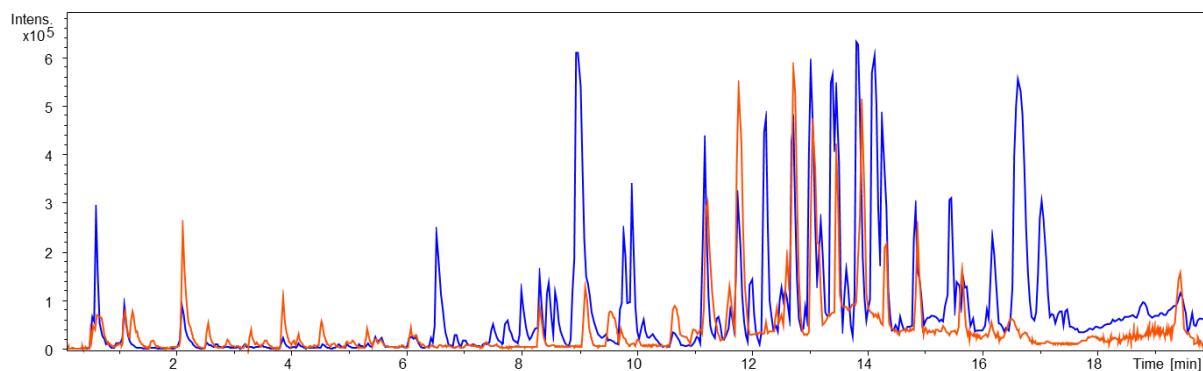
Figure 65 : Le BPC (Base Peak Chromatogram) de l'extrait éthanolique d'*Eclipta alba* obtenu en UHPLC-HRMS en mode d'ionisation (+) ESI

Une méthode chromatographique a été développée et optimisée pour éluer et séparer au mieux les composés contenus dans l'extrait brut. En effet, une première méthode a été développée sur une colonne Fortis C18 de type FPP nécessitant une élution avec un gradient riche en isopropanol, car la majorité des composés d'*Eclipta alba* sont apolaires, générant une forte contre pression. Par conséquent, des colonnes de type SPP (*Fused Core*) ont été testées pour réduire la contre pression du fait de leurs noyaux solides.

Une colonne C18 Kinetex (Phenomenex) a donc d'abord été testée et a permis de réduire la contre pression, néanmoins, un gradient riche en isopropanol était également nécessaire pour éluer les composés apolaires générant des pics interférant en (+) ESI de par sa forte élution. Pour pallier à cela, une colonne C18 peu rétentive Aeris *widapore* (Phenomenex), également avec une technologie "*Fused Core*", a été utilisée afin d'éluer les composés apolaires avec l'acétonitrile. C'est une colonne plus adaptée à la séparation des protéines que des petites molécules, avec un greffon permettant d'éluer des composés apolaires tout en gardant une faible pression à 40°C.

La richesse en composés apolaires peut être observée en Figure 65 sur le chromatogramme de 8 à 18 min, par rapport aux composés polaires et moyennement polaires qui sont élués entre 0,5 et 7 min.

L'extrait a ensuite été analysé en (-) ESI (Figure 66), afin de mieux ioniser les composés polaires à moyennement polaires élués entre 0,5 et 7 min représentés par des pics très peu intenses sur le chromatogramme obtenu en (+) ESI. Il permet ainsi de détecter des composés non ionisés en mode positif tel que le pic à 1,4 min, et d'aider l'identification des composés ionisés dans les deux modes.



*Figure 66 : Superposition des BPCs (Base Peak Chromatograms) d'extrait éthanolique d'*Eclipta alba* obtenu par UHPLC-HRMS en utilisant les modes d'ionisation (+) ESI en bleu, et (-) ESI en orange*

La complémentarité entre les modes positif et négatif peut être observée sur la Figure 66 qui présente une superposition des chromatogrammes de l'extrait éthanolique d'*Eclipta alba* en modes (+) ESI et (-) ESI. En effet, les composés élués entre 0,5 et 7 min sont mieux ionisés en mode négatif contrairement à ceux élués plus tard qui sont davantage ionisés en mode positif. Certains composés sont détectés dans les deux modes, et d'autres dans un seul mode confirmant ainsi leur complémentarité. Néanmoins, pour la création des réseaux moléculaires, le mode (+) ESI a été utilisé car il permet de détecter davantage de composés et il offre des meilleurs spectres MS/MS. En effet, 3 précurseurs sont sélectionnés pour chaque spectre MS et fragmentés en MS/MS à deux énergies de collision, 20 et 60 eV. Les fragments résultants des deux énergies sont groupés dans un seul spectre MS/MS pour chaque ion précurseur sélectionné. L'application de ces deux énergies extrêmes a permis d'obtenir un maximum d'information sur les ions parents puisque les structures rigides ont été fragmentées à 60 eV et les molécules plus fragiles ont été fragmentées à 20 eV. Ces énergies de collision ont offert un grand nombre de spectres MS/MS exploitables avec des informations sur les schémas de fragmentation, néanmoins, quelques composés très rigides n'ont pas été fragmentés et à l'inverse, d'autres très fragiles ont donné des spectres MS/MS non exploitables.

Un prétraitement des données a ensuite été effectué avant leur exportation sur la plateforme GNPS dans l'objectif de construire le réseau moléculaire de l'extrait éthanolique d'*Eclipta alba*. Pour ce faire, le logiciel MetaboScape de Bruker a été utilisé.

Les paramètres ont été optimisés sur l'analyse UHPLC-HRMS de l'extrait d'*Eclipta alba*. Tous les pics chromatographiques ayant une intensité supérieure à 2000 ua avec un nombre minimum de 3 spectres MS ont été sélectionnés pour la création de la *peaklist* afin de récupérer le maximum d'information tout en limitant le bruit de fond. L'aire a été choisie pour la quantification des caractéristiques spectrales et l'onglet de calibration a été désactivé puisque cette étape a été appliquée sur l'analyse directement après l'acquisition (DataAnalysis).

Lorsque MetaboScape trouve plusieurs spectres MS/MS d'un même ion précurseur, il les moyenne pour générer un seul spectre. En outre, pendant l'acquisition, si des spectres MS/MS avec des énergies de collision multiples ont été stockés séparément, ils sont additionnés au cours de l'importation des spectres et le meilleur est importé. Les caractéristiques spectrales ont été extraites dans un intervalle de temps de 0-17,5 min et de *m/z* 50-1200. Ensuite, MetaboScape regroupe les caractéristiques  $[M+H]^+$ ,  $[M+Na]^+$ ,  $[M+K]^+$ ,  $[M-H_2O+H]^+$  d'un même composé ce qui réduit la redondance d'information. Enfin, la table créée, appelée *bucket*, contenant les caractéristiques *m/z* des composés, leurs spectres MS/MS, temps de rétention, aires de pic, propositions de formules brutes, est directement exportée vers GNPS (via la plateforme de transfert de données WinSCP).

Une fois les données exportées sur GNPS, le réseau moléculaire est généré en liant deux spectres lorsqu'ils présentent un score cosine  $\geq 0,65$  avec 3 ions et/ou pertes de neutres identiques avec une tolérance de 0,01 Da pour les valeurs *m/z*. Le nombre maximum de voisin d'un composé est de 99 en désactivant l'outil MSClustering de GNPS puisque les données ont été retraitées avec MetaboScape avant l'export vers GNPS. La recherche dans les bases de données GNPS a été activée, avec des consignes plus restreintes soit un score cosine  $\geq 0,75$  et un nombre minimum de 4 ions qui concordent.

La composition phytochimique de l'extrait d'*Eclipta alba* est présentée sous forme de réseau moléculaire sur la Figure 67. Le réseau moléculaire global (A) est divisé en deux parties ; une partie entourée en bleu comportant des clusters de groupes moléculaires (composés ayant des structures similaires) et une partie entourée en orange avec des singletons (nœuds solitaires).

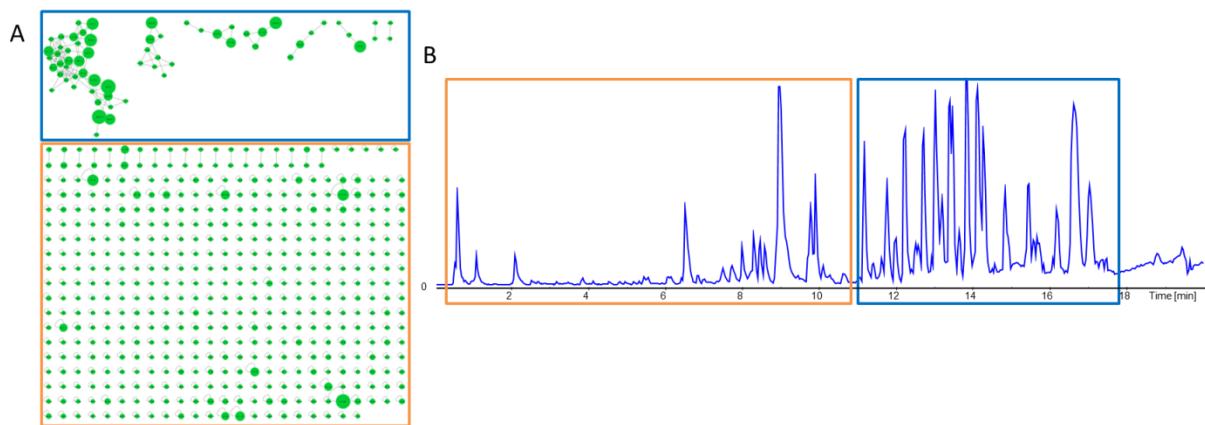


Figure 67 : A) Réseau moléculaire visualisé sur Cytoscape et B) BPC (Base Peak Chromatogram) de l'extrait éthanolique d'*Eclipta alba* obtenu par UHPLC-HRMS en mode (+) ESI

Le réseau moléculaire montre les composés sous forme de nœuds avec des tailles proportionnelles à leurs intensités (aires des pics chromatographiques). La comparaison du réseau (A) et du chromatogramme (B) de l'extrait et l'analyse des temps de rétention des nœuds présentés sur le réseau moléculaire permet d'associer les singltons dans le carré orange comportant des nœuds de faibles tailles aux composés polaires à moyennement polaires, puisque ces composés sont peu intenses sur le chromatogramme. Tandis que les clusters regroupant les nœuds dans le carré bleu ayant des tailles largement supérieures sont associés aux composés apolaires qui sortent à partir de 90% de phase organique (acétonitrile) puisqu'ils sont très intenses sur chromatogramme de l'extrait brut.

Ensuite, l'observation des spectres MS/MS des composés dans le carré orange a révélé qu'ils sont bien différents et quelques spectres présentent peu de fragments avec des valeurs  $m/z$  faibles ce qui explique le fait qu'ils soient sous forme de singltons dans le réseau moléculaire. Au contraire, les spectres des composés dans le carré bleu ont plusieurs ions fragments identiques.

Enfin, les données contenues dans les nœuds ont été vérifiées et ont montré un regroupement des caractéristiques  $[M+H]^+$ ,  $[M+Na]^+$ ,  $[M+K]^+$ ,  $[M-H_2O+H]^+$  pour la plupart des composés. Ceci permet de confirmer que le prétraitement de données par le logiciel MetaboScape a bien réduit la redondance d'information. Néanmoins, quelques composés apolaires ont des adduits spécifiques  $[2M+HCOOHNa]^+$ ,  $[M+NH_4]^+$  et n'ont pas été introduits dans la liste des adduits MetaboScape mais ont été repérés sur le réseau.

Toutes ces vérifications confirment que le réseau moléculaire représente correctement les données chromatographiques (temps de rétention et aires des pics) et spectrométriques (spectres MS/MS) et fournit des résultats fiables et cohérents.

L'annotation du réseau moléculaire a ensuite été effectuée par la combinaison des informations de la bibliographie, soit les composés déjà décrits dans la littérature comme présents dans *Eclipta alba*, des bases de données de la plateforme GNPS et de l'élucidation structurale *de novo*. Les résultats sont résumés dans le Tableau 8 comportant des composés numérotés selon l'ordre d'élution et leurs numéros sont reportés aux nœuds qui les représentent sur le réseau moléculaire de la Figure 68.

En effet, l'étude bibliographique de la composition chimique d'*Eclipta alba* a permis l'annotation des composés élués au début du gradient (composés 1→8, 10, 11 et 29) qui sont sous forme de nœuds singletons, et tous les autres composés plus lipophiles n'ont pas été décrits dans la littérature d'*Eclipta alba*. En effet, l'utilisation d'une colonne peu rétentive a permis l'élution et la séparation de composés non observés auparavant lors des études phytochimiques d'*Eclipta alba*.

De plus, cette plante contient majoritairement des composés spécifiques au genre *Eclipta*, telle que la famille des Eclalbasaponines regroupant une dizaine de saponines connues (composés 7, 8, 10, 11). Elle contient très peu de composés enregistrés dans les bases de données de GNPS, par conséquent, la déréplication faite par les bases spectrales n'a révélé qu'une dizaine de molécules. C'est pourquoi, la méthode d'élucidation structurale *de novo* a été adoptée pour l'annotation des autres composés.

Toutefois, comme mentionné précédemment, les clusters observés sur le réseau moléculaire d'*Eclipta alba* (Figure 68 C) représente les composés apolaires ce qui a permis de s'appuyer sur les similarités structurales pour propager l'annotation sur le réseau après identification de quelques composés. Ainsi, l'annotation des autres composés a été effectuée par les réseaux moléculaires et l'élucidation *de novo* sans pour autant identifier les structures de tous les composés.

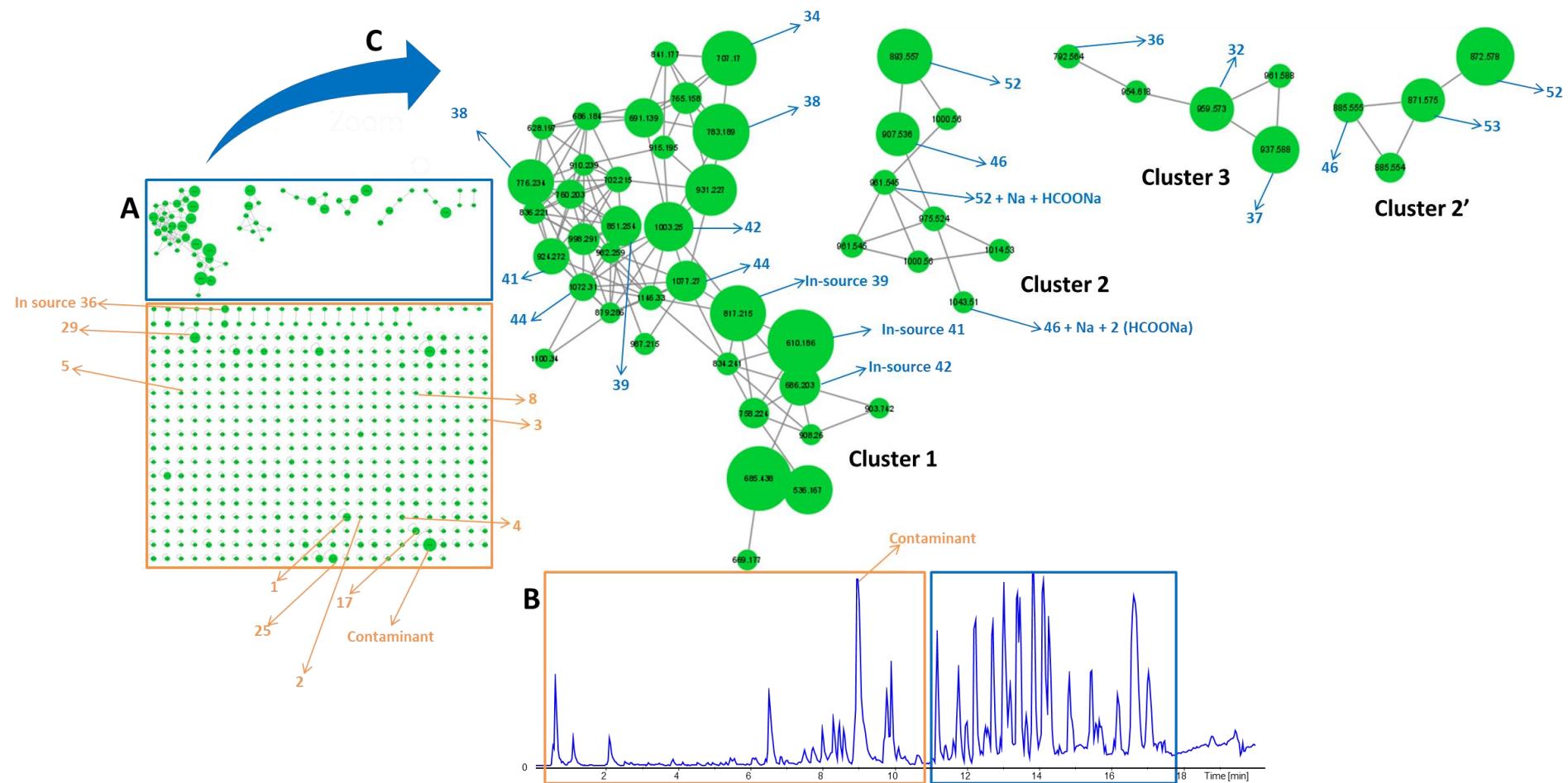


Figure 68 : A) Réseau moléculaire global de l'extrait éthanolique d'*Eclipta alba* visualisé sur Cytoscape. B) BPC (Base Peak Chromatogram) de l'extrait éthanolique d'*Eclipta alba* obtenu par UHPLC-HRMS en mode (+) ESI. C) Focalisation sur les clusters des composés majoritaires de l'extrait éthanolique d'*Eclipta alba*. Les numéros associés aux nœuds sont les composés annotés sur le réseau moléculaire et détaillés dans le Tableau 8

| N° | Rt<br>(min) | <i>m/z</i><br>expérimental<br>[M+H] <sup>+</sup> , [M+Na] <sup>+</sup> | Formule<br>brute [M]  | <i>m/z</i><br>théorique | Erreur<br>(ppm) | Spectre MS/MS<br>(%intensité<br>relative)  | <i>m/z</i><br>expérimental<br>[M-H] <sup>-</sup> , [M+Cl] <sup>-</sup> | Proposition de<br>structure     | Famille<br>moléculaire | Noeud dans<br>le réseau<br>moléculaire |
|----|-------------|--|---|-------------------------|-----------------|--|--|---------------------------------|------------------------|--|
| 1  | 0,64        | 226,154994   | C <sub>11</sub> H <sub>19</sub> N <sub>3</sub> O <sub>2</sub> | 226,155003              | 0,0             | 208,1427 (20),<br>181,2226 (10)  | -  | acide plantago-<br>guanidinique | Alcaloïde              | Singleton                              |
| 2  | 0,67        | 449,107616   | C <sub>21</sub> H <sub>20</sub> O <sub>11</sub>               | 449,107838              | 0,5             | 287,0550 (100)   | 447,093674   | Lutéoline-7-O-<br>glucoside     | Flavonoïde             | Singleton                              |
| 3  | 0,72        | 517,133853   | C <sub>25</sub> H <sub>24</sub> O <sub>12</sub>               | 517,134053              | 0,4             | 449,1236 (5),<br>319,0803 (7,54),<br>163,0394 (100),<br>145,0288 (9,27),<br>135,0442 (11,40),<br>117,0343 (7,70) | 515,1202   | Acide 3,5-<br>dicaféylquinique  | Acide<br>phénolique    | Singleton                              |
| 4  | 1,09        | 315,049867   | C <sub>16</sub> H <sub>10</sub> O <sub>7</sub>                | 315,049929              | 0,2             | 297,0412 (11),<br>287,0548 (36),<br>161,0234 (6)   | 313,035674   | wedelolactone                   | Coumestan              | Singleton                              |
| 5  | 1,23        | 351,016895   | C <sub>15</sub> H <sub>10</sub> O <sub>8</sub> S              | 351,016915              | 0,1             | 271,0610 (100),<br>153,0179 (16)   | 349,001615   | Apigénine sulfate               | Flavonoïde             | Singleton                              |
| 6  | 1,27        | -  | C <sub>16</sub> H <sub>10</sub> O <sub>7</sub>                | -                       | -               | -  | 313,034451   | Wedelolactone<br>isomère        | Coumestan              | -                                      |
| 7  | 1,62        | -  | C <sub>48</sub> H <sub>78</sub> O <sub>19</sub>               | -                       | -               | -  | 957,505846   | Ecalbasaponine<br>III et B      | Saponines              | -                                      |

|    |      |            |   |            |      |   |            |  |            |           |
|----|------|------------|---|------------|------|---|------------|--|------------|-----------|
| 8  | 2,14 | 819,449748 | C <sub>42</sub> H <sub>68</sub> O <sub>14</sub> | 819,450128 | 0,5  | 657,3973 (18),<br>613,4055 (10),<br>595,3957 (5)        | 831,431183 | Eclalbasaponine I<br>et C  | Saponines  | Singleton |
| 9  | 2,56 | 905,449467 | C <sub>45</sub> H <sub>70</sub> O <sub>17</sub> | 905,450521 | 1,2  | 861,465500 (15),<br>743,399796 (29),<br>699,406800 (10) | 881,453727 | -  | Saponine   | Singleton |
| 10 | 3,32 | -          | C <sub>42</sub> H <sub>68</sub> O <sub>14</sub> | -          | -    | -   | 831,430991 | Eclalbasaponine<br>IV  | Saponine   | -         |
| 11 | 3,87 | 657,3970   | C <sub>36</sub> H <sub>58</sub> O <sub>9</sub>  | 657,3973   | 0,5  | -   | 633,401231 | Eclalbasaponine<br>II, A et D  | Saponines  | -         |
| 12 | 4,13 | 721,414431 | C <sub>39</sub> H <sub>60</sub> O <sub>12</sub> | 721,415754 | 1,8  | 703,4032 (12),<br>455,3596 (100),<br>437,3415 (18)      | 719,399967 | Acide oléanique +<br>sucre   | Saponine   | Singleton |
| 13 | 5,08 | -          | C <sub>41</sub> H <sub>62</sub> O <sub>12</sub> | -          | -    | -   | 745,416897 | -  | Saponine   | -         |
| 14 | 5,32 | 699,3567   | C <sub>33</sub> H <sub>56</sub> O <sub>14</sub> | 699,3562   | -0,7 | 537,3014 (20)   | 675,360720 | -  | -          | Singleton |
| 15 | 6,09 | 537,303380 | C <sub>27</sub> H <sub>46</sub> O <sub>9</sub>  | 537,303404 | 0,0  | -   | 513,306745 | 1-O- $\alpha$ -linolenoyl-<br>3-O- $\beta$ -<br>galactopyranosyl-<br>sn-glycerol | Acide gras | -         |
| 16 | 6,37 | 393,260908 | C <sub>21</sub> H <sub>38</sub> O <sub>5</sub>  | 393,261145 | 0,6  | 371,2757 (42)<br>353,2696 (100)                         | -          | -  | -          | Singleton |
| 17 | 6,52 | 301,141394 | C <sub>16</sub> H <sub>22</sub> O <sub>4</sub>  | 301,141030 | -1,2 | 245,0776 (10)   | -          | -  | -          | Singleton |

|    |       |            |   |            |      |  |            |   |              |           |           |
|----|-------|------------|---|------------|------|--|------------|---|--------------|-----------|-----------|
| 18 | 7,28  | 331,224498 | C <sub>19</sub> H <sub>32</sub> O <sub>3</sub>                | 331,224366 | -0,4 | -  | -          | - | -            | -         | -         |
| 19 | 8,34  | 595,402878 | C <sub>30</sub> H <sub>58</sub> O <sub>11</sub>               | 595,405189 | 3,9  | 335,1881 (39),<br>317,1784 (100)   | 639,393438 | - | -            | -         | Singleton |
| 20 | 8,52  | 829,506620 | C <sub>45</sub> H <sub>74</sub> O <sub>12</sub>               | 829,507249 | 0,8  | 535,287017 (35)  | -          | - | -            | -         | Singleton |
| 21 | 9,21  | 625,265074 | C <sub>34</sub> H <sub>40</sub> O <sub>11</sub>               | 625,264339 | -1,2 | 523,2329 (6),<br>493,2591(9),<br>451,2123 (20),<br>421,2379 (7),<br>404,2063 (5) | -          | - | -            | -         | Singleton |
| 22 | 9,45  | 609,270400 | C <sub>35</sub> H <sub>36</sub> N <sub>4</sub> O <sub>6</sub> | 609,2708   | 0,6  | 591,2612 (11),<br>489,2276 (7),<br>475,2133 (10),<br>461,1972 (7)                | -          | - | Tétrapyrrole | Singleton |           |
| 23 | 9,55  | 975,5650   | C <sub>51</sub> H <sub>84</sub> O <sub>16</sub>               | 975,565157 | 0,1  | 813,5131 (6)<br><br>697,3391 (11),<br>681,3462 (17)                              | 951.568746 | - | Saponine     | Singleton |           |
| 24 | 9,74  | 257,247575 | C <sub>16</sub> H <sub>32</sub> O <sub>2</sub>                | 257,247507 | -0,3 | 187,1696 (25),<br>103,0748 (100)   | 255,233166 | - | Acide gras   | Singleton |           |
| 25 | 9,92  | 559,130980 | C <sub>39</sub> H <sub>20</sub> O <sub>3</sub>                | 559,130465 | -0,9 | -  | -          | - | -            | -         |           |
| 26 | 10,02 | -          | C <sub>18</sub> H <sub>34</sub> O <sub>2</sub>                | -          | -    | -  | 281,248967 | - | Acide gras   | -         |           |
| 27 | 10,67 | 813,5118   | C <sub>45</sub> H <sub>74</sub> O <sub>11</sub>               | 813.5123   | 0,7  | 535,2922 (16),   | 789,516525 | - | -            | Singleton |           |

|    |       |             |  |             |      |  |            |   |            |           |
|----|-------|-------------|--|-------------|------|--|------------|---|------------|-----------|
|    |       |             |  |             |      | 519,2882 (18)  |            |   |            |           |
| 28 | 11,08 | 1099,639538 | C <sub>57</sub> H <sub>94</sub> O <sub>20</sub>  | 1099,641122 | 1,4  | -  | -          | -   | Triterpène | -         |
| 29 | 11,18 | 597,412634  | C <sub>35</sub> H <sub>58</sub> O <sub>6</sub>   | 597,412560  | -0,1 | -  | 609,392616 | Stigmasterol-O-β-glucose  | Stérol     | -         |
| 30 | 11,61 | 599,4281    | C <sub>35</sub> H <sub>60</sub> O <sub>6</sub>   | 599,4282    | 0,2  | -  | 611,408295 | Daucosterol   | Stérol     | -         |
| 31 | 11,63 | 732,5618    | C <sub>40</sub> H <sub>77</sub> NO <sub>10</sub> | 732,5620    | 0,3  | 570,5074 (100),<br>552,4973 (22)   | 766,523950 | Hexadecanamide glycosylé  | Amide gras | Singleton |
| 32 | 11,77 | 959,570309  | C <sub>51</sub> H <sub>84</sub> O <sub>15</sub>  | 959,570243  | -0,1 | 797.517874 (5),<br>681.345588 (36),<br>519.293044 (7)  | 935,573512 | 1,2-di-(alpha-linolenoyl)-3-[alpha-D-galactosyl-(1->6)-beta-D-galactosyl]-sn-glycerol | Acide gras | Cluster 3 |
| 33 | 11,97 | 601,425295  | C <sub>40</sub> H <sub>56</sub> O <sub>4</sub>   | 601,425137  | -0,3 | 584,4236 (73),<br>568,4285 (100),<br>476,3618 (46),<br>145,0997 (42),<br>133,1010 (28),<br>119,0831 (29),<br>105,0694 (25) | -          | Olean-12-en-28-oic acid + 3 (acetyloxy) + 4-ethenylphenyl ester                       | Triterpène | Singleton |
| 34 | 12,21 | 707,168094  | C <sub>44</sub> H <sub>28</sub> O <sub>8</sub>   | 707,167639  | -0,6 | 445,1198 (10),<br>429,0883 (10),   | 719,1490   | -   | Flavonoïde | Cluster 1 |

|    |       |            |   |            |      |   |            |  |            |           |
|----|-------|------------|---|------------|------|---|------------|--|------------|-----------|
|    |       |            |   |            |      | 371,1008 (20),<br>355,0708 (11),<br>285,0091 (10),<br>225,0393 (3)  |            |  |            |           |
| 35 | 12,42 | 961,584666 | C <sub>51</sub> H <sub>86</sub> O <sub>15</sub> | 961,585893 | 1,3  | 281,3474 (23)   | -          | -  | -          | Singleton |
| 36 | 12,69 | 797,517516 | C <sub>45</sub> H <sub>74</sub> O <sub>10</sub> | 797,517419 | -0,1 | 613,4825 (16),<br>519,2922 (19),<br>335,2583 (11)   | 809,497972 | -  | Acide gras | Cluster 3 |
| 37 | 13,00 | 937,585544 | C <sub>49</sub> H <sub>86</sub> O <sub>15</sub> | 937,585893 | 0,4  | 775,5321 (5),<br>681,3453 (22),<br>659,3611 (13)  | 949,567058 | 1-(9Z,12Z,15Z-octadecatrienoyl)-2-hexadecanoyl-3-[α-D-galactosyl-(1→6)-β-D-galactosyl]-sn-glycerol | Acide gras | Cluster 3 |
| 38 | 13,06 | 781,1882   | C <sub>43</sub> H <sub>34</sub> O <sub>13</sub> | 781,1892   | 1,2  | 776,2321 (3)<br><br>503,1064 (8),<br>445,1193 (17),<br>429,0876 (8),<br>371,1006 (36),<br>355,0703 (13),<br>285,0087 (14) | 793,168583 | -  | Flavonoïde | Cluster 1 |
| 39 | 13,45 | 855,207644 | C <sub>49</sub> H <sub>36</sub> O <sub>13</sub> | 855,204812 | -1,2 | 817,2124 (2),   | 867,187718 | -  | Flavonoïde | Cluster 1 |

|    |       |             |   |             |      |   |             |   |            |           |
|----|-------|-------------|---|-------------|------|---|-------------|---|------------|-----------|
|    |       |             |   |             |      | 445,1211 (17),<br>373,0992 (27),<br>355,0718 (25),<br>285,0099 (15)                                     |             |   |            |           |
| 40 | 13,46 | -           | C <sub>27</sub> H <sub>54</sub> O <sub>7</sub>  | -           | -    | -   | 525,357064  | - | -          | -         |
| 41 | 13,91 | 929,226262  | C <sub>48</sub> H <sub>42</sub> O <sub>18</sub> | 929,226335  | 0,1  | 924,2803 (12),<br>445,1206 (18),<br>429,0889 (14),<br>371,1041 (28),<br>355,0713 (28),<br>285,0142 (16) | 941,205700  | - | Flavonoïde | Cluster 1 |
| 42 | 14,30 | 1003,2641   | C <sub>51</sub> H <sub>48</sub> O <sub>20</sub> | 1003,2631   | -0,9 | 445,1289 (6),<br>355,0781 (19),<br>285,0155 (10)  | 1025,2532   | - | Flavonoïde | Cluster 1 |
| 43 | 14,81 | 1079,304024 | C <sub>57</sub> H <sub>52</sub> O <sub>20</sub> | 1079,2944   | 0,6  | 615,1583 (36),<br>581,1310 (38),<br>541,1387 (100),<br>507, 1125 (13),<br>467,1178 (28)                 | 1091,242109 | - | -          | Singleton |
| 44 | 14,85 | 1077,2960   | C <sub>54</sub> H <sub>54</sub> O <sub>22</sub> | 1077,299894 | 3,6  | 1072,3483 (14),<br>503,1263 (3),<br>445,1368 (6),<br>371,1177 (8),                                      | 1089,242833 | - | Flavonoïde | Cluster 1 |

|    |       |             |   |             |      |   |            |  |              |              |
|----|-------|-------------|---|-------------|------|---|------------|--|--------------|--------------|
|    |       |             |   |             |      | 355,0846 (5),<br>341,0315 (2),<br>285,0211 (3)  |            |  |              |              |
| 45 | 15,01 | 939,587914  | C <sub>60</sub> H <sub>78</sub> N <sub>2</sub> O <sub>7</sub> | 939,588179  | 0,3  | 813,5861 (7),<br>535,2690 (71),<br>491,2752 (20),<br>463,2785 (8)                             | -          | -  | -            | Singleton    |
| 46 | 15,42 | 885,558880  | C <sub>55</sub> H <sub>72</sub> N <sub>4</sub> O <sub>6</sub> | 885,552462  | 3,8  | 629,268635 (57),<br>579,2732 (17),<br>519,2493 (7),<br>474,2146 (5)                           | -          | Pheophytin B   | Tétrapyrrole | Cluster 2-2' |
| 47 | 15,66 | 931,760367  | C <sub>57</sub> H <sub>102</sub> O <sub>9</sub>               | 931,759661  | -0,8 | 651,5059 (21),<br>635,5101 (12),<br>317,2323 (17)   | -          | 24-Norolean-12-en-28-oic acid, 3-[(27-carboxyheptacosyl)oxy]-2,16,23-trihydroxy + H <sub>2</sub> O | Triterpène   | Singleton    |
| 48 | 16,06 | 1032,607278 | C <sub>52</sub> H <sub>89</sub> NO <sub>19</sub>              | 1032,610156 | 2,8  | 947,591362 (100),<br>887,566785 (9),<br>609,2541 (58),<br>537,260438 (42),<br>521,265268 (24) | -          | -  | -            | Singleton    |
| 49 | 16,15 | -           | C <sub>54</sub> H <sub>78</sub> O <sub>11</sub>               | -           | -    | -   | 901,548033 | -  | -            | -            |

|    |       |            |   |            |      |  |            |                         |              |              |
|----|-------|------------|---|------------|------|--|------------|-------------------------|--------------|--------------|
| 50 | 16,18 | 887,628390 | C <sub>45</sub> H <sub>90</sub> O <sub>16</sub>               | 887,630163 | 2,0  | 609,3160 (18),<br>591,300810 (44),<br>559,2726 (29),<br>531,275516 (50),<br>515,280037 (14),<br>503,279090 (19),<br>485,2683 (9) | -          | -                       | -            | Singleton    |
| 51 | 16,48 | -          | C <sub>40</sub> H <sub>79</sub> NO <sub>5</sub>               | -          | -    | -  | 652,588938 | -                       | -            | -            |
| 52 | 16,52 | 871,574580 | C <sub>55</sub> H <sub>74</sub> N <sub>4</sub> O <sub>5</sub> | 871.573198 | -1,4 | 593,3121 (92),<br>533,2844 (71),<br>459,2473 (9)   | -          | Pheophytin A            | Tétrapyrrole | Cluster 2-2' |
| 53 | 17,03 | 871,574580 | C <sub>55</sub> H <sub>74</sub> N <sub>4</sub> O <sub>5</sub> | 871.573198 | -1,4 | 593,3117 (91),<br>533,2905 (74),<br>519,2695 (8),<br>459,2461 (14)   | -          | Pheophytin A<br>isomère | Tétrapyrrole | Cluster 2-2' |

Tableau 8 : Les composés chimiques contenus dans l'extrait éthanolique des parties aériennes d'*Eclipta alba*

## Cluster 1

Les composés les plus intenses du chromatogramme sont liés dans le cluster 1. Ils sont également détectés en UV et leurs longueurs maximales d'absorbance sont 242, 267 et 338 nm. Ces longueurs d'ondes sont caractéristiques de la famille des flavonoïdes<sup>282</sup>. Compte tenu de la polarité de ces composés (élués en fin de gradient), on peut supposer que ce sont des flavonoïdes avec une génine complexe ou associés à des motifs lipophiles.

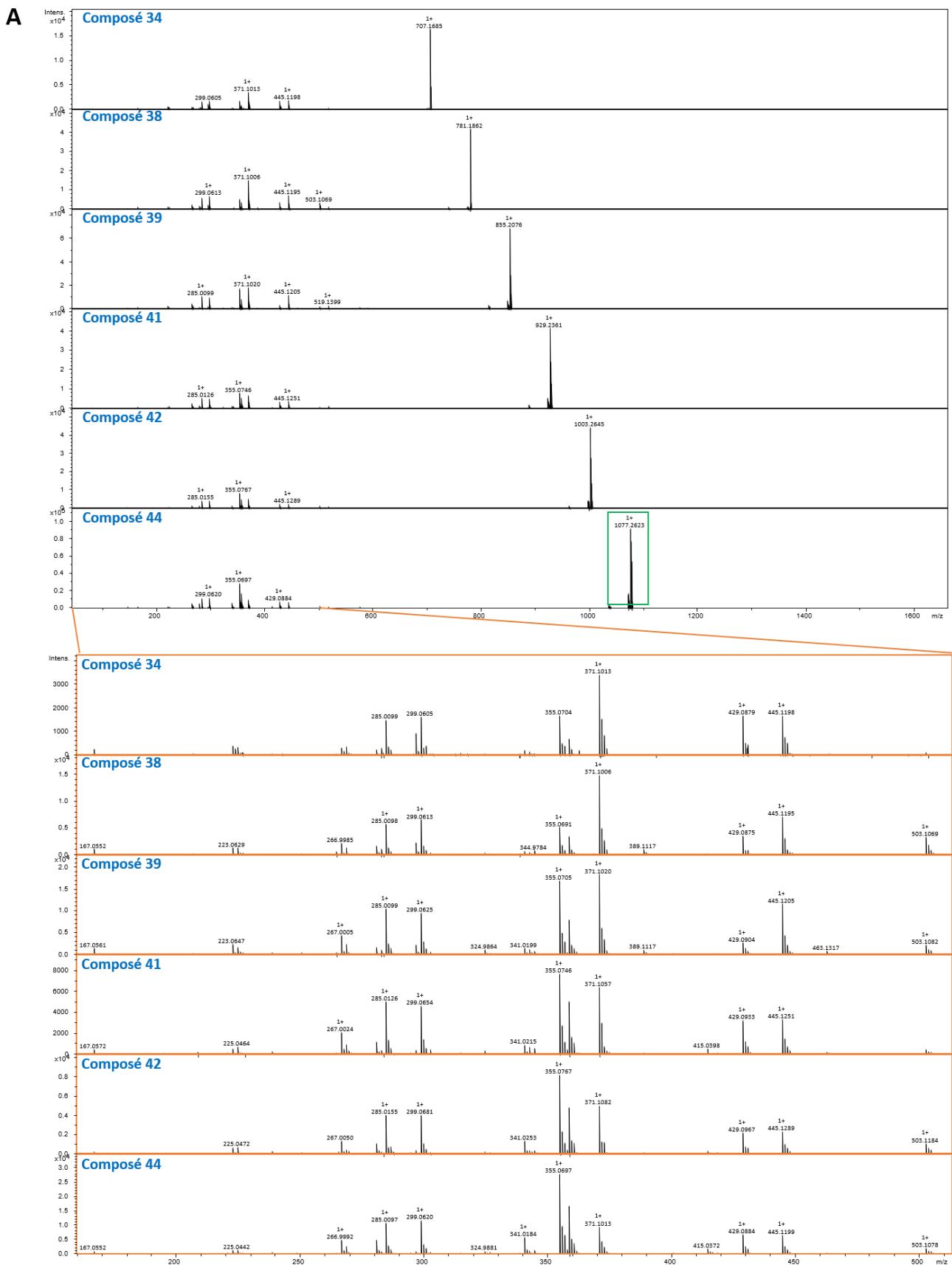
La Figure 69 présente les spectres MS/MS des composés du cluster 1, soit les composés 34, 38, 39, 41, 42 et 44. Les fragments observés sur le spectre MS/MS proviennent de la fragmentation des adduits  $[M+NH_4]^+$  puisque les ions moléculaires ont une très faible intensité et n'ont pas été fragmentés en MS/MS et les adduits  $[M+Na]^+$  ne se fragmentent pas aux énergies de collision utilisées, ils nécessitent une énergie plus importante pour se fragmenter.

Les spectres MS/MS des composés liés dans le cluster 1 avec une focalisation sur leurs fragments entre  $m/z$  100-550 sont présentés dans la Figure 69 A. Des ions fragments communs sont observés pour toutes ces molécules. De plus, les ions moléculaires ont une différence de 74 u qui pourrait correspondre à un motif  $C_3H_6O_2$ . Néanmoins, certains des composés qui sont liés dans le cluster 1 (composé 43, 44) sont coélusés avec deux autres molécules ayant une différence de 1 et 2 u, ce qui engendre des massifs isotopiques complexes Figure 69 B. Lors de l'isolation de ces ions moléculaires pour réaliser la fragmentation en CID, la fenêtre de la sélection englobe ces 3 molécules ce qui empêche de déterminer quel fragment provient de quel ion moléculaire et de déterminer le schéma de fragmentation. De plus, ces coelutions de composés très proches en  $m/z$  peuvent engendrer des décalages en décimales de la masse réduisant la confiance accordée aux formules brutes proposées.

Lors de la soumission des formules brutes proposées pour l'ensemble de ces composés liés dans le cluster 1 aux bases de données publiques, peu de formules développées sont proposées et celles proposées ne sont pas cohérentes avec les temps de rétention et spectres MS/MS obtenus. Toutefois, les fragments de ces molécules présentent des masses similaires à celles des ions observés sur un spectre MS/MS ajouté dans les bases de données de GNPS par Dorrestein (CCMSLIB00000847749 NCGC00385199-01) pour un flavonoïde diglycosylé. Bien que compte tenu des temps de rétention et du pourcentage de solvant organique nécessaire à l'élution des composés d'*Eclipta alba*, ces composés ne semblent pas glycosylés.

Les données MS et UV de ces composés permettent d'assoir leur appartenance à la classe de flavonoïde, néanmoins, les informations que nous avons ne permettent pas d'élucider les structures définitives de ces molécules. Avec un analyseur plus résolutif tel que le FT-ICR ou

permettant de mieux sélectionner un seul ion parent tel que le Triple Quad, il serait probablement possible de distinguer et de fragmenter séparément ces composés coélués et de relier les fragments à leurs ions moléculaires.



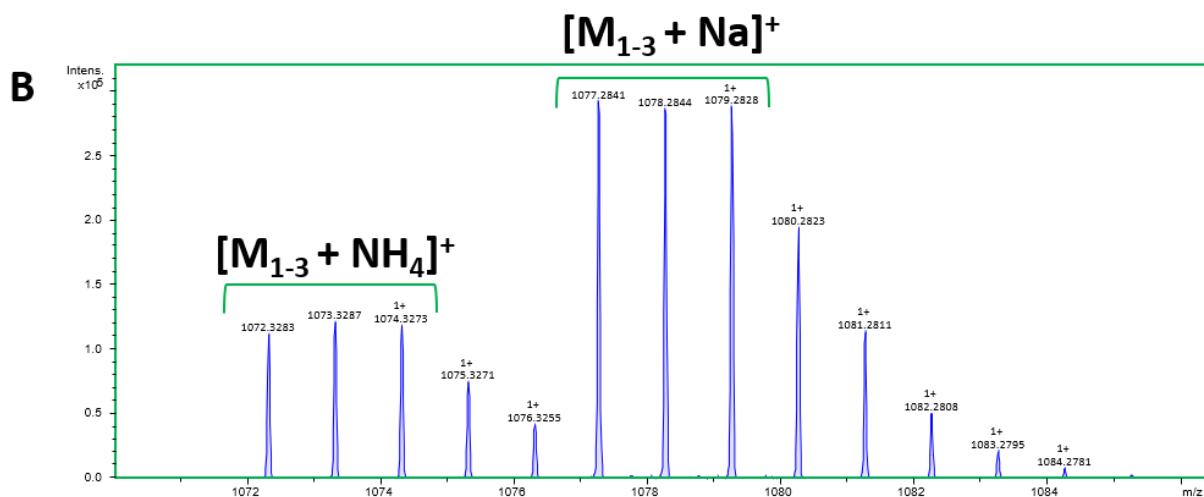


Figure 69 : A) Spectres MS/MS des composés liés dans le cluster 1 avec une focalisation sur la partie commune  $m/z$  100-550. B) Exemple d'un massif isotopique de l'ion moléculaire du composé 44

## Cluster 2 et 2'

Le composé 52 a été reconnu par les bases de données de GNPS comme étant la Phéophytine A (CCMSLIB00000471426) avec un score cosiné de 0,93, ce qui indique que le spectre expérimental est très proche du spectre enregistré dans la base de données. L'interprétation du spectre MS/MS a ensuite confirmé cette proposition de structure. Le nœud du composé 46, voisin de la Phéophytine A, comportant un même schéma de fragmentation (Figure 70) a ensuite été associé à la Phéophytine B. Ces deux composés, appartenant à la famille des tétrapyrroles, ont plusieurs adduits formés en source et fragmentés en MS/MS donnant des schémas de fragmentation différents, ce qui explique la formation de 2 clusters regroupant les ions moléculaires d'une part (cluster 2') et leurs adduits dans une autre part (cluster 2). Cette redondance d'information ne peut pas être gérée avec le logiciel MetaboScape. En effet, tous les fragments ou adduits qui sont générés en source et qui sont introduits dans la cellule de collision sont considérés comme des composés indépendants. Le logiciel ne peut pas faire la différence entre un composé et un adduit ou fragment, tous rentrés en cellule de collision de la même manière.

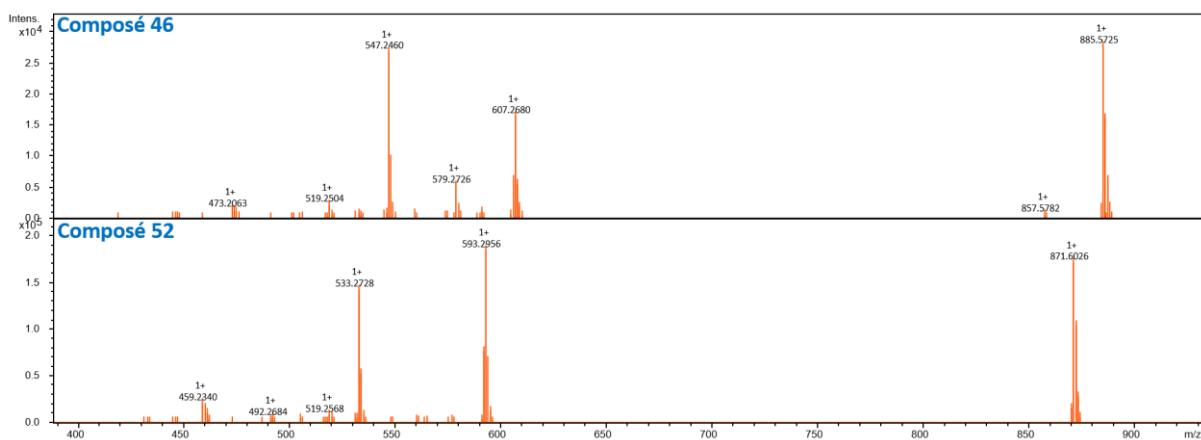


Figure 70: Spectres MS/MS des composés 46 et 52

### Cluster 3

La formule brute du composé 32 a été soumise aux bases de données publiques telles que PubChem et SciFinder et 67 propositions de structure ont été trouvées. Le spectre du composé 32 a ensuite été travaillé manuellement et les schémas de fragmentation ont été interprétés afin de l'identifier (Figure 71). Le spectre de l'adduit  $[M+Na]^+$  permet de distinguer entre les fragments parvenant de la fragmentation de l'adduit  $[M+Na]^+$  et ceux de l'adduit  $[M+NH_4]^+$ .

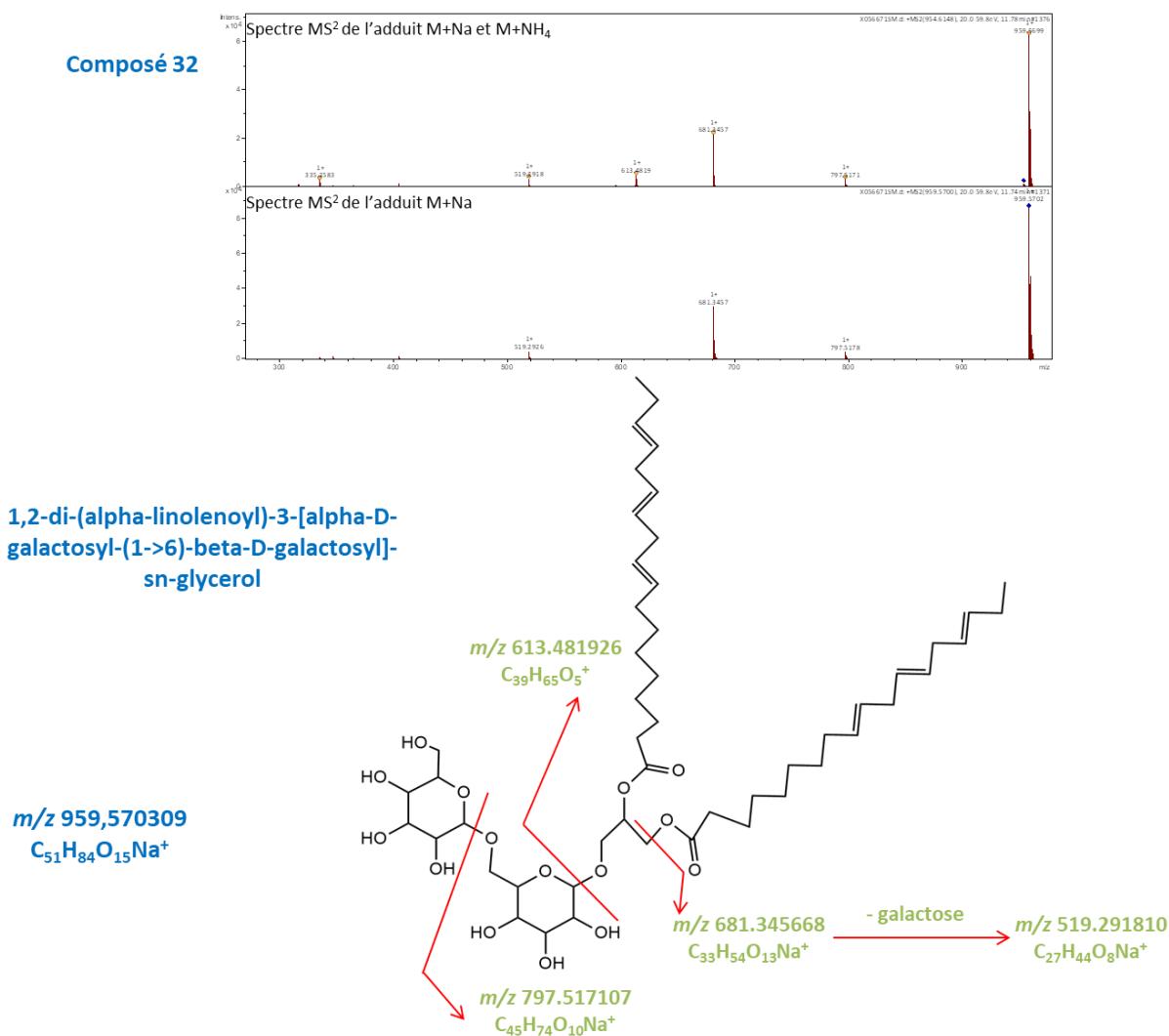


Figure 71 : Interprétation du spectre MS/MS du composé 32 et proposition de voies de fragmentations

Cette interprétation nous a permis d'identifier un composé du cluster, le 1,2-di-(alpha-linolenoyl)-3-[alpha-D-galactosyl-(1->6)-beta-D-galactosyl]-sn-glycerol.

Le composé 37 est le nœud voisin du composé 32, ils sont liés par une arrête car leurs spectres MS/MS présentent le même schéma de fragmentation avec des fragments communs. En soumettant la formule brute du composé 37 aux bases de données publiques, 24 propositions ont été trouvées sur SciFinder et 63 propositions sur PubChem. Le composé 1-(9Z,12Z,15Z-octadecatrienoyl)-2-hexadecanoyl-3-[α-D-galactosyl-(1→6)-β-D-galactosyl]-sn-glycerol a ensuite été sélectionné, parmi toutes les autres propositions, au vu de sa similarité structurale et spectrale avec le composé 32. Pareillement pour le composé 36, un autre nœud voisin du 32 et 37, présentant le même schéma de fragmentation faisant donc partie de la famille des acides gras (Figure 72). Les spectres des adduits  $[M+Na]^+$  permettent de distinguer entre les fragments parvenants de la fragmentation de l'adduit  $[M+Na]^+$  et ceux de l'adduit  $[M+NH_4]^+$ .

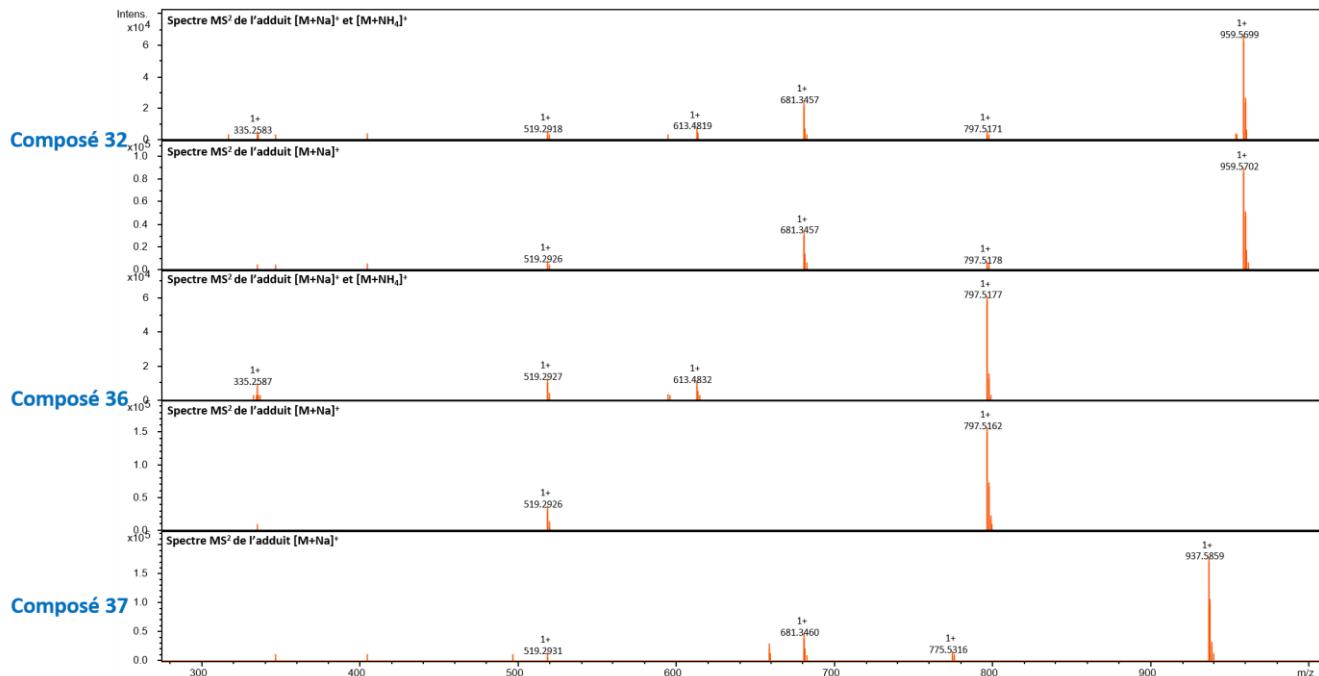


Figure 72 : Spectres MS/MS des composés 32, 36 et 37 liés en cluster 3

Enfin, l'annotation des composés liés en clusters a permis de mettre en évidence la présence des familles moléculaires majoritaires présentées sur la Figure 73. Ces composés liés en clusters n'ont pas été décrits dans la littérature comme étant présents dans *Eclipta alba*.

L'utilisation du réseau moléculaire et des similarités spectrales a permis de propager l'annotation des composés, les bases spectrales de GNPS ont notamment aidé à l'identification de certains composés. D'après les résultats obtenus, on peut conclure que les flavonoïdes, les triterpènes, notamment les saponines, les tétrapyrroles et acides gras sont les classes majoritaires dans l'extrait éthanolique d'*Eclipta alba*. Cependant, les triterpènes et saponines sont représentés par des nœuds singulons dans le réseau moléculaire et n'ont pas été liés en cluster car certains n'ont pas été fragmentés et d'autres ont des spectres MS/MS différents sans similarité spectrale.

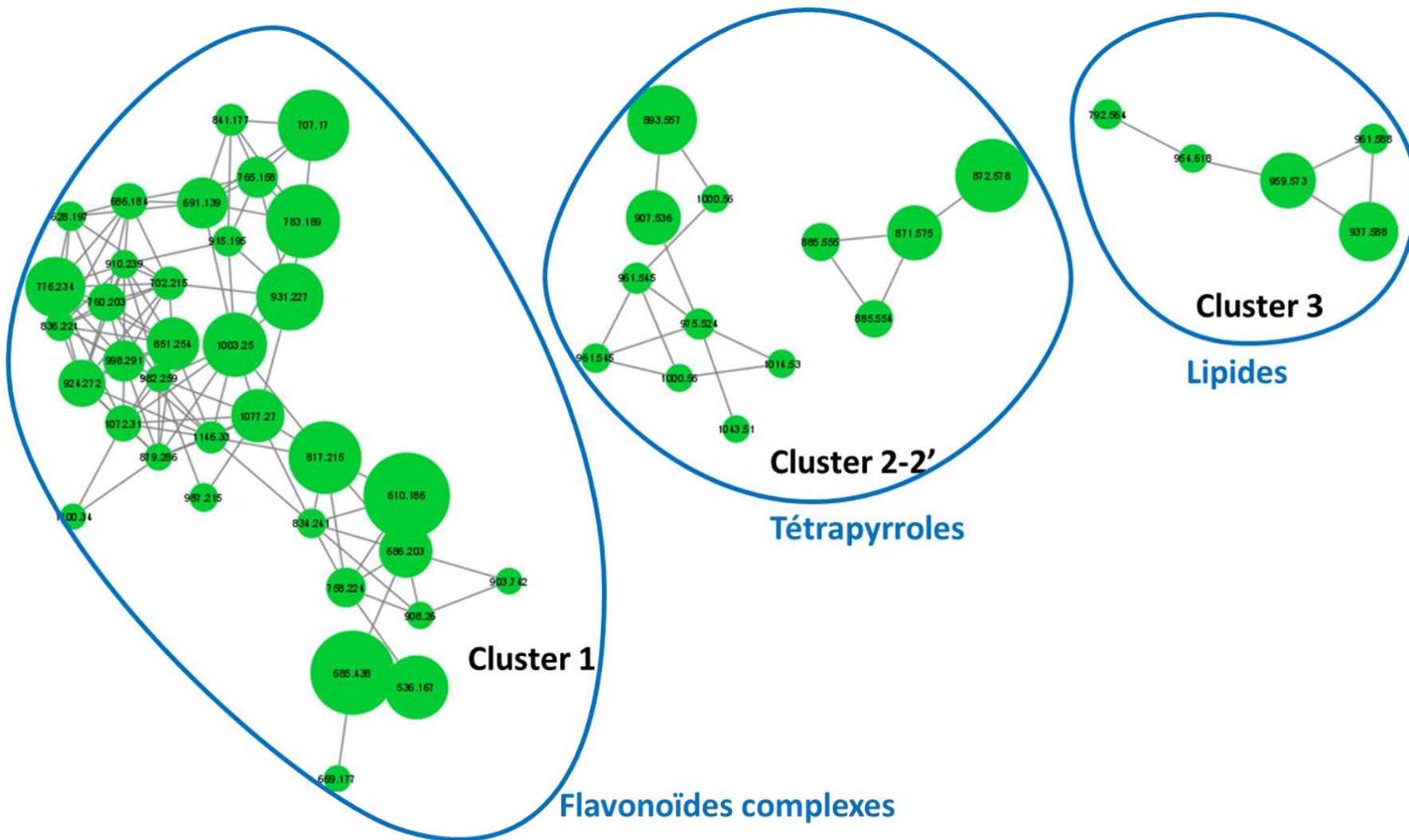


Figure 73 : Les familles moléculaires regroupant les composés les plus intenses détectés par UHPLC-HRMS en mode (+) ESI dans l'extrait éthanolique des parties aériennes d'*Eclipta alba*

Une fois l'étude phytochimique de l'extrait d'*Eclipta alba* réalisée, il a été fractionné par CPC en suivant le mode opératoire décrit en Annexe 2. L'extrait d'*Eclipta alba* et les 8 fractions (A→H) obtenues par CPC à partir de l'extrait brut, ont été évaporés et solubilisés dans l'éthanol à 1 mg/mL pour l'évaluation de leurs activités biologiques.

## 2) Activités biologiques

### a) Activité antioxydante

La Figure 74 montre les résultats des tests antioxydants de l'extrait *Eclipta alba* et de ses fractions CPC. L'absorbance a été tracée en fonction de la concentration des échantillons.

On peut remarquer que l'extrait d'*Eclipta alba* a une bonne capacité de réduction des radicaux DPPH<sup>·</sup> et ABTS<sup>·+</sup>, et de chélation du fer (II). L'activité antioxydante est proche du contrôle positif pour le test DPPH et l'ABTS, voire meilleure que le trolox pour la chélation du fer (II). Au contraire, sa capacité à réduire les ions cuivreux et ferriques est faible, surtout dans le cas des ions cuivreux. Cela signifie que les composés agissent principalement par transfert d'hydrogène plutôt que par transfert d'un seul électron. Néanmoins, les 5 graphiques montrent le même ordre d'activité entre les fractions, ce qui indique que les 5 évaluations classent les fractions dans le même ordre.

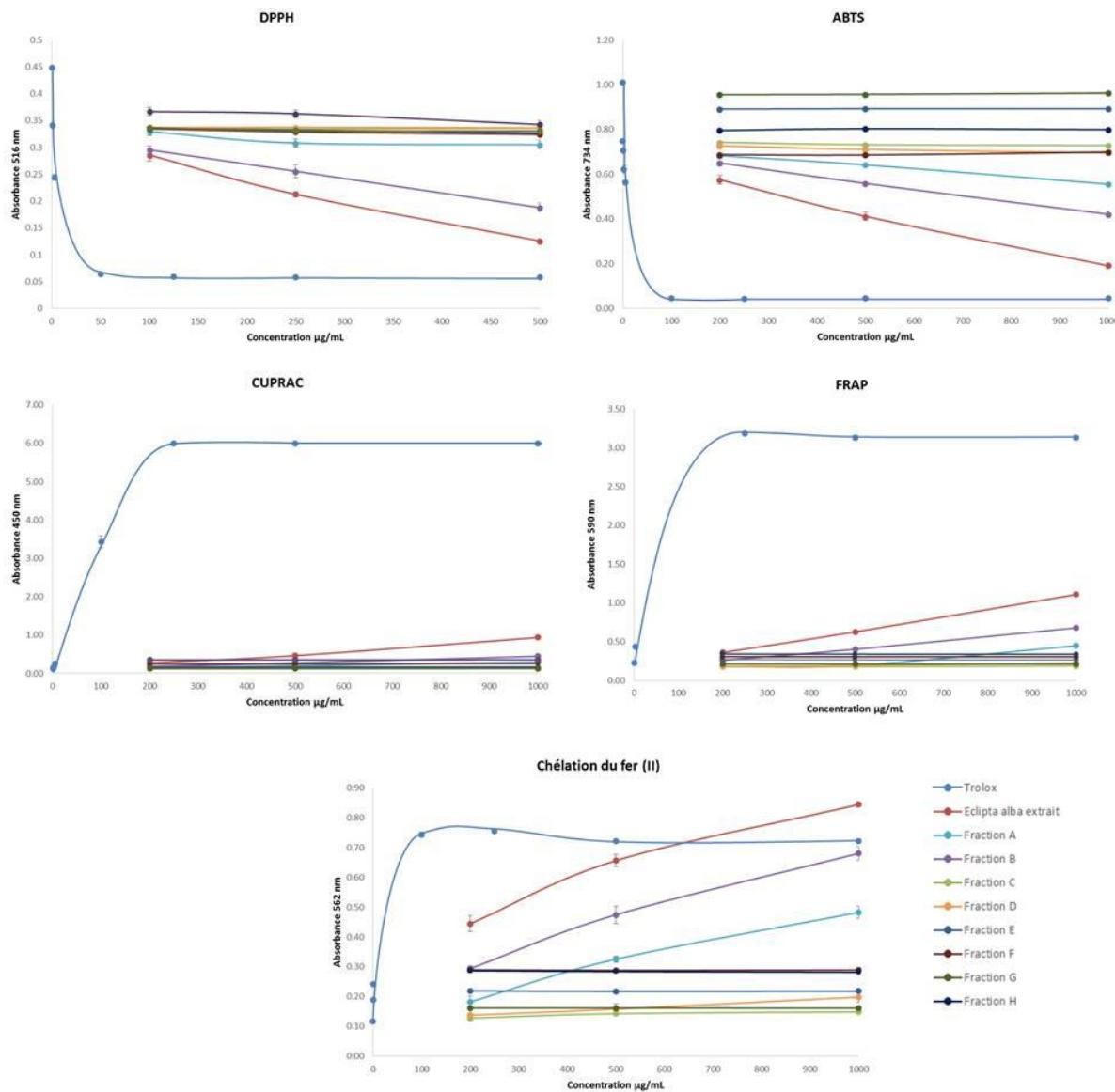


Figure 74 : Activité antioxydante de l'extrait *Eclipta alba* et ses fractions CPC. L'absorbance moyenne a été tracée  $\pm$  l'écart-type ( $n = 3$ )

Les résultats montrent que l'extrait brut et la Fraction B ont la meilleure capacité antioxydante, avec des activités élevées et proches ; cela signifie que les composés antioxydants de l'extrait brut sont principalement présents dans la Fraction B. De plus, la Fraction A présente une légère capacité antioxydante, inférieure à la Fraction B mais plus active que les autres fractions qui n'ont presque aucun pouvoir antioxydant.

**b) Activité anti-enzymatique**

## Activité anti-collagénase

L'activité d'inhibition de la collagénase par l'extrait d'*Eclipta alba* et des fractions CPC est présentée sur la Figure 75. Les résultats montrent une fluorescence élevée due à la

dégradation du substrat (la gélatine), par la collagénase. Cette fluorescence diminue en présence d'un inhibiteur enzymatique, la 1,10-phénanthroline, à 4 et 40 µM. On peut remarquer que l'extrait d'*Eclipta alba* et toutes les fractions CPC ont une activité d'inhibition de l'enzyme collagénase démontrée par la réduction de la fluorescence en présence de 50 µg/mL d'échantillon dans le puits. Tous les échantillons inhibent plus de 50% de l'activité enzymatique ; cependant, les fractions B et H présentent la meilleure activité anti-enzymatique soit 90% d'inhibition de la collagénase après 2 heures d'incubation.

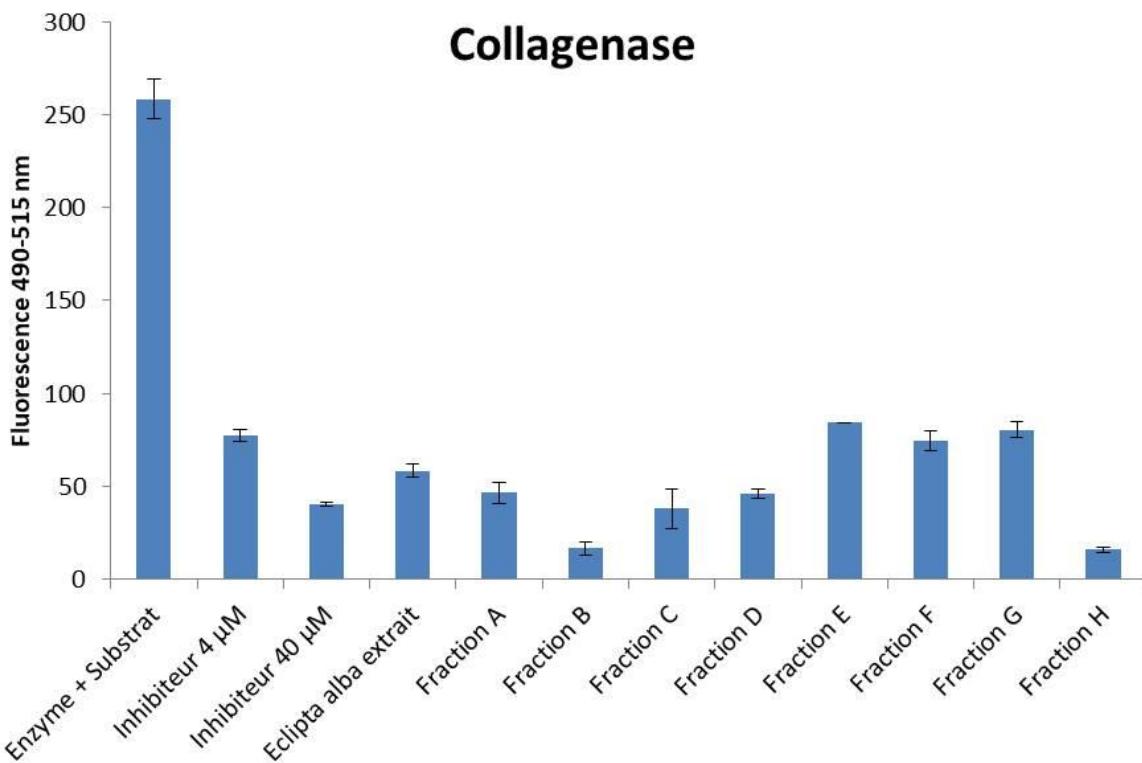


Figure 75 : Activité d'inhibition de collagénase par l'extrait *Eclipta alba* et ses fractions CPC. La fluorescence moyenne a été tracée  $\pm$  l'écart-type ( $n = 3$ )

Cette forte activité anti-collagénase d'*Eclipta alba* pourrait être due en partie à la présence de polyphénols puisqu'ils sont connus pour leur activité anti-collagénase<sup>283</sup>. En effet, les groupes hydroxyle des polyphénols jouent un rôle important en agissant comme accepteurs ou donneurs de liaisons hydrogène avec le squelette amide et d'autres groupes fonctionnels de la chaîne de l'enzyme<sup>284</sup>.

### Activité anti-élastase

L'activité d'inhibition de l'élastase par l'extrait d'*Eclipta alba* et ses fractions CPC est illustrée sur la Figure 76. Une forte fluorescence est observée après 2 heures d'incubation de l'enzyme et du substrat, elle est due à la dégradation du substrat, c'est-à-dire de l'élastine par l'élastase. Les résultats montrent une diminution de cette fluorescence en présence d'un inhibiteur

d'enzyme, à savoir la N-méthoxysuccinyl-Ala-Ala-Pro-Val-chlorométhylcétone à 125 µM. On peut voir que l'extrait d'*Eclipta alba* à 50 µg/mL dans le puits, inhibe 80% de l'activité de l'élastase. De plus, dans les Fractions A, B et C, on observe une fluorescence élevée, ce qui signifie que ces fractions n'ont pas d'effet sur l'activité enzymatique contrairement à d'autres fractions (D, E, F, G, H) qui inhibent plus de 80% de l'activité enzymatique. La Fraction H est la plus active avec 95% d'inhibition de l'élastase.

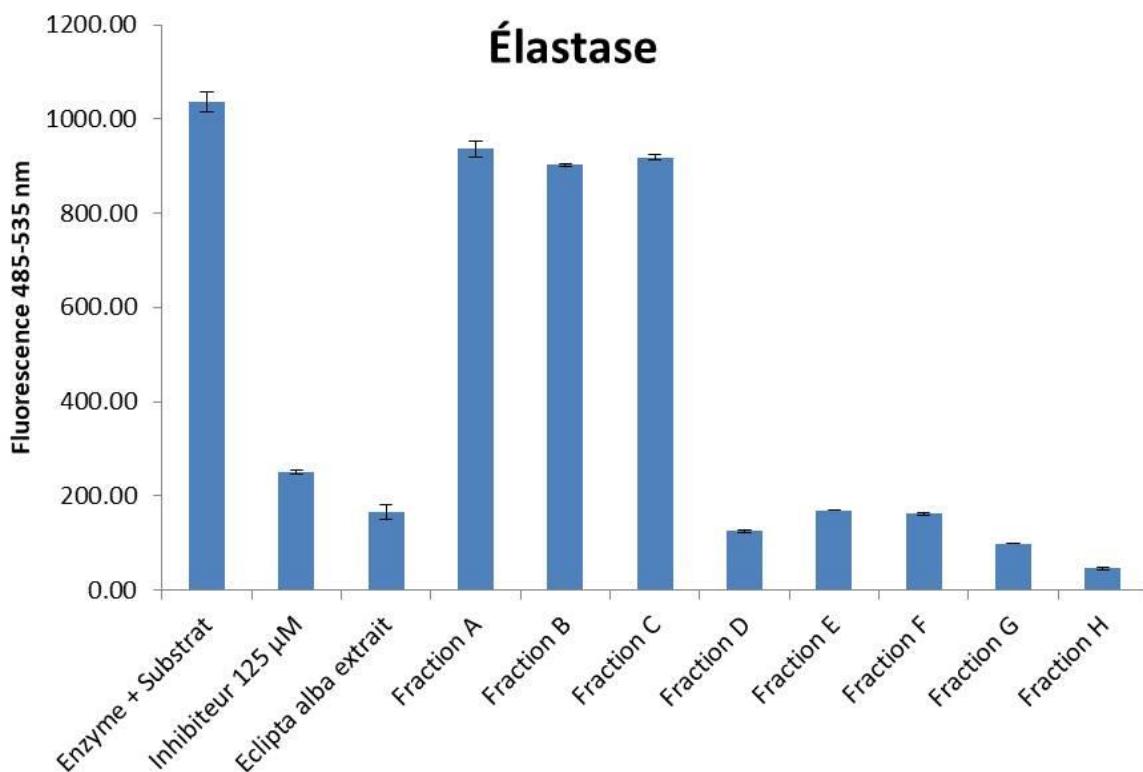


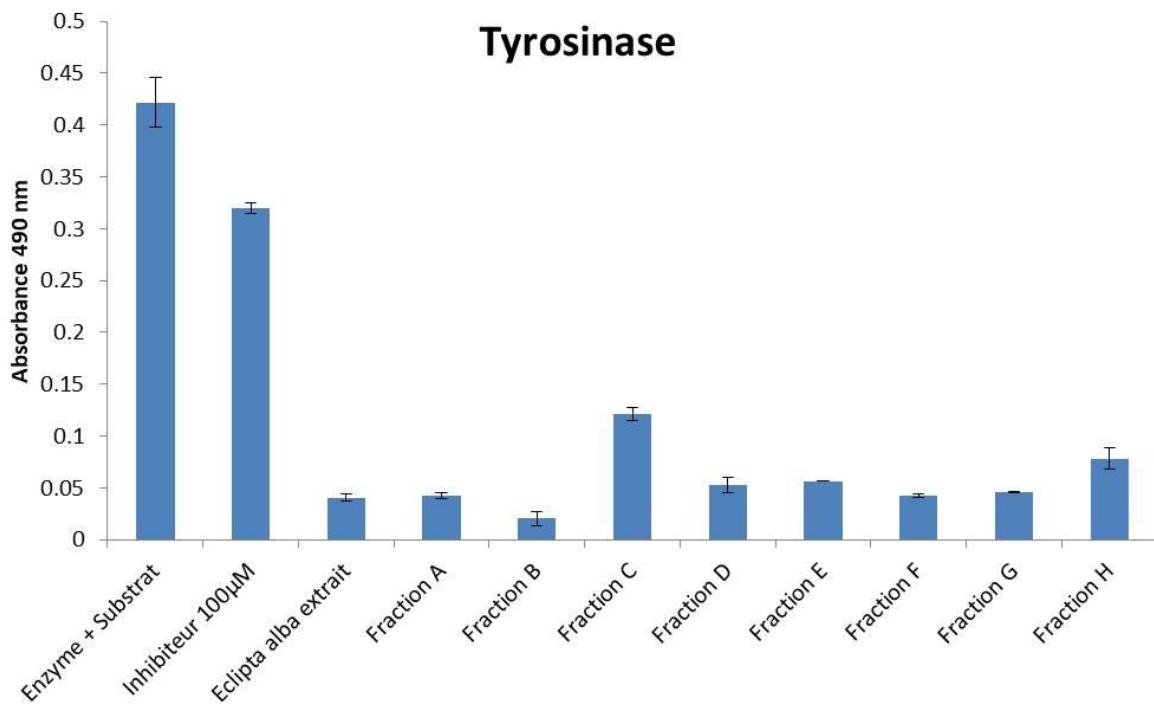
Figure 76 : Activité d'inhibition de l'élastase par l'extrait d'*Eclipta alba* et ses fractions CPC. La fluorescence moyenne a été tracée  $\pm$  l'écart-type ( $n = 3$ )

Des études ont rapporté que plusieurs métabolites triterpénoïdes pentacycliques d'origine végétale sont des inhibiteurs de l'hydrolyse des substrats peptidiques synthétiques et de l'élastine par l'élastase<sup>285</sup>. En effet, Le groupe carboxyle en position 28 dans le système pentacyclique des triterpènes contribue à la liaison aux élastases, permettant ainsi son inhibition<sup>286</sup>.

### Activité anti-tyrosinase

La Figure 77 montre les résultats d'inhibition de la tyrosinase par l'extrait d'*Eclipta alba* et ses fractions CPC. Une absorbance élevée est observée lorsque l'enzyme est mélangée avec le substrat, c'est-à-dire la tyrosinase et la L-Dopa, après une incubation de 30 minutes à 37°C. L'absorbance augmente en présence d'inhibiteur de la tyrosinase, la glabridine à 100 µM. On peut remarquer que l'extrait d'*Eclipta alba* et toutes les fractions CPC à 50 µg/mL inhibent plus

de 60% de l'activité de tyrosinase, en particulier la fraction B qui inhibe 95% de l'oxydation de la L-Dopa par la tyrosinase.



*Figure 77 : Inhibition de la tyrosinase par l'extrait d'*Eclipta alba* et ses fractions CPC. L'absorbance moyenne a été tracée  $\pm$  l'écart-type ( $n = 3$ )*

Des études ont montré que les flavonoïdes sont des chélateurs de métaux en raison de leur structure polyhydroxyphénolique, qui peut faire interagir les ions de cuivre dans le site actif de la tyrosinase<sup>287</sup>. En effet, les polyphénols agissent contre la tyrosinase par un mécanisme compétitif d'inhibition, en accord avec le rôle biologique de cette enzyme, qui est en fait une polyphénoloxydase<sup>296</sup>.

En conclusion, les tests biologiques *in vitro* présentés ci-dessus se concentrent sur les activités liées à la peau, soit activité antioxydante, anti-collagénase, anti-élastase et anti-tyrosinase de l'extrait éthanolique d'*Eclipta alba*, afin de cibler ses composés bioactifs. Le Tableau 9 regroupe les résultats des tests chimiques et enzymatiques avec une comparaison de l'activité des fractions CPC d'*Eclipta alba*.

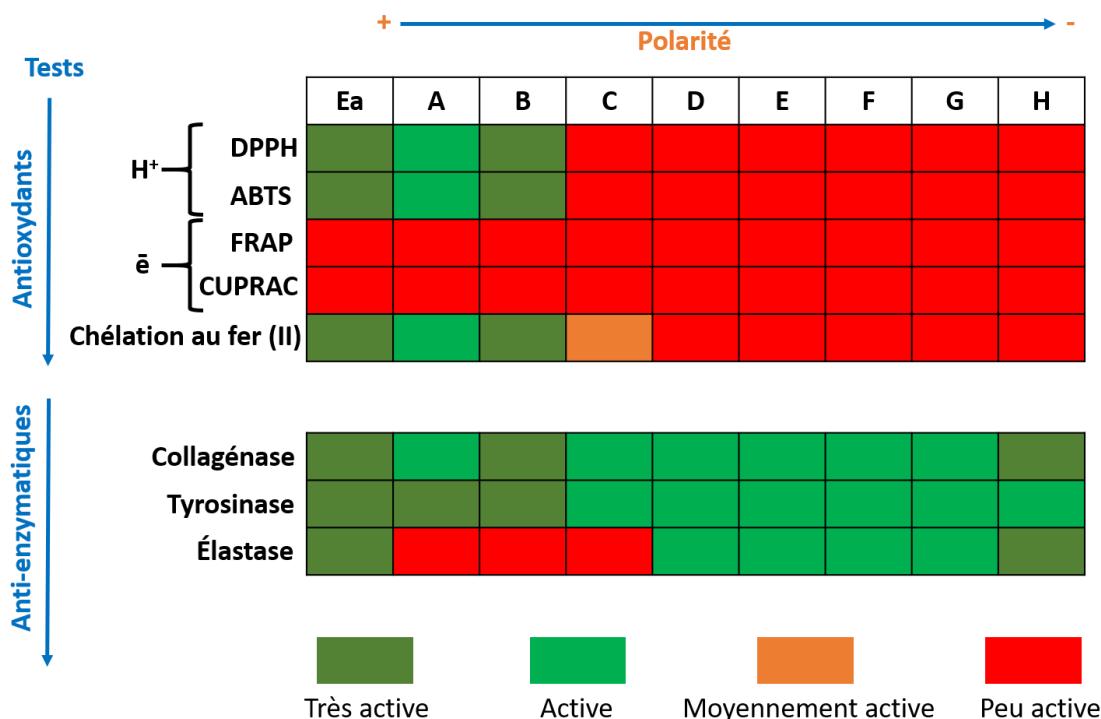


Tableau 9 : Récapitulatif des résultats antioxydants et anti-enzymatiques des fractions CPC d'*Eclipta alba*

La fraction B montre une activité antioxydante, anti-collagénase et anti-tyrosinase plus importante que les autres fractions de l'extrait. La fraction H présente les activités anti-collagénase et anti-élastase les plus intéressantes. Toutes les fractions ont été analysées par spectrométrie de masse et on s'est focalisé sur les fractions actives afin de cibler les composés bioactifs.

### 3) Identification des molécules bioactives

Les résultats de l'évaluation des activités biologiques ont mis en évidence la présence de plusieurs fractions CPC bioactives. Afin de cibler les composés bioactifs qui participent à l'activité globale de l'extrait brut d'*Eclipta alba*, les fractions CPC ont été analysées en UHPLC-MS/HRMS et leurs réseaux moléculaires ont été superposés à celui de l'extrait brut afin de comparer les proportions des composés. En effet, le suivi des éluctions des composés indépendamment exprimé par l'évolution de leurs intensités dans les différentes fractions, permet de mettre en évidence rapidement les composés bioactifs ayant des intensités proportionnelles à l'activité biologique des fractions.

La Figure 78 montre les chromatogrammes de l'extrait brut d'*Eclipta alba* et de ses fractions CPC ainsi qu'un échantillon blanc qui permet de repérer les contaminants présents dans le système d'analyse afin de ne pas les prendre en compte lors de la comparaison. L'analyse non exhaustive des chromatogrammes permet de suivre l'évolution de l'élution des composés.

En effet, les composés polaires élués au début de l'analyse UHPLC-HRMS (composés phénoliques et terpènes glycosylées) sont intenses dans les premières fractions et disparaissent au bout de la fraction E. Ceux élués au milieu du chromatogramme, ne sont pas observés dans les premières fractions et sortent principalement dans la fraction D. En revanche, les flavonoïdes apolaires (liés dans le cluster 1) sortant vers la fin du gradient en UHPLC-HRMS, sont élués majoritairement dans les premières fractions CPC (B et C). Ceci montre la différence de sélectivité entre les deux techniques chromatographiques CPC et HPLC puisque ces composés sont fortement retenus sur la colonne C18 alors qu'ils présentent une forte solubilité dans la phase mobile polaire CPC. Tandis que les autres composés apolaires sortant également à la fin de l'analyse UHPLC-HRMS (acides gras et tétrapyrroles) sont retrouvés dans la fraction H obtenue à l'étape d'extrusion. La fraction F quant à elle, présente une empreinte moléculaire assez riche car c'est la fraction CPC pour laquelle les coefficients de partage de la majorité des composés sont équivalents.

Certains composés sont donc retrouvés dans plusieurs fractions mais leurs intensités sont différentes dans les fractions. Enfin, ces chromatogrammes montrent une répartition différente des composés de l'extrait brut dans les fractions CPC avec une récupération de tous les composés détectés dans l'extrait brut d'*Eclipta alba*.

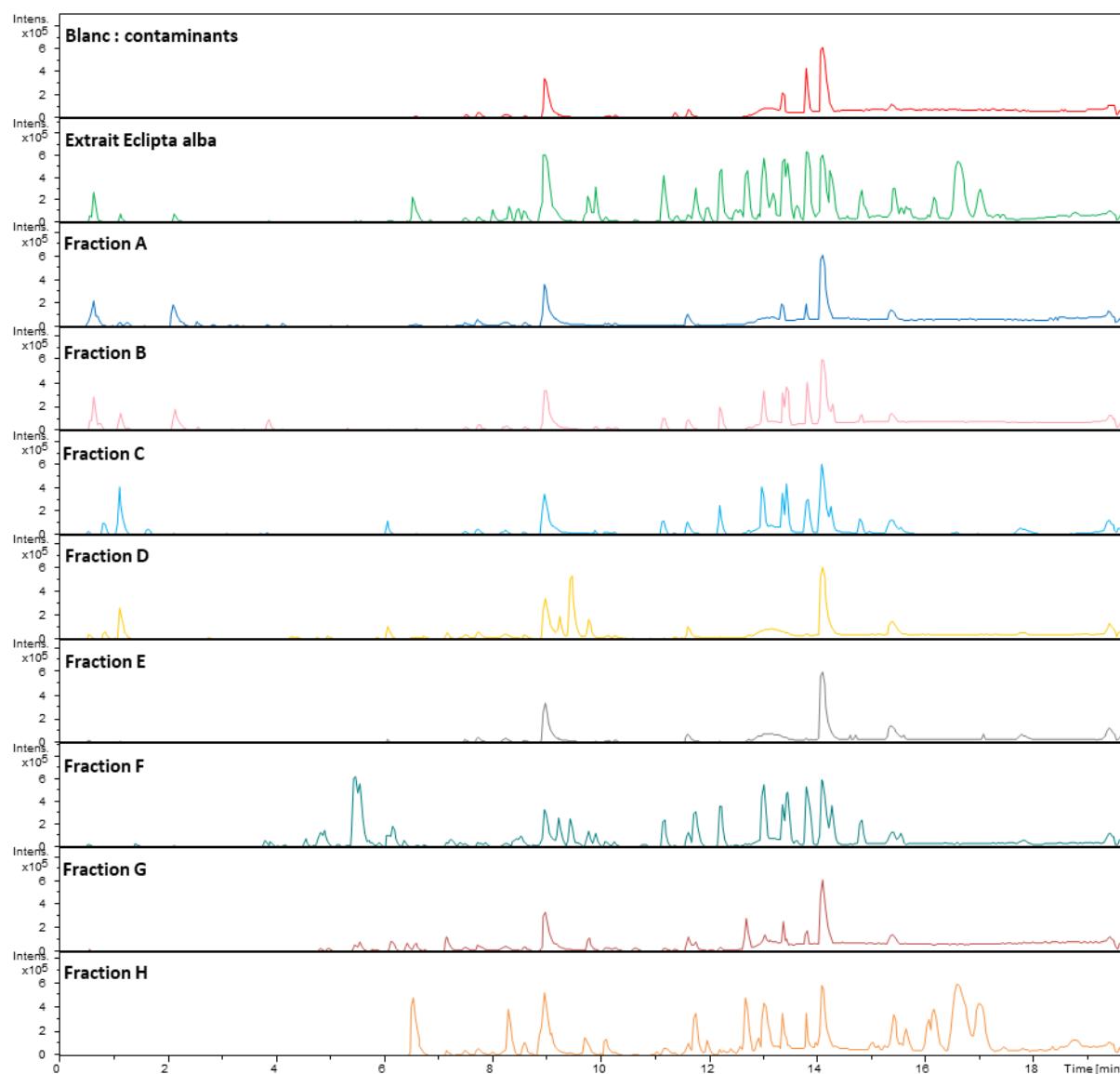


Figure 78 : BPC (Base Peak Chromatogram) de l'extrait d'*Eclipta alba* et des fractions CPC

Les réseaux moléculaires de l'extrait brut d'*Eclipta alba* avec chacune des fractions CPC ont ensuite été créés pour une comparaison plus exhaustive de la répartition des composés dans les différentes fractions.

Comme mentionné précédemment, lors de l'étude phytochimique de l'extrait brut, les composés élués au début du gradient sont sous forme de nœuds singletons dans le réseau moléculaire tandis que les composés plus lipophiles ont formé les clusters 1, 2-2' et 3.

Les résultats de suivi de l'évolution d'élution des composés sont présentés ci-après en conservant la numérotation des composés décrits précédemment (Tableau 8). La Figure 79 regroupe les composés singletons élués entre 0,5 et 6 min. La Figure 80 présente les composés du cluster 1 tandis que la Figure 81 montre les composés des clusters 2-2', 3. La

couleur verte représente l'extrait brut d'*Eclipta alba* tandis que les autres couleurs sont les fractions CPC.

À l'intérieur de chaque nœud, il est mentionné la valeur *m/z* de l'ion parent qui peut être soit l'ion moléculaire  $[M+H]^+$  soit l'adduit sodium  $[M+Na]^+$  soit l'ion fragment après une perte d'eau  $[M-H_2O+H]^+$ . La taille de chaque nœud est proportionnelle à l'intensité globale du composé dans l'extrait brut, et le camembert à l'intérieur du nœud représente la proportion des composés dans les échantillons, définies par les valeurs de leurs aires de pics en extrait d'ion parent.

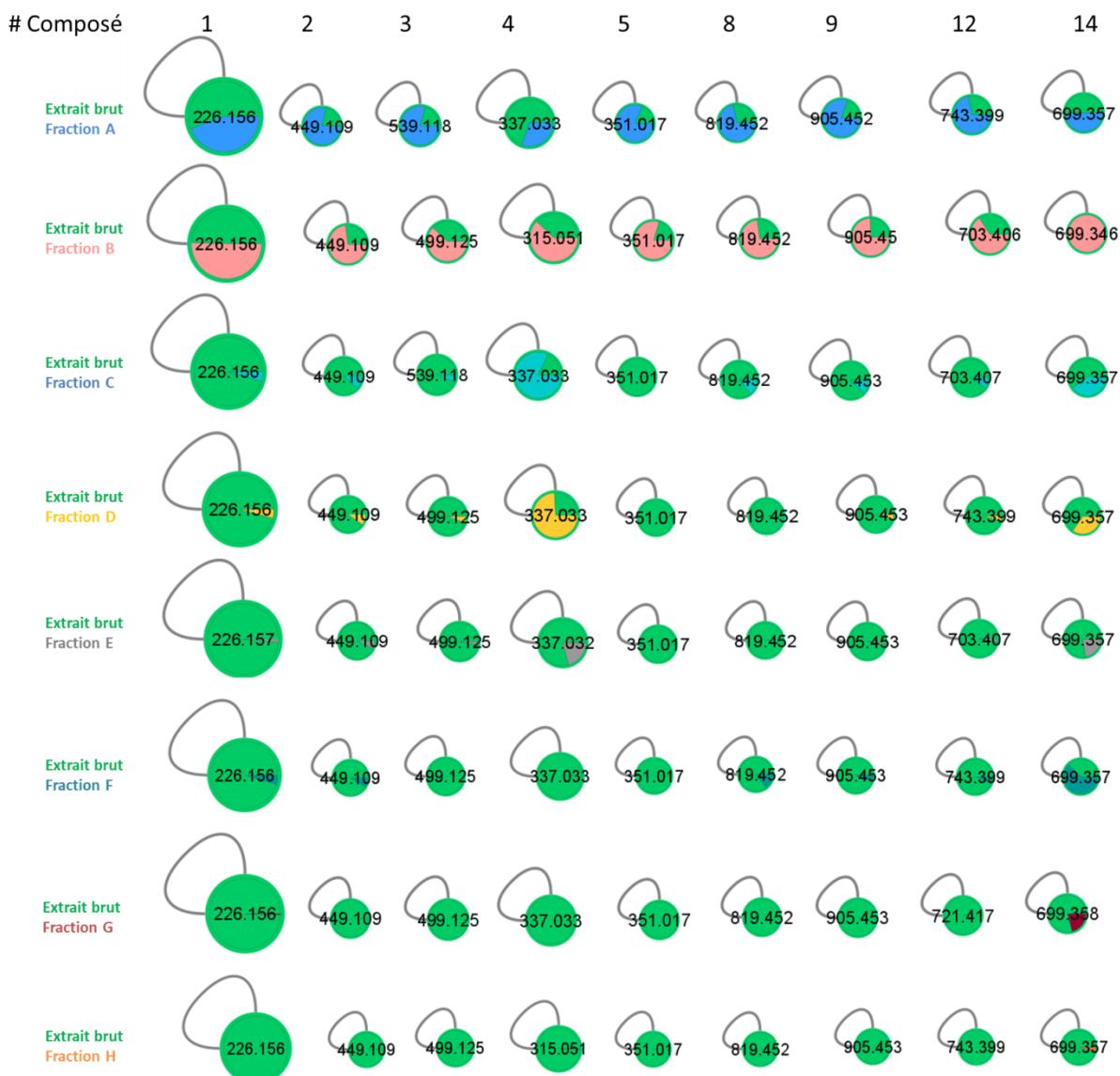


Figure 79 : Évolution de l'élution des composés polaires à moyennement polaires, présentés en nœuds singletons sur le réseau moléculaire, dans les différentes fractions CPC de l'extrait d'*Eclipta alba*

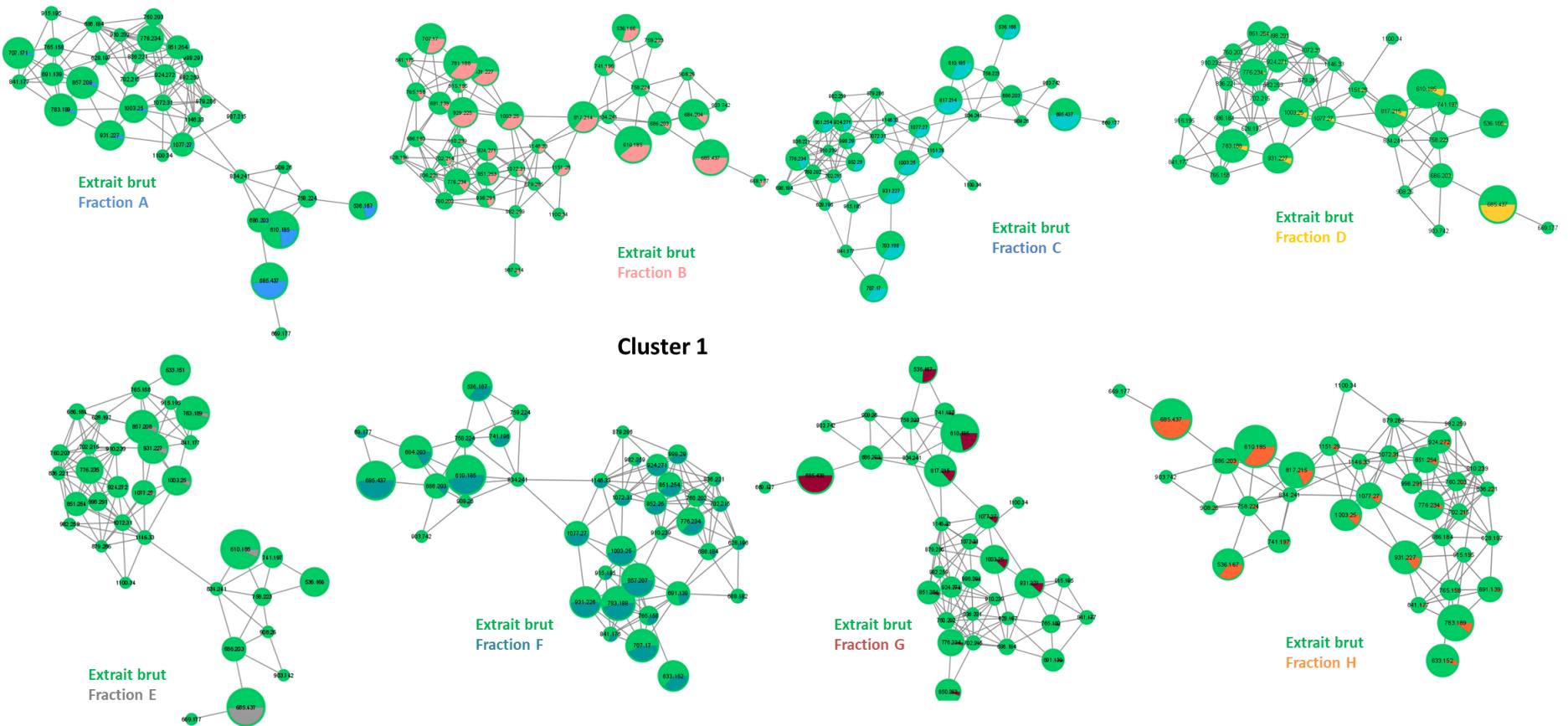


Figure 80 : Évolution de l'élution des composés du cluster 1 dans les différentes fractions CPC de l'extrait d'*Eclipta alba*

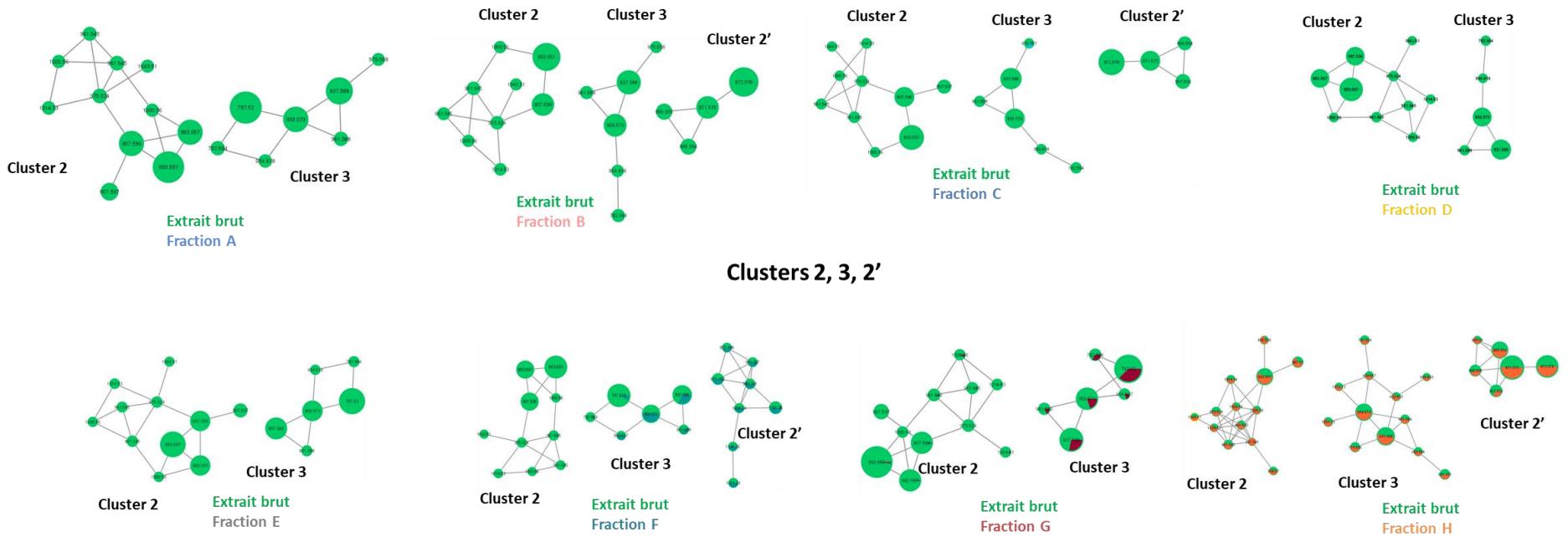


Figure 81 : Évolution de l'élution des composés des clusters 2, 3 et 2' dans les différentes fractions CPC de l'extrait d'*Eclipta alba*

L'utilisation des réseaux moléculaires quantitatifs permet une analyse précise et une quantification des composés puisque le logiciel MetaboScape fait un extrait d'ion de l'ion parent et relève l'aire du pic chromatographique. Il permet ainsi de gérer les coélutions puisqu'il relève les aires des pics en faisant des extraits d'ions composé par composé. Nous pouvons ainsi suivre l'évolution de l'abondance des composés dans chacune des fractions CPC. Ces résultats appuient l'observation constatée sur les chromatogrammes des fractions CPC avec l'extrait brut. En effet, la Figure 79 montre d'importantes intensités des composés (1,2,3,4,5,8,9,12 et 14) élués au début de l'analyse dans les fractions A et B, les composés 4 et 14 sont retrouvés à des abondances assez intenses dans les fractions C et D puis la teneur en composés polaires diminue drastiquement après la fraction E. Par exemple, le composé 1, acide plantagoguanidinique, un alcaloïde ayant un ion parent  $[M+H]^+$  de 226,156 sur le nœud est élué dans les fractions A et B avec une intensité proche de celle de l'extrait brut. On le retrouve ensuite à des concentrations très faibles, sur le chromatogramme, dans les fractions D, E et F. Enfin, ce composé disparaît dans les dernières fractions G et H.

Le cluster 1 quant à lui, regroupant des flavonoïdes complexes qui sont élués majoritairement dans les fractions B, C et F. Certains d'entre eux sont observés dans les dernières fractions G et H.

Contrairement aux composés polaires et moyennement polaires, les composés plus lipophiles formants les clusters 2-2', 3 présentés dans la Figure 81, sont non détectés dans les premières fractions. Ces composés commencent à être élués dans la fraction F et G en fin d'élution, et sont très abondants dans la fraction H de l'extraction. Les clusters 2 et 2' représentent les mêmes composés avec des schémas de fragmentation différents selon les adduits fragmentés. Les deux clusters peuvent être détectés, par exemple pour la fraction H, ou un seul des deux comme pour la fraction G.

L'objectif de la mise en réseaux moléculaires des fractions CPC est de mettre en évidence les familles moléculaires qui sont présentes dans les fractions CPC afin de corrélérer l'activité biologique avec la composition chimique des fractions. Cependant, comme l'activité biologique est répartie sur plusieurs fractions, on s'appuie sur les proportions et les intensités des composés pour expliquer l'activité. C'est pourquoi les réseaux moléculaires quantitatifs des fractions ont aidé à suivre les proportions des composés et mettent en évidence les molécules même lorsqu'elles sont présentes à de faibles concentrations.

## Activité antioxydante

Les deux fractions présentant une activité antioxydante sont les fractions A et B avec une activité plus importante pour la fraction B, tandis que toutes les autres fractions, ne présentent pas d'activité antioxydante. Cela signifie que les composés responsables de l'activité

antioxydante de l'extrait d'*Eclipta alba* sont principalement élués dans les fractions A et B avec une concentration plus importante dans la fraction B puisqu'elle est plus active. Ces mêmes composés devraient être d'une intensité très faible voire absents dans les autres fractions. Ceci permet d'éliminer tous les composés moyennement polaires à apolaires qui ne sont pas détectés dans la fraction active A et qui sont très abondants dans d'autres fractions inactives, et de se focaliser sur les composés polaires.

Parmi les composés polaires présentés dans la Figure 79, on peut remarquer que les composés 4 et 14 sont présents en quantités importantes dans les fractions C et D qui ne présentent pas d'activité antioxydante, ces deux composés ne semblent donc pas être directement responsables de l'activité antioxydante de l'extrait brut d'*Eclipta alba*.

Il reste donc 7 candidats majoritaires potentiels ; un alcaloïde (composé 1), deux flavonoïdes (composés 2 et 5), un acide phénolique (composé 3) et trois saponines glycosylées (composés 8, 9 et 12) : acide plantagoguanidinique, Lutéoline-7-O-glucoside, Apigénine sulfate, Acide 3,5-dicaféylquinique, Eclalbasaponines et Acide oléanique glycosylé respectivement. N'ayant pas continué le fractionnement ou la purification des molécules contenues dans ces fractions nous ne pouvons pas savoir précisément si une ou plusieurs molécules contribuent à cette activité. Compte tenu de leurs structures qui présentent des protons labiles, ils pourraient tous être impliqués dans cette activité même si les polyphénols sont davantage associés à cette activité.

L'apigénine sulfate (composé 5) pourrait contribuer à la plus forte activité antioxydante de la fraction B car tous les autres composés sont présents à des quantités équivalentes dans les fractions A et B, mis à part ce composé 5 qui est environ 2 fois plus abondant dans la fraction B (Figure 79).

### **Activité anti-collagénase et anti-tyrosinase**

L'ensemble des fractions CPC a présenté une activité anti-collagénase et anti-tyrosinase, certaines ont été plus actives que d'autres mais toutes ont présenté une activité, on peut donc supposer que des molécules présentant des potentiels anti-collagénase et anti-tyrosinase sont présentes dans toutes les fractions.

Les composés liés dans le cluster 1 (flavonoïdes apolaires) sont notamment retrouvés dans toutes les fractions à des concentrations différentes. On peut donc supposer que ces composés contribuent aux activités anti-collagénase et anti-tyrosinase.

La fraction H présente aussi une activité anti-collagénase similaire à celle de la fraction B. Les composés des cluster 2-2' et 3 pourraient donc également contribuer à cette activité.

Dans ce cas, il est compliqué de cibler celui ou ceux qui contribue(nt) le plus à l'activité d'autant qu'ils peuvent fonctionner en synergie et les intensités des composés ne nous aident pas à en dire davantage, c'est l'inconvénient lorsque toutes les fractions sont actives.

### **Activité anti-élastase**

Les fractions A, B et C n'ont pas présenté d'activité anti-élastase ce qui signifie que les composés polaires et moyennement polaires ainsi que les composés liés dans le cluster 1 ne sont pas directement responsables de l'activité anti-élastase puisqu'ils sont abondants dans ces fractions. Cependant, l'autre famille abondante de l'extrait d'*Eclipta alba*, soit les triterpènes pourrait être responsable de cette activité. En effet, les triterpènes étudiés dans cet extrait, contiennent majoritairement l'acide oléanique comme structure de base. Cette génine est retrouvée soit sous forme glycosylée éluée au début de l'analyse mais avec de faibles intensités (les eclalbasaponines) ou alors libre voire liés à des motifs lipophiles, se trouvant ainsi plus retenue sur la colonne et éluée à partir de 9,5 min. L'acide oléanique est un isomère de position de l'acide ursolique, un composé déjà connu pour son pouvoir d'inhibition de l'élastase et a fait l'objet de plusieurs études<sup>288</sup>. Par conséquent, les triterpènes apolaires, avec un groupe hydroxyle libre sur la position 28 (composés 28, 33, 47) élués dans les dernières fractions D, E, F, G et H pourrait expliquer l'activité anti-elastase de l'extrait éthanolique d'*Eclipta alba*. Enfin, la fraction H présente une forte activité donc les composés liés dans les clusters 2-2' et 3 peuvent également contribuer à sa capacité anti-élastase.

### **4) Viabilité cellulaire**

La toxicité de l'extrait pour les cellules a été évaluée afin de déterminer la concentration maximale d'extrait d'*Eclipta alba* qui était bien tolérée. Les résultats d'un essai XTT effectué 24 heures après le traitement (Figure 82) ont montré une courbe dose-réponse de la viabilité des cellules HaCaT traitées de 31,25 à 1000 µg/mL d'extrait d'*Eclipta alba*. Ces résultats montrent que jusqu'à 250 µg/mL, l'extrait d'*Eclipta alba* est bien toléré par les cellules. En effet, la toxicité cellulaire relative de l'extrait pour les cellules n'est pas statistiquement différente de celle des cellules non traitées. Des résultats similaires ont été observés 48 heures après le traitement cellulaire (données non présentées). Ainsi, les concentrations en extrait d'*Eclipta alba* utilisés lors des tests chimiques et enzymatiques sont tolérées par les cellules et n'induisent pas de mort cellulaire.

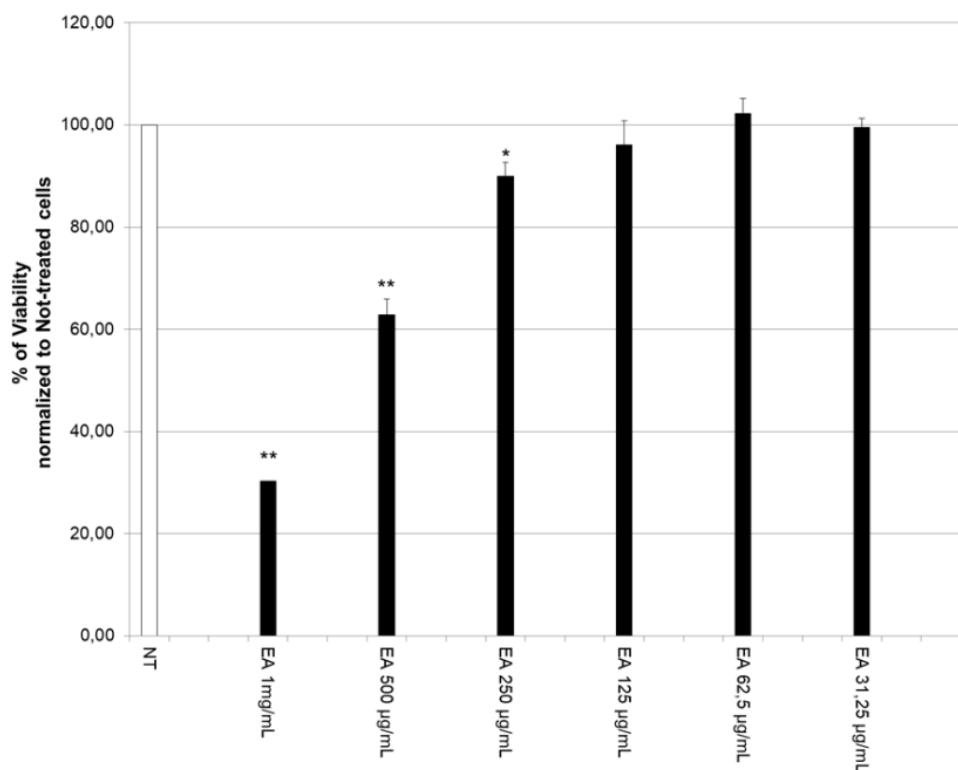
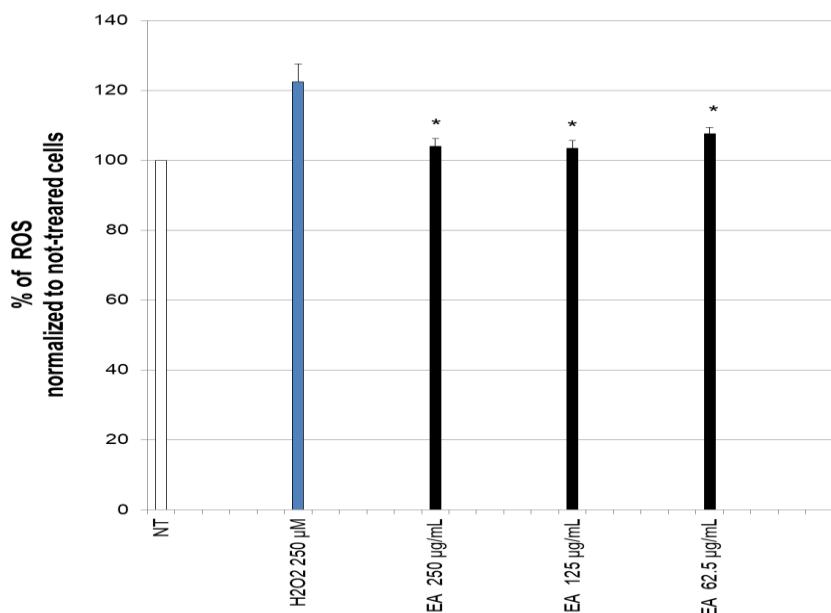


Figure 82 : Évaluation de la cytotoxicité XTT des cellules HaCaT à différentes concentrations d'extrait d'Eclipta alba après un traitement de 48 heures. Les valeurs sont exprimées en % de viabilité normalisée à 100 % des cellules non traitées. Les valeurs sont la moyenne  $\pm$  l'écart type d'une expérience réalisée au moins 3 fois en quintuplet, \* $p < 0,05$ , \*\* $p < 0,01$ , test T-student utilisant un test bilatéral non apparié

La propriété antioxydante de l'extrait d'Eclipta alba a ensuite été évaluée par rapport aux ROS produites dans les cellules après un traitement avec une solution de 250 µM de H<sub>2</sub>O<sub>2</sub> pendant une heure. Comme le montre la Figure 83, le traitement des cellules HaCaT avec H<sub>2</sub>O<sub>2</sub> a augmenté de 22 % la production de ROS dans les cellules. La pré-incubation des cellules avec une concentration élevée (250 µg/mL) à faible (62,5 µg/mL) d'extrait d'Eclipta alba a entraîné une réduction presque totale de la production de ROS dans les cellules. En effet, les pourcentages de production de ROS dans les cellules prétraitées avec l'extrait avant le traitement H<sub>2</sub>O<sub>2</sub> étaient similaires à ceux observés pour des cellules non traitées par H<sub>2</sub>O<sub>2</sub>.



*Figure 83 : Protection d'extrait d'Eclipta alba contre les espèces réactives de l'oxygène (ROS) dans les cellules HaCaT traitées avec 250 μM de H<sub>2</sub>O<sub>2</sub>. Les valeurs sont exprimées en % de la production de ROS normalisée à 100 % des cellules non traitées. Les valeurs sont la moyenne ± l'écart-type d'une expérience réalisée au moins 3 fois en quintuplet, \*p < 0,05, \*\*p<0,01, test T-student utilisant un test bilatéral non apparié*

Ces résultats montrent que l'extrait étudié est bien toléré par les cellules HaCaT et présente une forte propriété antioxydante protégeant les cellules contre la production de ROS induite dans les cellules par H<sub>2</sub>O<sub>2</sub>. Cet extrait pourrait donc être incorporé dans une formule à une concentration de 62,5 μg/mL présentant une efficacité et tolérée par les cellules.

#### IV. CONCLUSION

La combinaison du réseau moléculaire avec une approche de fractionnement CPC a été développée pour cibler et caractériser les composés bioactifs de deux extraits bruts ayant montré des propriétés biologiques d'intérêt cosmétique. D'une part l'extrait éthanolique d'*Artemisia annua* afin de mettre en évidence les molécules responsables de sa capacité antioxydante, et d'autre part l'extrait éthanolique d'*Eclipta alba* pour caractériser les molécules antioxydantes et anti-enzymatiques.

Les réseaux moléculaires conçus au moyen des spectres MS/MS des extraits bruts ont montré les principales familles moléculaires contenues dans les parties aériennes d'*Artemisia annua* et d'*Eclipta alba*. Les extraits complexes ont ensuite été divisés et simplifiés en fractions CPC avec différentes distributions moléculaires. Par la suite, les propriétés antioxydantes et anti-enzymatiques des fractions CPC ont été évaluées à l'aide de tests DPPH, ABTS, FRAP, CUPRAC, chélation du fer (II), tests d'inhibition de collagénase, d'élastase, de tyrosinase.

Les fractions actives ont été analysées spectrométrie de masse dans le but de caractériser les composés actifs. Pour *Artemisia annua*, une seule fraction antioxydante a été obtenue et analysée par FIA-HRMS mettant en évidence un composé principal d'intérêt appartenant à la famille des flavonoïdes. Tandis que pour *Eclipta alba*, plusieurs fractions antioxydantes et anti-enzymatiques ont été analysées par UHPLC-MS/HRMS et par les réseaux moléculaires quantitatifs mettant en évidence plusieurs composés bioactifs : des composés phénoliques et des flavonoïdes pour la capacité antioxydante de l'extrait, d'autres flavonoïdes plus complexes avec des propriétés anti-collagénase et anti-tyrosinase, ainsi que des triterpènes pour le pouvoir anti-élastase. Les composés ont été caractérisés à l'aide d'élucidation *de novo*, de réseaux moléculaires pour les similitudes spectrales et des bases de données de GNPS.

Ces applications nous ont notamment permis d'appréhender certaines limites de la plateforme GNPS. En effet, le système de quantification utilisé par GNPS est directement dépendant des paramètres d'acquisition appliqués lors de l'analyse data-dépendante des échantillons. La quantification par GNPS, étant le nombre de spectre fusionné en un seul spectre consensus grâce à l'outil MSClustering, est nettement influencée par le nombre d'isomère, le nombre de précurseurs fragmentés par scan, le seuil d'intensité, le temps d'exclusion dynamique, etc. Il n'est donc pas rare d'être confronté à des réseaux aberrants caractérisés par un très grand nombre de noeuds identiques partageant une grande similarité spectrale donc des scores cosine élevés.

Par conséquent, il est apparu nécessaire de trouver des alternatives à ce système de clustering inadapté à nos extraits. La première optimisation importante qui nécessite d'être effectuée, en fonction de l'échantillon à analyser et des familles moléculaires, est celle des

paramètres d'acquisition DDA afin de faire un compromis entre le taux et la qualité d'informations. Ensuite, la deuxième optimisation est celle de la méthode de prétraitement de données qui va permettre de remplacer l'outil MSClustering de GNPS apportant ainsi des informations plus fiables.

Le prétraitement des données d'*Artemisia annua* effectué par DataAnalysis en utilisant l'algorithme AutoMSn a permis de retirer le bruit de fond de l'analyse, d'exporter des spectres exploitables et de fusionner les spectres provenant d'un même composé, néanmoins, pour le cluster des terpènes par exemple, on retrouve une redondance d'information liant les adduits, les ions moléculaires et les ions fragments en source. De plus ce prétraitement a permis uniquement l'obtention d'un réseau qualitatif puisqu'il ne prend pas en considération les aires des pics ou les intensités et ne comprend pas de système d'alignement des analyses. Tous ces défauts ont été pris en compte afin de les améliorer.

Une autre méthode a été envisagée pour *Eclipta alba* en utilisant un nouveau logiciel de Bruker permettant d'améliorer davantage le prétraitement de données. En effet, il permet de superposer des échantillons, de quantifier correctement en prenant compte les intensités ou les aires (selon le paramétrage demandé par l'utilisateur), de fusionner les adduits (Na, K, Ca, HCOOH, etc.) et pertes de neutres souhaités (perte d'eau, de sucre, etc.). Néanmoins, la redondance d'information liée aux spectres MS/MS des ions qui proviennent de la fragmentation en source de l'ion parent reste difficile à gérer.

## **CHAPITRE 5 – CONCLUSION GÉNÉRALE ET PERSPECTIVES**

---



Les travaux décrits dans ce manuscrit traitent de l'élaboration, de la mise en place et l'optimisation de stratégies de recherche universelles et efficaces pour le criblage de biomarqueurs discriminant des extraits de plantes ou de molécules naturelles d'intérêt cosmétique.

Ces études ont démontré le grand intérêt de la présentation des données chromatographiques et spectrales d'un extrait de plante, seul ou en combinaison avec d'autres extraits ou des fractions simplifiées, sur une seule carte multi-informative pour explorer leur espace chimique. Cette stratégie de réseau moléculaire relativement facile à mettre en œuvre et applicable sur tout type d'échantillons de différents organismes (bactéries, champignons, organismes marins, plantes, etc.) représente une véritable avancée pour l'étude des mélanges complexes au sens large.

En effet, elle permet dans un premier temps d'avoir une vue globale de la composition chimique de l'échantillon, des composés structuralement proches associés en cluster, et de visualiser les composés majoritairement détectés en spectrométrie de masse. Dans un second temps, en se focalisant sur des composés, injectés sous forme d'étalon, décrits dans la littérature ou reconnus dans les bases de données, d'annoter ces composés sur le réseau, d'identifier les familles moléculaires puis les composés inconnus qui diffèrent des échantillons ainsi que des métabolites d'intérêt pour leur potentiel biologique ou leur nouveauté structurelle en propageant l'annotation de composés connus.

L'efficacité de cette stratégie d'analyse par réseau moléculaire a été exemplifiée sur différents extraits pour répondre à différentes problématiques. Elle a d'abord été utilisée seule pour étudier l'impact des conditions de culture de cals cellulaires de *Linum usitatissimum*, ensuite elle a été combinée à une approche métabolomique pour déterminer des biomarqueurs de croissance de racines de tomate *Solanum lycopersicum*, et enfin, elle a été associée au fractionnement et à l'évaluation d'activités biologiques pour le criblage de molécules actives d'*Artemisia annua* et d'*Eclipta alba*.

La première partie de notre travail a donc consisté en l'étude des paramètres influençant la création des réseaux moléculaires, à la comparaison de deux méthodes de prétraitement de données (DataAnalysis, MetaboScape), à l'évaluation de l'impact du réglage des paramètres d'acquisition et à l'utilisation d'un autre système de quantification plus fiable que MSClustering inclus dans la plateforme GNPS. Cette première étude nous a permis d'optimiser toutes les étapes permettant de créer les réseaux moléculaires d'extraits de cals cellulaires de *Linum usitatissimum*.

Les cals cellulaires de lin ont été élaborés à la lumière et à l'obscurité afin d'évaluer l'impact de la lumière sur la production des métabolites. Les réseaux moléculaires quantitatifs ont permis

d'analyser et de comparer les teneurs en composés extraits à partir des deux conditions de culture et ont mis en évidence les composés qui sont fortement impactés par le changement de luminosité. Cette étude a montré une influence bénéfique de la lumière sur l'accumulation des lignanes, métabolites principaux du lin, mais surtout sur celle de dérivés glycosylés non décrits dans la littérature, permettant ainsi de choisir la condition de culture en lumière pour les prochaines productions de cals de lin. Puisqu'ils permettent de quantifier tous les composés de l'échantillons indépendamment par l'utilisation de MetaboScape, les réseaux moléculaires quantitatifs ont apporté un réel avantage dans la comparaison d'extraits en l'absence de plusieurs répliques ne permettant pas une analyse comparative métabolomique ou statistique.

Dans une seconde partie, la stratégie de réseaux moléculaires a été combinée à une approche métabolomique dans le but de mettre en évidence et caractériser des biomarqueurs discriminant des groupes de racines de plants de tomate. Dans cette étude, les extraits provenant de 24 individus de racines de plants de tomate jeunes et matures ont été comparés par des analyses chimiométriques afin de mettre en évidence les molécules discriminantes de ces deux groupes d'échantillons. Puis les réseaux qualitatifs ont été élaborés dans le but de caractériser les biomarqueurs en propageant l'annotation établie grâce aux informations de déréPLICATION obtenues en utilisant les bases de données spectrales et la littérature. Cette combinaison d'approches a permis de mettre en évidence et d'identifier une vingtaine de métabolites impliqués lors du développement racinaire des plants de tomate. Certains d'entre eux sont induits à la hausse mais la majorité sont induits à la baisse après la maturité des plantes de tomates. En perspectives, la méthodologie développée sera appliquée à l'analyse de 800 extraits des parties aériennes et racinaires de plants de tomate matures (stabilisation des teneurs de métabolites) attaqués par différentes combinaisons de ravageurs dans le but de comprendre leur système de défense. Ainsi, nous pourrons proposer une liste de métabolites spécialisés biosynthétisés par les plants de tomate pour se protéger contre l'attaque des agents pathogènes, comme de potentiels composés d'intérêt écologique pouvant substituer les pesticides.

Dans une dernière partie, les réseaux moléculaires ont été combinés à un fractionnement des extraits bruts en fractions simplifiées et à l'évaluation de leurs activités biologiques dans le but de cibler et d'identifier des molécules bioactives des extraits d'*Artemisia annua* et d'*Eclipta alba*, plantes d'intérêt cosmétique cultivées en Centre-Val de Loire. Cette approche a permis de cibler et d'identifier i) un composé majoritaire de la famille de flavonoïdes, responsable de l'activité antioxydante de l'extrait d'*Artemisia annua*, ii) un ensemble de plusieurs composés phénoliques procurant à l'extrait d'*Eclipta alba* un pouvoir antioxydant iii) un groupe de flavonoïdes avec des similarités structurales conférant à l'extrait d'*Eclipta alba* une capacité à inhiber les enzymes collagénase et tyrosinase de la peau iv) des composés terpénoïques avec

une génine d'acide oléanolique permettant à l'extrait d'*Eclipta alba* d'inhiber l'enzyme élastase. Toutes ces propriétés biologiques ont un réel intérêt en cosmétique. Ces extraits qui n'ont pas présenté de toxicité cellulaire aux concentrations actives peuvent donc être intégrer dans les formules de produits cosmétiques à des concentrations permettant d'obtenir une efficacité tout en étant tolérés par les cellules de la peau.

Il serait intéressant de purifier ces composés bioactifs pour confirmer leurs structures en NMR, et de tester leurs activités biologiques dans le but confirmer qu'elles sont bien responsables des propriétés biologiques des extraits d'*Artemisia annua* et d'*Eclipta alba*. Néanmoins, ces extraits de plantes n'ayant pas montré une activité intense ni des activités plus originales sur d'autres cibles testées par nos partenaires du **CBM** durant le projet, les industriels collaborateurs de ValBioCosm' ne se projettent pas dans une industrialisation de produits à base de ces extraits de plantes. Toutefois la preuve de concept de faisabilité de cultiver en Région Centre-val de Loire des espèces végétales d'intérêt provenant d'autres pays en conservant un rendement en biomasse et en métabolites satisfaisant pour assurer un approvisionnement local, traçable et permettant de valoriser, mutualiser et rentabiliser des équipements couteux est démontrée et l'ensemble du processus pourra être appliqué à d'autres plantes d'intérêt.

En outre, l'ensemble de ces études réalisées au cours de cette thèse ont fait ressortir l'importance d'une sélection appropriée des paramètres d'acquisition en mode DDA sur le résultat global des réseaux moléculaires et ont permis de montrer que la stratégie d'analyse par réseaux moléculaires peut être utilisée sous différents angles en fonction de la problématique et de la question à laquelle on souhaite répondre.

Ces études ont également permis d'appréhender le logiciel MetaboScape et de mettre en évidence l'intérêt qu'il apporte lors du traitement de données. Nous avons également repéré certains points qui pourraient être améliorés tels que la création d'un format de fichier exportable directement sur Cytoscape pour la visualisation des résultats. En effet, MetaboScape permet également de comparer des spectres MS/MS des échantillons, entre eux et avec des spectres enregistrés dans les bases de données internes, et de créer des clusters de composés structuralement liés. Néanmoins, il ne permet pas la visualisation du réseau moléculaire globale, ce qui rend le passage par GNPS nécessaire pour obtenir des fichiers visualisables sur Cytoscape. Cette amélioration pourrait faire de ce logiciel une plateforme complète et indépendante pour l'étude des compositions chimiques des échantillons (analyses statistiques, déréplication, réseau moléculaire). Un intérêt supplémentaire est celui de créer un seul fichier de traitement pour plusieurs analyses superposées permettant de s'affranchir de la limitation du nombre d'échantillons à 6 sur la plateforme GNPS. De plus, ce logiciel pourrait également proposer des schémas de

fragmentation *in-silico*, il serait donc avantageux de mettre en place d'optimiser cette partie dans MetaboScape.

En parallèle des études discutées précédemment, d'autres matrices végétales ont également été étudiées, avec le même objectif de cibler des molécules naturelles d'intérêt cosmétique. Ces matrices sont : le chardon Marie dans le cadre du projet ValBioCosm' et les huiles d'olives Européennes dans le cadre d'un projet Européen OliveNet.

#### ➤ **Chardon Marie**

Dans le cadre du projet **ValBioCosm'**, une plateforme de criblage cellulaire a été développée au **CBM**, afin d'identifier de nouveaux extraits de plantes capables de réguler le miR-21 dans les kératinocytes humains. En effet, Le miR-21 joue un rôle majeur dans le maintien de l'homéostasie de la peau. Des études ont montré que l'activation de son expression et/ou de sa fonction contribue à la cicatrisation cutanée.

Nos collaborateurs du **CBM** ont montré qu'un extrait brut de graines de chardon Marie induit la fonction du miR-21 dans les kératinocytes placés dans un contexte de cicatrisation. De manière intéressante, l'activité de l'extrait du chardon Marie sur le miR-21 est corrélée à la silymarine, un mélange de 7 flavonolignanes et un flavonoïde contenu majoritairement dans les graines de chardon Marie. Nous nous sommes donc intéressés à la silymarine dans l'objectif de séparer ces flavonolignanes de structures très similaires afin d'évaluer leur activité cicatrisante sur la plateforme de criblage. Pour ce faire, un fractionnement CPC de la silymarine a été réalisé à l'**ICOA**. Les fractions obtenues avec différentes proportions en flavonolignanes ont été évaluées et une fraction contenant majoritairement un de ces flavonolignanes a montré une activité biologique comparable à celle de la silymarine sur le miR-21 et dépendant de la concentration en cette molécule.

En perspective, nous souhaitons comprendre le mécanisme d'action de cette molécule en suivant au niveau cellulaire le site d'action de la molécule et en étudiant les interactions développées entre la cellule et le flavonolignane. Pour cela, des études ciblées seront menées par spectrométrie de masse afin de comprendre son mécanisme et lieu de réaction (par exemple, intérieur ou extérieur du noyau de la cellule) en partenariat avec le **CBM**.

#### ➤ **Huiles d'olives Européennes**

L'objectif de cette étude qui s'inscrit dans le projet Européen OliveNet réalisée à ADSI (*Austrian Drug Screening Institute*, Innsbruck) était d'évaluer l'activité cosmétique des composés phénoliques de différentes huiles d'olive européennes afin d'intégrer un actif potentiel dans une formule cosmétique (huile ou extrait de composés phénoliques).

L'extrait entier étant généralement utilisé directement dans les formules de produits cosmétiques, nous avons cherché à identifier les molécules qui contribuent le plus à l'activité de l'extrait (activités antioxydante et anti-enzymatique) sans pour autant identifier tous les composés.

Un extrait de polyphénols d'huile d'olive obtenu par CPE a montré une forte activité antioxydante et anti-collagénase, il a été fractionné en HPLC semi-préparative. En perspective, il est prévu d'évaluer l'activité biologique des fractions et cibler les composés bioactifs, et de créer les réseaux moléculaires quantitatifs des extraits phénoliques des huiles d'olives étudiées afin de comparer les teneurs en composés bioactifs et pour mettre en évidence les variations chimiques entre les huiles.



# ANNEXES

---



## ANNEXE 1

### ➤ Analyses chromatographiques

#### ***Linum usitatissimum***

Les analytes ont été séparés sur une colonne Luna omega C18 (150 x 2,1 mm ; 1,6 µm, Phenomenex) à 40°C. La phase mobile à un débit de 500 µL/min était composée du solvant A : eau avec 0,1% d'acide formique, et du solvant B : acétonitrile avec 0,08% d'acide formique. Le volume d'injection était de 2 µL.

| %B   | min |
|------|-----|
| 0    | 3   |
| 0,2  | 3   |
| 12   | 45  |
| 14   | 90  |
| 15   | 90  |
| 15,5 | 3   |
| 18,5 | 3   |

Tableau 10 : Gradient utilisé pour l'analyse UHPLC-HRMS des extraits de cals cellulaires de *Linum usitatissimum* (B : acétonitrile avec 0,08% d'acide formique).

#### ***Solanum lycopersicum***

Les composés de racines de tomates ont été séparés sur une colonne Luna Omega C18 (150 x 2,1 mm ; 1,6 µm, Phenomenex) à 40°C. La phase mobile à un débit de 500 µL/min était composée d'eau + 0,1% d'acide formique (A) et d'acétonitrile + 0,08% d'acide formique (B) (Tableau 11). Les échantillons ont été analysés aléatoirement, le volume d'injection était de 1 µL.

| min  | % B  |
|------|------|
| 0    | 3    |
| 4,8  | 26,5 |
| 8,75 | 36,2 |
| 10,6 | 65   |
| 11,1 | 90   |
| 12,6 | 90   |
| 15,6 | 3    |

Tableau 11 : Gradient utilisé pour l'analyse UHPLC-HRMS des racines de tomate. (B : acétonitrile avec 0,08% d'acide formique).

***Artemisia annua***

La séparation a été réalisée à 60°C sur une colonne Fortis C18 (150 x 2,1 mm ; 1,7 µm) avec une élution par gradient à un débit de 300 µL/min de (A) eau et (B) acétonitrile : 2-propanol (50:50) tous les deux acidifiés avec 0,1% d'acide formique (Tableau 12). Le volume d'injection était de 0,2 µL d'extrait à 1 mg/mL d'éthanol.

| min  | % B |
|------|-----|
| 0    | 10  |
| 0,1  | 10  |
| 9    | 100 |
| 11,5 | 100 |
| 12   | 10  |
| 15   | 10  |

Tableau 12 : Gradient utilisé pour l'analyse UHPLC-HRMS d'*Artemisia annua*. (B : acétonitrile : 2-propanol à 50:50 avec 0,1% d'acide formique).

***Eclipta alba***

La séparation a été réalisée à 40°C sur une colonne Aeris WIDEPORÉ XB C18 (150 x 2,1 mm ; 3,6 µm, Phenomenex) avec une élution par gradient à un débit de 500 µL/min de (A) eau et (B) acétonitrile tous les deux acidifiés avec 0,1 et 0,08% d'acide formique respectivement (Tableau 13). Le volume d'injection était de 0,2 µL d'extrait ou de fraction à 1 mg/mL d'éthanol.

| min  | % B |
|------|-----|
| 0    | 30  |
| 0,6  | 30  |
| 11,4 | 90  |
| 12   | 100 |
| 18   | 100 |
| 18,5 | 30  |
| 20,5 | 30  |

Tableau 13: Gradient utilisé pour l'analyse UHPLC-HRMS d'*Eclipta alba*. (B : acétonitrile avec 0,08% d'acide formique)

## **ANNEXE 2**

### ➤ **Fractionnement CPC**

#### ***Artemisia annua***

Le fractionnement par CPC a été réalisé avec un système semi-préparative FCPC en utilisant une colonne de 50 mL (Rousselet Robatel Kromaton, Annonay, France), associé à une pompe LC-8A préparatoire (Shimadzu, Japon). L'échantillon a été introduit dans la colonne par une vanne d'injection Rheodyne haute pression à 6 ports (AIT, Houilles, France) équipée d'une boucle d'injection de 10 mL. Le système de solvant biphasique Arizona M (heptane : acétate d'éthyle : méthanol : eau) avec un rapport volumique de (5:6:5:6) a été préparé en mélangeant les différents solvants dans une ampoule à décanter. Les deux phases ont été séparées après agitation et équilibrage. La phase aqueuse a été choisie comme phase mobile et la phase organique comme phase stationnaire (mode descendant). La phase stationnaire a d'abord été introduite dans la colonne (9 mL/min, 500 rpm), puis la phase mobile a été pompée à travers la phase stationnaire (3 mL/min, 2000 rpm) jusqu'à la stabilisation du système biphasique (volume de la phase stationnaire restée dans la colonne est de 30 mL soit 60% de rétention). 116 mg de l'extrait éthanolique d'*Artemisia annua* ont été solubilisés dans 4 mL d'un mélange phase stationnaire : phase mobile (50:50) à une concentration de 29 mg/mL et ont été injectés après l'étape d'équilibrage. Les étapes d'élution (3 mL/min, 2000 rpm) et d'extrusion (9 mL/min, 500 rpm) ont été réalisées pour récupérer tous les composés de la colonne. Au total, 13 fractions ont été obtenues, évaporées et solubilisées dans l'éthanol.

#### ***Eclipta alba***

Le fractionnement par CPC a été réalisé sur un nouveau système de purification PLC 2250 (Gilson, Villiers-Le-Bel, France) acquis durant le projet ValBioCosm', en utilisant une colonne de 50 mL. L'instrument est équipé d'une boucle de 10 mL pour introduire l'échantillon dans la colonne, un système de pompage intégré, des détecteurs DAD et ELSD, un collecteur de fraction et un logiciel Gilson Glider Prep pour contrôler l'instrument et traiter les données. Le système de solvant biphasique Arizona M (heptane : acétate d'éthyle : méthanol : eau) avec un rapport volumique de (5:6:5:6) a été utilisé. La phase aqueuse a été choisie comme phase mobile et la phase organique comme phase stationnaire (mode descendant).

La phase stationnaire a d'abord été introduite dans la colonne (20 mL/min, 500 rpm), puis la phase mobile a été pompée à travers la phase stationnaire (5 mL/min, 2000 rpm) jusqu'à atteindre un stade d'équilibre (volume de la phase stationnaire restant dans la colonne après stabilisation du système est de 24 mL soit 48% de rétention). 175 mg de l'extrait brut d'*Eclipta alba* ont été solubilisés dans 6 mL d'un mélange phase stationnaire : phase mobile (50:50) à

une concentration de 29 mg/mL et ont été injectés après l'étape d'équilibrage. Le temps d'élution était de 73 min (5 mL/min, 2000 rpm) et le temps d'extrusion était de 5 min (10 mL/min, 500 rpm) afin de récupérer tous les composés restés dans la colonne et qui n'ont pas été élués. Au total, 8 fractions (A, B, C, D, E, F, G, H) ont été collectées. Elles ont été évaporées, et solubilisées dans l'éthanol à 1 mg/mL.

## **ANNEXE 3**

### ➤ **Activité antioxydante**

#### **Test d'activité de piégeage des radicaux DPPH**

L'activité de piégeage des radicaux libres de DPPH a été déterminée par la méthode décrite par Lee *et al.*<sup>289</sup> avec de légères modifications. Brièvement, 10 µL d'échantillon ont été mélangés à 190 µL de réactif DPPH à 60 µM dans l'éthanol et incubés 20 min dans le noir à température ambiante. L'absorbance a été enregistrée à l'aide d'un lecteur de microplaques (MULTISCAN GO, ThermoFisher, USA) à 516 nm.

#### **Test d'activité de piégeage des cations radicaux ABTS**

L'activité de piégeage des radicaux ABTS a été déterminée selon la méthode décrite par Tagliazucchi *et al.*<sup>290</sup> avec quelques modifications. La solution mère d'ABTS a été préparée en faisant réagir une solution aqueuse d'ABTS de 7 mM avec une solution de persulfate de potassium de 2,45 mM à volumes équivalents et en les laissant réagir dans le noir pendant 16 heures avec une agitation de 700 tr/min à température ambiante. Ensuite, le mélange a été dilué avec un mélange d'éthanol : eau à (25:75) dans les proportions volumiques suivantes (1:12,5). Ensuite, 10 µL d'échantillon ont été mélangés avec 190 µL du mélange dilué et l'absorbance a été enregistrée après 20 min d'incubation dans le noir à température ambiante à l'aide d'un lecteur de microplaques à 734 nm.

#### **Test FRAP de réduction des ions ferriques**

Le pouvoir antioxydant réduisant les ions ferriques a été réalisé selon la méthode décrite par Benzie et Strain<sup>291</sup> avec quelques modifications. Le réactif FRAP est composé de FeCl<sub>3</sub> (20 mM) et de TPTZ (10 mM, solubilisés dans du HCl 40 mM) avec un tampon d'acétate (300 mM) dans un rapport volumique de 1:1:10 (v:v:v). Par la suite, 10 µL d'échantillon ont été mélangés avec 190 µL de solution FRAP et incubés 20 min dans le noir à température ambiante. L'absorbance a été mesurée à l'aide d'un lecteur de microplaques à 590 nm.

#### **Test CUPRAC de réduction des ions cuivre**

L'activité antioxydante réduisant les ions cuivreux a été réalisée en réajustant la méthode décrite par Apak *et al.*<sup>292</sup>. En résumé, 190 µL de réactif CUPRAC, préparé en mélangeant le Cu(II) 10 mM avec le Néocuproine à 7,5 mM et le tampon d'acétate d'ammonium à 1 M dans un rapport volumique de 1:1:1 (v:v:v), ont été mélangés avec 10 µL d'échantillon. Ensuite, le mélange a été incubé 20 min dans l'obscurité à température ambiante et l'absorbance a été mesurée à l'aide d'un lecteur de microplaques à 450 nm.

## Test de chélation du fer (II)

La capacité des composés à chélater le Fe<sup>2+</sup> a été évaluée à l'aide d'une méthode décrite par Denis *et al.*<sup>293</sup> avec quelques modifications. 10 µL d'échantillon ont été mélangés avec 190 µL de réactif préparé en combinant 1 mM de FeCl<sub>3</sub> et 0,3 mM de ferrozine à un rapport de 1:1 (v:v). Après une incubation de 30 min dans l'obscurité à température ambiante, l'absorbance a été mesurée à l'aide d'un lecteur de microplaques à 562 nm

### ➤ Activité anti-enzymatique

## Test d'inhibition de la collagénase

L'extrait brut et les fractions CPC d'*Eclipta alba* ont été testés contre l'enzyme collagénase provenant de la bactérie (*Clostridium histolyticum*) en raison de sa disponibilité commerciale et de sa caractérisation biochimique et cinétique bien étudiée.

Le kit de dosage de la collagénase contient de la gélatine, un conjugué de fluorescéine et de gélatine si abondamment marqué par la fluorescéine que la fluorescence est éteinte. Ce substrat est digéré par la collagénase et induit l'augmentation de la fluorescence qui est proportionnelle à l'activité photolytique. Le chélateur de métaux et inhibiteur général des métalloprotéases, la 1,10-phénanthroline (inhibiteur Pht) est fourni dans le kit de dosage et sert de témoin positif pour optimiser les conditions de dépistage des inhibiteurs potentiels de la collagénase. Le kit de collagénase contient un tampon de réaction, pour que l'enzyme soit dans les conditions optimales.

Le standard 1,10-phénanthroline (inhibiteur de Pht) a été diluée dans le tampon de réaction pour obtenir une solution à 4 mM et 40 mM et a été utilisé comme témoin positif, tandis que le tampon de réaction comme témoin négatif. L'extrait d'*Eclipta alba* et les fractions CPC ont été solubilisés dans de l'éthanol à 1 mg/mL.

Dans chaque puits, 80 µL d'échantillon, de contrôle positif ou négatif ont été mélangés avec 20 µL de gélatine, le substrat enzymatique à 125 µg/mL, et 100 µL de solution enzymatique de collagénase à 0,4 U/ml, à l'exception des puits sans enzyme. Ensuite, la plaque a été incubée 2 heures à température ambiante dans l'obscurité avant la mesure de la fluorescence ( $\lambda_{exc}$  495 nm -  $\lambda_{em}$  515 nm) à l'aide de l'instrument VARIOSKAN LUX Microplate Reader (Thermo Fisher Scientific). Le logiciel SkanIt a été utilisé pour récupérer les résultats.

## Test d'inhibition de l'élastase

Les échantillons ont été testés contre l'élastase du pancréas porcin afin de découvrir les inhibiteurs potentiels de l'élastase dans les fractions. Le kit d'élastase contient une élastine soluble du ligament cervical bovin qui a été marquée avec un fluorophore de telle sorte que la fluorescence du conjugué soit éteinte. Le substrat non fluorescent peut être digéré par

l'élastase et entraîne l'augmentation de la fluorescence qui est proportionnelle à l'activité photolytique. Le kit comprend un inhibiteur sélectif de l'élastase, la N-méthoxysuccinyl-Ala-Ala-Pro-Val-chlorométhylcétone qui a servi de témoin positif pour optimiser les conditions de criblage des inhibiteurs potentiels de l'élastase. Un tampon de réaction est également inclus dans le kit pour optimiser le milieu de réaction enzymatique.

La N-méthoxysuccinyl-Ala-Ala-Pro-Val chlorométhylcétone a été utilisée comme témoin positif et diluée dans le tampon de réaction d'élastase pour obtenir une solution à 125 mM. Le tampon de réaction a été utilisé comme témoin négatif et les échantillons ont été solubilisés dans de l'éthanol à 1 mg/mL.

Dans chaque puits de la plaque, 50 µL d'échantillon ou de contrôle positif/négatif ont été mélangés à 50 µL de substrat enzymatique, qui est de l'élastine à 100 µg/mL et 100 µL d'enzyme élastase à 0,5 U/mL. La plaque a été incubée 2 heures à température ambiante dans l'obscurité, puis la fluorescence ( $\lambda_{exc}$  485 nm -  $\lambda_{em}$  535 nm) a été mesurée à l'aide de l'instrument VARIOSKAN LUX Microplate Reader (Thermo Fisher Scientific). Le logiciel SkanIt a été utilisé pour récupérer les résultats.

### **Test d'inhibition de la tyrosinase**

L'activité anti-tyrosinase a été évaluée en testant les échantillons contre l'enzyme tyrosinase de champignon. Elle catalyse l'oxydation de la tyrosine, produisant un chromophore détectable à  $\lambda = 490$  nm. En présence d'un inhibiteur de la tyrosinase, le taux d'oxydation du substrat L-Dopa est réduit et l'absorbance en est affectée. La glabridine, un inhibiteur de la tyrosinase, a été utilisée comme témoin positif.

La solution de tampon réactionnel (pH=6,8) a été préparée en mélangeant du NaOH à 2 M et du H<sub>3</sub>PO<sub>4</sub> à 1 M et de l'eau avec un rapport de (1:1,15:250). La Glabridine a été diluée dans le tampon de réaction de la tyrosinase pour obtenir des solutions à 100 µM et a été utilisée comme témoin positif, tandis que le tampon de réaction comme témoin négatif. Les échantillons ont été solubilisés dans de l'éthanol à 1 mg/mL.

Dans chaque puits, 40 µL d'échantillon, d'inhibiteur ou de contrôle positif/négatif ont été mélangés à 40 µL de solution enzymatique de tyrosinase à 125 U/mL. La plaque a été incubée pendant 5 minutes à température ambiante. Ensuite, 120 µL de solution de substrat de tyrosinase L-dopa à 1,65 mg/mL ont été ajoutés au mélange et la plaque a été incubée 30 min à 37°C dans l'obscurité. La lecture de l'absorbance ( $\lambda=490$  nm) a été effectuée à l'aide de l'instrument de lecture de microplaques VARIOSKAN LUX (Thermo Fisher Scientific).

## ➤ Activité cellulaire

### Test de viabilité cellulaire XTT

La lignée cellulaire immortalisée de kératinocytes humains HaCaT a été obtenue de l'ATCC. Les cellules ont été cultivées dans un milieu de *Dulbecco Eagle's media* (DMEM) modifié à haute teneur en glucose, complété par 10 % de sérum de veau fœtal inactivé par la chaleur, 50 IU/mL de pénicilline, 50 mg/mL de streptomycine et 2 mM de glutamine. La culture a été réalisée dans une atmosphère humidifiée contrôlée de 5 % de CO<sub>2</sub> à 37°C. Pour évaluer la toxicité relative d'extrait d'*Artemisia annua* et d'*Eclipta alba*, le kit de prolifération cellulaire a été utilisé conformément aux instructions du fabricant (XTT- Roche). Les cellules ont été chargées dans des plaques de 96 trous à une densité de 2x10<sup>4</sup> cellules/trou pour atteindre 90 % de confluence le lendemain. Ensuite, les extraits d'*Artemisia annua* et d'*Eclipta alba* dilués en série a été préparés dans un milieu sans sérum et incubés pendant 24, 48 ou 72 heures à 37°C, 5 % de CO<sub>2</sub>. Au moment indiqué après le traitement, la solution XTT à la concentration finale de 0,3 mg/mL a été ajoutée à chaque trou et incubée pendant 4 heures à 37°C pour la formation de *formazan* de couleur orange généré par des cellules viables. L'intensité de la coloration a été quantifiée en mesurant les valeurs d'absorbance de chaque trou à l'aide d'un spectrophotomètre Victor (PerkinElmer, Waltham, MA, USA) et réglé avec un filtre d'absorbance à 450 nm. Les données finales ont été exprimées en pourcentage de viabilité en normalisant la valeur d'absorbance détectée dans les cellules non traitées à la valeur arbitraire de 100 %. Tous les échantillons ont été déposés en n = 3 pour chaque série d'expériences et toute la série d'expériences a été répétée indépendamment au moins trois fois.

### Test de pouvoir antioxydant cellulaire

Pour évaluer les propriétés antioxydantes potentielles des extraits d'*Eclipta alba*, les cellules ont été traitées avec du H<sub>2</sub>O<sub>2</sub> pour produire des ROS dans les cellules et la sonde 2' 7' 2',7'-Dichlorofluorescin diacetate (DCF-DA Sigma Aldrich, France) a été utilisée pour quantifier la quantité de ROS produite dans les cellules. Les cellules ont été plaquées dans des plaques de 96 trous (2x10<sup>4</sup> cellules/trou) avant de charger le lendemain des dilutions en série d'extrait d'*Eclipta alba* préparées dans un milieu sans sérum. Les cellules ont été traitées avec les extraits pendant 6 heures à 5 % de CO<sub>2</sub>, 37°C. Ensuite, les monocouches de cellules ont été lavées deux fois dans du PBS avant de charger 200 µL de solution fraîche de H<sub>2</sub>O<sub>2</sub> à 250 µM pendant 60 min à 37°C et 5 % de CO<sub>2</sub>. Après un lavage rapide, 200 µL de la sonde DCF-DA utilisée à la concentration finale de 10 µM en PBS (Sigma Aldrich, France) ont été chargés dans les cellules et incubés pendant 30 min. Les valeurs de fluorescence dans chaque trou ont été quantifiées à ( $\lambda_{\text{ex}}=485$  nm,  $\lambda_{\text{em}}=535$  nm) à l'aide du spectrophotomètre Victor. Des

cellules non traitées ont été utilisées comme contrôle interne pour mesurer le niveau basal de la teneur en ROS dans les cellules. Les données finales ont été exprimées en pourcentage de la production de ROS en normalisant la valeur de fluorescence détectée dans les cellules non traitées à la valeur arbitraire de 100 %. Tous les échantillons ont été déposés en  $n = 3$  pour chaque série d'expériences et toute la série d'expériences a été répétée indépendamment au moins trois fois.

## **ANNEXE 4**

### **Culture en mésocosme**

Des plants de tomates, *Solanum lycopersicum* ; cultivar "San Marzano Nano", ont été cultivés en mésocosme ( $23\pm0.3^\circ\text{C}$  le jour,  $18\pm0.2^\circ\text{C}$  la nuit, HR :  $66\pm1\%$ , irradiance :  $663\pm48 \text{ W/m}^2$ ). Les semis d'une semaine ont été séparés par des pots et plantés dans un mélange de terre et de perlite et ont reçu une solution nutritive standard (pH=6, conductivité électrique (CE) = $1,6 \text{ mS/cm}$ , concentration en ions :  $620 \text{ mg/L } \text{NO}_3^-$ ,  $11 \text{ mg/L } \text{NH}_4^+$ ,  $170 \text{ mg/L } \text{H}_2\text{PO}_4^-$ ,  $115 \text{ mg/L } \text{SO}_4^{2-}$ ,  $180 \text{ mg/L } \text{K}^+$ ,  $212 \text{ mg/L } \text{Ca}^{2+}$ ,  $12 \text{ mg/L } \text{Mg}^{2+}$ ). Le 34<sup>e</sup> jour après le semis (plants de tomates à 3 feuilles complètes), les racines de tomates ont été échantillonnées et stockées à  $-80^\circ\text{C}$  en tant qu'échantillons au stade jeune ( $n = 12$  plants). De la même manière, des racines ont été prélevées sur 12 autres plants en tant qu'échantillons de vieux plants (plants en fleur avec 9 feuilles complètes) le 62<sup>e</sup> jour, également stockés à  $-80^\circ\text{C}$  (12 plants matures). Tous les échantillons ont été lyophilisés (Serail RP2V) pendant 48 heures pour éliminer l'humidité.

## RÉFÉRENCES BIBLIOGRAPHIQUES

---

- 1 Fraenkel, G. S. (1959). The Raison d’Etre of Secondary Plant Substances : These odd chemicals arose as a means of protecting plants from insects and now guide insects to food. *Science*, 129 (3361), 1466-1470. <https://doi.org/10.1126/science.129.3361.1466>.
- 2 Pichersky, E., Lewinsohn, E. (2011). Convergent evolution in plant specialized metabolism. *Annual Review of Plant Biology*, 62, 549-566. <https://doi.org/10.1146/annurev-arplant-042110-103814>.
- 3 Wink, M. (2003). Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. *Phytochemistry*, 64(1), 3-19. [https://doi.org/10.1016/S0031-9422\(03\)00300-5](https://doi.org/10.1016/S0031-9422(03)00300-5).
- 4 Kingston, D. G. I. (2011). Modern natural products drug discovery and its relevance to biodiversity conservation. *Journal of Natural Products*, 74(3), 496-511. <https://doi.org/10.1021/np100550t>.
- 5 Cantharide, cantharidine, grenouilles et légionnaires. (s. d.). Consulté 30 septembre 2020, à l'adresse <https://www.sciencepresse.qc.ca/blogue/2012/04/08/cantharide-cantharidine-grenouilles-legionnaires>.
- 6 <http://www7.inra.fr/opie-insectes/pdf/i150-goureau.pdf>. (Consulté le 02/08/2020).
- 7 Cosmétique : De la bave d'escargots pour lutter contre les rides. (s. d.). France 3 Nouvelle-Aquitaine. Consulté 30 septembre 2020, à l'adresse <https://france3-regions.francetvinfo.fr/nouvelle-aquitaine/2013/12/14/cosmetique-de-la-bave-d-escargots-pour-lutter-contre-les-rides-377669.html>.
- 8 Curare en Amazonie. (2020). In Wikipédia. [https://fr.wikipedia.org/w/index.php?title=Curare\\_en\\_Amazonie&oldid=174652409](https://fr.wikipedia.org/w/index.php?title=Curare_en_Amazonie&oldid=174652409).
- 9 La belladone, herbe au diable ou empoisonneuse au vertus médicinales. (s. d.). Binette & Jardin. Consulté 30 septembre 2020, à l'adresse <https://jardinage.lemonde.fr/dossier-1502-belladone.html>.
- 10 Kingston, D. G. I. (2011). Modern natural products drug discovery and its relevance to biodiversity conservation. *Journal of Natural Products*, 74(3), 496-511. <https://doi.org/10.1021/np100550t>.
- 11 Khan, R. A. (2018). Natural products chemistry : The emerging trends and prospective goals. *Saudi Pharmaceutical Journal*, 26(5), 739-753. <https://doi.org/10.1016/j.jsp.2018.02.015>.
- 12 Chapman Hall. Dictionary of Natural Products, <http://dnp.chemnetbase.com/>.
- 13 Grotewold, E. (2005). Plant metabolic diversity: a regulatory perspective. *Trends in Plant Science* 10: 57–62.
- 14 Lattanzio, V. (2013). Phenolic compounds : Introduction. In K. G. Ramawat & J.-M. Mérillon (Éds.), *Natural Products* (p. 1543-1580). Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-642-22144-6\\_57](https://doi.org/10.1007/978-3-642-22144-6_57).
- 15 Macheix, J.-J., Fleuriet, A., & Jay-Allemand, C. (2005). Les composés phénoliques des végétaux : Un exemple de métabolites secondaires d'importance économique. PPUR presses polytechniques.
- 16 Verpoorte, R., Choi, Y.H., and Kim, H.K. (2005). Ethnopharmacology and systems biology: a perfect holistic match. *Journal of Ethnopharmacology* 100: 53–56.
- 17 Tatti, S., Stockfleth, E., Beutner, K.R. et al. (2010). Polyphenon E (R): a new treatment for external anogenital warts. *British Journal of Dermatology* 162: 176–184.
- 18 Clarke, C. J., & Haselden, J. N. (2008). Metabolic profiling as a tool for understanding mechanisms of toxicity. *Toxicologic Pathology*, 36(1), 140-147. <https://doi.org/10.1177/0192623307310947>.

- 19 Van der Greef, J., van Wietmarschen, H., van Ommen, B., & Verheij, E. (2013). Looking back into the future : 30 years of metabolomics at TNO. *Mass Spectrometry Reviews*, 32(5), 399-415. <https://doi.org/10.1002/mas.21370>.
- 20 Horning, E. C., & Horning, M. G. (1971). Human metabolic profiles obtained by gc and gc/ms. *Journal of Chromatographic Science*, 9(3), 129-140. <https://doi.org/10.1093/chromsci/9.3.129>.
- 21 Pauling, L., Robinson, A. B., Teranishi, R., & Cary, P. (1971). Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography. *Proceedings of the National Academy of Sciences of the United States of America*, 68(10), 2374-2376. <https://doi.org/10.1073/pnas.68.10.2374>
- 22 Glauser, G., Guillarme, D., Grata, E., Boccard, J., Thiocone, A., Carrupt, P.-A., Veuthey, J.-L., Rudaz, S., & Wolfender, J.-L. (2008). Optimized liquid chromatography-mass spectrometry approach for the isolation of minor stress biomarkers in plant extracts and their identification by capillary nuclear magnetic resonance. *Journal of Chromatography A*, 1180(1-2), 90-98. <https://doi.org/10.1016/j.chroma.2007.12.021>.
- 23 Sumner, L. W., Mendes, P., & Dixon, R. A. (2003). Plant metabolomics : Large-scale phytochemistry in the functional genomics era. *Phytochemistry*, 62(6), 817-836. [https://doi.org/10.1016/s0031-9422\(02\)00708-2](https://doi.org/10.1016/s0031-9422(02)00708-2).
- 24 Dettmer, K., Aronov, P. A., & Hammock, B. D. (2007). Mass spectrometry-based metabolomics. *Mass Spectrometry Reviews*, 26(1), 51-78. <https://doi.org/10.1002/mas.20108>.
- 25 Leiss, K. A., Choi, Y. H., Verpoorte, R., & Klinkhamer, P. G. L. (2011). An overview of NMR-based metabolomics to identify secondary plant compounds involved in host plant resistance. *Phytochemistry Reviews*, 10(2), 205-216. <https://doi.org/10.1007/s11101-010-9175-z>.
- 26 Dunn, W. B. (2008). Current trends and future requirements for the mass spectrometric investigation of microbial, mammalian and plant metabolomes. *Physical Biology*, 5(1), 011001. <https://doi.org/10.1088/1478-3975/5/1/011001>.
- 27 De Vos, R. C. H., Moco, S., Lommen, A., Keurentjes, J. J. B., Bino, R. J., & Hall, R. D. (2007). Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nature Protocols*, 2(4), 778-791. <https://doi.org/10.1038/nprot.2007.95>.
- 28 Liseć, J., Schauer, N., Kopka, J., Willmitzer, L., & Fernie, A. R. (2006). Gas chromatography mass spectrometry-based metabolite profiling in plants. *Nature Protocols*, 1(1), 387-396. <https://doi.org/10.1038/nprot.2006.59>.
- 29 Ibáñez, C., Simó, C., García-Cañas, V., Cifuentes, A., & Castro-Puyana, M. (2013). Metabolomics, peptidomics and proteomics applications of capillary electrophoresis-mass spectrometry in Foodomics : A review. *Analytica Chimica Acta*, 802, 1-13. <https://doi.org/10.1016/j.aca.2013.07.042>.
- 30 Shulaev, V., & Isaac, G. (2018). Supercritical fluid chromatography coupled to mass spectrometry – A metabolomics perspective. *Journal of Chromatography B*, 1092, 499-505. <https://doi.org/10.1016/j.jchromb.2018.06.021>.
- 31 Wolfender, J.-L., Marti, G., Thomas, A., & Bertrand, S. (2015). Current approaches and challenges for the metabolite profiling of complex natural extracts. *Journal of Chromatography A*, 1382, 136-164. <https://doi.org/10.1016/j.chroma.2014.10.091>.
- 32 Wolfender, J.-L., Queiroz, E. F., & Hostettmann, K. (2005). Phytochemistry in the microgram domain—A LC-NMR perspective. *Magnetic Resonance in Chemistry: MRC*, 43(9), 697-709. <https://doi.org/10.1002/mrc.1631>.
- 33 Richardin P. La chromatographie. <http://www.culture.gouv.fr/culture/conservation/fr/methodes/chromato.htm>.
- 34 Caude M., Jardy A. Chromatographie en phase liquide – Théorie et méthodes de séparation. In Techniques de l'Ingénieur, 1994, P1 456.

- 35 Kingston, D. G. I. (1979). High performance liquid chromatography of natural products. *Journal of Natural Products*, 42(3), 237-260. <https://doi.org/10.1021/np50003a001>.
- 36 Natishan, T. K. (2004). Recent developments of achiral hplc methods in pharmaceuticals using various detection modes. *Journal of Liquid Chromatography & Related Technologies*, 27(7-9), 1237-1316. <https://doi.org/10.1081/JLC-120030603>.
- 37 Allwood, J. W., & Goodacre, R. (2010). An introduction to liquid chromatography-mass spectrometry instrumentation applied in plant metabolomic analyses. *Phytochemical Analysis: PCA*, 21(1), 33-47. <https://doi.org/10.1002/pca.1187>.
- 38 VEUTHEY, J.-L. (2016, septembre 10). Particules superficiellement poreuses en chromatographie liquide [Text]. Ref: TIP630WEB - « Techniques d'analyse ». <https://www.techniques-ingenieur.fr/base-documentaire/mesures-analyses-th1/chromatographie-et-techniques-separatives-42385210/particules-superficiellement-poreuses-en-chromatographie-liquide-in187/>.
- 39 Knox, J. H. (1977). Practical aspects of lc theory. *Journal of Chromatographic Science*, 15(9), 352-364. <https://doi.org/10.1093/chromsci/15.9.352>.
- 40 Fekete, S., Grand-Guillaume-Perrenoud, A., & Guillarme, D. (2014). Evolution and current trends in liquid and supercritical fluid chromatography. *Current Chromatography*, 1, 15.
- 41 De Villiers, A., Lestremau, F., Szucs, R., Gélébart, S., David, F., & Sandra, P. (2006). Evaluation of ultra performance liquid chromatography. Part I. Possibilities and limitations. *Journal of Chromatography. A*, 1127(1-2), 60-69. <https://doi.org/10.1016/j.chroma.2006.05.071>.
- 42 Guillarme, D., Casetta, C., Bicchi, C., & Veuthey, J.-L. (2010). High throughput qualitative analysis of polyphenols in tea samples by ultra-high pressure liquid chromatography coupled to UV and mass spectrometry detectors. *Journal of Chromatography. A*, 1217(44), 6882-6890. <https://doi.org/10.1016/j.chroma.2010.08.060>.
- 43 Wolfender, J.-L., Allard, P.-M., Kubo, M., & Queiroz, E. F. (2019). Metabolomics strategies for the dereplication of polyphenols and other metabolites in complex natural extracts. In H. Halbwirth, K. Stich, V. Cheynier, & S. Quideau (Éds.), *Recent Advances in Polyphenol Research* (p. 183-205). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781119427896.ch7>.
- 44 Wang, X., Sun, H., Zhang, A., Wang, P., & Han, Y. (2011). Ultra-performance liquid chromatography coupled to mass spectrometry as a sensitive and powerful technology for metabolomic studies. *Journal of Separation Science*, 34(24), 3451-3459. <https://doi.org/10.1002/jssc.201100333>.
- 45 Glish, G. L., & Vachet, R. W. (2003). The basics of mass spectrometry in the twenty-first century. *Nature Reviews. Drug Discovery*, 2(2), 140-150. <https://doi.org/10.1038/nrd1011>.
- 46 Hall, R. D. (2006). Plant metabolomics : From holistic hope, to hype, to hot topic. *The New Phytologist*, 169(3), 453-468. <https://doi.org/10.1111/j.1469-8137.2005.01632.x>.
- 47 Sanglier S. Cours de spectrométrie de masse. <http://www-esbs.u-strasbg.fr/notesdecours> /3eme-annee/gap/MmeSanglier05-10-05/Intro\_SS.pdf, 2005.
- 48 Colas C. (2007). Développement de méthodes physico-chimiques pour le contrôle de la médication par l'*Harpagophytum* et l'*Eleutherococcus*, principes actifs utilisés en phytothérapie équine. <https://pastel.archives-ouvertes.fr/pastel-00002923/document>.
- 49 Marti, G., Erb, M., Rudaz, S., Turlings, T., & Wolfender, J.-L. (2012). Search for low-molecular-weight biomarkers in plant tissues and seeds using metabolomics : Tools, strategies, and applications. In G. K. Agrawal & R. Rakwal (Éds.), *Seed Development : OMICS Technologies toward Improvement of Seed Quality and Crop Yield* (p. 305-341). Springer Netherlands. [https://doi.org/10.1007/978-94-007-4749-4\\_16](https://doi.org/10.1007/978-94-007-4749-4_16).

- 50 Hu, Q., Noll, R. J., Li, H., Makarov, A., Hardman, M., & Graham Cooks, R. (2005). The Orbitrap : A new mass spectrometer. *Journal of Mass Spectrometry: JMS*, 40(4), 430-443. <https://doi.org/10.1002/jms.856>.
- 51 Feng, X., & Siegel, M. M. (2007). FTICR-MS applications for the structure determination of natural products. *Analytical and Bioanalytical Chemistry*, 389(5), 1341-1363. <https://doi.org/10.1007/s00216-007-1468-8>.
- 52 Tsybin, Y. O., Fornelli, L., Kozhinov, A. N., Vorobyev, A., & Miladinovic, S. M. (2011). High-resolution and tandem mass spectrometry – the indispensable tools of the xxi century. *CHIMIA International Journal for Chemistry*, 65(9), 641-645. <https://doi.org/10.2533/chimia.2011.641>.
- 53 Guillarme, D., Schappler, J., Rudaz, S., & Veuthey, J.-L. (2010). Coupling ultra-high-pressure liquid chromatography with mass spectrometry. *TrAC Trends in Analytical Chemistry*, 29(1), 15-27. <https://doi.org/10.1016/j.trac.2009.09.008>.
- 54 Benton, H. P., Ivanisevic, J., Mahieu, N. G., Kurczy, M. E., Johnson, C. H., Franco, L., Rinehart, D., Valentine, E., Gowda, H., Ubhi, B. K., Tautenhahn, R., Gieschen, A., Fields, M. W., Patti, G. J., & Siuzdak, G. (2015). Autonomous metabolomics for rapid metabolite identification in global profiling. *Analytical Chemistry*, 87(2), 884-891. <https://doi.org/10.1021/ac5025649>.
- 55 Broeckling, C. D., Hoyes, E., Richardson, K., Brown, J. M., & Prenni, J. E. (2018). Comprehensive tandem-mass-spectrometry coverage of complex samples enabled by data-set-dependent acquisition. *Analytical Chemistry*, 90(13), 8020-8027. <https://doi.org/10.1021/acs.analchem.8b00929>.
- 56 Kind, T., Tsugawa, H., Cajka, T., Ma, Y., Lai, Z., Mehta, S. S., Wohlgemuth, G., Barupal, D. K., Showalter, M. R., Arita, M., & Fiehn, O. (2018). Identification of small molecules using accurate mass MS/MS search. *Mass Spectrometry Reviews*, 37(4), 513-532. <https://doi.org/10.1002/mas.21535>.
- 57 Ludwig, C., Gillet, L., Rosenberger, G., Amon, S., Collins, B. C., & Aebersold, R. (2018). Data-independent acquisition-based SWATH - MS for quantitative proteomics : A tutorial. *Molecular Systems Biology*, 14(8). <https://doi.org/10.15252/msb.20178126>.
- 58 Li, H., Cai, Y., Guo, Y., Chen, F., & Zhu, Z.-J. (2016). Metdia : Targeted metabolite extraction of multiplexed ms/ms spectra generated by data-independent acquisition. *Analytical Chemistry*, 88(17), 8757-8764. <https://doi.org/10.1021/acs.analchem.6b02122>.
- 59 Révész, Á., Rokob, T. A., Jeanne Dit Fouque, D., Turiák, L., Memboeuf, A., Vékey, K., & Drahos, L. (2018). Selection of collision energies in proteomics mass spectrometry experiments for best peptide identification : Study of mascot score energy dependence reveals double optimum. *Journal of Proteome Research*, 17(5), 1898-1906. <https://doi.org/10.1021/acs.jproteome.7b00912>.
- 60 Eugster, P. J., Guillarme, D., Rudaz, S., Veuthey, J.-L., Carrupt, P.-A., & Wolfender, J.-L. (2011). Ultra high pressure liquid chromatography for crude plant extract profiling. *Journal of AOAC International*, 94(1), 51-70.
- 61 Wilson, I. D., Nicholson, J. K., Castro-Perez, J., Granger, J. H., Johnson, K. A., Smith, B. W., & Plumb, R. S. (2005). High resolution « ultra performance » liquid chromatography coupled to oa-TOF mass spectrometry as a tool for differential metabolic pathway profiling in functional genomic studies. *Journal of Proteome Research*, 4(2), 591-598. <https://doi.org/10.1021/pr049769r>.
- 62 Grata, E., Guillarme, D., Glauser, G., Boccard, J., Carrupt, P.-A., Veuthey, J.-L., Rudaz, S., & Wolfender, J.-L. (2009). Metabolite profiling of plant extracts by ultra-high-pressure liquid chromatography at elevated temperature coupled to time-of-flight mass spectrometry. *Journal of Chromatography. A*, 1216(30), 5660-5668. <https://doi.org/10.1016/j.chroma.2009.05.069>.
- 63 Grata, E., Boccard, J., Guillarme, D., Glauser, G., Carrupt, P.-A., Farmer, E. E., Wolfender, J.-L., & Rudaz, S. (2008). UPLC-TOF-MS for plant metabolomics : A sequential approach for wound marker analysis in *Arabidopsis thaliana*. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences*, 871(2), 261-270. <https://doi.org/10.1016/j.jchromb.2008.04.021>.

- 64 Wolfender, J.-L. (2009). HPLC in natural product analysis : The detection issue. *Planta Medica*, 75(7), 719-734. <https://doi.org/10.1055/s-0028-1088393>.
- 65 Creek, D. J., Dunn, W. B., Fiehn, O., Griffin, J. L., Hall, R. D., Lei, Z., Mistrik, R., Neumann, S., Schymanski, E. L., Sumner, L. W., Trengove, R., & Wolfender, J.-L. (2014). Metabolite identification : Are you sure? And how do your peers gauge your confidence? *Metabolomics*, 10(3), 350-353. <https://doi.org/10.1007/s11306-014-0656-8>.
- 66 Bitzer, J., Köpcke, B., Stadler, M., Hellwig, V., Ju, Y.-M., Seip, S., & Henkel, T. (2007). Accelerated dereplication of natural products, supported by reference libraries. *CHIMIA International Journal for Chemistry*, 61(6), 332-338. <https://doi.org/10.2533/chimia.2007.332>.
- 67 Bakiri, A., Plainchont, B., de Paulo Emerenciano, V., Reynaud, R., Hubert, J., Renault, J.-H., & Nuzillard, J.-M. (2017). Computer-aided dereplication and structure elucidation of natural products at the university of reims. *Molecular Informatics*, 36(10), 1700027. <https://doi.org/10.1002/minf.201700027>.
- 68 Blondeau, S., Do, Q. T., Scior, T., Bernard, P., & Morin-Allory, L. (2010). Reverse pharmacognosy : Another way to harness the generosity of nature. *Current Pharmaceutical Design*, 16(15), 1682-1696. <https://doi.org/10.2174/138161210791164036>.
- 69 Jeannerat, D. (2017). Human- and computer-accessible 2D correlation data for a more reliable structure determination of organic compounds. Future roles of researchers, software developers, spectrometer managers, journal editors, reviewers, publisher and database managers toward: Human- and computer-accessible 2D correlation data. *Magnetic Resonance in Chemistry*, 55(1), 7-14. <https://doi.org/10.1002/mrc.4527>.
- 70 Cordell, G. A., Beecher, C. W. W., Douglas Kinghorn, A., Pezzuto, J. M., Constant, H. L., Chai, H.-B., Fang, L., Seo, E.-K., Long, L., Cui, B., & Slowing-Barillas, K. (1996). The dereplication of plant-derived natural products. In Atta-ur-Rahman (Ed.), *Studies in Natural Products Chemistry* (Vol. 19, p. 749-791). Elsevier. [https://doi.org/10.1016/S1572-5995\(96\)80018-4](https://doi.org/10.1016/S1572-5995(96)80018-4).
- 71 Farnsworth, N. R. (1966). Biological and phytochemical screening of plants. *Journal of Pharmaceutical Sciences*, 55(3), 225-276. <https://doi.org/10.1002/jps.2600550302>.
- 72 Marini-Bettolo, G. B., Nicoletti, M., Patamia, M., Galeffi, C., & Messana, I. (1981). Plant screening by chemical and chromatographic procedures under field conditions. *Journal of Chromatography A*, 213(1), 113-127. [https://doi.org/10.1016/S0021-9673\(00\)80639-1](https://doi.org/10.1016/S0021-9673(00)80639-1).
- 73 Rios, J. L., Recio, M. C., & Villar, A. (1988). Screening methods for natural products with antimicrobial activity : A review of the literature. *Journal of Ethnopharmacology*, 23(2), 127-149. [https://doi.org/10.1016/0378-8741\(88\)90001-3](https://doi.org/10.1016/0378-8741(88)90001-3).
- 74 Beutler, J. A., Alvarado, A. B., Schaufelberger, D. E., Andrews, P., & McCloud, T. G. (1990). Dereplication of phorbol bioactives : Lyngbya majuscula and Croton cuneatus. *Journal of Natural Products*, 53(4), 867-874. <https://doi.org/10.1021/np50070a014>.
- 75 Hubert, J., Nuzillard, J.-M., & Renault, J.-H. (2017). Dereplication strategies in natural product research : How many tools and methodologies behind the same concept? *Phytochemistry Reviews*, 16(1), 55-95. <https://doi.org/10.1007/s11101-015-9448-7>.
- 76 Nguyen, D. H., Nguyen, C. H., & Mamitsuka, H. (2019). Recent advances and prospects of computational methods for metabolite identification : A review with emphasis on machine learning approaches. *Briefings in Bioinformatics*, 20(6), 2028-2043. <https://doi.org/10.1093/bib/bby066>.
- 77 Dark matter. (2008). *Nature*, 455(7213), 698-698. <https://doi.org/10.1038/455698a>.
- 78 Da Silva, R. R., Dorrestein, P. C., & Quinn, R. A. (2015). Illuminating the dark matter in metabolomics. *Proceedings of the National Academy of Sciences of the United States of America*, 112(41), 12549-12550. <https://doi.org/10.1073/pnas.1516878112>.

- 79 Wolfender, J.-L., Nuzillard, J.-M., van der Hooft, J. J. J., Renault, J.-H., & Bertrand, S. (2019). Accelerating metabolite identification in natural product research : Toward an ideal combination of liquid chromatography high-resolution tandem mass spectrometry and nmr profiling, in silico databases, and chemometrics. *Analytical Chemistry*, 91(1), 704-742. <https://doi.org/10.1021/acs.analchem.8b05112>.
- 80 Kuhl, C., Tautenhahn, R., Böttcher, C., Larson, T. R., & Neumann, S. (2012). CAMERA : An integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Analytical Chemistry*, 84(1), 283-289. <https://doi.org/10.1021/ac202450g>.
- 81 Scheubert, K., Hufsky, F., & Böcker, S. (2013). Computational mass spectrometry for small molecules. *Journal of Cheminformatics*, 5(1), 12. <https://doi.org/10.1186/1758-2946-5-12>.
- 82 Kind, T., & Fiehn, O. (2006). Metabolomic database annotations via query of elemental compositions : Mass accuracy is insufficient even at less than 1 ppm. *BMC Bioinformatics*, 7, 234. <https://doi.org/10.1186/1471-2105-7-234>.
- 83 Kind, T., & Fiehn, O. (2007). Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinformatics*, 8, 105. <https://doi.org/10.1186/1471-2105-8-105>.
- 84 Wolfender, J.-L., Marti, G., Thomas, A., & Bertrand, S. (2015). Current approaches and challenges for the metabolite profiling of complex natural extracts. *Journal of Chromatography A*, 1382, 136-164. <https://doi.org/10.1016/j.chroma.2014.10.091>.
- 85 Klitgaard, A., Iversen, A., Andersen, M. R., Larsen, T. O., Frisvad, J. C., & Nielsen, K. F. (2014). Aggressive dereplication using UHPLC–DAD–QTOF : Screening extracts for up to 3000 fungal secondary metabolites. *Analytical and Bioanalytical Chemistry*, 406(7), 1933-1943. <https://doi.org/10.1007/s00216-013-7582-x>.
- 86 Yang, J., Liang, Q., Wang, M., Jeffries, C., Smithson, D., Tu, Y., Boulos, N., Jacob, M. R., Shelat, A. A., Wu, Y., Ravu, R. R., Gilbertson, R., Avery, M. A., Khan, I. A., Walker, L. A., Guy, R. K., & Li, X.-C. (2014). UPLC-MS-ELSD-PDA as a powerful dereplication tool to facilitate compound identification from small-molecule natural product libraries. *Journal of Natural Products*, 77(4), 902-909. <https://doi.org/10.1021/np4009706>.
- 87 Kildgaard, S., Mansson, M., Dosen, I., Klitgaard, A., Frisvad, J. C., Larsen, T. O., & Nielsen, K. F. (2014). Accurate dereplication of bioactive secondary metabolites from marine-derived fungi by UHPLC-DAD-QTOFMS and a MS/HRMS library. *Marine Drugs*, 12(6), 3681-3705. <https://doi.org/10.3390/md12063681>.
- 88 Bertrand, S., Azzollini, A., Schumpp, O., Bohni, N., Schrenzel, J., Monod, M., Gindro, K., & Wolfender, J.-L. (2014). Multi-well fungal co-culture for de novo metabolite-induction in time-series studies based on untargeted metabolomics. *Molecular BioSystems*, 10(9), 2289-2298. <https://doi.org/10.1039/c4mb00223g>.
- 89 Funari, C. S., Eugster, P. J., Martel, S., Carrupt, P.-A., Wolfender, J.-L., & Silva, D. H. S. (2012). High resolution ultra high pressure liquid chromatography-time-of-flight mass spectrometry dereplication strategy for the metabolite profiling of Brazilian Lippia species. *Journal of Chromatography. A*, 1259, 167-178. <https://doi.org/10.1016/j.chroma.2012.03.069>.
- 90 Pluskal, T., Castillo, S., Villar-Briones, A., & Orešič, M. (2010). MZmine 2 : Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics*, 11(1), 395. <https://doi.org/10.1186/1471-2105-11-395>.
- 91 Allard, P.-M., Genta-Jouve, G., & Wolfender, J.-L. (2017). Deep metabolome annotation in natural products research : Towards a virtuous cycle in metabolite identification. *Current Opinion in Chemical Biology*, 36, 40-49. <https://doi.org/10.1016/j.cbpa.2016.12.022>.
- 92 Shahaf, N., Rogachev, I., Heinig, U., Meir, S., Malitsky, S., Battat, M., Wyner, H., Zheng, S., Wehrens, R., & Aharoni, A. (2016). The WEIZMASS spectral library for high-confidence metabolite identification. *Nature Communications*, 7, 12423. <https://doi.org/10.1038/ncomms12423>.

- 93 Sawada, Y., Nakabayashi, R., Yamada, Y., Suzuki, M., Sato, M., Sakata, A., Akiyama, K., Sakurai, T., Matsuda, F., Aoki, T., Hirai, M. Y., & Saito, K. (2012). RIKEN tandem mass spectral database (Respect) for phytochemicals: A plant-specific MS/MS-based data resource and database. *Phytochemistry*, 82, 38-45. <https://doi.org/10.1016/j.phytochem.2012.07.007>.
- 94 Wang, M., Carver, J. J., Phelan, V. V., Sanchez, L. M., Garg, N., Peng, Y., Nguyen, D. D., Watrous, J., Kapono, C. A., Luzzatto-Knaan, T., Porto, C., Bouslimani, A., Melnik, A. V., Meehan, M. J., Liu, W.-T., Crüsemann, M., Boudreau, P. D., Esquenazi, E., Sandoval-Calderón, M., Bandeira, N. (2016). Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nature Biotechnology*, 34(8), 828-837. <https://doi.org/10.1038/nbt.3597>.
- 95 Schymanski, E. L., Singer, H. P., Slobodnik, J., Ipolyi, I. M., Oswald, P., Krauss, M., Schulze, T., Haglund, P., Letzel, T., Grosse, S., Thomaidis, N. S., Bletsou, A., Zwiener, C., Ibáñez, M., Portolés, T., de Boer, R., Reid, M. J., Onghena, M., Kunkel, U., Hollender, J. (2015). Non-target screening with high-resolution mass spectrometry : Critical review using a collaborative trial on water analysis. *Analytical and Bioanalytical Chemistry*, 407(21), 6237-6255. <https://doi.org/10.1007/s00216-015-8681-7>.
- 96 Cabral, R. S., Allard, P.-M., Marcourt, L., Young, M. C. M., Queiroz, E. F., & Wolfender, J.-L. (2016). Targeted isolation of indolopyridoquinazoline alkaloids from *conchocarpus fontanesianus* based on molecular networks. *Journal of Natural Products*, 79(9), 2270-2278. <https://doi.org/10.1021/acs.jnatprod.6b00379>.
- 97 Laatsch, H. AntiBase (2013): The Natural Compound Identifier, Upgrade; Wiley-VCH, 2013.
- 98 Asakura, K. Yuki Gosei Kagaku Kyokaishi (2015), 73 (12), 1247–1252.
- 99 Shinbo, Y., Nakamura, Y., Altaf-Ul-Amin, M., Asahi, H., Kurokawa, K., Arita, M., Saito, K., Ohta, D., Shibata, D., & Kanaya, S. (2006). Knapsack : A comprehensive species-metabolite relationship database. In K. Saito, R. A. Dixon, & L. Willmitzer (Éds.), *Plant Metabolomics* (p. 165-181). Springer. [https://doi.org/10.1007/3-540-29782-0\\_13](https://doi.org/10.1007/3-540-29782-0_13).
- 100 MarinLit, a Database of the Marine Natural Products Literature; Dabb, S., Potter, H., Eds.; Royal Society of Chemistry, 2014.
- 101 Allard, P.-M., Péresse, T., Bisson, J., Gindro, K., Marcourt, L., Pham, V. C., Roussi, F., Litaudon, M., & Wolfender, J.-L. (2016). Integration of molecular networking and in-silico ms/ms fragmentation for natural products dereplication. *Analytical Chemistry*, 88(6), 3317-3323. <https://doi.org/10.1021/acs.analchem.5b04804>.
- 102 Demuth, W., Karlovits, M., & Varmuza, K. (2004). Spectral similarity versus structural similarity : Mass spectrometry. *Analytica Chimica Acta*, 516 (1-2), 75-85. <https://doi.org/10.1016/j.aca.2004.04.014>.
- 103 Schollée, J. E., Schymanski, E. L., Stravs, M. A., Gulde, R., Thomaidis, N. S., & Hollender, J. (2017). Similarity of high-resolution tandem mass spectrometry spectra of structurally related micropollutants and transformation products. *Journal of the American Society for Mass Spectrometry*, 28(12), 2692-2704. <https://doi.org/10.1007/s13361-017-1797-6>.
- 104 Cuyckens, F., & Claeys, M. (2004). Mass spectrometry in the structural analysis of flavonoids. *Journal of Mass Spectrometry: JMS*, 39(1), 1-15. <https://doi.org/10.1002/jms.585>.
- 105 Kerzaon, I., Pouchus, Y. F., Monteau, F., Le Bizec, B., Nourrisson, M.-R., Biard, J.-F., & Grovel, O. (2009). Structural investigation and elucidation of new communesins from a marine-derived *Penicillium expansum* Link by liquid chromatography/electrospray ionization mass spectrometry. *Rapid Communications in Mass Spectrometry: RCM*, 23(24), 3928-3938. <https://doi.org/10.1002/rcm.4330>.
- 106 Watrous, J., Roach, P., Alexandrov, T., Heath, B. S., Yang, J. Y., Kersten, R. D., van der Voort, M., Pogliano, K., Gross, H., Raaijmakers, J. M., Moore, B. S., Laskin, J., Bandeira, N., & Dorrestein, P. C. (2012). Mass spectral molecular networking of living microbial colonies. *Proceedings of the National Academy of Sciences of the United States of America*, 109(26), E1743-1752. <https://doi.org/10.1073/pnas.1203689109>.

- 107 Naake, T., & Gaquerel, E. (2017). MetCirc : Navigating mass spectral similarity in high-resolution MS/MS metabolomics data. *Bioinformatics* (Oxford, England), 33(15), 2419-2420. <https://doi.org/10.1093/bioinformatics/btx159>.
- 108 Dayringer, H. E., Pesyna, G. M., Venkataraghavan, R., & McLafferty, F. W. (1976). Computer-aided interpretation of mass spectra. Information on substructural probabilities from stirs. *Organic Mass Spectrometry*, 11(5), 529-542. <https://doi.org/10.1002/oms.1210110512>.
- 109 Depke, T., Franke, R., & Brönstrup, M. (2017). Clustering of MS/MS spectra using unsupervised methods to aid the identification of secondary metabolites from *Pseudomonas aeruginosa*. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences*, 1071, 19-28. <https://doi.org/10.1016/j.jchromb.2017.06.002>.
- 110 Treutler, H., Tsugawa, H., Porzel, A., Gorzolka, K., Tissier, A., Neumann, S., & Balcke, G. U. (2016). Discovering regulated metabolite families in untargeted metabolomics studies. *Analytical Chemistry*, 88(16), 8082-8090. <https://doi.org/10.1021/acs.analchem.6b01569>.
- 111 Van der Hooft, J. J. J., Wandy, J., Barrett, M. P., Burgess, K. E. V., & Rogers, S. (2016). Topic modeling for untargeted substructure exploration in metabolomics. *Proceedings of the National Academy of Sciences of the United States of America*, 113(48), 13738-13743. <https://doi.org/10.1073/pnas.1608041113>.
- 112 Nothias-Scaglia, L.-F., Esposito, M., Costa, J., Paolini, J., Touboul, D., & Litaudon, M. (2015). Les réseaux moléculaires, une approche bio-informatique globale pour interpréter les données de spectrométrie de masse tandem. *Spectra Analyse*, 307, 73.
- 113 Quinn, R. A., Nothias, L.-F., Vining, O., Meehan, M., Esquenazi, E., & Dorrestein, P. C. (2017). Molecular networking as a drug discovery, drug metabolism, and precision medicine strategy. *Trends in Pharmacological Sciences*, 38(2), 143-154. <https://doi.org/10.1016/j.tips.2016.10.011>.
- 114 Olivon, F., Allard, P.-M., Koval, A., Righi, D., Genta-Jouve, G., Neyts, J., Apel, C., Pannecouque, C., Nothias, L.-F., Cachet, X., Marcourt, L., Roussi, F., Katanaev, V. L., Touboul, D., Wolfender, J.-L., & Litaudon, M. (2017). Bioactive natural products prioritization using massive multi-informational molecular networks. *ACS Chemical Biology*, 12(10), 2644-2651. <https://doi.org/10.1021/acscchembio.7b00413>.
- 115 Frank, A. M., Monroe, M. E., Shah, A. R., Carver, J. J., Bandeira, N. F., Moore, R. J., Anderson, G. A., Smith, R. D., & Pevzner, P. A. (2011). Spectral archives : Extending spectral libraries to analyze both identified and unidentified spectra. *Nature methods*, 8(7), 587-591. <https://doi.org/10.1038/nmeth.1609>.
- 116 Frank, A. M., Bandeira, N., Shen, Z., Tanner, S., Briggs, S. P., Smith, R. D., & Pevzner, P. A. (2008). Clustering millions of tandem mass spectra. *Journal of Proteome Research*, 7(1), 113-122. <https://doi.org/10.1021/pr070361e>.
- 117 Cosine similarity—An overview | sciencedirect topics. (s. d.). Consulté 30 septembre 2020, à l'adresse <https://www.sciencedirect.com/topics/computer-science/cosine-similarity>.
- 118 Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape : A software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11), 2498-2504. <https://doi.org/10.1101/gr.1239303>.
- 119 Nguyen, D. D., Wu, C.-H., Moree, W. J., Lamsa, A., Medema, M. H., Zhao, X., Gavilan, R. G., Aparicio, M., Atencio, L., Jackson, C., Ballesteros, J., Sanchez, J., Watrous, J. D., Phelan, V. V., van de Wiel, C., Kersten, R. D., Mehnaz, S., De Mot, R., Shank, E. A., Dorrestein, P. C. (2013). MS/MS networking guided analysis of molecule and gene cluster families. *Proceedings of the National Academy of Sciences of the United States of America*, 110(28), E2611-2620. <https://doi.org/10.1073/pnas.1303471110>.
- 120 MassBank <http://www.massbank.jp/>
- 121 Metlin <https://metlin.scripps.edu/>.

- 122 <http://spectra.psc.riken.jp/menta.cgi/respect/index>.
- 123 The National Institute of Standards and Technology. NIST <http://www.nist.gov/srd/nist1a.cfm>.
- 124 Fox Ramos, A. E., Alcover, C., Evanno, L., Maciuk, A., Litaudon, M., Duplais, C., Bernadat, G., Gallard, J.-F., Jullian, J.-C., Mouray, E., Grellier, P., Loiseau, P. M., Pomel, S., Poupon, E., Champy, P., & Beniddir, M. A. (2017). Revisiting previously investigated plants : A molecular networking-based study of *geissospermum laeve*. Journal of Natural Products, 80(4), 1007-1014. <https://doi.org/10.1021/acs.jnatprod.6b01013>.
- 125 Yang, J. Y., Sanchez, L. M., Rath, C. M., Liu, X., Boudreau, P. D., Bruns, N., Glukhov, E., Wodtke, A., de Felicio, R., Fenner, A., Wong, W. R., Linington, R. G., Zhang, L., Debonsi, H. M., Gerwick, W. H., & Dorrestein, P. C. (2013). Molecular networking as a dereplication strategy. Journal of Natural Products, 76(9), 1686-1699. <https://doi.org/10.1021/np400413s>.
- 126 Naman, C. B., Rattan, R., Nikoulina, S. E., Lee, J., Miller, B. W., Moss, N. A., Armstrong, L., Boudreau, P. D., Debonsi, H. M., Valeriote, F. A., Dorrestein, P. C., & Gerwick, W. H. (2017). Integrating molecular networking and biological assays to target the isolation of a cytotoxic cyclic octapeptide, samoamide a, from an american samoan marine cyanobacterium. Journal of Natural Products, 80(3), 625-633. <https://doi.org/10.1021/acs.jnatprod.6b00907>.
- 127 Luzzatto-Knaan, T., Garg, N., Wang, M., Glukhov, E., Peng, Y., Ackermann, G., Amir, A., Duggan, B. M., Ryazanov, S., Gerwick, L., Knight, R., Alexandrov, T., Bandeira, N., Gerwick, W. H., & Dorrestein, P. C. (2017). Digitizing mass spectrometry data to explore the chemical diversity and distribution of marine cyanobacteria and algae. eLife, 6. <https://doi.org/10.7554/eLife.24214>.
- 128 Nguyen, D. D., Melnik, A. V., Koyama, N., Lu, X., Schorn, M., Fang, J., Aguinaldo, K., Lincecum, T. L., Ghequire, M. G. K., Carrion, V. J., Cheng, T. L., Duggan, B. M., Malone, J. G., Mauchline, T. H., Sanchez, L. M., Kilpatrick, A. M., Raaijmakers, J. M., De Mot, R., Moore, B. S., Dorrestein, P. C. (2016). Indexing the *Pseudomonas* specialized metabolome enabled the discovery of poaeamide B and the bananamides. Nature Microbiology, 2, 16197. <https://doi.org/10.1038/nmicrobiol.2016.197>.
- 129 Wolfender, J.-L., Litaudon, M., Touboul, D., & Queiroz, E. F. (2019). Innovative omics-based approaches for prioritisation and targeted isolation of natural products – new strategies for drug discovery. Natural Product Reports, 36(6), 855-868. <https://doi.org/10.1039/C9NP00004F>.
- 130 Di Ottavio, F., Gauglitz, J. M., Ernst, M., Panitchpakdi, M. W., Fanti, F., Compagnone, D., Dorrestein, P. C., & Sergi, M. (2020). A UHPLC-HRMS based metabolomics and chemoinformatics approach to chemically distinguish ‘super foods’ from a variety of plant-based foods. Food Chemistry, 313, 126071. <https://doi.org/10.1016/j.foodchem.2019.126071>.
- 131 Da Silva, R. R., Wang, M., Nothias, L.-F., van der Hooft, J. J. J., Caraballo-Rodríguez, A. M., Fox, E., Balunas, M. J., Klassen, J. L., Lopes, N. P., & Dorrestein, P. C. (2018). Propagating annotations of molecular networks using in silico fragmentation. PLoS Computational Biology, 14(4), e1006089. <https://doi.org/10.1371/journal.pcbi.1006089>.
- 132 Le Pogam, P., Poupon, E., Genta-Jouve, G., Gallard, J.-F., Turpin, V., Skiredj, A., Leblanc, K., & Beniddir, M. (2020). Caffeine alkaline hydrolysis and in silico anticipation reveal the origin of camellimidazoles. <https://doi.org/10.26434/chemrxiv.12110862.v1>.
- 133 Zohary, D., Hopf, M., Weiss, E. (2012). Domestication of Plants in the Old World, 4th ed., Oxford University Press: New York, United States, pp. 101-103.
- 134 Oomah, B. D. (2001). Flaxseed as a functional food source. *J. Sci. Food Agric.* 81, 889-894. [doi.org/10.1002/jsfa.898](https://doi.org/10.1002/jsfa.898).
- 135 Shim, Y. Y., Gui, B., Arnison, P. G., Wang, Y., Reaney, M. J. (2014). Flaxseed (*Linum usitatissimum* L.) bioactive compounds and peptide nomenclature: a review. Trends Food Sci Technol. 38, 5-20.

- 136 Palla, A. H., Khan, N. A., Bashir, S., Iqbal, J., Gilani, A. H. (2015). Pharmacological basis for the medicinal use of *Linum usitatissimum* (Flaxseed) in infectious and non-infectious diarrhea. *J Ethnopharmacol.* 160, 61-68. [doi.org/10.1016/j.jep.2014.11.030](https://doi.org/10.1016/j.jep.2014.11.030).
- 137 Bakht, J., Ali, H., Khan, M. A., Khan, A., Saeed, M., Shafi, M., Islam, A., Tayyab, M. (2011). Antimicrobial activities of different solvents extracted samples of *Linum usitatissimum* by disc diffusion method. *Afr J Biotechnol.* 10, 19825-19835. [doi.org/10.5897/AJB11.229](https://doi.org/10.5897/AJB11.229).
- 138 Zanwar, A., Hegde, M., Bodhankar, S. (2011). Cardioprotective activity of flax lignan concentrate extracted from seeds of *Linum usitatissimum* in isoprenalin induced myocardial necrosis in rats. *Interdiscip Toxicol.* 4, 90-97.
- 139 Lainé, E., Hano, C., Lamblin, F. (2009). Phytoestrogens: Lignans. In Chemoprevention of Cancer and DNA Damage by Dietary Factors (eds S. Knasmüller, D. M. DeMarini, I. Johnson and C. Gerhäuser). WILEY-VCH, Weinheim, pp. 555-577. [doi.org/10.1002/9783527626588.ch31](https://doi.org/10.1002/9783527626588.ch31).
- 140 [China Food & Drug Administration \(2015\)](#) China IECIC - Inventory of Existing Cosmetic Ingredients in China. Accessible on: [https://www.chemsafetypro.com/Topics/Cosmetics/China\\_IECIC\\_Finder.html](https://www.chemsafetypro.com/Topics/Cosmetics/China_IECIC_Finder.html).
- 141 Dedičová, B., Hricova, A., Šamaj, J., Obert, B., Bobák, M., Pret'ová, A. (2000). Shoots and embryo-like structures regenerated from cultured flax (*Linum usitatissimum* L.) hypocotyl segments. *J Plant Physiol.* 157, 327-334. [doi.org/10.1016/S0176-1617\(00\)80055-3](https://doi.org/10.1016/S0176-1617(00)80055-3)
- 142 Salaj, J., Petrovská, B., Obert, B., Pret'ová, A. (2005). Histological study of embryo-like structures initiated from hypocotyl segments of flax (*Linum usitatissimum* L.). *Plant Cell Rep.* 24, 590-595. [doi.org/10.1007/s00299-005-0029-x](https://doi.org/10.1007/s00299-005-0029-x).
- 143 Cunha, A., Ferreira, M. F. (1999). Influence of medium parameters on somatic embryogenesis from hypocotyl explants of flax (*Linum usitatissimum* L.): effect of carbon source, total inorganic nitrogen and balance between ionic forms and interaction between calcium and zeatin. *J Plant Physiol.* 155, 591-597. [doi.org/10.1016/S0176-1617\(99\)80059-5](https://doi.org/10.1016/S0176-1617(99)80059-5).
- 144 Nichterlein, K., Umbach, H., Friedt, W. (1991). Genotypic and exogenous factors affecting shoot regeneration from anther callus of linseed (*Linum usitatissimum* L.). *Euphytica.* 58, 157-164. [doi.org/10.1007/BF00022816](https://doi.org/10.1007/BF00022816)
- 145 Rutkowska-Krause, I., Mankowska, G., Lukaszewicz, M., Szopa, J. (2003). Regeneration of flax (*Linum usitatissimum* L.) plants from another culture and somatic tissue with increased resistance to *Fusarium oxysporum*. *Plant Cell Rep.* 22, 110-116. [doi.org/10.1007/s00299-003-0662-1](https://doi.org/10.1007/s00299-003-0662-1).
- 146 Anjum, S., Abbasi, B. H., Hano, C. (2017a). Trends in accumulation of pharmacologically important antioxidant-secondary metabolites in callus cultures of *Linum usitatissimum* L. *Plant Cell Tissue Organ Cult.* 129, 73-87. [doi.org/10.1007/s11240-016-1158-3](https://doi.org/10.1007/s11240-016-1158-3).
- 147 Zahir, A., Ahmad, W., Nadeem, M., Giglioli-Guivarc'h, N., Hano, C., Abbasi, B. H. (2018). In vitro cultures of *Linum usitatissimum* L.: Synergistic effects of mineral nutrients and photoperiod regimes on growth and biosynthesis of lignans and neolignans. *J Photochem Photobiol B, Biol.* 187, 141-150. [doi.org/10.1016/j.jphotobiol.2018.08.009](https://doi.org/10.1016/j.jphotobiol.2018.08.009).
- 148 Hano, C., Addi, M., Bensaddek, L., Crônier, D., Baltora-Rosset, S., Doussot, J., Maury, S., Mesnard, F., Chabbert, B., Hawkins, S., Lainé, E., Lamblin, F. (2006). Differential accumulation of monolignol-derived compounds in elicited flax (*Linum usitatissimum*) cell suspension cultures. *Planta.* 223, 975-989. [doi.org/10.1007/s00425-005-0156-1](https://doi.org/10.1007/s00425-005-0156-1).
- 149 Attoumbre, J., Hano, C., Mesnard, F., Lamblin, F., Bensaddek, L., Raynaud-Le Grandic, S., Lainé, E., Baltora-Rosset, S., Fliniaux, M.A. (2006a). Identification by NMR and accumulation of a neolignan, the dehydrodiconiferyl alcohol-4-b-D-glucoside, in *Linum usitatissimum* cell cultures. *C. R. Chimie.* 9, 420-425. [doi.org/10.1016/j.crci.2005.06.012](https://doi.org/10.1016/j.crci.2005.06.012).

- 150 Barbulova, A., Apone, F., Colucci, G. (2014). Plant cell cultures as source of cosmetic active ingredients. *Cosmetics*, 1, 94-104. [doi.org/10.3390/cosmetics1020094](https://doi.org/10.3390/cosmetics1020094).
- 151 Intérêt des lignanes dans la prévention et le traitement de cancers. (s. d.). Consulté 30 septembre 2020, à l'adresse [http://www.ipubli.inserm.fr/bitstream/handle/10608/6454/MS\\_2008\\_5\\_511.html?sequence=27&isAllowed=y](http://www.ipubli.inserm.fr/bitstream/handle/10608/6454/MS_2008_5_511.html?sequence=27&isAllowed=y).
- 152 Beejmohun, V., Fliniaux, O., Hano, C., Pilard, S., Grand, E., Lesur, D., Cailleu, D., Lamblin, F., Lainé, E., Kovensky, J., Fliniaux, M.-A., & Mesnard, F. (2007). Coniferin dimerisation in lignan biosynthesis in flax cells. *Phytochemistry*, 68(22), 2744-2752. <https://doi.org/10.1016/j.phytochem.2007.09.016>.
- 153 Differential accumulation of silymarin induced by exposure of *Silybum marianum* L. callus cultures to several spectra of monochromatic lights. (2018). *Journal of Photochemistry and Photobiology B: Biology*, 184, 61-70. <https://doi.org/10.1016/j.jphotobiol.2018.05.018>.
- 154 Zahir, A., Ahmad, W., Nadeem, M., Giglioli-Guivarc'h, N., Hano, C., & Abbasi, B. H. (2018). In vitro cultures of *Linum usitatissimum* L.: Synergistic effects of mineral nutrients and photoperiod regimes on growth and biosynthesis of lignans and neolignans. *Journal of Photochemistry and Photobiology. B, Biology*, 187, 141-150. <https://doi.org/10.1016/j.jphotobiol.2018.08.009>.
- 155 Quinet, M., Angosto, T., Yuste-Lisbona, F. J., Blanchard-Gros, R., Bigot, S., Martinez, J.-P., & Lutts, S. (2019). Tomato Fruit Development and Metabolism. *Frontiers in Plant Science*, 10(1554). <https://doi.org/10.3389/fpls.2019.01554>.
- 156 Walling, L. L. (2000). The Myriad Plant Responses to Herbivores. *Journal of Plant Growth Regulation*, 19(2), 195-216. <https://doi.org/10.1007/s003440000026>.
- 157 Dyer, L. A., Philbin, C. S., Ochsenrider, K. M., Richards, L. A., Massad, T. J., Smilanich, A. M., Jeffrey, C. S. (2018). Modern approaches to study plant-insect interactions in chemical ecology. *Nature Reviews Chemistry*, 2(6), 50-64. <https://doi.org/10.1038/s41570-018-0009-7>.
- 158 Flint, M. L. (1998). Integrated Pest Management for Tomatoes—Fourth Edition: University of California, California.
- 159 Lange, W., & Bronson, L. (1981). Insect Pests of Tomatoes. *Annual Review of Entomology*, 26(1), 345-371. <https://doi.org/10.1146/annurev.en.26.010181.002021>.
- 160 Cárdenas, P. D., Sonawane, P. D., Heinig, U., Jozwiak, A., Panda, S., Abebie, B., Aharoni, A. (2019). Pathways to defense metabolites and evading fruit bitterness in genus *Solanum* evolved through 2-oxoglutarate-dependent dioxygenases. *Nature communications*, 10(1), 5169-5169. <https://doi.org/10.1038/s41467-019-13211-4>.
- 161 Chowański, S., Adamski, Z., Marciniak, P., Rosiński, G., Büyükgüzel, E., Büyükgüzel, K., Bufo, S. A. (2016). A Review of Bioinsecticidal Activity of Solanaceae Alkaloids. *Toxins*, 8(3), 60. <https://doi.org/10.3390/toxins8030060>.
- 162 Friedman, M. (2002). Tomato Glycoalkaloids: Role in the Plant and in the Diet. *Journal of Agricultural and Food Chemistry*, 50(21), 5751-5780. <https://doi.org/10.1021/jf020560c>.
- 163 Larbat, R., Paris, C., Le Bot, J., & Adamowicz, S. (2014). Phenolic characterization and variability in leaves, stems and roots of Micro-Tom and patio tomatoes, in response to nitrogen limitation. *Plant Science*, 224, 62-73. <https://doi.org/https://doi.org/10.1016/j.plantsci.2014.04.010>.
- 164 Kaplan, I., Halitschke, R., Kessler, A., Rehill, B. J., Sardanelli, S., & Denno, R. F. (2008). Physiological integration of roots and shoots in plant defense strategies links above- and belowground herbivory. *Ecology Letters*, 11(8), 841-851. <https://doi.org/10.1111/j.1461-0248.2008.01200.x>.

- 165 Vuckovic, D. Current trends and challenges in sample preparation for global metabolomics using liquid chromatography–mass spectrometry. (2012) *Anal. Bioanal. Chem.* 403, 1523–1548. <https://doi.org/10.1007/s00216-012-6039-y>.
- 166 Sangster, T., Major, H., Plumb, R., Wilson, A.J., Wilson, I.D. A pragmatic and readily implemented quality control strategy for HPLC-MS and GC-MS-based metabonomic analysis. (2006). *Analyst* 131, 1075–1078. <https://doi.org/10.1039/B604498K>.
- 167 Dudzik, D., Barbas-Bernardos, C., García, A., Barbas, C. Quality assurance procedures for mass spectrometry untargeted metabolomics. a review. (2017). *J. Pharm. Biomed. Anal.* 147, 149-173. <https://doi.org/10.1016/j.jpba.2017.07.044>.
- 168 Pedrioli, P. G. A., Eng, J. K., Hubley, R., Vogelzang, M., Deutsch, E. W., Raught, B., Pratt, B., Nilsson, E., Angeletti, R. H., Apweiler, R., Cheung, K., Costello, C. E., Hermjakob, H., Huang, S., Julian, R. K., Kapp, E., McComb, M. E., Oliver, S. G., Omenn, G., Aebersold, R. (2004). A common open representation of mass spectrometry data and its application to proteomics research. *Nature Biotechnology*, 22(11), 1459-1466. <https://doi.org/10.1038/nbt1031>.
- 169 Rew, R., & Davis, G. (1990). NetCDF : An interface for scientific data access. *IEEE Computer Graphics and Applications*, 10(4), 76-82. <https://doi.org/10.1109/38.56302>.
- 170 Vinaixa, M., Samino, S., Saez, I., Duran, J., Guinovart, J.J., Yanes, O., Vinaixa, M., Samino, S., Saez, I., Duran, J., Guinovart, J.J., Yanes, O., 2012. A Guideline to Univariate Statistical Analysis for LC/MS-Based Untargeted Metabolomics-Derived Data. *Metabolites* 2, 775–795. <https://doi.org/10.3390/metabo2040775>.
- 171 Galaxy training : Visualization of rna-seq results with volcano plot. (s. d.). Consulté 30 septembre 2020, à l'adresse <https://galaxyproject.github.io/training-material/topics/transcriptomics/tutorials/rna-seq-viz-with-volcanoplot/tutorial.html#:~:text=A%20volcano%20plot%20is%20a,the%20most%20biologically%20significant%20genes>.
- 172 Worley, B., Powers, R., 2013. Multivariate Analysis in Metabolomics. *Curr. Metabolomics* 1, 92-107. <https://doi.org/10.2174/2213235X11301010092>.
- 173 Gika, H.G., Theodoridis, G.A., Plumb, R.S., Wilson, I.D. (2014). Current practice of liquid chromatography–mass spectrometry in metabolomics and metabonomics. *J. Pharm. Biomed. Anal.* 87, 12–25. <https://doi.org/10.1016/j.jpba.2013.06.032>.
- 174 Viant, M.R., Kurland, I.J., Jones, M.R., Dunn, W.B. (2017). How close are we to complete annotation of metabolomes *Curr Opin. Chem. Biol.* 36, 64–69. <https://doi.org/10.1016/j.cbpa.2017.01.001>.
- 175 Rochat, B. (2017). Proposed Confidence Scale and ID Score in the Identification of Known Unknown Compounds Using High Resolution MS Data. *J. Am. Soc. Mass Spectrom.* 28, 709–723. <https://doi.org/10.1007/s13361-016-1556-0>.
- 176 Choi, Y. H., & Verpoorte, R. (2014). Metabolomics : What you see is what you extract: editorial. *Phytochemical Analysis*, 25(4), 289-290. <https://doi.org/10.1002/pca.2513>.
- 177 Martin, A. C., Pawlus, A. D., Jewett, E. M., Wyse, D. L., Angerhofer, C. K., & Hegeman, A. D. (2014). Evaluating solvent extraction systems using metabolomics approaches. *RSC Adv.*, 4(50), 26325-26334. <https://doi.org/10.1039/C4RA02731K>.
- 178 Gómez-Romero, M., Segura-Carretero, A., & Fernández-Gutiérrez, A. (2010). Metabolite profiling and quantification of phenolic compounds in methanol extracts of tomato fruit. *Phytochemistry*, 71(16), 1848-1864. <https://doi.org/10.1016/j.phytochem.2010.08.002>.

- 179 Han, P., Desneux, N., Michel, T., Le Bot, J., Seassau, A., Wajnberg, E., Amiens-Desneux, E., & Lavoir, A.-V. (2016). Does plant cultivar difference modify the bottom-up effects of resource limitation on plant-insect herbivore interactions? *Journal of Chemical Ecology*, 42(12), 1293-1303. <https://doi.org/10.1007/s10886-016-0795-7>.
- 180 U. Heinig, A. Aharoni, Analysis of steroid alkaloids and saponins in *Solanaceae* plant extracts using UPLC-qTOF mass spectrometry, in: M. Rodríguez-Concepción (Ed.) *Plant Isoprenoids: Methods and Protocols*, Springer New York, New York, NY, 2014, pp. 171-185.
- 181 Giacomoni, F., Le Corguillé, G., Monsoor, M., Landi, M., Pericard, P., Pétéra, M., Duperier, C., Tremblay-Franco, M., Martin, J.-F., Jacob, D., Goulitquer, S., Thévenot, E. A., & Caron, C. (2015). Workflow4Metabolomics : A collaborative research infrastructure for computational metabolomics. *Bioinformatics* (Oxford, England), 31(9), 1493-1495. <https://doi.org/10.1093/bioinformatics/btu813>.
- 182 Goecks, J., Nekrutenko, A., Taylor, J., & The Galaxy Team. (2010). Galaxy : A comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology*, 11(8), R86. <https://doi.org/10.1186/gb-2010-11-8-r86>.
- 183 Guittion, Y., Tremblay-Franco, M., Le Corguillé, G., Martin, J.-F., Pétéra, M., Roger-Mele, P., Delabrière, A., Goulitquer, S., Monsoor, M., Duperier, C., Canlet, C., Servien, R., Tardivel, P., Caron, C., Giacomoni, F., & Thévenot, E. A. (2017). Create, run, share, publish, and reference your LC-MS, FIA-MS, GC-MS, and NMR data analysis workflows with the Workflow4Metabolomics 3.0 Galaxy online infrastructure for metabolomics. *The International Journal of Biochemistry & Cell Biology*, 93, 89-101. <https://doi.org/10.1016/j.biocel.2017.07.002>.
- 184 Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R., & Siuzdak, G. (2006). XCMS : Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Analytical Chemistry*, 78(3), 779-787. <https://doi.org/10.1021/ac051437y>.
- 185 Larbat, R., Paris, C., Le Bot, J., & Adamowicz, S. (2014). Phenolic characterization and variability in leaves, stems and roots of Micro-Tom and patio tomatoes, in response to nitrogen limitation. *Plant Science*, 224, 62-73. <https://doi.org/https://doi.org/10.1016/j.plantsci.2014.04.010>.
- 186 Zhang, P., Chan, W., Ang, I. L., Wei, R., Lam, M. M. T., Lei, K. M. K., & Poon, T. C. W. (2019). Revisiting Fragmentation Reactions of Protonated  $\alpha$ -Amino Acids by High-Resolution Electrospray Ionization Tandem Mass Spectrometry with Collision-Induced Dissociation. *Scientific Reports*, 9(1), 6453. <https://doi.org/10.1038/s41598-019-42777-8>.
- 187 Cataldi, T. R. I., Lelario, F., & Bufo, S. A. (2005). Analysis of tomato glycoalkaloids by liquid chromatography coupled with electrospray ionization tandem mass spectrometry. *Rapid Communications in Mass Spectrometry*, 19(21), 3103-3110. <https://doi.org/10.1002/rcm.2176>.
- 188 Hodge, A., Berta, G., Doussan, C., Merchan, F., & Crespi, M. (2009). Plant root growth, architecture and function. *Plant and Soil*, 321(1), 153-187. <https://doi.org/10.1007/s11104-009-9929-9>.
- 189 Rasman, S., & Agrawal, A. A. (2008). In defense of roots: A research agenda for studying plant resistance to belowground herbivory. *Plant Physiology*, 146(3), 875.
- 190 Roldan, M. V. G., Engel, B., de Vos, R. C. H., Vereijken, P., Astola, L., Groenenboom, M., Hall, R. D. (2014). Metabolomics reveals organ-specific metabolic rearrangements during early tomato seedling development. *Metabolomics*, 10(5), 958-974. <https://doi.org/10.1007/s11306-014-0625-2>.
- 191 Sung, J., Lee, S., Lee, Y., Ha, S., Song, B., Kim, T., Krishnan, H. B. (2015). Metabolomic profiling from leaves and roots of tomato (*Solanum lycopersicum* L.) plants grown under nitrogen, phosphorus or potassium-deficient condition. *Plant Science*, 241, 55-64. <https://doi.org/https://doi.org/10.1016/j.plantsci.2015.09.027>.
- 192 Van Dam, N. M., De Jong, T. J., Iwasa, Y., & Kubo, T. (1996). Optimal distribution of defences: Are plants smart investors? *Functional Ecology*, 10(1), 128-136. <https://doi.org/10.2307/2390271>.

- 193 Zhao, Y. (2012). Auxin Biosynthesis: A Simple Two-Step Pathway Converts Tryptophan to Indole-3-Acetic Acid in Plants. *Molecular Plant*, 5(2), 334-338. <https://doi.org/https://doi.org/10.1093/mp/ssr104>.
- 194 Kazan, K. (2013). Auxin and the integration of environmental signals into plant root development. *Annals of botany*, 112(9), 1655-1665. <https://doi.org/10.1093/aob/mct229>.
- 195 Saini, S., Sharma, I., Kaur, N., & Pati, P. K. (2013). Auxin: a master regulator in plant root development. *Plant Cell Reports*, 32(6), 741-757. <https://doi.org/10.1007/s00299-013-1430-5>.
- 196 Thème : L'industrie cosmétique en France. Statista. Consulté 30 septembre 2020, à l'adresse <https://fr.statista.com/themes/2735/le-secteur-des-produits-de-beaute-et-d-hygiene-en-france/>.
- 197 <file:///C:/Users/NTIC-trading/Downloads/France-Fiche-sectorielle-cosmetique.pdf>. (Consulté le 29/06/2020)
- 198 Bora, K. S., & Sharma, A. (2011). The genus *artemisia* : A comprehensive review. *Pharmaceutical Biology*, 49(1), 101-109. <https://doi.org/10.3109/13880209.2010.497815>.
- 199 Ćavar, S., Maksimović, M., Vidic, D., & Parić, A. (2012). Chemical composition and antioxidant and antimicrobial activity of essential oil of *Artemisia annua* L. from Bosnia. *Industrial Crops and Products*, 37(1), 479-485. <https://doi.org/10.1016/j.indcrop.2011.07.024>.
- 200 Artemisia annua – recherche google. (s. d.). Consulté 30 septembre 2020, à l'adresse [https://www.google.com/search?q=artemisia+annua&sxsrf=ALeKk02B1Fy4OP5m0vYUv-sfjbsOsM-hTw:1596385828567&source=lnms&tbo=isch&sa=X&ved=2ahUKEwikuPHR-PzqAhUpxYUKHaOtBBMQ\\_AUoAnoECBkQBA&biw=1280&bih=610#imgrc=YrX684meVmpGsM](https://www.google.com/search?q=artemisia+annua&sxsrf=ALeKk02B1Fy4OP5m0vYUv-sfjbsOsM-hTw:1596385828567&source=lnms&tbo=isch&sa=X&ved=2ahUKEwikuPHR-PzqAhUpxYUKHaOtBBMQ_AUoAnoECBkQBA&biw=1280&bih=610#imgrc=YrX684meVmpGsM).
- 201 The forests of China with a survey of grassland and desert vegetation. (1962). *The Forestry Chronicle*, 38(4), 511-512. <https://doi.org/10.5558/tfc38511b1-4>.
- 202 Zalkow, L. H., Harris, R. N., Van Derveer, D., & Bertrand, J. A. (1977). Isocomene : A novel sesquiterpene from *Isocoma Wrightii*. X-Ray crystal structure of the corresponding diol. *Journal of the Chemical Society, Chemical Communications*, 13, 456. <https://doi.org/10.1039/c39770000456>.
- 203 Informal Consultation on Artemisinin and its Derivatives (1993 : Geneva, S., Unit, W. H. O. M. C., & Diseases, U. B. S. P. for R. and T. in T. (1994). The role of artemisinin and its derivatives in the current treatment of malaria (1994-1995 : Report of an informal consultation, Geneva, 27-29 September 1993. <https://apps.who.int/iris/handle/10665/61512>.
- 204 Brisibe, E. A., Umoren, U. E., Brisibe, F., Magalhäes, P. M., Ferreira, J. F. S., Luthria, D., Wu, X., & Prior, R. L. (2009). Nutritional characterisation and antioxidant capacity of different tissues of *Artemisia annua* L. *Food Chemistry*, 115(4), 1240-1246. <https://doi.org/10.1016/j.foodchem.2009.01.033>.
- 205 van der Kooy, F., & Sullivan, S. E. (2013). The complexity of medicinal plants : The traditional *Artemisia annua* formulation, current status and future perspectives. *Journal of Ethnopharmacology*, 150(1), 1-13. <https://doi.org/10.1016/j.jep.2013.08.021>.
- 206 From Artemisia annua 1. To artemisinins—1st edition. (s. d.). Consulté 30 septembre 2020, à l'adresse <https://www.elsevier.com/books/from-artemisia-annua-1-to-artemisinins/tu/978-0-12-811655-5>.
- 207 El-Askary, H., Handoussa, H., Badria, F., El-Khatib, A. H., Alsayari, A., Linscheid, M. W., & Abdel Motaal, A. (2019). Characterization of hepatoprotective metabolites from *Artemisia annua* and cleome droserifolia using hplc/pda/esi/ms–ms. *Revista Brasileira de Farmacognosia*, 29(2), 213-220. <https://doi.org/10.1016/j.bjfp.2018.10.001>.
- 208 Li, K.-M., Dong, X., Ma, Y.-N., Wu, Z.-H., Yan, Y.-M., & Cheng, Y.-X. (2019). Antifungal coumarins and lignans from *Artemisia annua*. *Fitoterapia*, 134, 323-328. <https://doi.org/10.1016/j.fitote.2019.02.022>.

- 209 Lang, S. J., Schmiedch, M., Hafner, S., Paetz, C., Steinborn, C., Huber, R., Gaafary, M. E., Werner, K., Schmidt, C. Q., Syrovets, T., & Simmet, T. (2019). Antitumor activity of an *Artemisia annua* herbal preparation and identification of active ingredients. *Phytomedicine: International Journal of Phytotherapy and Phytopharmacology*, 62, 152962. <https://doi.org/10.1016/j.phymed.2019.152962>.
- 210 Wan, X. L., Niu, Y., Zheng, X. C., Huang, Q., Su, W. P., Zhang, J. F., Zhang, L. L., & Wang, T. (2016). Antioxidant capacities of *Artemisia annua* L. leaves and enzymatically treated *Artemisia annua* L. in vitro and in broilers. *Animal Feed Science and Technology*, 221, 27-34. <https://doi.org/10.1016/j.anifeedsci.2016.08.017>.
- 211 Gouveia, S. C., & Castilho, P. C. (2013). *Artemisia annua* L.: Essential oil and acetone extract composition and antioxidant capacity. *Industrial Crops and Products*, 45, 170-181. <https://doi.org/10.1016/j.indcrop.2012.12.022>.
- 212 Iqbal, S., Younas, U., Chan, K. W., Zia-Ul-Haq, M., & Ismail, M. (2012). Chemical composition of *Artemisia annua* L. Leaves and antioxidant potential of extracts as a function of extraction solvents. *Molecules*, 17(5), 6020-6032. <https://doi.org/10.3390/molecules17056020>.
- 213 Eclipta in flora of north america efloras. Org. (s. d.). Consulté 30 septembre 2020, à l'adresse [http://www.efloras.org/florataxon.aspx?flora\\_id=1&taxon\\_id=111273](http://www.efloras.org/florataxon.aspx?flora_id=1&taxon_id=111273).
- 214 *Eclipta alba* (Bhringraj)—Ms full hair ingredients. (s. d.). Ms Full Hair. Consulté 30 septembre 2020, à l'adresse <http://www.msfullhair.com/ingredients/eclipta-alba/>.
- 215 L'Éclipte prostrée (*Eclipta prostrata* (L.) L. 1771) : Une espèce en progression en Nouvelle-Aquitaine ? (s. d.). Centre de ressources. Consulté 30 septembre 2020, à l'adresse <http://especes-exotiques-envahissantes.fr/eclipta-prostrata-progression-en-nouvelle-aquitaine/>.
- 216 Chung, I.-M., Rajakumar, G., Lee, J.-H., Kim, S.-H., & Thiruvengadam, M. (2017). Ethnopharmacological uses, phytochemistry, biological activities, and biotechnological applications of *Eclipta prostrata*. *Applied Microbiology and Biotechnology*, 101(13), 5247-5257. <https://doi.org/10.1007/s00253-017-8363-9>.
- 217 Jadhav V.M., Thorat R. M., Kadam V.J., Salaskar K.P. (2009). Chemical composition, pharmacological activities of *Eclipta alba*. *Journal of Pharmacy Research*, 2(8), 1229-1231.
- 218 Bhattacharjee A., Chakraborty A., Dutta D., Sau A., Chakraborty S., Samanta N., Biswas P.S. and Mukhopadhyay G. (2017). Standardization of *Eclipta alba* by hptlc, hplc and aas. *Pharmaceutica Analytica Acta*, 08(04). <https://doi.org/10.4172/2153-2435.1000541>.
- 219 Cheruvu, H. S., Yadav, N. K., Valicherla, G. R., Arya, R. K., Hussain, Z., Sharma, C., Arya, K. R., Singh, R. K., Datta, D., & Gayen, J. R. (2018). LC-MS/MS method for the simultaneous quantification of luteolin, wedelolactone and apigenin in mice plasma using hansen solubility parameters for liquid-liquid extraction : Application to pharmacokinetics of *Eclipta alba* chloroform fraction. *Journal of Chromatography B*, 1081-1082, 76-86. <https://doi.org/10.1016/j.jchromb.2018.01.035>.
- 220 Yahara, S., Ding, N., & Nohara, T. (1994). Studies on the constituents of compositae plants. Part ii. Oleanane glycosides from *Eclipta alba*. *CHEMICAL & PHARMACEUTICAL BULLETIN*, 42(6), 1336-1338. <https://doi.org/10.1248/cpb.42.1336>.
- 221 Liste A des plantes médicinales utilisées traditionnellement. Consulté le 06 juin 2020, à l'adresse [https://ansm.sante.fr/var/ansm\\_site/storage/original/application/dc6398f1f676936f296909ec52fc2213.pdf](https://ansm.sante.fr/var/ansm_site/storage/original/application/dc6398f1f676936f296909ec52fc2213.pdf).
- 222 Manvar, D., Mishra, M., Kumar, S., & Pandey, V. N. (2012). Identification and evaluation of anti Hepatitis C virus phytochemicals from *Eclipta alba*. *Journal of Ethnopharmacology*, 144(3), 545-554. <https://doi.org/10.1016/j.jep.2012.09.036>.
- 223 Saxena, A. K., Singh, B., & Anand, K. K. (1993). Hepatoprotective effects of *Eclipta alba* on subcellular levels in rats. *Journal of Ethnopharmacology*, 40(3), 155-161. [https://doi.org/10.1016/0378-8741\(93\)90063-B](https://doi.org/10.1016/0378-8741(93)90063-B).

- 224 Iqbal Hussain. (2011). Phytochemical, physicochemical and anti-fungal activity of *Eclipta alba*. African Journal of Pharmacy and Pharmacology, 5(19). <https://doi.org/10.5897/AJPP11.453>.
- 225 Kim, H.-Y., Kim, H. M., Ryu, B., Lee, J.-S., Choi, J.-H., & Jang, D. S. (2015). Constituents of the aerial parts of *Eclipta prostrata* and their cytotoxicity on human ovarian cancer cells in vitro. Archives of Pharmacal Research, 38(11), 1963-1969. <https://doi.org/10.1007/s12272-015-0599-2>.
- 226 Chaudhary, H., Dhuna, V., Singh, J., Kamboj, S. S., & Seshadri, S. (2011). Evaluation of hydro-alcoholic extract of *Eclipta alba* for its anticancer potential : An in vitro study. Journal of Ethnopharmacology, 136(2), 363-367. <https://doi.org/10.1016/j.jep.2011.04.066>.
- 227 Satheesh Naik K, Gurushanthaiah M, Nagesh Raju G, Lokanadham S, & Seshadri Reddy V. (2019). Protective role of *Eclipta alba* against hyperlipidemia induced by high-fat diet in albino rats. International Journal of Research in Pharmaceutical Sciences, 10(2), 1181-1184. <https://doi.org/10.26452/ijrps.v10i2.403>.
- 228 Gupta, A., Kumar, A., Kumar, D., Singh, R., Shankar, K., Varshney, S., Rajan, S., Srivastava, A., Gupta, S., Narender, T., & Gaikwad, A. N. (2018). Ecliptal, a promising natural lead isolated from *Eclipta alba* modulates adipocyte function and ameliorates metabolic syndrome. Toxicology and Applied Pharmacology, 338, 134-147. <https://doi.org/10.1016/j.taap.2017.11.016>.
- 229 Gupta, A., Kumar, A., Kumar, D., Nandan, S., Shankar, K., Varshney, S., Rajan, S., Srivastava, A., Gupta, S., Kanjojiya, S., Narender, T., & Gaikwad, A. N. (2017). Ethyl acetate fraction of *Eclipta alba* : A potential phytopharmaceutical targeting adipocyte differentiation. Biomedicine & Pharmacotherapy, 96, 572-583. <https://doi.org/10.1016/j.biopha.2017.10.002>.
- 230 Gautam, R., Saklani, A., & Jachak, S. M. (2007). Indian medicinal plants as a source of antimycobacterial agents. Journal of Ethnopharmacology, 110(2), 200-234. <https://doi.org/10.1016/j.jep.2006.12.031>.
- 231 Abdel Kader MS., Bahler BD., Malone S., Werkhoven MC., van Troon F., Wisse JH., Bursuker I., Neddermann KM., Mamber SW., Kingston DG. (1998). DNA-damaging steroid alkaloids from *Eclipta alba* from the Suriname rainforest. Journal of Natural Products, 61, 1202-1208.
- 232 Bapna S., Asdule S., Shirhat MS., Jadhav S., Patil LS., Deskmukh RA. (2007). Anti-malarial activity of *Eclipta alba* against Plasmodium berghei infection in mice. Journal of Communicable Diseases, 39, 91-94.
- 233 Ananthi J., Prakasam A., Pugalendi KV. (2003). Antihyperglycemic Activity of *Eclipta alba* Leaf on Alloxan-induced Diabetic Rats. Yale Journal of Biological Medicine, 1, 97-102.
- 234 Amritpal S., Samir M., Ravi S. (2008). Anti-inflammatory and analgesic agents from Indian medicinal plants. International journal of integrative biology, 3(1), 58-72.
- 235 Pithayanukul, P., Lapett, B., Bavovada, R., Pakmanee, N., & Suttisri, R. (2007). Inhibition of proteolytic and hemorrhagic activities by ethyl acetate extract of *eclipta prostrata*. Against malayan pit viper venom. Pharmaceutical Biology, 45(4), 282-288. <https://doi.org/10.1080/13880200701214805>.
- 236 Sawant, M., Isaac, J. C., & Narayanan, S. (2004). Analgesic studies on total alkaloids and alcohol extracts of *Eclipta alba* (Linn.) Hassk. Phytotherapy Research, 18(2), 111-113. <https://doi.org/10.1002/ptr.1165>.
- 237 Thakur, V. D., & Mengi, S. A. (2005). Neuropharmacological profile of *Eclipta alba* (Linn.) Hassk. Journal of Ethnopharmacology, 102(1), 23-31. <https://doi.org/10.1016/j.jep.2005.05.037>.
- 238 Banji O., Banji D., Annamalai AR., Manavalan R. (2007). Investigation on the effect of *Eclipta alba* on animal models of learning and memory. Indian Journal of Physiology and Pharmacology, 51, 274-278.
- 239 Christyapita D., Divyagnaneswari M., Dinakaran MR. (2007). Oral administration of *Eclipta alba* leaf aqueous extract enhances the non-specific immune responses and disease resistance of *Oreochromis mossambicus*. Fish and shellfish immunology, 23(4), 840-52.

- 240 Zida, P. E., Néya, B. J., Stokholm, M. S., Jensen, S. M., Soalla, W. R., Sérémé, P., & Lund, O. S. (2018). Increasing sorghum yields by seed treatment with an aqueous extract of the plant *Eclipta alba* may involve a dual mechanism of hydropriming and suppression of fungal pathogens. *Crop Protection*, 107, 48-55. <https://doi.org/10.1016/j.cropro.2018.01.001>.
- 241 Roy, R. K., Thakur, M., & Dixit, V. K. (2008). Hair growth promoting activity of *Eclipta alba* in male albino rats. *Archives of Dermatological Research*, 300(7), 357-364. <https://doi.org/10.1007/s00403-008-0860-3>.
- 242 Datta, K., Singh, A. T., Mukherjee, A., Bhat, B., Ramesh, B., & Burman, A. C. (2009). *Eclipta alba* extract with potential for hair growth promoting activity. *Journal of Ethnopharmacology*, 124(3), 450-456. <https://doi.org/10.1016/j.jep.2009.05.023>.
- 243 Yadav, N. K., Arya, R. K., Dev, K., Sharma, C., Hossain, Z., Meena, S., Arya, K. R., Gayen, J. R., Datta, D., & Singh, R. K. (2017). Alcoholic extract of *Eclipta alba* shows in vitro antioxidant and anticancer activity without exhibiting toxicological effects. *Oxidative Medicine and Cellular Longevity*, 2017, 1-18. <https://doi.org/10.1155/2017/9094641>.
- 244 Rao, D.B., Kiran, C.R., Madhavi, Y., Rao, P.K., Rao, T.R. (2009). Evaluation of antioxidant potential of *Clitoria ternata* L. and *Eclipta prostrata* L. *Indian J. Biochem. Biophys.* 46 (3), 247–252.
- 245 Karthikumar S., Vigneswari K., Jegatheesan K. (2007). Screening of antibacterial and antioxidant activities of leaves of *Eclipta prostrata* (L). *Sci. Res. Essays*, 2 (4), 101-04.
- 246 Bhringaraja (*Eclipta alba*) - Plante Ayurvédique Traditionnelle Croissance des cheveux. Consulté le 02 août 2020, à l'adresse <https://www.amazon.fr/BHRINGARAJA-Eclipta-alba-g%C3%A9ules-500mg/dp/B00QBQXZXG>.
- 247 Chemat, F., Rombaut, N., Sicaire, A.-G., Meullemiestre, A., Fabiano-Tixier, A.-S., & Abert-Vian, M. (2017). Ultrasound assisted extraction of food and natural products. Mechanisms, techniques, combinations, protocols and applications. A review. *Ultrasonics Sonochemistry*, 34, 540-560. <https://doi.org/10.1016/j.ultsonch.2016.06.035>.
- 248 Introduction to Centrifugal Partition Chromatography. Consulté le 30 septembre 2020, à l'adresse [https://fr.gilson.com/pub/static/frontend/Gilson/customtheme/en\\_US/images/docs/TN215-02\\_Gilson\\_Introduction\\_CPC.pdf](https://fr.gilson.com/pub/static/frontend/Gilson/customtheme/en_US/images/docs/TN215-02_Gilson_Introduction_CPC.pdf).
- 249 Oka, F., Oka, H., & Ito, Y. (1991). Systematic search for suitable two-phase solvent systems for high-speed counter-current chromatography. *Journal of Chromatography A*, 538(1), 99-108. [https://doi.org/10.1016/S0021-9673\(01\)91626-7](https://doi.org/10.1016/S0021-9673(01)91626-7).
- 250 Berthod, A., Hassoun, M., & Ruiz-Angel, M. J. (2005). Alkane effect in the Arizona liquid systems used in countercurrent chromatography. *Analytical and Bioanalytical Chemistry*, 383(2), 327-340. <https://doi.org/10.1007/s00216-005-0016-7>.
- 251 Berthod, A., Hassoun, M., & Harris, G. (2005). Using the liquid nature of the stationary phase : The elution-extrusion method. *Journal of Liquid Chromatography & Related Technologies*, 28(12-13), 1851-1866. <https://doi.org/10.1081/JLC-200063489>.
- 252 Berthod, A., Ruiz-Angel, M. J., & Carda-Broch, S. (2003). Elution-extrusion countercurrent chromatography. Use of the liquid nature of the stationary phase to extend the hydrophobicity window. *Analytical Chemistry*, 75(21), 5886-5894. <https://doi.org/10.1021/ac030208d>.
- 253 Agnely, M., & Thiébaut, D. (1997). Dual-mode high-speed counter-current chromatography : Retention, resolution and examples. *Journal of Chromatography A*, 790(1), 17-30. [https://doi.org/10.1016/S0021-9673\(97\)00742-5](https://doi.org/10.1016/S0021-9673(97)00742-5).

- 254 Hamzaoui, M., Renault, J.-H., Reynaud, R., & Hubert, J. (2013). Centrifugal partition extraction in the pH-zone-refining displacement mode: An efficient strategy for the screening and isolation of biologically active phenolic compounds. *Journal of Chromatography B*, 937, 7-12. <https://doi.org/10.1016/j.jchromb.2013.07.024>.
- 255 Hamzaoui, M., Hubert, J., Hadj-Salem, J., Richard, B., Harakat, D., Marchal, L., Foucault, A., Lavaud, C., & Renault, J.-H. (2011). Intensified extraction of ionized natural products by ion pair centrifugal partition extraction. *Journal of Chromatography A*, 1218(31), 5254-5262. <https://doi.org/10.1016/j.chroma.2011.06.018>.
- 256 Lee, J., Koo, N., & Min, D. B. (2004). Reactive oxygen species, aging, and antioxidative nutraceuticals. *Comprehensive Reviews in Food Science and Food Safety*, 3(1), 21-33. <https://doi.org/10.1111/j.1541-4337.2004.tb00058.x>.
- 257 Quideau, S., Deffieux, D., Douat-Casassus, C., & Pouységu, L. (2011). Plant polyphenols: Chemical properties, biological activities, and synthesis. *Angewandte Chemie International Edition*, 50(3), 586-621. <https://doi.org/10.1002/anie.201000044>.
- 258 Dayan, N. (Éd.). (2011). Formulating, packaging, and marketing of natural cosmetic products. Wiley.
- 259 Ribeiro, A. S., Estanqueiro, M., Oliveira, M. B., & Sousa Lobo, J. M. (2015). Main benefits and applicability of plant extracts in skin care products. *Cosmetics*, 2(2), 48-65. <https://doi.org/10.3390/cosmetics2020048>.
- 260 Pharm'Aging. (2019, mars 11). Radicaux libres, stress oxydatif et vieillissement cutané. Pharm'Aging. <https://www.pharmaging.com/radicaux-libres-stress-oxydatif-vieillissement-cutane/>.
- 261 Zheng, W., & Wang, S. Y. (2001). Antioxidant activity and phenolic compounds in selected herbs. *Journal of Agricultural and Food Chemistry*, 49(11), 5165-5170. <https://doi.org/10.1021/jf010697n>.
- 262 Ghosh, P. K., & Gaba, A. (2013). Phyto-extracts in wound healing. *Journal of Pharmacy & Pharmaceutical Sciences*, 16(5), 760. <https://doi.org/10.18433/J3831V>.
- 263 Nimse, S. B., & Pal, D. (2015). Free radicals, natural antioxidants, and their reaction mechanisms. *RSC Advances*, 5(35), 27986-28006. <https://doi.org/10.1039/C4RA13315C>.
- 264 Chen, Z., Bertin, R., & Froldi, G. (2013). EC50 estimation of antioxidant activity in DPPH assay using several statistical programs. *Food Chemistry*, 138(1), 414-420. <https://doi.org/10.1016/j.foodchem.2012.11.001>.
- 265 Sridhar, K., & Charles, A. L. (2019). In vitro antioxidant activity of Kyoho grape extracts in DPPH and ABTS assays: Estimation methods for EC50 using advanced statistical programs. *Food Chemistry*, 275, 41-49. <https://doi.org/10.1016/j.foodchem.2018.09.040>.
- 266 Wootton-Beard, P. C., Moran, A. G., & Ryan, L. (2011). Stability of the total antioxidant capacity and total polyphenol content of 23 commercially available vegetable juices before and after in vitro digestion measured by FRAP, DPPH, ABTS and Folin-Ciocalteu methods. *Food Research International*, 44(1), 217-224. <https://doi.org/10.1016/j.foodres.2010.10.033>.
- 267 Tsuji, N., Moriwaki, S., Suzuki, Y., Takema, Y., & Imokawa, G. (2001). The role of elastases secreted by fibroblasts in wrinkle formation: Implication through selective inhibition of elastase activity. *Photochemistry and Photobiology*, 74(2), 283-290. [https://doi.org/10.1562/0031-8655\(2001\)0740283TROESB2.0.CO2](https://doi.org/10.1562/0031-8655(2001)0740283TROESB2.0.CO2).
- 268 Varani, J., Spearman, D., Perone, P., Fligel, S. E. G., Datta, S. C., Wang, Z. Q., Shao, Y., Kang, S., Fisher, G. J., & Voorhees, J. J. (2001). Inhibition of type I procollagen synthesis by damaged collagen in photoaged skin and by collagenase-degraded collagen in vitro. *The American Journal of Pathology*, 158(3), 931-942. [https://doi.org/10.1016/S0002-9440\(10\)64040-0](https://doi.org/10.1016/S0002-9440(10)64040-0).
- 269 Lee, S. Y., Baek, N., & Nam, T. (2016). Natural, semisynthetic and synthetic tyrosinase inhibitors. *Journal of Enzyme Inhibition and Medicinal Chemistry*, 31(1), 1-13. <https://doi.org/10.3109/14756366.2015.1004058>

- 270 Mikšík, I., Sedláková, P., Mikulíková, K., & Eckhardt, A. (2006). Capillary electromigration methods for the study of collagen. *Journal of Chromatography B*, 841(1-2), 3-13. <https://doi.org/10.1016/j.jchromb.2006.02.043>.
- 271 Tsuji, N., Moriwaki, S., Suzuki, Y., Takema, Y., & Imokawa, G. (2001). The role of elastases secreted by fibroblasts in wrinkle formation : Implication through selective inhibition of elastase activity. *Photochemistry and Photobiology*, 74(2), 283-290. [https://doi.org/10.1562/0031-8655\(2001\)074<0283:troesb>2.0.co;2](https://doi.org/10.1562/0031-8655(2001)074<0283:troesb>2.0.co;2).
- 272 Kamkaen, N., Mulsri, N., Treesak, C. (2007). Screening of some tropical vegetables for antityrosinase activity. *Thail. Pharm. Health Sci. J.* 2 (1), 15–19.
- 273 Van Meerloo, J., Kaspers, G. J. L., & Cloos, J. (2011). Cell sensitivity assays : The mtt assay. In I. A. Cree (Ed.), *Cancer Cell Culture : Methods and Protocols* (p. 237-245). Humana Press. [https://doi.org/10.1007/978-1-61779-080-5\\_20](https://doi.org/10.1007/978-1-61779-080-5_20).
- 274 Borges Bubols, G., da Rocha Vianna, D., Medina-Remon, A., von Poser, G., Maria Lamuela-Raventos, R., Lucia Eifler-Lima, V., & Cristina Garcia, S. (2013). The antioxidant activity of coumarins and flavonoids. *Mini Reviews in Medicinal Chemistry*, 13(3), 318-334. <https://doi.org/10.2174/138955713804999775>.
- 275 Ferreira, J. F. S., Luthria, D. L., Sasaki, T., & Heyerick, A. (2010). Flavonoids from *Artemisia annua* L. as antioxidants and their potential synergism with artemisinin against malaria and cancer. *Molecules* (Basel, Switzerland), 15(5), 3135-3170. <https://doi.org/10.3390/molecules15053135>.
- 276 Y. Tu, From *Artemisia annua* L. to Artemisinins : The Discovery and Development of Artemisinins and Antimalarial Agents (2017).
- 277 Chen, Z., Bertin, R., & Froldi, G. (2013). EC50 estimation of antioxidant activity in DPPH assay using several statistical programs. *Food Chemistry*, 138(1), 414-420. <https://doi.org/10.1016/j.foodchem.2012.11.001>.
- 278 A. Shalaby , S.M.M. Shanab. (2013). Comparison of DPPH and ABTS assays for determining antioxidant potential of water and methanol extracts of *Spirulina platensis*, *Indian J. Geo-Mar. Sci.* 42 (5) 556–564.
- 279 Tirzitis, G., & Bartosz, G. (2010). Determination of antiradical and antioxidant activity : Basic principles and new insights. *Acta Biochimica Polonica*, 57(2), 139-142.
- 280 Mira, L., Fernandez, M. T., Santos, M., Rocha, R., Florêncio, M. H., & Jennings, K. R. (2002). Interactions of flavonoids with iron and copper ions : A mechanism for their antioxidant activity. *Free Radical Research*, 36(11), 1199-1208. <https://doi.org/10.1080/1071576021000016463>.
- 281 Wang, S., Cai, T., Liu, H., Yang, A., & Xing, J. (2019). Liquid chromatography-tandem mass spectrometry assay for the simultaneous determination of three major flavonoids and their glucuronidated metabolites in rats after oral administration of *Artemisia annua* L. extract at a therapeutic ultra-low dose. *Journal of Separation Science*, 42(21), 3330-3339. <https://doi.org/10.1002/jssc.201900668>.
- 282 Vihakas, M. (2014). Flavonoids and other phenolic compounds : Characterization and interactions with lepidopteran and sawfly larvae. *Undefined*. <https://www.semanticscholar.org/paper/Flavonoids-and-other-phenolic-compounds%3A-and-with-Vihakas/665273ecd8ac58765f3ea6051efca116b7180115>.
- 283 Miao YP, Wen R, Aoshima H, Zhou PG. (2004). Synthesis and antioxidative activity of 2-substituted phenyl-5-(3' indolyl)-oxazole derivatives. *Yao Xue Xue Bao*;39:37–40.
- 284 Madhan, B., Krishnamoorthy, G., Rao, J. R., & Nair, B. U. (2007). Role of green tea polyphenols in the inhibition of collagenolytic activity by collagenase. *International Journal of Biological Macromolecules*, 41(1), 16-22. <https://doi.org/10.1016/j.ijbiomac.2006.11.013>.
- 285 Gilabert, M., Marcinkevicius, K., Andujar, S., Schiavone, M., Arena, M. E., & Bardón, A. (2015). Sesqui- and triterpenoids from the liverwort *Lepidozia chordulifera* inhibitors of bacterial biofilm and elastase activity of human pathogenic bacteria. *Phytomedicine*, 22(1), 77-85. <https://doi.org/10.1016/j.phymed.2014.10.006>.

- 286 Ying, Q. L., Rinehart, A. R., Simon, S. R., & Cheronis, J. C. (1991). Inhibition of human leucocyte elastase by ursolic acid. Evidence for a binding site for pentacyclic triterpenes. *Biochemical Journal*, 277(Pt 2), 521-526.
- 287 Jacob, V., Hagai, T., & Soliman, K. (2011). Structure-activity relationships of flavonoids. *Current Organic Chemistry*, Volume 15, Number 15, 2011, pp. 2641-2657(17). <https://doi.org/info:doi/10.2174/138527211796367309>.
- 288 López-Hortas, L., Pérez-Larrán, P., González-Muñoz, M. J., Falqué, E., & Domínguez, H. (2018). Recent developments on the extraction and application of ursolic acid. A review. *Food Research International*, 103, 130-149. <https://doi.org/10.1016/j.foodres.2017.10.028>.
- 289 Lee, S. K., Mbwambo, Z. H., Chung, H., Luyengi, L., Gamez, E. J., Mehta, R. G., Kinghorn, A. D., & Pezzuto, J. M. (1998). Evaluation of the antioxidant potential of natural products. *Combinatorial Chemistry & High Throughput Screening*, 1(1), 35-46.
- 290 Tagliazucchi, D., Verzelloni, E., Bertolini, D., & Conte, A. (2010). In vitro bio-accessibility and antioxidant activity of grape polyphenols. *Food Chemistry*, 120(2), 599-606. <https://doi.org/10.1016/j.foodchem.2009.10.030>.
- 291 Benzie, I. F. F., & Strain, J. J. (1996). The ferric reducing ability of plasma (Frap) as a measure of “antioxidant power” : The frap assay. *Analytical Biochemistry*, 239(1), 70-76. <https://doi.org/10.1006/abio.1996.0292>.
- 292 Apak, R., Güçlü, K., Özyürek, M., & Karademir, S. E. (2004). Novel total antioxidant capacity index for dietary polyphenols and vitamins c and e, using their cupric ion reducing capability in the presence of neocuproine : Cuprac method. *Journal of Agricultural and Food Chemistry*, 52(26), 7970-7981. <https://doi.org/10.1021/jf048741x>.
- 293 Dinis, T. C. P., Madeira, V. M. C., & Almeida, L. M. (1994). Action of phenolic derivatives (Acetaminophen, salicylate, and 5-aminosalicylate) as inhibitors of membrane lipid peroxidation and as peroxy radical scavengers. *Archives of Biochemistry and Biophysics*, 315(1), 161-169. <https://doi.org/10.1006/abbi.1994.1485>.
- 294 Florent Olivon (2018). Nouvelle stratégie de priorisation pour l'étude des produits naturels par l'approche des réseaux moléculaires multi-informatifs. <http://www.theses.fr/2018SACLS383>
- 295 Pierre Barbier (2019). Utilisation de la métabolomique pour l'étude de l'encéphalopathie hépatique et développement de nouvelles approches pour l'acquisition de données par spectrométrie de masse. <http://www.theses.fr/2019SACLS193>.
- 296 Chiocchio, I., Mandrone, M., Sanna, C., Maxia, A., Tacchini, M., & Poli, F. (2018). Screening of a hundred plant extracts as tyrosinase and elastase inhibitors, two enzymatic targets of cosmetic interest. *Industrial Crops and Products*, 122, 498-505. <https://doi.org/10.1016/j.indcrop.2018.06.029>.

## Souhila MESSAILI

### Criblage de molécules naturelles d'intérêt cosmétique par combinaison d'outils analytiques, bioinformatiques et métabolomiques

#### Résumé :

Les produits naturels d'origine végétale sont devenus une source majeure de composés bioactifs pour les industries cosmétiques. Cependant, la détermination du principe actif, c'est-à-dire l'identification, l'isolement et le test de chaque molécule, est complexe, longue et fastidieuse compte tenu de la diversité moléculaire des extraits. Dans ce contexte, ces travaux de thèse portent sur la combinaison d'outils analytiques, bioinformatiques et métabolomiques dans le but de développer des approches simples, rapides et efficaces pour cibler et identifier des molécules d'intérêt.

Par conséquent, ces travaux ont conduit d'abord à la mise en place de réseaux moléculaires quantitatifs afin d'étudier l'accumulation de métabolites spécialisés dans des callus cellulaires de lin. Ensuite, au développement d'une étude métabolomique par UHPLC-HRMS, associée aux informations de déréplication et de similarités structurales obtenues par les réseaux moléculaires afin de mettre en évidence les variations chimiques par des analyses statistiques, et de caractériser les composés discriminants de deux extraits de plante à différents stades de développement. Enfin, à l'association des réseaux moléculaires et du fractionnement des extraits par chromatographie de partage centrifuge pour la caractérisation phytochimique de deux extraits de plante ayant montré des propriétés d'intérêt pour des applications cosmétiques par des tests chimiques, enzymatiques et cellulaires. Cette combinaison a permis d'obtenir et de tester des fractions simplifiées facilitant ainsi le criblage et l'identification des molécules responsables de l'activité biologique des extraits de plante étudiés.

Mots clés : Chimie analytique, Produits naturels, Phytochimique, Spectrométrie de masse, Métabolomique, Réseaux moléculaires.

### Screening of natural molecules of cosmetic interest by combining analytical, bioinformatic and metabolomic tools

#### Summary :

Natural products of plant origin have emerged a major source of bioactive compounds for the cosmetic industry. However, determination of the active ingredient, i.e. identification, isolation and testing of each molecule, is complex, long and tedious taking into account molecular diversity of the extracts. In this context, this PhD work focuses on the combination of analytical, bioinformatic and metabolomic tools in order to develop simple, rapid and efficient approaches to target and identify molecules of interest.

Consequently, this work initially conducted to the implementation of quantitative molecular networks in order to study the accumulation of specialized metabolites in flax cell callus. Then, to develop a metabolomic study by UHPLC-HRMS, associated with the information of dereplication and structural similarities obtained by molecular networks in order to highlight chemical variations by statistical analyses, and to characterize the discriminating compounds of two plants extracts at different stages of development. Finally, to establish a combination of molecular networks and fractionation by centrifugal partition chromatography for the phytochemical characterization of two plant extracts having shown biological properties interesting for cosmetic applications by chemical, enzymatic and cellular assays. This combination made it possible to obtain and test simplified fractions thus facilitating the screening and identification of the molecules responsible for the biological activity of studied plant extracts.

Keywords : Analytical Chemistry, Natural Products, Phytochemistry, Mass spectrometry, Metabolomic, Molecular Networks.



Institut de Chimie Organique et Analytique  
Pôle de chimie, Rue de Chartres, 45100  
Orléans

CoMetosciences  
Recherche & Industrie, innovons ensemble