Stratégies innovantes pour une valorisation d'extraits de plantes en cosmétique : Mise en oeuvre d'un outil de profilage métabolique et recherche de nouvelles activités biologiques

Glorianne Jouravel

▶ **To cite this version:**

Glorianne Jouravel. Stratégies innovantes pour une valorisation d'extraits de plantes en cosmétique : Mise en oeuvre d'un outil de profilage métabolique et recherche de nouvelles activités biologiques. Alimentation et Nutrition. Université d'Orléans, 2018. Français. NNT : 2018ORLE2017 . tel-04213588

HAL Id: tel-04213588
https://theses.hal.science/tel-04213588

Submitted on 21 Sep 2023

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

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

UNIVERSITÉ D'ORLÉANS

ÉCOLE DOCTORALE
SANTE, SCIENCES BIOLOGIQUES ET CHIMIE DU VIVANT

Laboratoires : Institut de Chimie Organique et Analytique / Greenpharma

THÈSE présentée par :
Glorianne JOURAVEL

soutenue le : **11 juin 2018**

pour obtenir le grade de : **Docteur de l'université d'Orléans**
Discipline/ Spécialité : Chimie Analytique - Phytochimie

Stratégies innovantes pour une valorisation d'extraits de plantes en cosmétique :
Mise en œuvre d'un outil de profilage métabolique et recherche de nouvelles activités biologiques

THÈSE dirigée par :
Claire ELFAKIR Professeur, Université d'Orléans

RAPPORTEURS :
Jane HUBERT Maître de conférences-HDR, Université de Reims
Sylvie MICHEL Professeur, Université de Paris Descartes

JURY :
Marcel HIBERT Professeur, Université de Strasbourg
Jane HUBERT Maître de conférences-HDR, Université de Reims
Sylvie MICHEL Professeur, Université de Paris Descartes
Pascal BONNET Professeur, Université d'Orléans
Bénédicte SENNELIER-PORTET Docteur, Groupe ROCHER
Emilie DESTANDAU Maître de conférences-HDR, Université d'Orléans
Claire ELFAKIR Professeur, Université d'Orléans
Franck HIMBERT Docteur, Greenpharma

Communications écrites et orales

PUBLICATION

JOURAVEL, G.; GUÉNIN, S.; BERNARD, F.-X.; ELFAKIR, C.; BERNARD, P.; HIMBERT, F. New Biological Activities of *Lythrum salicaria* L.: Effects on Keratinocytes, Reconstructed Epidermis and Reconstructed Skins, Applications in Dermo-Cosmetic Sciences. *Cosmetics* **2017**, *4*, 52.

COMMUNICATIONS ORALES*

- **Biotechnocentre**, 13-14 octobre 2016, SEILLAC, 29ème colloque Biotechnocentre

Développement d'un outil analytique GAINS 1.1 (GREENPHARMA AUTOMATED IDENTIFICATION of NATURAL STRUCTURES)

- **SEP 2017**, 28-30 mars 2017, PARIS, 12ème congrès francophone sur les sciences séparatives et les couplages de l'AFSEP

Développement d'une approche globale de profilage métabolique d'extraits végétaux

COMMUNICATIONS PAR AFFICHES*

- **Cosm'INNOV**, 24-25 mai 2016, ORLEANS, 4th Edition Cosmetic Innovation Days

Development of GAINS 1.0 [Greenpharma Automated Identification of Natural Structures]

- **SEP 2017**, 28-30 mars 2017, PARIS, 12ème congrès francophone sur les sciences séparatives et les couplages de l'AFSEP.

Concept d'un outil support au phytochimiste : GAINS 'GREENPHARMA AUTOMATED IDENTIFICATION OF NATURAL STRUCTURES'

* Pour l'ensemble de ces communications, les auteurs sont : JOURAVEL Glorianne, BLONDEAU Sylvain, COLAS Cyril, DESTANDAU Emilie, DO Quoc-Tuan, BERNARD Philippe, ELFAKIR Claire, HIMBERT Franck.

Remerciements

Le travail de doctorat a été réalisé au sein de la société Greenpharma S.A.S et en collaboration avec l'Institut de Chimie Organique et Analytique de l'université d'Orléans.

A ce titre, je remercie **Franck HIMBERT** qui a été mon encadrant direct au sein de l'entreprise. Nous avons réussi à avancer ensemble sur des projets transversaux et il m'a toujours accordé le temps nécessaire pour que nous réussissions à atteindre les objectifs que nous nous fixions, et aussi pour répondre à mes nombreuses interrogations. Bien souvent, sa vision globale sur les projets m'a éclairé et aidé à avancer. J'adresse à **Philippe BERNARD**, PDG de Greenpharma, ma reconnaissance pour avoir pensé que je serai à la hauteur du travail de recherche à effectuer et de m'avoir proposé de faire le doctorat au sein de l'entreprise. Philippe et Franck m'ont parfois ''challengé'' et cela m'a permis d'acquérir une certaine autonomie et appris à toujours trouver une solution pour avancer.

Aussi, au sein de l'ICOA dont je remercie la direction (**Olivier MARTIN** au moment du début de la thèse et **Pascal BONNET** actuellement), j'ai eu l'opportunité d'être encadrée par **Claire ELFAKIR** en tant que directrice de thèse et **Emilie DESTANDAU** en tant qu'encadrante. Elles ont apporté un regard neuf et pertinent à mon projet et m'ont toujours bien conseillée notamment lors de la préparation de communications écrites et orales. Après qu'elles aient été mes enseignantes en master, nous avons appris à nous connaître d'une nouvelle façon et je peux dire que nous avons réussi à bien nous accorder. Elles ont su se rendre disponibles lorsque j'avais besoin d'elles –même en urgence-. Je remercie Emilie pour tout ce qu'elle fait pour les étudiants qu'elle suit avec intérêt. Pour ma part, elle m'a souvent écoutée et a participé à ce que je sois confortée, au moment de master, que l'étude des plantes et leur valorisation était ce que je voulais faire.

Je remercie l'ensemble des personnes ayant accepté de faire partie de mon jury de thèse : **Sylvie MICHEL**, **Jane HUBERT**, **Marcel HIBERT**, **Bénédicte SENNELIER-PORTET** et **Pascal BONNET**. J'exprime ma reconnaissance à Sylvie MICHEL et Jane HUBERT qui ont accepté d'être les rapporteurs. Je souhaite à tous une agréable lecture du manuscrit.

Concernant les travaux effectués, j'ai eu la chance d'interagir avec de nombreuses personnes de différents laboratoires. Je souhaite remercie le **Groupe ALES** et le **groupe ROCHER** avec qui nous avons collaboré notamment autour de la Tulipe noire et de la Réglisse glabre. Ils nous ont fait confiance et ont accepté que nous communiquions les résultats obtenus lors de ces deux études. Je

tiens également à évoquer le partenariat avec **Bioalternatives** pour les tests d'activités biologiques. Merci à **François-Xavier BERNARD** et **Samuel GUENIN**. Ce dernier a travaillé plus particulièrement avec moi à la publication autour de la Salicaire. Ses conseils de présentation des résultats et ses explications en biologie cellulaire ont été précieux.

J'ai aussi beaucoup appris à propos de la spectrométrie de masse aux côtés de **Cyril COLAS** de la plateforme HRMS située à l'ICOA. Il a une très grande connaissance sur de nombreux sujets scientifiques et aime la partager. Sa pédagogie et ses réponses à mes nombreuses questions m'ont beaucoup aidé. Nous avons souvent pu échanger autour des méthodes actuelles d'analyses des composés naturels. Je le remercie pour l'ensemble des analyses de molécules et d'extraits qu'il a effectué pour mon travail de recherche.

Je tiens à remercier **Sylvain BLONDEAU**, informaticien de Greenpharma avec qui j'ai travaillé à l'élaboration de l'outil GAINS. Il a été très à l'écoute pour comprendre exactement ce qu'il fallait faire des données de spectrométrie de masse dont il ne savait pas grand-chose. Son travail a été au cœur du développement de l'interface et des algorithmes de l'outil.

Au sein de Greenpharma, je commencerai par remercier l'ensemble de l'équipe que ce soit du côté ''Bureaux'' ou ''Labo''. J'ai beaucoup aimé rejoindre quotidiennement d'abord en stage puis en doctorat : **Florence USO** (Super Florence), **Matthieu FRIAS** (El Muchacho), **Emilie GENTY** (la chirurgienne HPLC), **Magali LORION**, **Christelle PILLARD**, **Franck HIMBERT**, **Sylvain BLONDEAU**, **Eric PAYAT**, **Véronique CHIQUIER**, **Quoc-Tuan DO**, **Alexandra COIC** et **Philippe BERNARD**. J'espère qu'eux aussi ont apprécié travailler avec moi.

Je vais remercier globalement les membres de l'ICOA qu'ils soient enseignants-chercheurs, administratifs ou étudiants. Il y a beaucoup de personnes sympathiques avec qui il est simple et agréable d'aller discuter. Un merci plus particulier à **Pascal BOUYSSOU** et **Christel LOPIN-BON** d'avoir répondu à mes questions autour de la RMN. Je remercie aussi **Julien DIHARCE** pour ses conseils de rédaction, ses rétro-plannings, et ses encouragements permanents.

Il me reste à dire simplement à **Daniel**, **Claudine**, **Grace** et **Gabrielle** qu'ils ont été réconfortants et aussi patients de m'avoir écouté parler du doctorat et de phytochimie -même s'ils ne comprenaient pas toujours de quoi je parlais -. Ils m'ont motivé et encouragé en direct et plus souvent par téléphone. La même chose est vraie pour **Béthy** et **Marie** un peu loin géographiquement mais quand on se parle, la plupart du temps, j'ai le sourire et elles me font rire. Et puis, je remercie mon ami d'Orléans et binôme de master **Baptiste** qui est également un super binôme dans la vie de tous les jours.

Sommaire

Liste des abréviations ***1***

Table des illustrations ***2***

Table des tableaux ***7***

Introduction générale ***9***

Chapitre 1 Valorisation des substances naturelles : De l'analyse phytochimique à l'usage cosmétique -Données bibliographiques- ***13***

1. Introduction **14**

2. Recherche & Développement sur les produits naturels pour le domaine cosmétique **14**

2.1 Recherche et Développement en cosmétique 14

2.2 Cosmétiques et ingrédients actifs 15

2.3 Valorisation dans les domaines Cosmétique, Dermatologique et Nutraceutique 15

2.4 Aspects réglementaires ou toxicologiques à prendre en compte 17

3. La peau, organe ciblé par les produits cosmétiques **17**

3.1 Structuration globale 17

3.2 Fonctions essentielles de la peau 19

3.3 Intéraction avec les cosmétiques 20

3.3.1. Les différents types cellulaires ciblés par les cosmétiques 20

3.3.2. Fonctions cosmétiques recherchées 20

4. Analyse phytochimique **21**

4.1 Qu'est-ce que la phytochimie ? 21

4.1.1. Définition de la phytochimie 21

4.1.2. Objet d'étude de la phytochimie : les métabolites 21

4.1.3. Méthodes générales employées en phytochimie 23

4.2 Techniques analytiques majeures 25

4.2.1. Spectrométrie de masse et résonance magnétique nucléaire : définitions, performances et intérêts 26

4.2.2. Focus sur la technique UHPLC-ESI-HRMS 28

4.2.3. Performances et intérêts principaux de la détection en ESI-Q-TOF 32

4.3 Profilage métabolique : méthodes fondamentales et approches actuelles 33

4.3.1. Problématique : défis rencontrés en recherche sur les produits naturels 33

4.3.2. Déréplication : Retrouver les molécules connues 34

4.3.3. Stratégies et plateformes pour le traitement, l'organisation et l'annotation des données 37

5. Conclusion 51

Chapitre 2 Développement de GAINS : (Greenpharma Automated Identification of Natural Structures) Approche et Mise en œuvre 53

1. Introduction 54

2. Approche globale de GAINS 54

2.1 Analyses en LC-HRMS 55

2.1.1. Principe et techniques mis en jeu 55

2.1.2. Modes d'ionisation 56

2.2 Prétraitement des données brutes 56

2.2.1. Applications de fonctions de calibration et de bruit de fond spectral 56

2.2.2. Génération des pics et des listes de masses des ions détectés 57

2.3 Calculs des masses monoisotopiques 58

2.3.1. Principe et paramétrages techniques des calculs automatiques 58

2.3.2. Processus de recherche et de calcul des masses monoisotopiques 59

2.3.3. Application d'une fonction de score 61

2.4 Requête au sein de la Greenpharma Database (GPDB) 62

2.4.1. GPDB, une base de données dédiée aux produits naturels 62

2.4.2. Requête au sein de la base GPDB 63

2.5 Exploration des données de MS/MS 63

2.5.1. Analyse des données de fragmentation (MS/MS) : principe 63

2.5.2. Utilisation d'un outil automatique de fragmentation *in silico* : Metfrag 64

2.5.3. Gestion des groupements glycosylés ou acylés 65

3. Développement de l'outil 69

3.1 Traduction du travail du phytochimiste en code informatique 69

3.1.1. Démarche de travail pour développer l'outil 69

3.1.2. Langages de programmation 69

3.2 Mise en œuvre technique : mise au point de l'outil 70

3.2.1. Constitution d'un mélange standard 70

3.2.2. Analyses des standards pour valider l'outil 72

3.2.3. Mise au point : choix et réglages des poids statistiques des adduits 73

3.3 Résultats de validation : données de MS 74

3.3.1. Performances de GAINS sur les données MS obtenues en mode FIA -1er réplicat- 74

3.3.2. Performances de GAINS sur les données MS obtenues en mode FIA -2nd réplicat- 80

3.3.3. Comparaison des résultats de deux réplicats en FIA-MS 82

3.3.4. Performances de GAINS sur les données MS obtenues en LC-HRMS 82

3.4 Validation de l'outil : traitement des données MS/MS 84

3.4.1. Principe de l'évaluation 84

3.4.2. Réalisation de l'évaluation 85
3.4.3. Résultats du test de Metfrag 85
3.4.4. Perspectives d'amélioration dans le traitement des données MS/MS 90
3.5 Durées des étapes de traitement et de calcul 90

4. Limites de l'outil 91
4.1 Limites théoriques intrinsèques à la structure de l'outil 91
4.2 Limites pratiques liées aux performances de l'outil en fonctionnement 93

5. Conclusion 95

Chapitre 3 Applications de GAINS Différentes fonctionnalités appliquées à l'étude d'extraits de plantes 97

1. Introduction 98

2. Etude d'un extrait de pétales de Tulipe noire : Identification de chémo-marqueurs et intérêt de l'hydrolyse acide de composés 99
2.1 Contexte de l'étude et objectifs 99
2.2 Présentation de la Tulipe noire 99
2.3 Hydrolyse acide de l'extrait et analyse par HPLC-UV 99
2.4 Analyses LC-HRMS et résultats obtenus avec GAINS 100
2.5 Déréplication et proposition d'identification de chémo-marqueurs 102
2.5.1. Note explicative de la numérotation des pics dans les tableaux de résultats 102
2.5.2. Résultats 102
2.5.3. Raisonnement tenu à partir des données analytiques et apport des données MS/MS : exemple détaillé pour le composé le plus intense 106
2.5.4. Déréplication ciblée et non ciblée pour les trois composés majoritaires et apport de Metfrag 108
2.6 Conclusion 109

3. Etude phytochimique d'un extrait de fleurs de Primevère officinale : de l'approche classique à l'approche moderne 111
3.1 Contexte de l'étude 111
3.2 Description de la plante 111
3.3 Etude préliminaire de l'extrait hydro-alcoolique de fleurs de Primevère officinale 113
3.4 Approche classique : isolement, purification et identification par RMN 114
3.4.1. Prétraitement par SPE (extraction sur phase solide) 114
3.4.2. Fractionnement en HPLC-préparative 117
3.4.3. Purification de composés par HPLC semi-préparative 119
3.4.4. Analyses FIA-HRMS, RMN et hydrolyse acide 121

3.4.5. Conclusion sur l'approche classique pour décrire le contenu de l'extrait de Primevère vraie 129
3.5 Approche globale : analyse en LC-HRMS et utilisation de GAINS 130
3.5.1. Résultats 132
3.6 Comparaison entre la méthode classique et la méthode utilisant GAINS 138
3.7 Conclusion 140
3.7.1. Bilan sur la comparaison des techniques d'exploration du contenu 140
3.7.2. Bilan sur les composés mis en évidence 141

4. Etude d'un extrait de réglisse glabre : Obtention d'une vue globale du contenu phytochimique et recherche d'isoflavones 143
4.1 Contexte de l'étude et objectifs 143
4.2 Présentation de la réglisse glabre 143
4.3 Analyses LC-HRMS 144
4.4 Identification des pics principaux 146
4.4.1. Note concernant la synthèse des résultats et numérotations des pics 146
4.4.2. Résultats du traitement des données de LC-HRMS 146
4.4.3. Confirmation de la détection d'isoflavones : différence de fragmentation par rapport aux flavones 152
4.5 Fonction de recherche d'une famille phytochimique particulière : exemple avec les isoflavones 153
4.5.1. Présentation et intérêts 153
4.5.2. Exemple d'application : cas des isoflavones 153
4.5.3. Focus sur une molécule particulière : présence ou absence de la DAIDZEINE 159
4.6 Conclusion de l'étude de l'extrait de réglisse 161

5. Conclusion du chapitre d'applications de GAINS 163

Chapitre 4 Etude de Lythrum salicaria L. Caractérisation et valorisation 165

1. Introduction 166

2. Etudes bibliographiques 167
2.1 Présentation de la salicaire (*Lythrum salicaria* L.) 167
2.1.1. Aspect botanique 167
2.1.2. Connaissances phytochimiques 168
2.1.3. Usages connus et activités biologiques reconnues 170
2.2 La peau et sa physiologie 172

3. Première approche exploratoire sur les potentialités dermo-cosmétiques de la salicaire 173

4. Etude phytochimique de *Lythrum salicaria* L. 175

4.1 Matériel végétal et extraction 175

4.2 Evaluation rapide des teneurs en composés polaires 175

4.2.1. Teneur en sucres neutres totaux 175

4.2.2. Contenu en composés phénoliques 176

4.3 Caractérisation de l'extrait par HPLC-DAD-MS 176

4.4 Détermination structurale des composés majoritaires 179

4.4.1. Isolement par HPLC semi-préparative 179

4.4.2. Détermination structurale du composé **D** 180

4.4.3. Détermination structurale des composés majoritaires (**B** et **C**) 181

4.5 Quantification des composés majoritaires 185

5. Evaluation des activités biologiques de l'extrait 186

5.1 Cultures cellulaires et traitements 186

5.2 Effet pro-différenciant de l'extrait de *Lythrum salicaria* 187

5.2.1. Niveaux d'expressions de gènes impliqués dans la différenciation suivis par PCR quantitative en temps réel 187

5.2.2. Expression de protéines caractéristiques de kératinocytes différenciés suivie par immunofluorescence 188

5.2.3. Stimulation de l'expression de la protéine TGK dans les kératinocytes suivie par immunofluorescence 189

5.2.4. Caractérisation histologique d'épidermes humains reconstruits 190

5.3 Effet de l'extrait sur la structure de la peau 192

5.3.1. Amélioration de la fonction barrière de la peau sur des modèles de peaux reconstruites 192

5.3.2. Effets bénéfiques sur la matrice extracellulaire 193

6. Conclusion 197

Conclusion générale 199

Références bibliographiques 203

ANNEXES 215

Liste des abréviations

R&D Recherche et Développement
MS Spectrométrie de Masse (*Mass Spectrometry*)
MS/MS Spectrométrie de Masse en tandem
RMN Résonance Magnétique Nucléaire
CPC Chromatographie de Partage Centrifuge
SPE Extraction sur phase solide (*Solid Phase Extraction*)
FIA Analyse par injection en flux continu (*Flow Injection Analysis*)
HPLC Chromatographie Liquide Haute Performance (*High Performance Liquid Chromatography*)
UHPLC Chromatographie Liquide Ultra Haute Performance
RP Reverse-Phase (Phase Inverse)
HRMS Spectrométrie de Masse Haute Résolution (*High Resolution Mass Spectrometry*)
ESI Ionisation par electrospray (*Electrospray Ionisation*)
TOF Temps de vol (*Time Of Flight*)
Q-TOF Analyseurs en série de type Quadripôle et Temps de vol
CCM Chromatographie sur Couche Mince
DAD Détecteur à Barrette de Diodes
UV Ultraviolet
DEDL Détecteur évaporatif à diffusion de la lumière
GAINS *Greenpharma Automated Identification of Natural Structures*
GPDB Base de Données Greenpharma *(GreenPharma DataBase)*
MSI *Metabolomics Standards Initiative*
CO_2 Dioxyde de carbone
H_2O Eau
D_2O Eau deutérée
MeOH Méthanol
MeOH-d_4 Méthanol deutéré
EtOH Ethanol
ACN Acétonitrile
DMSO Dimethyl Sulfoxide
DMSO-d_6 Dimethyl Sulfoxide deutéré
Glu : Glucose
Gal : Galactose
Rha : Rhamnose

Table des illustrations

Figure 1 Schéma de coupe de peau humaine montrant les 3 grands blocs constitutifs d'après [5] 18

Figure 2 Schéma de coupe de peau : détail des éléments et annexes cutanées ancrées dans le derme, d'après [6]. 19

Figure 3 Représentation schématique des voies de biosynthèse rencontrées dans le règne végétal. Les encadrés gris indiquent les grandes classes phytochimiques qui regroupent diverses familles phytochimiques (encadré vert). A partir du glucose, différentes voies de biosynthèse se sont développées (encadré marron). Les liaisons existantes entre les voies, les classes et les familles sont représentées par les flèches..... 22

Figure 4 Schéma d'une source ESI (figures provenant du site de l'université de Bristol [31]). A : Vue générale de la source d'ionisation couplée en amont avec un système de chromatographie ou une seringue. B : Principe de formation d'ions positifs dans la source ESI..... 31

Figure 5 Utilisation de la RMN pour la recherche sur les produits naturels : avancées techniques et stratégies récentes. Schéma tiré de la revue de Halabalaki de 2014 [11]. 40

Figure 6 Schéma de spectre MS en mode d'ionisation positif ; illustration du principe d'annotation par évaluation de la différence (Δ_x) entre les différents ions générés par formation d'adduits, de dimères, de pertes de neutres, de fragments en source et présence d'isotopes..... 43

Figure 7 Exemple de réseaux moléculaires obtenus à partir de données MS/MS d'extraits de plante, d'après un papier de *Nothias-Scaglia et al*. ([93]..... 49

Figure 8 Approche globale de GAINS : étapes mises en jeu..... 54

Figure 9 Exemple de chromatogrammes LC-MS en mode positif et mode négatif reconstruits avec l'algorithme DISSECT sous *Data Analysis* (axe x : intensité, axe y : temps de rétention, axe z : *m/z)* .. 57

Figure 10 Exemple du visuel des résultats de calculs des masses monoisotopiques effectués au sein de l'interface GAINS : la molécule analysée à une masse théorique de 258,13683 $g.mol^{-1}$. La première ligne montre l'hypothèse moyenne proposée d'après la correspondance de 3 ions distincts issus des deux polarités. 60

Figure 11 Schéma de la méthode employée pour effectuer une requête dans la base de données afin de trouver les candidats structuraux pour caractériser une molécule inconnue ; x la valeur de masse monoisotopique moyenne expérimentale dotée du meilleur score et issue de l'annotation des ions du spectre et Δx la tolérance en masse pour effectuer la requête..... 63

Figure 12 Structure de deux dissaccharides isomères de position constitués par un glucose et un rhamnose: le rutinoside et le neohesperidoside. 68

Figure 13 Répartition des composés du mélange test en fonction des grandes classes phytochimiques (données en % sur un total de 150 composés) 70

Figure 14 Distribution normalisée du nombre de composés en fonction de leurs masses molaires pour l'ensemble des molécules de GPDB et pour celles du mélange standard évalué 71

Table des illustrations.

Figure 15 Performances de GAINS en FIA-HRMS obtenues lors du 1er réplicat pour le mélange de test de 150 molécules standard 75

Figure 16 Exemple de molécule détectée mais non retrouvée par GAINS : cas d'un sesquiterpène pour lequel l'ion pseudo-moléculaire (en pointillé) n'est pas présent dans le spectre de masse expérimental 76

Figure 17 Exemple de molécule détectée dont la masse est retrouvée sans être classée en meilleure hypothèse : cas d'un composé phénolique dont la fonction ester subit une fragmentation en source 78

Figure 18 Spectre FIA-ESI-MS positif du composé n° 10 de masse monoisotopique 318,073955 $g.mol^{-1}$ dont la fragmentation en source est majoritaire 79

Figure 19 Spectre FIA-ESI-MS négatif du composé n° 10 de masse monoisotopique 318,073955 $g.mol^{-1}$ dont la fragmentation en source est majoritaire 79

Figure 20 Performances de GAINS en FIA obtenues lors du second réplicat sur le mélange test de 150 molécules standard 80

Figure 21 Performances de GAINS en couplage LC-HRMS obtenues pour le mélange test 83

Figure 22 Bilan de l'utilisation de Metfrag sur le classement des structures fragmentées *in silico* : pourcentage de molécules retrouvées en fonction du score considéré. Le score est calculé et normalisé entre 0 et 1 par le module Metfrag. 86

Figure 23 Bilan de l'utilisation de Metfrag sur le classement des structures fragmentées *in silico* : pourcentage de molécules retrouvées en fonction du rang de classement considéré. 87

Figure 24 Sélectivités atteintes avec Metfrag en fonction du nombre de candidats structuraux associés à une masse donnée 89

Figure 25 Chromatogrammes de l'extrait de tulipe noire avant hydrolyse (tracé noir) et après hydrolyse (tracé bleu) en UV 365 nm 100

Figure 26 Chromatogrammes UV 254 nm (tracé bleu) et MS-ESI^+ de l'extrait de tulipe noire à 1000 ppm 101

Figure 27 Chromatogrammes UV 254 nm (tracé vert) et MS-ESI^- de l'extrait de tulipe noire à 1000 ppm 101

Figure 28 Structure d'un noyau flavonol dont la caractéristique est l'hydroxylation en C3 du cycle 103

Figure 29 Structure du kaempférol (tétrahydroxyflavone) 103

Figure 30 Structure de la quercétine (pentahydroxyflavone) 103

Figure 31 Chromatogrammes UV 254 nm (tracé bleu) et MS-ESI^+ de l'extrait de tulipe noire. Zoom de 10 à 30 minutes 104

Figure 32 Spectre MS-ESI positif du pic 72P 107

Figure 33 Spectre MS/MS du pic 72P à partir de l'ion précurseur $[M+H]^+$ 107

Figure 34 Fragmentation de l'adduit sodé d'un flavonoïde diglycosylé. Deux ruptures distinctes sont observées en MS/MS: 1) perte de la génine et 2) perte du sucre terminal 108

Figure 35 *Primula veris* en fleurs à Orléans, Loiret, Mars 2018. A gauche : à l'état naturel, à droite : hampe florale 111

Figure 36 Structures des saponosides décrites dans *Primula veris*, illustration d'après [109] 112

Figure 37 Chromatogrammes (UV 365 nm) de l'extrait de *Primula veris* avant hydrolyse (bleu) et après hydrolyse (noir). 113

Figure 38 Chromatogrammes HPLC-DEDL de l'extrait de *Primula veris* avant purification (noir) et après purification sur SPE C_{18} (bleu) 116

Figure 39 Chromatogrammes HPLC-UV 254 nm de l'extrait de *Primula veris* avant purification (noir) et après purification sur SPE C_{18} (bleu) 116

Figure 40 Superposition des chromatogrammes (normalisés) de chaque fraction (F1 à F9) issue du traitement en HPLC préparative. Analyse en HPLC analytique en mode isocratique MeOH 40 %. Sur la droite : quantité obtenue après séchage de chaque de fraction. 118

Figure 41 Traitement de fractionnement HPLC et bilan massique pour l'extrait de *Primula veris* : rendement de chaque étape mise en jeu. 119

Figure 42 Chromatogramme de l'extrait de primevère vraie après traitement SPE analysé en RP-HPLC – UV 254 nm. Les pics **A** à **G** correspondent aux composés purifiés par HPLC semi-préparative. 120

Figure 43 Spectre MS/MS isCID (*In Source Collision Induced Dissociation*) du produit **D** 125

Figure 44 Structure du kaempférol 3-O-(2''-O-α-rhamnosyl-6''-O-β-glucosyl)-β-glucoside 126

Figure 45 Structure de l'isorhamnetin- 3-O-(2''-O-α-rhamnosyl-6''-O-α-rhamnosyl)-β-galactoside.. 126

Figure 46 Structure de l'isorhamnetin-3O-(2,6-di-O-β-D-glucopyranosyl-β-D-glucopyranoside) 127

Figure 47 Structure de l'isorhamnetin-3-O-rutinoside 128

Figure 48 Structure de l'isorhamnetin-3-O-robinoside 128

Figure 49 Chromatogrammes UV 254 nm (tracé bleu) et MS-ESI^+ de l'extrait de *Primula veris* à 1000 ppm 131

Figure 50 Chromatogrammes UV 254 nm (tracé vert) et MS-ESI^- de l'extrait de *Primula veris* à 1000 ppm 131

Figure 51 Structure de a) l'acide glycyrrhizique et b) la glabridine 143

Figure 52 Chromatogrammes UV 250 nm (tracé bleu) et MS en mode positif de l'extrait liquide de réglisse analysé à 0,1 % de matière sèche. Indications des pics à identifier (**A** à **F**) 145

Figure 53 Chromatogrammes UV 250 nm (tracé vert) et MS en mode négatif de l'extrait liquide de réglisse analysé à 0,1 % de matière sèche. Indications des pics à identifier (**A** à **F**) 145

Figure 54 Structure de l'échinatine (= (E)-4,4'-dihydroxy-2-methoxy-chalcone) et voies de fragmentation proposées pour les dérivés de type chalcone d'après [121] 148

Figure 55 Spectre MS en mode positif du pic **D** (53P) 149

Figure 56 Spectre MS/MS en mode positif du pic **D** (53P) : fragmentation de l'ion $[M+H]^+$ 149

Table des illustrations.

Figure 57 Structure de l'ononin (décrite pour *Glycyrrhiza glabra*) 150

Figure 58 Structure de l'isoononin (décrite dans le genre *Glycyrrhiza*) 150

Figure 59 Structure de la formononétine (= 7-hydroxy-4'-methoxyisoflavone) 150

Figure 60 Structure de l'isoformononétine (= 4'-hydroxy-7-methoxyisoflavone) 150

Figure 61 Spectre MS/MS en mode négatif du pic F (49N) : fragmentation de l'ion $[M-H]^-$ 151

Figure 62 Structure de afromosin 7-O-glucoside 151

Figure 63 Structure de afromosin 7-O-galactoside 151

Figure 64 Structure de 4',7,8-trihydroxyisoflavone; 4',8-di-me ether, 7-o-beta-d-glucopyranoside 152

Figure 65 Structure de 3',4',7-trihydroxyisoflavone; 3',4'-di-me ether, 7-o-beta-d-glucopyranoside 152

Figure 66 Structure de 4',5,7-trihydroxyflavone; 5,7-di-me ether, 4'-o-beta-d-galactopyranoside ... 152

Figure 67 Structure des flavones et des isoflavones d'après [127] : enchaînement en C6-C3-C6. La liaison entre le cycle C et le cycle B différencient ces deux classes de flavonoïdes..... 152

Figure 68 Fragmentation d'une di-hydroxyisoflavone selon une réaction rétro-Diels-Alder ; cas de la Daïdzeine à partir de $[M+H]^+$ 160

Figure 69 Spectre MS/MS du pic 48P à partir de l'ion $[M+H]^+$: observation des ions issus d'un réaction rétro-Diels-Alder 161

Figure 70 Salicaire a) en milieu naturel –Saint Jean le Blanc dans le Loiret en juillet 2014- et détails d'organes de la plante : b) inflorescence fleurie et c) tige avec feuilles 167

Figure 71 Structures de molécules décrites dans la littérature pour la salicaire 170

Figure 72 Schéma descriptif de l'organisation de l'épiderme de la peau humaine : les couches cellulaires sont représentées à gauche (traduction du bas vers le haut : lame basale, couche basale, couche spineuse, couche granulaire, couche granulaire supérieure et couche cornée). Au centre, sont indiquées les protéines exprimées dans chaque couche cellulaire : TG, transglutaminase ; SPRs, Small Proline-Rich proteins. A droite, la composition moléculaire de la couche cornée est détaillée (traduction : involucrine, loricrine, fillagrine, kératines, SPRs : petites protéines riches en proline, et lipides). Illustration tirée de ''The cornified envelope: a model of cell death in the skin'', Candi et al., 2005, Nature Reviews Molecular Cell Biology, 6, 328-340. 172

Figure 73 Analyses en immunofluorescence de l'expression de TGK dans les kératinocytes épidermiques à l'état basal ou traités par des extraits de *Lythrum salicaria*. Après 72 heures de traitement, les cellules sont fixées et la protéine d'intérêt est révélée en utilisant un anticorps dirigé contre elle. Ces résultats sont issus d'une expérience représentative de trois. 174

Figure 74 Analyses en immunofluorescence de l'expression de TGK dans les kératinocytes épidermiques à l'état basal ou en présence de CaCl2 à 1,5mM ou traité par l'extrait EtOH 30% de *Lythrum salicaria* dans une gamme de concentration croissante. Après 72 heures de traitement, les cellules sont fixées et la protéine d'intérêt est révélée en utilisant un anticorps dirigé contre elle. Ces résultats sont issus d'une expérience représentative de trois. 174

Figure 75 Chromatogramme de l'extrait de salicaire, colonne POLAR-RP (150x4.6 mm, 4µm) – Détection UV 254 nm- 176

Table des illustrations.

Figure 76 Structures de la vescaline (R_1 = OH, R_2 = H) et de la castaline (R_1 = H, R_2 = OH) 178

Figure 77 Structures de la vescalagine et de la castalagine 181

Figure 78 Chromatogramme de l'extrait de *Lythrum salicaria* L. obtenu avec le gradient utilisé pour le dosage de la vescalagine et la castalagine - UV 235 nm - 185

Figure 79 Effets de l'extrait de *Lythrum salicaria* sur l'expression d'ARN messagers des marqueurs de différenciation de kératinocytes. Les NHEK sont traités pendant 48 heures avec $CaCl_2$ à 1,5mM, l'extrait de *Lythrum salicaria* (extrait de Lyth. S.) à 20µg.mL^{-1} et les niveaux de ARNm sont évalués par RT-qPCR : a) TGK, b) KRT1, c) KRT10, d) Filaggrine, e) Loricrine, f) Involucrine. Ces résultats sont issus d'une expérience représentative de deux. 187

Figure 80 Analyses en immunofluorescence de l'expression de a) TGK, b) KRT10 and c) Filaggrine dans les kératinocytes épidermiques traités ou non avec $CaCl_2$ à 1,5mM ou l'extrait de Lythrum salicaria à 20 µg.mL^{-1}. Après 72 heures de traitement, les cellules sont fixées et les protéines d'intérêts sont révélées en utilisant un anticorps dirigé contre les protéines d'intérêts. Ces résultats sont issus d'une expérience représentative de trois. 189

Figure 81 Expression de TGK dans les monocouches de NHEK suite à l'incubation des cellules avec le témoin positif ($CaCl_2$), l'extrait de *Lythrum salicaria*, l'acide gallique (A), la vescalagine (B), la castalagine (C) ou la salicarinine (D). (Significativités statistiques; ns : non significatif ($p > 0.05$), * : significatif ($0.01 < p < 0.05$), ** : très significatif ($0.001 < p < 0.01$), *** : extrêmement significatif ($p < 0.001$)..... 190

Figure 82 Différenciation des RHE : 1) sans vitamine C, 2) avec vitamine C (contrôle positif) et 3) en présence de l'extrait de *Lythrum salicaria*. Après 7 jours de traitement, l'expression des marqueurs de différenciation de l'épiderme : a) KRT10, b) filaggrine et c) TGK est évaluée par immunohistochimie. Ces résultats sont issus d'une expérience représentative de trois..... 191

Figure 83 Evaluation de l'impact de l'extrait *de Lythrum salicaria* et de la vescalagine sur l'architecture de peaux reconstruites. Les coupes transversales histologiques des peaux sont obtenues après 9 jours d'incubation de la peau : a) en milieu complet, b) en milieu carencé et c) en présence de l'extrait de *Lythrum salicaria* ou d) de vescalagine. 193

Figure 84 Effet de l'extrait de *Lythrum salicaria* et de la vescalagine sur l'expression de MMP-1 sur FTSK (Significativités statistiques; ns : non significatif ($p > 0.05$), * : significatif ($0.01 < p < 0.05$), ** : très significatif ($0.001 < p < 0.01$), *** : extrêmement significatif ($p < 0.001$)). 194

Figure 85 Effet de l'extrait de *Lythrum salicaria* et de la vescalagine sur l'expression de MMP-3 sur FTSK (Significativités statistiques ; voir détails à la Figure 84) 195

Figure 86 Effet de l'extrait de *Lythrum salicaria* et de la vescalagine sur l'expression du procollagène I sur FTSK (Significativités statistiques ; voir détails à la Figure 84) 196

Table des tableaux

Tableau 1 Principaux détecteurs utilisés en chimie analytique et pouvant être couplés à l'HPLC..... 23

Tableau 2 Résumé des niveaux de confiance en identification de métabolites définis par Sumner *et al*. en 2007..... 25

Tableau 3 Base de données de produits naturels ou généralistes utiles à la recherche sur les produits naturels..... 36

Tableau 4 Outils d'annotation des données de MS 44

Tableau 5 Référencement non exhaustif d'outils d'intérêts pour l'étude et l'annotation de données de MS/MS. Classement selon principe utilisé ; Comparaison à une base spectrale (A), comparaison à une base structurale avec utilisation d'outils computationnels de fragmentation *rule-based* (B) ou d'outils de fragmentation combinatoire (C), génération et comparaison d'arbres de fragmentation ou d'empreintes moléculaires (D), réseaux moléculaires (E)..... 46

Tableau 6 : Types d'ions pouvant être sélectionnés pour permettre l'annotation des ions des spectres traités sous GAINS..... 59

Tableau 7 Oses considérés (type, formule brute et masse monoisotopique) pour l'exploitation des données de fragmentation..... 66

Tableau 8 Groupements acylés pris en compte pour la recherche de leur présence au sein de structures de composés naturels 67

Tableau 9 Adduits, clusters ou pertes de neutres considérés pour les calculs des masses monoisotopiques..... 73

Tableau 10 : Poids attribué à chaque type d'ions pour le calcul du score de classement des masses monoisotopiques de GAINS..... 74

Tableau 11 Récapitulatif des résultats obtenus pour valider GAINS sur le traitement des données MS : comparatif sur les données collectées en FIA lors de deux essais indépendants..... 82

Tableau 12 Exemple pour illustrer la limite de la normalisation du score au sein de Metfrag 87

Tableau 13 Données analytiques avec éléments structuraux déterminés et propositions d'identification des trois pics principaux de l'extrait de tulipe noire, et des pics qui ont des candidats attribués par déréplication ciblée dans Tulipa gesneriana. 105

Tableau 14 Résumé de données MS et MS/MS observées pour le composé apparenté aux pics 72P/62N..... 106

Tableau 15 Résultats de recherche de candidats structuraux pour identifier les chémo-marqueurs de l'extrait de tulipe noire. * : éléments structuraux déterminés par l'étude (nature de la génine et des sucres). ** : Metfrag fragmente des candidats de la base obtenus après requête via le critère de masse monoisotopique, on détaille le nombre de candidat ayant obtenu un score de 1 et parmi eux, ceux correspondant aux éléments structurant déjà déterminés) 109

Tableau 16 Dosage en HPLC-UV de flavonols avant hydrolyse et après hydrolyse de l'extrait EtOH 60 % de *Primula veris*. 114

Tableau 17 Prétraitement par SPE-C_{18} : Etapes, rôle de l'étape, solvant et volume de solvant utilisé. 115

Tableau 18 Estimation de la teneur approximative dans l'extrait de chaque pic purifié en considérant l'aire en UV du composé et la teneur en flavonoïdes glycosylés dans l'extrait de primevère. 121

Tableau 19 Récapitulatif des informations obtenues pour les 7 molécules isolées à partir de l'extrait EtOH 60 % de Primevère vraie (HA : Hydrolyse Acide). [α]Données complétées par analyse FIA-HRMS/MS pour tenter d'observer l'enchaînement de la partie glycosylée sous forme sodée. 123

Tableau 20 Résultats de propositions d'identification issus de GAINS : déréplication ciblée parmi les molécules connues de l'organisme *Primula veris* et du genre *Primula* 134

Tableau 21 Résultats de GAINS pour des pics d'intensités non négligeables du profil avec des candidats non décrits parmi ceux du genre *Primula* 134

Tableau 22 Résultats obtenus avec Metfrag pour les pics dont la masse monoisotopique correspond à celle de composés purifiés à partir de l'extrait. 136

Tableau 23 Comparaison des deux méthodes utilisées pour décrire l'extrait : mise en parallèle des résultats obtenus. 139

Tableau 24 Données analytiques des principaux pics chromatographiques (A à F) de l'extrait glycériné de réglisse glabre (nd : non déterminé, sh : *shoulder* épaulement) 147

Tableau 25 Données LC-HRMS sur l'extrait glycériné de réglisse glabre : masses monoisotopiques et recherche de la famille des isoflavones 159

Tableau 26 Données LC-ESI-MS collectées par analyse de l'extrait de *Lythrum salicaria* en mode positif et en mode négatif d'ionisation 177

Tableau 27 Isolement des composés majoritaires de l'extrait de *Lythrum salicaria* : quantité obtenues, pureté et rendement de purification de chaque composé à partir de l'extrait. 180

Tableau 28 : Résultats HRMS pour le pic D 180

Tableau 29 Données HRMS et MS/MS pour les composés B et C obtenues par analyses en ESI-Q-TOF (HHDP : unité hexahydroxydiphenoyl = 302.0057 g.mol-1). Pour chaque ion : les intensités relatives apparaissent entre parenthèses, l'attribution entre crochets. Les ions isolés pour la MS/MS ont un m/z en gras. Les ions de MS/MS sans attribution (entre crochets) étaient tous monochargés. 183

Tableau 30 Données RMN ^{1}H et ^{13}C-NMR (400 et 100 MHz) des composés B et C ; déplacements chimiques (δ) exprimés en ppm et constantes de couplage (J) en Hz. 184

Tableau 31 Tableau de correspondance entre les modèles cellulaires utilisés pour l'étude et les types d'analyses qu'ils rendent possibles et qui ont été effectués dans l'étude qui va être présentée. 186

Introduction générale

A l'image de notre société actuelle, la nature use de stratégies de communication et de réseaux complexes pour fonctionner[*]. Dans la nature, le règne végétal et le règne animal sont constitués d'organismes vivants qui cohabitent et interagissent, de manière conflictuelle (prédation, défense) ou coopérative (nourriture, pollinisation, dispersion)[†]. Ces interactions ont lieu par le biais de langages subtils, les plantes silencieuses agissent grâce à la biochimie en libérant des signaux odorants, des goûts ou en prenant des aspects spécifiques. Les plantes sont autotrophes puisqu'à partir de l'énergie du soleil et du peu de ressources qu'elles trouvent partout (H_2O, CO_2 et minéraux), elles arrivent à produire tout ce dont elles ont besoin. Véritables virtuoses de la biochimie, les plantes ont su développer un large panel de voies biochimiques par lesquelles elles synthétisent les métabolites secondaires en plus des métabolites primaires (sucres, protéines et lipides). Le métabolisme secondaire est spécifique et répond à la contrainte majeure que les plantes connaissent : leur fixité. Ainsi, n'ayant pas la possibilité de fuir ou de se défendre lors d'une prédation par un animal, de se mettre à l'abri lorsque les conditions météorologiques leur sont peu favorables -au cours du cycle des saisons, les plantes ne peuvent pas migrer- ou encore de partir à la recherche de nourriture, les plantes investissent l'énergie qu'elles ont à leur disposition dans la synthèse de composés sophistiqués et cela leur permet de compenser leur immobilité.

Si la nature peut nous apparaître complexe et secrète, nous savons néanmoins l'apprivoiser et l'explorer depuis des millénaires. Il y a longtemps, l'homme n'avait que les ressources naturelles pour se nourrir, se guérir, s'habiller et même pour se loger. Les explorateurs et botanistes ont aussi exploré des contrées inconnues à la recherche de trésors naturels. C'est ainsi que nos jardins sont peuplés de plantes étrangères introduites au cours du temps principalement pour leur beauté. Aujourd'hui, bien qu'elle ne soit plus l'unique source de médicaments ou de soins, la nature nous inspire toujours du fait de sa grande diversité chimique que nous tentons d'imiter notamment par biomimétisme.

Dans le cadre de l'élaboration de produits cosmétiques à partir des substances naturelles, il est aujourd'hui fréquent que les chercheurs s'appuient sur les savoirs traditionnels accumulés par les peuples au cours du temps. Néanmoins, cette démarche ethnobotanique peut être complexe du fait

[*] Les langages secrets de la nature, Jean-Marie Pelt, Fayard, 1996

[†] Eloge de la nature, Francis Hallé, Editions du Seuil, 1999. Les deux premiers paragraphes sont inspirés de cet ouvrage.

du manque de données écrites et standardisées s'expliquant par une transmission des savoirs ancestraux souvent orale et par la multiplicité de noms communs pour une même espèce du fait des différents dialectes par exemple. C'est pourquoi, les laboratoires travaillent aussi en revisitant les plantes de façon moderne et en s'appuyant sur des techniques actuelles. C'est un travail multidisciplinaire qui réunit des spécialités qui vont de la botanique à la biologie cellulaire en passant par la chimie analytique ou la formulation.

Parmi les différentes actions à mener pour qu'un extrait de plante devienne un actif cosmétique, l'emploi de la phytochimie est fondamental. Ceci doit permettre la connaissance des constituants de l'extrait de plante, matière première des actifs développés. Grâce à ces explorations phytochimiques, les chémomarqueurs d'un extrait peuvent être identifiés. De plus, les travaux de phytochimie associés à des tests d'activités biologiques sur l'extrait et/ou sur des composés purifiés conduisent à la détermination de biomarqueurs de l'extrait.

Un projet de thèse a vu le jour dans le cadre d'un contrat CIFRE entre l'entreprise Greenpharma basée à Orléans et l'Institut de Chimie Organique et Analytique de l'Université d'Orléans (ICOA, UMR CNRS 7311). Greenpharma est spécialisée dans la recherche et le développement d'ingrédients cosmétiques, pharmaceutiques voire agro-alimentaires à partir de produits naturels et plus particulièrement issus de végétaux. L'ICOA est un institut dont l'objectif est d'étudier et d'identifier des molécules bioactives pour les valoriser dans les domaines de la santé et de la cosmétique. Les thématiques de recherche du laboratoire sont la conception, la synthèse organique et l'analyse des composés d'intérêt biologique. L'équipe SAAB (Stratégies Analytiques, Affinités et Bioactifs) travaille à la caractérisation moléculaire du végétal, l'analyse de traces dans des matrices complexes, l'évaluation des activités de composés bioactifs ainsi qu'au choix raisonné de systèmes séparatifs.

Dans le contexte d'études phytochimiques, la thèse avait pour enjeu central d'améliorer les méthodes de travail en intégrant de nouvelles approches basées sur des moyens informatiques et automatisés de traitement des données analytiques. Si d'un point de vue technique, il est aujourd'hui relativement aisé d'obtenir de nombreuses informations précises et capitales en vue de caractériser les structures de composés, le travail d'interprétation des données demande d'y consacrer un temps conséquent et les analyses phytochimiques restent des travaux chronophages.

Le projet de recherche doctoral a principalement été axé autour des actions suivantes :

- développement et intégration de méthodes semi-automatisées pour optimiser les études phytochimiques ;
- études phytochimiques d'extraits de façon globale et approfondie pour obtenir des informations fiables et précises ;
- coordination entre la phytochimie et l'évaluation des activités biologiques d'extraits ou de composés purifiés.

Pour Greenpharma, ces travaux ont vocation à se concrétiser par le développement d'ingrédients cosmétiques. Pour Greenpharma et l'ICOA, la mise en place de stratégies d'études phytochimiques efficaces, les plus exhaustives possibles et intégrant des aspects d'automatisation est également un enjeu important afin de faire évoluer les pratiques tout en profitant des progrès technologiques actuels. A l'heure de la génération de données en masse et du traitement informatique omniprésent, il est important d'évaluer leur potentiel et d'intégrer les méthodes modernes qui peuvent aider, simplifier et améliorer notre façon de découvrir de nouvelles applications.

La présentation de ce projet de recherche industrielle est répartie en 4 chapitres. Pour commencer, le chapitre 1 annoncera le contexte global de la recherche et du développement appliqués aux produits naturels et tourné majoritairement vers la cosmétique. L'enjeu est de développer des ingrédients entrant dans la composition de produits cosmétiques. Ces derniers étant destinés à être utilisés et actifs sur la peau, une présentation générale de cet organe sera donnée. Ce chapitre permettra aussi de décrire ce qu'est l'analyse phytochimique et quelles sont les méthodes et outils à disposition des phytochimistes aujourd'hui.

Le chapitre 2 sera dédié à la description détaillée de GAINS, un outil d'aide à l'identification de structures naturelles qui a été mis en place au cours de cette thèse au sein de Greenpharma. Cet outil est la réponse apportée à la problématique de gestion et de traitements de données analytiques. Il fonctionne pour le traitement de données obtenues en chromatographie liquide couplée à la spectrométrie de masse haute résolution. Dans le chapitre 2, l'outil est présenté d'un point de vue théorique et méthodologique. De plus, les résultats de la validation de l'outil sur un mélange reconstitué de composés connus sont proposés.

C'est le chapitre 3 qui exposera les applications pratiques de l'outil GAINS. Cela sera possible par l'intermédiaire d'exemples concrets d'étude d'extraits de trois plantes : la tulipe noire *Queen of the Night*, la primevère vraie et la réglisse glabre. L'étude de la primevère constitue la preuve de concept de l'outil car un extrait de la plante a été étudié en parallèle de façon classique (isolement et

caractérisation des composés majoritaires) et d'une façon plus automatisée avec l'emploi de GAINS. Les deux autres plantes ont été étudiées dans le cadre de projets de partenaires industriels qui ont fait appel à Greenpharma.

Le chapitre 4 présentera l'étude tant phytochimique que biologique de la salicaire. Le point de départ de ce projet a été de revisiter la phytochimie et les activités de cette plante traditionnellement utilisée pour de toutes autres applications que celle des soins cosmétiques. Le travail effectué en phytochimie dans le cadre de cette plante a une dimension plus classique et n'a pas fait appel à GAINS. En partant d'un extrait de la plante, nous avons purifié et identifié ses chémomarqueurs. Ces derniers se sont révélés être des biomarqueurs suite à des tests d'activité biologique sur des monocouches de kératinocytes, des épidermes reconstruits et des peaux reconstruites.

Une conclusion générale résumera l'ensemble des travaux et les principaux résultats obtenus. Les perspectives envisagées en termes d'amélioration de GAINS et de finalisation d'études phytochimiques seront exposées.

Chapitre 1
Valorisation des substances naturelles :
De l'analyse phytochimique à l'usage cosmétique
-Données bibliographiques-

1. Introduction

Le travail de thèse présenté s'inscrit dans le cadre d'une collaboration entre un laboratoire privé de recherche et un laboratoire de recherche publique dont le but commun est l'innovation à partir de produits naturels. Pour cela, les deux entités se doivent notamment de développer de nouvelles méthodes d'analyses pertinentes et performantes. L'analyse chimique est au cœur des travaux de ces deux laboratoires.

Ainsi, ce premier chapitre va introduire les notions et thématiques essentielles pour comprendre le projet dans sa globalité mais aussi sur certains aspects plus pointus. Dans un premier temps, je vais décrire ce qu'est la recherche et le développement à visée cosmétique axés sur les produits naturels en donnant des définitions inhérentes à ce domaine. De plus, la peau sera présentée puisqu'elle constitue la cible finale des produits naturels étudiés qui sont destinés à être utilisé comme actif cosmétique. Une description de la structure générale de la peau et de ses différents rôles sera donnée. Dans un second temps, l'approche phytochimique utilisée dans le cadre de profilage métabolique sera décrite au travers de la présentation de techniques d'analyse fondamentales utilisées et des tendances actuelles qui guident la manière de gérer au mieux les études sur les produits naturels et de traiter les données qu'elles génèrent.

2. Recherche & Développement sur les produits naturels pour le domaine cosmétique

2.1 Recherche et Développement en cosmétique

D'un point de vue général, la recherche et développement (R&D) a pour but de rationaliser les études qui peuvent être menées dans un laboratoire. Elle permet à la démarche de conception de cosmétiques d'être une science à part entière appelée cosmétologie [1] et ce, par le biais de méthodes scientifiques basées sur des hypothèses, des expériences, des validations et des connaissances scientifiques ou technologiques. Créer et proposer des produits cosmétiques nécessite de prouver et d'assurer leur efficacité, leur qualité et leur innocuité. S'agissant de l'étude des ingrédients cosmétiques, cela passe par la connaissance approfondie des matières premières utilisées, la compréhension des mécanismes d'actions des substances sur les cellules de la peau, la vérification de leur innocuité et le suivi de leur stabilité dans le temps par exemple. Concrètement, le champ d'investigation de la R&D au sein d'une entreprise est large : elle va jusqu'à coordonner les aspects marketing et réglementaires avec les notions de viabilité et de faisabilité d'un projet.

2.2 Cosmétiques et ingrédients actifs

La réglementation française définit ce qui peut être considéré comme un produit cosmétique. Selon le code de santé publique (article L.531-1) [2], un produit cosmétique est ‘’toute substance ou tout mélange destiné à être mis en contact avec les parties superficielles du corps humain (épiderme, système pileux et capillaire, ongles, lèvres et organes génitaux externes) ou avec les dents et les muqueuses buccales en vue, exclusivement ou principalement, de les nettoyer, de les parfumer, d’en modifier l’aspect, de les protéger, de les maintenir en bon état ou de corriger les odeurs corporelles’’. Les cosmétiques sont donc des produits de soin destinés à embellir, maquiller ou parfumer par exemple et ils sont des produits destinés à la peau normale sur laquelle ils ont une action à durée limitée. Ils n’ont pas de propriétés curatives ou préventives à l’égard de maladies, et ne doivent pas induire d’effets secondaires [1]. Ils ne constituent en aucun cas des médicaments, qui sont régis par une autre réglementation.

Une formule cosmétique se compose d’un ensemble de constituants supports pour lui donner sa texture, son aspect, sa couleur ou son parfum. Il existe de nombreuses formes en fonction du produit final à obtenir (crèmes, émulsions, eaux micellaires ou florales, gels, parfums, savons, produits de maquillage). Aussi, deux grandes familles de produits cosmétiques existent : les cosmétiques d’embellissement et ceux de soin. Les produits d’embellissement ou de parure colorent la peau, les cils, les lèvres. Ils servent à maquiller certaines zones, unifier le teint par exemple, camoufler des imperfections ou cacher une maladie dermatologique. Les produits de soin, quant à eux, doivent être actifs. Leurs formules contiennent un ou des composés choisis pour leur activité. Ces actifs peuvent être d’origine naturelle : minérale ou végétale (issus d’organismes terrestres ou marins). Ce sont ces actifs végétaux qui constituent le cœur du travail du phytochimiste. L’actif cosmétique est la part de la formule cosmétique qui cible une action sur la peau et qui permet de revendiquer une propriété particulière du produit. Si l’ensemble des produits incorporés dans la formule peuvent participer de façon globale aux bienfaits du produit sur la peau, les ingrédients dits actifs sont ceux dont l’effet est clairement évalué et validé par des tests. Ils fournissent la preuve claire et objective des propriétés du produit. Ces derniers peuvent dans un premier temps être faits *in vitro* sur des cellules de peaux cultivées et *ex vivo* sur des biopsies de peaux puis validés *in vivo* lors d’essais cliniques (panel de volontaires).

2.3 Valorisation dans les domaines Cosmétique, Dermatologique et Nutraceutique

Anciennement, avant l’essor de la chimie, la nature était seule pourvoyeuse de remèdes et soins. Ainsi, de nombreux usages traditionnels des plantes existent et sont encore aujourd’hui

connus et utilisés. La chimie de synthèse a modifié notre rapport aux produits naturels, qui pouvaient paraître limités. Depuis quelques années, un retour à la nature s'est amorcé par rapport à ce qu'elle peut nous enseigner et apporter, que nous ne pouvons pas réinventer ou surpasser. En effet, la nature et les plantes en particulier peuvent être considérées comme une immense collection de molécules dont la grande diversité est reconnue [3]. La tendance à revisiter les plantes est grandissante en recherche sur les médicaments par exemple [4]. Les plantes sont donc des matières d'intérêt dont les positionnements cosmétiques sont encore à découvrir ou peut-être à rafraîchir par rapport à ce qu'ont fait de nombreuses civilisations avant nous. Il s'agit de rassembler des données diverses pour aboutir à de vraies solutions de bien-être. On peut s'appuyer sur les usages traditionnels et l'histoire de la plante, les capacités d'adaptation de celle-ci à un environnement particulier et revisiter ces informations à la lumière des savoirs et techniques actuels.

De nombreux laboratoires explorent les substances naturelles pour diverses applications : pharmaceutique, dermatologique, cosmétique mais aussi nutraceutique (compléments alimentaires) et agroalimentaire (arômes, colorants). Un grand intérêt est porté à ce qui est d'origine naturelle. La nature est pour beaucoup gage de qualité, de sécurité et signe d'authenticité. Elle renvoie également à la valeur de durabilité de plus en plus recherchée ou appréciée de nos jours. Les produits naturels sont également des ressources renouvelables -à condition de ne pas être surexploitées. C'est ainsi qu'on peut aussi valoriser des produits à la fois naturels et co-produits ou « déchets » d'une activité (éco-valorisation). L'intégration de produits naturels dans les produits cosmétiques est grandissante. On note qu'un très large panel d'entreprises y a recours avec des marques dites grands publics mais aussi de luxe.

Les actifs cosmétiques naturels sont capables d'avoir des effets multiples du fait de la diversité des propriétés des plantes et des molécules qu'elles contiennent. On trouve des actifs naturels qui agissent sur l'hydratation, le vieillissement cutané (anti-âge), l'inflammation impliquée lors de problèmes d'acné, de dermatite atopique, la pigmentation ou dépigmentation ou encore l'amincissement.

L'utilisation de produits naturels dans les domaines cosmétiques ou de bien-être est différente de celle faite au niveau pharmaceutique. En cosmétique, c'est un extrait qui est bien souvent valorisé, c'est-à-dire un mélange complexe de composés issus d'une matrice végétale. En pharmaceutique, il est plus fréquent que ce soit une seule molécule pure qui soit au cœur du médicament. Ceci s'explique par la divergence des contraintes et applications entre les deux domaines.

2.4 Aspects réglementaires ou toxicologiques à prendre en compte

Le développement en vue de la mise sur le marché d'un produit cosmétique est soumis à des règles afin d'assurer son innocuité, d'encadrer sa qualité, de prouver son efficacité ainsi que de vérifier qu'il ne contient pas de substances interdites par les lois en vigueur dans le pays où il sera commercialisé. En France, les produits cosmétiques sont régis par le code de santé publique et le règlement (CE) n° 1223/2009 du parlement européen et du conseil du 30 novembre 2009 relatif aux produits cosmétiques. Il est mis en application depuis juillet 2013.

La phytochimie peut aider à répondre à ces enjeux réglementaires et toxicologiques. Les actifs issus d'extraits de plantes sont étudiés afin que leur contenu soit connu, le but est de les décrire le plus finement et précisément possible. Cela permet aussi de s'assurer de l'absence de certaines molécules ou familles de substances naturelles controversées ou connues pour être toxiques. Caractériser l'extrait permet également de s'assurer de l'authenticité de la plante ou de l'extrait étudié.

3. La peau, organe ciblé par les produits cosmétiques

Dans un contexte d'innovation pour la cosmétique à partir de produits naturels, la peau constitue la cible principale des produits de soin. La peau est amenée à être en contact de façon répétitive avec les composants du produit. C'est sur la peau que les effets des produits naturels seront visibles mais c'est aussi elle qui permettra une plus ou moins bonne efficacité du produit final formulé puisque de nombreuses interactions vont y avoir lieu. Plus grand organe du corps humain, la peau doit assurer des fonctions essentielles. Une présentation rapide est proposée ci-dessous tandis que des aspects plus précis seront abordés au dernier chapitre présentant le développement et les activités d'un actif anti-âge.

3.1 Structuration globale

La peau est organisée en trois grands blocs que sont l'épiderme, le derme et l'hypoderme. Leur cohésion rend l'organe de la peau solide puisqu'elle forme une véritable barrière naturelle entre l'intérieur du corps et l'environnement extérieur, restant toutefois souple et perméable. La résistance mécanique de la peau est donc un équilibre fragile.

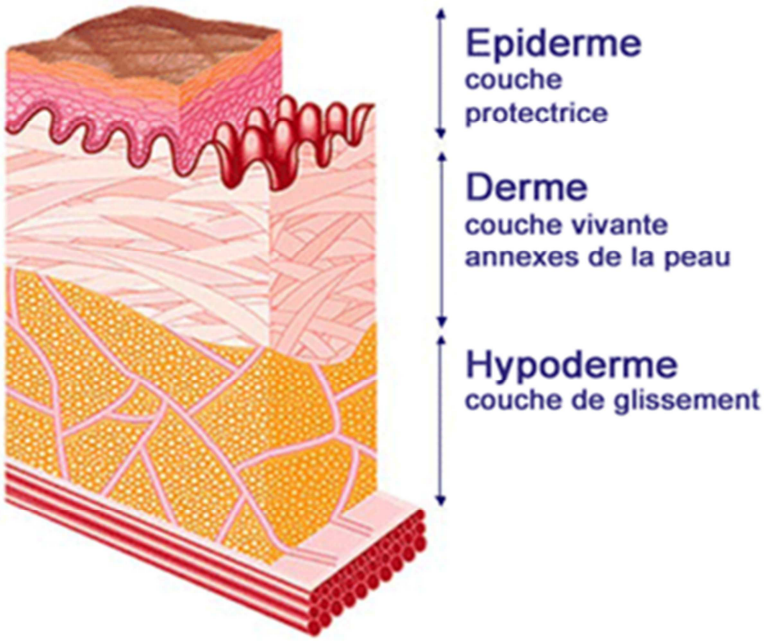


Figure 1 Schéma de coupe de peau humaine montrant les 3 grands blocs constitutifs d'après [5]

L'épiderme est la partie externe de la peau. Il est constitué majoritairement de kératinocytes et contient également les mélanocytes. Si de nombreuses terminaisons nerveuses se trouvent dans l'épiderme et nous permettent de ressentir la douleur, la température et la sensation du toucher en général, il n'y a en revanche aucun vaisseau sanguin dans cette zone de la peau. La couche cornée[*] (*stratum corneum)* est la couche supérieure de l'épiderme où les kératinocytes totalement différenciés sont devenus des cornéocytes (cornification). Très résistante, la couche cornée régule l'hydratation de la peau et forme une barrière protectrice.

Le derme, principalement formé de fibroblastes, est un tissu de soutien de l'épiderme. Il sert d'implantation aux annexes cutanées (glandes sébacées ou sudoripares, poils). Le derme assure l'apport nutritionnel en eau, vitamines et sels via les nombreux capillaires sanguins. Il produit aussi les fibres de la peau : collagène et matrice extracellulaire, ce qui en fait un acteur mécanique de la peau.

L'hypoderme est le socle des cellules adipeuses qui stockent et accumulent les graisses sous forme de triglycérides. C'est un réservoir énergétique et il a aussi un rôle de thermorégulation.

[*] Au chapitre 4, une description détaillée du *stratum corneum* complète les généralités du présent chapitre.

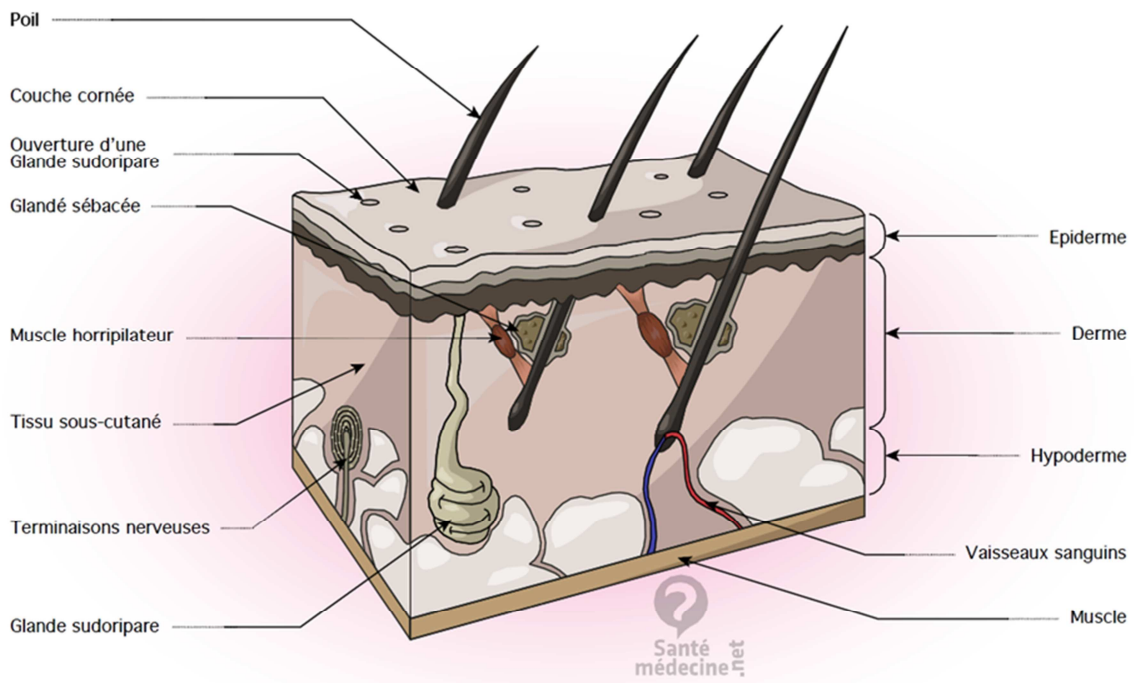


Figure 2 Schéma de coupe de peau : détail des éléments et annexes cutanées ancrées dans le derme, d'après [6].

3.2 Fonctions essentielles de la peau

Physiologiquement, la peau forme la frontière entre l'intérieur du corps et l'environnement : elle empêche les pertes de liquides (sang, eau). Définie comme une véritable barrière, c'est une protection contre les agressions extérieures diverses que l'on peut subir : blessures, attaques par les rayons UV du soleil, pollutions, air. Toutefois, la peau reste perméable. Elle respire, nous permet d'éliminer les toxines accumulées dans le corps et de réguler la température du corps grâce à la transpiration. La peau produit aussi les cheveux, poils et ongles, ce qui en fait le siège d'un ensemble d'éléments importants liés à notre apparence.

Socialement, la peau est au cœur de nos interactions avec les autres et l'environnement extérieur. En effet, c'est l'organe qui permet le sens du toucher grâce aux terminaisons nerveuses. Mais c'est aussi une enveloppe corporelle. C'est la partie du corps qui est visible. La peau constitue une part importante de notre apparence ; elle a une couleur, son état peut indiquer notre tranche d'âge mais aussi un état de santé particulier ou encore participer à notre beauté.

Elle peut aussi être l'expression d'un message ou d'attitude en fonction de notre façon de la montrer ou cacher, de la marquer ou maquiller ou de lui apporter des soins. C'est ainsi que chaque marque de cosmétiques va définir son identité, ses valeurs et ses produits pour qu'il y ait une correspondance entre l'utilisation de ces produits et ce qu'elle souhaite apporter à ses clients.

3.3 Intéraction avec les cosmétiques

Un autre grand pan de la R&D en cosmétique concerne les études des effets biologiques des actifs (ou formules) sur la peau. De plus en plus, l'activité et l'efficacité doivent être démontrées. Ces tests d'activités biologiques sur des cellules de peau ou des modèles de peaux font appel à des techniques d'analyse biologique, biochimique et des protocoles de plus en plus évolués[†].

3.3.1. Les différents types cellulaires ciblés par les cosmétiques

Globalement, les types cellulaires majoritaires de l'épiderme ou du derme peuvent être intéressants à cibler du fait de leur large représentation. C'est le cas des kératinocytes et cornéocytes au sein de la couche cornée, ou des fibroblastes. Mais on peut aussi travailler sur les mélanocytes par exemple. Le type cellulaire ciblé est forcément lié à la propriété recherchée. Il faut donc connaître le fonctionnement des éléments entre eux au sein de la peau : développement, réactions, signalisations.

3.3.2. Fonctions cosmétiques recherchées

Les soins cosmétiques (sans parler des maquillages ou produits nettoyants) sont aujourd'hui proposés pour répondre à des problématiques variées concernant l'aspect et le bon fonctionnement de la peau. Ils peuvent avoir les propriétés suivantes : effet hydratant, effet apaisant, effet anti-âge, effet antioxydant et anti-inflammatoire, effet pigmentant ou dépigmentant, effet amincissant, effet protecteur contre les UV, effet régulateur de la séborrhée [1]. Le premier effet d'un cosmétique est avant tout d'être hydratant puisqu'une peau déshydratée ne peut fonctionner correctement (perte de souplesse, des propriétés mécaniques et de défense). Pour un même effet ciblé, de nombreux actifs différents sont éligibles. L'ensemble des produits cosmétiques de soin garantissant un même effet global au final peuvent y parvenir selon différents modes d'action. Prenons l'exemple du vieillissement cutané. Cette altération de la peau dans le temps présente une grande variété de signes : rides et modification du relief de la peau, diminution de son épaisseur, diminution de la souplesse et de la fermeté de la peau, baisse de l'hydratation, pigmentations (tâches). Ainsi pour contrer ces aspects du vieillissement, il existe différentes cibles et différentes actions possibles adaptées aux problèmes rencontrés dans les peaux vieillies ou dites matures. C'est pourquoi de nombreux types de produits seront considérés comme soins anti-âge : antirides, tenseurs, lissants du relief, améliorateurs de l'éclat du teint, régénérants du tissu.

[†] Les tests d'activités biologiques et les méthodes d'analyses employées ne sont pas développés dans ce mémoire. C'est en effet une discipline parallèle à l'analyse phytochimique.

4. Analyse phytochimique

Les plantes constituent les matières premières à travailler pour développer des actifs naturels. La description du contenu de ces matières est donc essentielle pour savoir comment les valoriser et pour justifier leur efficacité.

4.1 Qu'est-ce que la phytochimie ?

4.1.1. Définition de la phytochimie

La phytochimie est littéralement la chimie des plantes. Dans le cadre de projets de recherche et développement, elle regroupe des aspects techniques d'analyse qui doivent permettre de décrire le contenu de plantes, et plus particulièrement d'extraits obtenus à partir de parties de ces plantes. Ce qui est extrait de la plante constitue une matrice naturelle complexe composée d'un grand nombre de composés. Ces derniers sont classés en deux grands types : les métabolites primaires et les métabolites secondaires. Les premiers regroupent les protéines, les lipides et les sucres. Ces composés gèrent les fonctions essentielles de la plante (croissance et développement, constitution des cellules, respiration ou stockage par exemple) et sont donc des composés ubiquitaires au sein du règne végétal comme au sein du règne animal. Par ailleurs, les plantes ont un second niveau de métabolisme qui a un caractère plus spécifique [7,8]. Les métabolites secondaires permettent aux plantes de s'adapter à leur environnement. En effet, dépourvues de mobilité, les plantes ont développés des moyens chimiques pour gérer cette contrainte afin d'interagir avec leur environnement : protection face à des prédateurs, gestion de stress extérieurs, communication avec d'autres plantes ou insectes [8,9].

4.1.2. Objet d'étude de la phytochimie : les métabolites

Au sens large, un métabolite est un composé chimique synthétisé par un organisme vivant (ici une plante). Toutefois, dans le contexte de notre travail de recherche sur les produits naturels issus de plantes, lorsque le terme de métabolite est employé, il a un sens plus restrictif puisque nous évoquerons principalement les métabolites impliqués dans le métabolisme secondaire. Ils peuvent aussi être désignés par l'ensemble des molécules de faible poids moléculaire ou ''petites molécules''. La gamme de masses de ces composés peut aller de 50 à 1500 $g.mol^{-1}$. Ces produits naturels présentent une très grande diversité que ce soit en termes de structures, de propriétés physico-chimiques (polarité par exemple) et de propriétés biologiques.

Les métabolites (primaires et secondaires) sont représentés par des classes de familles phytochimiques variées. La Figure 3 illustre la diversité des métabolites au sein des végétaux et les voies de biosynthèse desquelles ils sont issus. Les grandes familles constitutives des métabolites secondaires sont les alcaloïdes, les terpènes et stéroïdes et les composés phénoliques.

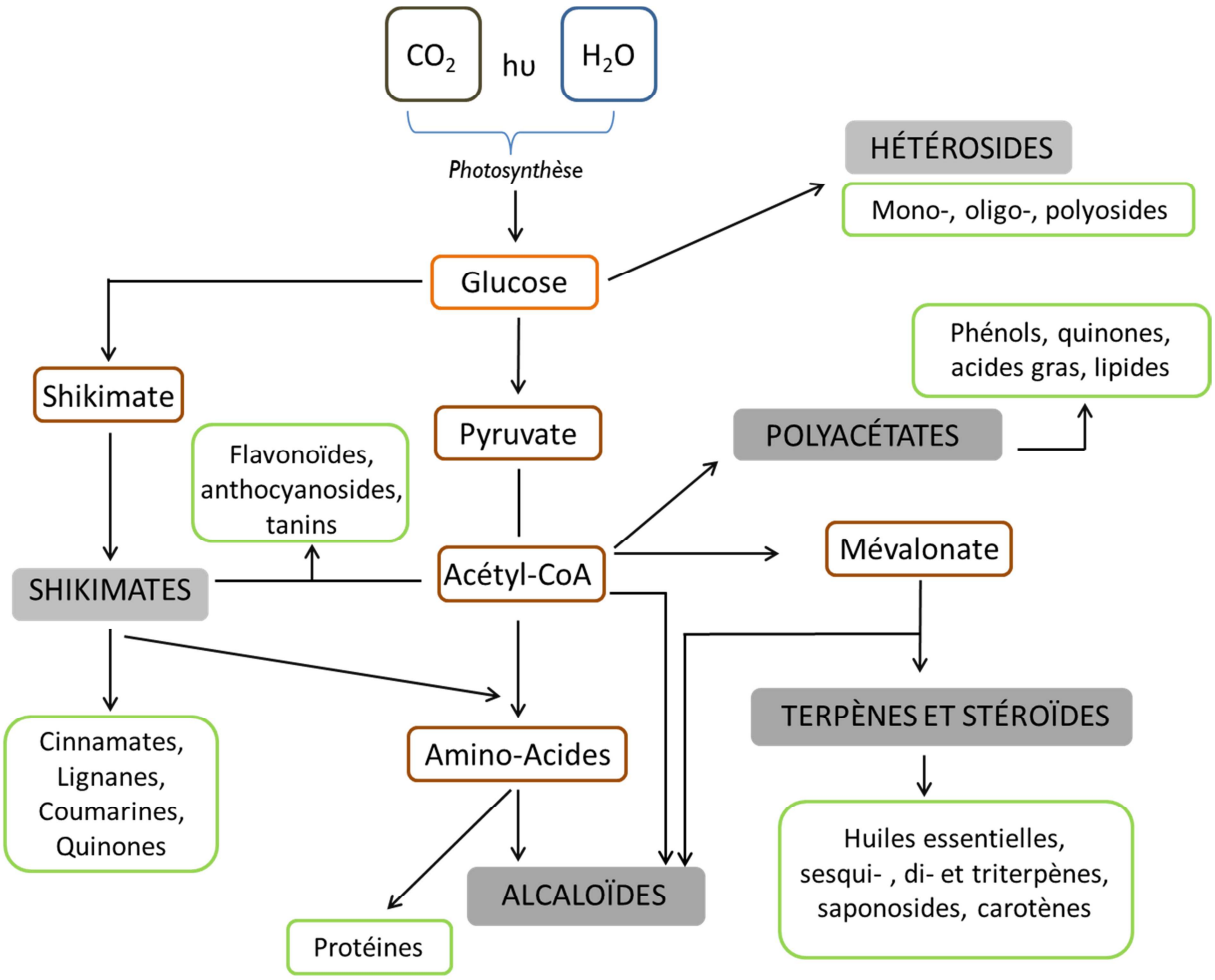


Figure 3 Représentation schématique[‡] des voies de biosynthèse rencontrées dans le règne végétal. Les encadrés gris indiquent les grandes classes phytochimiques qui regroupent diverses familles phytochimiques (encadré vert). A partir du glucose, différentes voies de biosynthèse se sont développées (encadré marron). Les liaisons existantes entre les voies, les classes et les familles sont représentées par les flèches.

Point de départ fondamental des réactions biochimiques, la photosynthèse est un processus de synthèse de composés à partir de l'énergie du Soleil qui combine les éléments naturels tels que l'eau et le dioxyde carbone. La photosynthèse est assurée par la chlorophylle, pigment vert spécifique des végétaux qui capte la lumière et absorbe les radiations bleues et rouges [8].

[‡] Représentation adaptée à partir d'un schéma proposé par Jean Bruneton dans Pharmacognosie, phytochimie, plantes médicinales à la page 4 (5ème édition, 2016).

Il existe certaines classes de composés qui peuvent être considérées comme des métabolites secondaires alors qu'elles sont apparentées au métabolisme primaire [7]. On peut évoquer des sucres de faible poids moléculaires que sont les oligosaccharides et plusieurs classes de composés issues d'acides aminés ayant un rôle de défense telles que les glucosinolates (hétérosides soufrés), des hétérosides cyanogènes ou les acides-aminés non constitutifs de protéines. Il a été recensé plus de 300 acides aminés chez les végétaux pour une vingtaine chez les animaux. Leurs fonctions ne sont pas clairement définies : il est supposé qu'ils participent au stockage de l'azote et soient toxiques envers des prédateurs.

4.1.3. Méthodes générales employées en phytochimie

Pour connaître le contenu d'un extrait naturel, il existe des méthodes classiques au sens où elles sont utilisées depuis longtemps et qu'elles s'appuient sur des techniques analytiques bien établies. Parmi elles, on peut citer deux niveaux de caractérisation. Le premier est un niveau global qui correspond à la mise en évidence de familles phytochimiques par l'observation de réactions ou l'obtention d'informations caractéristiques (maxima UV, spectres de masse par exemple). Pour cela, le phytochimiste peut s'appuyer sur des tests colorimétriques caractéristiques ou sur la chromatographie sur couche mince (CCM) suivie de révélations spécifiques. Il peut également analyser l'échantillon en chromatographie en phase gazeuse pour les composés volatils ou en chromatographie liquide haute performance (HPLC pour *High Performance Liquid Chromatography*) (détails au paragraphe 4.2.2). Cette dernière peut être couplée avec des détecteurs qui s'appuient sur différentes propriétés physico-chimiques des composés [10] comme le résume le tableau qui suit :

Détecteur	Propriété mise en jeu	Avantages principaux
Détecteur monochromatique Détecteur à barrettes de diodes (DAD)	Absorption lumineuse	Sélectivité
Détecteur évaporatif à diffusion de la lumière (DEDL)	Diffusion de la lumière	Quasi-universalité
Détecteur à aérosol chargé (CAD pour *Charged Aerosol Detector*)	Ionisation par décharge corona	Quasi-universalité
Spectromètre de masse	Ionisation	Quasi-universalité, grande sensibilité, très grande sélectivité
Résonance magnétique nucléaire	Résonance magnétique	Quasi-universalité, très grande sélectivité

Tableau 1 Principaux détecteurs utilisés en chimie analytique et pouvant être couplés à l'HPLC

Un second niveau plus précis de description du contenu correspond au fait d'isoler les composés de l'échantillon pour les identifier formellement. Cela passe par des étapes de traitement

de l'échantillon, de purification, d'analyses multiples et d'interprétation. Ce travail nécessite des temps plus longs que ceux concernant les analyses du premier niveau mettant en jeu des procédures plus globales, générales et standards.

De façon plus récente, des techniques alternatives de déréplication ou d'analyse métabolomique sont également utilisées. Ces méthodes modernes et innovantes en termes d'approches sont possibles notamment grâce à l'évolution des instruments d'analyses, des moyens informatiques et des capacités en termes de collecte et de stockage des données analytiques. Ce sont des approches qui vont s'affranchir d'étapes complexes telles que celles de purification pour décrire un extrait (la partie 4.3 de ce chapitre développe ces approches). On peut obtenir un grand nombre d'informations à partir de méthodes de déréplication et d'interrogation de bases de données par exemple. Sans être opposées aux techniques plus classiques, elles les complètent et les rendent plus pertinentes. Grâce à elles, il est possible de focaliser le travail d'isolement et de caractérisation sur des molécules particulières, présentant un intérêt particulier au regard des informations collectées plus rapidement. Il faut préciser que, même aujourd'hui avec les avancées technologiques, l'identification de composés naturels sans ambiguïté demeure un véritable défi et une tâche complexe [11].

Aussi, depuis 2007, des critères ont été proposés pour normaliser les niveaux d'identification qui sont obtenus en fonction des données collectées. Il s'agit des initiatives de standards métabolomiques (*Metabolomics Standards Initiative* ou MSI) [12] inspirées de la littérature scientifique en métabolomique. Les MSI ont été publiées par le groupe de travail en chimie analytique (*Chemical Analysis Working Group ou CAWG*) affilié à la *Metabolomics Society* [13]. Ces standards permettent une normalisation rigoureuse et valide de ce qui est proposé à l'issue d'une étude phytochimique et ont pour but d'établir un consensus pour la communauté scientifique grâce à des descripteurs analytiques appropriés et cohérents. En effet, le niveau de confiance dans l'identification d'un composé est fonction des données obtenues lors de son analyse. Quatre niveaux sont renseignés comme le résume le Tableau 2.

Niveau de confiance	Description	Données analytiques requises
Niveau 1	**Composés identifiés**	Au moins 2 données expérimentales orthogonales correspondant aux données obtenues avec un composé standard dans les mêmes conditions d'analyse
Niveau 2	**Composés annotés de façon supposée** (*putatively*)	Absence de référence standard, annotation basée sur propriétés physico-chimiques et/ou similarité de données expérimentales avec des données publiées
Niveau 3	**Classes de composés caractérisées de façon supposée**	Similarité de données expérimentales avec des données publiées caractéristiques d'une classe de composés
Niveau 4	**Composés inconnus**	Donnée expérimentale correctement détectée et reproductible (RMN, MS ou autre) sans similarité avec des composés connus

Tableau 2 Résumé des niveaux de confiance en identification de métabolites définis par Sumner *et al*. en 2007 [12].

Les niveaux 1 à 3 concernent des composés connus et non identifiés pour la première fois et s'appliquent à des travaux où les composés ne sont pas nécessairement isolés. Dans le cas de composés nouveaux c'est-à-dire à identifier pour la première fois, le niveau 4 correspond à des composés dont les données sont insuffisantes pour être correctement caractérisés. En effet, pour identifier formellement un nouveau composé encore jamais identifié par ailleurs, un ensemble conséquent d'éléments est à réunir. Classiquement, le composé est extrait, isolé, purifié avant les étapes d'analyse élémentaire, de mesure de la masse monoisotopique, d'observations des chemins de fragmentation, d'analyses en RMN (proton, carbone et deux dimensions) ou d'autres analyses spectrales en Ultraviolet ou infrarouge par exemple [12].

Si de nombreuses données analytiques peuvent être collectées avec des techniques de plus en plus pointues qui permettent l'accès à des informations très précises, il n'en reste pas moins vrai qu'au final, l'identification structurale d'un composé est confirmée par un chimiste qui doit mobiliser ses connaissances et s'appuyer sur son expertise pour valider l'ensemble [14].

4.2 Techniques analytiques majeures

Pour mener des études phytochimiques, plusieurs techniques analytiques peuvent être nécessaires et la combinaison d'un ensemble d'analyses est souvent utile. Aujourd'hui, les deux techniques majeures et largement utilisées en phytochimie pour l'identification de composés sont la spectrométrie de masse (MS pour *Mass Spectrometry*) et la résonance magnétique nucléaire (RMN) [15,16]. Elles permettent la détection des composés et peuvent être couplées à des techniques séparatives (LC-MS, GC-MS, LC-RMN) ou utilisées sans séparation (Infusion directe et FIA –*Flow*

Injection Analysis- en MS ou DIMS -pour *Direct-Infusion Mass Spectrometry*-, profilage RMN) [16]. Aussi, pour la MS comme pour la RMN, de nombreuses modalités existent s'agissant tant du type d'instrument que des modes d'acquisition. Le but n'est pas de toutes les présenter ici mais d'énoncer en quoi la MS et la RMN sont informatives grâce aux principes qui les régissent et comment elles aident le phytochimiste à mener à bien ses recherches.

4.2.1. Spectrométrie de masse et résonance magnétique nucléaire : définitions, performances et intérêts

La spectrométrie de masse (MS pour *Mass Spectrometry*) et la résonance magnétique nucléaire (RMN) sont deux techniques incontournables pour étudier les produits naturels. Elles donnent en effet accès à des informations structurales sur les molécules. Largement complémentaires, la spectrométrie de masse renseigne sur la composition élémentaire (formule brute, masse de la molécule) et sur des éléments structuraux *via* la fragmentation tandis que la RMN permet de retrouver l'organisation spatiale des atomes.

En spectrométrie de masse, des ions fournissent des informations sur la nature et la structure de la molécule qui les a produits [17]. Les composés sont détectés sous formes ionisées et l'appareil mesure un rapport *m/z* qui correspond au nombre de masse de l'ion détecté sur le nombre de charges élémentaires portées par cet ion. Ce rapport de nombres est sans unité. Toutefois, lorsque l'on parle du rapport masse sur charge, il peut être exprimé en Thomson (Th) du nom du Prix Nobel de Physique de 1906 qui a découvert l'électron et déterminé son rapport *m/z* [17].

En chimie analytique, il existe une terminologie correspondant à différentes considérations de la masse des molécules : on peut évoquer la masse moléculaire (moyenne) et la masse monoisotopique (aussi appelée masse exacte) [17,18]. La masse moléculaire est calculée avec la masse atomique de chaque élément qui prend en compte l'abondance relative de chaque isotope de l'élément. La masse monoisotopique considère uniquement la masse exacte de l'isotope le plus abondant de chaque élément. C'est cette valeur qui est mesurée par les spectromètres de masse à haute résolution. La précision en masse (ou erreur de mesure de la masse) peut être exprimée de façon absolue ou relative :

Erreur absolue en $g.mol^{-1}$ = Masse expérimentale – Masse théorique

Erreur relative en ppm = [(Masse expérimentale – Masse théorique)] / Masse théorique) x 10^6

Un spectromètre de masse comporte trois éléments essentiels : (i) **la source** permet l'ionisation des molécules de l'échantillon analysé, (ii) **l'analyseur** sépare les ions en fonction de leur rapport *m/z*, (iii) **le détecteur** enregistre les signaux correspondants aux ions et les envoie aux

systèmes de traitement des données. Pour chacun des trois modules, différents types ou géométries ont été développés au cours du temps pour répondre à des défis analytiques variés. Pour les sources, un des paramètres importants est l'énergie qu'elles transfèrent au cours de l'ionisation. Certaines sont très énergétiques et ceci implique la fragmentation des ions directement dans la source tandis que d'autres sources occasionnent peu de fragmentations grâce à un processus moins énergétique et plus doux. Concernant les analyseurs, un premier type est celui des analyseurs à balayage, ils détectent les ions de façon successive au cours du temps tel le quadripôle. Le second type concerne les analyseurs qui sont capables d'analyser tous les ions simultanément, comme l'analyseur à temps de vol ou la trappe d'ions. Les analyseurs de masses sont définis selon cinq caractéristiques : plage de masses explorables, vitesse d'analyse, transmission, précision en masse et résolution [17].

Utiliser la spectrométrie de masse présente de multiples intérêts. Globalement, la méthode est : très sensible, versatile en termes de diversité moléculaire détectable et de domaine de masses mesurées, très précise et permet l'accès à de hautes résolutions [17].

Lorsque deux analyseurs au minimum sont en série, on peut réaliser des analyses de spectrométrie de masse en tandem (MS/MS). L'analyseur situé entre la source et l'analyseur qui permettra l'analyse des ions jouera le rôle de cellule de collision. Dans cette cellule, les ions sont fragmentés et cette décomposition de la structure donne accès à des informations plus précises à son sujet. Par ailleurs, la spectrométrie de masse ne rend pas toujours possible la détermination de façon univoque de la position des groupements constitutifs de la molécule. C'est le cas notamment pour les isomères de position qui présentent ses schémas de fragmentation identiques.

La résonance magnétique nucléaire (RMN) est indispensable dans les procédures d'élucidation de structures [11] : elle contribue largement à l'étude des métabolismes primaires et secondaires des plantes [19]. C'est une technique spectroscopique d'une grande puissance d'investigation qui donne accès à des renseignements structuraux importants. En effet, la résonance magnétique nucléaire est une propriété de la matière soumise à un champ magnétique. On étudie les interactions qui existent entre les atomes de la matière. Ainsi, la RMN rend possible la description de l'arrangement des atomes des molécules entre eux. Elle est considérée comme la technique analytique la plus pertinente pour identifier sans ambiguïté des composés [14,20]. Elle s'applique à des composés organiques ou inorganiques. Elle permet leur analyse à l'état solide ou en solution. Des quantités plus ou moins importantes de produit purifié sont nécessaires pour mener à bien les analyses en RMN [14]. De plus, les instruments évoluent avec l'utilisation de champs magnétiques plus élevés et de cryo-sondes, ce qui permet d'accroître la sensibilité de la RMN tout en n'atteignant pas encore celle de la MS [20]. Contrairement à la MS, la RMN offre les avantages d'être

reproductible et de donner accès à des informations structurales non ambiguës [19]. D'autant plus que la MS nécessite de composés avec des fonctions ionisables pour la détection alors que la RMN est non spécifique (universelle) et qu'elle détecte n'importe quel composé contenant des hydrogènes ou carbones.

Pour obtenir les informations structurales, il faut interpréter les spectres de RMN qui représentent les signaux de résonance. Ces derniers sont caractérisés par une valeur de déplacement chimique, une intensité et une forme. Ils constituent les éléments essentiels pour comprendre l'environnement dans lequel se trouve chaque atome d'une molécule. A partir de ces informations, on peut remonter jusqu'à la structure moléculaire. Les noyaux détectés en RMN sont ceux dont le nombre de spin est non nul. C'est le cas des éléments dont le nombre Z de protons ou le nombre A de nucléons (ou les deux à la fois) sont impairs. Ainsi, concernant les éléments organiques, on peut étudier : ${}^{1}_{1}\mathrm{H}$, ${}^{2}_{1}\mathrm{H}$ (deutérium), ${}^{13}_{6}\mathrm{C}$, ${}^{14}_{7}\mathrm{N}$ par exemple tandis que les noyaux ${}^{12}_{6}\mathrm{C}$ ou ${}^{16}_{8}\mathrm{O}$ par exemple ne peuvent pas être détectés. Cette règle conditionne aussi la sensibilité de détection pour chaque élément qui est liée à l'abondance relative de l'isotope.

Les analyses RMN peuvent être monodimensionnelles ou bidimensionnelles (RMN 2D). Les premières sont par exemple les analyses du proton (RMN ^{1}H) ou du carbone (RMN ^{13}C). Le spectre ^{13}C est généralement accompagné de spectres de séquences DEPT (*Distorsionless Enhancement of Polarization Transfer*) pour différencier les signaux de CH, CH_2, CH_3 et les carbones quaternaires. Les analyses RMN homonucléaires (proton-proton, carbone-carbone) ou hétéronucléaires (proton-carbone) sont utiles pour corréler les signaux des noyaux couplés entre eux.

Les deux techniques (RMN et MS) sont de plus en plus intégrées dans des stratégies globales et automatisées de description moléculaire (voir partie 4.3.3).

4.2.2. Focus sur la technique UHPLC-ESI-HRMS

Nous allons détailler les techniques analytiques au cœur de ce travail de thèse axé autour du développement d'un outil automatisé d'identification de métabolites secondaires par MS (voir chapitre suivant). Ainsi, le système d'ultra-haute performance en chromatographie liquide (UHPLC) couplée à la spectrométrie de masse de haute résolution (HRMS pour *High Resolution Mass Spectrometry*) équipée d'une source électrospray (ESI) et d'une détection par un quadripôle à temps de vol (Q-TOF pour *Quadrupole-Time Of Flight*) sera présenté. Largement plébiscitées à l'heure actuelle et disponibles au sein du laboratoire, ces techniques utilisées en couplage présentent de nombreux intérêts pour l'étude des ''petites'' molécules organiques et naturelles.

Cette combinaison de techniques permet de nombreuses applications avec par exemple la conduite d'analyses tant ciblées que non ciblées. Un très large emploi de cette technique concerne le profilage métabolique [14,15,21–23]. L'UHPLC-ESI-HRMS donne accès à une grande sensibilité, à une grande précision et à une large gamme de composés en termes de structures. D'après Theodoridis *et al*. dans une revue sur le profilage global de métabolites basé sur la LC-MS [24], la RP-LC-ESI-MS (RP pour *Reverse-Phase*) est considérée comme une des meilleures méthodologies disponibles bien qu'elle ne permette pas à elle seule de satisfaire l'objectif d'une description phytochimique exhaustive. D'un point de vue pratique, un compromis est souvent fait car il est lourd d'analyser un même échantillon en réplicats dans différents modes de séparation et d'ionisation. On ajoute qu'il est également complexe d'avoir à interpréter d'autant plus de résultats.

Pour commencer, afin de simplifier l'étude de matrices naturelles complexes, il est recommandé d'utiliser, en amont de la détection des composés, une technique de séparation avec une importante résolution chromatographique [16]. L'HPLC est reconnue comme un moyen versatile de séparation de produits naturels à partir de mélanges complexes sans nécessiter d'importantes préparations d'échantillon. La séparation d'un large panel de composés est possible car l'HPLC offre des mécanismes de séparation variés incluant la phase normale (silice), la phase inverse (C_8, C_{18}), la phase inverse non aqueuse dite NARP (*Non-Aqueous Reverse Phase*), la chromatographie d'interactions hydrophiles (HILIC) ou encore la chromatographie d'échange d'ions ou de paires d'ions. La phase inverse (RP), très employée, offre une très grande variété de phases stationnaires de différents types (C_8, C_{12}, C_{18}, phenyl, flurophenyl par exemple) qui présentent des spécificités, les rendant adaptées à différents types de composés [25].

Une avancée majeure en chromatographie liquide est l'apparition de l'UHPLC. Au niveau de la phase stationnaire des colonnes, l'emploi de particules sphériques de moins de 2 µm de diamètre a permis à la technique d'atteindre de meilleures performances comparées à l'HPLC. En effet, plus les colonnes sont remplies de particules de faibles diamètres, plus elles sont efficaces et génèrent des pics chromatographiques fins. Il est ainsi possible d'obtenir des temps d'analyses plus courts par rapport à l'HPLC tout en gagnant en résolution [16,26], néanmoins, au prix d'une augmentation importante de pression dans le système, nécessitant l'utilisation d'un matériel spécifique.

La chromatographie liquide peut directement être couplée aux détecteurs spectroscopiques (UV, visible et fluorimètre) et spectrométriques (MS) [16,27], ainsi les analyses en UHPLC-MS sont faciles à mettre en place.

La source électrospray (ESI pour *ElectroSpray Ionisation*) s'est imposée comme la technique d'ionisation la plus utilisée depuis la fin des années 1990 [17]. Décrite et développée par Fenn –qui a obtenu le prix Nobel de Chimie en 2002 pour cette innovation rapidement appliquée à l'analyse des biomolécules-, la source ESI fait partie des sources d'ionisation à pression atmosphérique. L'ionisation est dite douce et il en résulte la faible fragmentation des ions et donc l'observation d'espèces quasi-moléculaires. On observe des ions avec un nombre pair d'électrons [28]. De plus, l'efficacité d'ionisation à pression atmosphérique est meilleure qu'à pression réduite. On notera également que la simplicité de couplage direct entre le système de séparation des composés avec la source et la facilité d'introduction de l'échantillon rendent les sources d'ionisation à pression atmosphérique attrayantes [17].

La source ESI est appropriée pour une variété importante de molécules de polarité moyenne à élevée, jusqu'à des composés ioniques. C'est pourquoi elle est considérée comme une des sources les plus versatiles [22,23]. Les problèmes d'ionisation peuvent intervenir pour les composés peu ou non polaires (il faut alors se tourner vers d'autres sources d'ionisation à pression atmosphérique telles que l'APPI -*Atmospheric Pressure Photo Ionization*- ou APCI - *Atmospheric Pressure Chemical Ionization*-). De plus, elle permet l'analyse de composés sur une très large amplitude de masse, notamment du fait de la formation d'ions multichargés. Toutefois, même pour les composés monochargés, elle est efficace pour des masses à hauteur de 1000 unités de masse [28].

Avec une source ESI (Figure 4), l'ionisation s'applique à des composés en solution après nébulisation. L'électrospray est produit grâce à un fort champ électrique appliqué sur le liquide qui passe à travers un tube capillaire. C'est une différence de potentiel entre le capillaire et une contre-électrode qui est à l'origine du champ. Son effet en sortie de capillaire est la formation du cône de Taylor au niveau du liquide : les charges s'accumulent du fait de l'attraction par la contre-électrode tout en subissant un phénomène de répulsion [29,30]. Le cône de Taylor finit par se rompre et former des gouttelettes fortement chargées dans la source en présence d'un gaz de nébulisation. C'est ainsi qu'est obtenu un spray de l'échantillon introduit initialement à l'état liquide. En présence d'un gaz séchant, les gouttelettes s'évaporent et leurs composés sont ainsi désolvatés. Lorsque la densité de charge devient trop importante les gouttelettes finissent par se rompre créant des ions [10,17]. L'ionisation dans la source ESI est donc régie par deux principes : celui lié à l'évaporation du solvant et celui lié à la force coulombienne trop importante qui fait éclater les gouttelettes (dépassement de la limite de Rayleigh) [29].

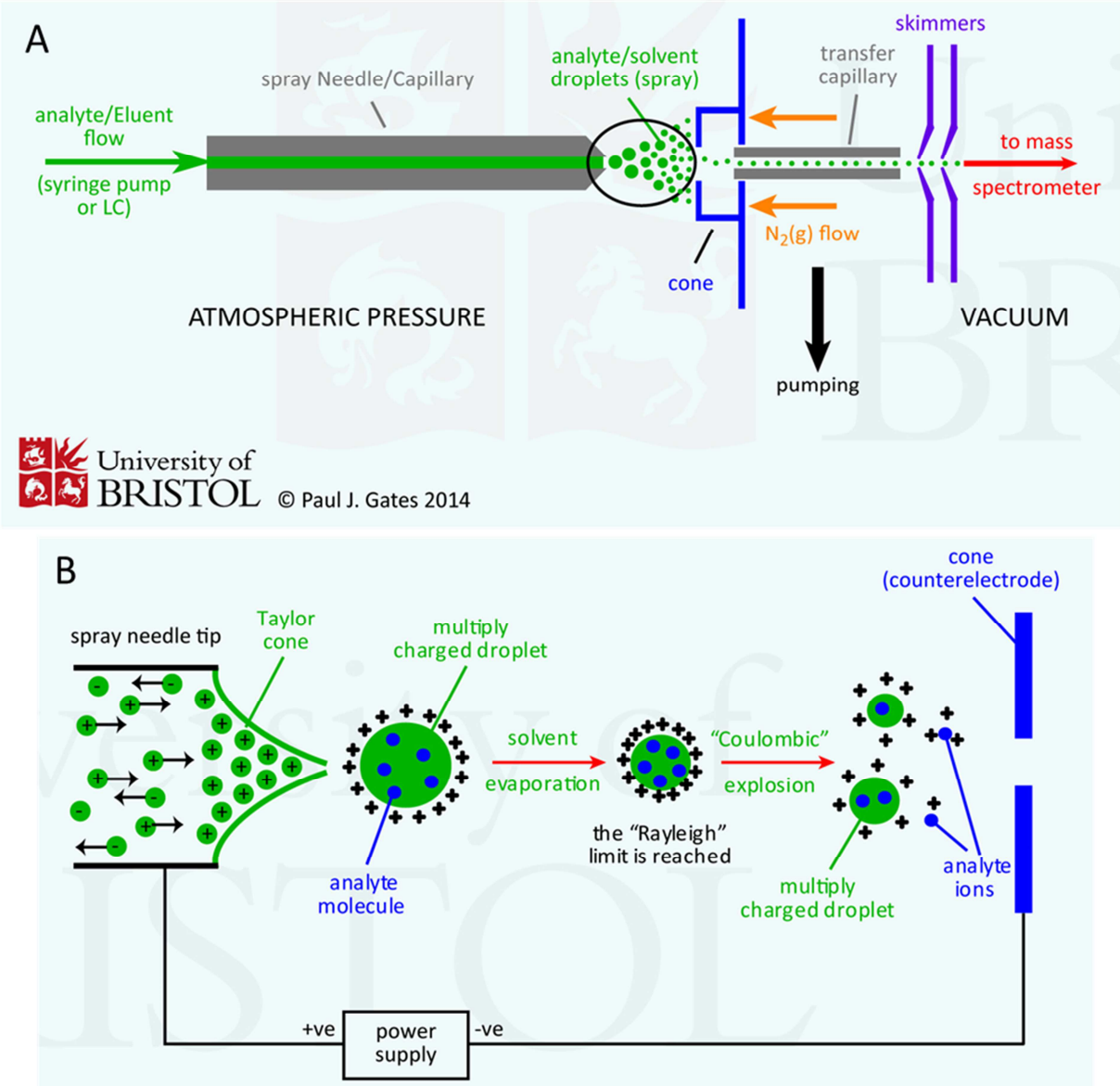


Figure 4 Schéma d'une source ESI (figures provenant du site de l'université de Bristol [31]). A : Vue générale de la source d'ionisation couplée en amont avec un système de chromatographie ou une seringue. B : Principe de formation d'ions positifs dans la source ESI

Lorsque des composés sont ionisés, il arrive qu'un phénomène de fragmentation en source intervienne. La fragmentation en source peut provenir notamment de la collision des ions avec des molécules du gaz séchant résiduel [29]. C'est un flux d'azote envoyé en sens opposé à celui du trajet des ions qui permet l'évaporation totale du solvant avant l'entrée des ions dans l'analyseur. Le gaz séchant joue aussi le rôle de limitation des possibles contaminations de l'analyseur. Cette éventualité de fragmentation en source peut complexifier les spectres de masse et rendre leur interprétation plus longue, notamment de façon manuelle. En effet, la fragmentation en source correspond à de la pseudo- MS/MS et rend plus difficile pour l'analyste, la détermination de l'ion pseudo-moléculaire ou de la masse moléculaire du composé détecté. Néanmoins, ces fragmentations apportent aussi du sens aux analyses car elles répondent à des mécanismes physico-chimiques précis et ne sont donc

pas aléatoires. Elles participent à la signature d'une famille phytochimique ou de types de fonctions particulières.

Des systèmes hybrides donnant accès à de bonnes performances existent. Pour nos travaux, l'analyse et la détection ont été menées sur un système Q-TOF (ou plus précisément QhTOF). Dans cette configuration, un quadripôle Q, un hexapôle h et un analyseur de masse à temps de vol muni d'un réflectron se succèdent (voir schéma du maXis de Bruker en Annexe 3.1). L'emploi d'un Q-TOF rend possible le mode d'analyse MS/MS : le quadripôle sert de filtre pour sélectionner et guider les ions précurseurs à fragmenter [32]. Ces derniers entrent dans la cellule de collision située entre le quadripôle et le TOF, il s'agit de l'hexapôle où ils sont fragmentés par application d'une énergie de collision d'intensité variable. Les ions fragments et les ions précurseurs restants passent dans le TOF [17]. Enfin un détecteur d'ions enregistre le signal des ions, ce qui permet d'obtenir les spectres de masse.

Le fonctionnement d'un analyseur TOF est le suivant : il sépare les ions en fonction de leur vitesse lors de leur déplacement dans la zone libre de champ que constitue le tube de vol. Un des intérêts lié au TOF provient de sa capacité à enregistrer la totalité des ions produits de façon simultanée. Lorsque le TOF est muni d'un réflectron, il a une résolution améliorée par rapport à un temps de vol linéaire. Ce réflectron est un élément qui permet de focaliser les ions de même rapport *m/z* dont l'énergie cinétique est légèrement différente au moment de quitter la source. Grâce au réflectron en milieu de parcours, les ions de rapport *m/z* identique atteindront le détecteur exactement en même temps [17].

4.2.3. Performances et intérêts principaux de la détection en ESI-Q-TOF

Avec un système de détection de spectrométrie de masse haute résolution, nos études sont effectuées avec un ensemble performant, précis et adapté à nos échantillons de produits naturels. Notre but est l'identification de métabolites secondaires. La source ESI est la plus adaptée grâce à la large gamme de composés qu'elle peut ioniser par rapport à d'autres sources et l'ionisation douce qu'elle permet.

Les analyseurs peuvent être combinés et utilisés en tandem. C'est le cas de l'association d'un quadripôle avec un analyseur à temps de vol (Q-TOF). Cette combinaison rend la détection très résolutive et précise [21,22]. La technologie TOF et les instruments hybrides Q-TOF permettent des mesures dont la précision en masse est en dessous de 5 ppm via l'utilisation de calibration interne ou de *lock mass* [21]. L'avantage du système, par rapport à un triple quadripôle, est lié aux

performances supérieures obtenues grâce au TOF pour la résolution et l'exactitude en masse notamment [17].

La résolution est un paramètre important des performances des analyseurs puisqu'elle correspond à la capacité de l'instrument à différencier deux pics de masse. Pour un pic gaussien, la résolution est calculée en divisant le rapport *m/z* du pic par la largeur du pic Δ(*m/z)* repérée à mi-hauteur du pic. Pour le TOF, la résolution est de 30 000 à *m/z* = 1000 à mi-hauteur du pic. L'analyseur hybride Q-TOF permet donc des analyses haute résolution, il est capable de distinguer des *m/z* de valeurs très proches, ceci permet d'enregistrer la masse des ions de façon très précise. A partir de ces masses monoisotopiques, on peut obtenir la formule brute des ions et déduire celle du composé analysé tout comme on peut avoir accès aux formules brutes de chaque fragment pour l'ensemble des fragments de MS/MS ce qui peut aider à les interpréter.

4.3 Profilage métabolique : méthodes fondamentales et approches actuelles

4.3.1. Problématique : défis rencontrés en recherche sur les produits naturels

Dans le contexte d'études phytochimiques, la caractérisation des composés constituants l'extrait demeure un défi majeur à relever [20,33] et engendre des étapes longues et laborieuses. Le profilage métabolique consiste à identifier un large groupe de composés pouvant être liés à une voie métabolique ou une classe de composés en particulier [16,34]. Ce profilage nécessite l'obtention d'informations précises à partir de matrices complexes. La description d'extraits végétaux n'est pas un travail évident que ce soit en utilisant des techniques classiques passant par la purification et la caractérisation formelle des molécules ou que ce soit en choisissant des méthodes analytiques couplant une méthode séparative avec une détection universelle (LC-HRMS, EC-MS, LC-RMN). En effet, dans ce dernier cas, le traitement de données reste conséquent s'il est manuel compte-tenu de la quantité d'informations à gérer. Aujourd'hui d'un point de vue technologique, il est possible d'accumuler un grand nombre d'informations mais le vrai travail reste celui de leur traitement et croisement pour accéder aux informations pertinentes et interprétables. Il faut donc avoir recours à des techniques sophistiquées et des méthodes avancées [16].

De plus, du fait de leur transversalité, les projets sont globalement complexes puisqu'il faut faire face à de multiples problèmes qui impliquent la coopération d'experts de différents secteurs tels que celui de la chimie analytique, la biochimie, l'informatique ou les statistiques [15].

A l'heure du *Big Data* qui regroupe les enjeux dits « 4V » de volume, variété, vélocité et véracité, une des préoccupations majeures est de traiter les volumes massifs de données de façon pertinente et rapide. Il faut donc tendre vers des méthodes efficaces, robustes et automatisées (ou semi-automatisées) pour limiter les erreurs possibles liées à des traitements manuels répétitifs. Pour répondre à cet enjeu d'efficacité dans la description moléculaire d'une plante, la tendance actuelle est d'adopter une stratégie basée sur l'automatisation d'une part importante des différentes étapes. Il s'agit donc de connaître les méthodes existantes et les processus performants, de les maîtriser pour les utiliser à bon escient afin que les résultats générés soient corrects. Le but est de réduire le temps consacré à ces travaux de recherche sans limiter leur pertinence et leur justesse.

4.3.2. Déréplication : Retrouver les molécules connues

D'après une définition initiale donnée en 1990 [20,35], la déréplication consiste en l'identification rapide des composés connus c'est-à-dire ceux déjà caractérisés et constituant l'espace chimique des molécules naturelles au moment de l'analyse. La déréplication permet d'annoter les composés connus sans avoir à les purifier. Ainsi, le travail d'identification peut être concentré sur un nombre restreint de nouveaux composés et cela peut guider l'isolement uniquement vers des composés inconnus et/ou d'intérêt [36]. De plus, la déréplication présente l'intérêt de pouvoir mettre en lumière la présence des composés dans une espèce ou un genre pour lesquels ils n'avaient pas encore été décrits.

Actuellement, différentes approches permettent de mettre en œuvre la déréplication sur un extrait ou une collection d'extraits. Pour décrire un échantillon par déréplication, un moyen commun est d'obtenir son profil chimique grâce à un système analytique pour ensuite comparer les spectres obtenus avec ceux de bases spectrales. Cela permet d'avoir une vue globale de la composition de l'échantillon et notamment des familles phytochimiques en présence. Toutefois, cette manière de procéder est limitée par la taille des bases spectrales disponibles. Elles restent très limitées ne représentant environ qu'un dixième des composés connus aujourd'hui (voir page 36 paragraphe 4.3.2 b)). La déréplication peut aussi impliquer des étapes de tests d'activités biologiques et de fractionnements pour chercher à identifier les composés de fractions biologiquement actives par exemple. Aussi, dans le cadre d'analyse ciblée, on peut effectuer la déréplication uniquement sur un nombre restreint de molécules d'intérêt. Ces molécules sont ciblées pour différentes raisons : pour leur bioactivité intéressante ou au contraire leur toxicité connue ou encore comme marqueur de l'authenticité de l'échantillon [20].

Globalement, les différentes méthodes de déréplication employées tendent toutes vers le même but : accélérer la découverte des produits naturels biologiquement actifs [14,20–23,28]. Si des stratégies innovantes sont nécessaires pour limiter le temps passé à la découverte de produits naturels, intégrer une étape de déréplication dans le processus de profilage chimique est devenu fondamental [20].

4.3.2 a) *Base de données de produits naturels*

A partir des données analytiques, identifier des molécules nécessite souvent l'interrogation de bases de données de métabolites [21] qui répertorient les structures naturelles connues et formellement identifiées. Combiner l'analyse de données de MS avec la consultation de bases de données de produits naturels rend possible l'identification de signaux discriminants [21]. De nombreuses bases de données de composés existent sans être toutes accessibles librement ni complètement différentes les unes des autres [33,37,38]. Aussi, la communauté scientifique tente de rassembler et d'harmoniser ces données afin de rendre leur partage et leur utilisation de plus en plus simple et efficace. Les bases générales les plus importantes en termes de taille ou d'intérêt pour la recherche sur les produits naturels sont mentionnées dans le Tableau 3. Certaines bases sont dédiées à un type de ressource en particulier.

Base de données	Type de données	Accessibilité	Intérêt
PubChem	Petits composés organiques (structure, nomenclature) et activités biologiques	Recherche libre www.ncbi.nlm.nih.gov/pccompound?cmd=search	Richesse et versatilité
Scifinder	Composés organiques, résumés de publications, brevets	Recherche sous licence scifinder.cas.org	Richesse Lien composés-références/brevets/fournisseurs
KEGG	Métabolites et voies métaboliques	Recherche libre www.genome.jp/kegg/	Liens entre métabolites et informations sur les protéines/enzymes
***CHEMnetBase* (DNP)**	Composés naturels	Recherche libre Export de résultats sous licence dnp.chemnetbase. com	Richesse Fréquemment mise à jour
Universal Natural Products Database [39]	Composés naturels	Libre accès pkuxxj.pku.edu.cn/UNPD	Richesse
Supernatural II [40,41]	Composés naturels	Libre accès bioinformatics.charite.de/supernatural/	Richesse

Base de données	Type de données	Accessibilité	Intérêt
KNApSAck [42]	Composés naturels, Organismes, Pairs composé-organisme	Libre accès kanaya.naist.jp/knapsack_jsp/top.html	Fréquemment mise à jour
GPDB	Composés naturels, Organismes, activités	Recherche interne Greenpharma	Richesse Lien composé-organisme-activité
Antibase	Composés naturels de microorganismes et champignons	Accès sous licence application.wiley-vch.de/stmdata/antibase.php	Spécificité
MarinLit	Composés d'origine marine	Accès sous licence pubs.rsc.org/marinlit/	Spécificité

Tableau 3 Base de données de produits naturels ou généralistes utiles à la recherche sur les produits naturels

Les bases de données constituent des ensembles finis de molécules et ceci créé une limite à leur utilisation lors de l'analyse de matrices complexes dans la mesure où l'on sait que l'espace chimique des molécules naturelles est plus vaste que celui que nous sommes en capacité de décrire aujourd'hui. Ainsi, une tendance est à la génération de bases de données virtuelles qui s'appuient sur les bases existantes. Cela permet une expansion des bases de données structurales de composés connus avec des produits naturels plausibles générés *in silico* selon des voies biosynthétiques [33]. Un exemple est celui de MINEs (*Metabolic In Silico Network Expansions*) [43] proposé en 2015 et basé sur l'algorithme '*Biochemical Network Integrated Computational Explorer*' (BNICE) [44,45]. Cette base de données se veut une extension des bases de données et inclut des molécules pas encore observées mais qui existent probablement puisque ces nouvelles structures sont issues de l'application de réactions biochimiques.

4.3.2 b) *Base de données spectrales*

La revue de Johnson et Lange de 2015 [37] à propos des bases de données d'intérêt pour la recherche sur les produits naturels référence et détaille l'ensemble des outils utiles et accessibles librement. Un nombre important de bases spectrales tant de MS que de RMN existent aujourd'hui. Toutefois, elles ne sont pas toutes très conséquentes en nombre de composés décrits et comparées les unes aux autres, elles peuvent contenir des doublons.

Ce type de bases rassemble des spectres de masses expérimentaux de référence pouvant être de niveau MS, MS/MS ou MS^n pour la spectrométrie de masse; ^{13}C, ^{1}H ou spectres 2D pour la RMN. Elles permettent d'annoter des composés d'après la comparaison des spectres expérimentaux

avec les spectres référencés dans la base. Les principales bases de données spectrales sont : Metlin [46], Massbank [47], NIST [48], mzCloud [49], ReSpect [50] ou encore GNPS [51], les deux dernières étant plus spécifiques des produits naturels. GNPS est une plateforme récente qui a pour but de rassembler de nombreuses données et a vocation à les rendre disponibles publiquement. GNPS réunit les bases spectrales disponibles et libres et des collections de spectres nouvellement acquis et fournis par des contributeurs de la communauté GNPS. Des outils dédiés à la RMN existent également. Les bases de spectres comme NMR ShiftDB [52], NAPROC-13 [53], MMCD [54], BMRB [55] sont également accessibles en ligne.

Actuellement, les bases de spectres de masses restent incomplètes et, de ce fait, pas totalement pertinentes. Elles décrivent environ 20 000 molécules (tandis que le nombre de composés connus est de l'ordre de 200 000) [33,56]. De plus, en comptant sur ce genre de répertoires, on se confronte à la difficulté liée au fait que le spectre de masse obtenu est très dépendant des différentes conditions opératoires employées. En effet, en fonction de la nature de la source d'ionisation, du mode d'ionisation, de l'énergie d'ionisation et de l'ion précurseur choisi ainsi que de l'énergie de collision utilisée pour la fragmentation, les spectres de masse obtenus seront différents sans compter les effets de matrices ou de suppression d'ions qui peuvent également survenir [20]. Ainsi, la comparaison est impossible si les conditions expérimentales ne sont pas rigoureusement équivalentes. Ceci est vrai aussi pour les répertoires de spectres RMN où le solvant notamment a une influence sur l'allure des spectres.

Pour pallier le nombre limité de spectres de masses dans les bases existantes, une tendance émerge : celle de la génération et l'utilisation de bases de données spectrales *in silico* [36,57]. Les spectres référencés sont théoriques et obtenus grâce à des outils de modélisation de la fragmentation (MS/MS) ou de la résonance (RMN). Le paragraphe 4.3.3 d) en page 45 de ce chapitre développera ces aspects pour la MS. Concernant la RMN, des solutions ont été mises au point tels que les logiciels de prédiction de spectres 1D, 2D des noyaux ^{13}C et ^{1}H principalement [58,59]. C'est le cas des outils tels que *NMR predictors* proposés par ACD/Labs [60] qui développe de nombreux logiciels concernant des problématiques de chimie ou encore *NMR Prediction* de PERCH [61].

4.3.3. Stratégies et plateformes pour le traitement, l'organisation et l'annotation des données

Afin de conduire des études approfondies sur les produits naturels, il apparaît évident d'utiliser –si ce n'est de mettre au point- une stratégie méthodologique pour organiser le travail afin qu'il soit pertinent et performant. Des stratégies globales et semi-automatisées impliquant tant la

MS que la RMN existent : elles s'appuient sur la puissance des bases de données et sur les performances des outils computationnels pour devenir plus complètes et prendre un caractère holistique [11], c'est-à-dire où le contexte global de l'étude est pris en compte pour donner du sens à l'ensemble des données. Les supports computationnels servent dans les étapes de manipulation ou traitement de données et élucidation de structures [11].

La manipulation et le traitement de données pour la recherche sur les produits naturels ou en métabolomique nécessitent différentes étapes incontournables à partir des données brutes et avant qu'une interprétation ne soit possible. Le prétraitement transforme tout d'abord les données brutes en données gérables pour les algorithmes à suivre. Puis le traitement réfère à l'analyse des données par l'application des algorithmes (calculs). Enfin, le traitement final des résultats permet de les synthétiser ou de les visualiser pour faciliter leur interprétation [62].

La déréplication et la métabolomique sont des exemples de stratégies exploratoires et sont de plus en plus utilisées dans les sciences du vivant et dans le domaine des produits naturels.

La métabolomique est une approche dite compréhensive, c'est-à-dire qui a pour but d'identifier et de quantifier l'ensemble des métabolites d'un système biologique [16]. Néanmoins, l'objectif de la métabolomique étant ambitieux, certaines études tente de le relever tandis que dans d'autres se concentrent sur les différences entre différents échantillons [34]. L'emploi d'outils statistiques est également important en métabolomique pour réduire le jeu de données et mettre en évidence les variations ou différences au niveau du métabolome dans des échantillons [16].

La métabolomique est aussi définie comme un profilage chimique global et non ciblé [24]. Toutefois, le profilage chimique est une des stratégies pouvant permettre la conduite d'études métabolomiques mais n'est pas forcément inscrit dans une démarche métabolomique qui met en jeu un grand nombre d'échantillons et concerne souvent l'étude à grande échelle du contenu métabolique d'un organisme en réponse à des variations de son environnement ou dans différents états physiologiques [16,20].

4.3.3 a) *Stratégies adossées à la résonance magnétique nucléaire*

Pour décrire de façon générale ce qu'il est possible de réaliser en utilisant la RMN, la Figure 5 présente les différents aspects à prendre en considération et les moyens actuels utilisés. La RMN demeure une technique d'intérêt qui évolue tant en termes d'instrumentations que d'approches scientifiques. En ce qui concerne sa mise en œuvre, le couplage à des méthodes de pré-concentration (SPE[§]-RMN) ou de fractionnement (LC-RMN, CPC[**]-RMN) devient plus fréquent pour pallier les limitations de la technique de RMN. Toutefois, les instruments ont également évolué notamment pour gagner en sensibilité. Des moyens computationnels et automatiques (ou semi-automatiques) ont commencé et devront continuer à avoir un rôle clé dans l'utilisation de la RMN dédiée à l'identification de composés naturels [11]. Parmi eux on trouve des outils de prédiction de spectres mais également d'élucidation automatique. Les méthodologies d'assistance informatique pour l'élucidation de structures (CASE) cherchent, grâce à des algorithmes, les correspondances entre des données RMN expérimentales et des spectres standards enregistrés dans des bases [11,58].

Des approches de déréplication ont également été développées en s'appuyant sur la signature des composés en RMN du carbone. En 2014, l'équipe de Hubert *et al*. a développé une approche qui consistait à fractionner un extrait en CPC avant d'enregistrer les spectres RMN de ^{13}C de chaque fraction. L'emploi d'une classification ascendante hiérarchique permet de regrouper les déplacements chimiques communs à plusieurs fractions et correspondant au même métabolite. Afin d'identifier les composés de l'extrait, la dernière étape est l'interrogation d'une base de données de spectres de RMN de ^{13}C [63]. En 2017, la même équipe a publié une méthodologie de déréplication par analyses en RMN ^{13}C sans fractionnement préalable [64]. Cette évolution a été possible grâce à l'implémentation d'un algorithme d'évaluation de la correspondance entre les données expérimentales et des données de RMN prédites stockées dans une base de données. L'algorithme de recherche propose une liste de molécules potentiellement présentes dans l'échantillon.

La RMN en deux dimensions a également une place prépondérante. Elle apporte un degré supplémentaire à la RMN à une dimension. Pour parvenir à élucider des structures de produits naturels qui sont complexes (plus de 20-25 atomes) et isolés pour la première fois, le recours à la RMN 2D est indispensable afin d'établir les connexions qui existent entre les atomes [58].

[§] SPE : Solid Phase Extraction
[**] CPC : Chromatographie de Partage Centrifuge

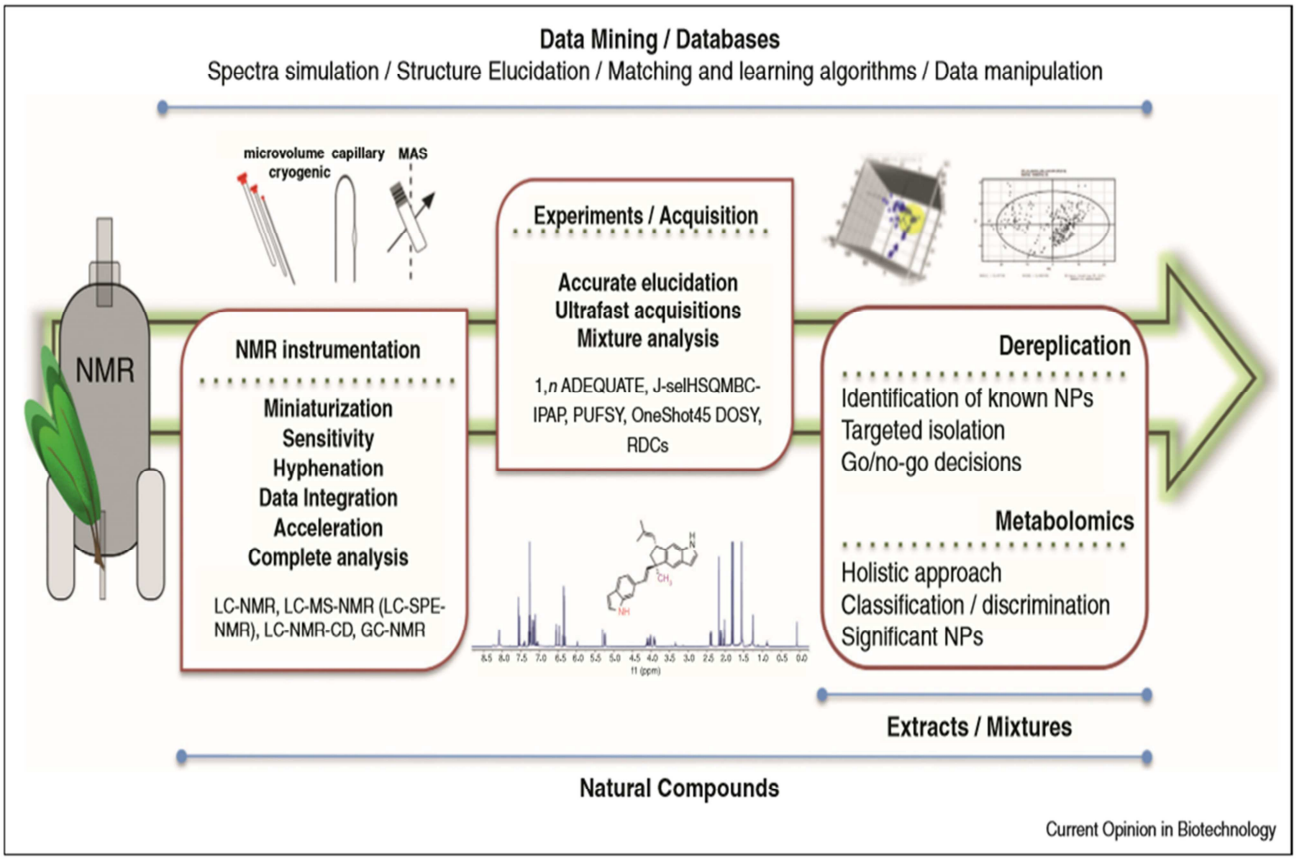


Figure 5 Utilisation de la RMN pour la recherche sur les produits naturels : avancées techniques et stratégies récentes. Schéma tiré de la revue de Halabalaki de 2014 [11].

4.3.3 b) *Plateformes de gestion et de traitement de données de LC-MS*

Dans la suite du chapitre seront détaillées les méthodologies - nombreuses et très utilisées- adossées à la chromatographie en phase liquide couplée à la spectrométrie de masse puisque cette technique est au cœur du projet de thèse présenté. Les stratégies et plateformes développées pour traiter les larges jeux de données sont notamment employées dans des démarches de métabolomique mais aussi pour d'autres applications d'études de composés organiques.

Les plateformes d'exploration (*data analysis/mining*) et de traitements (*data processing*) de larges jeux de données ont été conçues pour gérer chaque étape des procédures de traitements de données LC-MS. Elles permettent la conversion de données brutes en liste de pics annotés d'un temps de rétention et d'une formule brute par exemple [26,33]. Les fonctionnalités vont du prétraitement des données LC-MS (détection de pics, alignements des pics, prédiction de formules, calculs d'adduits) à la normalisation et aux études statistiques ainsi qu'à la mise en forme pour une visualisation simplifiée des jeux de données. Elles constituent une alternative aux traitements

possibles avec les logiciels de retraitement des constructeurs mais surtout elles permettent le traitement de données de tout constructeur du moment que les données sont importées dans un format qu'elles prennent en charge. Parmi les plateformes pour traiter et organiser les données avant de pouvoir les analyser, on peut accéder librement à : XCMS [65], MZmine 2 [66], Open MS [67], MetAlign [68] et le portail collaboratif Workflow4Metabolomic 3.0 [69]. XCMS est une plateforme très largement utilisée notamment parce qu'elle propose une interface web XCMS Online (xcmsonline.scripps.edu) avec de nombreuses fonctionnalités génériques pour traiter et visualiser les données de MS notamment dans le cadre d'études métabolomiques. MZmine 2 constitue une autre plateforme populaire qui a pour but principal d'être un logiciel flexible et facile d'accès avec un ensemble de modules couvrant le processus complet d'analyses de données LC-MS.

Une des principales vocations de ces plateformes est de pouvoir effectuer des statistiques sur les jeux d'échantillons analysés. Ces analyses statistiques mettent en exergue les différences ou les corrélations entre les échantillons afin de décrire les variations du contenu métabolique d'échantillons. En effet, si l'annotation de l'ensemble des métabolites demeure un défi complexe à relever, il est plus judicieux de chercher à mettre en évidence les analytes différentiant d'échantillons et de chercher à identifier ces composés ; c'est le cas de nombreux travaux effectués en métabolomique dite ciblée.

Ces plateformes illustrent la tendance actuelle de plus en plus suivie du partage des traitements, des données et des résultats pour compiler des informations et aller plus loin dans le développement des outils. A ce titre, Workflow4Metabolomics (W4M) [69], défini comme un portail collaboratif en ligne, est destiné à la fois à des utilisateurs sans compétences en programmation tout comme à des équipes de développeurs en métabolomique ou bioinformatique. Pour les premiers utilisateurs, il est possible d'effectuer des analyses computationnelles de données brutes résultant en l'annotation de composés à l'aide d'un ensemble de modules gérant les étapes de traitement et d'analyses de données LC-HRMS et d'interrogation de bases de données. Pour les développeurs, W4M est une ressource accessible, libre et ouverte au développement d'autres infrastructures adaptées à leur propre recherche.

4.3.3 c) *Outils spécifiques des données MS*

Une des premières étapes dans l'identification de composés est la détermination de la formule brute ou de la masse monoisotopique [33]. C'est à cette information capitale que doit conduire l'analyse MS. Toutefois, les données de MS et en particulier la redondance des données issues de la même molécule nécessite un travail important et conséquent pour identifier formellement la masse moléculaire de la structure analysée. L'aspect redondant est lié à l'ionisation notamment en source *electrospray* , processus d'ionisation douce à pression atmosphérique qui peut générer différents ions pour une même molécule à savoir des adduits, des *clusters*, des pertes de neutres, des fragments en source ou des dimères [21,70,71]. Les adduits sont issus de l'interaction non covalente entre le composé et les ions présents dans la phase mobile (Na^+, K^+, Cl^-) après l'étape d'évaporation dans la source [72]. Les clusters sont des ensembles formés par une molécule ionisée avec des molécules neutres présentes dans le milieu. Les dimères (et en général les ions polymériques) se forment d'autant plus que la concentration du composé est importante dans l'échantillon [28]. La fragmentation en source se produit lorsque des liaisons se rompent dans la source dans les étapes d'ionisation ; ces ruptures de liaisons peuvent avoir lieu à différents endroits sur la molécule, ainsi de nombreux fragments peuvent être libérés. Les isotopes (un même élément chimique avec un nombre différent de neutrons) sont aussi à l'origine des signaux distincts pour une même formule brute. Chaque signal correspondant à une formule brute se présente donc sous la forme d'un massif isotopique lié à l'abondance naturelle des différents isotopes [72]. D'une part, cette multitude de signaux reliés au même composé est précieuse pour la détermination de la masse monoisotopique du composé analysé. D'autre part, elle complique l'étude puisqu'il faut passer un certain temps pour déterminer ce que représente chaque ion et être capable d'établir les relations entre les ions détectés, ce qui constitue l'étape d'annotations des ions [21].

Pour faciliter cette annotation, les données MS peuvent être gérées à l'aide d'outils automatiques [71,72]. Ces outils d'annotations des spectres MS sont d'un intérêt majeur pour déterminer rapidement et efficacement les masses neutres des composés, données indispensables pour l'identification de composés et aussi pour utiliser les outils de fragmentation *in silico* dans la suite du processus [73]. Plusieurs approches sont présentées dans une revue récente publiée par le *Scripps Center for Metabolomics* [71] : nous allons nous concentrer sur l'approche la plus communément utilisée. Il s'agit de l'annotation accomplie en comparant les distances entre les masses pour les ions expérimentaux avec les distances décrites pour des combinaisons théoriques d'adduits connus, de pertes de neutres, d'ions multichargés ou aussi de fragments en source (Figure

6). L'établissement de la relation entre deux ions se fait lorsque la différence entre les valeurs de *m/z* de chaque ion correspond à une valeur définie à plus ou moins l'erreur relative.

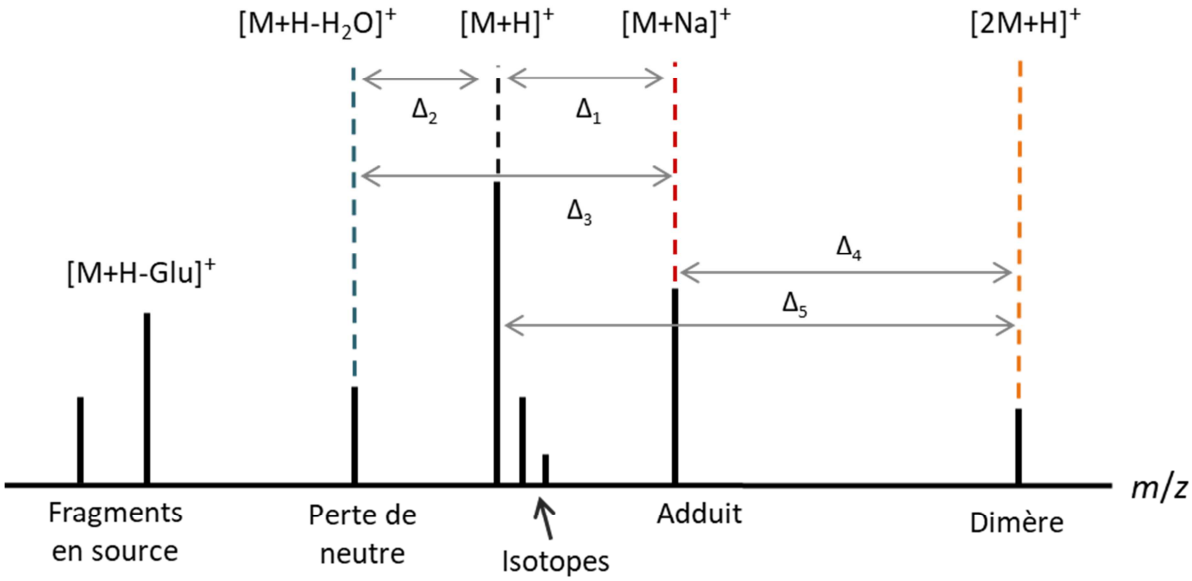


Figure 6 Schéma de spectre MS en mode d'ionisation positif ; illustration du principe d'annotation par évaluation de la différence (Δ_x) entre les différents ions générés par formation d'adduits, de dimères, de pertes de neutres, de fragments en source et présence d'isotopes.

Quelques modules d'interprétation et d'annotations des données MS -parmi les plus proches de la stratégie qui nous a inspiré ou qui sont largement cités aujourd'hui dans les revues- sont présentés au sein du Tableau 4. Ces différentes approches sont disponibles sous forme de *script* informatique ou de paquet (ensemble de fonctions codées pour effectuer des traitements de données sur une problématique ciblée) qu'il est nécessaire d'installer et de savoir utiliser pour les faire fonctionner (exécution). Le langage R est destiné aux programmeurs et bloque potentiellement les non-initiés à son usage pour l'analyse de données [74]. Cela requiert donc des compétences informatiques.

Outil	Référence Date de publication	Intérêt(s)	Limitation(s)	Accessibilité / format et langage informatique
CAMERA	[75] 2012	Si absence de correspondance entre ions d'un spectre, pas d'utilisation d'heuristique (règle arbitraire par défaut) pour annoter les ions Outil reconnu et utilisé	Validations de l'outil sur nombre restreint de composés (39 standards) ou par analyse ciblée	Oui / Paquet R

Outil	Référence Date de publication	Intérêt(s)	Limitation(s)	Accessibilité / format et langage informatique
Precursor mass prediction by clustering ionization products	[76] 2013	Outil allant du prétraitement des données analytiques (groupement et alignement des spectres et temps de rétention) à l'interrogation de bases de données (type généraliste) Prétraitement des données (avant recherche des masses neutres) s'appuie sur des analyses répétées de l'échantillon : Robustesse	Score attribué à une masse déterminée : uniquement fonction du nombre d'ions qui expliquent la même masse neutre Validation de l'outil sur nombre restreint de composés (32 standards)	Non / Non communiqué
MetAssign	[77] 2014	Annotation avec approche statistique de groupement des ions Mesure probabiliste du niveau de confiance de l'annotation	Méthodologie mise en place un peu complexe	Oui / Codes R
mz.unity	[72] 2016	Procédure d'annotation étendue à des relations complexes : adduits complexes, hétérodimères, fragmentations	Améliorations en perspectives : annotation de fragments rares, discrimination entre fragments et composés distincts	Oui / Paquet R
FindMAIN	[73] 2017	Détermination de la nature de tous les ions majeurs (ions d'intensité supérieure à un seuil relatif à l'ion majoritaire) Score dépendant de multi-paramètres (intensité, précision en masse, motif isotopique) Validé sur plus de 1000 spectres en mode positif	Non testé et non validé en mode négatif (autre publication à venir)	Oui / Paquet R

Tableau 4 Outils d'annotation des données de MS

Identifier un composé à partir de la MS avec la masse monoisotopique ou la formule brute qu'on peut en déduire n'est pas suffisant. Une masse monoisotopique donnée (ou de la même manière pour une formule brute donnée) recherchée au sein d'une base de données peut correspondre à un grand nombre de candidats. Il devient alors nécessaire de traiter les données de MS/MS afin d'accéder à des éléments plus discriminants.

4.3.3 d) *Outils spécifiques des données MS/MS*

Les données de MS/MS présentent un intérêt certain puisqu'elles donnent accès à un niveau d'information plus détaillé des structures que les données MS. Les informations structurales tirées de l'analyse des fragments (et/ou des chemins de fragmentation) permettent d'aller au-delà de la masse du composé associée à sa formule brute puisqu'elles rendent possible l'établissement de la structure en partie ou totalement. Les fragments d'une structure ne sont pas issus de ruptures aléatoires : elles sont régies par des mécanismes et il y a des liaisons plus faibles que d'autres qui auront tendance à se rompre plus facilement. Il est notamment possible d'observer les ions issus de ruptures particulières et informatives telles que la perte de parties glycosylées, la perte de fonctions hydroxyle ou acide carboxylique mais également rencontrer des ions obtenus par des réactions de réarrangements, d'éliminations de type rétro-Diels-Alder. La revue de Demarque *et al*. [78] aborde ces mécanismes de façon théorique, illustrée et exhaustive dans le cas de fragmentations dans les cellules de collisions après ionisation en ESI pour différentes familles de produits naturels.

Le champ d'investigation des fragmentations est large, de nombreuses équipes y travaillent et le nombre de publications et outils en découlant est important. Des revues très récentes comme celle de Misra sortie en 2018 [79] référencent les approches développées ces dernières années. Ainsi sans chercher à être exhaustif, nous évoquerons quelques outils d'intérêt par rapport à notre travail qui sont des solutions reconnues car de nombreux acteurs du domaine les utilisent et les citent dans leurs travaux.

Afin de traduire ce que les données MS/MS contiennent en termes d'éléments structuraux, un premier aspect concerne l'organisation des données de fragmentation comme l'expose la revue de Allard [33]. Cette organisation peut être gérée soit grâce à des réseaux moléculaires (plateforme GNPS) soit à l'aide d'arbres de fragmentation (XCMS, MEtITree). L'organisation correspond au fait de lier les masses neutres obtenues en amont lors du traitement des informations MS selon un paramètre tel que la similarité spectrale ou structurale ou la présence d'une sous-structure particulière [33]. Un second aspect aborde l'annotation de données de fragmentations. On peut classer les différentes méthodes semi-automatisées de traitement des données de MS/MS en deux grandes catégories: la première est la comparaison avec des données de bases spectrales, la seconde étant l'utilisation de bases de données de structures associées à des programmes informatiques. Parmi ces outils informatiques affiliés à l'approche prédictive dite *in silico*, certains ont un fonctionnement basé sur des règles connues de fragmentation et d'autres sont des fragmenteurs combinatoires. La prédiction de fonctions structurales et d'empreintes moléculaires ainsi que la génération d'arbres de fragmentation ou la mise en réseau moléculaire appartiennent à d'autres

approches prédictives aujourd'hui développées et utilisées. Le Tableau 5 référence des outils d'aide à l'annotation des données de fragmentation et indique le principe avec lequel il fonctionne.

Approche / Outil	Référence / Date	Principe				
		A	B	C	D	E
Metfrag et Metfrag *relaunched*	[80,81] / 2010-2016					
Metfusion	[82] / 2013					
XCMS^2	[83]/ 2008					
MIDAS et MIDAS-G	[84,85] / 2014-2017					
Mass Frontier	[86] / Version 7.0 2017					
HAMMER	[87] / 2014					
CFM-ID	[88] / 2015					
SIRIUS + CSI:FingerID	[56] / 2015					
iMet	[89] / 2017					
GNPS	[51] / 2016					

Tableau 5 Référencement non exhaustif d'outils d'intérêts pour l'étude et l'annotation de données de MS/MS. Classement selon principe utilisé ; Comparaison à une base spectrale (A), comparaison à une base structurale avec utilisation d'outils computationnels de fragmentation *rule-based* (B) ou d'outils de fragmentation combinatoire (C), génération et comparaison d'arbres de fragmentation ou d'empreintes moléculaires (D), réseaux moléculaires (E).

Metfrag, un des premiers outils développés, est un système de fragmentation combinatoire utilisant le principe de dissociation de liaisons. CFM-ID (*Competitive fragmentation modeling for metabolite identification*) tout comme Metfrag, prédit des candidats structuraux possibles pour un composé donné et attribue un rang à chaque candidat en fonction de la correspondance des spectres prédits avec le spectre expérimental. La principale différence de fonctionnement de CFM-ID d'avec Metfrag (hormis la fonction de score pour classer les candidats) est la façon dont il génère les spectres *in silico* puisqu'ils sont obtenus suivant le principe de fragmentation compétitive. Ceci consiste en une prédiction de fragmentation basée sur un modèle probabiliste appris (*machine learning*) d'après des données expérimentales. Le modèle indique les fragments les plus probables d'être observés, augmentant ainsi la précision de l'outil [90].

Avant les années 2010 environ, l'analyse automatique de ces données consistait majoritairement à interroger des bases spectrales. Le spectre MS/MS expérimental est comparé à des spectres de référence principalement via la méthode dite du compte de pics partagés qui dénombre le nombre d'ions en commun. XCMS^2 fait partie de cette catégorie d'outils et interroge la base de spectres METLIN de deux façons. La première restreint la recherche aux spectres de référence ayant le même ion précurseur que pour le spectre expérimental. La seconde ne restreint

pas la recherche, et cherche des spectres ayant une grande similarité avec celui du composé inconnu, ce qui permet d'indiquer des sous-structures pouvant être présentes dans le composé inconnu. Très récemment (2018), METLIN (base de données de spectres MS/MS rendue accessible dès 2005) est devenue une plateforme pour identifier les composés connus et inconnus [91] : des recherches simples dans la base de données aident à annoter les composés, la comparaison des données MS/MS permet l'identification de métabolites et des fonctionnalités de recherche de similarité ou de pertes de neutres particulières sont proposées pour tenter de caractériser des composés inconnus.

Depuis 2010, les outils d'analyse des spectres de fragmentation pour l'identification des petites molécules ont été diversifiés pour remplacer les librairies spectrales par des libraires plus grandes afin de dépasser les frontières des bases spectrales [57,92]. Néanmoins, s'appuyer sur ces bases de structures se heurte toujours à la même limite absolue liée au fait que l'espace chimique naturel n'y est pas totalement décrit. Toutefois, ces outils apportent une aide précieuse pour donner du sens aux données MS/MS et accéder à une première approche du type de composé analysé. Par ailleurs, une approche utilisant des arbres de fragmentation ne requiert pas de bases de données du tout [57]. A partir d'un spectre de masse de fragmentation et d'une formule brute, les arbres sont constitués de nœuds représentant les différents ions (précurseur et fragments) reliés par des ponts qui symbolisent la perte en masse. L'approche nommée HAMMER (*High-throughput AutoMation of Mass frontiER*) [87] est un moyen de générer une base spectrale *in* silico d'intérêt au moment de l'étude à partir de bases telles que PubChem, ChemSpider ou KEGG. Les exemples d'application de l'outil sont restreints à une voie métabolique particulière dans le premier cas et à un type de composés particuliers à mettre en évidence dans le second cas. C'est donc un travail de recherche plus ciblé qui peut être mis en place avec cet outil.

Des approches hybrides sont apparues également : elles combinent deux méthodes afin d'inclure le plus d'informations possibles dans le processus et de tirer parti des avantages de chaque méthode pour élucider les structures [90]. C'est le cas de Metfusion qui allie la fragmentation combinatoire via l'intégration de Metfrag avec l'interrogation de bases spectrales. Cette combinaison a permis l'amélioration des résultats avec Metfusion par rapport à Metfrag utilisé seul. En effet, Metfrag permet d'atteindre le niveau 3 des MSI tandis qu'avec Metfusion, le niveau d'identification se situe entre les niveaux 2 et 3 (revoir Tableau 2 du paragraphe 4.1.3). De même, CSI:FingerID incorpore les informations d'arbres de fragmentation (obtenus grâce à l'outil SIRIUS) avec la prédiction d'empreintes moléculaires. CSI:FingerID comporte trois grandes étapes : la première est celle de l'apprentissage où des données MS/MS de référence sont utilisées pour entraîner des calculs de prédictions de fonctions moléculaires (empreintes). La seconde phase consiste en l'obtention de l'empreinte moléculaire du composé inconnu pour troisièmement, comparer l'empreinte de

l'inconnu avec celles de composés contenus dans des bases de données (de type PubChem) et trouver la meilleure correspondance.

Plus récemment, iMet tente de faciliter l'annotation de composés absents de bases de données. A partir du spectre MS/MS et de la masse exacte du composé, l'outil retrouve les composés similaires au composé inconnu au sein d'une base spectrale. Pour cela, il passe par la création de réseaux de composés voisins. L'utilisation des réseaux moléculaires semble grandissante et est plébiscitée notamment grâce au développement de la plateforme GNPS. Les grandes fonctions proposées par GNPS sont la déréplication (par interrogation de la base spectrale communautaire), la déréplication variable (recherche de spectres tolérant une modification pour permettre la détection d'analogues de composés connus qui peuvent appartenir à la même classe de composés ou différer d'une modification ou substitution d'un groupe chimique uniquement) et enfin la création de réseaux moléculaires. Ils correspondent à une cartographie de similarité spectrale. Ces réseaux moléculaires sont construits à partir de l'alignement entre deux spectres. Les ponts qui connectent les nœuds (c'est-à-dire des spectres MS/MS) sont définis par un score de cosinus modifié (similarité cosinus) qui permet d'évaluer la similarité de deux objets, ici, deux spectres MS/MS. L'hypothèse de base qui est faite est que deux molécules de structures proches vont avoir des spectres de masse similaires. La fonction clé de GNPS de mise en réseaux des composés détectés en fait un outil d'organisation et de visualisation des données MS/MS, mettant en lumière l'existence de groupes spectraux ainsi que leurs degrés de similarité [93].

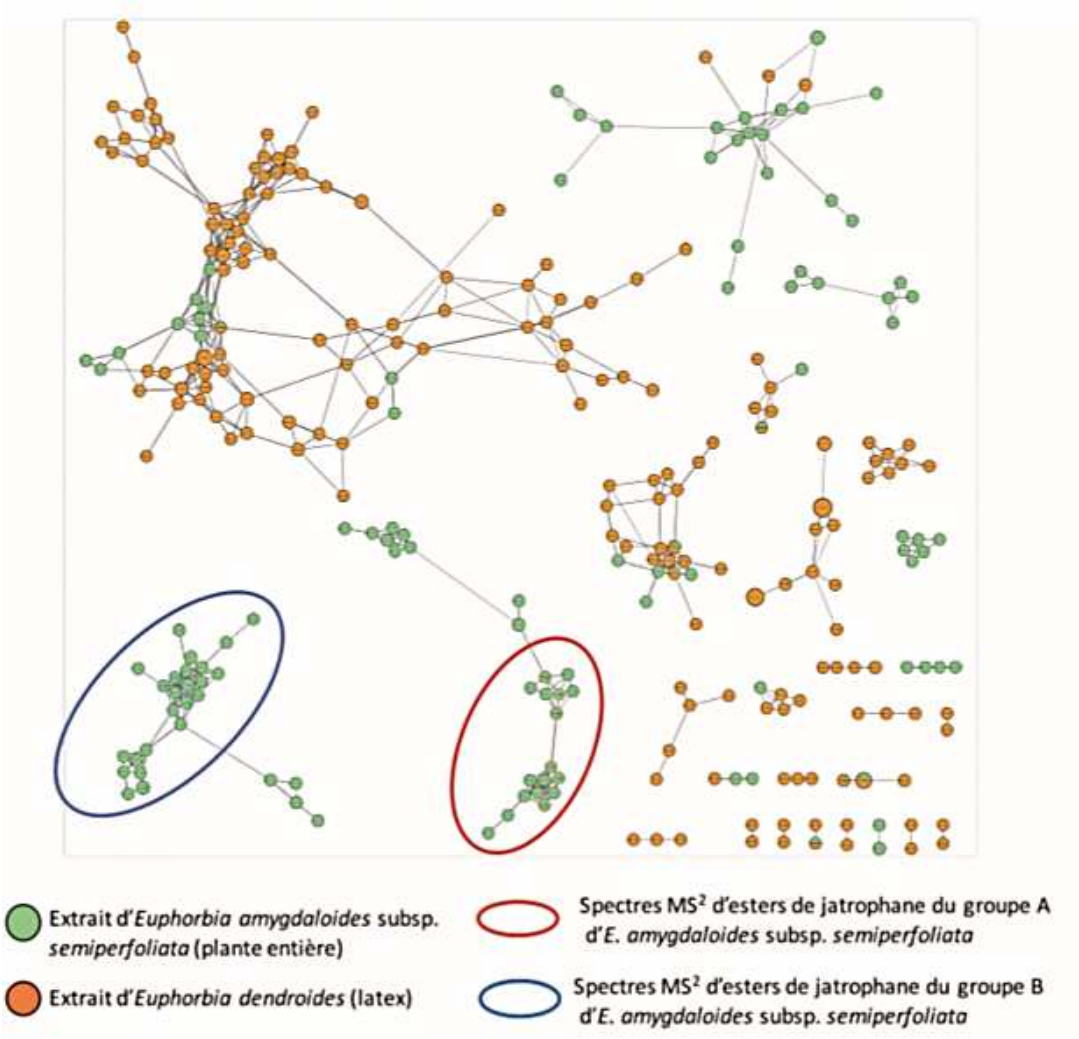


Figure 7 Exemple de réseaux moléculaires obtenus à partir de données MS/MS d'extraits de plante, d'après un papier de *Nothias-Scaglia et al*. ([93]

Un exemple intéressant d'application des réseaux moléculaires associé à la fragmentation *in silico* a été décrit par Allard *et al*. [36]. L'étude tire parti des avantages de l'association des deux approches dans une démarche de déréplication. Dès lors qu'un nœud a été associé à un type particulier de structure, le nombre de composés pouvant être annotés augmente du fait même de la mise en relation de composés au sein du réseau. De plus, ces réseaux rendent possible la recherche d'analogues structuraux grâce à la recherche de fragmentation similaire de composés de masses différentes. A cela s'ajoute la fragmentation *in silico* de composés de bases de produits naturels qui permet de dépasser les frontières de bases de données spectrales expérimentales.

L'utilisation des réseaux moléculaires via GNPS a été améliorée grâce à la proposition d'une étape de prétraitement gérée avec MZmine 2 pour contourner certains inconvénients intrinsèques à GNPS : cette combinaison a été publiée par Olivon en 2017 [94]. Grâce à cette approche conjointe, les auteurs proposent un moyen de déréplication plus pertinent avec : la conservation de l'information des isomères séparés en LC grâce à la prise en compte du temps de rétention pour

annoter les nœuds et, l'annotation des nœuds via l'interrogation de bases de composés ou la prédiction des formules brutes.

Extraire les structures moléculaires à partir des données MS/MS, pas seulement les structures contenues dans des bases de données mais aussi celles pouvant être proposées de façon computationnelle, est actuellement l'approche qui se rapproche le plus de l'identification de nouveaux composés [90]. De nombreuses approches tentent de remplacer la recherche au sein de bases spectrales par l'interrogation de bases de structures moléculaires. Les méthodes de fragmentation combinatoire et les arbres de fragmentation génèrent une "hypothèse" sur la façon dont s'effectue la fragmentation d'un composé. Prédire le spectre de masse à partir d'une structure moléculaire est puissant et ne nécessite pas de spectres expérimentaux au départ du processus. La prédiction d'empreintes moléculaires est, selon Hufsky & Böcker [90], ce qui actuellement conduit aux meilleures performances. Ces différentes stratégies basées sur des méthodes informatiques, modélisées et s'appuyant sur des techniques analytiques pointues sont de plus en plus utilisées, présentent des avantages certains sans pour autant résoudre tous les problèmes rencontrés. Elles nécessitent d'être encore éprouvées, testées et sont toujours en évolution notamment grâce aux travaux collaboratifs d'équipes multidisciplinaires qui cherchent à partager les techniques, les programmes pour aller plus loin dans l'exploration des données et relever les défis qui restent non résolus quant à l'identification formelle de façon plus rapide de composés naturels.

4.3.3 e) *Bilan sur les stratégies existantes*

Le phytochimiste a à sa disposition un large panel de méthodes analytiques pour décrire les produits naturels contenus dans un extrait. Il existe des bases de données de produits naturels, des bases de spectres (MS ou RMN), des outils de prédiction de spectres et des programmes de gestion des données. En s'inspirant de ce qui existe en termes de traitement des données LC-MS, nous avons travaillé à la mise en place de notre propre stratégie pour qu'elle réponde à nos objectifs de profilage métabolique et soit la plus simple possible en termes d'utilisation. Un élément important concerne l'accessibilité de l'outil pour un chimiste analytique sans connaissances particulières en informatique (programmation). Les outils d'annotations des données de LC-MS et ceux de prédiction *in silico* de spectres MS/MS sont actuellement utilisés dans les démarches de déréplication et de mises en évidence de nouveaux composés.

5. Conclusion

Les analyses phytochimiques sont au cœur des projets de recherche et développement ciblant des applications pour les produits de soins, de cosmétique voire de santé. Elles font appel aux techniques de laboratoire pour analyser le contenu des matières premières que sont les plantes.

La recherche est basée sur des méthodes classiques et modernes. Les instruments et techniques analytiques évoluent et se perfectionnent. Il est important de pouvoir en bénéficier pour rendre les travaux plus performants. Les méthodes générales de phytochimie sont celles du laboratoire de chimie analytique. L'usage de la spectrométrie de masse et de la résonance magnétique nucléaire reste majeur et constant puisque les deux techniques fournissent des informations structurales fines sur les composés. Ces méthodes se perfectionnent constamment notamment en termes d'instrumentations mais aussi par rapport aux stratégies d'études mises en place. L'essor des techniques liées à l'usage de moyens informatiques change aussi le rapport des chercheurs face aux problématiques qu'ils ont à résoudre. La tendance générale depuis quelques années est d'intégrer ces outils computationnels dans des approches plus globales d'identification des composés qui nous mènent vers le profilage métabolique ou les études en métabolomique. En effet, la rapidité de calculs, la quantité de données traitées et la puissance d'algorithmes ou logiciels dégage du temps pour le phytochimiste et simplifie certaines étapes. Toutefois, cela lui demande d'élargir ses compétences et de réfléchir encore plus en amont au processus complet d'analyse qu'il doit mettre en place.

Parmi les perspectives pour améliorer le traitement des données et relever le défi de la caractérisation phytochimique, l'utilisation d'algorithmes d'extension de bases de données apparaît comme une voie prometteuse. En effet, dans le cas des composés non identifiés et donc potentiellement nouveaux dans l'espace chimique connu, la recherche de solutions dans l'espace chimique non connu est une nouvelle forme de réponse. Cela pourrait aussi aider à l'identification dans le cas où la purification du composé ne serait pas possible (présence en trop faible quantité dans l'échantillon par exemple).

Globalement, les grandes étapes à suivre pour décrire un ou des échantillons de produits naturels sont similaires quel que soit le procédé choisi : préparer et analyser l'échantillon, prétraiter et traiter les données, visualiser et interpréter les résultats, compléter l'étude avec un nouveau cycle d'analyses complémentaires. Ensuite, ce sont les spécificités de l'étude, du matériel choisi et du résultat recherché qui guident le travail.

L'enjeu pour effectuer le profilage métabolique d'un extrait est de choisir les méthodes appropriées et informatives qui permettent une étude approfondie la plus rapide possible. Il est donc nécessaire qu'une stratégie soit mise en place, notamment du fait que de nombreuses données sont générées et qu'elles doivent être traitées pour être informatives.

Dans cette optique, nous avons choisi de développer un outil (GAINS) qui soit adapté à notre travail de caractérisation, aux techniques et appareils disponibles au sein du laboratoire Greenpharma et de la plateforme HRMS de l'ICOA. De plus, nous l'avons axé autour de la base de données de produits naturels de Greenpharma (GPDB) et ce qu'elle pouvait directement nous apporter (proposition de candidats suite à des requêtes de différents types). Nous avons la possibilité de modifier et d'adapter l'outil à notre convenance en fonction des travaux à réaliser et en intégrant les modules existants qui nous sont utiles. Ce travail a été initié en 2015. Il a été réalisé en parallèle et de manière simultanée au travail de nombreuses équipes qui se sont penchées sur ces problématiques pour élaborer des solutions de traitements et d'annotations des données en vue d'identifier les composés contenus dans des extraits naturels.

Le développement de l'outil GAINS a aussi été nécessaire face au nombre important d'outil développés dans des langages de programmation non maîtrisés en interne et cela permet également de s'affranchir des limites liées au non-maintien de certains outils dans le temps. L'emploi de réseaux moléculaires proposés par la plateforme GNPS constitue un outil très plébiscité actuellement, Annoter les nœuds du réseau n'est pas directement proposée par la plateforme même si cela est possible à l'image de la démarche proposée par Allard en 2016 [36]. Toutefois, la démarche mise en place fait appel à une base de données de spectres MS/MS *in silico*. Par rapport à l'approche des réseaux moléculaires, l'apport de GAINS est la possibilité d'annoter les données directement avec des candidats puisque l'outil est un système intégré et connecté à la base de données GPDB.

Dans les chapitres suivants, la stratégie que nous avons choisie et mise en place va être présentée de façon théorique et pratique. Nous avons élaboré une approche qui s'est inspirée ou a intégré certains des outils existants et libres. Le prétraitement des données MS est fait sous *DataAnalysis,* un logiciel Bruker disponible à la plateforme HRMS de l'ICOA, l'annotation des données MS est réalisée grâce à un algorithme, développé en interne, dont la démarche est proche de celle de CAMERA et enfin l'exploration des données de MS/MS utilise Metfrag, l'outil de prédiction de MS/MS historique proposé en 2010 et mis à jour en 2016. De plus, l'outil est articulé autour de la base de données de produits naturels de Greenpharma (GPDB).

L'outil mis en place va être présenté d'abord d'une façon théorique et méthodologique puis aux travers d'exemples d'utilisation.

Chapitre 2

Développement de **GAINS** :

*(**G**reenpharma **A**utomated **I**dentification of **N**atural **S**tructures)*

Approche et Mise en œuvre

1. Introduction

GAINS (*Greenpharma Automated Identification of Natural Structures*) est un outil de profilage métabolique d'extraits végétaux. L'objectif du développement de cet outil est de tendre vers une description la plus exhaustive possible du contenu d'extraits de plante grâce à l'utilisation de modules automatiques efficaces et de techniques adaptées et précises. La méthode analytique sélectionnée pour mener nos travaux est la chromatographie liquide couplée à la spectrométrie de masse haute résolution (LC-HRMS) et en mode tandem (MS/MS). Face au défi majeur que représente toujours la caractérisation des composés d'un extrait, GAINS est une réponse possible que nous avons conçue, développée et appliquée. Le développement a été effectué pour que l'outil soit performant pour un usage général de description des métabolites secondaires d'un extrait avec des méthodes génériques. Ce chapitre commencera par décrire les grandes étapes mises en jeu dans le processus de l'outil GAINS. Ensuite, nous aborderons comment l'outil a été développé et testé pour être ensuite validé. Les limites de l'outil seront discutées en fin de chapitre. La présentation d'exemples concrets de caractérisation phytochimique de différents extraits végétaux grâce à l'outil GAINS validé fera l'objet du chapitre suivant.

2. Approche globale de GAINS

GAINS est un processus d'analyse et d'interprétation de la composition chimique d'extraits de plante avec deux grandes composantes que sont la **collecte de données analytiques** et le **traitement de ces données**. Le processus de GAINS est schématisé ci-dessous (Figure 8). Chaque étape est représentée par une icône qui est elle-même annotée par le résultat auquel elle donne accès.

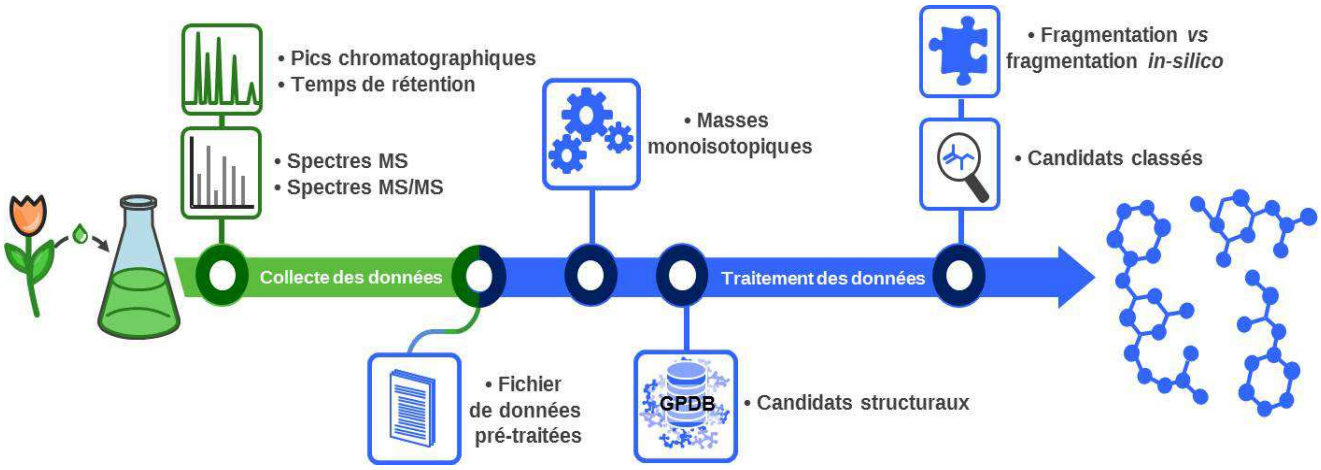


Figure 8 Approche globale de GAINS : étapes mises en jeu

A partir d'un extrait de plante, des analyses en LC-HRMS permettent de collecter les données analytiques : des spectres de masse sont acquis en continu sur toute la durée de l'analyse. A partir de ces spectres, le logiciel d'acquisition permet de reconstruire l'allure du chromatogramme et de visualiser des pics chromatographiques caractérisés par un temps de rétention (t_R) et des spectres de masse associés. Ces données (t_R et spectres de masse) sont prétraitées avant d'être compilées dans un fichier unique. A partir de ce fichier, le traitement des données commence par des calculs pour obtenir une liste de masses monoisotopiques. Des requêtes dans une base de données appelée GPDB (*GreenPharma DataBase* –décrite dans la suite du chapitre) sont alors réalisées pour relier les masses à des structures possibles de composés naturels. Les données de fragmentation (MS/MS) sont ensuite explorées pour classer les différents candidats entre eux. En fin de processus, le profil de l'extrait est caractérisé par une empreinte chromatographique pour laquelle chaque pic chromatographique est associé à une masse monoisotopique. Dans le cas où cette valeur correspond à la masse de molécules naturelles connues, des propositions de structures sont effectuées.

2.1 Analyses en LC-HRMS

2.1.1. Principe et techniques mis en jeu

Pour commencer le travail d'étude du contenu phytochimique, il est nécessaire de collecter des données analytiques sur lesquelles l'exploration se basera. La technique analytique mise en jeu dans notre approche est la chromatographie liquide couplée à la spectrométrie de masse haute résolution (LC-HRMS). Pour gérer la dimension séparative des molécules présentes dans un échantillon, la priorité est donnée à l'utilisation de la chromatographie liquide à ultra haute performance (UHPLC) pour ses capacités résolutives supérieures à celles de la chromatographie liquide haute performance (HPLC).

Pour détecter les molécules, l'échantillon est ionisé dans une source électrospray (ESI) car c'est une source d'ionisation dite douce qui génère peu de fragmentation en source, ce qui facilite l'identification des composés. De plus, c'est la source qui permet d'ioniser la plus large gamme de molécules que ce soit en termes de taille ou de polarité. Après ionisation, l'échantillon est analysé dans un quadripôle à temps de vol (Q-TOF) doté d'un réflectron. Cet ensemble permet une haute résolution massique, une grande exactitude en masse et permet de travailler aussi bien en MS qu'en MS/MS.

Dans le processus, les spectres de masse (MS) et les données de fragmentation (MS/MS) sont systématiquement enregistrés grâce à une acquisition en mode *AutoMS/MS*. Ce mode sélectionne automatiquement l'ion le plus intense du spectre MS pour qu'il soit l'ion précurseur, c'est-à-dire celui

fragmenté afin d'obtenir le spectre MS/MS. Des paramètres d'acquisition sont définis : plage de masses à suivre et à transmettre au détecteur, durée des cycles d'acquisition en MS et en MS/MS, fréquence des cycles d'acquisition en MS par rapport à ceux de MS/MS, et énergie de fragmentation pour la MS/MS.

2.1.2. Modes d'ionisation

L'échantillon est systématiquement analysé en modes d'ionisation positif et négatif pour maximiser le nombre de métabolites détectés puisque les deux modes peuvent être complémentaires [26]. On sait que certaines familles de molécules répondent plus ou moins selon le mode d'ionisation voire ne répondent que dans un seul mode. Pour cette raison, on ne peut pas prendre le parti d'analyser l'extrait dans un seul mode même si cela résulterait en une quantité moins importante de données à traiter. L'idée est d'obtenir le maximum d'informations sans à priori sur le contenu moléculaire d'un extrait en tirant parti de la complémentarité des deux polarités d'ionisation. Le système d'analyse doit être suffisamment générique pour pouvoir passer aux étapes suivantes du processus quel que soit l'état de connaissances sur l'extrait avant analyse. De plus, pour les composés qui s'ionisent dans les deux modes, une plus grande confiance est attribuée à la détermination de la masse monoisotopique.

2.2 Prétraitement des données brutes

2.2.1. Applications de fonctions de calibration et de bruit de fond spectral

Après analyses, les données brutes sont prétraitées grâce au logiciel *Data Analysis* (Bruker) dans sa version 4.4. En effet, il faut effectuer plusieurs opérations qui sont détaillées ci-après et ce pour chacune des analyses effectuées pour un extrait : deux analyses en positif et deux en négatif : celle de l'extrait et celle du solvant d'injection qui constitue le blanc d'analyse. Ces blancs d'analyse permettent de vérifier que les composés détectés sont bien présents dans l'extrait et ne sont pas dus à d'éventuelles pollutions du système analytique utilisé.

L'application des *lock mass* ou ''masses verrous'' permet de corriger l'étalonnage du tube de vol. Cela passe par un réajustement d'éventuelles dérives de la précision en masses au cours de l'analyse. Les composés étalons sont infusés en continu dans le système : l'appareil enregistre leur rapport *m/z* expérimental. La fonction de calibrage appelée *Apply Lock Mass* de *Data Analysis* vérifie la différence avec le rapport *m/z* théorique du produit afin de corriger l'ensemble des masses enregistrées qui pourraient être déviées. Cette fonction s'applique dans les deux modes d'ionisation.

Il est également nécessaire de retirer tous les ions qui sont lents à monter en intensité ou qui sont constants au cours de l'analyse. Ainsi on exécute un algorithme de soustraction du bruit au niveau du chromatogramme, basé sur l'observation d'une montée progressive des contaminants avec le gradient, appelé *Spectral background* et disponible dans *Data Analysis*.

2.2.2. Génération des pics et des listes de masses des ions détectés

Après les étapes de prétraitement générique, on peut procéder à des étapes de prétraitement plus spécifiques. Il s'agit de repérer les composés potentiellement présents dans l'extrait analysé en utilisant l'algorithme *DISSECT*.

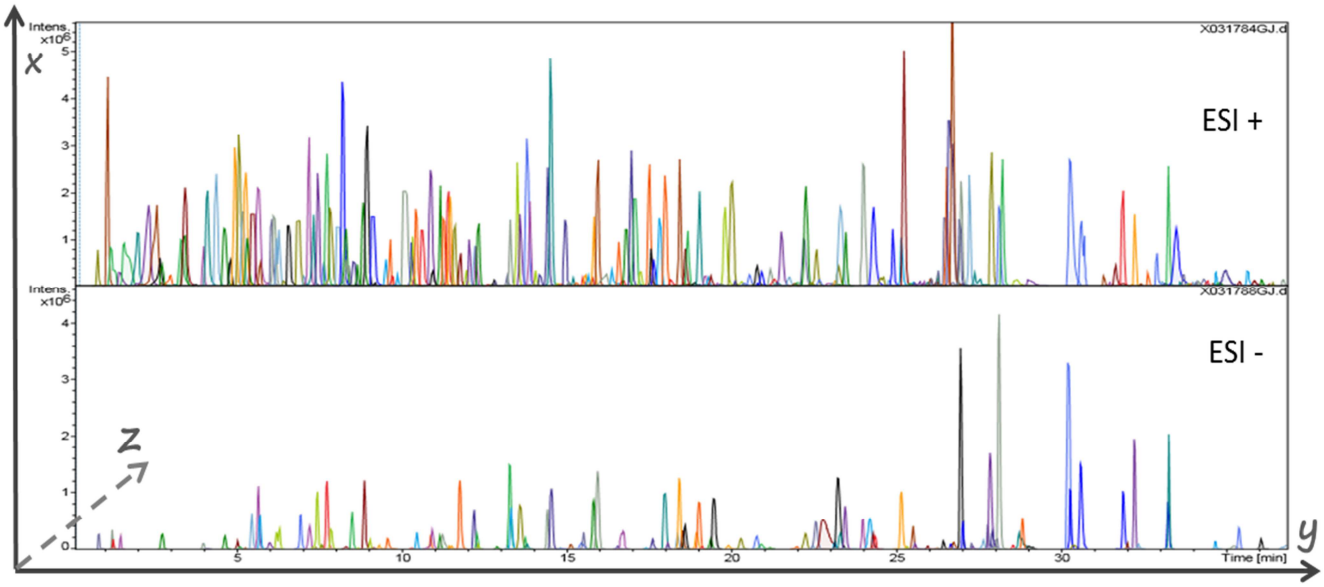


Figure 9 Exemple de chromatogrammes LC-MS en mode positif et mode négatif reconstruits avec l'algorithme DISSECT sous *Data Analysis* (axe x : intensité, axe y : temps de rétention, axe z : *m/z)*

L'algorithme DISSECT permet de repérer dans les spectres acquis, les ions qui sont apparentés et issus de la même molécule. Il indique les différents composés sous la forme de pics de couleur distincte et ce, même si ces composés sont pratiquement co-élués dans un même pic chromatographique. L'algorithme apporte de la sélectivité. L'algorithme est basé sur le principe que tous les ions qui ont une intensité maximale au même moment appartiennent au même composé. Chaque composé repéré résulte donc de la mise en commun de tous les ions devant correspondre à la même molécule mais détectée sous différentes formes ionisées (ions pseudo-moléculaires, ions adduits, ions fragments). *DISSECT* gère la liste des masses des ions de chaque composé selon les réglages d'un second algorithme. La génération des listes de masses des ions d'un spectre moyen d'un composé s'effectue avec l'algorithme SNAP (*Sophisticated Numerical Annotation Procedure*). C'est un algorithme de *deisotoping*, il repère le massif isotopique de chaque ion et indique une seule

valeur d'isotope par massif : celle correspondant à la masse monoisotopique. On a choisi de l'utiliser pour simplifier les listes de masses, ce qui facilitera la détermination de la masse monoisotopique du composé par la suite.

L'ensemble des informations d'une même analyse est compilé dans un fichier de sortie qui constitue le dernier élément obtenu dans la phase de collecte des données analytiques. On obtient ainsi les fichiers d'export au format XML. Ces fichiers référencent chaque composé caractérisé par un temps de rétention et par la liste de masses de ses ions (rapport *m/z* et charge *z*) de niveaux MS et MS/MS.

2.3 Calculs des masses monoisotopiques

2.3.1. Principe et paramétrages techniques des calculs automatiques

Le traitement des données collectées et prétraitées se poursuit avec le calcul des masses monoisotopiques associées à chaque pic chromatographique généré à l'étape précédente. Après avoir chargé les données prétraitées de chaque mode d'ionisation dans le module de calculs de GAINS, le croisement des données entre les modes positif et négatif est effectué. En effet, l'enregistrement des chromatogrammes de façon indépendante dans les deux polarités implique deux considérations. D'une part le nombre total de pics n'est pas forcément équivalent dans chacun des deux modes. D'autre part, les pics n'ont pas exactement le même temps de rétention puisqu'on fait coïncider des pics de deux analyses distinctes. Expérimentalement, il est observé un léger décalage de temps de rétention entre deux analyses consécutives d'un même échantillon. Ainsi, un alignement des chromatogrammes reconstruits est prévu en considérant une fenêtre de temps que nous avons paramétrée à 5 secondes. La valeur que nous avons choisie de fixer est basée sur des décalages observés lors de différentes analyses d'extraits effectuées en LC-HRMS et elle a été margée pour ne pas être trop restrictive. Combiner les données issues des deux polarités d'ionisation augmente la capacité à décrire les composés du fait de la complémentarité des modes. En effet, certaines molécules ne sont détectées que dans un seul mode d'ionisation tandis que de nombreuses autres le sont dans les deux modes.

Ensuite, le système va pouvoir chercher les relations entre les ions détectés dans les deux polarités d'ionisation pour chaque composé. A l'issue des calculs, deux pics chromatographiques de polarité opposée et élués à des temps de rétention différents de moins de 5 secondes pourront alors être comparés. S'il y a correspondance d'ions, un composé est alors défini. S'il n'y pas d'ions en correspondance entre les deux polarités (deux analyses distinctes), le système considérera que ces

deux pics chromatographiques ne sont pas liés. Deux composés différents sont alors élués au même temps de rétention.

Pour démarrer les calculs, l'opérateur doit paramétrer l'outil. Il peut définir la valeur de tolérance d'erreur des calculs de masse -par défaut à 5 ppm-. Il sélectionne également les adduits, clusters, pertes de neutres possibles en fonction des solvants utilisés pour l'analyse. Les différentes espèces ioniques pouvant être sélectionnées sont mentionnées dans le Tableau 6. Elles ont été intégrées au système de calculs d'après leur description dans la littérature et résumées entre autres dans la revue de Holčapek [28]. Ce sont des formes répandues et les différences de masse qu'elles induisent sur la molécule neutre sont bien décrites. Les différences de masse induites par chaque adduit, *cluster* ou perte de neutres ont été calculées d'après les masses des isotopes les plus abondants de chaque élément chimique et sont disponibles en Annexe 2.1.

Types	Mode positif	Mode négatif
Ions pseudo-moléculaires	H^+	$-H^+$
Adduits	Na^+ ; NH_4^+ ; K^+; Li^+	$HCOO^-$; Cl^- ; (Na^+-2H^+)
Clusters	ACN ; MeOH ; HCOOH ; H_2O	
Pertes de neutres	H_2O ; $2H_2O$; CO_2	CO_2

Tableau 6 : Types d'ions pouvant être sélectionnés pour permettre l'annotation des ions des spectres traités sous GAINS

Très souvent, en électrospray notamment, une même entité chimique est détectée sous plusieurs formes d'espèces ionisées au sein du même spectre de masse. Les spectres de masses peuvent présenter des adduits, des dimères, des ions multichargés ou encore des fragments en source [73]. Comme l'évoque la revue de Werner *et al.* [21], c'est cette multitude et redondance d'informations qui rend le travail compliqué et long car il n'est pas évident de repérer à quelle espèce ionisée correspond chaque ion. Néanmoins, croiser ces informations est aussi le moyen d'annoter ces ions et ceci est nécessaire pour connaitre la masse du composé détecté. C'est à cette étape cruciale que le système automatique est précieux pour annoter les ions afin de retrouver la masse neutre du composé détecté et ce de façon rapide et efficace par rapport à un travail manuel.

2.3.2. Processus de recherche et de calcul des masses monoisotopiques

Notre système fonctionne à l'image de la stratégie employée dans CAMERA [75], outil d'annotations de jeux de données obtenues en LC-MS. Des similarités sont observées avec FindMAIN [73] : une approche d'interprétation des spectres ESI-MS décrite en 2017 (*cf*. description au chapitre 1). Pour chaque ion du spectre d'un composé, GAINS fait l'hypothèse de l'ensemble des formes

ionisées possibles et suppose que les molécules peuvent être sous forme de monomères ou de dimères. Il calcule la masse neutre correspondante pour chaque ion. Il compare ensuite les listes de masses neutres possibles pour tous les ions. Lorsqu'au moins deux ions présentent des masses neutres égales avec une tolérance de différence de moins de 5 ppm (valeur paramétrée par défaut), l'outil peut proposer une masse qui caractérise le composé. Tous les ions avec la même masse neutre sont combinés en une hypothèse de masse monoisotopique moyenne. Tous les ions dont la masse monoisotopique neutre est équivalente sont ainsi reliés. GAINS enregistre les informations qui ont permis de faire l'hypothèse de masse monoisotopique en précisant la nature des ions détectés et associés entre eux et la valeur de leur rapport *m/z*. Pour chaque pic, GAINS obtient plusieurs hypothèses de masses monoisotopiques possibles du fait de multiples correspondances d'ions (*cf*. Figure 10). Chaque hypothèse est associée à un score.

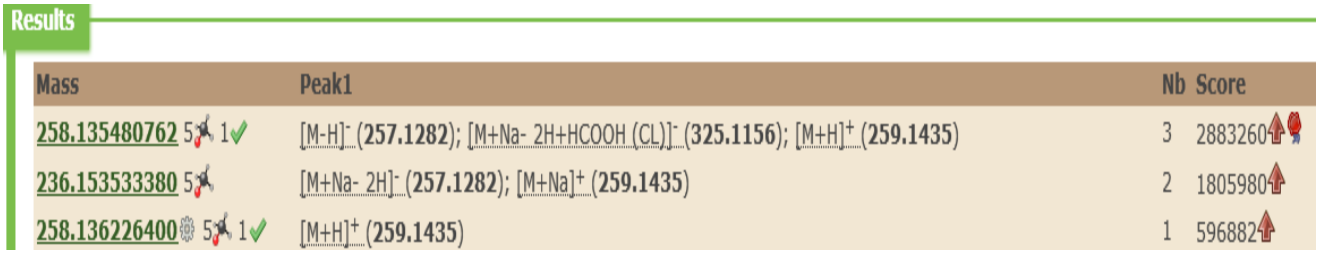


Figure 10 Exemple du visuel des résultats de calculs des masses monoisotopiques effectués au sein de l'interface GAINS : la molécule analysée à une masse théorique de 258,13683 g.mol^{-1}. La première ligne montre l'hypothèse moyenne proposée d'après la correspondance de 3 ions distincts issus des deux polarités.

Dans l'exemple, GAINS propose trois masses monoisotopiques possibles. La première masse (258.1355 g.mol^{-1}) résulte de la correspondance de trois ions dont deux négatifs et un positif. Les ions de *m/z* 257 et 259 sont attribués aux ions pseudo-moléculaires. La seconde masse monoisotopique (236.1535 g.mol^{-1}) proposée fait intervenir un ion positif et un ion négatif mais ici, les ions de *m/z* 257 et 259 sont attribués différemment par l'outil qui propose des adduits sodés. La masse neutre résultante (236 g.mol^{-1}) pour une telle attribution des ions arrive en seconde proposition car il n'y a que 2 ions pour l'expliquer contre 3 pour la première hypothèse et que les ions attribués sont moins probables. Enfin, la troisième proposition est générée automatiquement par le système sur l'ion le plus intense du spectre (*m/z* 259,1435) en l'attribuant à un ion pseudo-moléculaire.

Le système génère automatiquement des hypothèses sur l'ion le plus intense du spectre. Ainsi, en mode positif, quatre hypothèses sont faites considérant que l'ion peut correspondre à la forme $[M+H]^+$, $[M+Na]^+$, $[M+NH_4]^+$ ou $[M]^+$ (molécule avec charge permanente). Dans le mode négatif, l'ion d'intensité maximale est considéré comme pouvant être $[M-H]^-$, $[M+HCOO]^-$. Ces hypothèses automatiques permettent de gérer deux situations. La première est celle où aucun des

ions du spectre ne correspondent entre eux, alors des hypothèses de masses sont quand même effectuées. GAINS utilise cet heuristique qui donne de l'importance à l'ion majoritaire en lui attribuant les formes les plus fréquentes d'après la revue d'Holčapek [28]. Cela dit, cet heuristique n'est pas pertinent dans toutes les situations, notamment si la molécule est largement fragmentée en source. Ces hypothèses automatiques permettent de gérer une seconde situation : l'opérateur intervient pour déterminer et sélectionner la vraie hypothèse s'il considère qu'elle est liée à l'ion majoritaire et non à la correspondance d'ions qui seraient d'intensités plus faibles et dans lequel l'ion majoritaire ne serait pas inclus. Les opérations d'annotation et de calculs d'hypothèses de masses sont répétées pour chaque composé de l'extrait.

2.3.3. Application d'une fonction de score

Pour un composé donné, les multiples masses monoisotopiques sont ensuite classées selon l'équation de score suivante :

$$\text{score (MW moy.)} = \frac{\sum_i (I \times W)^2}{\sum_i I} \times n \times \text{Opp} \quad \text{(Équation 1)}$$

Avec **MW moy.** : masse monoisotopique moyenne obtenue par correspondance de n ions i ;

I : Intensité de l'ion i ;

W : Poids statistique de l'espèce ionisée attribué à l'ion i. (Si l'espèce est dichargée sous forme : adduit avec cluster ou adduit avec perte de neutre ; $W = W_1 \times W_2$ où W_1 est le poids de l'adduit 1, et W_2 est le poids du second adduit, du cluster ou de la perte de neutre) ;

n : Nombre d'ions i correspondants pour une même hypothèse de masse monoisotopique ;

Opp : Facteur prenant en compte la correspondance entre ions issus d'un seul mode d'ionisation ou des deux modes d'ionisation.

Cette fonction de score permet de prendre en compte différents facteurs importants. Son établissement a eu pour but premier de générer un classement dans le cas général. L'équation 1 donne donc de l'importance à l'intensité des ions engagés pour établir l'hypothèse de masse, au type et nombre de formes ionisées permettant de générer cette même hypothèse ainsi qu'à la corrélation entre des ions des modes d'ionisation positif et négatif. Ces paramètres expérimentaux ont été choisis pour établir la fonction de score du fait de leur caractère fondamental en spectrométrie de masse. Un rapport *m/z* d'intensité très élevée est représentatif de l'abondance importante de l'ion correspondant détecté pour un instant donné. Aussi, toute concordance entre espèces ioniques est plus certaine si le nombre d'ions impliqués dans la concordance est important. Le type de l'espèce ionique considérée est également primordial : il est plus probable de détecter les molécules sous la

forme d'adduits que de *clusters* par exemple, ainsi le poids statistique pour un adduit est plus grand que pour un *cluster*. Enfin, on donne plus de crédit à une correspondance entre des ions issus des deux modes d'ionisation opposés $(\mathrm{Opp} = 2)$ par rapport à une correspondance intra-mode $(\mathrm{Opp} = 1)$. En effet, cela permet de donner plus de confiance à des hypothèses provenant de deux analyses différentes par rapport à des hypothèses basées sur des données d'un seul mode d'ionisation.

La résultante de cette étape est une liste de masses monoisotopiques classées selon un score établi en prenant en compte des éléments analytiques pouvant indiquer la probabilité d'une masse donnée (intensité des ions et notamment de l'ion pseudo-moléculaire s'il est détecté, correspondance intra ou inter-mode d'ionisation). Ce score permet d'avoir une hypothèse de masse classée numéro 1 pour laquelle le niveau de confiance et de probabilité est le plus élevé et basé sur les données expérimentales.

2.4 Requête au sein de la Greenpharma Database (GPDB)

2.4.1. GPDB, une base de données dédiée aux produits naturels

Pour attribuer des candidats structuraux à chaque masse monoisotopique calculée, on lance des requêtes au sein de la *Greenpharma Database*. C'est une base de données spécifique des produits naturels, riche d'environ 160 000 molécules qui sont reliées aux organismes dans lesquels elles sont présentes ainsi qu'à leurs propriétés biologiques. Chaque molécule naturelle possède une fiche regroupant un ensemble d'informations telles que la structure, le nom, les synonymes, le numéro CAS, le code SMILES, la formule brute, la masse moléculaire et la masse monoisotopique, la solubilité, le log de P, la famille phytochimique. Pour les organismes reliés à la molécule, des indications telles que l'organe et la concentration à laquelle on peut s'attendre peuvent aussi être renseignées. Les composés référencés sont très majoritairement des métabolites secondaires des plantes incluant les trois grandes classes proposées par Bruneton [95] à savoir celles des composés phénoliques, shikimates et acétates, celles des terpènes et stéroïdes et enfin celles des alcaloïdes. S'agissant des métabolites primaires, on trouve des glucides tels que les oses simples et des oligosaccharides ou des lipides tels que les acides gras. Enfin les acides aminés et peptides sont représentés sous diverses formes : acides aminés non constitutifs des protéines, autres composés azotés (hétérosides cyanogènes) ou soufrés (thiosulfinates par exemple). Les acides aminés non protéiques sont des acides aminés végétaux non impliqués dans la formation de protéines et qui peuvent être apparentés à des métabolites secondaires.

2.4.2. Requête au sein de la base GPDB

L'interrogation de la base est effectuée avec une tolérance relative (en ppm) ou absolue (en g/mol) sur la masse monoisotopique. L'utilisateur peut l'ajuster mais par défaut la tolérance relative est paramétrée à 5 ppm. L'espace interrogé est également ajustable : recherche dans la totalité de la GPDB ou choix de sous-espaces tels que ceux définis par les molécules d'une famille phytochimique particulière, d'une famille ou genre de plantes ou encore par les molécules d'un seul organisme (en général, celui correspondant à l'espèce végétale étudiée). La Figure 11 schématise l'étape de requête dans la base de données qui n'est autre que le référencement de molécules (mol A, mol B, mol C …) caractérisées, entre autres, par une masse monoisotopique (m_a, m_b …).

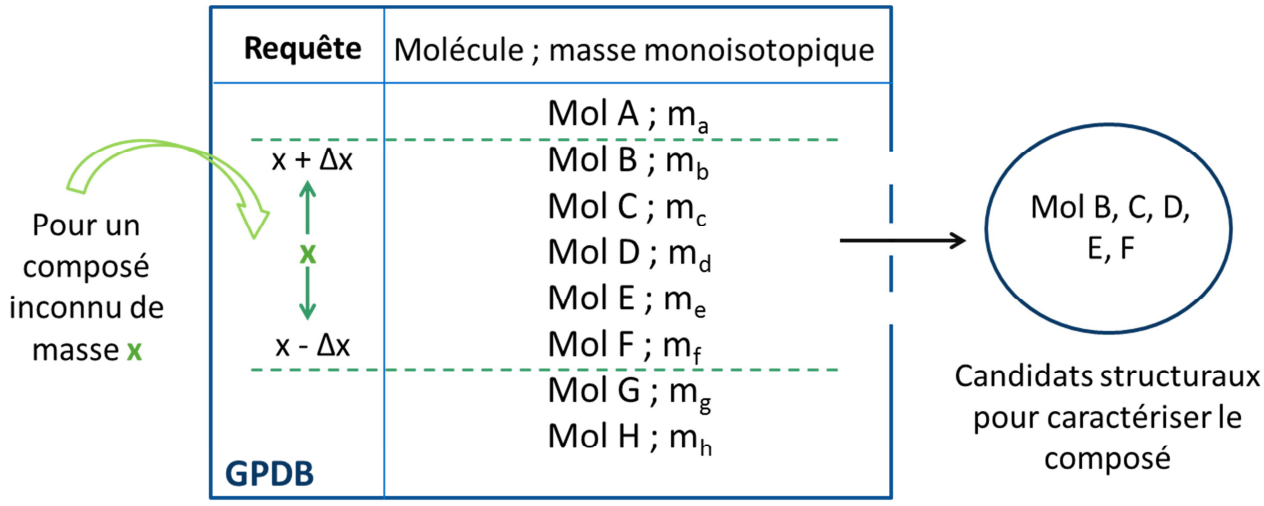


Figure 11 Schéma de la méthode employée pour effectuer une requête dans la base de données afin de trouver les candidats structuraux pour caractériser une molécule inconnue ; x la valeur de masse monoisotopique moyenne expérimentale dotée du meilleur score et issue de l'annotation des ions du spectre et Δx la tolérance en masse pour effectuer la requête.

La recherche de structures est effectuée pour chaque masse monoisotopique classée en première position de chaque pic détecté. A l'issue de la requête pour l'ensemble des pics, on obtient une première vue d'ensemble de l'extrait avec pour chaque pic chromatographique, la liste des masses monoisotopiques classées et pour la masse la plus probable, une liste de candidats structuraux possibles.

A ce stade, plusieurs molécules sont proposées pour une masse monoisotopique donnée. Pour affiner les résultats, on fait appel aux informations complémentaires que constituent les données de MS/MS.

2.5 Exploration des données de MS/MS

2.5.1. Analyse des données de fragmentation (MS/MS) : principe

Pour distinguer les candidats structuraux proposés pour chaque pic, les données de fragmentation sont précieuses et utiles puisqu'elles fournissent des indications structurales supplémentaires par rapport à la MS et constituent une sorte d'empreinte de la molécule [57,80]. En effet, en MS/MS, un ion tend à se fragmenter au niveau de certaines liaisons pour donner des fragments de molécule. Ces fragmentations ne sont pas aléatoires et permettent de repérer la présence de certains groupements dans la molécule. Pour explorer ces informations structurales, deux techniques sont utilisées dans l'approche GAINS : une approche de fragmentation *in silico* versus celle de fragmentation expérimentale obtenue en MS/MS et une approche de vérification des règles de fragmentation décrites dans la littérature avec une gestion plus manuelle.

2.5.2. Utilisation d'un outil automatique de fragmentation *in silico* : Metfrag

La première approche consiste à utiliser un module de fragmentation *in silico* qui propose un classement des candidats moléculaires en fonction de la similarité du spectre expérimental de fragmentation avec le spectre MS/MS modélisé par traitement informatique.

C'est l'outil libre et *open-source* METFRAG[*] qui a été testé dans notre processus. Son principe est le suivant : pour chaque composé doté d'une masse monoisotopique déterminée en amont et d'un spectre expérimental de fragmentation, Metfrag modélise le spectre MS/MS des différents candidats structuraux reliés au composé étudié.

L'outil fragmente *in silico* le composé candidat sur le principe de l'énergie de dissociation des liaisons, c'est-à-dire que les ruptures effectuées virtuellement sont en premier celles de plus faible énergie de liaison. Metfrag génère les fragments possibles pour chaque candidat. Le nombre de ruptures de liaisons (*maximum tree depth*) est limité pour réduire les faux-positifs liés à un nombre trop important de fragments générés.

La seconde étape de Metfrag consiste en un classement des candidats en fonction de la ressemblance de leur spectre modélisé avec le spectre expérimental du composé. Les fragments *in silico* sont comparés avec ceux du spectre expérimental de la molécule. L'outil classe les différentes structures selon une fonction de score prenant en compte le nombre d'ions et leur intensité ainsi que l'énergie nécessaire à la rupture de liaison [80]. L'idée avec ce ''fragmenteur'' combinatoire n'est pas de prédire les voies de fragmentation correctement d'un point de vue mécanistique mais plutôt de considérer les fragments comme des descripteurs additionnels des composés analysés.

[*] c-ruttkies.github.io /MetFrag

2.5.3. Gestion des groupements glycosylés ou acylés

Après la gestion automatique des données au sein de GAINS, nous obtenons des hypothèses de masses et de structures. Toutefois, l'exploration des données de MS/MS par le module combinatoire Metfrag ne cible aucun groupement particulier. Or, la présence de glycosylation et/ou acylation sur les molécules naturelles est très fréquente. Ainsi, nous avons voulu intégrer à notre démarche la mise en évidence de ces groupements. Ces recherches étant gérées manuellement, nous avons axé le travail principalement sur la détermination de parties glycosylées ou acylées. Nous avons ciblé la recherche des O-glycosylations et O-acylations qui sont des substitutions davantage répandues par rapport aux C-substitutions dont les voies de fragmentations sont différentes. De plus, la recherche de ces fonctions concerne uniquement les fragments les plus intenses des spectres MS/MS. Le travail est réalisé à l'aide d'un formulaire qui regroupe les valeurs des masses de pertes de neutres ou d'ions à chercher. Le document contient les masses exactes de formes osidiques simples ou multiples et les valeurs des ions correspondants sous formes protonées et sodées pour le mode positif, déprotonées pour le mode négatif. Notons qu'une des perspectives pour améliorer le fonctionnement de l'outil est d'automatiser ce travail.

Par opposition avec les moyens *in silico* mis en œuvre précédemment, c'est une autre façon d'explorer les données de fragmentation qui est abordée ici en s'intéressant à la façon dont l'ion a pu se rompre d'après des règles de fragmentations établies. Pour repérer des fragmentions d'intérêt, deux possibilités s'offrent à l'observateur : soit il cherche les pertes de neutres entre l'ion précurseur et l'ion fragment et évalue si cette perte correspond à des groupements particuliers, soit il se concentre sur les ions que ces groupements peuvent eux-mêmes générer après fragmentation et conservation de la charge. Ces deux possibilités de traitement de données MS/MS vont être décrites dans la suite de ce chapitre.

2.5.3 a) *Recherche de pertes de neutres : cas des molécules O-glycosylées / O-acylées*

On peut calculer les pertes à partir de l'espèce ionique fragmentée et vérifier si cela explique la présence de certains éléments structuraux connus ou visibles sur des candidats (structures proposées). Il est possible de s'appuyer sur de nombreuses études publiées jusqu'à présent au sujet de la glycosylation des composés phénoliques et notamment celles des flavonoïdes qui est encore aujourd'hui largement étudiée. Il existe un outil nommé FlavonQ [96] dédié à l'identification de flavonoïdes et principalement des flavones et flavonols glycosylés, deux sous-classes importantes des flavonoïdes. Le système permet le traitement automatique de données brutes issues d'analyses en

UHPLC-DAD-HRMS et MS^n. Il établit des propositions d'identification des composés flavonoïdiques notamment à l'aide de la combinaison suivante :

$$M_{flavone/flavonol\ glycosylé} = M_{\ aglycone} + a \times M_{\ glycosyl} + b \times M_{\ acyl}$$

L'outil n'a pas pu être utilisé dans notre approche n'étant pas directement destiné à la mise en évidence des groupements glycosylés ou acylés mais à l'identification ciblée des flavonoïdes. Toutefois, nous nous sommes inspirés des oses ou groupements acylés considérés dans les calculs de l'outil FlavonQ.

Les oses simples pris en compte sont parmi les plus fréquemment rencontrés dans les molécules naturelles et sont également pris en compte dans la démarche de FlavonQ. Ainsi, on considère 9 monosaccharides listés dans le Tableau 7. On sait que dans le cas d'une O-glycosylation, la réaction entre une génine et un sucre libère une molécule d'eau. Ainsi, en tant que groupement attaché à une génine, les pertes ou les différences de masses recherchées correspondent donc à la masse monoisotopique d'un ose auquel est retranchée la masse d'une molécule d'eau. De plus, sachant que les enchaînements d'oses peuvent être de type monosaccharidique, disaccharidique et plus rarement trisaccharidique ou tetra-saccharidique [95], le formulaire répertorie les masses correspondantes à ces multiples combinaisons.

Type d'ose	Ose	Formule brute	Masse de la perte de neutre ($g.mol^{-1}$)
Acide uronique	Acide glucuronique	$C_6H_8O_6$	176,03209
Hexose	Glucose Galactose Mannose	$C_6H_{10}O_5$	162,05283
Desoxyhexose	Fucose Rhamnose	$C_6H_{10}O_4$	146,05791
Pentose	Xylose Arabinose Apiose	$C_5H_8O_4$	132,04226

Tableau 7 Oses considérés (type, formule brute et masse monoisotopique) pour l'exploitation des données de fragmentation

Pour les groupements O-acylés, la démarche est identique à celle présentée pour les parties osidiques. Ainsi, les recherches sont axées sur des pertes de masse correspondant à celle du groupement acylé moins celle d'une molécule d'eau. Les groupements acylés considérés avec les

formules brutes et les masses recherchées, sont répertoriés dans le Tableau 8. La liste des groupements est largement basée sur celle proposée dans la démarche de FlavonQ [96].

Type de groupement	Groupement	Formule brute	Masse de la perte de neutre ($g.mol^{-1}$)
Hydroxycinnamoyls	p-Coumaroyl	$C_9H_6O_2$	146,03678
	Caffeoyl	$C_9H_6O_3$	162,03169
	Feruloyl	$C_{10}H_8O_3$	176,04735
	Sinapoyl	$C_{11}H_{10}O_4$	206,05791
Hydroxybenzoyls	p-Hydroxybenzoyl	$C_7H_4O_2$	120,02113
	Galloyl	$C_7H_4O_4$	152,01096
Acides dicarboxyliques acylés	Oxalyl	C_2O_3	71,98475
	Malonyl	$C_3H_2O_3$	86,00039
	Succinoyl	$C_4H_4O_3$	100,01605
	Maloyl	$C_4H_4O_4$	116,0109
Autres aliphatiques ou acylés	Acetyl	C_2H_2O	42,01057
	Propionyl	C_3H_4O	56,02621
	Butyryl	C_4H_6O	70,04187

Tableau 8 Groupements acylés pris en compte pour la recherche de leur présence au sein de structures de composés naturels

2.5.3 b) *Recherche d'ions sodés : intérêt pour l'étude des glycosylations*

Une génine peut porter les oses en une seule position (monosubstitution) ou en plusieurs positions. A cet égard, parmi les formes ionisées sous lesquelles les oses sont recherchés, les adduits sodés sont considérés avec intérêt car leur formation et interprétation permet d'accéder à des informations particulières du fait de leur stabilité notamment. Les sucres sont riches en oxygène et ceci les rend prompts à capter l'ion sodium [97]. A ce titre, la littérature évoque différents types de molécules riches en oxygène et présentant cette aptitude à l'image de saponines [98], de pyranocoumarines [99], de terpénoïdes lactones [100], d'oligosaccharides [101] ou encore de phospholipides [102]. Le cation sodium est présent dans le milieu d'analyse même s'il n'y a pas été ajouté volontairement par l'opérateur. Il peut provenir des solvants utilisés ou de la verrerie par exemple [99]. La résultante –molécule glycosylée sodée- est une forme ionique très stable qui ne se dissocie pas facilement [97,103–105]. Ainsi, lors d'étapes de fragmentation avec l'application d'une énergie de fragmentation suffisante pour provoquer des ruptures, il est fréquent d'observer l'ion correspondant à la partie osidique complexée à un cation sodium. Dans le cas de composés portant au moins deux oses, ceci a un grand intérêt pour aider à décrire la partie osidique de la molécule. On peut notamment déterminer les sucres qui sont sous forme continue ou pas. Dans le premier cas,

cela correspond au fait que les oses s'enchaînent sur le composé selon une monosubstitution des sucres sur la génine, on peut alors observer par exemple l'ion correspondant au dissaccharide sodé tandis que dans le second cas, la génine est polysubstituée et les ions correspondants aux différents monosaccharides sodés ne sont pas forcément observés.

Certaines structures restent malgré tout difficiles à décrire avec uniquement les données de HRMS/MS. Il est impossible de définir l'enchaînement précis des oses entre eux. Par exemple, les données MS/MS ne permettent pas de différencier un enchaînement rutinoside d'un enchaînement neohesperidoside dont la seule différence est la position de substitution du glucose par le rhamnose comme l'illustre la Figure 12.

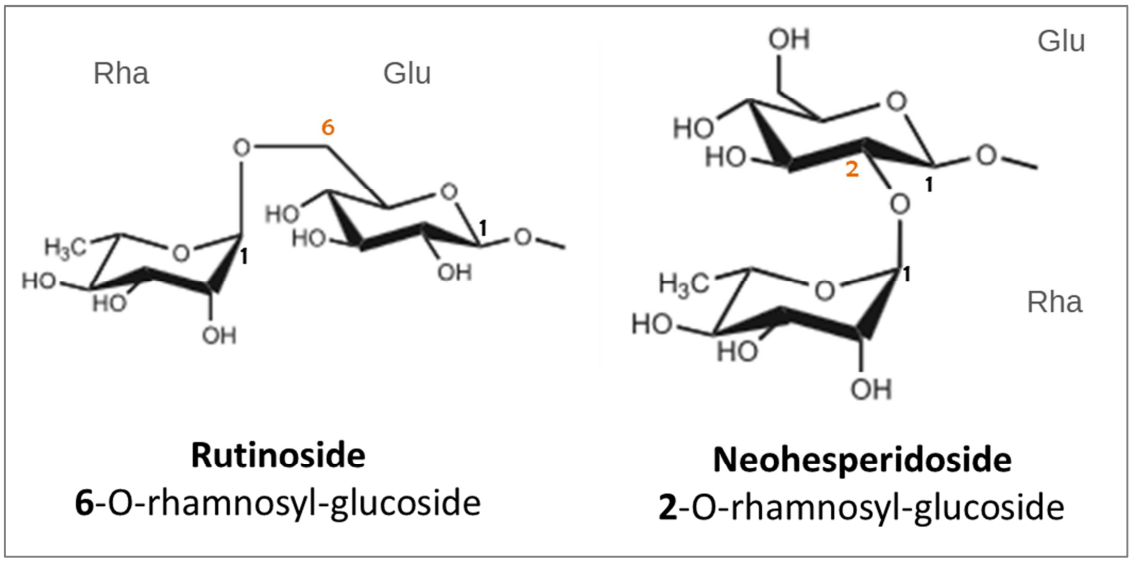


Figure 12 Structure de deux dissaccharides isomères de position constitués par un glucose et un rhamnose: le rutinoside et le neohesperidoside.

Dans le cas de 3 oses ou plus, on ne sait donc pas s'ils sont sous forme linéaire (un sucre substitué par un sucre) ou ramifiée (un sucre substitué par deux sucres). De plus, la position précise de la substitution sur la génine reste une information compliquée à déterminer.

3. Développement de l'outil

3.1 Traduction du travail du phytochimiste en code informatique

3.1.1. Démarche de travail pour développer l'outil

GAINS est une démarche globale qui mime ce que ferait manuellement un phytochimiste face à des données analytiques brutes issus d'un échantillon d'intérêt. En revanche, GAINS apporte une dimension d'automatisation à ce travail d'exploration et en ce sens, l'outil permet une analyse plus complète et plus approfondie des données que celle qui serait effectuée manuellement. De plus, l'emploi de méthodes automatiques (ou semi-automatiques) permet de gagner un temps considérable sur des étapes longues et laborieuses, de limiter les erreurs possibles et d'harmoniser le rendu des résultats. En pratique, GAINS se présente sous la forme d'une interface web, implémentée dans la base de données de Greenpharma (GPDB), qui a été développée au sein de Greenpharma. Pour y parvenir, GAINS est le fruit d'une collaboration entre le spécialiste informatique qui avait déjà développé la GPDB et l'équipe phytochimie de Greenpharma. Sur la base de ce que le phytochimiste effectuait manuellement jusque-là, et après avoir expliqué et traduit les différentes étapes d'analyses et de traitements en actions simples à l'informaticien, ce dernier a pu les transformer en code informatique. J'ai donc explicité chaque étape le plus simplement et fidèlement possible afin qu'elles puissent être comprises par un non-initié aux techniques analytiques mises en jeu. L'élaboration de GAINS résulte ainsi de concertations entre l'expert informatique et l'expert phytochimiste pour permettre compréhension, explications, décisions et validations du modèle. Des vérifications au niveau analytique contrôlaient que les résultats proposés suite à l'automatisation étaient corrects.

3.1.2. Langages de programmation

GAINS est un algorithme pourvu d'une interface qui permet de gérer les actions à mener pour le screening phytochimique, de stocker les résultats analytiques et les traitements effectués pour les consulter autant que nécessaire. Le code informatique est écrit dans le langage php (*personal home page*). C'est un langage de *script* généraliste, libre et gratuit, qui est très largement utilisé pour le développement d'applications web. De plus, il s'interface très bien avec le langage SQL (*structured query language*) qui supporte et organise les bases de données. La base de données GPDB est gérée avec MySQL. C'est un système de gestion de bases de données largement utilisé, simple d'utilisation et qui au moment de la création de GPDB était disponible et le plus classique.

Grâce à l'interface web de la base de données GPDB contenant le module GAINS, l'utilisateur peut aisément traiter, visualiser et gérer les données analytiques.

3.2 Mise en œuvre technique : mise au point de l'outil

3.2.1. Constitution d'un mélange standard

3.2.1 a) *Familles phytochimiques représentées*

Pour mettre au point l'outil et vérifier son fonctionnement, nous avons constitué un mélange de 150 molécules naturelles[†]. Les concentrations des molécules en solution dans le méthanol variaient entre 5 et 10 ppm. L'idée était d'obtenir un mélange avec une diversité importante dans sa composition pour se confronter à une matrice complexe et étudier les capacités de GAINS sans se restreindre à un type particulier de composés naturels. Ainsi, avec les 150 composés, les grandes classes phytochimiques sont représentées dans des proportions différentes, illustrées en Figure 13.

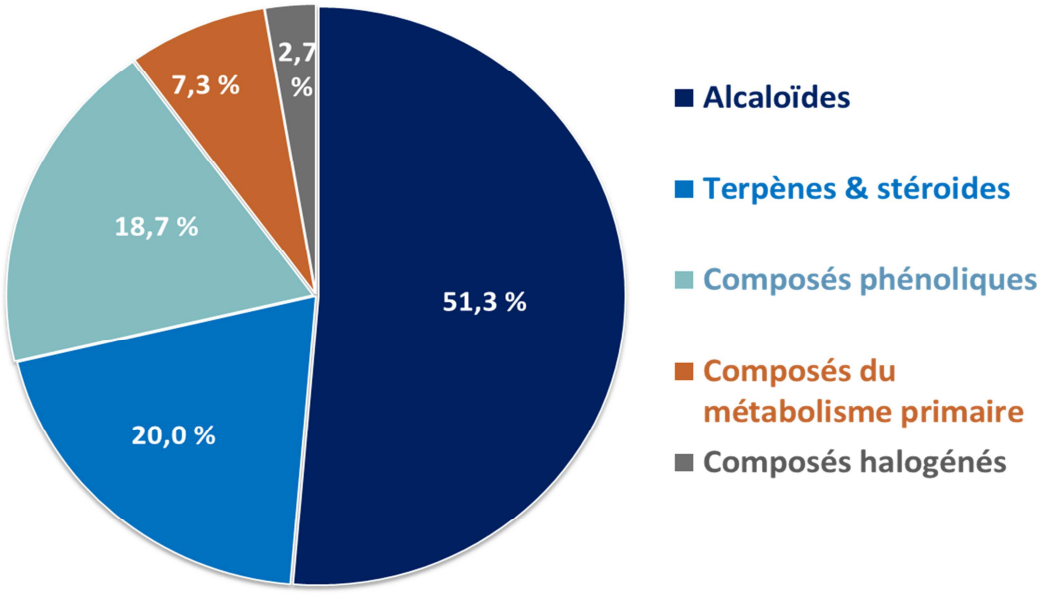


Figure 13 Répartition des composés du mélange test en fonction des grandes classes phytochimiques (données en % sur un total de 150 composés)

Plus de la moitié des composés sont de type alcaloïdique, 20 % environ appartiennent aux terpènes et stéroïdes, 20 % encore sont des composés phénoliques, shikimates et acétates, les 10 % restants se répartissent entre les molécules du métabolisme primaire et les halogénés que nous avons choisi de distinguer des autres composés. La détection des composés halogénés est relativement aisée en spectrométrie de masse. Ils sont relativement rares ; on notera que les

[†] Les structures, noms et classes phytochimiques des 150 molécules sont présentées en Annexe 2.2.1

composés bromés sont plutôt spécifiques du monde marin par exemple. Ces molécules relativement exotiques sont donc réunies dans une même catégorie. La répartition globale est représentative de la proportion de chaque classe au sein de notre base de données qui décrit principalement l'espace chimique des métabolites secondaires. Ceci explique que les composés apparentés au métabolisme primaire soient les moins présents. La forte proportion d'alcaloïdes est cohérente avec le fait que cette classe est l'une des plus diverses et importantes des substances naturelles. Les composés phénoliques regroupent des métabolites issus des voies du shikimate et de l'acétate. Dans le mélange étudié se trouvent des phénols, acides phénoliques, coumarines, xanthones, flavonoïdes, lignanes, tanins ou encore quinones. En termes de répartition dans le règne végétal, les molécules du mélange correspondent à 154 familles de plantes différentes (au moins une molécule présente par famille) et plus de 700 espèces de plantes.

3.2.1 b) *Gamme de masses molaires analysées*

La distribution en masses moléculaires est également un critère intéressant à mentionner. En effet, l'outil est principalement destiné à l'analyse des métabolites secondaires des plantes, c'est-à-dire des molécules de faible masse moléculaire par rapport aux métabolites primaires qui peuvent avoir de très haute masse moléculaire (protéines notamment). La distribution en masses des composés de la base de données Greenpharma et du mélange testé pour l'étude ont été comparées : voir Figure 14.

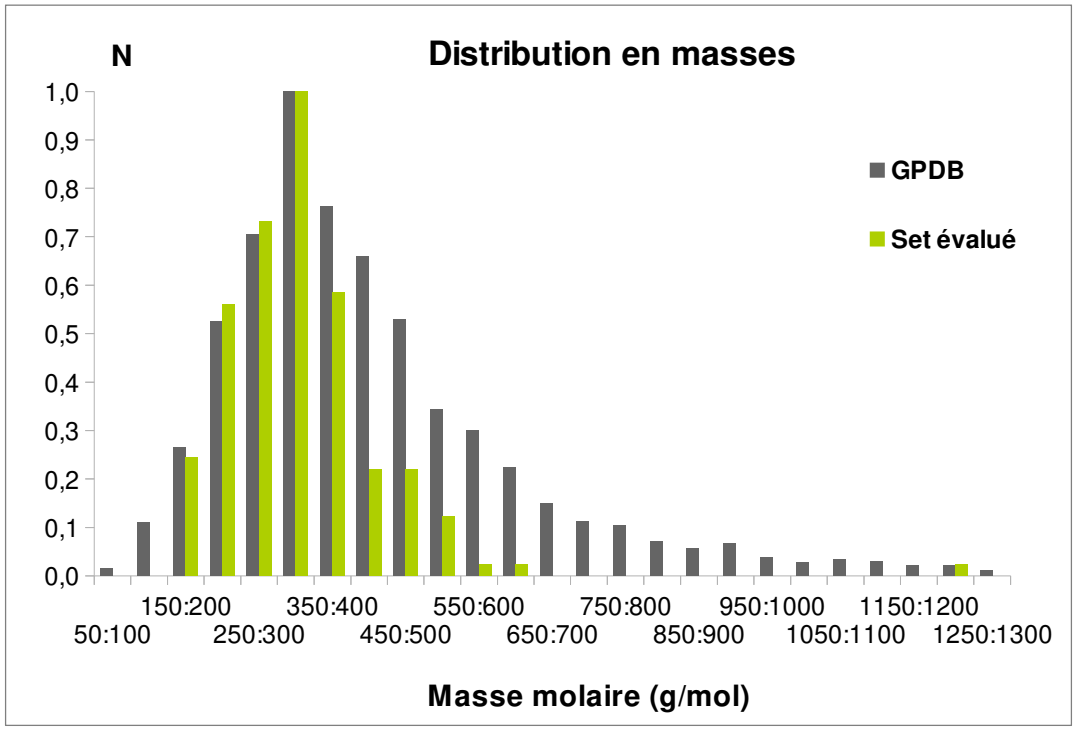


Figure 14 Distribution normalisée du nombre de composés en fonction de leurs masses molaires pour l'ensemble des molécules de GPDB et pour celles du mélange standard évalué

On observe que la gamme de masses molaires des molécules de GPDB s'étend de 50 à 1300 $g.mol^{-1}$ et est centrée sur l'intervalle de [300 ; 350] $g.mol^{-1}$. Ces éléments sont cohérents avec le fait que la base de données est spécifique aux produits naturels de type métabolites secondaires. Le mélange test a une distribution très proche de celle de la base de données complète avec une forte proportion de masses entre 200 et 400 $g.mol^{-1}$. Au-delà de 650 $g.mol^{-1}$, une seule molécule de masse 1240 $g.mol^{-1}$ est représentée. Le mélange test est ainsi représentatif de la distribution des composés naturels que l'on peut être amené à étudier.

3.2.2. Analyses des standards pour valider l'outil

3.2.2 a) *Séquence d'analyses*

Afin de mettre au point et valider l'outil, nous avons procédé aux analyses des molécules standards selon deux phases de validation. La première correspond à l'analyse des 150 molécules injectées séparément via le mode *Flow Injection Analysis* (FIA). La FIA (*Flow Injection Analysis* ou Analyse par injection en flux continu) permet d'injecter un faible volume d'échantillon dans la phase mobile qui est en flux continu. L'échantillon est transporté par la phase mobile vers le détecteur sans passer par une colonne. L'ordre de passage des molécules était défini et connu. L'idée était de collecter les données analytiques de chaque composé et de vérifier à chaque étape, les éléments importants et si les calculs mis en place tout comme le score étaient pertinents. La seconde phase a consisté à injecter les 150 molécules en mélange en couplage UHPLC-HRMS.

3.2.2 b) *Conditions analytiques*

Que ce soit pour la validation en FIA-HRMS ou en couplage LC-HRMS, les molécules ont été analysées en mode d'ionisation positif et en mode d'ionisation négatif avec enregistrement des spectres MS et MS/MS. Le mode d'acquisition des données HRMS employé est le mode *Data Dependant Analysis* (DDA) aussi appelé *AutoMS/MS*. Dans ce mode, les ions envoyés en fragmentation sont forcément les espèces d'intensité maximale du spectre MS : ils sont isolés pour l'analyse en MS/MS [106]. Le détail de ces conditions d'analyse est disponible en Annexe 3.2. Un prétraitement des données avec *Data Analysis* 4.4 et un traitement des données par le module de calculs GAINS ont été appliqués aux données. Les formes d'ions prises en compte sont résumées dans le Tableau 9 :

Adduits, clusters ou pertes de neutres considérés pour les calculs	**Mode positif**	H^+ ; Na^+ ; NH_4^+ ; H_2O ; $2H_2O$
	Mode négatif	$-H^+$; $HCOO^-$; Cl^- ; (Na^+-2H^+)
	Modes positif et négatif	ACN ; HCOOH ; H_2O / CO_2

Tableau 9 Adduits, clusters ou pertes de neutres considérés pour les calculs des masses monoisotopiques

3.2.3. Mise au point : choix et réglages des poids statistiques des adduits

La première phase d'évaluation de GAINS a été menée par l'analyse de l'ensemble de 150 molécules de façon individuelle en mode d'injection FIA. Des ajustements de l'outil dans sa version originale de calculs ont pu être effectués suite à l'observation des premiers résultats. Certains cas n'avaient pas été pris en compte comme certaines formes d'adduits particuliers ou certains calculs étaient mal exécutés pour des cas précis (perte de neutre H_2O par exemple), ou encore, la possibilité que les composés soient des molécules à charge permanente n'avait pas été intégrée.

3.2.3 a) *Détermination du poids statistique des espèces ionisées prise en compte*

En FIA, chaque molécule ayant été injectée séparément et dans un ordre précis, nous savions ce que nous attendions comme résultat pour chaque pic. GAINS a pu effectuer les calculs de masses. Pour les pics où la meilleure hypothèse n'était pas celle attendue, nous avons sélectionné manuellement la bonne hypothèse. Puis, nous avons pu déterminer la fréquence des espèces ionisées. Ce compte nous a permis de leur attribuer un poids. Nous avons pu classer les types d'ions par ordre décroissant en fonction de leur fréquence de formation. Le poids le plus élevé est accordé aux espèces les plus fréquentes tandis qu'un poids plus faible correspond à celles qui sont moins probables. Le poids attribué à chaque type d'ions (ou espèces ioniques) a été enregistré pour le calcul du score décrit précédemment (*cf*. paragraphe 2.3.3 de ce chapitre). Ces poids, listés dans le Tableau 10, ont été fixés pour être cohérent avec une utilisation de l'outil en routine, c'est-à-dire dans les conditions d'utilisation les plus fréquentes.

Mode positif		Mode négatif	
Type d'ions	**Poids**	**Type d'ions**	**Poids**
H^+	**0,95**	$-H^+$	**0,95**
Na^+	**0,90**	$HCOO^-$	**0,90**
NH_4^+	**0,85**	CH_3COO^-	**0,8**
K^+	**0,8**	$Na^+ - 2H^+$	**0,8**
Li^+	**0,7**	Cl^-	**0,7**
Cluster (solvants) ou Perte de Neutre (H_2O ou CO_2) : **0,5**			

Tableau 10 : Poids attribué à chaque type d'ion pour le calcul du score de classement des masses monoisotopiques de GAINS

Dans le cadre d'études particulières, il pourrait être envisagé d'ajouter un ion métallique (sodium, lithium…) aux phases mobiles pour favoriser la formation d'adduits sodés ou lithiés, il faudrait alors revoir ces valeurs en maximisant le poids de l'adduit concerné.

3.3 Résultats de validation : données de MS

Les analyses des 150 standards en FIA-HRMS ont été répétées deux fois de façon indépendante. A l'issue des calculs de masses monoisotopiques, la vérification du travail effectué par l'outil a pu être menée via deux actions : lister les molécules non détectées par le système analytique ou non retrouvées par le système automatisé de calculs puis, compter les molécules retrouvées – c'est-à-dire correctement identifiées par le système pour calculer le recouvrement. Le recouvrement de l'outil mesure sa capacité à retrouver automatiquement les molécules présentes dans l'échantillon.

On a défini deux critères de validation : le **critère #1** et le **critère #2**. Le **critère #1** permet de comptabiliser les propositions correctes de molécules retrouvées par l'outil pour l'ensemble des hypothèses de masse. Le **critère #2** permet de comptabiliser les propositions correctes de molécules uniquement en considérant celles classées en première hypothèse (c'est-à-dire celles dotées du meilleur score). Le **critère #2** est donc plus exigeant que le **critère #1** puisqu'il permet d'évaluer les performances de l'outil en prenant en compte la fonction de score établie et implique qu'elle soit pertinente.

3.3.1. Performances de GAINS sur les données MS obtenues en mode FIA -1er réplicat-

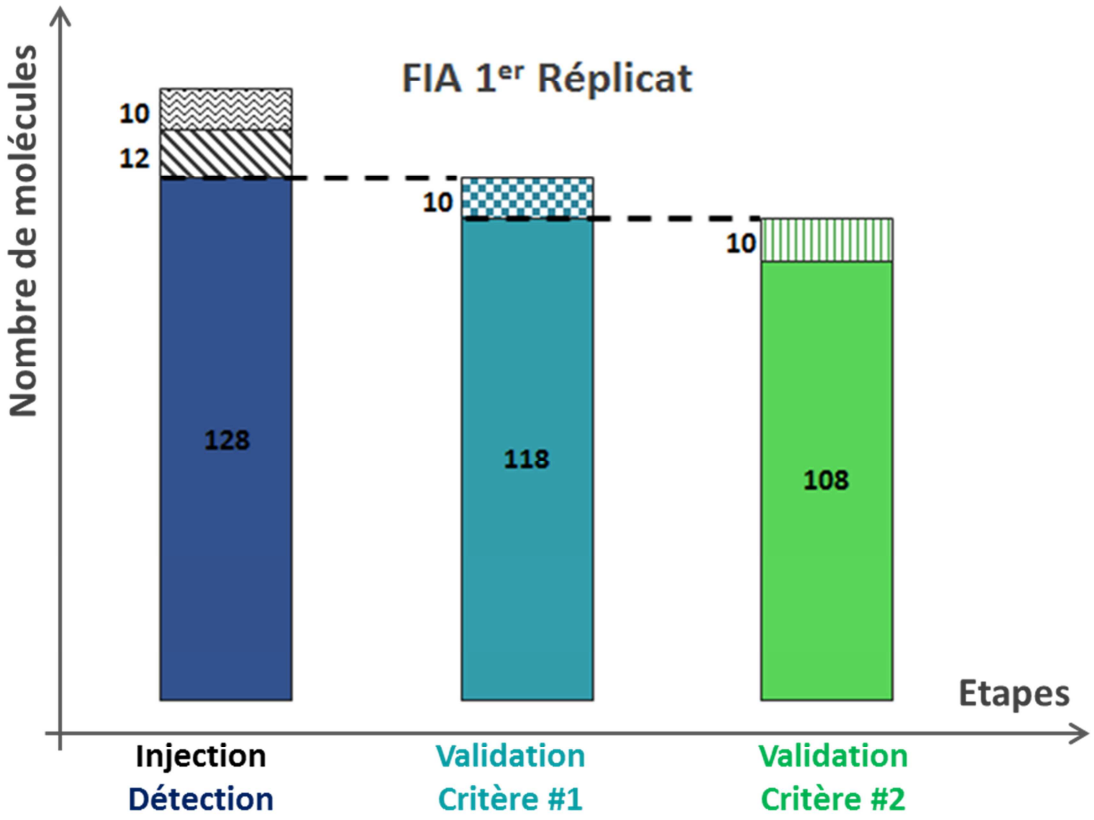


Figure 15 Performances de GAINS en FIA-HRMS obtenues lors du 1er réplicat pour le mélange de test de 150 molécules standard

3.3.1 a) *Molécules injectées et non détectables*

Sur 150 molécules injectées, 22 molécules ne sont pas détectables par GAINS : 10 d'entre elles ont subi un défaut à l'étape de prétraitement en amont de GAINS (molécules pour lesquelles les performances ne peuvent pas être évaluées à cause d'un problème de reconnaissance des massifs isotopiques ou de problèmes au moment du prétraitement des données sans qu'ils soient clairement identifiés) et 12 d'entre elles n'ont pas été détectées (Structures présentées en annexe 2.2.2). Les raisons qui peuvent expliquer l'absence de détection sont variées : absence d'ionisation avec la source électrospray, standard avec la mauvaise structure, ou dégradation du composé en solution.

3.3.1 b) *Validation selon critère #1*

Sur 150 molécules injectées, les performances de l'outil ont donc été évaluées sur un total de 128 molécules détectées. Parmi celles-ci, GAINS retrouve 118 molécules au total. Pour les 10 molécules non validées selon le critère #1, c'est-à-dire non retrouvées par GAINS (128 détectées - 118 retrouvées) : 8 d'entre elles se sont fragmentées en source et aucun ion du spectre n'a permis de faire la bonne hypothèse. Les composés concernés portent les numéros 15, 44, 89 et 98 (alcaloïdes)

et 34, 109, 133, 153 (terpènes et stéroïdes). Les structures sont présentées en annexe 2.2.1. D'un point de vue structural, les huit composés portent au moins une fonction hydroxyle, carbonyle ou ester. Un exemple de fragmentation empêchant de retrouver la masse du composé entier est illustré en Figure 16.

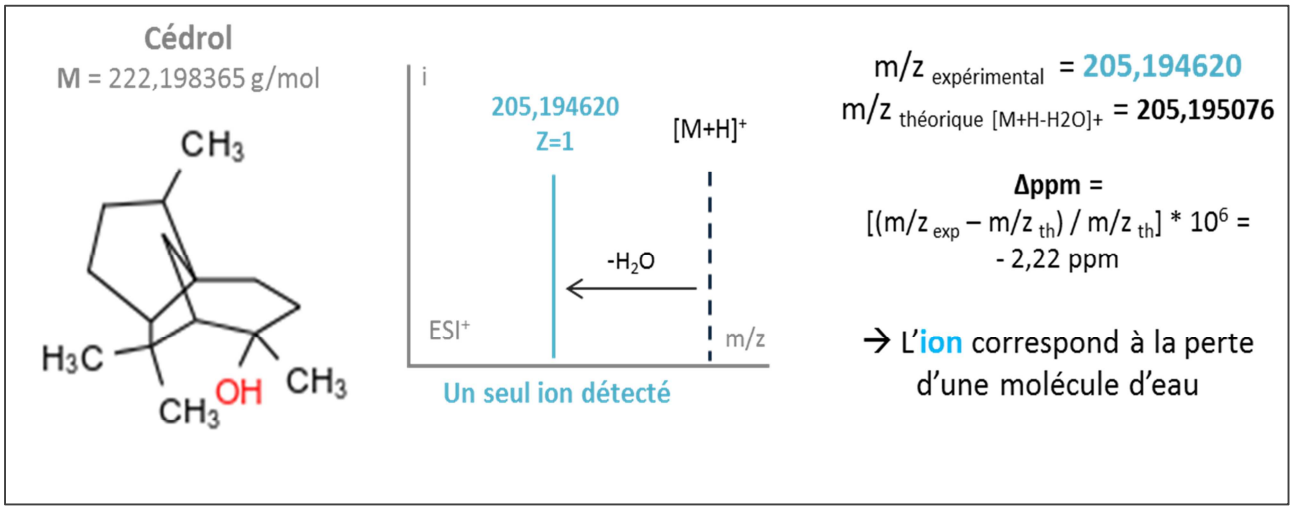


Figure 16 Exemple de molécule détectée mais non retrouvée par GAINS : cas d'un sesquiterpène pour lequel l'ion pseudo-moléculaire (en pointillé) n'est pas présent dans le spectre de masse expérimental

Dans le cas du cédrol (composé n°109), un seul ion est détecté en ESI^+. La valeur du rapport *m/z* correspond à la perte d'H_2O à partir de l'ion pseudo-moléculaire non observé. Ceci est confirmé par le calcul de la différence entre la masse monoisotopique du cédrol et celle mesurée expérimentalement (ion corrigé de la perte) : les masses sont équivalentes à environ 2 ppm près. Retrouver la masse de ce sesquiterpène est donc impossible dans ces conditions. Aucune information ne permet de savoir que le seul ion détecté n'est qu'un fragment de la structure moléculaire.

Globalement, la fragmentation en source a posé un problème pour environ 6 % des composés détectés. Le système n'est pas capable de gérer le cas où la molécule se fragmente en source de façon totale ou très importante. Toutefois, il est important de préciser que dans le cas d'une recherche manuelle non ciblée, ce problème reste le même et il n'est aucunement lié au traitement automatique. La fragmentation en source même en électrospray dite source d'ionisation douce, est liée à la propension d'une structure à se casser et dépend donc de sa fragilité.

Il reste deux composés non retrouvés parmi ceux détectés. Pour l'un d'eux (n°32), le problème venait de l'ionisation sous forme $[M+2H]^{2+}$ sans être l'ion le plus intense du spectre ; aucune correspondance n'a été trouvée avec cet ion et l'hypothèse de sa masse n'a pu être faite. Pour le dernier composé manquant (n°41), les ions pseudo-moléculaires en positif et en négatif portaient un *cluster* méthanol, ce qui a rendu impossible la mise en évidence du *cluster* et la bonne

hypothèse de masse. Les cas problématiques présentés avec ces deux composés restent des exceptions représentant 1,5 % des molécules détectées.

3.3.1 c) *Validation selon critère #2*

Pour aller plus loin dans la description des performances, il faut voir comment le score a classé les multi-hypothèses de masses générées pour un même spectre. L'outil a retrouvé 118 molécules dont 108 avec le meilleur score, c'est-à-dire classées en première position dans la liste des hypothèses de masse. La fonction de score a donc permis que 84 % des molécules détectées soient correctement classées.

Pour l'évaluation de GAINS via les analyses FIA, les calculs sont menés avec une tolérance de 10 ppm. Cette valeur a été ajustée suite à l'observation en FIA que l'erreur entre les masses proposées par GAINS et les masses théoriques ne dépassait pas 5 ppm. Cette valeur de tolérance fixée à 10 ppm en FIA avait permis dans un premier temps de ne pas appliquer un critère trop faible de correspondance entre les ions.

Les 10 composés retrouvés sans être bien classés (non validées selon critère #1) représentent 8 % des composés détectés (10 / 128). L'invalidation selon le critère #1 correspond à trois cas de figures. Le premier est celui où il n'y a eu aucune correspondance entre ions d'un même spectre. Des hypothèses de masse sont proposées pour l'ion majoritaire et le meilleur score est attribué à l'hypothèse obtenue en considérant l'ion sous la forme $[M+H]^+$. Or, il arrive que l'hypothèse correcte soit, en fait celle en considérant l'adduit sodé (composés concernés numéros 12, 74, 76, 80 et 86). Pour le composé n°12, molécule avec charge permanente positive répondant sous la forme $[M]^+$ de façon principale, il n'y a pas eu de correspondance entre différents ions et l'hypothèse de la masse en considérant la forme $[M]^+$ est la moins bien classée (*cf*. explications sur la mise en place de la fonction de score dans la partie 2.3 de ce chapitre). Ainsi, l'hypothèse de masse correcte pour le composé 12 est faite par le système mais n'est pas dotée d'un score élevé. Le cas des molécules chargées de façon permanente n'est pas facilement gérable puisque cette caractéristique est indétectable. En effet, il n'y a pas de différences entre un ion provenant d'une molécule de ce type et une molécule qui capte une charge au moment de l'ionisation. Dans ces cas-là, ce sont les informations externes à l'analyse qui peuvent orienter vers les propositions de masses correctes. Prenons l'exemple suivant : si la littérature rapporte la présence d'anthocyanes (molécules à charge positive permanente) pour une plante alors un extrait de cette plante est susceptible de contenir ces

molécules. Ainsi, les hypothèses de masse proposées pour la forme $[M]^+$ pourront être considérées avec plus d'intérêt.

Le second cas de figure, où les hypothèses correctes sont faites mais pas proposées en meilleure hypothèse, est lié à des spectres un peu complexes où de nombreux ions ont été référencés. Potentiellement, de nombreuses correspondances peuvent alors être trouvées et interfèrent avec la fonction de score (composés numéros 23 et 124).

Le troisième et dernier cas dans la difficulté à classer les hypothèses apparait lorsque la molécule est fragmentée pas totalement mais majoritairement : le spectre de la molécule contient des fragments source de hautes intensités et des ions de la molécule non fragmentée d'intensité plus faibles que les fragments. Ce cas a été vérifié pour 3 molécules (numéros 4, 10 et 103). Ces trois structures n'ont pas de fonction particulière en commun. Un exemple pour le composé 10 est donné avec la Figure 17 qui indique la liaison rompue sur la structure du composé analysé et les Figure 18 et Figure 19 qui montrent les spectres de masses expérimentaux correspondants à cette structure.

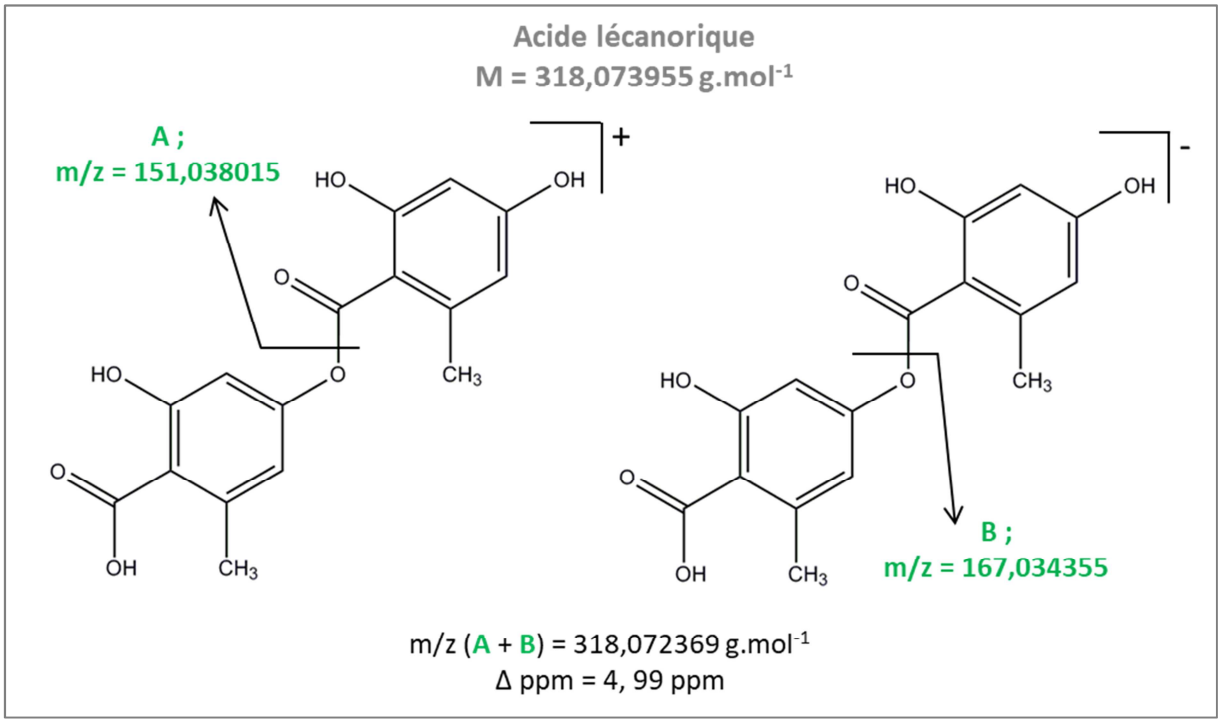


Figure 17 Exemple de molécule détectée dont la masse est retrouvée sans être classée en meilleure hypothèse : cas d'un composé phénolique dont la fonction ester subit une fragmentation en source

Dans le cas de l'acide lécanorique, la fonction ester qui relie les deux monomères d'acide phénolique se rompt dans la source. Ceci résulte en la détection de deux ions fragments majoritaires que sont A et B, respectivement détectés en mode positif et négatif. La somme de deux fragments est égale à la masse monoisotopique de la molécule analysée à 5 ppm près. On détecte donc bien la

structure mais majoritairement de façon fragmentée. Ces ions sont les plus intenses des spectres présentés à la figure suivante.

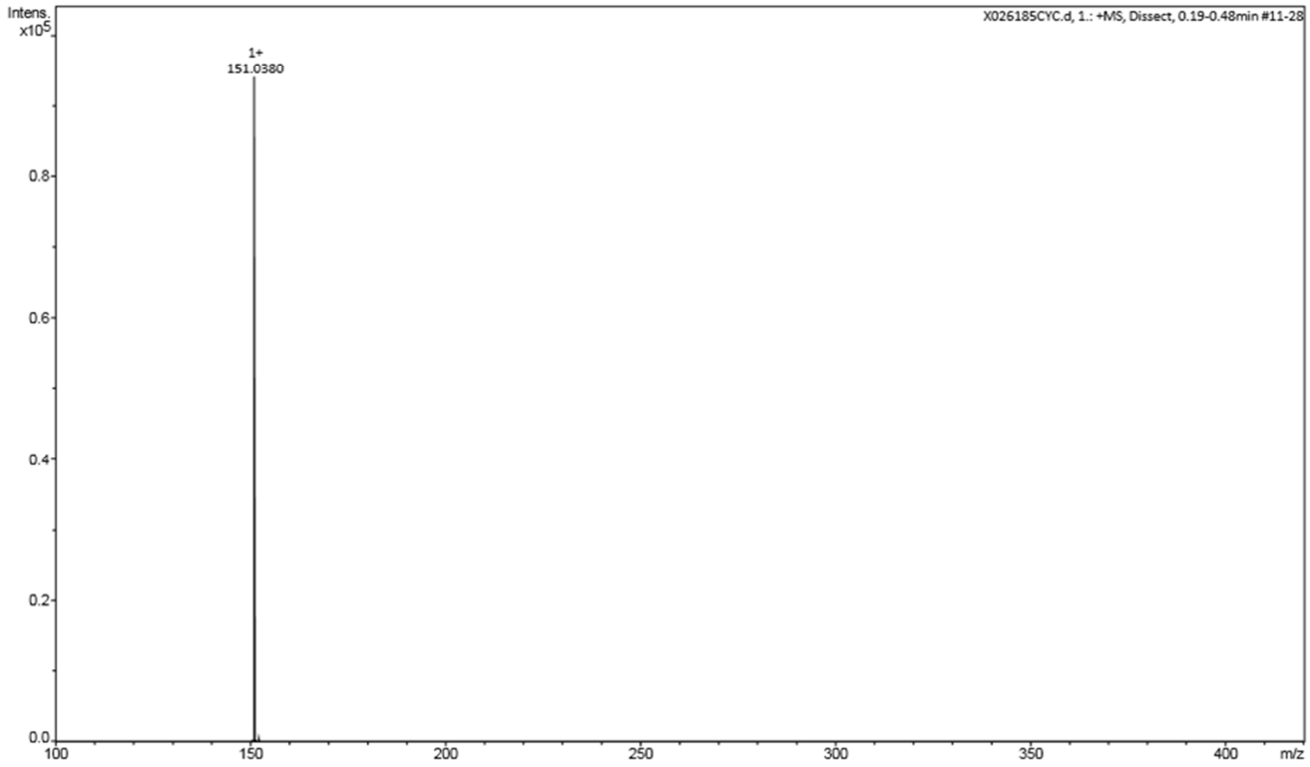


Figure 18 Spectre FIA-ESI-MS positif du composé n° 10 de masse monoisotopique 318,073955 g.mol^{-1} dont la fragmentation en source est majoritaire

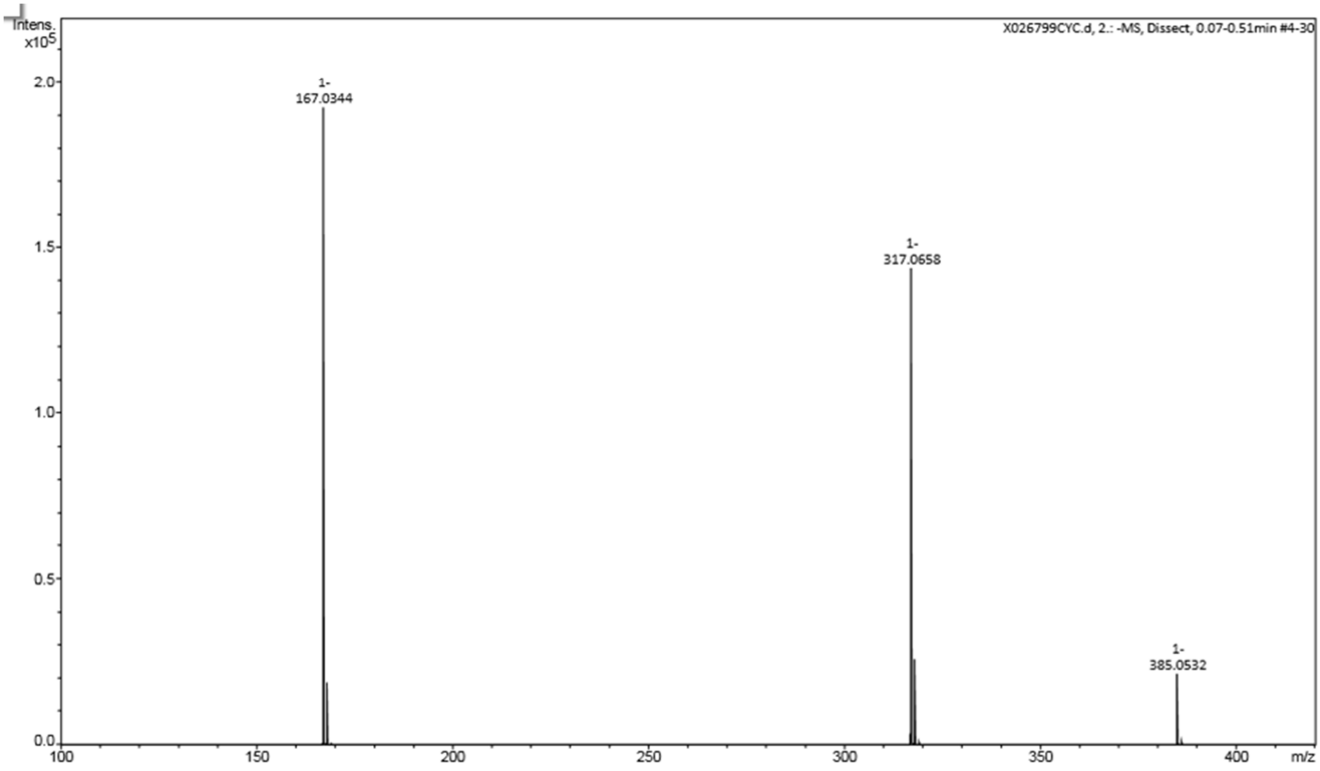


Figure 19 Spectre FIA-ESI-MS négatif du composé n° 10 de masse monoisotopique 318,073955 g.mol^{-1} dont la fragmentation en source est majoritaire

Pour le spectre du mode négatif, la fragmentation est moins radicale puisque l'ion $[M-H]^-$ et le plus exotique $[M+(Na^+-2H^+)+HCOOH]^-$ sont également visibles avec respectivement des valeurs de rapport *m/z* = 317,065783 et 385,053232. Toutefois, ils sont moins intenses que les ions fragments. Ainsi, la bonne hypothèse de masse est effectivement faite par le système GAINS mais n'est pas proposée en première hypothèse puisque les ions permettant cette hypothèse correcte sont moins intenses que les ions fragments.

3.3.2. Performances de GAINS sur les données MS obtenues en mode FIA -2nd réplicat-

Les résultats obtenus pour le second réplicat sont présentés ci-dessous :

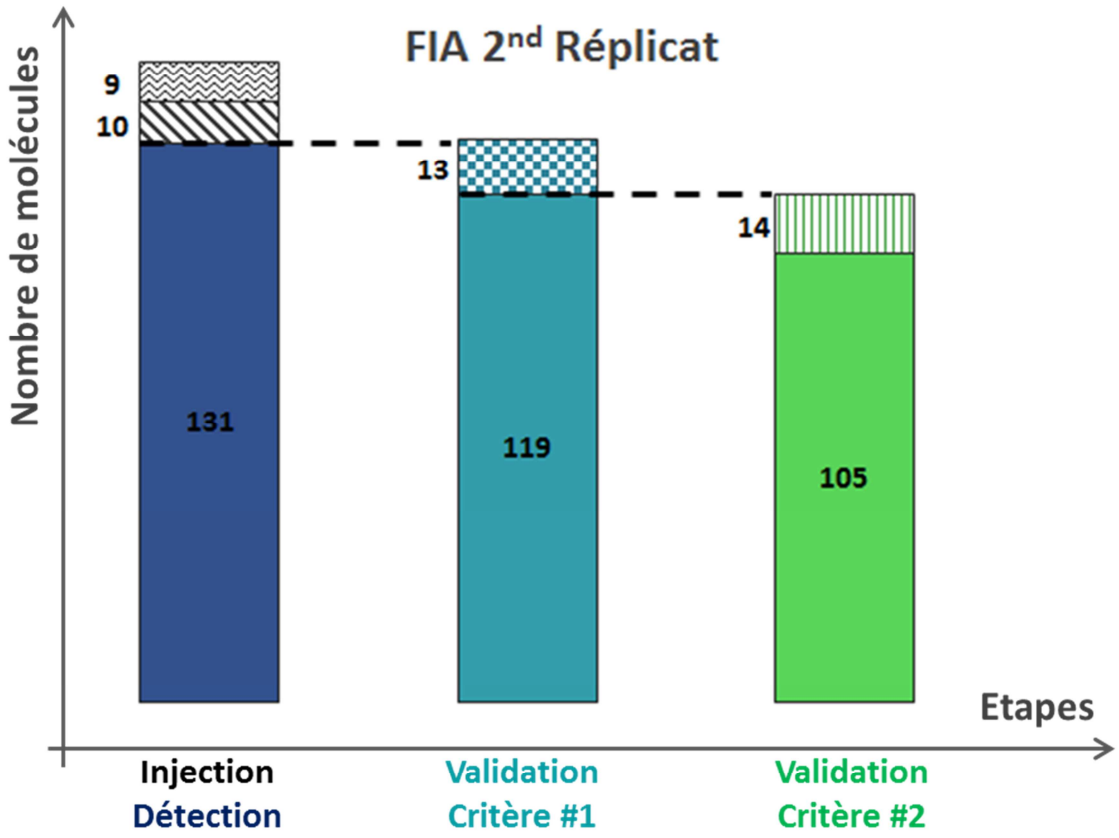


Figure 20 Performances de GAINS en FIA obtenues lors du second réplicat sur le mélange test de 150 molécules standard

3.3.2 a) *Molécules non détectées ou non détectables*

Sur 150 molécules injectées, il manque 19 molécules : 9 d'entre elles ont subi un défaut à l'étape de prétraitement en amont de GAINS (molécules pour lesquelles les performances ne peuvent être évaluées à cause d'un problème de reconnaissance des massifs isotopiques ou de

problèmes au moment de prétraitement des données sans qu'ils soient clairement identifiés) et 10 d'entre elles n'ont pas été détectées (Structures présentées en annexe 2.2.3).

3.3.2 b) *Validation selon critère #1*

Sur 150 molécules injectées, les performances de l'outil ont donc été évaluées sur un total de 131 molécules constituant le nombre maximum de molécules détectées. Ce sont 119 molécules qui ont été retrouvées lorsqu'on considérait toutes les hypothèses faites. Le recouvrement atteint donc 91 % si l'on prend en compte la totalité des hypothèses émises sans restreindre la recherche à la meilleure hypothèse.

Pour les 12 molécules non retrouvées : 10 d'entre elles se sont fragmentées en source et aucun ion du spectre n'a permis de faire la bonne hypothèse. Les composés concernés portent les numéros : 37 (halogénés) ; 148 (composés phénoliques) ; 25, 34, 109 (terpènes et stéroïdes) ; 15, 44, 89, 93, 98 et 103 (alcaloïdes). Les 2 autres molécules détectées et non retrouvées constituent les cas limites que GAINS ne gère pas : pour les molécules n° 32 et 41, les raisons sont identiques à ce qui s'est passé lors du premier réplicat.

3.3.2 c) *Validation selon critère #2*

Le recouvrement est de 80 % si l'on limite les performances aux seules hypothèses les mieux classées : 105 composés sur 131 détectés se trouvent dans ce cas de figure. Sur ces 119 composés ayant été validés par le critère #1, le nombre de composés validés selon le critère #2 s'élève à 105 composés. En effet, ils sont retrouvés avec une proposition classé en première position par GAINS. Les 14 composés, non validés selon critère #2, sont retrouvés dont la bonne hypothèse de masse n'est pas en première position. Parmi ces 14 composés, dans la majorité des cas, l'hypothèse correcte est la seconde de la liste des masses proposées par GAINS. Le fait que certains spectres contiennent un nombre important d'ions (dû à des fragmentations source par exemple) permet que d'autres hypothèses de masse soient mieux classées.

3.3.3. Comparaison des résultats de deux réplicats en FIA-MS

Sur 150 molécules injectées :	**1er réplicat**	**2nd réplicat**
Molécules détectées	**128**	**131**
Molécules identifiées au total	**118 (92 %)**	**119 (91 %)**
Molécules identifiées en 1ère hypothèse	**108 (84 %)**	**105 (80 %)**

Tableau 11 Récapitulatif des résultats obtenus pour valider GAINS sur le traitement des données MS : comparatif sur les données collectées en FIA lors de deux essais indépendants

Les performances globales en FIA sont comparables pour les deux réplicats. Le nombre total de molécules retrouvées ainsi que le nombre de molécules retrouvées et classées en meilleure hypothèse sont très proches. Une majorité des composés non détectés et non retrouvés le sont dans les deux cas.

Cependant, certaines molécules étaient retrouvées dans le premier réplicat et ne le sont pas pour le second. Elles sont non retrouvées pour cause d'absence de correspondance d'ions pertinents (un seul ion détecté) ou souvent à cause de fragmentation en source : cas des molécules (n° 25, 37, 93 et 103). Ce phénomène de fragmentation n'est pas constant. Au contraire, les molécules n° 65, 78, 79, 119 et 129 sont détectées pour le 2nd réplicat mais ne l'étaient pas pour le 1er. Pour quatre d'entre elles, ceci est lié à un problème rencontré au moment du prétraitement et une entité n'est pas du tout détectée dans le 1er réplicat.

En conclusion, s'agissant du traitement des données pour proposer des masses monoisotopiques associées à chaque composé à partir de son spectre de masse, la première phase d'évaluation des performances a permis de montrer que le système atteignait les mêmes niveaux de performances pour deux réplicats indépendants. De plus, ce travail a abouti sur une première validation de l'outil qui s'est révélé performant grâce à un grand niveau de recouvrement (autour de 90 % et une fonction de score efficace qui a engendré un classement de la bonne hypothèse en première position dans plus de 80 % des molécules détectées.

3.3.4. Performances de GAINS sur les données MS obtenues en LC-HRMS

GAINS est un outil d'analyse de matrices complexes qu'il est nécessaire de séparer pour être en mesure de les décrire plus aisément. Ainsi, pour aller plus loin dans l'évaluation de l'outil, nous avons ajouté la dimension séparative pour analyser le mélange test.

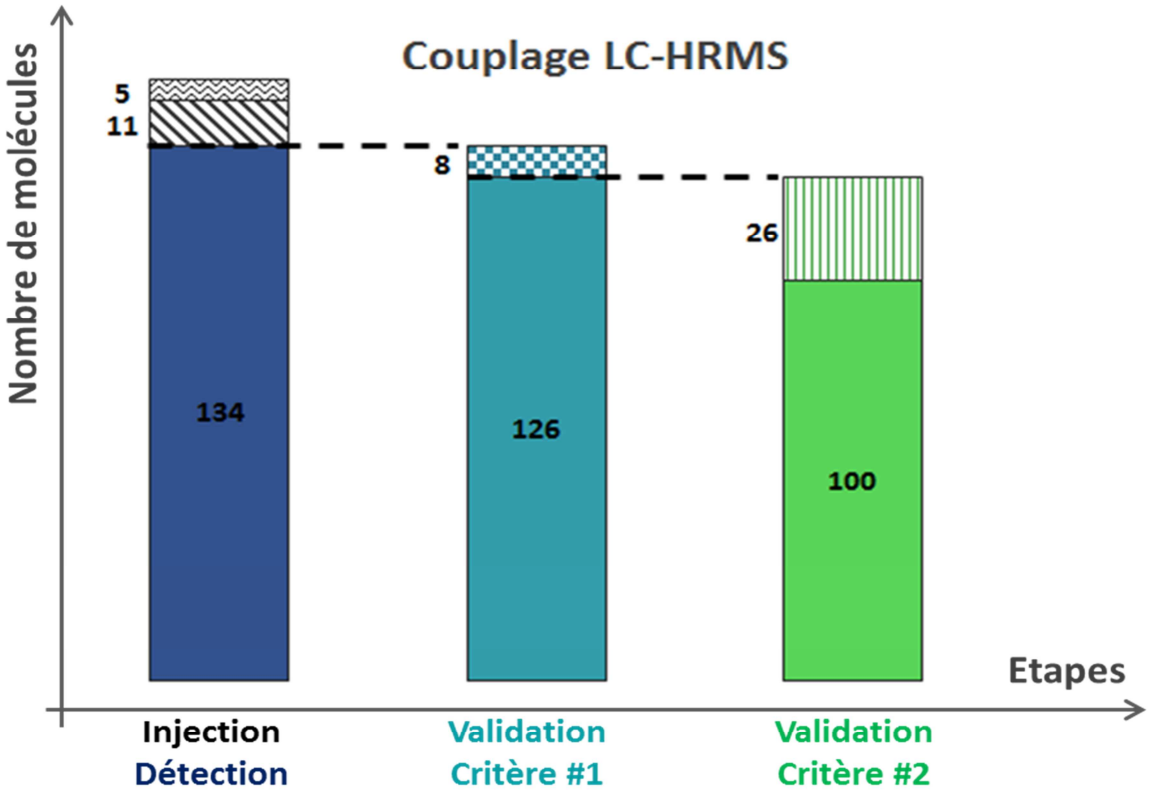


Figure 21 Performances de GAINS en couplage LC-HRMS obtenues pour le mélange test

3.3.4 a) *Molécules non détectées ou non détectables*

Dans le cas de l'analyse en LC-HRMS des 150 molécules en mélange, les performances de GAINS ont été très satisfaisantes. Pour commencer, on remarque que 16 molécules ne sont pas détectées sur les 150 injectées avec la répartition suivante : 11 non détectées physiquement et 5 non prétraitées correctement. Au total, ce sont 6 molécules détectées de plus qu'en FIA. Pour expliquer cette différence, c'est le nombre d'erreur au prétraitement qui a diminué avec l'apport de la dimension séparative LC. La majorité des molécules non détectées en LC-HRMS est commune à celles non observées en FIA.

3.3.4 b) *Validation selon critère #1*

Ensuite, sur les 134 molécules pouvant être retrouvées au maximum, GAINS en a retrouvé 126 lorsqu'on a considéré la totalité des hypothèses que l'outil a faites. Ceci équivaut à 94 % de recouvrement et ce taux est égal à celui obtenu en FIA.

Les molécules détectées mais non retrouvées par le système sont au nombre de 8 (134 détectées - 126 retrouvées). Parmi elles, 5 ont subi une fragmentation importante en source (composés n° 34, 44, 96, 109 et 153) et 3 molécules échappent au système de GAINS et constituent

des exemples de limites auxquelles l'outil est confronté. Il s'agit des composés n° 32, 41 et 139. Les composés n°32 et 41 ne sont pas retrouvés et ce, pour les mêmes raisons que celles exposées dans le cas de leur analyse en FIA (*cf*. paragraphe 3.2.2 de ce chapitre). S'agissant du composé n° 139, il est bien détecté sous la forme $[M+H]^+$ qui n'est ni l'ion majoritaire du spectre (fragments plus intenses), et qui n'est pas en correspondance avec un autre ion définissant une même masse neutre.

3.3.4 c) *Validation selon critère #2*

Enfin, si l'on prend en compte uniquement les hypothèses les mieux classées, GAINS a proposé le résultat correct pour 100 composés sur 134 soit 75 % des composés. Ce taux de recouvrement est légèrement inférieur à celui obtenu en FIA.

Au sujet des composés détectés, retrouvés par GAINS mais dont la masse n'est pas classée en meilleure hypothèse, on en compte 26 (n° 10, 12, 13, 25, 29, 33, 65, 66, 74, 79, 87, 92, 111, 113, 124, 125, 130, 136, 147, 149, 150, 152 et 15 ou 63, 17 ou 39, 47 ou 49, 89 ou 136 avec une incertitude pour 4 composés appartenant à des paires d'isobares et dont un seul a été détecté). C'est un nombre plus important qu'en FIA lorsque chaque composé était injecté individuellement. Dans le cas de l'analyse d'un mélange complexe, il se peut que les spectres n'aient pas parfaitement été prétraités par la fonction DISSECT qui retrouve les composés après analyse et avant les calculs opérés par GAINS.

Le recouvrement atteint est le signe que le système et le score sont définis de manière satisfaisante et fonctionne parfaitement pour plus de trois quarts des molécules. Ainsi, on peut avoir relativement confiance en l'outil. GAINS nous permet de traiter de larges jeux de données analytiques de façon plus rapide et plus complète que s'il avait fallu le faire manuellement tout en évitant notamment de probables erreurs de calculs.

3.4 Validation de l'outil : traitement des données MS/MS

3.4.1. Principe de l'évaluation

La dernière étape semi-automatique au sein de GAINS fait intervenir les données de fragmentation (MS/MS). Le but est d'aller plus loin dans la tentative d'identification des composés analysés puisque la fragmentation révèle certains groupements et permet donc de décortiquer la structure plutôt que de l'analyser de façon globale comme le fait l'analyse en MS. L'outil Metfrag intervient dans le processus pour gérer la similarité entre les données *in silico* qu'il est capable de générer et les données de fragmentation expérimentales.

Pour appréhender l'apport de l'outil Metfrag dans notre stratégie, nous allons évaluer sa capacité à être sélectif en présence du bon candidat parmi un certain nombre d'autres candidats pour un composé. Il est nécessaire que l'outil dote le bon candidat d'un bon score.

3.4.2. Réalisation de l'évaluation

L'essai de Metfrag est effectué sur les données MS/MS obtenues en FIA. L'énergie de fragmentation a été réglée à 25 eV. C'est une valeur moyenne qui permet une fragmentation suffisante mais pas trop importante des composés. Les lignes de commandes de la version 2.3 de Metfrag ont été implémentées dans notre script. Celles utilisées ainsi que leur réglage sont décrits en 0.

Pour rappel, en FIA, la fragmentation *in silico* est lancée sur l'ensemble des 118 molécules identifiées en MS par GAINS. Au démarrage, le nombre de molécules éligibles au fonctionnement de Metfrag s'est vu réduit de 118 molécules à 93 molécules. En effet, en pratique, différentes contraintes ont été rencontrées soit par rapport au fonctionnement de Metfrag et aux données d'entrée dont il a besoin pour lancer les modélisations de spectres, soit du fait de la qualité insuffisante des données collectées expérimentalement. Ainsi, pour un certain nombre de composés, le lien des données de fragmentation avec la molécule n'était pas établi pour Metfrag. En effet, soit l'ion précurseur fragmenté expérimentalement n'était pas parmi ceux ayant permis l'hypothèse de masse effectuée par GAINS pour la molécule, soit l'ion précurseur ne faisait pas parti des types d'ions que l'outil accepte comme possible précurseur. Les ions correspondants à une perte d'eau, les formes dimères ou les formes dichargées ne sont pas envisagées comme précurseur par Metfrag. Par ailleurs, pour certaines molécules, aucune donnée de fragmentation n'a été associée au moment de l'application de l'algorithme DISSECT pour traiter les données brutes (*cf*. début du chapitre paragraphe 2.2.2). Alors, les données modélisées n'ont pas de données expérimentales de référence pour établir la comparaison. Dans notre expérience, avec 25 molécules non prises en compte, ce sont 21 % des composés qui n'ont pas pu être traités par Metfrag.

3.4.3. Résultats du test de Metfrag

Pour évaluer l'apport de Metfrag à notre démarche, nous avons mesuré le taux de structures correctes proposées en fonction du score considéré : la Figure 22 présente les résultats obtenus sur le jeu de 93 molécules analysées en FIA et dont les données permettaient l'exploitation par Metfrag.

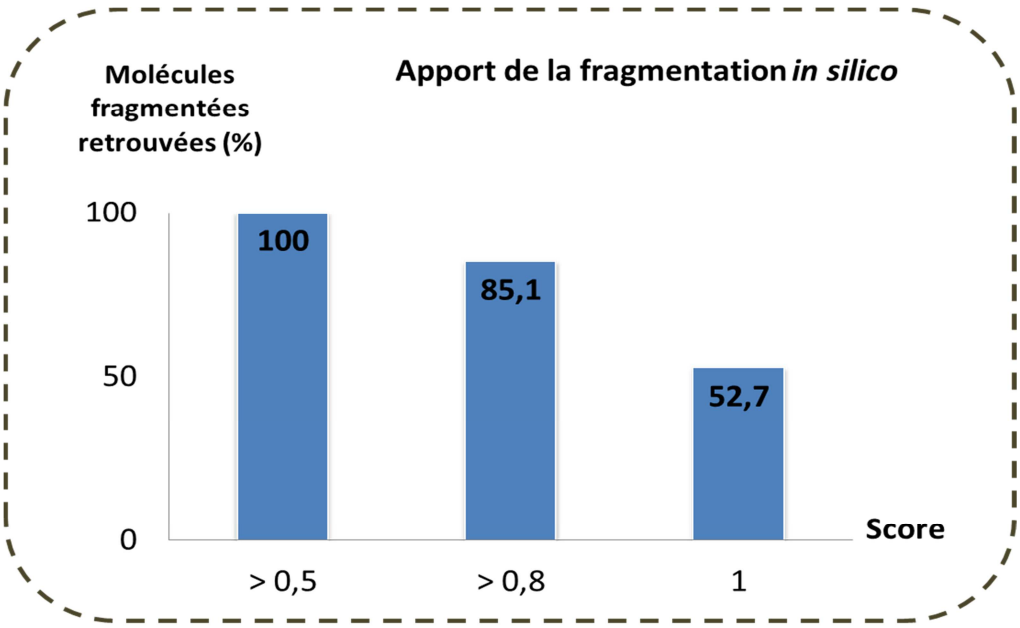


Figure 22 Bilan de l'utilisation de Metfrag sur le classement des structures fragmentées *in silico* : pourcentage de molécules retrouvées en fonction du score considéré. Le score est calculé et normalisé entre 0 et 1 par le module Metfrag.

Le critère de score est assimilable au niveau d'exigence que l'opérateur fixe au sujet de la pertinence du module Metfrag dans sa capacité à trier les candidats structuraux. Plus le score considéré est élevé, plus la similarité de l'empreinte spectrale MS/MS entre le spectre modélisé et le spectre expérimental est censé être importante. En termes de résultats, si on se limite aux structures ayant obtenu le score maximum (score = 1), nous retrouvons un peu plus de la moitié (53 %) des structures analysées pour l'évaluation. En revanche, pour que 100 % des structures attendues soient proposées, la limite de score à considérer doit être abaissée jusqu'à une valeur de 0,5. Ceci rend l'application de Metfrag un peu limitée. En effet, le nombre de molécules pouvant obtenir un score entre 0,5 et 1 est potentiellement plus important que si l'on pouvait se limiter à ne considérer que les structures dotées d'un score de 1. Le tri opéré via la modélisation des fragmentations des candidats n'est donc pas totalement sélectif même s'il permet de franchir un pas de plus dans la description des molécules potentiellement présentes.

Par ailleurs, Metfrag propose un score pour l'ensemble des structures possibles. Celui-ci est normalisé entre 0 et 1 sur la base du meilleur score obtenu. Ceci créé une limite certaine quant à la pertinence des résultats proposés. Pour saisir cette limite, le

Parmi n candidats, résultats pour le meilleur candidat :	Composé 1	Composé 2
Nombre d'ions équivalents entre spectre modélisé et spectre expérimental	1 / 6	5 / 5
Score normalisé Metfrag	1	1

Parmi n candidats, résultats pour le meilleur candidat :	Composé 1	Composé 2
Nombre d'ions équivalents entre spectre modélisé et spectre expérimental	1 / 6	5 / 5
Score normalisé Metfrag	1	1

Tableau 12 donne un exemple pour deux composés différents de l'analyse.

Tableau 12 Exemple pour illustrer la limite de la normalisation du score au sein de Metfrag

Pour deux composés différents, chacun a un candidat structural obtenant un score de 1. Pour le composé 1, le meilleur candidat n'a qu'un seul fragment en commun avec ceux du spectre expérimental. Pour le composé 2, la similarité entre le spectre expérimental et modélisé est bien plus importante puisque tous les fragments sont expliqués. Ainsi, le score normalisé n'est pas le reflet direct de la similarité des spectres comparés. La valeur du score normalisé n'est pas un gage absolu de la qualité du candidat proposé.

On propose un autre paramètre d'évaluation de l'apport de Metfrag. Il s'agit du rang auquel est classée la structure correcte. Pour les 93 molécules utilisées pour l'évaluation, la Figure 23 présente le pourcentage de molécules retrouvées en fonction du rang de classement.

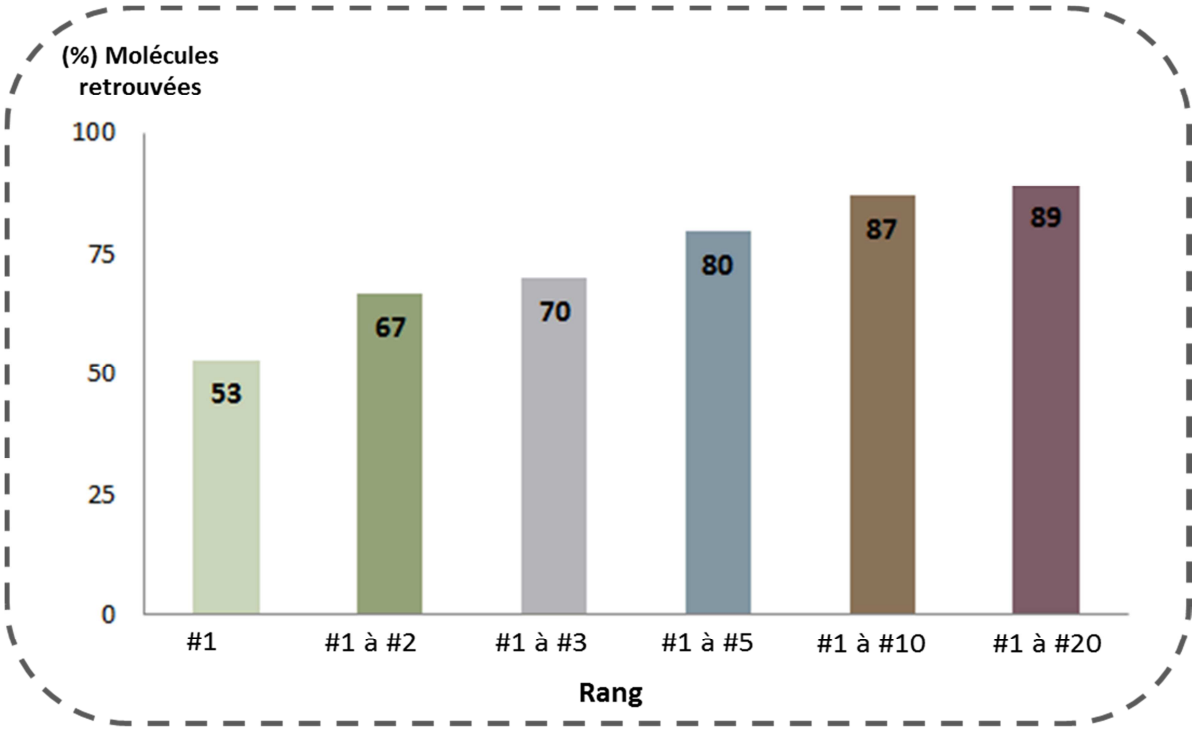


Figure 23 Bilan de l'utilisation de Metfrag sur le classement des structures fragmentées *in silico* : pourcentage de molécules retrouvées en fonction du rang de classement considéré.

Metfrag a classé les candidats structuraux pour chaque molécule à caractériser. Si l'opérateur choisit de ne considérer que les molécules classées au rang #1, on retrouve 53 % des molécules du

mélange. L'outil a établi un classement correct des structures uniquement pour la moitié des composés. On retrouve le même pourcentage que dans la Figure 22 pour les molécules dotées du meilleur score. Ceci est normal car on représente le même résultat mais de deux façons différentes. Ensuite, on observe que plus on regarde un rang élevé de classement, plus la probabilité que la molécule attendue soit décrite augmente. Toutefois, cette progression a tendance à plafonner vers 90 %. Globalement, la figure précédente indique dans 70 % des cas, la molécule correcte est retrouvée dans les 3 premières places de classement et que ce taux atteint 89 % si on considère les 20 premiers candidats de la liste. Du point de vue de l'opérateur qui doit tirer des conclusions, ces résultats sont positifs car pour 90 % des molécules, on peut trouver la molécule correcte parmi un nombre de candidats restreint.

Toutefois, ce recouvrement (nombre de molécules analysées et effectivement retrouvées) en fonction du classement n'indique pas le nombre total de molécules qui seraient classées au même rang (scores égaux). En effet, une ou plusieurs structures peuvent obtenir le même score Metfrag, ce qui les classe au même rang. Ceci nous interroge sur le niveau de sélectivité qu'apporte Metfrag pour classer les candidats.

Nous avons mis en place un moyen de déterminer la sélectivité du système. On entend par sélectivité la capacité du système à réduire le nombre de candidats possibles pour un composé détecté : on veut voir dans quelle proportion un outil parvient à discriminer les faux candidats. Pour chaque molécule à caractériser, la sélectivité a été calculée selon l'équation suivante :

$$\frac{[Nombre\ total\ de\ candidats - Nombre\ de\ candidats\ de\ score\ supérieur\ ou\ égal\ \text{à}\ celui\ du\ candidat\ correct]}{Nombre\ total\ de\ candidats} \text{ x } 100$$

La Figure 24 montre la sélectivité obtenue pour les 93 composés traités par Metfrag et présentée en fonction du nombre de candidats associés à une molécule à caractériser.

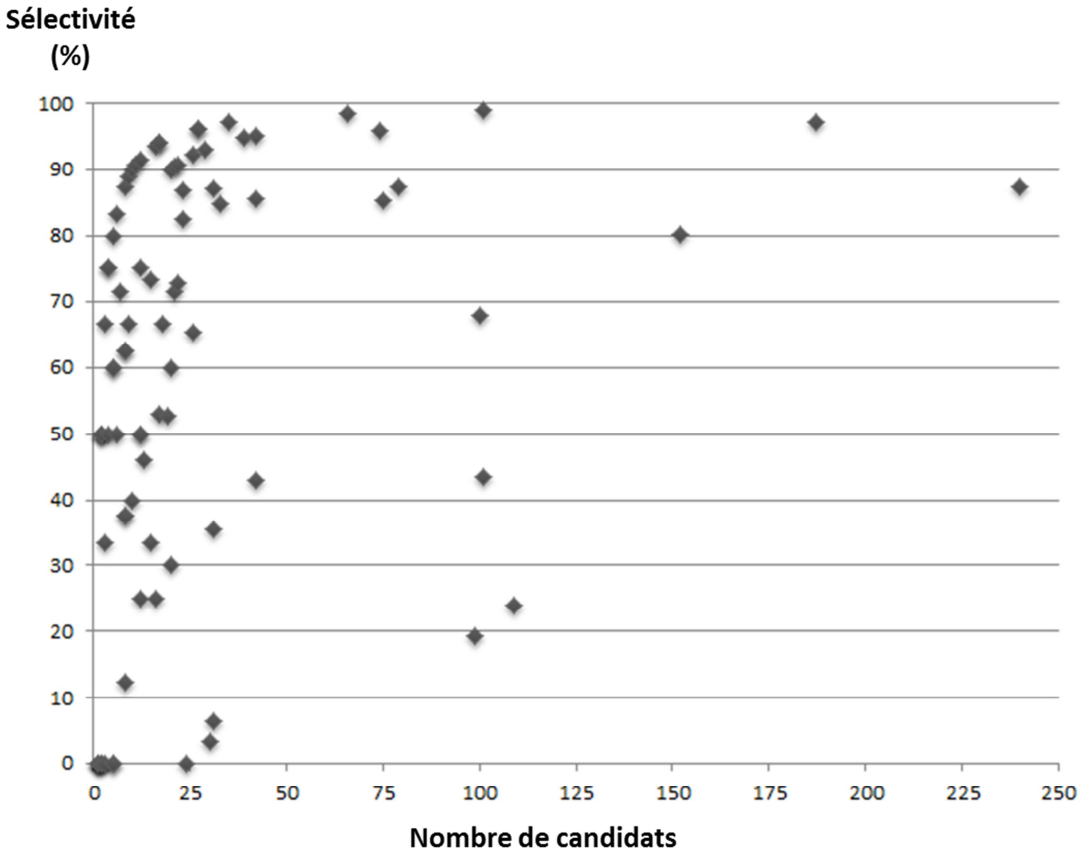


Figure 24 Sélectivités atteintes avec Metfrag en fonction du nombre de candidats structuraux associés à une masse donnée

Pour les cas où le nombre de candidats est réduit (1 à 50), les sélectivités sont variées, avec une gamme de valeurs très larges (de 0 % à plus de 90 %). De plus, quand il y a peu de candidats, ne pas être doté du score le plus élevé impacte plus fortement la pertinence du classement : être troisième parmi 3 candidats signifie être dernier.

Les cas où le nombre de candidats est supérieur à 50 sont moins fréquents. Toutefois, dans cette situation, les valeurs de sélectivité sont majoritairement élevées (entre 70 et 99 %). Il semble que le système apporte de la sélectivité puisqu'à nombre conséquent de candidats à différencier, les valeurs de sélectivité sont élevées. En d'autres termes, l'outil est capable de bien classer (à un rang élevé sans forcément être le mieux classé) le candidat correct lorsqu'il a un nombre important de candidats à distinguer. En résumé, lorsque l'on a peu de candidats, ce qui est en soit une situation favorable, Metfrag apporte parfois peu de sélectivité. Par contre, quand le nombre de candidat augmente, l'utilisation de Metfrag prend tout son sens.

Pour exprimer la valeur de sélectivité de façon résumée, nous avons calculé la moyenne pondérée de l'ensemble des sélectivités et pris en compte le nombre de candidats selon l'équation suivante :

Pour les i molécules à caractériser ; $\frac{\sum_i (\boldsymbol{sélectivité\ x\ nombre\ de\ candidats})}{\sum_i \boldsymbol{nombre\ de\ candidats}}$

Plus le nombre de candidats possibles pour caractériser une molécule est important, plus il est difficile pour le système d'être sélectif. Ainsi, le nombre de candidats est le facteur de pondération. La moyenne pondérée des sélectivités est égale à 71 %.

Cette valeur indique que le système parvient, en moyenne, à écarter 70 % des molécules candidates. Cette moyenne ne permet pas de régler un paramètre de l'outil puisqu'on ne peut pas gérer le nombre de candidats. Mais ce résultat est un moyen d'évaluer l'outil et par conséquent, de le comparer à d'autres outils si nécessaire.

3.4.4. Perspectives d'amélioration dans le traitement des données MS/MS

S'agissant de l'exploitation des informations, des études complémentaires liées à la comparaison d'autres outils d'aide à l'identification à partir des spectres de fragmentation seraient nécessaires à condition qu'ils soient en libre-accès. Ils pourraient être intégrés aux étapes de GAINS. Les outils existants sont basés sur le même principe que Metfrag à savoir la fragmentation combinatoire ou d'autres principes tels que la modélisation par fragmentation compétitive ou la prédiction d'empreintes moléculaires (Détails sur ces techniques en chapitre 1).

De plus, la mise en place d'une recherche automatisée de la présence ou non de groupements glycosylés et/ou acylés demeure nécessaire (voir détails au paragraphe 2.5 de ce chapitre). D'une part, ces groupements sont très fréquents au sein des molécules naturelles. D'autre part, ils sont aisément fragmentés du fait des liaisons chimiques mises en jeu puisque les groupements liés aux génines sous formes O-glycosylées ou O-acylées se rompent relativement facilement en électrospray. Pour l'heure, ce travail est effectué manuellement à l'aide d'un formulaire pour vérifier ces éléments et de façon restreinte en ciblant les molécules majoritaires ou des composés d'intérêt particulier.

3.5 Durées des étapes de traitement et de calcul

S'agissant du temps que la démarche peut prendre, on peut donner une idée générale pour chaque grande étape. Après avoir effectué les analyses d'un extrait dans les modes positif et négatif, le prétraitement et la récupération de fichiers des données prétraitées requiert environ 2h. Il faut charger les données d'entrée au sein de GAINS et qu'elles s'affichent avant les calculs, ce qui dure environ 1 ou 2 minutes. Les calculs de détermination des masses monoisotopiques par GAINS nécessitent au maximum quelques heures. Les requêtes au sein de la GPDB sont faites en quelques

minutes dans le cas de recherche la plus complète c'est-à-dire au sein de la totalité de la base de données. Les modélisations *in silico* via Metfrag sur toutes les molécules d'un extrait sont de l'ordre de quelques heures maximum.

Chaque étape de calculs peut être exécutée de façon indépendante et est à l'initiative de l'utilisateur. Ceci apporte de la flexibilité au système et l'opérateur peut suivre les différents avancements. En fonction des résultats, il peut relancer des calculs avant de passer à une nouvelle étape dans le cas d'erreurs de paramétrages ou de fichiers d'entrée par exemple et ce, sans avoir attendu le terme de tout le processus.

S'agissant du temps nécessaire au traitement des données, l'outil apporte un véritable avantage comparé à un traitement manuel. Au-delà du temps, l'amplitude de données traitées ne peut pas être aussi importante lorsqu'on travaille manuellement. Toutefois, les temps de calcul pourraient être optimisés via l'utilisation des langages de programmation plus adaptés notamment et la durée de certaines étapes de GAINS reste un point à améliorer à l'heure où les capacités techniques pour générer des résultats très rapidement sont disponibles.

4. Limites de l'outil

4.1 Limites théoriques intrinsèques à la structure de l'outil

GAINS est un outil à la disposition du phytochimiste pour l'aider à caractériser le plus efficacement la composition phytochimique d'extraits naturels. Le développement de l'outil a été mené pour fonctionner de façon pertinente dans le cas général. La majorité des travaux concernent l'étude de métabolites secondaires dont les polarités peuvent être de faibles à importantes. Les limites de GAINS exposées dans la suite du paragraphe sont donc celles rencontrées au cours d'un fonctionnement de GAINS dans le cas général.

La première limite dont il faut avoir conscience est que l'identification proposée par l'outil n'est pas formelle. Cette limite est inhérente à la démarche elle-même qui est d'abord destinée à permettre une meilleure appréhension des données analytiques denses à aborder manuellement. L'outil dégage du temps à l'utilisateur pour qu'il puisse confirmer les éléments phytochimiques les plus complexes et mener son étude de façon pertinente. Une caractérisation idéale de l'extrait qui consisterait à identifier formellement la totalité des composés et ce, le plus automatiquement possible, n'existe pas à l'heure actuelle.

S'agissant du domaine de polarité des molécules qui peuvent être caractérisées, celui-ci est limité par la méthode d'analyse mise en jeu. En effet, l'ionisation avec une source électrospray est adaptée à des composés de polarité importante à faible mais n'est pas très adaptée aux molécules apolaires. En choisissant cette source d'ionisation, on sait que certaines familles de molécules ne seront pas détectables. Toutefois, l'électrospray reste la source d'ionisation qui offre le spectre d'applications le plus large aujourd'hui. Elle est adaptée à la majeure partie de nos études dans le cas de l'analyse de métabolites secondaires de végétaux.

Concernant le traitement des données, pour chaque pic chromatographique, le système propose des hypothèses multiples de masses monoisotopiques. Or, dans la majorité des cas, le pic chromatographique correspond à l'élution d'un seul composé, donc une seule masse parmi celles proposées est correcte. Etablir un classement pertinent des différentes propositions de masse est d'une grande nécessité. Le score mis en place est basé sur des éléments empiriques et fondamentaux liés à la spectrométrie de masse. La problématique est d'être sûr de la fonction de score. Nos étapes de validation ont montré que la fonction était bien adaptée. Toutefois, les résultats générés automatiquement ne sont pas totalement figés puisque l'hypothèse classée en première position peut être sélectionnée manuellement au sein de l'interface. On peut citer le cas des molécules qui vont se fragmenter en source de façon importante. Les fragments seront les ions les plus intenses du spectre et GAINS va faire des hypothèses de masses à partir de ces fragments. Un traitement manuel aurait probablement conduit aux mêmes résultats. Par contre, après avoir recherché les structures correspondant à la masse obtenue, l'opérateur peut se rendre compte de l'incohérence entre la polarité des structures et le temps de rétention du composé. Dans l'exemple d'une molécule glycosylée, si le système propose une structure de génine à un temps de rétention faible et/ou proche d'autres composés glycosylés, on peut supposer que la proposition n'est pas la bonne. La prise en compte de ces éléments reste une limite difficile à gérer de façon automatique et c'est là que l'expert doit intervenir.

Enfin, l'espace interrogé pour la recherche des candidats structuraux est celui de la GPDB qui est une large base de données mais qui reste non exhaustive par rapport à l'ensemble de l'espace chimique des molécules naturelles. En effet, si une molécule n'est pas connue, elle ne peut être présente dans la base de données. Ainsi, le contenu de celle-ci borne toute requête pour trouver des candidats. De ce fait, il est primordial que des mises à jour régulières et importantes soient effectuées pour insérer les nouveaux composés découverts et éviter que l'espace chimique couvert ne reste figé. Enfin, concernant la GPDB, on précisera que les métabolites primaires y sont très peu représentés (absence totale de protéines) car elle est spécifique aux métabolites secondaires des plantes. Ceci est cohérent avec l'usage qui en est fait puisque la majorité des molécules naturelles

valorisées dans les domaines dermo-cosmétique et nutraceutique font partie du métabolisme secondaire des plantes.

Concernant l'exploitation automatique des données de MS/MS, les limites assimilées à l'utilisation d'un modèle de fragmentation combinatoire comme Metfrag sont à mentionner. La fragmentation combinatoire s'appuie sur le principe de Dissociation Energétique de Liaison, c'est-à-dire que l'outil va rompre virtuellement les liaisons de moindre énergie. Ces ruptures ne sont donc pas forcément celles qui se font au sein d'une fragmentation réelle. Pour autant, ce type de modèle a fait ses preuves et fait partie des moyens computationnels actuellement utilisés pour interpréter les données de MS/MS de façon automatisée [57]. L'emploi de ce module permet un gain de temps pour interpréter les données de MS/MS des molécules puisque l'obtention de résultats de fragmentation *in silico* de même que la comparaison entre les fragments modélisés et les données expérimentales sont des étapes complètement automatisées.

4.2 Limites pratiques liées aux performances de l'outil en fonctionnement

Actuellement, GAINS est un outil qui permet de traiter de façon automatique et efficace les données de niveau MS. Les prochaines étapes importantes et primordiales pour l'améliorer de façon conséquente concernent les données de niveau MS/MS. Un seul outil de fragmentation *in silico* a été envisagé et son apport n'a pas complètement été évalué.

Premièrement, une limite liée au score de l'outil existe. Metfrag normalise les scores des différents candidats structuraux et propose des scores allant de 0 à 1. Ce qui signifie que si toutes les structures évaluées sont de ''mauvais'' candidats, certaines seront tout de même scorées à 1. Il est impossible de vérifier si le score de 1 attribué par Metfrag est significatif d'une très bonne correspondance entre le spectre expérimental et le spectre *in silico,* ou le moins mauvais parmi des correspondances peu significatives. Ainsi, nous n'avons pas de moyen simple de se focaliser sur les molécules dotées des meilleurs scores (avec un score absolu minimum qui serait à atteindre). Comme Metfrag est intégré à notre outil sous forme de lignes de commandes, nous ne disposons pas de l'interface graphique qui est proposée en ligne (possibilité de traiter un seul composé à la fois) et qui permettrait d'observer les différents spectres. Cette solution ajouterait à l'exploitation une étape manuelle longue.

Deuxièmement, l'utilisation d'outils de fragmentation combinatoire signifie de faire appel à une modélisation de la fragmentation sans aucune intégration des règles de fragmentation tirées d'observations expérimentales sur des molécules. Aussi, nous voulons intégrer dans la démarche

automatisée, la recherche que nous exécutons manuellement pour l'heure et qui concerne les parties glycosylées et acylées des molécules car ces groupements sont très souvent présents dans les molécules naturelles.

Troisièmement, Metfrag permet de classer les différentes structures correspondant à une masse monoisotopique donnée mais il ne distingue pas les isomères, de même qu'on ne les distingue pas lors de traitement manuel.

5. Conclusion

Le premier intérêt de GAINS est d'atteindre un niveau précis et fin de caractérisation. La précision des résultats est liée aux techniques employées. La finesse, c'est-à-dire la capacité à décrire un grand nombre de composés, est liée à l'ensemble des étapes coordonnées pour y arriver. L'enjeu est que l'outil soit efficace. Comparé à un traitement manuel, un gain de temps important est permis par l'automatisation du processus. De plus, l'automatisation permet une gestion des données reproductibles puisque le programme et les actions réalisés par l'outil sont formatés et standardisés. L'intervention de l'opérateur dans le début du processus est réduite. Cela limite les risques d'erreurs potentielles.

Le second intérêt réside dans le fait que GAINS s'appuie sur les analyses en spectrométrie de masse obtenues avec seulement deux injections de l'extrait, dans les deux modes de polarité. De ce fait, on ne procède aucunement à l'isolement des composés avant d'accéder à des informations sur leur identité. Ainsi, on s'affranchit des longues étapes de purification pour mener l'étude.

Avec cette démarche, le niveau d'identification n'est pas formel puisqu'il n'y a ni étapes de purification et caractérisation du composé ni comparaison de plusieurs données expérimentales avec celles de standards. D'après les *Metabolomics Standards Initiative* [12], la correspondance d'au moins deux données expérimentales avec celles d'une standard peut aussi constituer une identification formelle. Toutefois, l'approche qui a été mise en place est destinée à permettre des propositions d'identification fiables et avec un gain de temps dans les processus de profilage métabolique pour obtenir une vision globale du contenu de l'extrait.

Ensuite, pour aller plus loin dans la caractérisation, des étapes supplémentaires peuvent être envisagées, au cas par cas et de façon pertinente sur la base des informations déjà obtenues. En conséquence, via GAINS, le phytochimiste est en capacité de mieux cibler le travail à effectuer au laboratoire pour gérer l'exploitation des données et décrire le contenu d'un extrait. L'outil doit être considéré comme un moyen d'accéder à une vision claire et globale de l'échantillon à étudier pour pouvoir découvrir ce qu'il contient.

La mise en évidence de molécules connues pour l'organisme étudié, aussi appelée déréplication, permet de ne pas perdre de temps à purifier des molécules qui n'avaient pas besoin d'être isolée pour être identifiées. De plus, l'outil permet de mettre en lumière la présence de molécules inconnues. En effet, tout signal chromatographique caractérisé par une masse monoisotopique sans structure candidate ou, dont les données de fragmentation ne correspondent

pas aux candidats envisagés pour la caractériser, est potentiellement une nouvelle molécule naturelle à découvrir.

Suite à la mise en œuvre et aux étapes d'optimisation des paramètres expérimentaux ou de programmation, GAINS est aujourd'hui un outil semi-automatique, fonctionnel permettant de répondre aux objectifs que nous nous étions fixés concernant les gains de traitements des données de spectrométrie de masse de niveau MS. L'intervention de l'opérateur reste nécessaire et importante pour exploiter les données de niveau MS/MS notamment, mais aussi de façon plus générale pour valider les traitements, gérer les exceptions et prendre les décisions d'action à mettre en place pour la suite de la description du contenu de l'extrait. L'outil rend service à l'expert pour aller plus vite et rendre ses études plus pertinentes sans avoir vocation à le remplacer. L'application de GAINS pour la caractérisation phytochimique de trois extraits de plante est présentée dans le chapitre 3.

Chapitre 3
Applications de GAINS
Différentes fonctionnalités appliquées à l'étude d'extraits de plantes

1. Introduction

Ce troisième chapitre aborde les résultats obtenus grâce à l'utilisation de l'outil GAINS pour caractériser différents extraits de plante. Il sera présenté les différentes fonctions que l'outil peut proposer notamment grâce à sa connexion à la base de données de produits naturels GPDB. Nous verrons qu'en fonction des cas, la stratégie globale peut varier mais que GAINS reste un support central. Ainsi, trois études se différentiant par leurs objectifs, leur déroulement et leur résultante sont détaillées en guise d'exemples d'utilisations de GAINS. Deux des études ont été réalisées à la demande d'industriels qui avaient besoin de connaître plus précisément le contenu d'extraits.

La première partie du chapitre concernera un extrait de tulipe noire qui entre dans la composition d'un ingrédient cosmétique que Greenpharma a développé pour Alès Groupe. Ce dernier a mandaté Greenpharma pour identifier des chémo-marqueurs de l'extrait. Dans cet exemple, nous détaillerons l'utilisation de GAINS combiné à l'hydrolyse acide de l'extrait pour répondre à l'objectif.

La seconde partie concernera l'étude d'un extrait de fleurs de primevère officinale que Greenpharma a voulu décrire d'un point de vue phytochimique dans le but d'évaluer son potentiel en tant qu'actif cosmétique. L'extrait de fleurs a été décrit selon les deux méthodes parallèles suivantes : l'approche dite classique d'isolement et de purification des composés principaux pour effectuer leur identification par HRMS et RMN et, l'approche plus globale faisant appel à des techniques émergentes de traitement automatisé de données que constitue GAINS. Avec ce cas d'étude, nous pourrons comparer les deux approches entre elles, décrire leurs intérêts et limites respectives.

La troisième et dernière partie sera destinée à mettre en évidence l'intérêt de la base de données GPDB, notamment du fait de son organisation, lorsqu'elle est associée à GAINS pour la recherche de familles phytochimiques particulières. L'étude d'un extrait de réglisse illustrera cette complémentarité puisque l'objectif fixé par le groupe ROCHER était d'avoir une vue d'ensemble de la composition phytochimique de l'extrait et de mettre en évidence la présence ou absence d'isoflavones.

2. Etude d'un extrait de pétales de Tulipe noire : Identification de chémo-marqueurs et intérêt de l'hydrolyse acide de composés

2.1 Contexte de l'étude et objectifs

L'étude de l'extrait hydro-alcoolique (EtOH 30 %) de pétales de tulipe noire entrant dans la composition d'un ingrédient cosmétique, la Dumaflorine, a été menée. L'objectif principal était d'identifier un chémo-marqueur de l'extrait. Par la suite, la teneur de ce composé pourrait entrer dans les spécifications de l'ingrédient. Dans le cadre du développement de l'ingrédient pour Alès Groupe, Greenpharma a proposé d'effectuer des analyses en chromatographie liquide couplée à la spectrométrie de masse haute résolution (LC-HRMS) associées à l'interrogation de la base GPDB sur l'ensemble des molécules décrites dans *Tulipa gesneriana*. Compte tenu de la mise en évidence antérieure de la présence de flavonoïdes de cette espèce, une analyse spécifique des flavonoïdes présents dans l'extrait de tulipe noire a complété l'étude LC-HRMS. Cette famille phytochimique est très riche, de nombreux isomères existent et ces composés sont souvent présents sous forme glycosylée. Une analyse HPLC avant et après hydrolyse de l'extrait de tulipe noire a été réalisée pour définir les noyaux flavonoïdiques présents.

2.2 Présentation de la Tulipe noire

La tulipe noire '*Queen of the Night'*, dont un extrait a été étudié, est la plus sombre des tulipes dites ''noires''. Reconnue pour ses qualités ornementales avec une grande tige pouvant atteindre 60 cm, la Tulipe '*Queen of the Night'* est un cultivar (variété obtenue par culture) apparenté à *Tulipa gesneriana* L. (tulipe des jardins ou tulipe des fleuristes) comme c'est le cas pour de nombreuses variétés de tulipes créées et dédiées à l'horticulture et à l'ornement des jardins. Les tulipes sont des plantes vivaces qui poussent à partir d'un bulbe. Elles appartiennent à la famille des Liliacées [107]. Le travail de caractérisation a été fait en considérant les informations disponibles au sujet de *Tulipa gesneriana*.

Suite à une recherche au sein de la base GPDB (et après veille bibliographique sur la plante afin d'assurer un référencement à jour des molécules), les 74 molécules connues et présentes dans les tulipes du genre *gesneriana* (*Tulipa gesneriana* ou cultivars) ont été recensées parmi lesquelles des flavonoïdes, des anthocyanes, des diterpénoïdes (gibbérellines) ou de petits acides organiques. C'est une plante plutôt ordinaire mais relativement peu décrite d'un point de vue moléculaire.

2.3 Hydrolyse acide de l'extrait et analyse par HPLC-UV

L'extrait est analysé en HPLC-UV avant et après une étape d'hydrolyse acide (conditions d'hydrolyse en 4.5.1). Avant hydrolyse, on observe les flavonoïdes aglycones présents. Après l'étape d'hydrolyse, on met en évidence les génines correspondant aux flavonoïdes glycosylés. Les chromatogrammes de l'extrait avant et après hydrolyse sont présentés en Figure 25 (conditions d'analyse en 4.5.3).

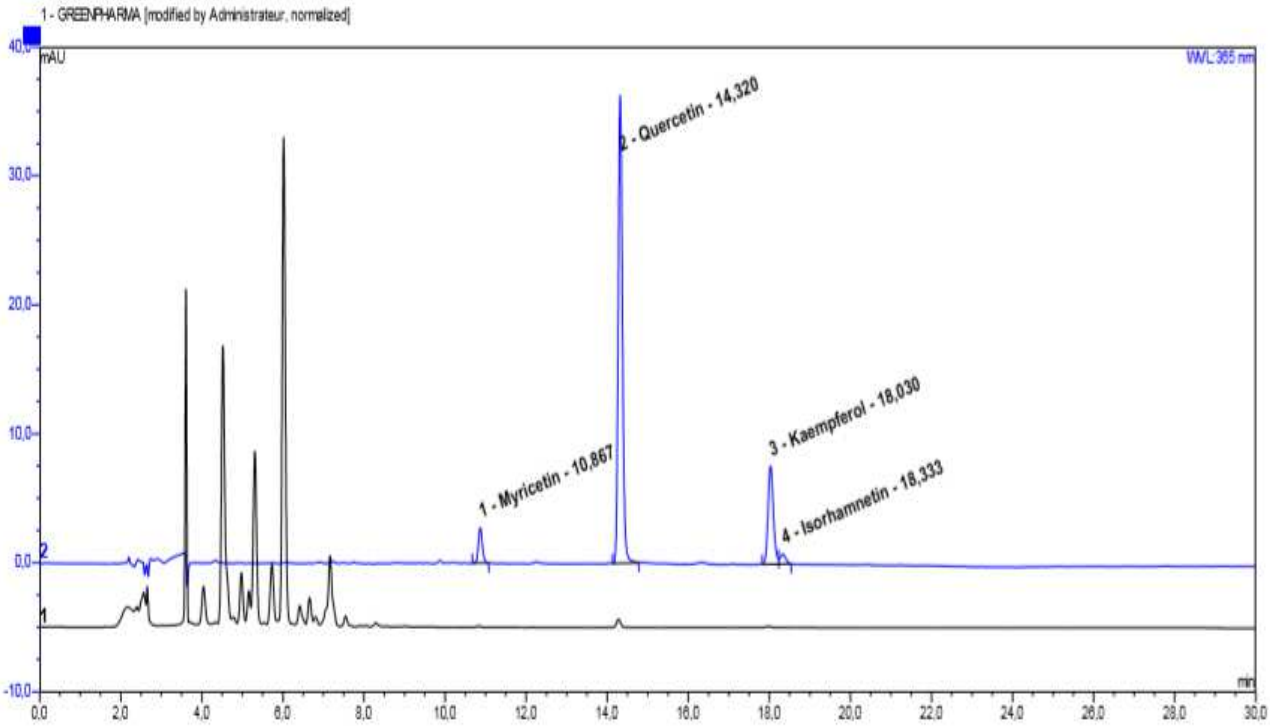


Figure 25 Chromatogrammes de l'extrait de tulipe noire avant hydrolyse (tracé noir) et après hydrolyse (tracé bleu) en UV 365 nm

L'extrait analysé avant hydrolyse est constitué principalement de composés sous forme glycosylée élués en début de gradient. Après hydrolyse, les aglycones ont été libérés des sucres comme le montre le chromatogramme bleu. L'extrait de tulipe noire contient sous forme libre et glycosylée les noyaux flavonoïdiques suivants : la quercétine, le kaempférol, la myricétine et l'isorhamnétine. D'après des données semi-quantitatives, la quercétine représente 70% de la totalité des noyaux flavonoïdiques présents dans l'extrait. Le kaempférol est le deuxième noyau le plus présent (environ 20%) tandis que la myricétine et l'isorhamnétine interviennent chacune dans environ 5% des composés formés sur un noyau flavone/flavonol. Des composés non flavonoïdiques sont observés après hydrolyse. Un premier pic associé à un temps de rétention de 4,33 minutes possède un spectre UV caractéristique d'un anthocyane (maxima à 276 et 525 nm). Un second pic (temps de rétention = 8,47 minutes) dont l'unique maximum d'absorbance est 290 nm constitue probablement un tanin. Les pics de ces composés qui n'absorbent pas à 365 nm n'apparaissent pas sur la Figure 25.

2.4 Analyses LC-HRMS et résultats obtenus avec GAINS

L'extrait hydro-alcoolique de pétales de tulipe noire a été analysé en LC-DAD-HRMS à 1000 ppm dans les modes positif et négatif d'ionisation. Les conditions d'analyses sont détaillées en Annexe 3.3.

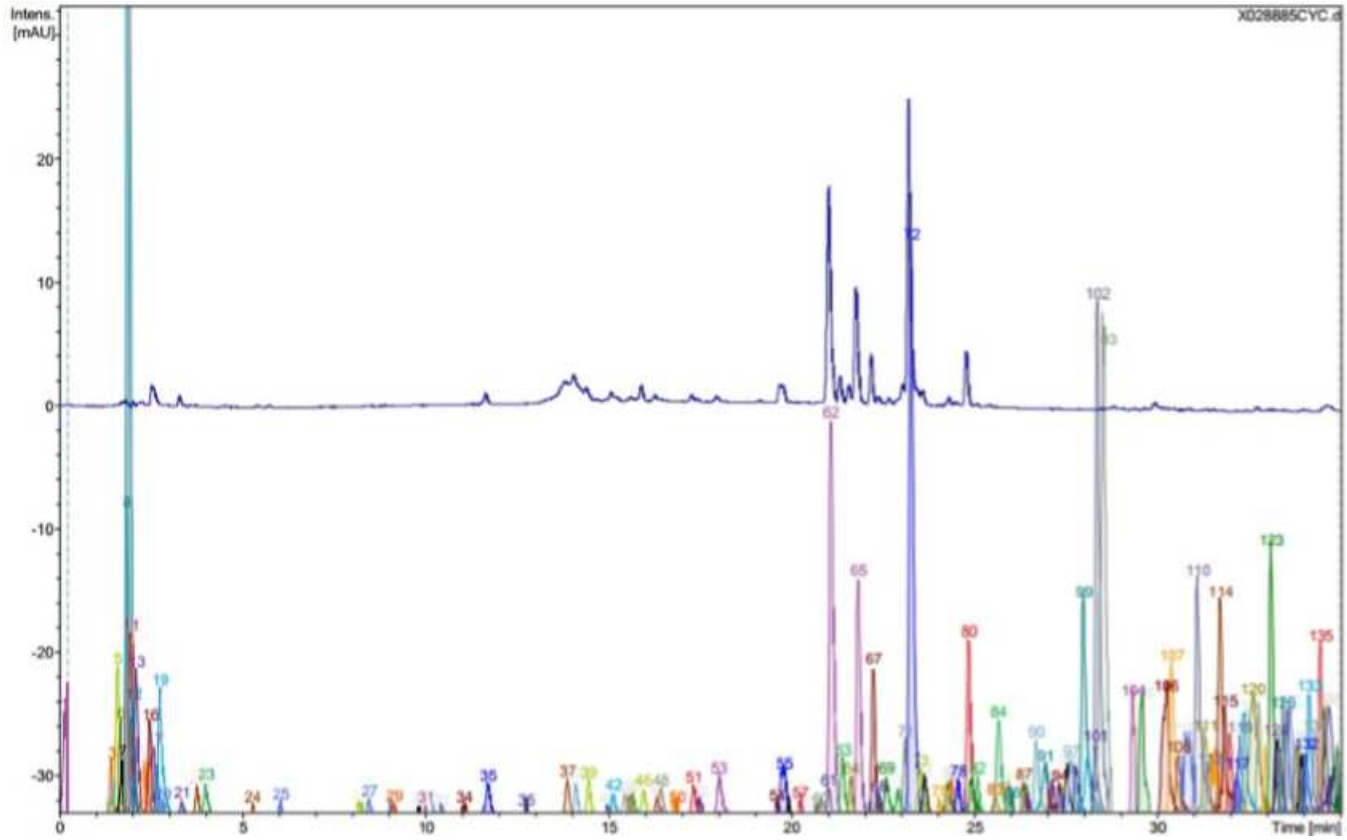


Figure 26 Chromatogrammes UV 254 nm (tracé bleu) et MS-ESI$^+$ de l'extrait de tulipe noire à 1000 ppm

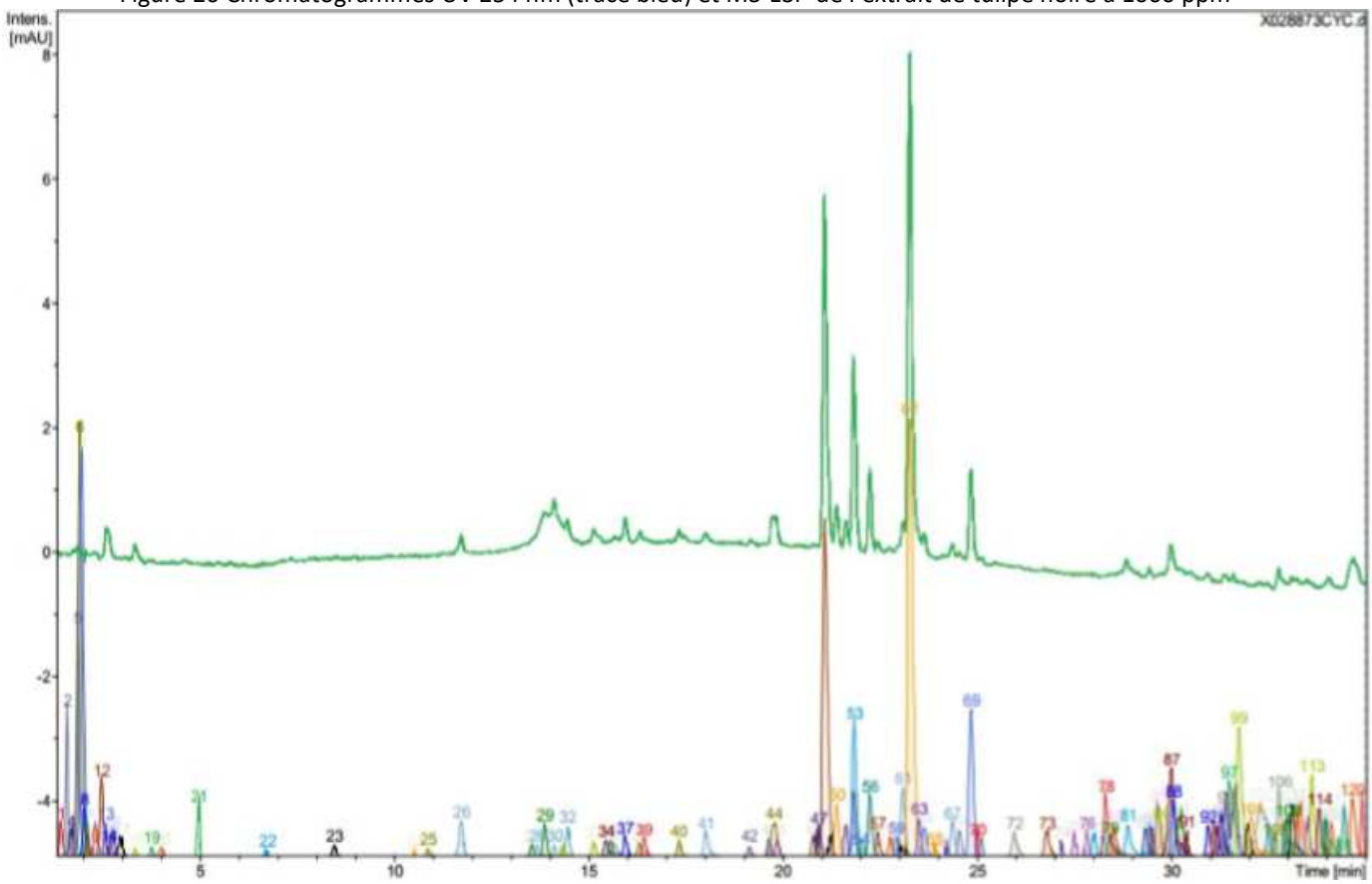


Figure 27 Chromatogrammes UV 254 nm (tracé vert) et MS-ESI$^-$ de l'extrait de tulipe noire à 1000 ppm

L'extrait possède un riche profil UV à 254nm et répond de façon satisfaisante en spectrométrie de masse que ce soit en mode positif ou négatif d'ionisation.

2.5 Déréplication et proposition d'identification de chémo-marqueurs

Pour connaître le contenu de l'extrait, nous avons mis en place le profilage de l'extrait en utilisant GAINS pour permettre sa description par déréplication et pour proposer des identifications de chémo-marqueurs.

2.5.1. Note explicative de la numérotation des pics dans les tableaux de résultats

Les pics ont été numérotés dans l'ordre de leur rétention (Figure 26 et Figure 27). Les numérotations des modes positif et négatif d'ionisation sont indépendantes. En conséquence, pour un même temps de rétention, un pic détecté en mode positif peut avoir un numéro différent du pic détecté en mode négatif. Pour distinguer ces deux numérotations, nous avons ajouté à la suite du numéro du pic, la lettre « P » en mode positif et la lettre « N » en mode négatif.

2.5.2. Résultats

Les masses monoisotopiques sont proposées grâce à la concordance de plusieurs ions issus d'un mode d'ionisation ou des deux pour un composé donné. Les maxima UV lorsqu'ils sont mesurables sont également indiqués. Dans certains cas, les données de fragmentation (obtenues en Auto MS/MS) nous ont permis de confirmer le type de génines ou monosaccharides en présence.

Dans un premier temps, les masses monoisotopiques ont été confrontées aux données de la base GPDB : lorsqu'une structure pouvait expliquer la masse d'un pic, les données MS/MS ont été utilisées pour confirmer ou infirmer la proposition. Dans un second temps, des propositions d'identification ont été faites à partir de l'ensemble des données (UV, masse monoisotopique et fragmentation). L'exploitation des données MS/MS a été gérée manuellement. Le traitement a été fait de façon manuelle car la détermination automatiquement de la nature des parties aglycones et des parties glycosylées des molécules n'est pas encore intégrée dans l'outil GAINS.

Le Tableau 13 présente les données correspondant aux trois composés principaux ou celles des pics qui ont un candidat connu pour le genre *gesneriana* (déréplication ciblée). Les lignes sur fond bleu sont celles des 3 composés majoritaires en UV tandis que les propositions en **rouge** correspondent aux molécules décrites dans la littérature pour *Tulipa gesneriana*. L'ensemble de ces pics présentent des données de fragmentations caractéristiques ayant rendu possible la confirmation de génine ou de monosaccharides en présence. Pour chaque pic chromatographique (un zoom de la zone entre 10 et 30 minutes est présentée en Figure 31), le tableau regroupe les informations suivantes : le temps de rétention, la masse monoisotopique, le maximum d'absorbance du composé, la génine et/ou les monosaccharides qui le constitue et enfin la

proposition d'identification. L'analyse des noyaux flavonoïdiques après hydrolyse permet d'indiquer la nature exacte de chaque flavonol, ainsi les types de noyaux tetrahydroxyflavone, pentahydroxyflavone ont été remplacés respectivement par le kaempférol et la quercétine (*cf*. respectivement Figure 29 et Figure 30) dans la colonne de proposition d'identification.

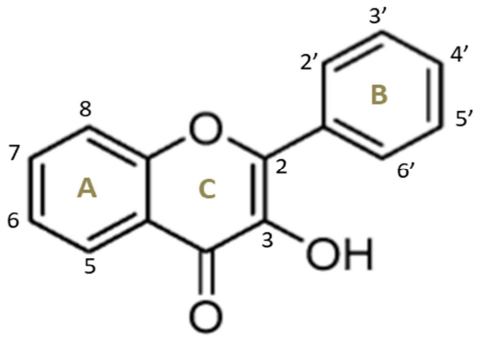


Figure 28 Structure d'un noyau flavonol dont la caractéristique est l'hydroxylation en C3 du cycle

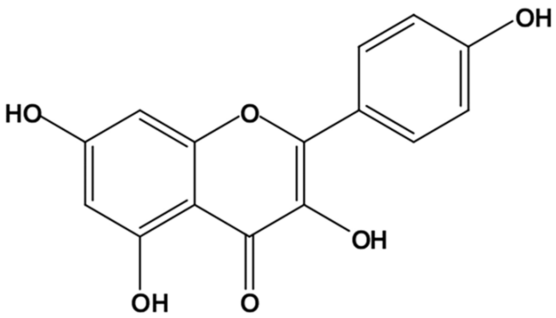


Figure 29 Structure du kaempférol (tétrahydroxyflavone)

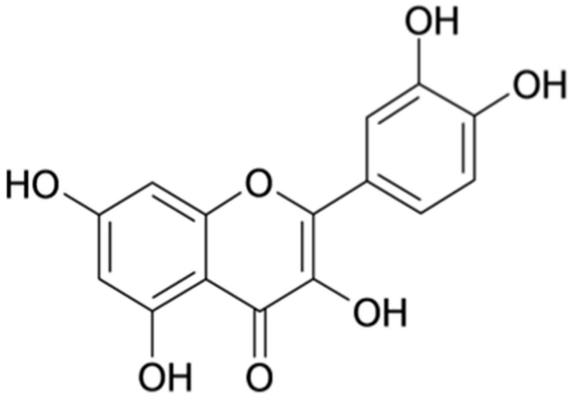


Figure 30 Structure de la quercétine (pentahydroxyflavone)

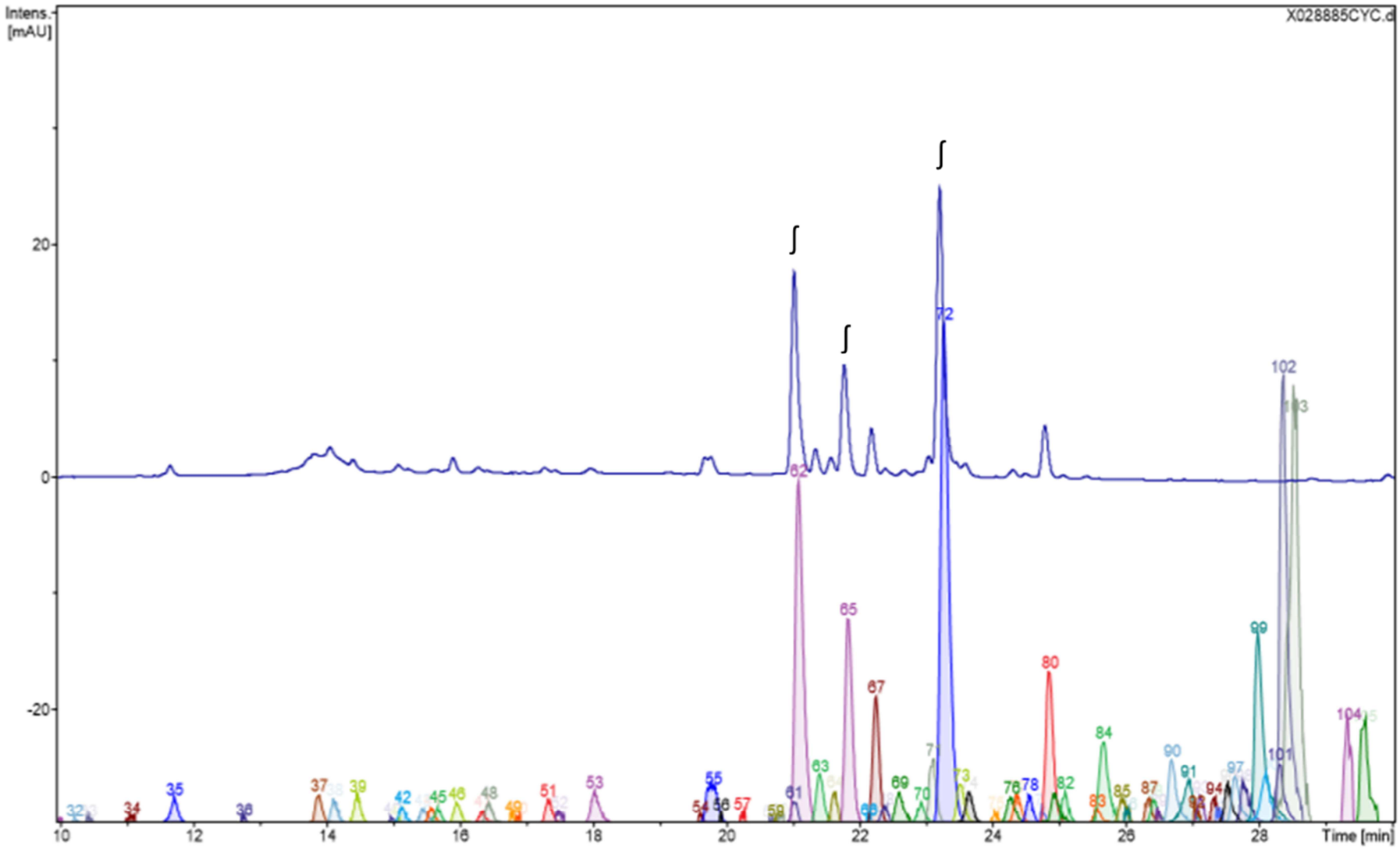


Figure 31 Chromatogrammes UV 254 nm (tracé bleu) et MS-ESI$^+$ de l'extrait de tulipe noire. Zoom de 10 à 30 minutes

t_R (minutes)	n°pic mode positif	n°pic mode négatif	Masse monoisotopique ($g.mol^{-1}$)	Maxima d'absorbance (nm) (>250nm)	Partie aglycone	Partie glycosylée	Proposition d'identification
15,95	46P	37N	**772,205841**	267 ; 360	Penta-hydroxyflavone	Hexose ; hexose ; desoxyhexose	**Quercétin-3-O-rutinoside-7-O-glucoside**
17,33	51P	40N	**756,211056**	266 ; 360	Tetra-hydroxyflavone	Hexose ; hexose ; desoxyhexose	**Kaempférol 7-O-glucoside-3-O-rutinoside**
21,09	62P	48N	**756,211446**	257 ; 355	Penta-hydroxyflavone	Hexose ; desoxyhexose ; desoxyhexose	Quercétine triglycosylé (1 hexose, 2 desoxyhexoses)
21,84	65P	53N	**596,137334**	260 ; 357	Penta-hydroxyflavone	Hexose_Aldopentose (terminal)	Quercétine diglycosylé (hexose, aldopentose terminal)
23,28	72P	62N	**610,153079**	255 ; 358	Penta-hydroxyflavone	Hexose_Desoxyhexose (terminal)	Quercétine diglycosylé (hexose ; desoxyhexose)
23,65	74P	64N	**464,094553**	267 ; 351	Penta-hydroxyflavone	Hexose	**Quercétin-3-O-glucoside**
24,85	80P	69N	**594,157980**	255 ; 360	Tetra-hydroxyflavone	Hexose_desoxyhexose (terminal)	**Kaempférol 3-O-rutinoside**
25,10	82P	71N	**448,099891**	267 ; 355	Tetra-hydroxyflavone	Hexose	**Kaempférol 3-O-glucoside ou Kaempférol 7-O-glucoside**

Tableau 13 Données analytiques avec éléments structuraux déterminés et propositions d'identification des trois pics principaux de l'extrait de tulipe noire ou des pics qui ont des candidats attribués par déréplication ciblée dans *Tulipa gesneriana*.

Les éléments déterminés pour la partie glycosylée des structures sont déterminés par l'exploitation manuelle des données de MS et MS/MS. Le raisonnement suivi est détaillée dans le paragraphe 2.5.3.

Les composés mis en évidence dans l'extrait de tulipe noire sont principalement des flavonols glycosylés. On retrouve cinq pics pour lesquels des candidats structuraux sont connus pour la plante : quercétin-3-O-rutinoside-7-O-glucoside (rutine-7-O-glucoside), kaempférol 7-O-glucoside-3-O-rutinoside (populnin 3-O-rutinoside), quercétine 3-O-glucoside (isoquercitrin), kaempférol-3-O-rutinoside ainsi que kaempférol 3-O-glucoside (astragalin) ou kaempférol 7-O-glucoside (populnin). Pour le composé 82P/71N, deux structures données dans la littérature et concordant avec les analyses sont possibles sans que les données de HRMS puissent en privilégier une par rapport à l'autre. En effet, le kaempférol 3-O-glucoside (astragalin) et le kaempférol 7-O-glucoside (populnin) sont deux isomères de substitution. Toutefois, une requête au sein de la totalité de la GPDB montre que d'autres structures isomères seraient également possibles pour le composé 82P/71N. Pour les trois pics majoritaires, aucun candidat n'est attribué par déréplication ciblée dans *Tulipa gesneriana*.

2.5.3. Raisonnement tenu à partir des données analytiques et apport des données MS/MS : exemple détaillé pour le composé le plus intense

Les données de fragmentation des molécules ont été exploitées –manuellement- de manière à atteindre un niveau supérieur d'informations concernant les structures. L'exemple suivant permet d'illustrer le travail effectué pour obtenir les informations concernant les génines et les monosaccharides constituant chaque composé. Il est représentatif de ce qui a été fait pour l'ensemble de l'étude. L'exemple correspond au composé 72P/62N majoritaire en UV. On sait que c'est un flavonoïde grâce à son spectre UV (λ_{max} 255, 358 nm).

Composé 72P / 62N		
Masse monoisotopique **610,15307924** g.mol^{-1}		
Mode d'ionisation	Positif	Négatif
Ion pseudo-moléculaire	$[M+H]^+$ = 611,160663	$[M+H]^-$ = 609,145767
Fragments MS/MS	$[Génine + H]^+$= 303,048974	$[Génine – H]^-$ = 301,033475

Tableau 14 Résumé de données MS et MS/MS observées pour le composé apparenté aux pics 72P/62N

La masse moléculaire de 610 g.mol^{-1} indique que le flavonoïde est glycosylé, cela est cohérent avec le fait que l'extrait contienne majoritairement des formes glycosylées de flavonoïdes. Les spectres MS et MS/MS du mode positif correspondant au pic 72P sont présentés en Figure 32 et Figure 33 :

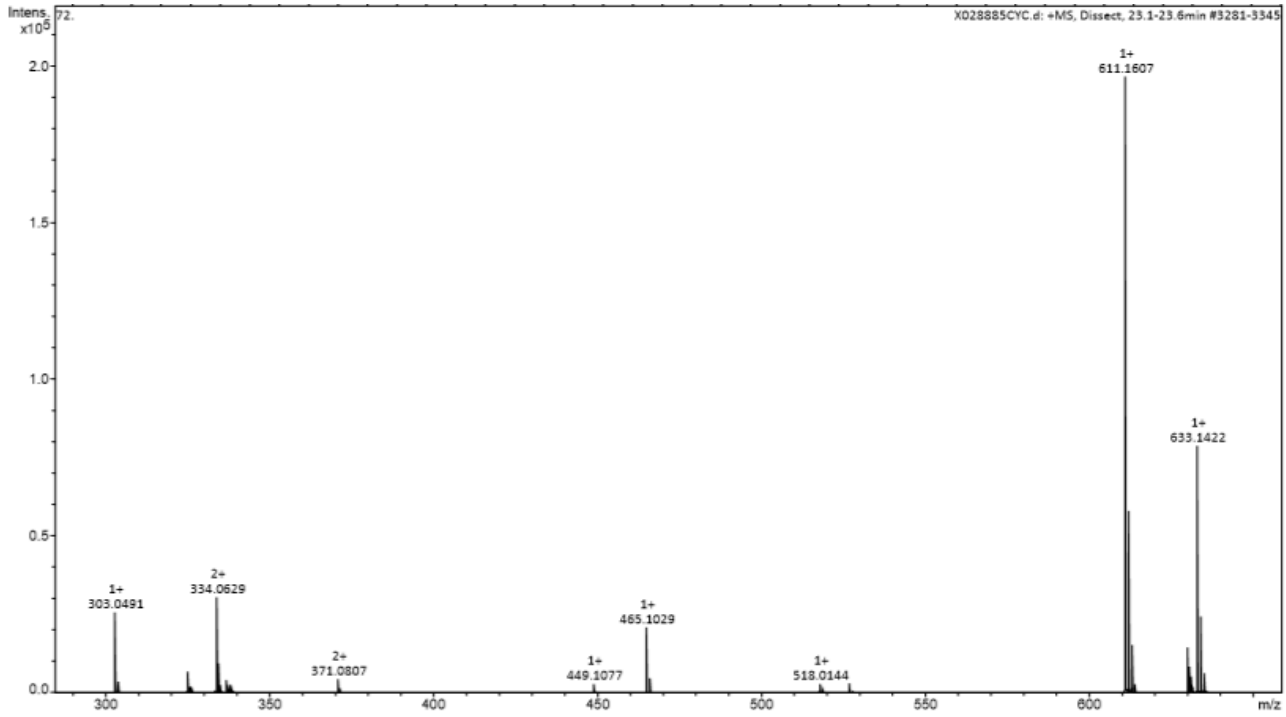


Figure 32 Spectre MS-ESI positif du pic 72P

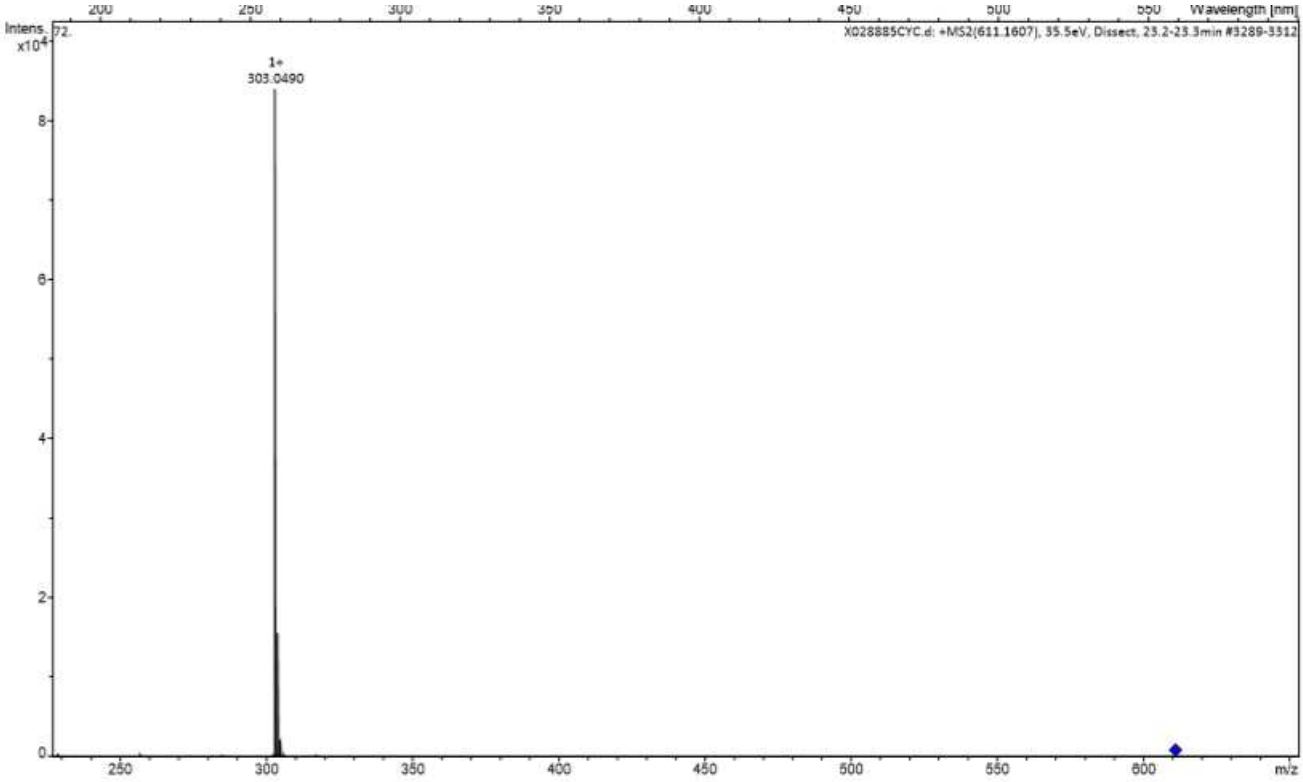


Figure 33 Spectre MS/MS du pic 72P à partir de l'ion précurseur $[M+H]^+$

Les données obtenues en HRMS indiquent que la génine a une masse de 302,04122 $g.mol^{-1}$ ce qui correspond à la masse monoisotopique d'une pentahydroxyflavone dans le groupe des aglycones de flavonoïdes. D'après l'analyse de l'extrait après hydrolyse, on sait que c'est la quercétine. La différence de masse entre l'ion pseudo-moléculaire et l'ion de la génine permet de déterminer que les monosaccharides portés sont un hexose et un désoxyhexose.

Un autre ion de la molécule a été fragmenté en Auto MS/MS : il s'agit de l'adduit sodé de la molécule $[M+Na]^+$ à *m/z* 633,1422. Les ions fragments observés sur le spectre MS/MS sont *m/z* 331,099218 et *m/z* 487,083485. Ils correspondent comme le montre la Figure 34 d'une part à la perte de la génine (on observe le dissaccharide complexé à l'ion sodium) et d'autre part à la perte d'un seul des deux sucres.

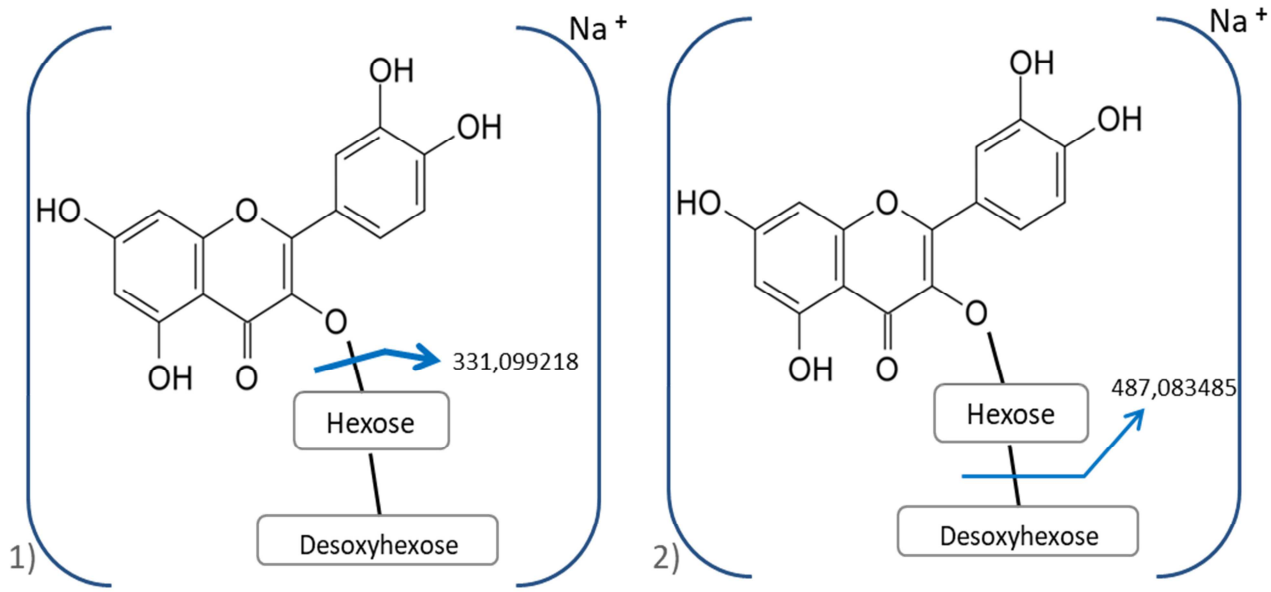


Figure 34 Fragmentation de l'adduit sodé d'un flavonoïde diglycosylé.
Deux ruptures distinctes sont observées en MS/MS: 1) perte de la génine et 2) perte du sucre terminal

De ces observations, on déduit que les deux oses sont enchaînés et que c'est l'hexose qui est lié à la génine, ce qui fait du desoxyhexose, le sucre terminal.

Sur la Figure 34, la génine de type flavonol est substituée par les sucres en position 3 du cycle C. Ici, la structure représentée permet d'illustrer ce qui est observé avec un exemple. Cette substitution a été choisie car les flavonols substitués en C3 sont plus communs que ceux substitués en C7. Le raisonnement aurait été le même avec un flavonol substitué en C7.

2.5.4. Déréplication ciblée et non ciblée pour les trois composés majoritaires et apport de Metfrag

Suite à la recherche de candidats, il est intéressant de noter que les trois pics majoritaires de l'extrait correspondent tous à des composés de structures inconnues pour la plante. Pour le pic 62P/48N, un candidat de même masse monoisotopique est connu dans *Tulipa gesneriana* mais ne correspond pas du tout en termes d'éléments structuraux. Ceci indique que le fait de déterminer les masses des pics d'un profil aussi précises soient elles, ne suffit pas pour faire une proposition d'identification.

Lorsque la recherche de candidats est étendue à la totalité de la base, on se retrouve confronté à un grand nombre de candidats potentiels. En s'appuyant sur les éléments structuraux déterminés grâce à l'étude de l'extrait, on réduit de façon conséquente le nombre de candidats possibles.

Le Tableau 15 résume ce qui ressort de la requête au sein de la base GPDB concernant les candidats structuraux pouvant correspondre pour chacun des 3 pics majoritaires.

	Nombre de candidats affiliés à *Tulipa gesneriana*		**Nombre de candidats parmi la base GPDB**		
Critère de recherche :	Masse	Masse + éléments structuraux*	Masse	Masse + éléments structuraux*	Metfrag**
Pics 62P/48N	1	Aucun	50	11	4 candidats, tous correspondent aux éléments
Pics 65P/53N	Aucun	Aucun	25	10	9 candidats, 1 candidat ne correspond pas (le sucre terminal n'est pas le bon)
Pics 72P/62N	Aucun	Aucun	101	18	Aucun candidat (tous les scores sont à zéro)

Tableau 15 Résultats de recherche de candidats structuraux pour identifier les chémo-marqueurs de l'extrait de tulipe noire. * : éléments structuraux déterminés par l'étude (nature de la génine et des sucres). ** : Metfrag fragmente des candidats de la base obtenus après requête via le critère de masse monoisotopique, on détaille le nombre de candidat ayant obtenu un score de 1 et parmi eux, ceux correspondant aux éléments structurant déjà déterminés)

En s'appuyant sur l'outil Metfrag, tous les candidats attribués à un pic, suite à une requête dans la base GPDB, sont classés en fonction de la concordance entre le spectre prédit *in silico* et le spectre expérimental. Si on considère uniquement la (ou les) structure(s) qui arrivent en meilleur rang du classement, on restreint de fait le nombre de candidats. Sur l'exemple des 3 chémo-marqueurs de l'extrait, on observe que Metfrag peut apporter un supplément d'informations. C'est le cas pour les pics 62P/48N, on obtient 4 candidats (score = 1) qui correspondent aux éléments structuraux déjà déterminés (génine, type de monosaccharides). Ces 4 structures ont les sucres enchainés et une génine substituée en une seule position. Ainsi, ce type de candidats semble privilégié à l'issue de la comparaison des spectres modélisés avec le spectre expérimental qu'a effectué Metfrag. Cet élément n'avait pu être déterminé lors de l'analyse manuelle des données de MS/MS.

2.6 Conclusion

L'étude de l'extrait de tulipe noire entrant dans la composition d'un ingrédient cosmétique a permis d'investiguer sa composition phytochimique avec, pour but principal d'identifier des chémo-marqueurs. Le travail était axé autour de l'examen des noyaux flavonoïdiques et d'analyses en LC-HRMS de l'extrait. Les

analyses en HPLC-UV donnent accès à de précieuses informations : elles mettent en évidence la présence de trois classes phytochimiques (anthocyane, flavonoïde et tanin). Ensuite, grâce aux analyses avant et après hydrolyse acide, on a mis en évidence que les flavonoïdes sont principalement présents sous des formes glycosylées, et que les noyaux flavonoïdiques majoritaires sont : la quercétine et le kaempférol et moins fréquemment, la myricétine et l'isorhamnétine. Ces informations sont concordantes avec les analyses issues du couplage LC-HRMS qui mettent en lumière la présence de nombreux flavonoïdes glycosylés. Ces derniers constituent le cœur de l'extrait en nombre et en intensité. L'extrait contient des flavonoïdes portant de 1 à 4 monosaccharides. Les principaux composés de l'extrait sont identifiés comme des flavonoïdes glycosylés. Ceci correspond aux données de la littérature pour *Tulipa gesneriana* dans laquelle cette famille de composés est décrite.

Les molécules majoritaires de cet extrait sont : **i)** une quercétine triglycosylée dont un hexose et deux desoxyhexoses (pics 62P/48N), l'arrangement des trois monosaccharides n'est pas déterminé, **ii)** une quercétine diglycosylée portant un hexose lié à un aldopentose terminal (pics 65P/53N) et **iii)** une quercétine diglycosylée portant un hexose lié à un desoxyhexose terminal (pics 72P/62N). Si ces composés n'ont jamais été décrits au sein de *Tulipa gesneriana*, les masses monoisotopiques de ces composés correspondent à un grand nombre de candidats lorsqu'on étend la recherche à la totalité de la base GPDB. En s'appuyant sur les éléments structuraux déterminés grâce à l'analyse fine des données de MS/MS, on parvient à réduire le nombre de propositions. Toutefois, différencier des isomères de position reste impossible avec ces données seules. Dans le cas d'une famille aussi riche que les flavonoïdes, le nombre de possibilités de candidats reste difficile à réduire avec les seules données UV et HRMS. Toutefois, si l'on décidait de les identifier formellement via une purification et des analyses RMN, les informations collectées seraient des bases précieuses de travail et simplifierait probablement la caractérisation.

3. Etude phytochimique d'un extrait de fleurs de Primevère officinale : de l'approche classique à l'approche moderne

3.1 Contexte de l'étude

Au sein de Greenpharma, un intérêt particulier est porté aux plantes locales, répandues dans l'environnement naturel de la région Centre-Val de Loire. Le choix d'étudier les fleurs de primevère officinale appartenant à cet environnement a été fait selon de multiples critères à la fois botaniques, scientifiques et marketing.

3.2 Description de la plante

La primulacée communément appelée Coucou correspond à la primevère vraie (*Primula veris* L.) aussi dite primevère officinale (*Primula officinalis* (L.) Hill.). C'est une primevère non rare dans les prairies d'Europe et d'Asie, qui se développe dans les climats tempérés à froids de l'hémisphère nord [108,109]. En région Centre, c'est une plante très bien représentée. Plante herbacée vivace, elle fleurit au printemps : ses fleurs jaune vif se développent dans un calice (ensemble des sépales, élément de protection externe de la fleur) en forme de tube et forment une hampe.

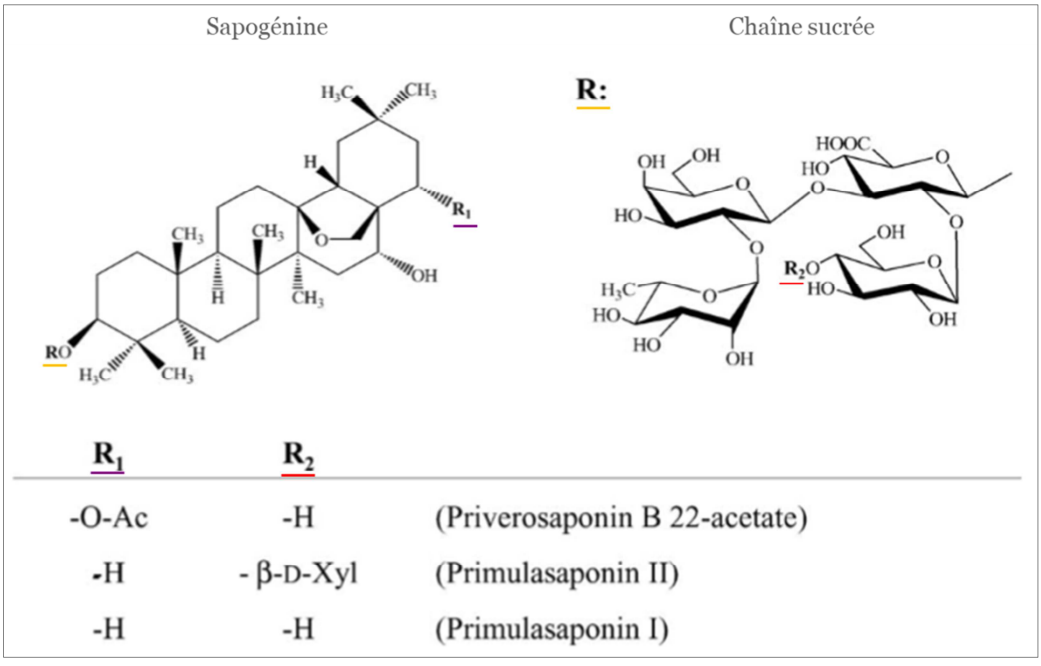
Figure 35 *Primula veris* en fleurs à Orléans, Loiret, Mars 2018.
A gauche : à l'état naturel, à droite : hampe florale

Les fleurs et les racines de la primevère officinale sont les principaux organes utilisés. Les premières figurent dans la pharmacopée française et les secondes sont inscrites à la pharmacopée européenne. De façon traditionnelle, la plante a été utilisée dès le moyen-âge pour traiter la goutte et les maux de tête, puis ensuite dans les cas de bronchites, rhumes et toux. Actuellement, elle reste employée pour gérer ces types de troubles respiratoires grâce à ses effets sécrétolytiques et sécrétomoteurs [109] (respectivement qui fluidifie les sécrétions et qui aide à les évacuer) et en tant qu'agent expectorant (agent qui augmente le rejet

des sécrétions) [7]. On notera que fleurs et racines sont indiquées pour le traitement d'appoint adoucissant (pour assouplir la peau et la rendre douce) et antiprurigineux (destiné à lutter contre les démangeaisons) des affections de la peau de type crevasses, gerçures d'après l'agence nationale de sécurité du médicament et des produits de santé (ANSM [110]). En cosmétique, l'ingrédient *Primula veris extract* (nom INCI*) est catégorisé comme agent d'entretien de la peau.

Le genre *Primula* compte plusieurs centaines d'espèces dont certaines ont un effet allergisant du fait de la présence de quinones allergisantes (primine principalement) dans les trichomes de la plante (poils développés en surface de feuilles, tiges ou racines) [108,112,113]. Jean Bruneton mentionne l'absence de la primine dans les parties souterraines mais n'évoque pas les autres organes de la plante [7]. Il apparaît que les primevères communes et notamment *Primula veris* ne sont pas concernées par ce problème [112].

D'un point de vue phytochimique, la fleur est connue pour contenir des flavonoïdes avec principalement des hétérosides du kaempférol et de la quercétine, des caroténoïdes (pigments) ainsi que des saponosides triterpéniques [7]. Ces dernières sont responsables des propriétés expectorantes de la plante avec notamment la primulasaponin 1 et primulasaponin 2 [114,115] (structures en Figure 36). Les parties souterraines (rhizome et fines racines) renferment des sucres, des glycosides d'esters salicyliques tels que la primulavérine et la primevérine) et des saponosides (mêmes saponosides que celles des fleurs ainsi que la priverosaponin-B-22 acétate) qui sont présentes en quantités importantes (3 à 10 %) [7,109,116].



R_1	R_2	
-O-Ac	-H	(Priverosaponin B 22-acetate)
-H	- β-D-Xyl	(Primulasaponin II)
-H	-H	(Primulasaponin I)

Figure 36 Structures des saponosides décrites dans *Primula veris*, illustration d'après [109]

* INCI : *International Nomenclature of Cosmetics Ingredients.* Le système INCI permet une nomenclature réglementaire qui permet de lister les produits contenus dans la formule d'un produit cosmétique de façon standardisée dans tous les pays. Le système INCI est géré et maintenu par le *Personal Care Products Council* aux Etats-Unis [111].

3.3 Etude préliminaire de l'extrait hydro-alcoolique de fleurs de Primevère officinale

Nous voulions travailler l'extrait hydro-alcoolique obtenu avec 60% d'éthanol (EtOH 60%) de parties de fleurs de Coucou. Constituant un extrait plutôt polaire du fait du solvant d'extraction, nous l'avons principalement analysé en HPLC en phase inverse (RP). Des analyses HPLC de l'extrait en NARP, HILIC et RP ont confirmé que le profil RP était le plus riche. D'autre part, les analyses en HPLC-RP-UV ont rapidement mis en évidence la présence de flavonoïdes grâce à leurs spectres UV caractéristiques. Une des premières expériences a donc été de déterminer la nature des noyaux flavonoïdiques présents et d'évaluer leur teneur.

Pour cela, l'extrait est analysé en HPLC-UV avant et après une étape d'hydrolyse acide. Les chromatogrammes de l'extrait avant et après hydrolyse sont présentés en Figure 37 et les résultats de dosage des génines dans le Tableau 16. Les conditions d'hydrolyse et d'analyse se trouvent en annexe 4.5.1. On détermine la nature des génines par comparaison des temps de rétention et des spectres UV des composés obtenus après hydrolyse à ceux des standards de flavonoïdes.

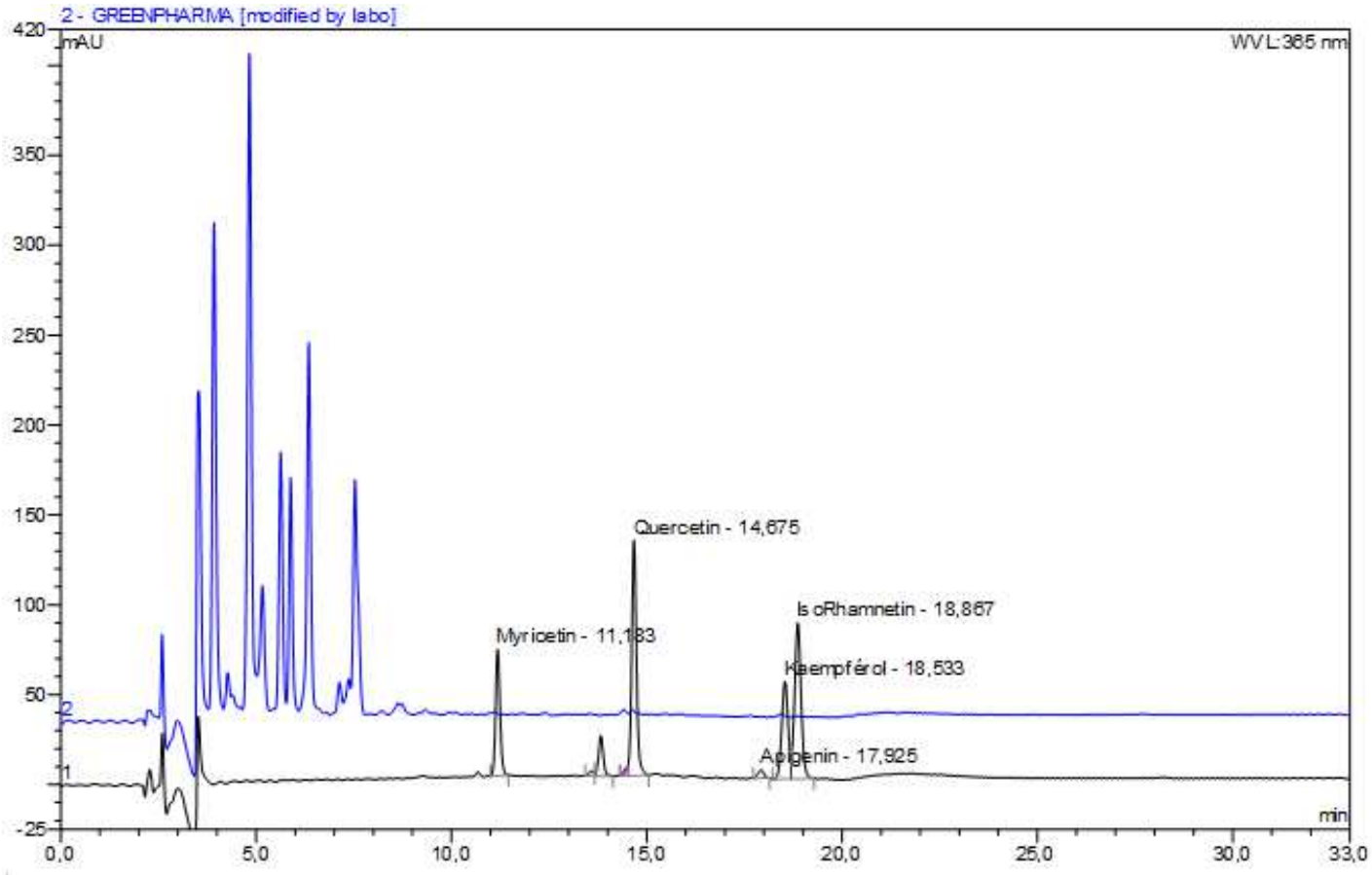


Figure 37 Chromatogrammes (UV 365 nm) de l'extrait de *Primula veris* avant hydrolyse (bleu) et après hydrolyse (noir).

Extrait	Condition	Teneur en apigénine	Teneur en isorhamnétine	Teneur en kaempférol	Teneur en quercétine	**Teneur totale**
EX1GJ4098	Avant hydrolyse	ND	ND	ND	0,0005 %	**0,0005 %**
EX1GJ4098	Après hydrolyse	0,15 %	6,80 %	1,32 %	2,55 %	**10,8 %**

Tableau 16 Dosage en HPLC-UV de flavonoïdes avant hydrolyse et après hydrolyse de l'extrait EtOH 60 % de *Primula veris*.

La première information importante retirée des profils chromatographiques et de l'hydrolyse acide est que l'extrait contient très peu de flavonoïdes sous forme libre mais beaucoup sous forme glycosylée. La teneur totale en flavonoïdes avoisine 11 %. En sachant que les flavonoïdes sont sous forme glycosylés (notamment di et tri-glycosylations d'après les données de la littérature sur la plante), on peut appliquer un coefficient multiplicateur de 2,12 pour avoir une idée approximative de la teneur en équivalent flavonoïdes glycosylés. Ce coefficient est le rapport de la moyenne de masses moléculaires des composés glycosylés sur la moyenne des masses molaires de génines. Ainsi, la teneur approximative en flavonoïdes glycosylés est estimée à 23 %. La seconde information est la répartition de ces noyaux : l'isorhamnétine représente presque les deux tiers de la teneur totale, puis le kaempférol et la quercétine constituent un peu plus d'un tiers de la teneur totale en flavonols. L'apigénine est également présente en proportion beaucoup plus faible. Le pic supplémentaire entre la myricétine et la quercétine est la lutéoline. Cette génine n'a pas été quantifiée.

3.4 Approche classique : isolement, purification et identification par RMN

A partir des premières informations obtenues pour l'extrait EtOH 60% de fleurs de *Primula veris*, nous avons entamé un travail d'investigation de son contenu par isolement des composés en vue de les identifier par RMN, FIA-HRMS et en appui avec des données de la littérature sur le type de composés en présence (flavonoïdes glycosylés).

La mise en œuvre de ce travail de purification a nécessité de multiples étapes. Pour commencer, nous avons produit une quantité d'extrait importante pour avoir suffisamment de matière au départ afin d'obtenir les molécules purifiées en quantité suffisante pour les analyses prévues. De plus, nous avons traité l'extrait par SPE afin de ne garder que la fraction flavonoïdique. Ensuite, nous avons pu entamer le processus d'isolement des composés par fractionnement en HPLC préparative, puis en HPLC semi-préparative. C'est alors que les composés purifiés ont été analysés par RMN, FIA-HRMS et également après hydrolyse acide pour une analyse en chromatographie ionique.

3.4.1. Prétraitement par SPE (extraction sur phase solide)

Pour commencer notre travail d'isolement des composés, nous avons effectué un prétraitement de l'extrait à l'aide de la SPE. En effet, la SPE est une méthode commune en préparation d'échantillon puisqu'elle permet d'éliminer certains composés de façon sélective et par conséquent de concentrer ceux qui sont conservés dans l'échantillon. C'est une méthode rapide et facile à mettre en œuvre qui nécessite des volumes de phases mobiles restreints. L'utilisation d'un prétraitement par SPE avait pour but l'obtention d'une fraction enrichie en composés d'intérêt de l'extrait et dépourvue des composés pouvant interférer sur la suite du travail de purification.

Nous avons utilisé une cartouche SPE C_{18} du fait de la polarité de l'extrait hydro-alcoolique étudié. Après une mise au point sur une cartouche de faible volume (500 mg/3 mL) (Phenomenex, Le Pecq, France), nous avons effectué le prétraitement de 4,84 grammes d'extrait sec obtenu par macération dans EtOH 60% de fleurs de *Primula veris* sur une cartouche de plus grande capacité SPE-C18 E-Strata (20 g/60 mL) (Phenomenex, Le Pecq, France). Les étapes, leurs rôles ainsi que les solvants et les volumes utilisés sont détaillés dans le Tableau 17.

Etape	Rôle de l'étape	Solvant	Volume
Mouillage	Rendre la phase opérationnelle en la solvatant	MeOH 100 %	120 mL
Conditionnement	Préparer la phase au solvant de dépôt (équilibrage)	MeOH 30 %	120 mL
Dépôt	Mettre l'échantillon en contact avec la phase : enrichissement des analytes d'intérêt par adsorption	MeOH 30 %	120 mL
Rinçage	Eliminer les composés non retenus par la phase	MeOH 30 %	60 mL
Collecte n°1	Récupérer les analytes d'intérêt par élution avec augmentation de la force éluante	MeOH 55 %	80 mL
Collecte n°2	Finaliser la récupération pour une élution quantitative	MeOH 55 %	80 mL
Nettoyage	Régénérer la phase pour recommencer le cycle à partir de l'étape de conditionnement	MeOH 100 %	240 mL

Tableau 17 Prétraitement par SPE-C_{18} : Etapes, rôle de l'étape, solvant et volume de solvant utilisé.

Cette suite d'étapes a été répétée à trois reprises de façon successive pour traiter l'ensemble de l'extrait. Il n'y avait pas d'étapes de séchage entre le nettoyage et le conditionnement suivant. En effet, pour éviter le phénomène de saturation de la cartouche SPE (qui n'aurait alors plus joué son rôle de rétention des composés d'intérêt pour les concentrer), on a limité le dépôt à 1,6 g d'extrait.

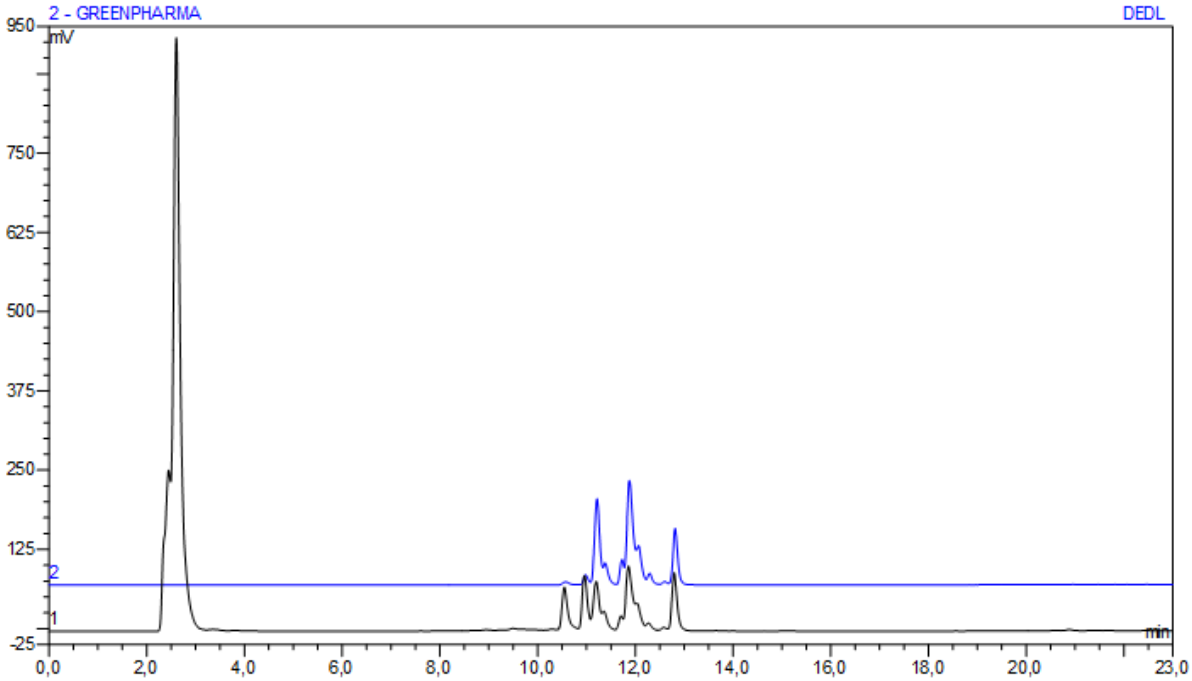


Figure 38 Chromatogrammes HPLC-DEDL de l'extrait de *Primula veris* avant purification (noir) et après purification sur SPE C_{18} (bleu)

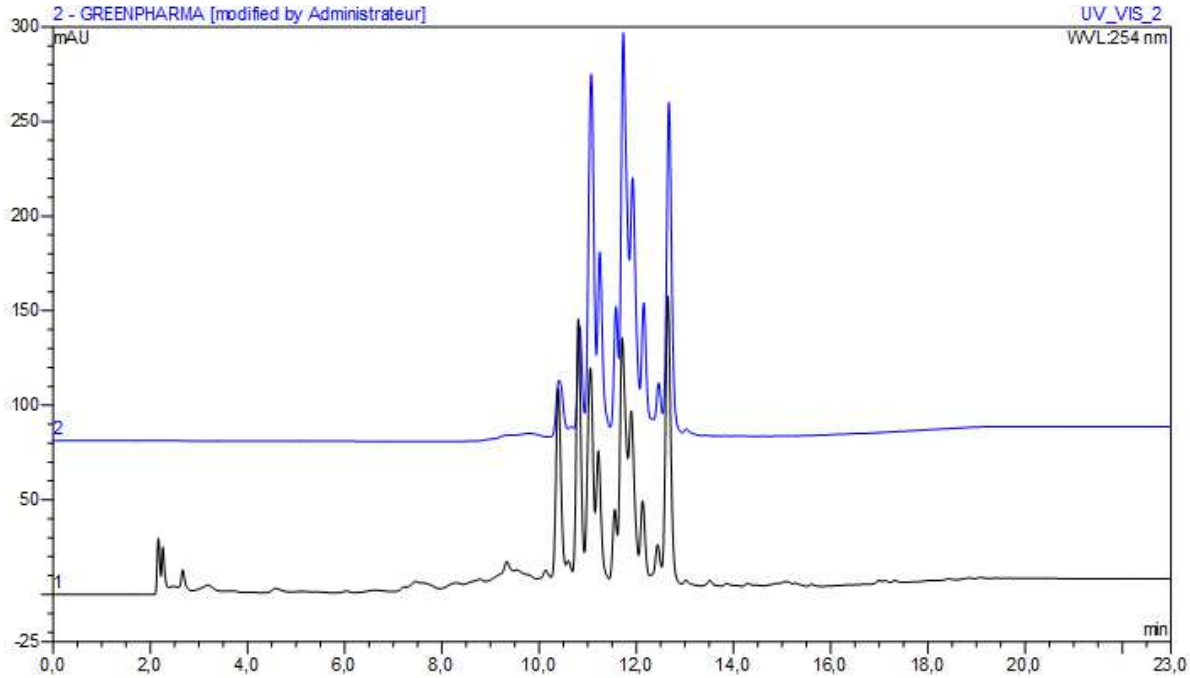


Figure 39 Chromatogrammes HPLC-UV 254 nm de l'extrait de *Primula veris* avant purification (noir) et après purification sur SPE C_{18} (bleu)

Le résultat de ce prétraitement a été l'obtention d'une version purifiée de l'extrait par élimination de la partie très polaire (non retenue sur C_{18}) et de la partie la plus apolaire (très retenue sur C_{18}) de l'extrait (zone d'élution après 15 minutes sur la Figure 39). Le rendement de l'étape SPE (ratio de la masse de fraction enrichie en flavonoïdes sur la masse injectée d'extrait) étant de 20 %, cela a conduit à la concentration des flavonoïdes glycosylés par un facteur 5.

3.4.2. Fractionnement en HPLC-préparative

L'objectif du fractionnement était de simplifier le travail de purification de molécules à partir d'un extrait uniquement traité par SPE. Le fractionnement constitue l'amorce d'un travail plus fin à venir qui sera fait en HPLC semi-préparative. Pour le fractionnement global, le choix de l'HPLC préparative est lié à sa capacité à traiter de grandes quantités d'échantillon grâce à des dimensions adaptées de la colonne et notamment de son diamètre interne. Nous l'avons mise en place pour fractionner l'extrait de *Primula veris* après le traitement par SPE. Nous avons récupéré une fraction de volume important qui après concentration était de 70 mL. La concentration de la fraction SPE ainsi obtenue est d'environ 14 000 ppm (20 % de la masse d'extrait traité / volume final).

Pour se faire, nous avons utilisé une HPLC préparative couplée à un détecteur UV muni d'une cellule de détection de dimensions adaptées à un grand débit de phase mobile. Après avoir testé un mode d'élution en isocratique à 40 % de MeOH puis à 30 % de MeOH, nous avons trouvé un compromis avec un isocratique à 38 % de MeOH. L'ensemble des conditions opératoires est détaillé en Annexe 4.3.

La Figure 40 illustre le fractionnement fait en HPLC préparative observé par HPLC analytique après concentration des fractions. On peut observer la répartition et les recoupements des différents pics de l'extrait. Les chromatogrammes des fractions F1 à F9 sont présentés, le chromatogramme de F10 n'a pas été ajouté puisque le signal est bruiteux du fait de sa faible concentration. Toutefois, il a le même profil que F9 et représente 15 mg de produit après séchage.

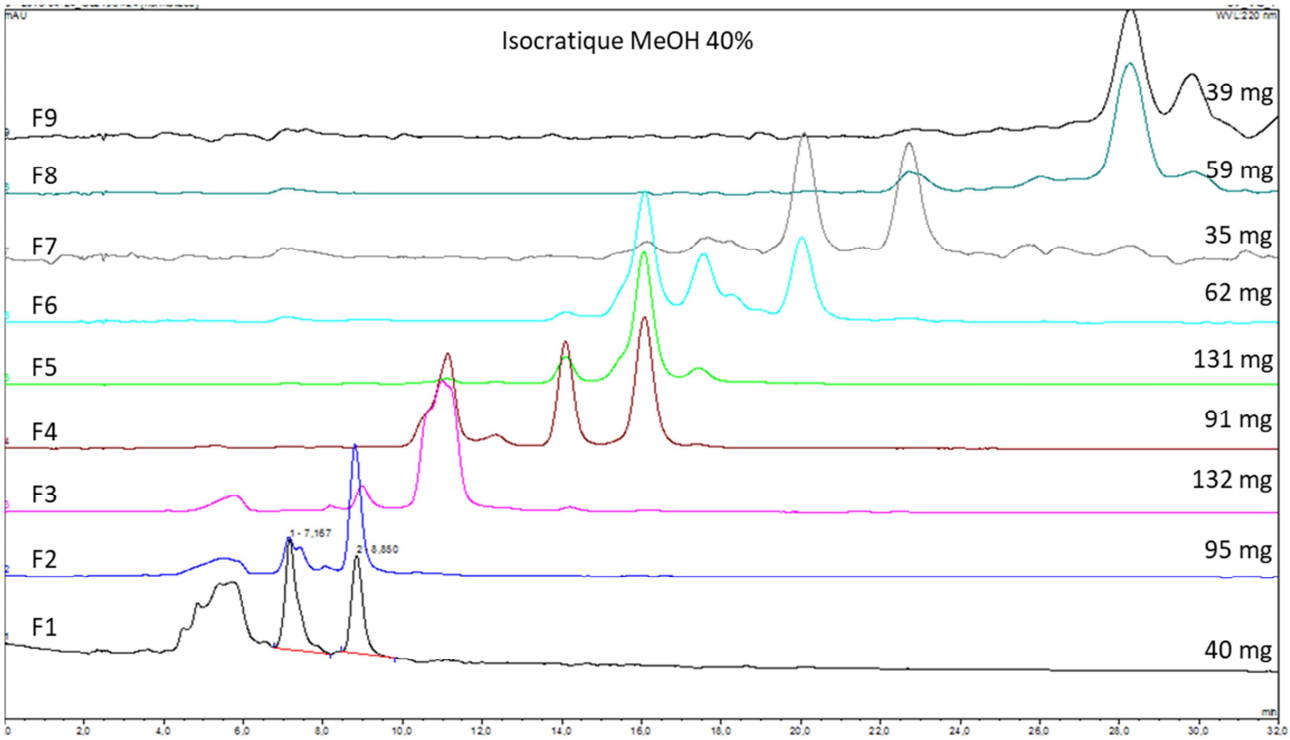


Figure 40 Superposition des chromatogrammes (normalisés) de chaque fraction (F1 à F9) issue du traitement en HPLC préparative. Analyse en HPLC analytique en mode isocratique MeOH 40 %. Sur la droite : quantité obtenue après séchage de chaque de fraction.

Le découpage de l'extrait en HPLC préparative a permis de répartir les composés dans des fractions avec une séparation satisfaisante. De plus, chaque pic est récupéré majoritairement au sein d'une seule et même fraction. Néanmoins, on remarque que des recoupements existent entre les fractions, ce qui implique qu'un même composé est présent dans différentes fractions. Toutefois, certaines fractions contiennent plus de matière que d'autres et il est possible que l'isolement d'un composé présent dans plusieurs fractions soit purifié en quantité suffisante à partir d'une seule de ces fractions.

Les fractions permettront d'appréhender plus simplement la suite des isolements. Les fractions constituent des sous-ensembles simplifiés de l'extrait et sont donc des échantillons moins complexes.

La Figure 41 représente les étapes de SPE, de mise au point et de collecte en HPLC préparative à travers le bilan massique du traitement de l'extrait de *Primula veris*.

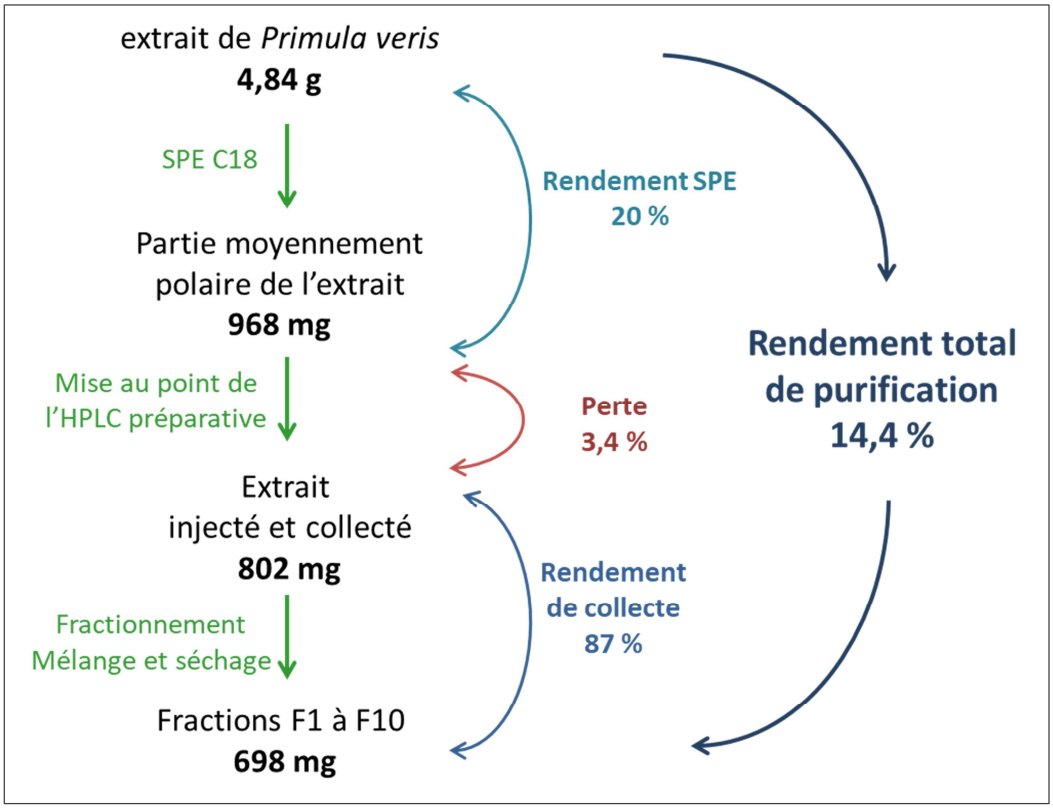


Figure 41 Traitement de fractionnement HPLC et bilan massique pour l'extrait de *Primula veris* : rendement de chaque étape mise en jeu.

Le fractionnement de l'extrait a abouti à l'obtention de dix fractions pour un rendement global de purification d'environ 15 %. A partir de près de 5 grammes d'extrait, les étapes successives ont limité la quantité finale obtenue. La perte de 3,4 % est liée à l'étape de mise au point où l'on a eu besoin de tester plusieurs conditions avant de les fixer pour l'ensemble des collectes. L'étape de fractionnement, de regroupement des collectes équivalentes (même pic collecté au cours des injections successives) et de séchage ont atteint un rendement important avec près de 90 % de recouvrement de la matière.

3.4.3. Purification de composés par HPLC semi-préparative

L'objectif final du processus est l'isolement de composés purs. Grâce à l'étape d'HPLC semi-préparative, il est possible de l'atteindre en effectuant une séparation plus efficace entre les différentes molécules et la matrice.

Les conditions semi-préparatives doivent être optimisées et on se base sur les informations d'HPLC analytique (avec notamment la zone de rétention en fonction du gradient) pour transposer les conditions à l'échelle semi-préparative. Dans le mode HPLC semi-préparative, nous avons effectué le travail de purification sur 8 fractions (F1, F2, F3, F5, F6, F7, F8 et F9). Les conditions opératoires pour les différents composés purifiés sont détaillées en annexe 4.4.1. Les purifications ont été effectuées en mode isocratique avec de l'acétonitrile sur des temps les plus courts possibles. En effet, les quantités de fractions injectées

doivent être limitées pour éviter les phénomènes de surcharge, ce qui implique de multiples injections. Toutefois, afin de limiter le nombre d'injections, l'optimisation consistait aussi à évaluer les concentrations et volumes maximaux qui garantissaient une séparation convenable des pics chromatographiques.

Pour chaque fraction, on collecte chaque pic en faisant en sorte de ne récupérer que le cœur du pic et pas les extrémités (début et fin de pic) pour assurer un maximum de pureté.

Après injection de la totalité de la fraction, collectes des différents pics, assemblages des collectes équivalentes, on procède au séchage par évaporation sous vide des différentes sous-fractions obtenues par HPLC semi-préparative. Ensuite, on vérifie la pureté des fractions par HPLC analytique. Il faut vérifier que le composé collecté soit pur à un niveau important. Idéalement une pureté supérieure à 95 % est souhaitée pour ne pas complexifier les données RMN ou HRMS.

Nous avons traité 8 des 10 fractions issues de l'HPLC préparative. Les fractions F4 et F10 n'ont pas été passées en HPLC semi-préparative puisqu'elles contiennent des composés que l'on a obtenus par le traitement d'autres fractions (F3 et F5 pour F4 ; F9 pour F10). Nous avons cherché à purifier les composés majoritaires de l'extrait et nous nous sommes concentrés sur les pics correspondants.

En utilisant l'approche classique consistant à isoler les composés de l'extrait, nous avons réussi à isoler 7 composés de pureté satisfaisante (5 composés de pureté supérieure ou égale à 98 %, 2 composés de pureté supérieure ou égale à 85 %) en utilisant près de 5 grammes d'extrait. La Figure 42 indique sur le profil RP-HPLC-UV de l'extrait de primevère, les composés qui ont pu être purifiés.

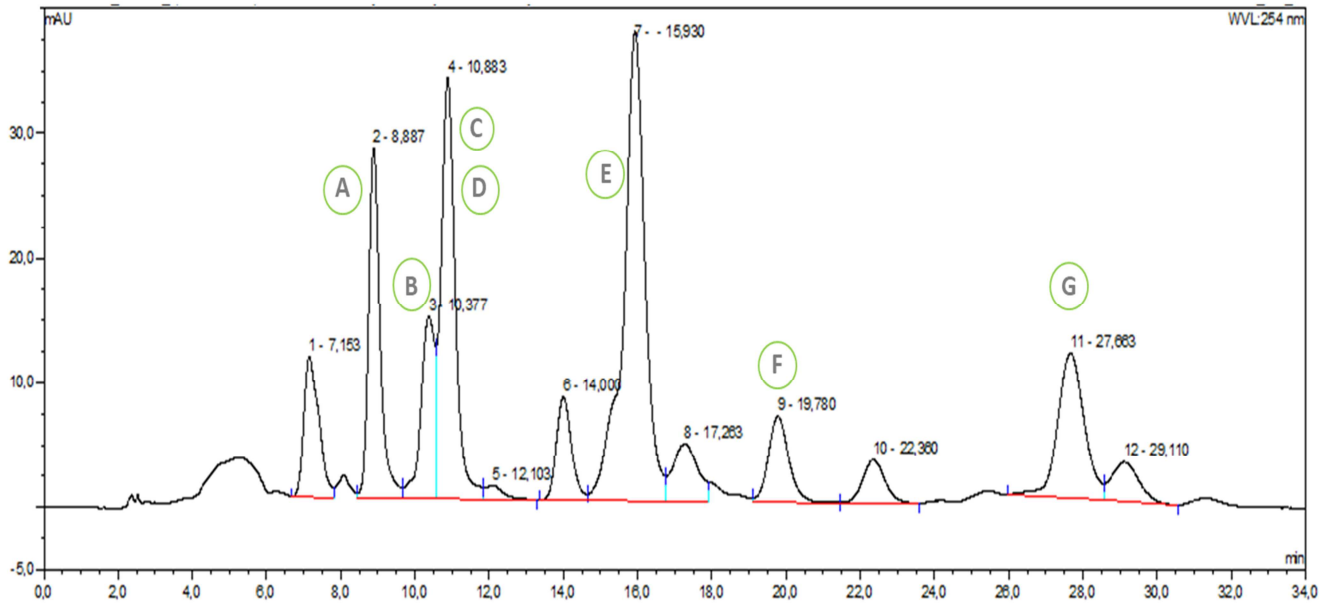


Figure 42 Chromatogramme de l'extrait de primevère vraie après traitement SPE analysé en RP-HPLC – UV 254 nm. Les pics **A** à **G** correspondent aux composés purifiés par HPLC semi-préparative.

A partir des aires des pics du chromatogramme, on a estimé les teneurs dans l'extrait de primevère de chaque composé. Pour cela, on a pris en compte le pourcentage de l'aire du pic purifié par rapport à celle

du massif dans lequel il se trouve (exemple avec **E** qui représente 1/5^{e} du massif à 15,93 minutes). On a ramené l'aire de chaque composé à la concentration de flavonoïdes glycosylées dans l'extrait (dosage des flavonoïdes et estimation de la teneur en équivalent flavonoïdes glycosylés en paragraphe 3.3). La teneur estimée indiquée dans le tableau suivant est le résultat de la multiplication de l'aire du composé par la teneur en flavonoïdes glycosylés dans l'extrait (à savoir environ 23 %). Cette méthode d'estimation est approximative mais permet d'avoir une idée des teneurs des composés les uns par rapport aux autres notamment.

Pic	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**TOTAL**
Teneur estimée	≈ 2,3 %	≈ 1,7 %	≈ 1,3 %	≈ 1,3 %	≈ 1,3 %	≈ 1,1 %	≈ 2,5 %	**11,5 %**

Tableau 18 Estimation de la teneur approximative dans l'extrait de chaque pic purifié en considérant l'aire en UV du composé et la teneur en flavonoïdes glycosylés dans l'extrait de primevère.

Les autres pics du profil (n°1, 5, 6, 8, 10 et 12 et la part non purifiée du n°4) représentent 6,7 % de flavonoïdes glycosylés. La somme des teneurs estimées des composés purifiés et non purifiés est égale à 18,2 %, on retrouve quasiment le rendement obtenu par SPE sur l'extrait (rendement de 20 %).

3.4.4. Analyses FIA-HRMS, RMN et hydrolyse acide

3.4.4 a) *Démarche globale suivie*

Une combinaison d'analyses a été effectuée pour chacune des molécules isolées à des niveaux satisfaisants de pureté et obtenues en quantité suffisante. L'analyse d'un composé en FIA-HRMS permet d'accéder à sa formule brute et à certains éléments structuraux visibles sur le spectre dans le cas de fragmentation en source.

Avec la FIA-HRMS, on a pu accéder aux formules brutes des molécules, confirmer les génines et sucres en présence. Pour 3 composés (**C**, **D** et **E**), une analyse de fragmentation MS/MS a complété *a posteriori* la première analyse en HRMS pour collecter des informations sur la liaison des sucres. Pour la RMN, les analyses en ^{13}C, ^{1}H, DEPT et en 2D (COSY, HSQC) ont été réalisées. Chaque composé a également été hydrolysé en milieu acide (conditions annexe 4.5.1). L'hydrolysat est analysé selon deux méthodes : i) en chromatographie ionique ce qui permet l'identification des oses portés par la molécule et ii) en RP-HPLC pour expliciter la génine. Le détail de chaque type d'analyse et des conditions utilisées est donné en annexes 4.5.3 et 4.5.4.

Les différents résultats issus de l'ensemble des analyses ont été traités manuellement, puis croisés pour déduire des éléments structuraux. Les composés connus pour la plante ont servi d'appui pour certaines hypothèses et les données analytiques de la littérature (notamment RMN) sur ces composés ont aidé à les interpréter. En effet, grâce à la détermination de la formule brute, de la génine et des sucres, nous pouvions

formuler des hypothèses de structures. Les données RMN publiées dans la littérature nous ont servi de support pour les comparer à nos données expérimentales et nous aider dans l'attribution des signaux. Les analyses en COSY n'ont pas permis de déterminer l'enchaînement des sucres. En effet, les nombreux signaux des sucres sortant dans une zone de déplacement chimique très restreinte forment un massif complexe impossible à interpréter.

3.4.4 b) *Résumé des propositions d'identification*

Le Tableau 19 référence le nom des produits isolés, les analyses effectuées, la masse monoisotopique, les éléments structuraux déduits par le croisement des différentes données analytiques, la correspondance de la masse et des éléments avec les composés connus pour la plante ou le genre *Primula*, ainsi que la teneur estimée de chaque produit dans l'extrait pour les 7 composés isolés. Les masses monoisotopiques indiquées sont les moyennes des masses neutres correspondant entre elles pour un même composé analysé en FIA-HRMS en mode positif et/ou négatif. Chaque composé a une fiche récapitulative des données analytiques collectées et interprétées en Annexe 5.

Désignation composé	Analyses effectuées	Masse monoisotopique ($g.mol^{-1}$)	Eléments structuraux du composé	Correspondance avec composés de *Primula veris* ou du genre *Primula*	Remarques complémentaires
A (F2-02GJ3022)	RMN, HRMS, HA	788,20113	Quercétine triglycosylé (3 Glucoses)	**Quercetin triglucosylé** (enchaînement sucre non défini) / **Quercetin 3-(2^{G}-glucosylgentiobioside)** / **Isoquercitrin; 6''-o-[beta-d-glucopyranosyl-(1->2)-beta-d-glucopyranoside]**	RMN non exploitables (signal bruiteux en ^{13}C et pas correctement résolu en ^{1}H)
B (F3-01GJ3121)	HRMS, HA	626,14815	Quercétine diglycosylé (2 Glucoses)	**Quercetin-3-O-gentiobioside** / **Quercetin-3-O-neohesperidoside**	Pas d'analyses RMN : impossibilité de solubiliser le composé
C (F3-02GJ3121)	RMN, HRMS$^{\alpha}$, HA	740,21680	Kaempférol triglycosylé en une seule position (2 Rhamnoses + 1 hexose)	Aucun candidat	Pas de confirmation des positions liaisons entre les sucres Intégration < 1 pour H-C_6 et H-C_8
D (F3-03GJ3121)	RMN, HRMS$^{\alpha}$, HA	770,22753	Isorhamnétine triglycosylé en une seule position (2 Rhamnoses + 1 hexose)	**Isorhamnetin-3-O-(2''-O-α-rhamnosyl-6''-O-α-rhamnosyl)-β-galactoside**	Pas de confirmation des positions de liaisons entre les sucres C quaternaire C_7 non observé ; Intégration < 1 pour H-C_6 et H-C_8
E (F5-02GJ3100)	RMN, HRMS$^{\alpha}$, HA	802,21719	Isorhamnétine triglycosylé en une seule position (3 Glucoses)	**Isorhamnetin 3-O-(2,6-di-O-β-D-glucopyranosyl-β-D-glucopyranoside)**	Pas de confirmation des positions de liaisons entre les sucres
F (F6-05GJ3056)	RMN, HRMS, HA	640,163998 / 670,174364	Flavonoïde diglycosylé (2 hexoses) / Flavonoïde glycosylé	**Isorhamnetin-dihexoside** (structure non définie totalement) / Aucun candidat	Faible pureté du produit purifié (85 %) HRMS indique 2 produits dont l'un est majoritaire et diglycosylé.
G (F8-01GJ3087)	RMN, HRMS, HA	624,16875	Isorhamnétine diglycosylé (1 rhamnose + 1 hexose)	**Narcissin** / **Isorhamnetin 3-O-robinobioside** / **Isorhamnetin 3-O-neohesperidoside**	Le pic purifié est peut-être un mélange de deux isomères connus pour la plante

Tableau 19 Récapitulatif des informations obtenues pour les 7 molécules isolées à partir de l'extrait EtOH 60 % de Primevère vraie (HA : Hydrolyse Acide).

$^{\alpha}$Données complétées par analyse FIA-HRMS/MS pour tenter d'observer l'enchaînement de la partie glycosylée sous forme sodée.

3.4.4 c) *Détails de l'interprétation de résultats analytiques*

Suite à l'isolement et aux analyses multiples effectuées pour chaque composé, l'enjeu était de déterminer la structure du produit isolé en faisant coïncider l'ensemble des résultats.

Pour les produits **C**, **D, F** et **G** : dans la colonne des éléments structuraux, on a indiqué un hexose puisque le monosaccharide en présence n'est pas déterminé plus précisément. En effet, après hydrolyse acide, on a détecté à la fois le glucose et le galactose (voir détails dans les fiches de chaque composé en Annexe 5). Pour **C**, **D** et **G**, nous émettons l'hypothèse que les produits isolés soient des mélanges de composés dont la seule différence serait la nature de l'hexose porté. Il est plausible que ces isomères ne soient pas séparés dans nos conditions de RP-HPLC. Pour **F**, on sait que la fraction n'est pure qu'à 85 % du composé majoritaire et en FIA-HRMS, on a effectivement observé la masse de deux composés. Les sucres détectés par hydrolyse sont cette fois encore : le glucose et le galactose. Ainsi, on ne peut dire si les deux hexoses portés par F sont deux glucoses ou deux galactoses ou un glucose et un galactose.

Pour le produit **A**, les spectres RMN sont non exploitables puisqu'ils sont de très faibles intensités. On ne détecte pas de signaux de carbones et les signaux des protons sont mal définis (manque de résolution notamment). Il y eu un problème de réponse du produit mis en solution et pourtant bien solubilisé dans D_2O. Une piste aurait été d'augmenter la température d'analyse RMN mais cela présente des inconvénients du fait de la structure envisagée (flavonoïde glycosylé) qui est sensible à la chaleur. Par ailleurs, les autres analyses nous ont permis d'envisager une quercétine triglucosylée. Une molécule pouvant correspondre a été décrite dans *Primula auricula*.

S'agissant du produit **F** ; il se trouve qu'il y deux composés en présence. D'après les intensités des ions en HRMS, on sait que le composé de masse moléculaire 640 $g.mol^{-1}$ est majoritaire. En FIA-HRMS, pour ce composé, on observe la perte de deux hexoses. D'après l'hydrolyse et l'analyse en chromatographie ionique, ces deux hexoses peuvent être deux glucoses, ou deux galactoses ou un glucose et un galactose. Pour le composé minoritaire (masse de 670 $g.mol^{-1}$), on n'observe pas les pertes de sucres mais du fait de la masse, il est certain que le composé est également glycosylé. Les ions détectés contiennent des atomes de deutérium (voir détails en 0). La raison de leur présence s'explique par le fait que l'échantillon a été solubilisé dans D_2O pour la RMN et a servi pour l'analyse FIA-HRMS. Il y a eu des échanges de protons avec le deutérium en solution.

Quant aux produits **C**, **D et E**, l'ensemble des données nous a permis d'aller plus loin dans leur caractérisation notamment avec l'analyse en MS/MS des adduits sodés qui nous a permis d'observer les sucres sous forme sodée : les 3 composés ont leurs trois sucres sous forme continue (lié entre eux). Par ailleurs, avec cette expérience, on ne peut pas dire s'ils sont linéaires ou branchés. L'expérience en MS/MS

est illustrée avec l'exemple du spectre de **D** (Figure 43) où l'ion de m/z 477,1579 correspond à la masse d'un hexose et deux desoxyhexoses complexés avec un sodium. Ainsi, on sait que les trois monosaccharides portés par le composé sont liés.

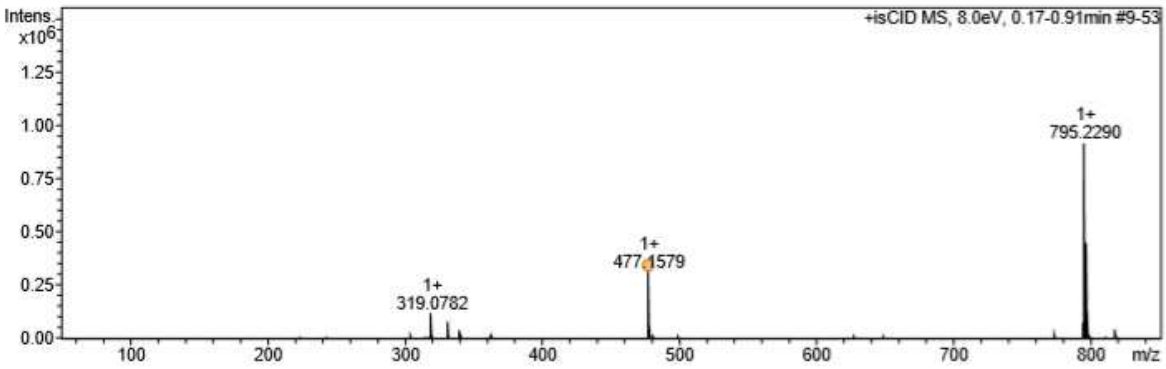


Figure 43 Spectre MS/MS isCID (*In Source Collision Induced Dissociation*) du produit **D**

L'ion majoritaire de *m/z* 795 correspond à l'adduit sodé et deutéré du composé isolé. En effet, on a déjà déterminé que la masse de la molécule était de 770 g.mol^{-1}, l'adduit sodé correspond à un *m/z* égal à 793. Le *m/z* majoritaire visible sur le spectre correspond à cet adduit sodé mais dont deux atomes d'hydrogènes ont été échangés avec du deutérium. Ces échanges sont liés au fait que le composé a été solubilisé dans D_2O pour l'analyse RMN et qu'un prélèvement de cette solution a servi pour l'analyse HRMS. L'ion à 319,0782 correspond l'ion $[M+H]^+$ de la génine isorhamnétine dont deux hydrogènes ont été échangés avec des deutériums (provenant du solvant utilisé en RMN).

3.4.4 d) *Interprétation des analyses RMN et/ou problèmes rencontrés*

Pour les composés **A** et **B**, les données RMN ne sont pas suffisamment exploitables pour pouvoir aider à l'identification des composés. Pour **A**, les analyses RMN du carbone correspondent à l'enregistrement d'un signal bruiteux sans pouvoir discerner de signal de carbone. De plus, l'analyse RMN du proton est mal résolue et les intégrations effectuées sont incohérentes. Quant au produit **B**, le composé n'a pas pu être analysé car il ne s'est pas solubilisé dans les solvants suivants testés : D_2O, D_2O/MeOH-d_4 50/50 et MeOH-d_4.

Pour le produit **C**, la référence bibliographique utilisée est ''Malonylated flavonol glycosides from the petals of *Clitoria ternatea*'' [117] qui décrit un kaempférol 3-O-(2''-O-α-rhamnosyl-6''-O-β-glucosyl)-β-glucoside (*cf*. structure page suivante : Figure 44). Le choix de s'appuyer sur les données RMN de cette structure est lié au fait que cet enchaînement de sucres est rencontré dans des molécules connues pour *Primula veris*. C'est un élément nous permettant de nous orienter vers un type d'enchaînement glycosidique et de nous aider à attribuer les signaux obtenus.

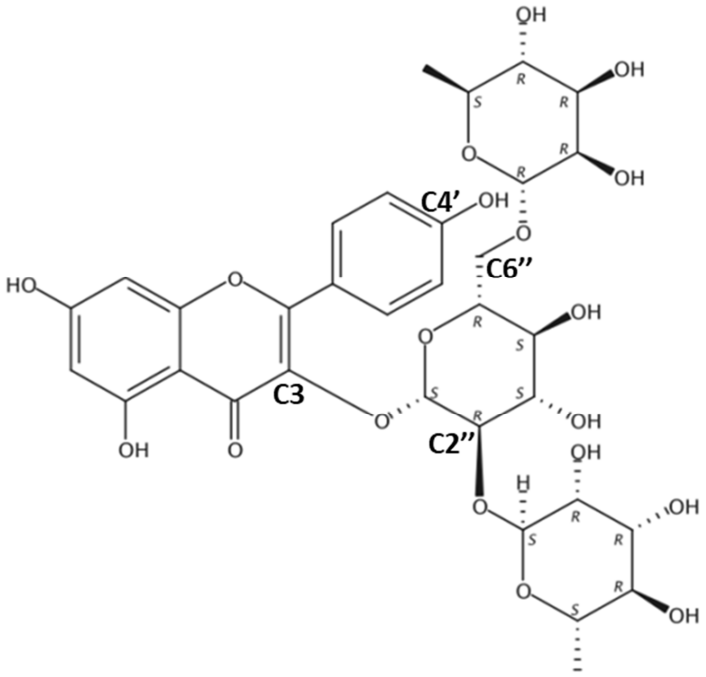


Figure 44 Structure du kaempférol 3-O-(2''-O-α-rhamnosyl-6''-O-β-glucosyl)-β-glucoside

La comparaison des données expérimentales avec celles de la référence (*cf*. Annexe 5.3) montre qu'un grand nombre de signaux sont équivalents. L'ajout d'acide formique dans le solvant deutéré a permis d'obtenir des spectres RMN mieux définis que lors d'une première analyse en méthanol deutéré pur. Toutefois, le déplacement de deux carbones ne correspond pas du tout avec la référence utilisée : il s'agit des carbones C2'' et C3'' appartenant au glucose : ils ont des déplacements chimiques plus élevés dans le spectre de référence. Cela semble indiquer que l'enchaînement des sucres n'est pas celui envisagé.

Pour l'analyse des données de RMN du composé **D**, la publication ayant servi de base à l'annotation des signaux de RMN est la référence [118] qui décrit le composé présenté en Figure 45. Ce composé est identique à une molécule isolée de *Primula veris* au cours la publication de Karl *et al.* de 1981 [119]. Dans cette autre référence, les signaux RMN de la partie glycosylée ne sont pas détaillés.

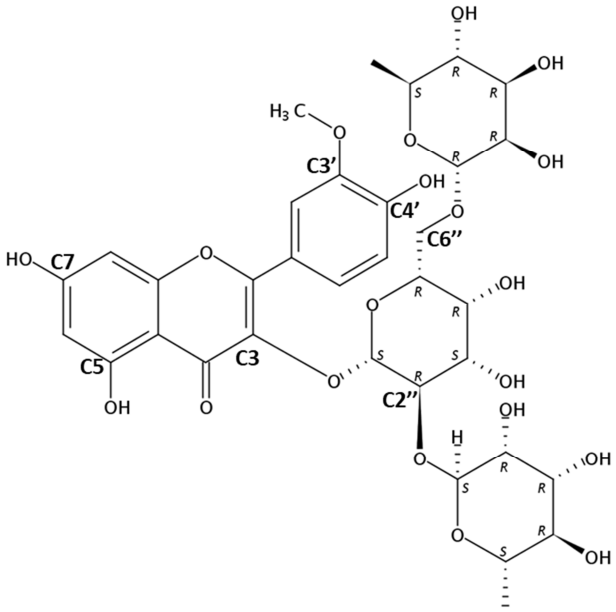


Figure 45 Structure de l'isorhamnetine-3-O-(2''-O-α-rhamnosyl-6''-O-α-rhamnosyl)-β-galactoside

Pour le composé **D**, une première analyse RMN dans D_2O montre une correspondance importante des signaux expérimentaux avec ceux de la référence à l'exception des signaux des carbones de l'hexose. Les différences de déplacements chimiques obtenus sont entre 1 et 2,5 ppm (*cf*. Annexe 5.4). Ainsi,

il est difficile de confirmer que D est le composé décrit. Ces différences s'expliquent par la différence de solvant car les données de la référence sont obtenues dans le DMSO-d_6. Le produit **D** a donc été analysé dans le DMSO-d_6 et dans ce cas les carbones de l'hexose correspondent parfaitement avec ceux de la publication, toutefois d'autres signaux (notamment de la génine) précédemment détectés dans D_2O ne sont plus détectés dans le DMSO-d_6 ou sont légèrement différents. L'emploi des deux solvants d'analyse nous permet de conclure que L'enchaînement des sucres est celui de la structure présentée dans la publication. Sans avoir une seule et même analyse RMN le confirmant, il semble que la structure isolée soit celle envisagée et déjà décrite dans *Primula veris* à savoir l'isorhamnetin-3-O-(2''-O-α-rhamnosyl-6''-O-α-rhamnosyl)-β-galactoside. Cette structure est assez originale du fait de la présence du galactose plutôt que du glucose qui est plus fréquent.

Par ailleurs, en RMN, certains signaux ont été détectés de façon incomplète. Pour les deux composés (**C** et **D**) les hydrogènes des carbones C_6 et C_8 ont une intégration inférieure à 1. Pour le composé **D**, le carbone C_7 n'est pas détecté. Ce phénomène survient parfois pour les carbones quaternaires.

L'identification du composé **E** s'appuie sur les données de la publication ''Comparative phytochemical and morphological analyses of three Italian *Primula species*'' [108] dont les analyses RMN ont été effectuées dans le DMSO-d_6. La molécule décrite qui pourrait correspondre au composé est l'isorhamnetin-3O-(2,6-di-O-β-D-glucopyranosyl-β-D-glucopyranoside) dont la structure est présentée en Figure 46.

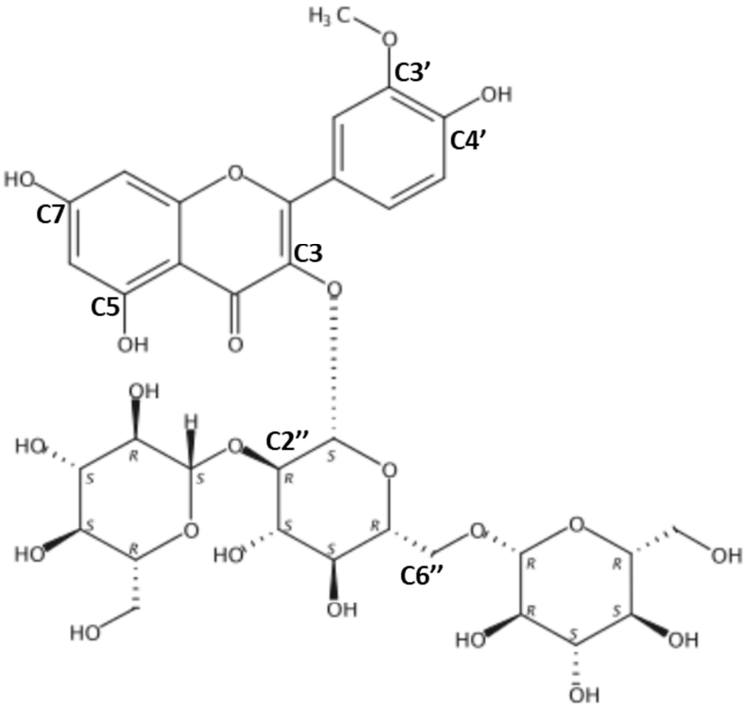


Figure 46 Structure de l'isorhamnetin-3O-(2,6-di-O-β-D-glucopyranosyl-β-D-glucopyranoside)

Les spectres RMN du carbone et du proton de la molécule isolée correspondent à ceux décrits dans la littérature même si les deux analyses ne sont pas faites dans le même solvant (D_2O *versus* DMSO-d_6) : les déplacements chimiques observés sont similaires (*cf*. Annexe 5.5). Ainsi, il est possible d'annoter les données expérimentales. Les données RMN tendent à valider que la structure de **E** est la molécule décrite pour le genre *Primula* dans la publication citée précédemment. Toutefois, il manque un signal de carbone quaternaire (C_7). De plus, les signaux des protons sont superposés dans le spectre RMN ^{1}H expérimental, ce qui n'est pas le cas pour l'analyse publiée et réalisée dans le DMSO-d_6. La bonne concordance des données permet quasiment d'avoir une proposition d'identification sans ambiguïté.

Pour le composé **F**, nous n'avons pas de publication de référence pour comparer les données expérimentales obtenues. Toutefois, ces dernières confirment la présence de deux sucres et indiquent 4 protons aromatiques (Annexe 5.6). Ce nombre de protons aromatiques fait penser que la génine portant les sucres n'est pas une isorhamnétine comme on l'envisageait d'après la FIA-HRMS et suite à l'hydrolyse acide. En effet, l'isorhamnétine est un flavonol doté de 5 protons aromatiques. Pour rappel, ce produit n'est pur qu'à 85 % et il est possible que l'isorhamnétine soit constitutive d'une autre structure présente dans le produit isolé.

Concernant le produit **G**, l'exploitation des données expérimentales est effectuée en appui des données de la littérature (références pour les données ^{13}C : [120] et pour les données ^{1}H : [121]) de l'isorhamnetin-3O-α-L-rhamnosyl (1→6) β-D-glucoside (communément appelée narcissine). Celle-ci est une des deux structures proposées pour le composé d'après la déréplication ciblée parmi les molécules connues dans l'organisme *Primula veris*. Cela a aidé à attribuer les signaux détectés. La structure de la narcissine est présentée en Figure 47. L'autre structure probable pour caractériser le produit est l'isorhamnetin-3O-α-L-rhamnosyl (1→6) β-D-galactoside (aussi appelée isorhamnetin-3-O-robinoside) (*cf*. Figure 48) d'après la recherche ciblée de composés de cette masse dans la plante *Primula veris*.

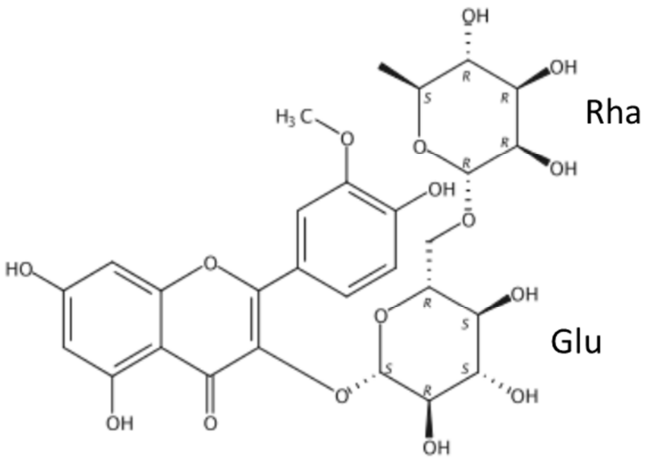


Figure 47 Structure de l'isorhamnetin-3-O-rutinoside

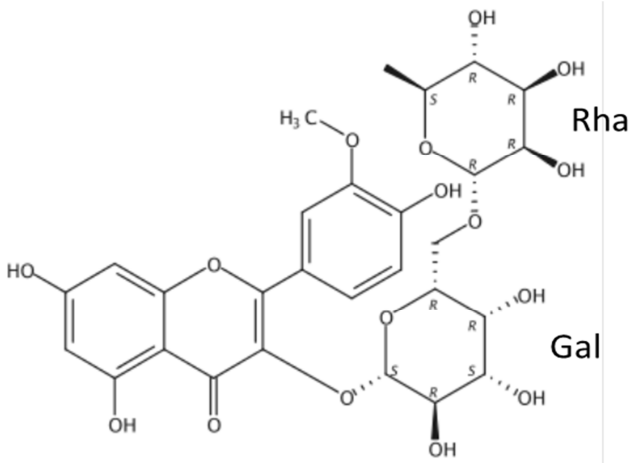


Figure 48 Structure de l'isorhamnetin-3-O-robinoside

Les données de RMN (*cf*. Annexe 5.7) permettent de confirmer nos hypothèses mais elles ne permettent pas la mise en évidence formelle d'un des deux isomères. Les données RMN, qui nous permettraient de différencier l'isorhamnetin-3-O-rutinoside de l'isorhamnetin-3-O-robinoside, sont les constantes de couplage entre les protons H4 et H3 ainsi que H4 et H5 de l'hexose du composé analysé qui seraient grandes pour le glucose et faibles pour le galactose. En effet la différence entre les deux composés proposés est la nature de l'hexose (glucose ou galactose). Entre ces deux hexoses épimères (différent par la configuration d'un seul centre asymétrique), c'est la configuration du carbone C4 qui est opposée et donc l'orientation du proton H4 par rapport au proton H3 et H5. Dans notre étude, les données collectées ne nous permettent pas de déterminer la constante du fait de la complexité du massif des protons des sucres, ces derniers ayant des signaux superposés dans une gamme de déplacements chimiques restreinte. D'autre part, le signal du H4 n'est pas indiqué dans la référence bibliographique. Enfin la RMN pourrait indiquer le fait que les 2 sucres sont liés ou non mais on ne peut exploiter les signaux de sucres dans le détail en proton.

Pour le composé **G**, une proposition d'identification sans ambiguïté ne peut être faite mais de fortes similarités entre les données expérimentales et celles de littérature permettent de penser que la molécule est une isorhamnétine portant deux sucres dont un hexose et un rhamnose.

3.4.5. Conclusion sur l'approche classique pour décrire le contenu de l'extrait de Primevère vraie

La conclusion de nos travaux est que nous avons isolé des flavonoïdes glycosylés portant de 2 à 3 oses. Pour 5 produits, une correspondance avec les molécules connues sur la plante ou son genre est trouvée. On retrouve un certain nombre de dérivés à base d'isorhamnétine qui constitue la génine majoritaire de l'extrait d'après les résultats d'hydrolyse. Il y a aussi une mise en évidence de nouveaux composés pour l'espèce *Primula veris*. Toutefois, certains problèmes ont été rencontrés pour établir l'identification complète ou formelle de certains des produits comme cela a été décrit précédemment du fait principalement de spectres RMN incomplets ou peu intenses et de la complexité à déterminer les positions des liaisons des sucres entre eux avec certitude.

Si le travail nécessaire à l'identification de moins d'une dizaine de composés issus de l'extrait a été conséquent et long, il présente un intérêt notable. En effet, l'isolement de ces composés majeurs de l'extrait permet l'obtention de molécules pures (avec des quantités allant de 15 mg à 30 mg en moyenne pour les composés étudiés). D'autres molécules (une dizaine) ont été obtenues en quantités plus faibles (entre 1,5 et 10 mg) : elles n'ont pas été étudiées dans le détail comme les 7 autres principales du fait de la nécessité d'avoir assez de produits pour effectuer l'ensemble des analyses. Ces molécules peuvent être onéreuses à l'achat voire non commerciales. Or pour définir leur activité biologique dans le but d'identifier des biomarqueurs de l'extrait, les molécules à l'état pur sont primordiales. Ceci est important lorsqu'on veut

justifier l'activité d'un ingrédient de façon précise au niveau moléculaire. Cela conduit à étayer et renforcer l'innovation autour d'un ingrédient.

3.5 Approche globale : analyse en LC-HRMS et utilisation de GAINS

Le contenu de l'extrait de *Primula veris* a également été étudié sans passer par les étapes d'isolement des molécules mais par UHPLC-HRMS. Nous voulions appliquer l'outil GAINS et évaluer le niveau et la quantité d'informations que notre approche seule pouvait nous apporter en comparaison avec l'approche précédente dite « classique ».

Pour rappel, le processus semi-automatisé mis en place est le suivant :

1) Analyses de l'extrait en UHPLC-HRMS en mode d'ionisation positif et négatif

2) Prétraitement des données brutes pour obtenir le chromatogramme de masse (pics MS) et un fichier avec les temps de rétention des pics associés à leur spectre de masse et leurs données de fragmentations

3) Traitement des données par les algorithmes de GAINS : annotations des ions observés en MS et exploitation des données de MS/MS

4) Interrogation de la base de données GPDB pour proposer des candidats structuraux.

L'ensemble des conditions analytiques sont données en 0. Les chromatogrammes des analyses sont proposés en Figure 49 et Figure 50.

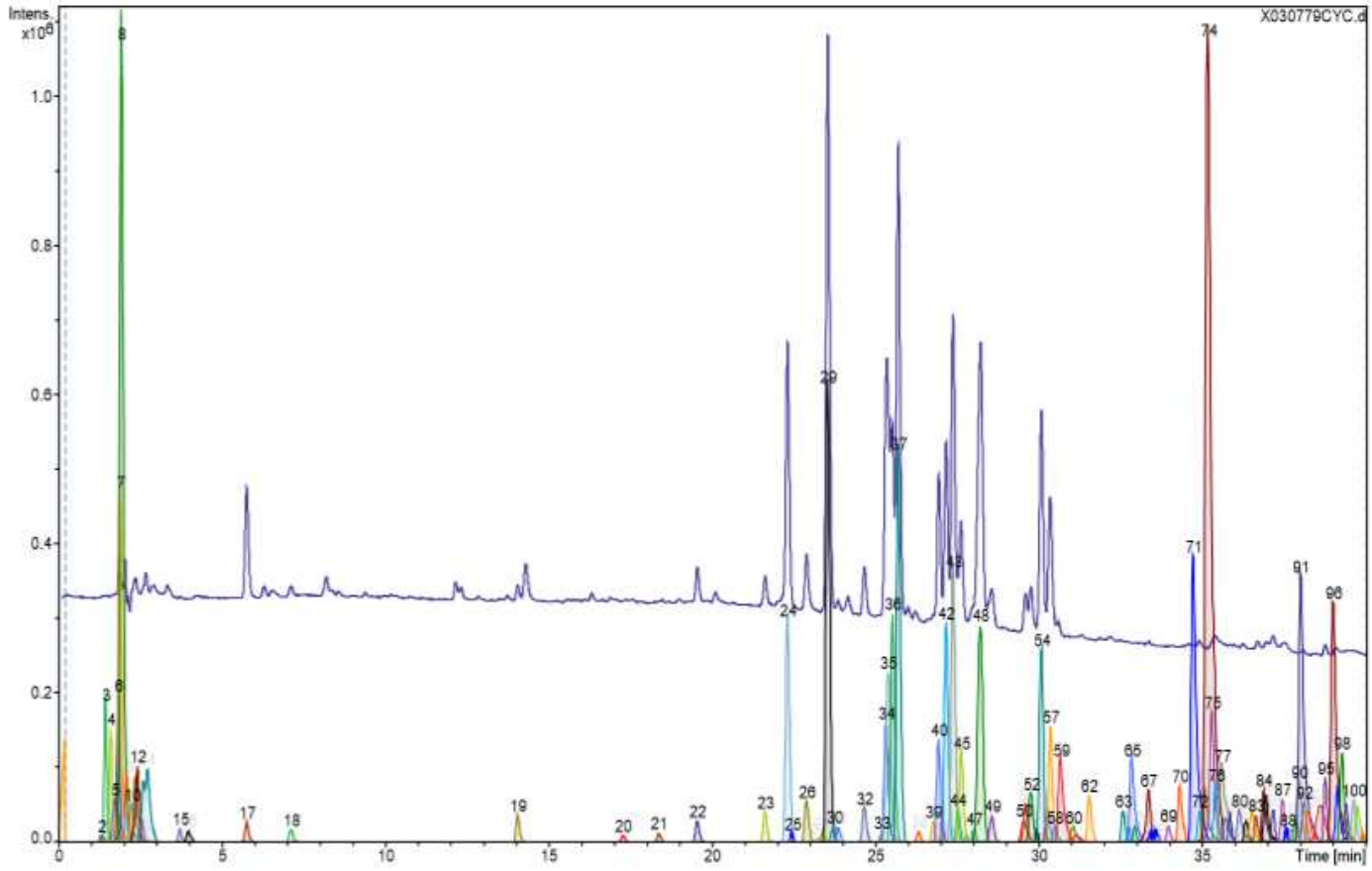


Figure 49 Chromatogrammes UV 254 nm (tracé bleu) et MS-ESI$^+$ de l'extrait de *Primula veris* à 1000 ppm

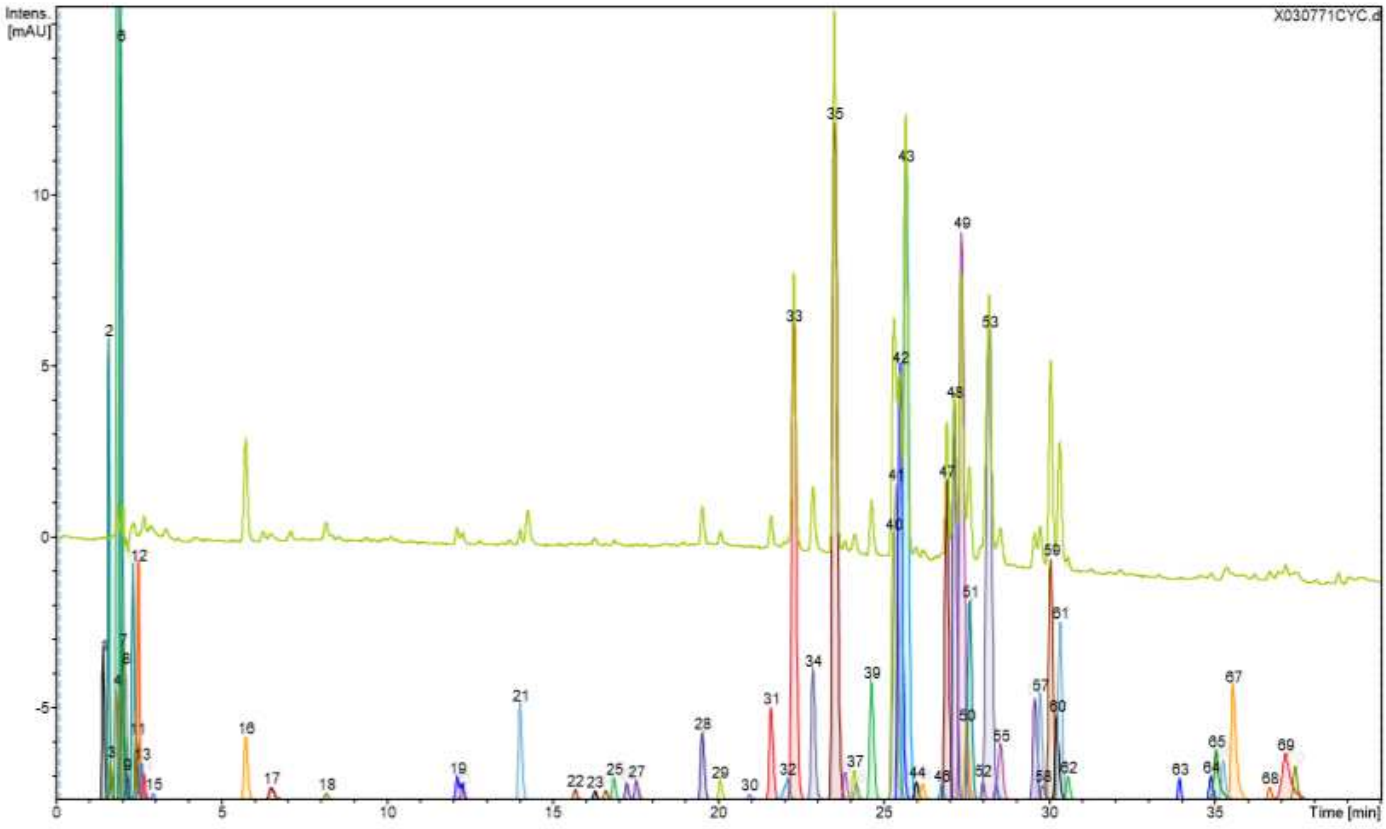


Figure 50 Chromatogrammes UV 254 nm (tracé vert) et MS-ESI$^-$ de l'extrait de *Primula veris* à 1000 ppm

3.5.1. Résultats

3.5.1 a) *Déréplication ciblée sur les candidats de Primula veris ou du genre Primula*

Dans un premier temps, on s'intéresse aux pics pour lesquels des candidats sont trouvés dans les molécules connues pour l'organisme *Primula veris* (proposition **en vert**) ou le genre *Primula* (proposition **en orange**). Le Tableau 20 référence les 30 pics pour lequel c'est le cas en donnant les numérotations des pics en mode positif (P) et/ou négatif (N), leur temps de rétention (lorsque un pic P et un N correspondent : temps de rétention moyen), la masse monoisotopique proposée par GAINS et le candidat proposé.

Pour chaque candidat, les données ont été consultées pour vérifier la cohérence de la proposition avec le temps de rétention (en fonction de la glycosylation par exemple). En tout début d'analyse, on a des propositions de petites molécules de la famille des sucres : leur grande polarité limite leur rétention en mode RP-HPLC. Ensuite, au cœur du gradient, les flavonols glycosylés sont élués. Pour les flavonoïdes glycosylés, la tendance observée est cohérente avec ce qui est attendu lors d'une séparation en mode phase inverse, à savoir : plus la partie aglycone est hydroxylée, plus le composé est élué rapidement. Globalement, les composés glycosylés basés sur des hexahydroxyflavones sortent avant ceux basés sur des pentahydroxyflavones, eux-mêmes mélangés aux composés tetrahydroxy-methoxyflavones. Enfin les tetrahydroxyflavones et les génines non glycosylés arrivent en fin de gradient. De plus, on remarque que pour une même génine, ce sont les composés de plus haut degré de glycosylation qui sont élués en premier. Ceci est cohérent puisqu'ils sont les plus polaires.

Pics	Temps de rétention (min)	Masse monoisotopique ($g.mol^{-1}$)	Candidats obtenus par déréplication ciblée dans *Primula veris* ou le genre *Primula*
8P/5N	1,91	212,088729	d-glycero-d-gluco-heptitol / d-glycero-d-manno-heptitol
8N	2,12	180,062515	myo-inositol
19P/21N	14,05	306,073118	epigallocatéchine
27N	17,54	290,078279	catéchine
22P/28N	19,55	804,195615	3,3',4',5,6,7-hexahydroxyflavone; 3-o-gentiotrioside / 3,3',4',5,7,8-hexahydroxyflavone; 3-o-gentiotrioside
23P/31N	21,63	788,200632	quercetin triglucosylé (enchaînement non défini) / quercetin 3-(2^G-glucosylgentiobioside) / isoquercitrin; 6''-O-[beta-d-glucopyranosyl-(1->2)-beta-d-glucopyranoside]
32N	22,10	626,147430	quercetin-3-O-gentiobioside
29P/35N	23,55	788,201007	quercetin 3-(2^G-glucosylgentiobioside) / isoquercitrin; 6''-O-[beta-d-glucopyranosyl-(1->2)-beta-d-glucopyranoside]
38N	24,20	610,152195	quercetin 3-robinobioside / rutine / kaempferol 3-O-gentiobioside
32P/39N	24,67	640,163540	isorhamnetin -dihexoside (structure non définie totalement)
34P/40N	25,32	626,147235	quercetin-3-O-gentiobioside
36P/42N	25,52	772,206069	kaempferol-3-O-diglucoside-7-O-glucoside (enchaînement non défini)
37P/43N	25,70	802,216431	isorhamnetin 3-O-(2,6-di-O-β-D-glucopyranosyl-β-D-glucopyranoside)
39P/46N	26,77	416,109566	macrophylloside (= 2'-hydroxy-7-*O*-β-D-glucopyranosyloxyflavone)
40P/47N	26,94	770,227175	isorhamnetin-3-O-(2''-O-α-rhamnosyl-6''-O-α-rhamnosyl)-β-galactoside
42P/48N	27,17	740,216484	kaempferol-3-O-galactoside-rhamnoside-7-O-rhamnoside (enchaînement non défini)
43P/49N	27,38	770,227071	isorhamnetin-3-O-(2''-O-α-rhamnosyl-6''-O-α-rhamnosyl)-β-galactoside
44P/50N	27,53	610,152399	quercetin 3-O-robinobioside / rutine / quercetin-3-O-neohesperidoside / kaempferol 3-O-gentiobioside
45P/51N	27,64	640,163553	isorhamnetin-dihexoside (structure non définie totalement)
47P/52N	28,02	594,157764	kaempferol 3-robinobioside / kaempferol-3-rutinoside / kaempferol-3-O-neohesperidoside
48P/53N	28,22	610,152746	quercetin 3-O-robinobioside / rutine / quercetin-3-O-neohesperidoside / kaempferol 3-O-gentiobioside
49P/55N	28,56	464,094524	isoquercitrin
51P/56N	29,61	594,157871	kaempferol 3-robinobioside / kaempferol-3-rutinoside / kaempferol-3-O-neohesperidoside
54P/59N	30,08	624,168658	narcissin / isorhamnetin 3-O-robinobioside / isorhamnetin 3-O-neohesperidoside

Pics	Temps de rétention (min)	Masse monoisotopique ($g.mol^{-1}$)	Candidats obtenus par déréplication ciblée dans *Primula veris* ou le genre *Primula*
57P	30,38	594,157838	**kaempferol 3-robinobioside** / **kaempferol-3-rutinoside**
67P	33,38	342,109919	**2',5,6,6'-tetramethoxyflavone**
79P	35,86	312,099433	**3', 4', 5'-trimethoxyflavone**
68N	36,67	286,046874	**kaempferol** / **luteoline**
92P	38,15	252,078486	**2'-methoxyflavone** / **3'-methoxyflavone**
95P	38,79	282,089193	**2',5'-dimethoxyflavone** / **3',4'-dimethoxyflavone**

Tableau 20 Résultats de propositions d'identification issus de GAINS : déréplication ciblée parmi les molécules connues de l'organisme *Primula veris* et du genre *Primula*

Les composés référencés sont des composés d'intensité UV intense à moyennement intense. On a donc collecté des informations pour un nombre conséquent de pics majoritaires de l'extrait.

3.5.1 b) *Déréplication avec les candidats de GPDB pour les autres pics intenses*

Un complément est apporté par l'interrogation de la base GPDB dans sa totalité. Dans le Tableau 21, on répertorie les 6 pics du profil qui n'ont pas de candidats dans le genre *Primula* mais qui font partie des pics relativement intenses du profil, en précisant pour chacun d'entre eux le temps de rétention, la masse monoisotopique, le nombre de candidats au sein de la base correspondant à cette masse et le(s) type(s) de candidats.

Pics	Temps de rétention (min)	Masse monoisotopique ($g.mol^{-1}$)	Nombre de candidats	Liste ou type de candidats
17P/16N	5,76	300,083873	5	1-(4-hydroxybenzoyl)glucose / 4-glucosyloxybenzoic acid / Amabiloside / di-methyl fukiic acid / salicylic acid beta-d-glucoside /
24P/33N	22,31	818,211165	1	3,4',5,5',7-pentahydroxy-3'-methoxyflavone; 3,5',7-tri-o-beta-d-glucopyranoside
26P/34N	22,90	656,158529	8	Pentahydroxy-methoxyflavones diglycosylés (avec Glucose et/ou Galactose)
35P/41N	25,32	756,211351	50	Flavonoïdes triglycosylés
52P/57N	29,77	654,179432	19	Flavonoïdes diglycosylés
57P/61N	30,36	594,157838	112	Flavonoïdes diglycosylés

Tableau 21 Résultats de GAINS pour des pics d'intensité non négligeable du profil avec des candidats non décrits parmi ceux du genre *Primula*

Après une recherche élargie à la base de données GPDB pour les pics principaux sans candidat connu dans le genre *Primula*, on aboutit à la mise en lumière de flavonoïdes glycosylés. Parmi les génines de la famille des flavonoïdes figurent des flavones, des flavonols, isoflavones.

Ces pics constituent une piste de travail pour identifier de nouveaux composés importants qui n'ont jamais été identifiés dans le genre *Primula* et plus particulièrement dans *Primula veris*. On confirme que l'extrait étudié est très riche en flavonoïdes. C'est une famille moléculaire très diverse dont les structures peuvent contenir de nombreux types de noyaux (génines) et de glycosylation.

3.5.1 c) *Apport de l'exploitation des données HRMS/MS*

Le Tableau 22 référence les pics LC-HRMS de masses équivalentes à celles des produits isolés par la méthode classique. Dans ce tableau, on a listé les pics, les masses monoisotopiques associées par LC-HRMS, les éléments structuraux confirmés par l'exploitation manuelle des données LC-HRMS/MS. Ensuite, on indique le type de candidats proposés par déréplication ciblée au sein de l'espace chimique décrit par les composés connus pour *Primula veris* et le genre *Primula*. Dans le cas où les éléments structuraux déterminés manuellement et le type de structure proposé par déréplication ne sont pas en adéquation, cela est indiqué par un surlignage en gris dans le tableau. Enfin, on référence le nombre de candidats trouvés pour la masse monoisotopique au sein de la totalité de la base de données GPDB (déréplication non ciblée). La dernière colonne indique les structures (parmi toutes celles associées à la masse après requête dans GPDB) ayant obtenu le score maximum grâce au traitement par Metfrag. Cela nous permet d'évaluer l'apport de Metfrag par rapport à la gestion automatique des données MS/MS pour distinguer l'ensemble des candidats.

Correspon-dance avec un produit isolé	Pics	Masse monoisotopique ($g.mol^{-1}$)	Eléments structuraux confirmés par analyses des données MS/MS manuelle	Type de candidats proposés par déréplication ciblée dans le genre *Primula*	Nombre de candidats obtenus par déréplication dans GPDB	Proposition avec score Metfrag de 1
A	**23P/31N**	788,200632	**Penta-hydroxyflavone avec 3 hexoses** (enchainement non déterminé)	Quercetin triglycosylé	13	4 Quercetin triglycosylé
A	**29P/35N**	788,201007	**Penta-hydroxyflavone** avec **3 hexoses liés**	Quercetin triglycosylé	13	4 Quercetin triglycosylé
B	**32N**	626,147430	**Penta-hydroxyflavone** avec **2 hexoses**	Quercetin diglucosylé	41	4 Hexahydroxyflavones diglycosylés ϟ
F	**32P/39N**	640,163540	**Penta-hydroxymethoxyflavone avec 1 hexose** et **1 desoxyhexose**	Isorhamnetin diglycosylé	41	Pentahydroxy-methoxyflavone rutinoside et robinobioside
B	**34P/40N**	626,147235	**Penta-hydroxyflavone** avec **2 hexoses liés**	Quercetin diglucosylé	41	Quercetin diglycosylé
E	**37P/43N**	802,216431	**Tetra-hydroxymethoxyflavone** avec **3 hexoses liés**	Isorhamnetin triglycosylé	12	1 Isorhamnetin triglycosylé
D	**40P/47N**	770,227174	**Tetra-hydroxymethoxyflavone avec 1 hexose** et **2 desoxyhexoses**	Isorhamnetin triglycosylé	16	3 Rhamnetine ou Isorhamnetine triglycosylés
C	**41P**	740,216135	**Tetra-hydroxyflavone** avec **1 hexose** et **2 desoxyhexoses**	Kaempférol triglycosylé	31	2 Kaempferol triglycosylé
C	**42P/48N**	740,216484	**Tetra-hydroxyflavone** avec **1 hexose** et **2 desoxyhexoses liés**	Kaempférol triglycosylé	31	2 Kaempferol triglycosylé
D	**43P/49N**	770,227071	**Tetra-hydroxymethoxyflavone** avec **1 hexose** et **2 desoxyhexoses**	Isorhamnetin triglycosylé	16	3 Rhamnetine ou Isorhamnetine triglycosylés
F	**45P/51N**	640,163553	**Tetra-hydroxymethoxyflavone** avec **2 hexoses liés**	Isorhamnetin diglycosylé	41	Pentahydroxy-methoxyflavone ϟ rutinoside et robinobioside
G	**54P/59N**	624,168658	**Tetra-hydroxymethoxyflavone** avec **1 hexose** et **1 desoxyhexose liés**	Isorhamnetin 3O-diglycosylé	76	1 Aurone diglycosylé ϟ

Tableau 22 Résultats obtenus avec Metfrag pour les pics dont la masse monoisotopique correspond à celle de composés purifiés à partir de l'extrait.

ϟ Dans les cas où Metfrag ne propose pas de candidats cohérents avec ce qui a été déterminé manuellement : le score obtenu, pour le type de candidats attendus d'après le traitement manuel, est convenable. Par exemple, pour le 32N, le meilleur score attribué à une quercétine diglycosylé est de 0,902. Pour le 54P/59N, le meilleur score attribué à une isorhamnetin diglycosylé est 0,998754 ($2^{ième}$ rang).

La 4ème colonne montre que les données HRMS/MS sont porteuses d'éléments précis et très informatifs quant aux structures. On parvient à déterminer le type de génine, le type et le nombre de sucres et parfois on retrouve aussi l'information concernant le fait que les sucres soient sous forme continue (tous liés entre eux) sans pour autant pouvoir déterminer leur arrangement précis. Toutefois, ces informations nécessitent encore une exploitation manuelle conséquente.

La dernière colonne montre que Metfrag fonctionne bien en proposant un nombre restreint de candidat souvent en adéquation avec ce qui est attendu. Pour l'ensemble des pics observés, les candidats les mieux classés par Metfrag ont des structures qui correspondent parfaitement aux éléments structuraux qui ont pu être déterminés manuellement (type de génine, type de sucres). Dans les deux tiers des cas, les structures dotées du meilleur score Metfrag sont cohérentes avec les structures proposées par déréplication.

Dans le cas du pic 23P/31N, Metfrag semble indiquer que la génine est glycosylée en une seule position tandis que l'ion correspondant aux trois sucres liés sous forme de l'adduit sodé, qui est le seul moyen lors du traitement manuel pour expliciter cette information, n'a pas été observé. Ici, l'outil combinatoire permet de proposer une information supplémentaire par rapport au traitement manuel.

Pour le pic 54P/59N, les molécules décrites pour la plante obtiennent le second meilleur score Metfrag (0,998754) pour lequel un grand nombre de candidats arrivent *ex-aequo*. Dans ce cas, Metfrag n'attribue pas le meilleur score aux propositions d'identification que nous pensons être correctes mais les classe avec un très bon score.

Pour le composé 32P/29N, Metfrag classe en première position des candidats correspondants aux éléments structuraux déterminés lors du traitement manuel des données MS/MS. En MS/MS, à partir de l'ion précurseur $[M+H]^+$ du composé, l'ion fragment majoritaire (*m/z* 333,06034) indique une génine de type penta-hydroxymethoxyflavone. La différence entre la masse monoisotopique du composé (640,163540 $g.mol^{-1}$) et celle de la génine (332,05322 $g.mol^{-1}$) est égale à la masse d'un hexose et d'un désoxyhexose à 1,3 ppm près. Ainsi, Metfrag confirme que la proposition d'identification (isorhamnétine portant deux hexoses) qui serait faite par déréplication ciblée dans *Primula veris* est fausse. On ajoutera que le type de génine mise en évidence pour ce pic est plus rare que les autres types de génines mentionnés dans cette étude et qu'il n'a jamais été décrit pour *Primula veris*.

Ainsi, en tant qu'outil automatique d'analyses des données MS/MS, Metfrag pallie la faiblesse du traitement manuel plus long. Pour autant, Metfrag modélise la fragmentation. Face à un nombre important de candidats à classer, il ne parvient pas toujours à les différencier et peut proposer de nombreux candidats avec le même score. De plus, il peut proposer des candidats dont les génines ou les substitutions ne sont pas toutes équivalentes. De ce cas-là, certains éléments structuraux ne sont pas forcément dégagés.

3.6 Comparaison entre la méthode classique et la méthode utilisant GAINS

Lorsque l'on compare les composés décelés avec la méthode classique aux données acquises par LC-HRMS, on retrouve les masses des composés purifiés. On aboutit aux mêmes éléments structuraux. Le Tableau 23 permet de comparer les résultats obtenus pour les composés isolés par la méthode classique.

La masse déterminée des produits isolés par la méthode classique constitue le critère de recherche en LC-HRMS. Une remarque concerne les pics du profil LC-HRMS qui sont de même masse monoisotopique. Ces pics sont indiqués sur la même ligne (exemple du pic 23P/31N avec le pic 29P/35N) puisqu'ils présentent tous deux une masse équivalente à celle recherchée. On précise que c'est l'un ou l'autre des deux pics qui correspond au produit isolé de même masse.

Résultats de la méthode classique			Résultats de la méthode utilisant GAINS		
Composé isolé	**Eléments structuraux**	**Candidats (déréplication ciblée dans le genre *Primula*)**	**Numéro du pic**	**Candidats (déréplication ciblée dans le genre *Primula*)**	**Score Metfrag**
A	Quercétine triglucosylé	Quercetin triglucosylé (enchaînement sucre non défini) Quercetin 3-(2^G-glucosylgentiobioside) Isoquercitrin; 6''-O-[beta-d-glucopyranosyl-(1->2)-beta-d-glucopyranoside]	23P/31N ou 29P/35N	Quercetin triglucosylé (enchaînement sucre non défini) Quercetin 3-(2^G-glucosylgentiobioside) Isoquercitrin; 6''-O-[beta-d-glucopyranosyl-(1->2)-beta-d-glucopyranoside]	NA 1 1
B	Quercétine diglucosylé	Quercetin-3-O-gentiobioside	32N ou 34P/40N	Quercetin-3-O-gentiobioside	0,992
C	Kaempférol-hexoside-rhamnoside-rhamnoside (forme continue)	Kaempferol-3-O-galactoside-rhamnoside-7-O-rhamnoside (structure non définie totalement)	42P/48N	Kaempferol-3-O-galactoside-rhamnoside-7-O-rhamnoside (structure non définie totalement)	NA
D	Isorhamnetin-hexoside-rhamnoside-rhamnoside (forme continue)	isorhamnetin-3-O-(2''-O-α-rhamnosyl-6''-O-α-rhamnosyl)-β-galactoside	40P/47N 43P/49N	isorhamnetin-3-O-(2''-O-α-rhamnosyl-6''-O-α-rhamnosyl)-β-galactoside	1
E	Isorhamnétine triglucosylé (forme continue)	Isorhamnetin 3-O-(2,6-di-O-β-D-glucopyranosyl-β-D-glucopyranoside)	37P/43N	Isorhamnetin 3-O-(2,6-di-O-β-D-glucopyranosyl-β-D-glucopyranoside)	1
F	Isorhamnétine hexoside-hexoside (composé majoritaire)	Isorhamnetin-dihexoside (structure non définie totalement)	32P/39N ou 45P/51N	Isorhamnetin-dihexoside (structure non définie totalement)	NA
G	Isorhamnétine-hexoside-rhamnoside	Narcissin Isorhamnetin 3-O-robinobioside Isorhamnetin 3-O-neohesperidoside	54P/59N ou 57P/61N	Narcissin Isorhamnetin 3-O-robinobioside Isorhamnetin 3-O-neohesperidoside	0,9988 0,9988 0,9972

Tableau 23 Comparaison des deux méthodes utilisées pour décrire l'extrait : mise en parallèle des résultats obtenus.

Une remarque est à faire pour le tableau précédent. Pour **G**, la masse monoisotopique déterminée correspond en UHPLC-HRMS à deux pics. Il s'agit du 54P/ 59N et du 57P/61N. Pour le pic 57P/61N, il faut préciser que la masse de 624 est proposée par GAINS mais seulement comme deuxième hypothèse de masse (la première étant 594 $g.mol^{-1}$). On a considéré cette deuxième hypothèse de façon exceptionnelle car elle a du sens par rapport aux composés majoritaires détectés dans l'extrait. En effet, le pic 57P/61N élué juste après le 54P/59N pourrait être un isomère comme on l'a envisagé lors de l'approche classique de l'extrait pour les produits où le galactose et le glucose avaient été détectés. Ainsi, le duo de pics 54P/59N et 57P/61N pourrait expliquer la détection de glucose et du galactose après l'hydrolyse du produit **G**. Le fait que la séparation des isomères soit effective en UHPLC et pas en HPLC est plausible. Une des perspectives du travail pour l'extrait concerne le produit **G**. L'analyse de standards commerciaux (narcissine et isorhamnetin-3-O-robinobioside) permettra d'évaluer si les deux composés sont présents dans l'extrait par injection en UHPLC des standards et de l'extrait.

La comparaison entre la méthode classique et la méthode plus moderne montre que les informations obtenues par la première sont aussi déterminées grâce à la seconde. L'outil GAINS permet d'effectuer un travail de déréplication pertinent et à la hauteur de ce que nous avions fait précédemment en isolant les composés majoritaires de l'extrait. Le module Metfrag permet de classer l'ensemble des candidats associés à une masse monoisotopique. Les candidats proposés sur la base des données LC-HRMS et MS/MS sont bien classés par Metfrag. Les candidats n'ayant pas de score attribué (score indiqué : NA) correspondent à des structures non définies totalement dans la littérature. Si elles sont référencées dans la base de données GPDB avec une structure non totalement définie, Metfrag ne peut donc fonctionner pour ces cas-là ne pouvant pas gérer des structures référencées en plusieurs sous-structures.

3.7 Conclusion

3.7.1. Bilan sur la comparaison des techniques d'exploration du contenu

Le contenu de l'extrait obtenu par extraction avec EtOH 60 % de fleurs de Coucou a été étudié selon deux approches indépendantes mettant en jeu des combinaisons de techniques analytiques différentes.

La première approche a utilisé des méthodes classiques de préparation d'échantillons, de purification et d'analyses des composés isolés : SPE, HPLC préparative et semi-préparative, FIA-HRMS et RMN. Cette démarche a été longue, complexe puisqu'il a fallu mettre au point de nombreuses étapes et conditions d'analyse ou de purification. Elle a nécessité une grande quantité d'extrait. En effet, différentes étapes successives ont été réalisées afin de pouvoir passer de l'extrait brut à un peu moins d'une dizaine de composés purifiés. L'isolement a ainsi dû être limité aux pics majoritaires. En revanche, ces purifications nous donnent ou devraient nous donner accès à un niveau de confiance élevé d'identification des composés

puisque nous nous appuyons sur les données RMN et de MS des composés isolés. Dans cette étude, les données RMN n'ont pas apporté beaucoup d'informations supplémentaires à celles déjà obtenues par le panel des autres analyses. En effet, il n'a pas été possible d'exploiter de façon approfondie les RMN du proton notamment dû à la superposition des signaux des protons des sucres. Les informations RMN interprétées ont plutôt validé les éléments structuraux déterminés et par comparaison avec des données de la littérature, certaines propositions de structures ont pu être infirmées ou étayées. Un avantage certain de ces étapes de purification est qu'à l'issue de ce travail, les composés peuvent être testés pour leur activité biologique *versus* l'extrait. Cela permet de déterminer s'ils sont des biomarqueurs ou non de l'extrait.

La deuxième approche ayant été mise en place pour caractériser l'extrait a été accès autour d'analyses en UHPLC-HRMS pour s'appuyer sur l'outil GAINS au niveau du traitement des données analytiques. La mise en œuvre a consisté en la détermination de conditions optimisées pour l'UHPLC puis les résultats ont été obtenus en une seule série d'analyses UHPLC-HRMS. Cette démarche a donné accès à de nombreuses informations qui après traitement ont permis des propositions d'identification sûres et plus rapidement qu'avec la première approche. C'est donc une vision globale du contenu de l'extrait qui est obtenue par cette deuxième approche et l'étude n'est pas limitée uniquement aux pics majoritaires. Pour les composés majoritaires isolés et caractérisés par la méthode classique, on note une très bonne concordance des résultats obtenus par les deux approches. Une limite de l'approche utilisant GAINS est l'absence d'un niveau d'identification formelle puisque la molécule n'est pas isolée. Pour pallier cet inconvénient, on peut de façon complémentaire envisager de recourir à l'injection de composés standards pour étayer les propositions d'identification et atteindre ainsi un niveau d'identification plus important en passant du niveau intermédiaire entre 2 et 3 au niveau 1 de caractérisation d'après les MSI (*cf.* paragraphe 4.1.3 du Chapitre 1). Par ailleurs, après l'utilisation de l'outil GAINS, l'isolement de composés peut toujours être envisagé comme perspective du travail en ciblant des pics particuliers qui présenteraient un intérêt. Par exemple, on peut prédire *in silico* des interactions potentielles de molécules avec des cibles biologiques et envisager que ce criblage virtuel guiderait le choix des composés à purifier.

3.7.2. Bilan sur les composés mis en évidence

Les flavonoïdes glycosylés sont majoritaires dans cet extrait hydro-alcoolique. Cette famille de composés est largement répandue dans le règne végétal, notamment des plantes à fleurs/fruits (angiospermes). Toutefois, les composés mis en évidence présentent un caractère assez intéressant pour deux raisons. La première concerne les glycosylations : les structures avec 3 sucres sont plus rares que les mono- et di-glycosylations or il y a 8 composés portant trois sucres dans l'extrait analysé (données LC-HRMS). De plus, parmi les sucres mis en évidence, le rhamnose est assez occurrent. Ces éléments apportent de l'originalité en termes de phytochimie. La deuxième raison qui rend le contenu flavonoïdique de l'extrait

intéressant concerne la présence de la génine isorhamnétine, elle aussi plus rare que les flavonols quercétine et kaempférol. Le contenu phytochimique de l'extrait présente donc des intérêts en vue de la valorisation de l'extrait et notamment en cosmétique.

4. Etude d'un extrait de réglisse glabre : Obtention d'une vue globale du contenu phytochimique et recherche d'isoflavones

4.1 Contexte de l'étude et objectifs

Le Groupe Rocher a produit un extrait de racines de réglisse (*Glycyrrhiza glabra*) et a mandaté Greenpharma pour réaliser des analyses LC-HRMS afin d'identifier les molécules des six principaux pics chromatographiques observés lors d'analyses préliminaires en RP-DAD-MS par le groupe Rocher. Le Groupe Rocher a souhaité que ce travail soit complété par la recherche spécifique de présence d'éventuelles isoflavones, dont la daïdzeine, sur l'ensemble du profil chromatographique de cet extrait. GAINS a été pour nous un outil de choix pour tenter de répondre aux deux grands objectifs à atteindre : traiter rapidement les données acquises lors des analyses LC-HRMS pour identifier les molécules et permettre la recherche ciblée des composés d'une famille particulière (isoflavones) en interrogeant la base de données GPDB.

4.2 Présentation de la réglisse glabre

Appartenant à la famille des Fabacées, la réglisse glabre (*Glycyrrhiza glabra* L.) est une herbacée vivace, de forme arbustive et dont les inflorescences en forme de grappes ont une couleur lilas (violet pâle). La plante est cultivée depuis des siècles pour ses racines. Séchées et principalement présentées sous la forme des bâtons de réglisse, les racines possèdent un goût prononcé et sucré du fait de l'acide glycyrrhizique (aussi appelé glycyrrhizine) dont le pouvoir sucrant est 50 fois supérieur à celui du saccharose [7,122]. D'autres molécules sont connues pour leur présence dans la réglisse glabre : la glabridine (isoflavane), la liquiritigénine (isoflavone) et l'isoliquiritigénine (chalcone) ainsi que leur forme glucosylée pour les deux dernières [7]. Globalement, les saponosides triterpéniques et plusieurs classes de composés phénoliques dont les flavanones, isoflavones, isoflavènes, chalcones ou coumarines sont trouvées dans les espèces du genre *Glycyrrhiza* [7,123]. De nombreuses molécules ont été décrites pour ces plantes largement étudiées en phytochimie. Dans la base GPDB, pour la réglisse glabre, 230 molécules sont référencées.

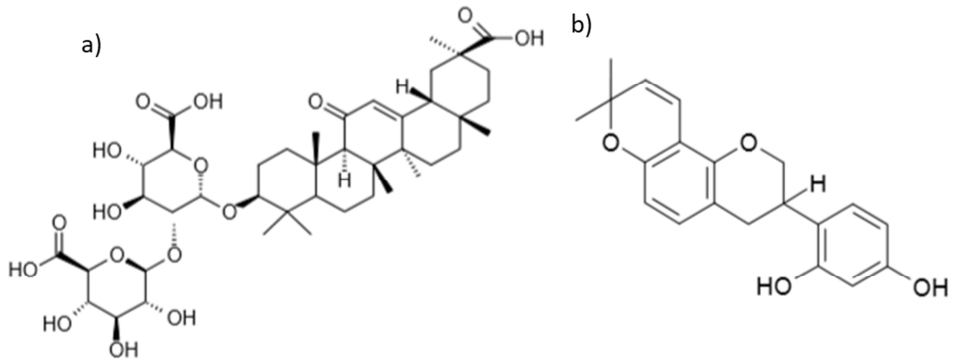


Figure 51 Structure de a) l'acide glycyrrhizique et b) la glabridine

La réglisse est une plante aromatique largement connue et réputée à la fois pour ses propriétés médicinales et ses usages alimentaires (édulcorant et exhausteur de goût dans les boissons, confiseries ou tabacs). Dans la médecine chinoise, la réglisse est considérée comme une herbe tonique, efficace pour harmoniser les caractéristiques des autres herbes d'une préparation et capable de réduire la toxicité de certaines plantes par son pouvoir détoxifiant [124]. On la trouve dans plus de la moitié des compositions traditionnelles chinoises [124]. Dans la pharmacopée européenne, les racines de réglisse doivent contenir au moins 4 % d'acide glycyrrhizique et sont inscrites pour le traitement de troubles digestifs ou comme expectorant. De façon traditionnelle, elle est utilisée dans le traitement de la toux, la grippe ou de problèmes liés au foie [123]. Les principales applications d'extrait de réglisse concernent des activités antitussive, anti-inflammatoire, anti-ulcéreuse gastrique, antioxydante et potentiellement éclaircissante de la peau [7]. Il s'agit donc d'une plante utilisée depuis longtemps en médecine, assez bien décrite en termes de contenu, dépourvue de toxicité (sauf hypertension et insuffisance rénale en cas de consommation excessive) et à fort potentiel pour différents champs d'applications dont celui de la cosmétique.

4.3 Analyses LC-HRMS

Pour assurer au mieux l'analyse des six pics chromatographiques annotés de l'extrait de réglisse tout en cherchant à mettre en évidence les isoflavones, nous avons mis au point une méthode d'analyse adaptée aux analyses UHPLC-HRMS à venir et qui permette aussi d'obtenir des informations sur les composés non séparés et peu retenus. Les conditions d'analyses chromatographiques sont indiquées en Annexe 3.7.

Les analyses permettent d'obtenir les profils chromatographiques en UV et en spectrométrie de masse. Ces derniers sont visibles en Figure 52 et Figure 53. L'extrait possède un profil plutôt riche en UV. Ses composés répondent de façon satisfaisante en spectrométrie de masse dans les deux modes d'ionisation. L'ensemble des pics (**A** à **F**) qu'on souhaite identifier répondent en ESI-MS.

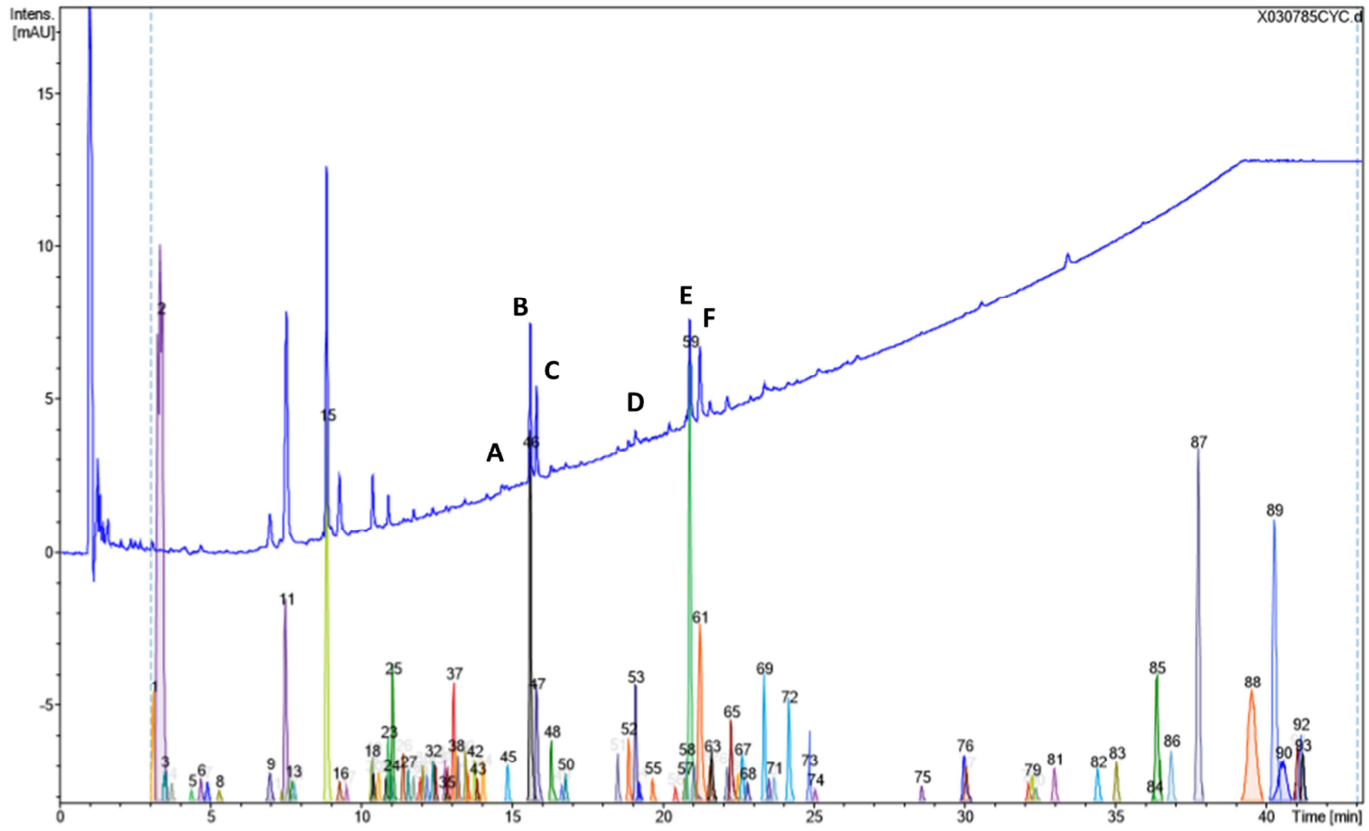


Figure 52 Chromatogrammes UV 250 nm (tracé bleu) et MS en mode positif de l'extrait liquide de réglisse analysé à 0,1 % de matière sèche. Indications des pics à identifier (A à F)

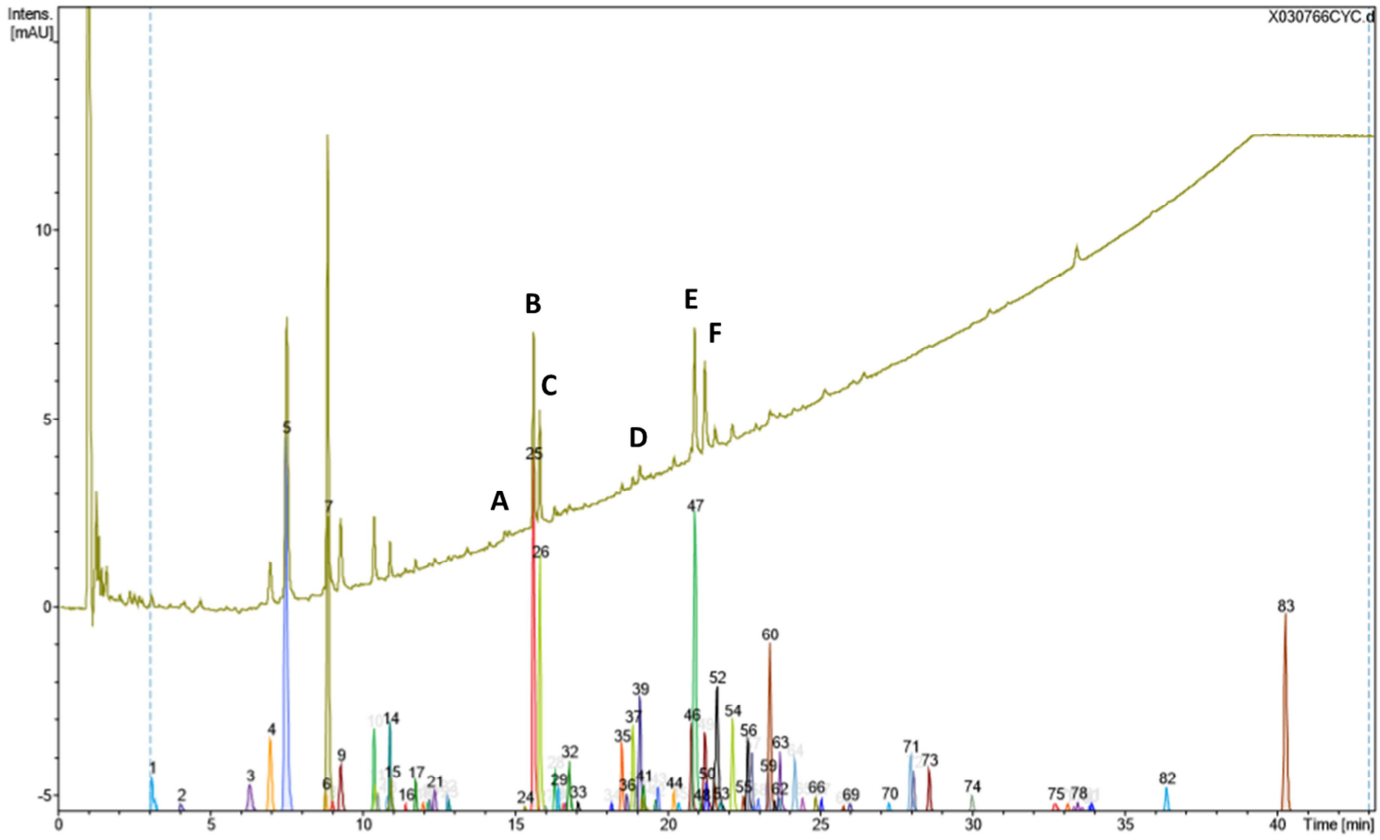


Figure 53 Chromatogrammes UV 250 nm (tracé vert) et MS en mode négatif de l'extrait liquide de réglisse analysé à 0,1 % de matière sèche. Indications des pics à identifier (**A** à **F**)

4.4 Identification des pics principaux

GAINS a été utilisé pour analyser l'extrait de racines de réglisse afin de proposer des structures possibles pour les six principaux pics chromatographiques. Les structures possibles mises en évidence ont notamment été confrontées aux composés connus de *Glycyrrhiza glabra* (données GPDB après veille bibliographique afin de mettre à jour le référencement des molécules connues pour la plante).

4.4.1. Note concernant la synthèse des résultats et numérotations des pics

Les tableaux de synthèse des paragraphes suivants rassemblent les données collectées en LC-UV-HRMS des chromatogrammes des Figure 52 et Figure 53. Comme pour les autres études, les pics ont été numérotés dans l'ordre de leur rétention et de façon indépendante en mode positif et en mode négatif. On conserve les deux numérotations avec à la suite du numéro du pic, la lettre « P » en mode positif et la lettre « N » en mode négatif.

4.4.2. Résultats du traitement des données de LC-HRMS

Dans un premier temps, le traitement des données a concerné les principaux pics chromatographiques (A à F) de l'extrait. Les masses monoisotopiques ont été confrontées aux données de la GPDB sur la plante et le genre *Glycyrrhiza*. Des propositions d'identification ont été faites et étayées par les données de fragmentation quand cela était possible.

Pour chaque pic chromatographique, le

Tableau 24 regroupe les informations suivantes : le temps de rétention moyen entre les deux modes (t_R), la masse monoisotopique, le maximum d'absorbance du composé, les informations structurales obtenues par les données de fragmentation et enfin la proposition d'identification. Dans cette même colonne, les molécules mentionnées en **vert** signalent que la structure est décrite dans la littérature pour ***Glycyrrhiza glabra* L.** et celles en **bleu** indiquent des molécules appartenant au **genre *Glycyrrhiza*** excepté *Glycyrrhiza glabra*.

Pic	t_R (min)	n°pic mode positif	n°pic mode négatif	Masse monoisotopique ($g.mol^{-1}$)	Maxima d'absorbance (nm) >250nm	Partie aglycone	Partie glycosylée	Proposition d'identification
A	14,87	45P	/	212,102161	nd	pas d'information	pas d'information	aucun candidat (faible réponse en masse)
B	15,61	46P	25N	430,126313	261 sh ; 301	Hydroxy-methoxy-isoflavone Génine 268,07287 $g.mol^{-1}$	Hexose	**Ononin** ou **Isoononin**
C	15,82	47P	26N	460,136530	256 sh ; 321	Hydroxy-dimethoxyisoflavone Génine 298,08379 $g.mol^{-1}$	Hexose	**Hydroxy-dimethoxyisoflavone glycosylé (hexose)**
D	19,10	53P	39N	270,088683	(300) ; 365	Dihydroxy-methoxy-chalcone	Aucune	**Echinatin**
E	20,90	59P	47N	268,073023	265 sh ; 301	Hydroxy-methoxy-isoflavone génine : 268,07287 $g.mol^{-1}$	Aucune	**Formononetin** ou **Isoformononetin**
F	21,23	61P	49N	298,083590	267 sh ; 312	Hydroxy-dimethoxy-isoflavone	Aucune	**Hydroxy-dimethoxyisoflavone**

Tableau 24 Données analytiques des principaux pics chromatographiques (**A** à **F**) de l'extrait glycériné de réglisse glabre (nd : non déterminé, sh : *shoulder* épaulement)

Les maxima d'absorbance des pics **B** à **F** sont cohérents avec les propositions effectuées.

Le pic **A** est un pic chromatographique de très faible intensité en UV et en masse. On n'a pas de données pertinentes pour le caractériser : faible réponse en masse avec une hypothèse faite sur l'ion le plus intense du spectre de masse positif considéré sous la forme $[M+H]^+$. Aucune structure candidate ne ressort dans la base de données Greenpharma pour la masse proposée.

Le pic **D** a un maximum d'absorbance de 365 nm ce qui peut indiquer la présence d'un flavonoïde de type chalcone. La masse monoisotopique -déduite des données de MS (Figure 55) et mesurée à 270,088683 $g.mol^{-1}$ - conduit à proposer l'échinatine (voir Figure 54) figurant dans la liste des candidats de l'organisme. La fragmentation en positif (voir Figure 56) étaye notamment la proposition d'une chalcone avec deux ruptures en rétro Diels-Alder qui peuvent survenir: en effet, on a observé les ions suivants de m/z 121,0284 (ion majoritaire) et 177,0539 correspondant respectivement aux fragments type $^{0,\alpha}$A et $^{0,1'}$B d' une chalcone. Les ions contenant les cycles A et B sont indiqués par les majuscules A et B et les indices indiquent les atomes de liaisons rompues (d'après [125–127]). Une représentation du noyau chalcone (avec l'indication des cycles A et B d'après [125,128]) constitue la Figure 54.

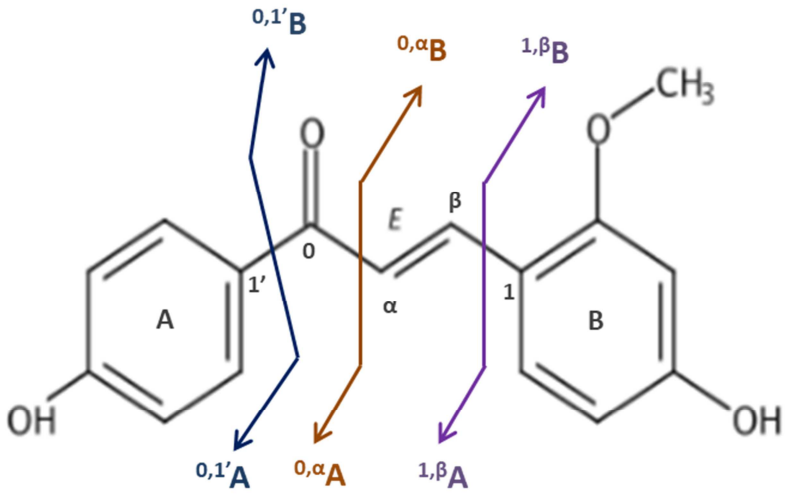


Figure 54 Structure de l'échinatine (= (E)-4,4'-dihydroxy-2-methoxy-chalcone) et voies de fragmentation proposées pour les dérivés de type chalcone d'après [125]

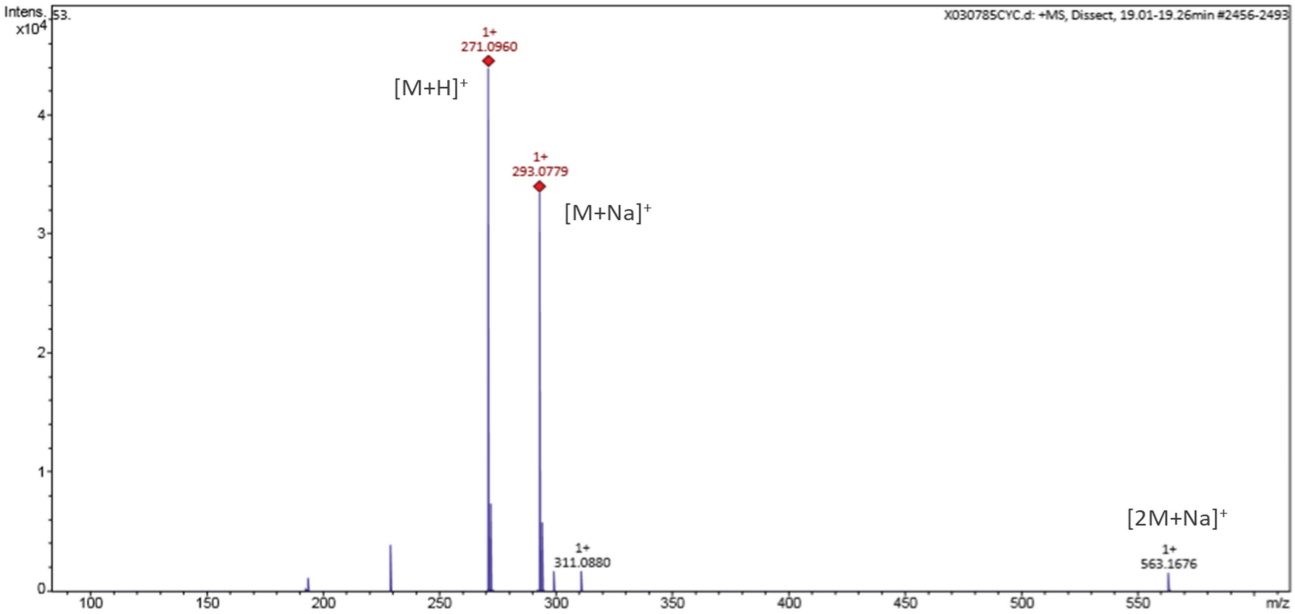


Figure 55 Spectre MS en mode positif du pic **D** (53P)

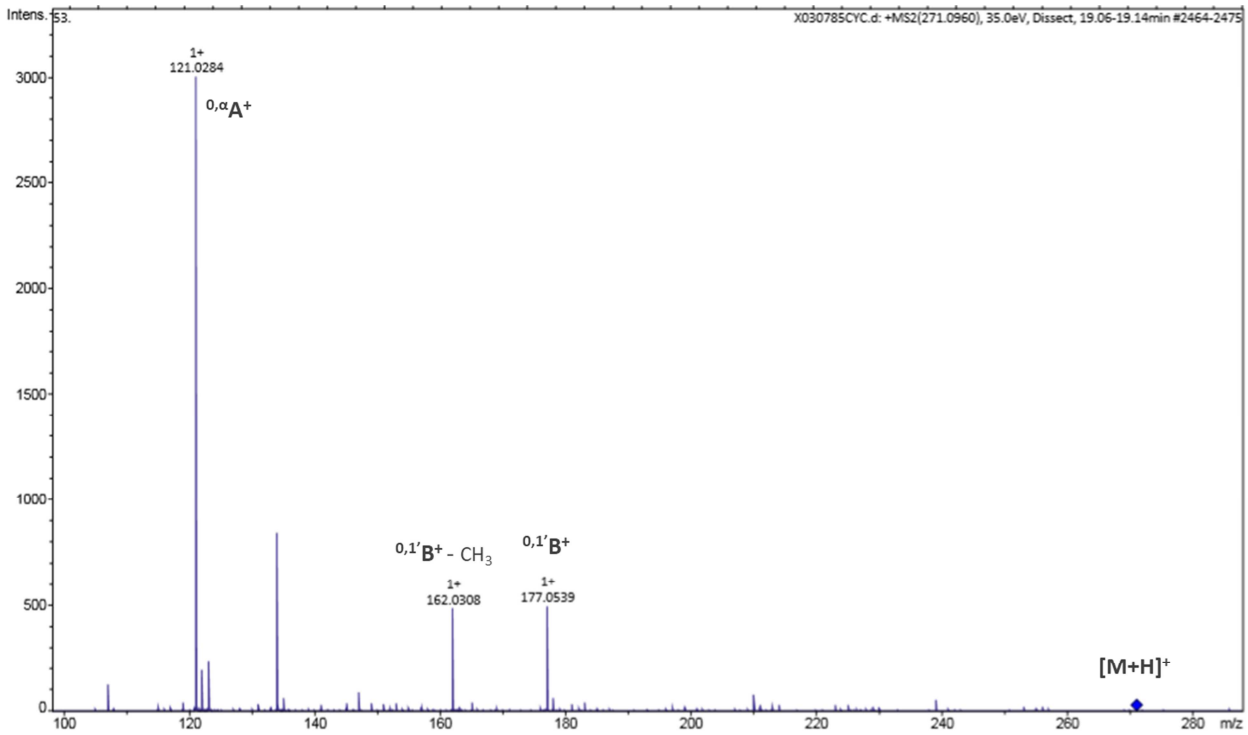


Figure 56 Spectre MS/MS en mode positif du pic **D** (53P) : fragmentation de l'ion $[M+H]^+$

Pour le pic **B,** le composé détecté a une masse exacte de 430,12631 $g.mol^{-1}$, ce qui dans la plante étudiée correspond à l'ononin (ou formononetin-7-O-glucoside). Le spectre MS présente les ions *m/z* 431,1337 $[M+H]^+$, 453,1158 $[M+Na]^+$ et un fragment source intense de *m/z* 269,0804 $[M+H\text{-Hexose}]^+$. Ce fragment source confirme le glucose de la structure proposée. Pour la partie génine, la fragmentation en MS/MS présente des ions conformes à ceux de la littérature pour la formononétine (masse neutre de 268 $g.mol^{-1}$). En effet, on observe un fragment majoritaire de *m/z* de 269,0802 ainsi que les ions *m/z* 254,0568

pour [Formononetin+H-CH_3]$^+$; *m/z* 237,0543 pour [Formononetin+H-CH_4O]$^+$ et *m/z* 213,0910 pour [Formononetin+H-CO-CO]$^+$ annotés d'après la comparaison aux données de la littérature [129].

Une seconde structure proche de celle évoquée précédemment est l'isoononin (aussi appelée isoformononetin-4'-O-glucoside). Elle n'est pas décrite pour l'organisme *Glycyrrhiza glabra* L. précisément mais l'est dans le genre *Glycyrrhiza.* Aussi, il est judicieux de la mentionner. Les données de fragmentation en masse ne permettent pas de déterminer quelle est la molécule en présence (les premiers fragments correspondant à la perte du sucre et celle du methyl de la fonction methoxy engendrent le même noyau). Les deux structures proposées sont visibles en Figure 57 et Figure 58.

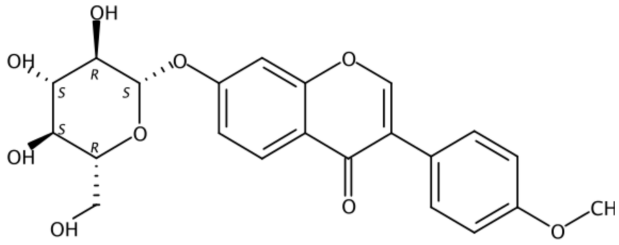


Figure 57 Structure de l'ononin
(décrite pour *Glycyrrhiza glabra*)

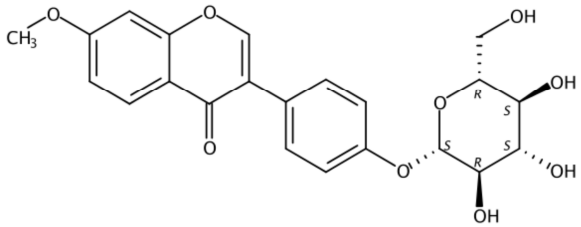


Figure 58 Structure de l'isoononin
(décrite dans le genre *Glycyrrhiza*)

Le pic **E,** de masse 268,073023 g.mol^{-1}, est une molécule non glycosylée. Cette masse correspond à la masse de la génine du pic B et les fragmentations en MS/MS à partir des ions pseudo-moléculaires sont équivalentes. Ainsi le pic **E** peut correspondre soit à la formononétine, soit à l'isoformononétine dont les structures sont présentées en Figure 59 et Figure 60. Dans la littérature, la formononétine est mentionnée pour la réglisse glabre, tandis que l'isoformononétine est décrite pour d'autres espèces du genre *Glycyrrhiza*.

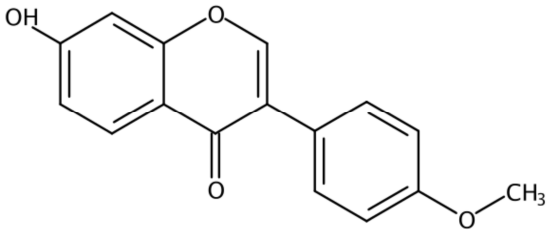


Figure 59 Structure de la formononétine
(= 7-hydroxy-4'-methoxyisoflavone)

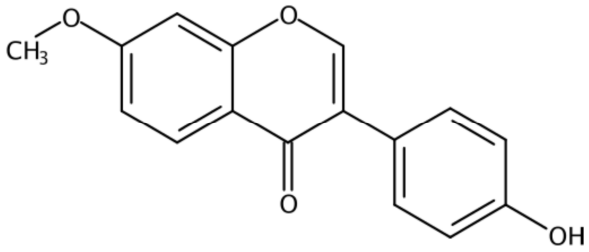


Figure 60 Structure de l'isoformononétine
(= 4'-hydroxy-7-methoxyisoflavone)

Le pic **F** est un flavonoïde qui serait basé sur une hydroxy-dimethoxyisoflavone. En effet, la fragmentation est cohérente avec ce type de structure : en MS/MS en mode négatif (Figure 61), on observe des pertes successives de deux fonctions methoxy et perte de deux fonctions carbonyl. Il y a 12 structures qui correspondent à ces éléments au sein de la base GPDB.

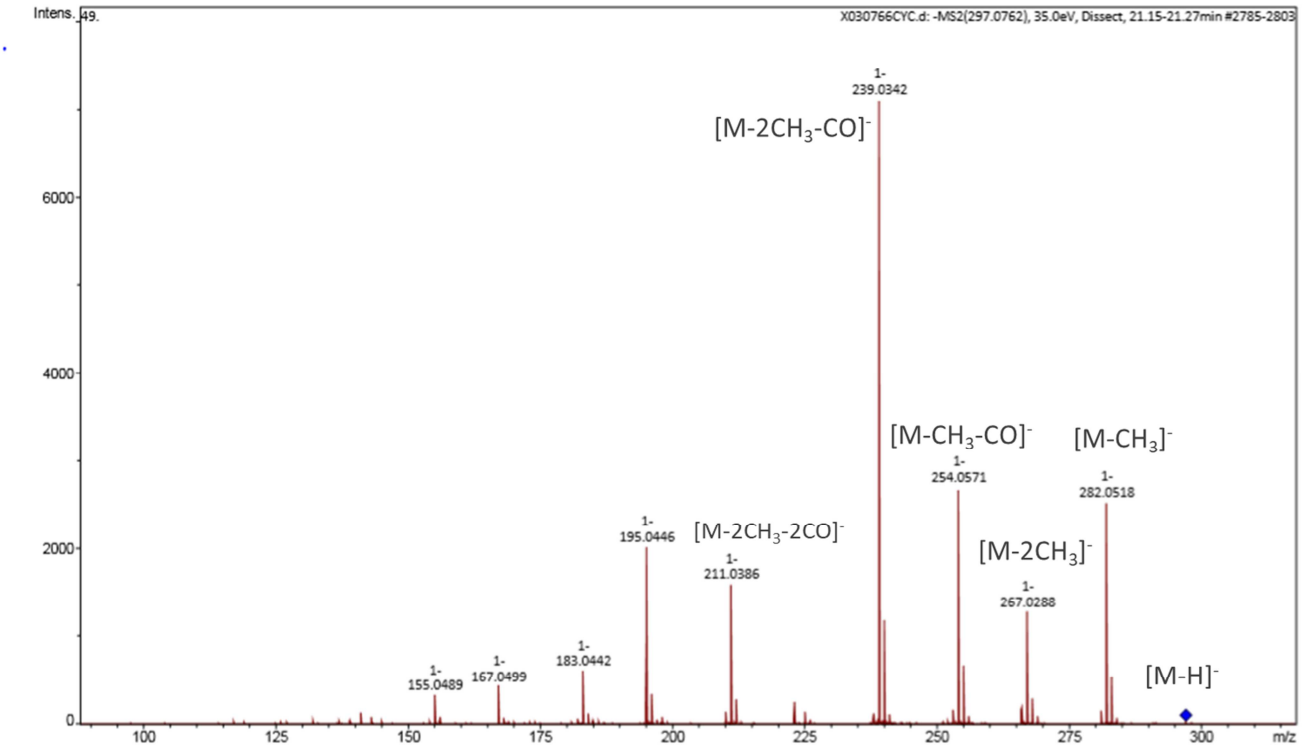


Figure 61 Spectre MS/MS en mode négatif du pic F (49N) : fragmentation de l'ion $[M-H]^-$

Enfin la masse monoisotopique mesurée pour le pic **C** est 460,13653 $g.mol^{-1}$. On détecte la perte d'un hexose. La partie aglycone a une masse de 298,08406 $g.mol^{-1}$. Parmi les molécules de la littérature pour la réglisse glabre, une proposition ressort : c'est la 6''-acetylliquiritin. Cette proposition se révèle non cohérente avec la fragmentation en MS/MS où l'on a observé la perte de l'hexose seul (-162 u), ce qui n'est pas possible s'il porte un acétyle. Aucun autre candidat ne correspond à cette masse dans la littérature de la plante. Par ailleurs, la génine de **C** correspond à celle déterminée pour **F** en termes de masses monoisotopiques et de fragmentations à partir des ions 299 et 297. Ainsi, on suggère que la molécule **C** serait une hydroxy-dimethoxyisoflavone glycosylée avec un hexose. Il y a cinq composés qui correspondent à ces éléments au sein de la base GPDB (structures en Figure 62 à Figure 66).

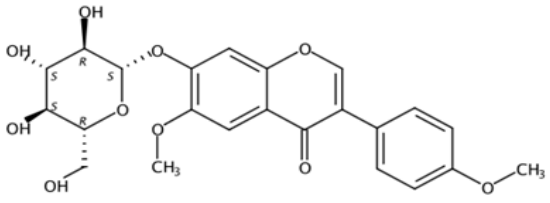


Figure 62 Structure de afromosin 7-O-glucoside

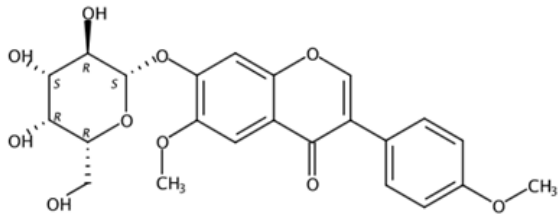


Figure 63 Structure de afromosin 7-O-galactoside

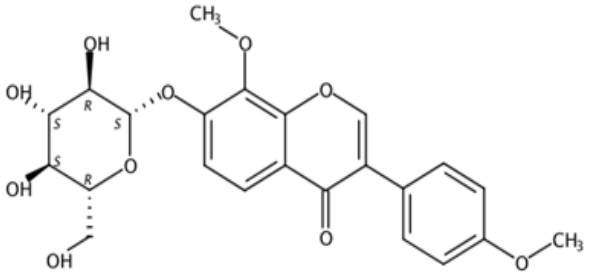


Figure 64 Structure de 4',7,8-trihydroxyisoflavone; 4',8-di-me ether, 7-o-beta-d-glucopyranoside

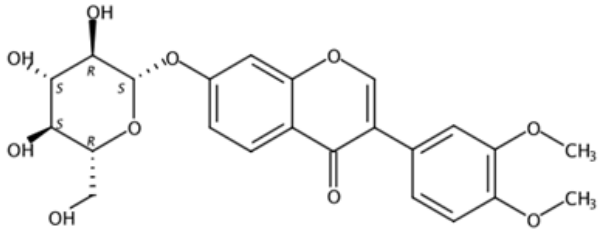


Figure 65 Structure de 3',4',7-trihydroxyisoflavone; 3',4'-di-me ether, 7-o-beta-d-glucopyranoside

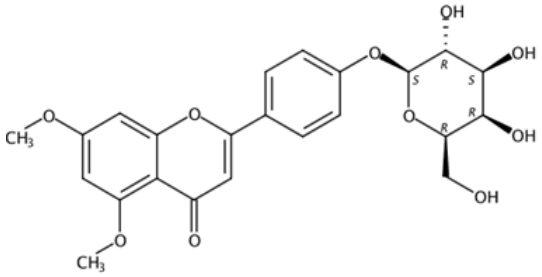


Figure 66 Structure de 4',5,7-trihydroxyflavone; 5,7-di-me ether, 4'-o-beta-d-galactopyranoside

4.4.3. Confirmation de la détection d'isoflavones : différence de fragmentation par rapport aux flavones

Concernant les pics **B**, **C**, **E** : on observe en MS/MS, une perte de 56 unités de masse à partir de l'ion pseudo-moléculaire $[M+H]^+$. Cette perte est caractéristique des isoflavones, elle correspond à la double élimination de CO à partir du cycle C des isoflavones. Ce type de fragmentation les distingue des flavones de façon non ambiguë, comme l'expose la publication de Kuhn au travers de la proposition des mécanismes de fragmentation [130]. En effet, les réarrangements impliqués ne sont pas équivalents entre les deux types de composés. La perte de 56 unités observée grâce aux données MS/MS nous a ainsi permis de proposer ou confirmer la présence des isoflavones plutôt que des flavones.

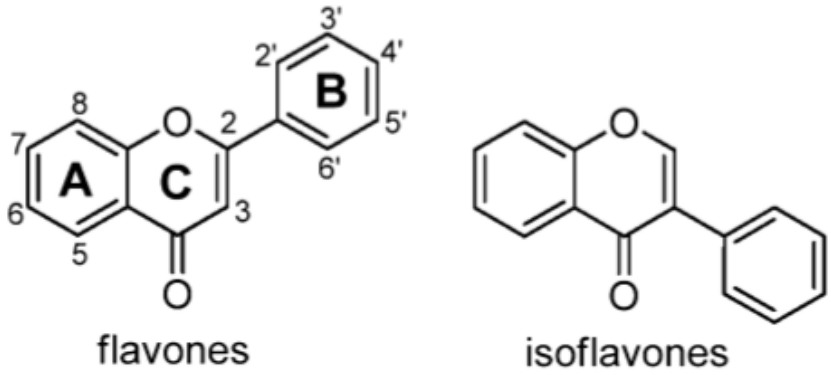


Figure 67 Structure des flavones et des isoflavones d'après [131] : enchaînement en C6-C3-C6. La liaison entre le cycle C et le cycle B différencient ces deux classes de flavonoïdes.

Chapitre 3.

Aucune donnée de fragmentation MS/MS en positif n'est disponible pour **F**. Par contre, on observe le *m/z* = 211,0386 en négatif qui correspond à une perte de deux CO consécutive à la perte de deux methyl ($[M-H-CH_3-CH_3-2CO]^-$). Ceci nous indique à nouveau une isoflavone puisque la perte de 2 CO a également été décrite lors de la fragmentation d'isoflavones analysées en mode négatif [132].

4.5 Fonction de recherche d'une famille phytochimique particulière : exemple avec les isoflavones

4.5.1. Présentation et intérêts

Lors de la description du contenu métabolique d'un extrait, utiliser une stratégie adossée à une base de données de composés naturels rend possible la mise en évidence de la présence ou l'absence de familles phytochimiques particulières. Au sein de la base GPDB, il nous est possible d'effectuer une requête des masses monoisotopiques au sein d'une famille phytochimique plutôt que dans la totalité de la base. Ceci constitue une fonction utile de GAINS pour gérer des aspects de veille toxicologique, réglementaire ou répondre à une volonté marketing. En effet, au moment de développer un ingrédient, la ou les familles phytochimiques constitutives de l'extrait peuvent représenter un avantage ou au contraire une contrainte en fonction de son activité potentielle ou connue, son originalité, sa probable toxicité ou encore ses propriétés physico-chimiques.

4.5.2. Exemple d'application : cas des isoflavones

Dans un second temps, l'étude de l'extrait de réglisse s'est focalisée sur la recherche d'isoflavones afin de confirmer ou infirmer leur présence. Les isoflavones appartiennent à la grande famille des flavonoïdes. Au niveau structural, elles ont la particularité de présenter un cycle B en position 3 du cycle C (structures des noyaux présentés en Figure 67).

Dans le cas de l'étude de l'extrait de réglisse, nous voulions avoir une vue d'ensemble générale sur les pics du profil et mettre en évidence les composés qui pourraient être des isoflavones. Nous avons donc étendu nos recherches de candidats pour l'ensemble des pics du chromatogramme à la totalité des composés de la base GPDB. Nous avons ainsi pu reporter la présence de candidats isoflavones ou l'absence de candidat de type isoflavone ou encore l'absence de candidat en général. Pour rappel, parmi les isoflavones, une attention particulière concernait la daïdzeine.

Le Tableau 25 répertorie le temps de rétention, la masse monoisotopique de l'ensemble des pics détectés en LC-HRMS (autres que les pics principaux **A** à **F**) et la présence potentielle d'isoflavones quand des structures de cette famille phytochimique sont présentes dans les résultats de la recherche. Dans ce dernier cas, la ligne apparaît en vert. Toutes les structures d'isoflavones détectées sont référencées dans le tableau.

La mention ''aucun candidat'' indique l'absence de structure de la masse correspondante dans la base de données interrogée et la mention ''aucune isoflavone'' signale qu'aucun des candidats issus de la base de données ne fait partie de cette classe. De plus, les molécules en **jaune** indiquent les structures décrites pour *Glycyrrhiza glabra* L. et ce, quelle que soit leur famille phytochimique (pas de restriction aux isoflavones). Ainsi, une molécule en jaune et surlignée en vert signale une isoflavone décrite pour la réglisse glabre.

Pic positif	Pic négatif	t_R positif (min)	t_R négatif (min)	Masse monoisotopique ($g.mol^{-1}$)	Candidats
	1N		3,09	**316,078542**	Aucune isoflavone
1P		3,11		**117,078834**	Aucune isoflavone **Triméthylglycine** *
2P		3,37		**156,075554**	Aucun candidat
	2N		4,05	**302,063046**	Aucun candidat
6P		4,69		**223,070250**	Aucune isoflavone
7P		4,91		**222,049975**	Aucun candidat
8P		5,30		**187,062980**	Aucune isoflavone
	3N		6,30	**286,067888**	Aucune isoflavone
9P	4N	6,99	6,98	**212,068118**	Aucune isoflavone
11P	5N	7,49	7,49	**210,05200**	Aucune isoflavone
12P		7,59		**120,021193**	Aucun candidat
13P		7,72		**187,062908**	Aucune isoflavone
14P		7,79		**196,121209**	Aucune isoflavone
	6N		8,79	**418,109863**	Aucune isoflavone
15P	7N	8,86	8,86	**212,068149**	Aucune isoflavone
	8N		9,02	**196,072880**	Aucune isoflavone
	9N		9,29	**154,025734**	Aucune isoflavone
18P		10,34		**210,136250**	Aucune isoflavone
19P	10N	10,38	10,38	**228,062477**	Aucune isoflavone
20P	11N	10,43	10,412	**162,030886**	Aucune isoflavone **7-Hydroxycoumarin**
	12N		10,50	**434,120746**	Aucune isoflavone
21P		10,59		**210,136118**	Aucune isoflavone
22P	13N	10,84	10,82	**240,099229**	Aucune isoflavone
23P	14N	10,91	10,90	**594,157857**	4',7-di-o-beta-d-glucopyranoside-2',4',7-trihydroxyisoflavone; 4'-o-d-glucopyranoside-8-glucopyranosyl-4',5,7-trihydroxyisoflavone ; genistein 4',7-diglucoside ; genistein 7-glucosylglucoside ; paniculatin
24P		10,99		**414,245799**	Aucun candidat
25P		11,06		**210,13628**	Aucune isoflavone

* La triméthylglycine est utilisée comme additif lors de l'extraction pour obtenir l'extrait analysé. La molécule est également décrite dans la littérature de la réglisse.

Pic positif	Pic négatif	t_R positif (min)	t_R négatif (min)	Masse monoisotopique ($g.mol^{-1}$)	Candidats
26P	16N	11,40	11,41	**462,115137**	7-o-beta-d-glucopyranoside-3'-me ether-3',4',5,7-tetrahydroxyisoflavone ; 8-o-beta-d-glucopyranoside-3-me ether-3',4',7,8-tetrahydroxyisoflavone ; 7-o-beta-d-glucopyranoside-8-me ether-4',5,7,8-tetrahydroxyisoflavone ; 4'-o-beta-d-glucopyranoside-4',5,7-trihydroxy-6-methoxyisoflavone ; 8-c-glucopyranosyl-4',5,7-trihydroxy-3'-methoxyisoflavone ; pratensein 7-o-glucoside ; tectoridin
27P		11,57		**458,272218**	Aucun candidat
28P	17N	11,74	11,74	**564,147301**	6''-o-apiosyl-8-glucopyranosyl-4',5,7-trihydroxyisoflavone; ambocin ; neobacin ; **isoschaftoside ou schaftoside**
29P		11,96		**564,275286**	Aucune isoflavone
	18N		11,99	**152,04705**	Aucune isoflavone **2,4'-dihydroxyacetophenone ; 4-hydroxyphenylacetic acid ; methyl salicylate**
30P		12,05		**502,298417**	Aucun candidat
	19N		12,15	**348,083215**	Aucune isoflavone
	20N		12,21	**436,136156**	Aucune isoflavone
32P	21N	12,39	12,38	**420,08398**	Aucune isoflavone
33P		12,48		**546,324572**	Aucune isoflavone
34P	22N	12,78	12,78	**450,151530**	Aucune isoflavone
35P	23N	12,83	12,83	**432,104205**	7-o-beta-d-glucopyranoside-2',5,7-trihydroxyisoflavone; 4'-o-beta-d-glucopyranoside-4',6,7-trihydroxyisoflavone; 8-glucopyranosyl-3',4',7-trihydroxyisoflavone ; 8-glucopyranosyl-4',5,7-trihydroxyisoflavone ; genistein 4'-glucoside ; genistein 5-glucoside ; genistein 7-glucoside ; **vitexin**
36P		12,87		**590,350777**	Aucune isoflavone
37P		13,08		**196,109703**	Aucune isoflavone
38P		13,15		**674,372805**	Aucune isoflavone
40P		13,47		**718,398812**	Aucun candidat
41P		13,55		**678,403450**	Aucune isoflavone
42P		13,77		**762,425245**	Aucune isoflavone
43P		13,85		**722,43049**	Aucune isoflavone
44P		14,05		**806,451557**	Aucune isoflavone
	24N		15,32	**238,083439**	Aucune isoflavone
	27N		16,00	**464,131107**	Aucune isoflavone
48P	28N	16,31	16,31	**254,057190**	4',8-dihydroxyisoflavone ; **4',7-dihydroxyisoflavone = daidzein ; 4',7-dihydroxyflavone**

Pic positif	Pic négatif	t_R positif (min)	t_R négatif (min)	Masse monoisotopique ($g.mol^{-1}$)	Candidats
	29N		16,42	**434,120663**	Aucune isoflavone
	30N		16,59	**270,052264**	2',4',7-trihydroxyisoflavone ; 2',5,7-trihydroxyisoflavone ; 3',4',7-trihydroxyisoflavone ; 3',5,7-trihydroxyisoflavone ; 4',6,7-trihydroxyisoflavone (6-hydroxydaidzein) ; 4',7,8-trihydroxyisoflavone ; genistein
49P	31N	16,65	16,65	**284,067530**	2',4'-dihydroxy-7-methoxyisoflavone ; 2',7-dihydroxy-4'-methoxyisoflavone ; 2',7-dihydroxy-6-methoxyisoflavone ; 3',7-dihydroxy-4'-methoxyisoflavone ; **4',5-dihydroxy-7-methoxyisoflavone** ; 4',6-dihydroxy-7-methoxyisoflavone ; 4',7-dihydroxy-2'-methoxyisoflavone ; 4',7-dihydroxy-3'-methoxyisoflavone ; 4',7-dihydroxy-5-methoxyisoflavone ; 4',7-dihydroxy-6-methoxyisoflavone ; 5,7-dihydroxy-4'-methoxyisoflavone ; 6,7-dihydroxy-4'-methoxyisoflavone ; 7,8-dihydroxy-4'-methoxyisoflavone ; biochanin A ; calycosin **genkwanin** (= **4',5-dihydroxy-7-methoxyflavone)**
50P	32N	16,79	16,78	**256,072365**	Aucune isoflavone **Liquiritigenin** **2',4,4'-trihydroxychalcone** **(S)-4',7-dihydroxyflavanone** **Pinocembrin (= 5,7-Dihydroxyflavanone)**
	33N		17,06	**286,083019**	Aucune isoflavone **Licochalcone B**
	34N		18,17	**280,129526**	Aucune isoflavone
51P	35N	18,51	18,51	**280,129844**	Aucune isoflavone
	36N		18,67	**292,130363**	Aucune isoflavone
52P	37N	18,87	18,87	**296,125448**	Aucune isoflavone
	38N		18,97	**838,397900**	Aucune isoflavone
54P	40N	19,22	19,18	**284,067158**	2',4'-dihydroxy-7-methoxyisoflavone ; 2',7-dihydroxy-4'-methoxyisoflavone ; 2',7-dihydroxy-6-methoxyisoflavone ; 3',7-dihydroxy-4'-methoxyisoflavone ; **4',5-dihydroxy-7-methoxyisoflavone** ; 4',6-dihydroxy-7-methoxyisoflavone ; 4',7-dihydroxy-2'-methoxyisoflavone ; 4',7-dihydroxy-3'-methoxyisoflavone ; 4',7-dihydroxy-5-methoxyisoflavone ; 4',7-dihydroxy-6-methoxyisoflavone ; 5,7-dihydroxy-4'-methoxyisoflavone ; 6,7-dihydroxy-4'-methoxyisoflavone ; 7,8-dihydroxy-4'-methoxyisoflavone ; biochanin A ;

Pic positif	Pic négatif	t_R positif (min)	t_R négatif (min)	Masse monoisotopique ($g.mol^{-1}$)	Candidats
					calycosin **genkwanin (4',5-dihydroxy-7-methoxyflavone)**
	41N		19,22	**914,450810**	Aucune isoflavone
55P	42N	19,66	19,62	**340,130180**	Aucune isoflavone **(S)-4',5,7-trihydroxy-3'-prenylflavanone**
	43N		19,69	**840,413916**	Aucune isoflavone
	44N		20,21	**206,092983**	Aucune isoflavone
	45N		20,38	**296,125234**	Aucune isoflavone
57P		20,74		**230,094228**	Aucune isoflavone
58P	46N	20,79	20,79	**296,125443**	Aucune isoflavone
60P		21,04		**268,073036**	4'-hydroxy-7-methoxyisoflavone ; 5-hydroxy-4'-methoxyisoflavone ; 5-hydroxy-7-methoxyisoflavone ; **formononetin (= 7-hydroxy-4'-methoxyisoflavone)**
	48N		21,11	**272,103464**	Aucune isoflavone
	50N		21,28	**488,145621**	Aucune isoflavone
	51N		21,55	**284,068050**	2',4'-dihydroxy-7-methoxyisoflavone ; 2',7-dihydroxy-4'-methoxyisoflavone ; 2',7-dihydroxy-6-methoxyisoflavone ; 3',7-dihydroxy-4'-methoxyisoflavone ; **4',5-dihydroxy-7-methoxyisoflavone** ; 4',6-dihydroxy-7-methoxyisoflavone ; 4',7-dihydroxy-2'-methoxyisoflavone ; 4',7-dihydroxy-3'-methoxyisoflavone ; 4',7-dihydroxy-5-methoxyisoflavone ; 4',7-dihydroxy-6-methoxyisoflavone ; 5,7-dihydroxy-4'-methoxyisoflavone ; 6,7-dihydroxy-4'-methoxyisoflavone ; 7,8-dihydroxy-4'-methoxyisoflavone ; biochanin A ; calycosin ; **genkwanin (4',5-dihydroxy-7-methoxyflavone)**
62P		21,57		**222,088604**	Aucune isoflavone
63P	52N	21,64	21,63	**222,124319**	Aucune isoflavone
	53N		21,74	**256,071982**	Aucune isoflavone **2',4,4'-trihydroxychalcone ;** **(S)-4',7-dihydroxyflavanone ;** **liquiritigenin ;** **pinocembrin**
64P		22,14		**148,015752**	Aucun candidat
	54N		22,14	**222,088410**	Aucune isoflavone
65P		22,27		**273,266609**	Aucun candidat
66P	55N	22,50	22,49	**322,11964**	4',7-dihydroxy-3'-prenylisoflavone ; 4',7-dihydroxy-8-prenylisoflavone ; 7-hydroxy-4'-prenyloxyisoflavone ; **shinpterocarpin ;**

Pic positif	Pic négatif	t_R positif (min)	t_R négatif (min)	Masse monoisotopique ($g.mol^{-1}$)	Candidats
					glabrene
67P	56N	22,64	22,63	**250,119325**	Aucune isoflavone
	57N		22,76	**988,524657**	Aucune isoflavone
	58N		22,97	**664,382355**	Aucune isoflavone
	59N		23,29	**912,508449**	Aucune isoflavone **soyasaponin II**
69P	60N	23,37	23,36	**324,135612**	Aucune isoflavone **glabridin ;** **glabranin ;** **(S)-5,7-dihydroxy-6-prenylflavanone ;** **4',7-dihydroxy-3'-prenylflavanone ;** **phaseollinisoflavan**
70P	61N	23,53	23,52	**322,119906**	4',7-dihydroxy-3'-prenylisoflavone ; 4',7-dihydroxy-8-prenylisoflavone ; 7-hydroxy-4'-prenyloxyisoflavone ; **Shinpterocarpin ;** **glabrene**
	62N		23,66	**354,145941**	Aucune isoflavone **3'-hydroxy-4'-methoxyglabridin**
	63N		23,70	**294,181964**	Aucune isoflavone
72P	64N	24,19	24,18	**310,17742**	Aucune isoflavone
	65N		24,44	**282,051689**	2'-hydroxy-6,7-methylenedioxyisoflavone ; 7-hydroxy-3',4'-methylenedioxyisoflavone ; corylinal 5-hydroxy-6,7-methylenedioxyflavone ;
73P	66N	24,85	24,87	**324,135583**	Aucune isoflavone **glabridin ;** **glabranin ;** **(S)-5,7-dihydroxy-6-prenylflavanone ;** **4',7-dihydroxy-3'-prenylflavanone ;** **phaseollinisoflavan**
74P	67N	25,05	25,05	**354,146040**	Aucune isoflavone **3'-hydroxy-4'-methoxyglabridin**
	69N		25,99	**495,97563**	Aucun candidat
	70N		27,28	**352,129792**	4',5-dihydroxy-7-methoxy-6-prenylisoflavone ; 3'-hydroxy-4'-methoxy-7-prenyloxyisoflavone ; 4',7-dihydroxy-5-methoxy-8-prenylisoflavone ; 5,7-dihydroxy-4'-methoxy-6-prenylisoflavone ; Erylatissin A ; **1-methoxyphaseollin**
	71N		27,99	**278,187535**	Aucune isoflavone
	72N		28,09	**340,166483**	Aucune isoflavone
	73N		28,59	**324,135670**	Aucune isoflavone **glabridin ;** **glabranin ;** **(S)-5,7-dihydroxy-6-prenylflavanone ;** **4',7-dihydroxy-3'-prenylflavanone ;** **phaseollinisoflavan**

Pic positif	Pic négatif	t_R positif (min)	t_R négatif (min)	Masse monoisotopique ($g.mol^{-1}$)	Candidats
75P		28,60		**268,072711**	4'-hydroxy-7-methoxyisoflavone ; 5-hydroxy-4'-methoxyisoflavone ; 5-hydroxy-7-methoxyisoflavone ; **formononetin (= 7-hydroxy-4'-methoxyisoflavone)**
76P	74N	30,00	30,00	**234,161543**	Aucune isoflavone
77P		30,09		**248,177262**	Aucune isoflavone
78P		32,14		**300,133447**	Aucune isoflavone
79P		32,25		**300,133554**	Aucune isoflavone
80P		32,37		**300,133560**	Aucune isoflavone
	75N		32,73	**533,930835**	Aucun candidat
81P		32,99		**294,180013**	Aucun candidat
	76N		33,14	**300,265818**	Aucune isoflavone
	78N		33,47	**312,207720**	Aucune isoflavone
	81N		33,92	**298,249951**	Aucune isoflavone
82P		34,43		**308,195519**	Aucune isoflavone
83P		35,05		**301,237451**	Aucun candidat
84P		36,29		**353,310345**	Aucun candidat
85P	82N	36,39	36,39	**228,208538**	Aucune isoflavone **tetradecanoic acid**
86P		36,86		**255,255457**	Aucune isoflavone
87P		37,76		**281,271654**	Aucune isoflavone
88P		39,53		**337,333968**	Aucune isoflavone
89P	83N	40,29	40,28	**256,239790**	Aucune isoflavone **hexadecanoic acid**
90P		40,58		**412,258614**	Aucune isoflavone
91P		41,07		**337,333942**	Aucune isoflavone
92P		41,17		**283,286791**	Aucun candidat
93P		41,23		**337,333880**	Aucune isoflavone

Tableau 25 Données LC-HRMS sur l'extrait glycériné de réglisse glabre : masses monoisotopiques et recherche de la famille des isoflavones

L'extrait de réglisse glabre contient probablement des composés de la famille des isoflavones aux vues de la correspondance entre les masses monoisotopiques détectées et les candidats de même masse (avec une tolérance de 5 ppm) référencés dans la base de données Greenpharma.

Ce sont une quinzaine de pics chromatographiques qui peuvent être des isoflavones et pour certains, les isoflavones possibles ont déjà été décrites dans la plante. Ces hypothèses d'isoflavones concernent des pics chromatographiques de second ordre dont les spectres de masses sont globalement d'intensité moyenne.

4.5.3. Focus sur une molécule particulière : présence ou absence de la DAIDZEINE

La daïdzeine (ou 4',7-dihydroxyisoflavone) a une masse monoisotopique de 254,057910 $g.mol^{-1}$. Le composé 48P/28N de temps de rétention de 16,31 minutes correspond à une masse monoisotopique mesurée de 254,057190 $g.mol^{-1}$ (soit un écart de 2,8 ppm entre masse théorique et expérimentale). Son maximum d'absorbance en UV, bien que faible en intensité, est mesuré à 330 nm.

Les données de fragmentation du pic 48P/28N suggèrent une molécule de type di-hydroxyisoflavone avec un hydroxyle porté par le cycle A et un autre par le cycle B. En effet, suite à une réaction rétro-Diels-Alder illustrée par la Figure 68 d'après la revue de Vacek [133] et suivant la nomenclature proposée pour la fragmentation des isoflavones selon Ma *et al*. en 1998 [134], on observe les ions $^{1,3}A^+$ (*m/z* 137) et $^{1,3}B^+$ (*m/z* 119) pour le pic 48P (voir spectre MS/MS en positif de la Figure 69). L'équivalent de cette réaction a été observé en négatif avec les ions de *m/z* 135 et 117 (spectre non montré). Pour les aglycones, les étiquettes $^{i,j}A^+$ et $^{i,j}B^+$ désignent les ions produits contenant respectivement les cycles A et B intacts. Les indices i et j indiquent les liaisons du cycle C ayant été rompues.

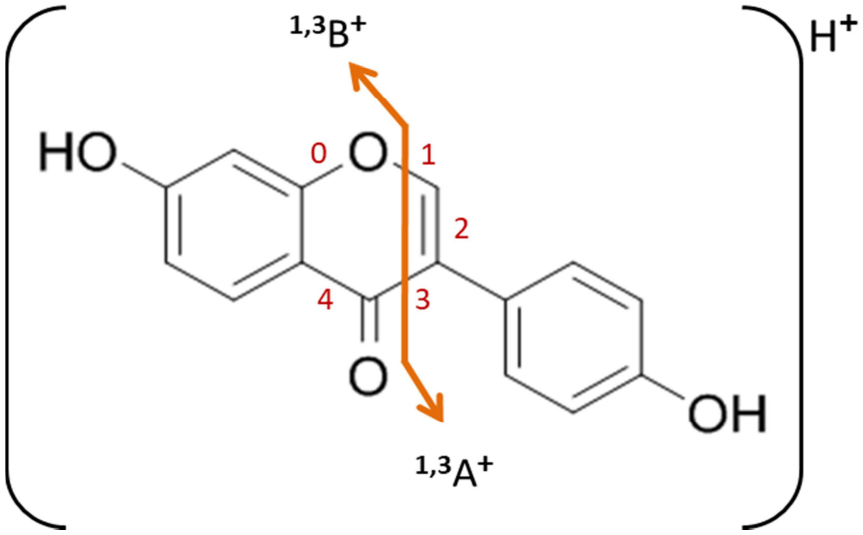


Figure 68 Fragmentation d'une di-hydroxyisoflavone selon une réaction rétro-Diels-Alder ; cas de la Daïdzeine à partir de $[M+H]^+$

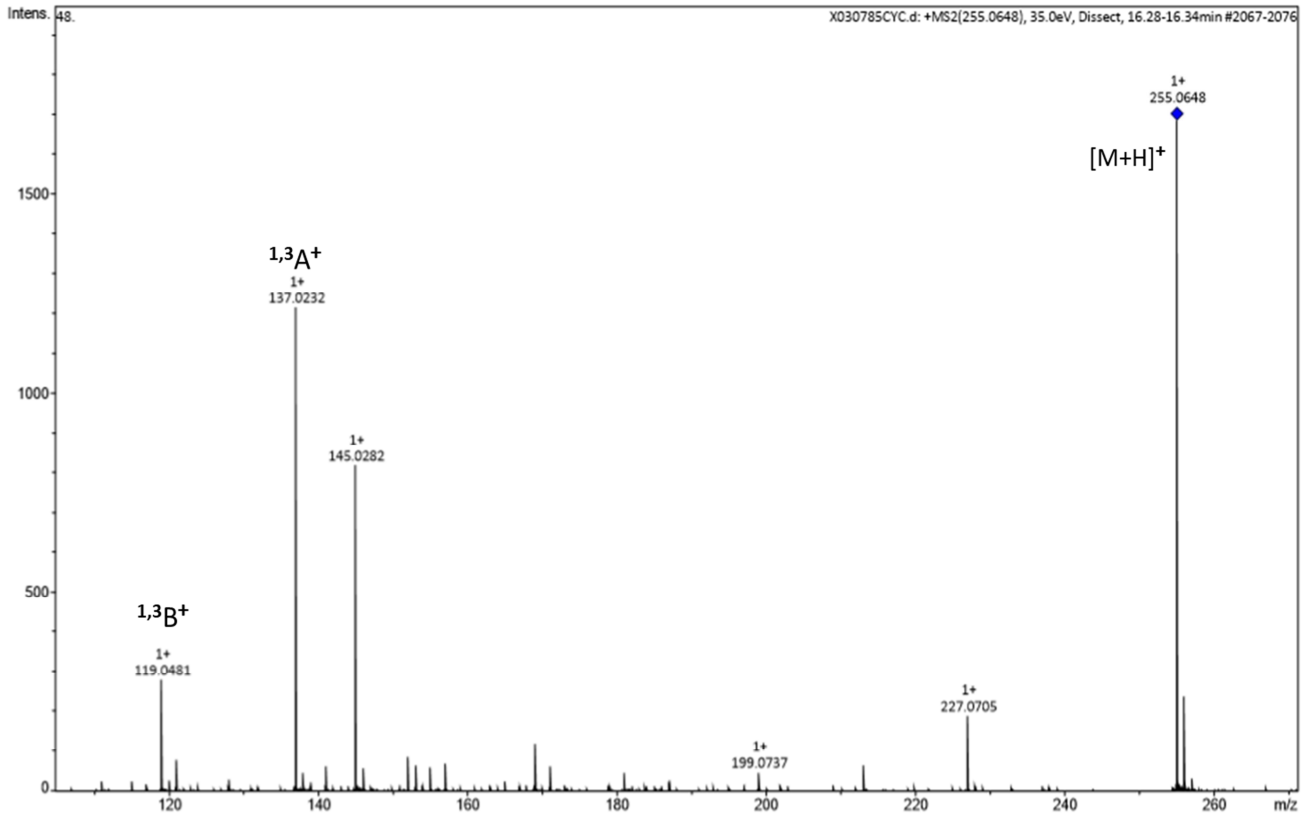


Figure 69 Spectre MS/MS du pic 48P à partir de l'ion $[M+H]^+$: observation des ions issus d'un réaction rétro-Diels-Alder

Seules quatre isoflavones de la base GPDB peuvent correspondre au pic 48P/28N. Parmi elles, la daïdzeine et la 4',8-dihydroxyisoflavone présentent un hydroxy sur chacun des cycles A et C, élément structural mis en évidence en MS/MS et expliqué au paragraphe précédent. Enfin, des deux possibilités restantes, seule la daïdzeine est décrite pour *Glycyrrhiza glabra* L. Ainsi, le pic 48P/28N est très probablement le signal de la daïdzeine.

On ajoute que la recherche de candidats de type isoflavones a mis en lumière une structure dérivée de la daïdzeine : la 6-hydroxydaïdzeine fait partie des structures possibles pour le pic 30N.

4.6 Conclusion de l'étude de l'extrait de réglisse

L'étude LC-HRMS d'un extrait glycériné de racines de réglisse glabre (*Glycyrrhiza glabra* L.) élaboré par notre partenaire avec le but de le valoriser en cosmétique, a été conduite afin de caractériser son contenu moléculaire. Le profil chromatographique de l'extrait est assez riche en UV et répond en spectrométrie de masse à un niveau satisfaisant.

Le premier objectif concernait la caractérisation des six pics chromatographiques (pics **A** à **F**) définis par notre partenaire. Pour cinq pics, une identification a été possible. Les molécules sont des isoflavones (**B**, **C**, **E**, **F**) et une chalcone (**D**). Le pic **A** n'a pas pu être caractérisé puisque seule sa masse monoisotopique a été déterminée, et aucun candidat ne correspond à cette masse dans la base GPDB.

Pour l'ensemble des composés attribués aux cinq pics principaux (**B** à **F**), les données de MS comme de MS/MS nous ont permis de déterminer leur masse monoisotopique et de préciser certains éléments structuraux avec un niveau de confiance élevé. Les propositions d'identification possibles pour les pics principaux de ce profil sont établies grâce au traitement des données issues de LC-HRMS et MS/MS croisées avec les spectres UV et la littérature sur la réglisse glabre ou plus largement sur le genre *Glycyrrhiza*.

Les propositions d'identification à l'issue de l'étude sont les suivantes : le pic **B** correspond à l'ononin ou peut-être à l'isoononin. Le pic **D** est attribué à l'échinatine, un composé qui fait partie de la famille des chalcones. Le pic **E** correspond à la formononétine ou peut-être à l'isoformononétine. Les pics **C** et **F** ont des structures qui ne font pas partie des molécules décrites pour l'organisme étudié. Ces deux structures sont probablement basées sur une hydroxydimethoxy-isoflavone. Pour **C**, la génine est complétée d'un hexose. Pour ces deux pics, la base de données GPDB permet de dresser une liste de candidats de 5 et 12 composés respectivement. Ce nombre restreint de candidats isoflavones est plus aisé à aborder si l'on veut tenter d'identifier le composé de façon plus fine (injection de standards par exemple), par rapport au nombre total de structures possibles (22) si aucun élément structural n'avait été confirmé par l'exploitation des données LC-HRMS/MS.

Le second objectif était la mise en évidence de la famille des isoflavones. Concernant les pics chromatographiques de second ordre, l'étude a permis la détermination de la masse monoisotopique de chaque pic détecté en spectrométrie de masse. De plus, la recherche de la famille phytochimique des isoflavones a été menée. L'interrogation de la base de données de produits naturels de Greenpharma via le critère de la masse monoisotopique a permis de mettre en évidence la présence probable d'isoflavones pour une quinzaine de pics du chromatogramme. L'isoflavone nommée communément daïdzeine est très probablement présente dans l'extrait étudié puisqu'on détecte un pic (48P/28N) dont la masse monoisotopique et la fragmentation en MS/MS correspondent à celles de structures de dihydroxyisoflavones dont seule la daïdzeine est référencée dans l'organisme étudié.

L'étude présente l'intérêt de proposer des résultats multiples que sont l'identification de pics majoritaires, la mise en évidence des pics sans candidats structuraux, l'accès à une vue d'ensemble rapide de la gamme étendue des masses détectées et la détection des composés de la famille des isoflavones plus particulièrement recherchée.

5. Conclusion du chapitre d'applications de GAINS

L'utilisation d'une stratégie telle que celle mise en œuvre avec GAINS permet d'obtenir des informations majeures sur le contenu d'un extrait, sans pour autant remplacer l'identification formelle du contenu moléculaire d'un extrait après l'obtention des composés par purification. Par contre, l'outil permet de ne pas commencer l'étude d'un extrait par le travail d'isolement et d'identification formelle. Ces étapes peuvent être rendues plus simples et efficaces en ciblant certaines molécules. Il est évident que ces étapes peuvent demeurer nécessaires en fonction du contexte et de ce qui doit être démontré. Par ailleurs, la stratégie mise en place avec GAINS représente aussi un moyen alternatif de caractérisation quand les teneurs faibles rendraient la purification d'un ou de composés inefficace ou trop laborieuse.

Afin de rendre l'utilisation de GAINS encore plus pertinente dans la description d'extraits naturels, les travaux en perspectives sont liés à l'amélioration de l'outil en poursuivant l'automatisation de certaines étapes. Il est notamment envisagé de mettre en place le traitement plus automatisé et pertinent des données MS/MS notamment pour exploiter les données de glycosylation des composés. Les perspectives concernent aussi l'intégration des fonctions spécifiques telles que la mise en évidence des molécules non naturelles mais pouvant être contenues dans les plantes (polluants, pesticides).

De plus, l'implémentation de modules de comparaison d'extraits entre eux sur la base du contenu moléculaire total (c'est-à-dire en considérant l'allure du profil chromatographique mais aussi les propositions de masses et donc les candidats associés) pourrait s'avérer utile. L'intérêt de ce type de fonction serait certain pour le suivi dans le temps de la composition phytochimique d'échantillons (étude de stabilité par exemple), pour la réalisation des comparaisons variétales ou la comparaison de modes d'extraction distincts.

Pour l'heure, GAINS est un véritable support pour le phytochimiste qui veut connaître le contenu d'extraits de plantes de façon rapide, globale et pertinente. C'est également en tant qu'outil proposant des pistes de travail plus fines que GAINS se positionne. Le travail du phytochimiste utilisant GAINS reste celui de croiser différentes informations de la littérature avec des données expérimentales. De cette façon, on valide la cohérence de la description qui peut être faite à partir d'outils automatiques et on peut accéder à une description plus exhaustive et approfondie de la composition des extraits à valoriser.

Chapitre 4
Etude de *Lythrum salicaria* L.
Caractérisation et valorisation

Aquarelle d'inflorescences de *Lythrum salicaria*,

Water-color sketches of plants of North America 1888 to 1910 by Helen Sharp

1. Introduction

L'étude de la salicaire commune - *Lythrum salicaria L.*- s'inscrit dans le cadre du projet régional de recherche ECOSMETOCENTRE. Ce projet visait l'étude de la flore locale afin de valoriser des plantes des zones humides naturelles –rivières, étangs, marécages- de la région Centre-Val de Loire. Le choix des plantes est également lié à la disponibilité en tant que ressource en vue d'un développement industriel. Répondant à ces critères, la salicaire commune a donc été sélectionnée. La valorisation s'est inscrite dans le domaine de la dermo-cosmétique avec la volonté de révéler ce que la plante pouvait apporter comme solution en tant qu'actif. Cette étude est la synthèse de ce que peuvent permettre la chimie analytique selon une démarche classique[*] associée à la biologie.

Avec la volonté de suivre une démarche écologique ancrée dans le territoire de la région Centre-Val de Loire, le projet ECOSMETOCENTRE a débuté en janvier 2013 pour une durée de 4 ans (clôture en janvier 2017) et a été labellisé par le pôle de compétitivité DREAM (Durabilité de la Ressource en Eau Associée aux Milieux). ECOSMETOCENTRE a réuni les sociétés Greenpharma (Orléans, 45) en tant que pilote du projet pour la gestion des études phytochimiques, BioAlternatives (Gençay, 86) pour les tests d'activités biologiques et l'Institut des Sciences de la Terre d'Orléans (Orléans, 45) en appui sur certains aspects d'analyses phytochimiques. La société Environnement 41 (Blois, 41) est intervenue en sous-traitance pour l'approvisionnement en plantes.

Ce dernier chapitre va donc présenter le travail effectué autour de la salicaire qui est la plante pour laquelle le travail a été le plus complet et transversal. Il a abouti au développement d'un ingrédient cosmétique commercialisé à l'heure actuelle.

[*] Au moment de ce travail, l'outil GAINS n'avait pas été encore développé.

2. Etudes bibliographiques

2.1 Présentation de la salicaire (*Lythrum salicaria* L.)

2.1.1. Aspect botanique

2.1.1 a) *Taxonomie*

La classification taxonomique de la salicaire est détaillée ci-dessous d'après la plateforme TelaBotanica [135] : Classe des **Dicotylédones ;** Ordre des **Myrtales ;** Famille des **Lythracées ;** Genre ***Lythrum*** **;** Espèce ***salicaria.***

Cette classification est attribuée à Carl von Linné (L.), un naturaliste et botaniste du 18^{e} siècle qui a notamment mis en place le système binominal de dénomination des plantes avec le nom de l'espèce et le nom du genre. Le genre *Lythrum* vient du grec sang, lié à la couleur des fleurs des plantes de ce genre.

2.1.1 b) *Description botanique*

La salicaire est une grande herbe vivace formant des bottes pouvant atteindre un mètre cinquante de hauteur et très souvent à proximité de zones humides. Pendant la floraison qui s'étend de juin à septembre, elle est très reconnaissable avec des fleurs rose violacé regroupées en longs épis terminaux qui forment les inflorescences. Ses feuilles vertes lancéolées sont proches de celles du saule et ses tiges rigides sont dites quadrangulaires [95]. Une aquarelle des parties aériennes en fleurs est présentée en première page de ce chapitre et des photos illustrent ces propos en Figure 70.

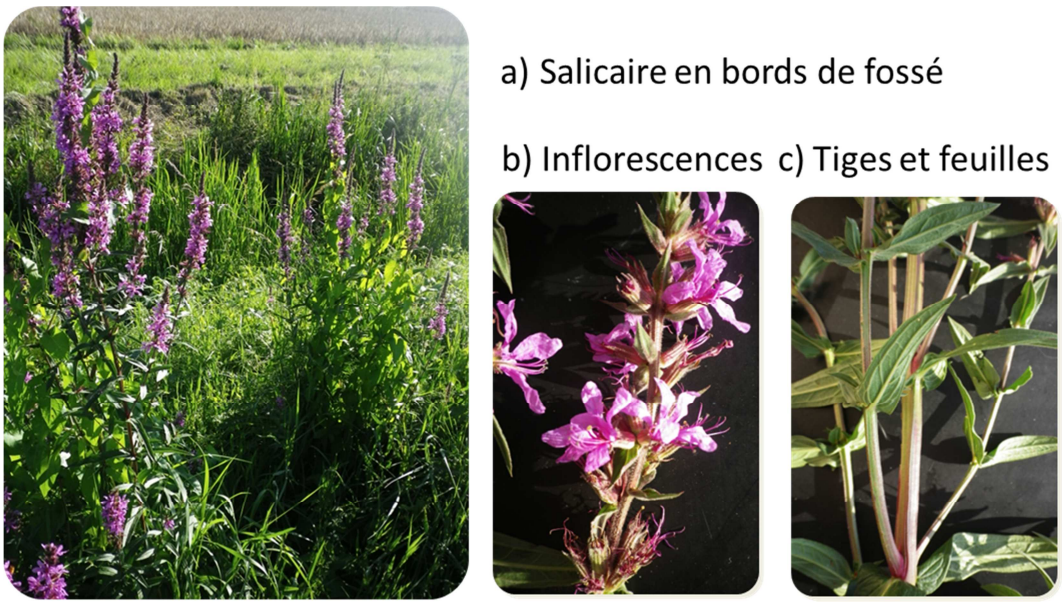


Figure 70 Salicaire a) en milieu naturel –Saint Jean le Blanc dans le Loiret en juillet 2014- et détails d'organes de la plante : b) inflorescence fleurie et c) tige avec feuilles

2.1.1 c) *Répartition géographique*

En France, on observe aisément la salicaire à l'état endémique en bords de rivières, d'étangs ou lacs ou encore le long de fossés car ces zones humides ou marécageuses constituent l'environnement privilégié où la plante s'installe et se développe facilement.

Originaire d'Europe, d'Asie et d'Afrique du Nord [136], la salicaire est actuellement une plante envahissante en Amérique du Nord où elle n'est pas endémique. Elle y a été introduite au début du 19ème siècle par les colons et les navires venant d'Europe [137]. La plante a notamment été utilisée dans les jardins en tant que fleurs d'ornements. En Amérique du nord, le problème réside dans le fait qu'elle n'a pas de prédateur efficace comme en Europe où des insectes limitent sa propagation comme le signalent des organisations telles que le groupe spécialiste des espèces invasives (ISSG) rattaché à l'UICN, le programme ''Ontario's Invading Species Awareness'' ou encore Conservation de la nature Canada [137–139]. L'herbe y déstabilise les écosystèmes et réduit la biodiversité : elle forme d'épais tapis de racines ce qui réduit l'espace pour d'autres espèces endémiques, dégrade les habitats des animaux et perturbe leur alimentation. D'après Lavoie [140], l'Union Internationale de la Conservation de la Nature la classe comme l'une des 100 espèces les plus invasives dans le monde. Des campagnes d'éducation et de prévention ont été mises en place au Canada et aux Etats-Unis dans les années 1990 afin de limiter les ventes pour l'horticulture et des millions de scarabées prédateurs de la plante ont été lâchés dans la nature pour lancer un programme de contrôle biologique. Ce procédé est toujours largement plébiscité et utilisé car il fait ses preuves : les insectes répandus se reproduisent et s'établissent dans les milieux de façon pérenne. Aussi, après quelques années (inférieurs à 5 ans), on peut observer soit un retrait quasi-total de la salicaire soit une vraie limitation de son expansion.

2.1.2. Connaissances phytochimiques

Lorsqu'on parcourt la littérature concernant le contenu moléculaire de la salicaire, on note qu'elle est presque exclusivement dédiée à la description des parties aériennes de la plante. Dans ces organes, ce sont les composés phénoliques et parmi eux les ellagitanins C-glycosidiques qui sont le plus fréquemment et largement décrits [141–146]. Appartenant aux tanins hydrolysables, les ellagitanins sont des polyesters de glucose qui, par hydrolyse acide, forment de l'acide hexahydroxydiphénique qui se lactonise en acide ellagique, d'où leur nom [95]. Les tanins ellagiques C-glycosylés ont des liaisons carbone-carbone entre le carbone C1 du sucre et un carbone d'une des unités galloyl constitutives. De plus, contrairement aux gallotanins, les ellagitanins présentent des liaisons C-C intra-moléculaires entre des unités galloyl voisines [147]. Dans la salicaire, les principaux tanins sont : la vescalagine, la castalagine, la pedunculagine ainsi que des dimères d'ellagitanins

appelés salicarinines et décrites formellement par Piwowarski et Kiss [148]. Les autres composés phénoliques décrits sont des acides phénoliques, des anthocyanes glycosylées (avec la cyanidine-3-O-glucoside et la malvidine 3,5-di-O-glucoside) notamment présentes dans les fleurs violettes de la plante ou encore des flavones et flavonoïdes [143,149–151]. S'agissant des métabolites secondaires, on peut mentionner également que quelques triterpénoïdes (acide oléanolique, acide bétulinique) sont évoqués dans la littérature. Une partie des structures sont représentées dans la Figure 71.

Les études phytochimiques menées sur la salicaire ont combinées plusieurs techniques pour décrire le contenu phytochimique de la plante. Avant isolement des composés, les analyses en RP-HPLC-DAD et UHPLC-CAD [144,145]ou bien en LC-DAD en combinaison avec de la LC-ESI-MS et LC-APCI-MS ont été décrites [142]. Globalement, les composés importants sont isolés et étudiés grâce à des techniques classiques de caractérisation que sont la RMN (proton, carbone et expériences en 2D de COSY ou NOE) [144,152] et la spectrométrie de masse (ESI-Q-TOF ou ESI-IT) [144].

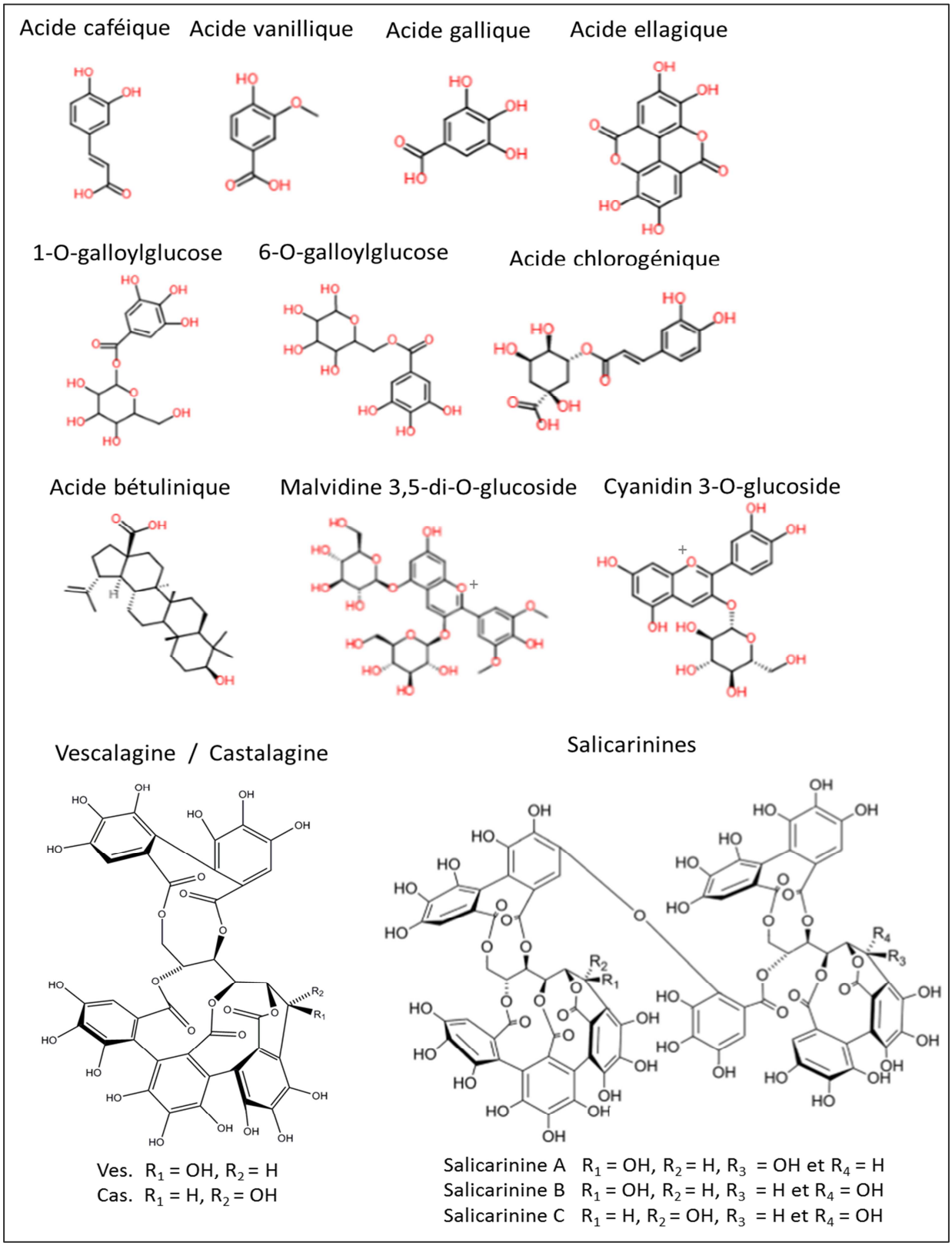


Figure 71 Structures de molécules décrites dans la littérature pour la salicaire

2.1.3. Usages connus et activités biologiques reconnues

Egalement appelée salicaire officinale, la plante est connue depuis l'antiquité et continue d'être utilisée de nos jours, même si l'intérêt lui étant porté a baissé. La salicaire fait partie de la médecine traditionnelle européenne. Elle est inscrite à la pharmacopée européenne. Selon sa

monographie, le contenu total en tanins dans le produit sec doit être de 5 % minimum en équivalent pyrogallol, structure simple de composé phénolique. Les tanins font partie des polyphénols. Ces derniers sont capables de présenter des activités anti-âges du fait de leurs importants effets antioxydants [153]. A travers le temps, l'utilisation principale s'est faite sous la forme d'infusions ou décoctions principalement des parties aériennes, des feuilles ou des fleurs pour leurs propriétés astringentes, hémostatiques et anti-diarrhéiques. Ces activités attribuées à la plante sont largement basées sur les usages traditionnels qui se sont révélés efficaces pour soigner les cas de diarrhées et même de dysenteries. Des études scientifiques apportent la preuve de ces effets à l'image d'une étude *in vivo* chez la souris, menée en 1998 par Brun *et al*. [154]. Ce travail a montré qu'une préparation à base d'extrait de salicaire réduisait les symptômes de diarrhées et se concluait en expliquant qu'un mode d'action probable de l'extrait de plante était lié à une augmentation de l'absorption de l'eau par le côlon. De manière générale, l'activité anti-diarrhéique est expliquée par la forte teneur en tanins dans la plante. C'est la principale activité pour laquelle la plante est connue et pour laquelle des médicaments à base d'extrait de salicaire sont actuellement disponibles à la vente en pharmacie en France [155]. Par ailleurs, une action anticoagulante a été décrite avec une première étude *ex vivo* par Vincent et Segonzac dès 1954 [156] mais en 2011, un papier de Pawlaczyk a décrit des résultats contradictoires sur l'effet de *Lythrum salicaria in vivo* et a indiqué que l'effet anticoagulant ou pro-coagulant de la plante restait à confirmer [157].

Quelques usages externes de la plante ont également été reportés entre 1700 et 1900 pour ses capacités astringentes, cicatrisantes ou vulnéraires [135,158]. Quelques études ethno-pharmacologiques mentionnent des usages externes actuels de la plante : en Italie comme nettoyant sur des peaux malades ou en Iran comme pansement de plaie par exemple [159,160]. Plus récemment, en 2011, une étude *in vitro* de plantes riches en tannins, a mis en évidence l'action anti-hyaluronidase et anti-elastase d'un extrait aqueux de parties aériennes de salicaire [161]. Les auteurs du papier précisaient qu'une interaction non spécifique des tanins avec les protéines était à exclure pour expliquer l'inhibition observée du fait des faibles concentrations auxquelles l'extrait avait montré une forte inhibition (de l'ordre de 10 $\mu g.mL^{-1}$). De plus, des tests de vérification sur le modèle hyaluronidase ont été effectués. Par ailleurs, un brevet[#] a été publié concernant les effets amincissants et de diminution des aspects inesthétiques de la cellulite d'un extrait de salicaire. Le brevet mentionne que l'extrait peut stimuler ou maintenir la production de collagène et activer la lipolyse.

Au vu de cette étude bibliographique, la salicaire commune –en tant que plante indigène en France, répandue et vivace- nous est apparue comme un matériel renouvelable de choix pour initier

[#]CITTADINI L., DANOUX L., GODARD N., Utilisation d'un extrait de *Lythrum salicaria*, WO2016102874, 2016.

une nouvelle stratégie plus moderne afin de valoriser cette espèce locale en cosmétique. En se basant sur les connaissances antérieures, notamment sur la composition phytochimique déjà décrite et les usages traditionnels répertoriés qui suggèrent un potentiel intéressant, l'objectif de notre travail a été d'innover en cherchant une nouvelle utilisation de *Lythrum salicaria*, et plus spécifiquement en cherchant des effets bénéfiques d'extraits de la plante sur la physiologie de la peau.

2.2 La peau et sa physiologie

La peau est un système complexe organisé en strates de cellules épidermiques reposant sur un tissu conjonctif dermique (derme et hypoderme). L'épiderme, dont un schéma est représenté dans la Figure 72, est composé de couches de kératinocytes qui sont constamment renouvelées grâce à la prolifération. Leur spécialisation, essentielle pour que la peau fonctionne parfaitement, est assurée par le processus de différenciation. L'état de différenciation augmente de la couche basale (avec les cellules les moins différenciées) jusqu'au *stratum corneum* (partie la plus externe de l'épiderme dont l'état est le plus différencié) [162,163]. Au sein de chaque couche, des protéines spécifiques sont exprimées telles que des cytokératines (kératines 5, 14, 1, 10, etc.), des marqueurs de différenciation tardive (transglutaminase M1 ou TGK, filaggrine), ou des protéines de l'enveloppe cornée (involucrine, petites protéines riches en proline). Ces protéines sont caractéristiques d'épidermes différenciés. Les kératinocytes migrent et se différencient de façon ultime en cornéocytes qui forment la couche cornée à la surface de la peau. Avant de devenir des cornéocytes, les kératinocytes produisent des protéines et des lipides qui forment une barrière imperméable à l'eau, le *stratum corneum* [164,165]. La peau a besoin d'être stable et en même temps de maintenir une dynamique pour permettre la régénération du tissu lui-même ou répondre aux blessures cutanées [162,163,166].

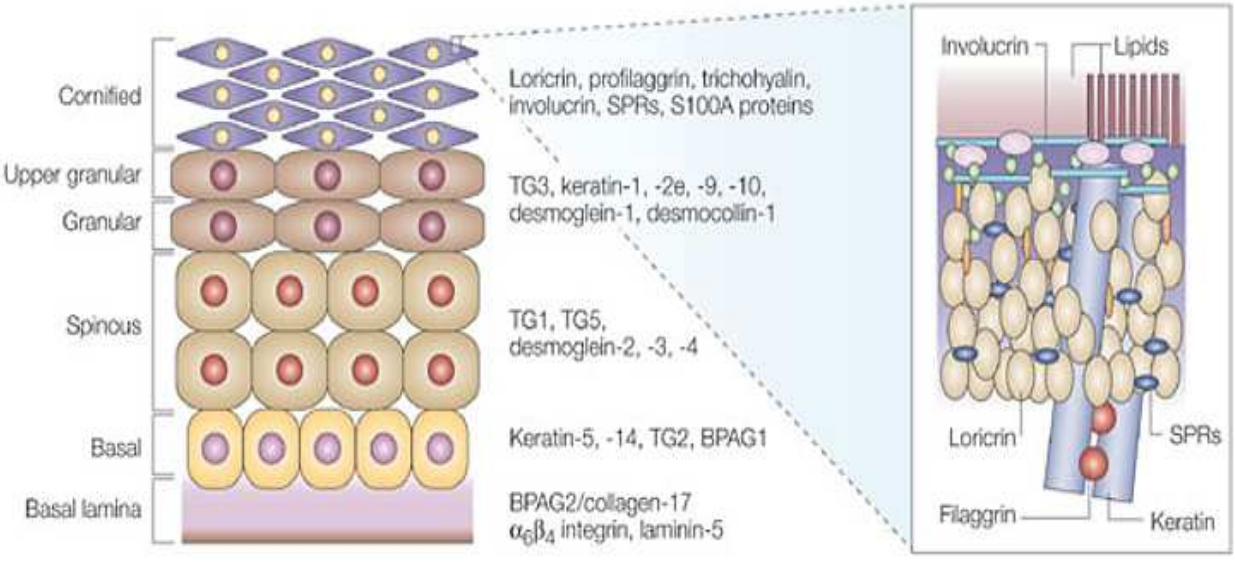


Figure 72 Schéma descriptif de l'organisation de l'épiderme de la peau humaine : les couches cellulaires sont représentées à gauche (traduction du bas vers le haut : lame basale, couche basale, couche spineuse, couche granulaire, couche granulaire supérieure et couche cornée). Au centre, sont indiquées les protéines exprimées dans chaque couche cellulaire : TG, transglutaminase ; SPRs, Small

Proline-Rich proteins. A droite, la composition moléculaire de la couche cornée est détaillée (traduction : involucrine, loricrine, fillagrine, kératines, SPRs : petites protéines riches en proline, et lipides). Illustration tirée de ''The cornified envelope: a model of cell death in the skin'', Candi et al., 2005, Nature Reviews Molecular Cell Biology, 6, 328-340.

Une des fonctions essentielles de la peau est de constituer une barrière protectrice pour éviter la perte des fluides internes et prévenir l'entrée d'agents externes nocifs. Quotidiennement, la peau doit faire face à des perturbations externes telles que les radiations ultraviolettes provenant du soleil, la pollution de l'air, ou les blessures ou irritations. A la fois barrière protectrice mais organe fragile (avec une épaisseur de quelques millimètres maximum), la peau a la faculté de s'auto-renouveler constamment. C'est cette régénération qui permet la réparation des dommages subis et le remplacement des vieilles cellules. Néanmoins, la peau ne peut pas échapper au vieillissement. Celui-ci induit des changements considérables dans la structure, le fonctionnement et l'apparence de la peau à l'image de la baisse de l'épaisseur de l'épiderme, de la dégradation de la matrice extracellulaire, de la perte de la capacité au renouvellement cellulaire ou de cicatrisation, du développement de l'élastose ou encore de la formation de rides [153,167,168].

La fonction protectrice de la peau est essentiellement assurée par le *stratum corneum* [165,169]. Une façon de maintenir l'intégrité de l'épiderme et plus généralement de prévenir ou limiter l'inévitable phénomène de vieillissement de la peau est d'utiliser des traitements topiques à base d'ingrédients actifs naturels. Le travail présenté ci-après s'inscrit dans cette démarche. Nous avons revisité *Lythrum salicaria* L. en investiguant à nouveau son contenu métabolique et ses activités biologiques envers la peau afin de mettre en lumière l'intérêt potentiel de ce végétal pour le domaine dermo-cosmétique.

3. Première approche exploratoire sur les potentialités dermo-cosmétiques de la salicaire

Pour introduire l'étude phytochimique et les tests d'activités biologiques qui vont être détaillés dans la suite de ce chapitre, les travaux antérieurs au doctorat qui ont constitués le début du projet sur la salicaire au sein de Greenpharma sont résumés ci-dessous.

Une série d'extraits hydro-alcooliques avec différentes proportions d'éthanol/eau a été réalisée par macération sur des parties aériennes de *Lythrum salicaria* L. et envoyée en tests d'activités biologiques. La série d'extraits a été testée pour évaluer leur potentiel à stimuler l'expression du marqueur de différenciation transglutaminase (TGK) dans des cultures de kératinocytes humains normaux (NHEK) avec un suivi en immunofluorescence (voir protocole en annexe 6.2.3). Les résultats de ces premiers travaux sont visibles en Figure 73 :

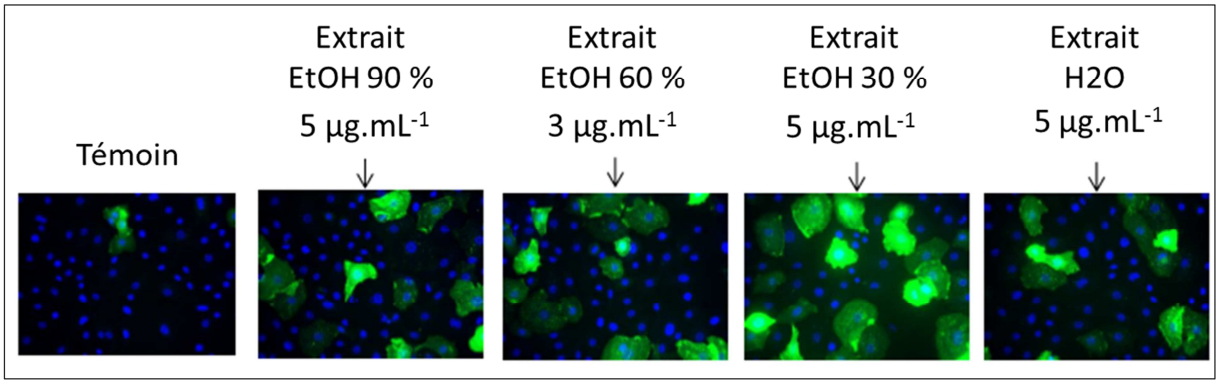


Figure 73 Analyses en immunofluorescence de l'expression de TGK dans les kératinocytes épidermiques à l'état basal ou traités par des extraits de *Lythrum salicaria*. Après 72 heures de traitement, les cellules sont fixées et la protéine d'intérêt est révélée en utilisant un anticorps dirigé contre elle. Ces résultats sont issus d'une expérience représentative de trois.

Dans ce test, les tâches fluorescentes vertes sont liées à la présence de TGK. L'expression de TGK est visible dans les kératinocytes ayant été traités par les différents extraits végétaux. Les quatre extraits sont donc actifs : ils stimulent la production de TGK par rapport au témoin (kératinocytes non traités). L'extrait EtOH 30 % s'est révélé être le plus actif des quatre extraits testés et de manière nette. Cet extrait a ensuite été appliqué sur les cellules avec une gamme de concentration de 1,6 à 20 µg.mL^{-1} choisies autour de la première concentration testée et toutes inférieures à la concentration cytotoxique de l'extrait. Ceci a permis de mettre en évidence un clair effet-dose de son effet de stimulation de la différenciation épidermique comme le montre la Figure 74 : plus la concentration en extrait pour traiter les cellules est élevée, plus la coloration verte est importante et donc plus la protéine TGK est exprimée.

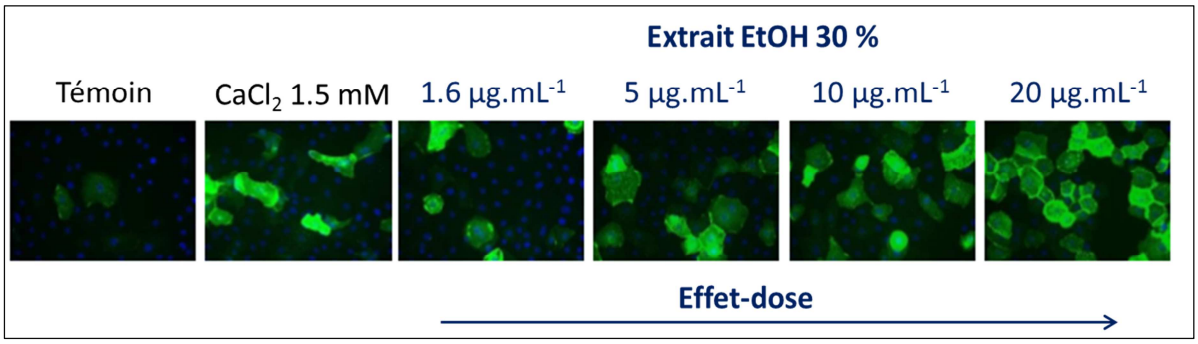


Figure 74 Analyses en immunofluorescence de l'expression de TGK dans les kératinocytes épidermiques à l'état basal ou en présence de CaCl2 à 1,5mM ou traité par l'extrait EtOH 30% de *Lythrum salicaria* dans une gamme de concentration croissante. Après 72 heures de traitement, les cellules sont fixées et la protéine d'intérêt est révélée en utilisant un anticorps dirigé contre elle. Ces résultats sont issus d'une expérience représentative de trois.

Ces premiers résultats ont donc orienté la poursuite de l'étude vers cet extrait. Dans la suite du chapitre, les éléments détaillés concernent donc exclusivement l'extrait hydro-alcoolique à 30 % d'éthanol.

4. Etude phytochimique de *Lythrum salicaria* L.

4.1 Matériel végétal et extraction

Le matériel végétal étudié se composait des parties aériennes séchées de *Lythrum salicaria* L. récoltées en cours de floraison pendant l'été 2012 dans le département du Loiret en France. Un spécimen est conservé sous la forme d'un herbier au sein de Greenpharma SAS.

La plante a été broyée puis extraite de façon simple et classique par macération dans un mélange éthanol-eau en proportion 30/70 (v/v) pendant 24 heures sous agitation magnétique et à l'abri de la lumière. Le ratio plante-solvant était de 1 pour 10 (m/v). Après 24 heures, l'extrait a été successivement filtré sur des filtres de porosité décroissante jusqu'à une filtration sur un filtre en fibre de verre de porosité de 1 µm. Ensuite, le filtrat a été traité sur charbon actif pendant 30 minutes sous agitation magnétique. Le charbon adsorbe les composés aromatiques et les pigments : le traitement permet de les éliminer partiellement et ainsi de décolorer l'extrait mais aussi de le simplifier. En fin de procédé, l'extrait sec a été obtenu après une ultime filtration sur un filtre en fibre de verre de porosité de 1 µm puis par l'évaporation de l'éthanol à l'aide d'un évaporateur rotatif à pression réduite et enfin une lyophilisation de l'eau.

L'extrait hydro-éthanolique obtenu par macération sera mentionné de la manière suivante dans la suite du texte : extrait de *Lythrum salicaria.*

4.2 Evaluation rapide des teneurs en composés polaires

4.2.1. Teneur en sucres neutres totaux

Faisant partie des métabolites primaires, les sucres sont des composés de grande polarité dont nous avons voulu évaluer la teneur du fait de la polarité de l'extrait étudié. La quantité de sucres neutres totaux (ou monosaccharides totaux) de l'extrait a été déterminée avec une méthode colorimétrique au résorcinol dans l'acide sulfurique adaptée des travaux de Monsigny [170] : les conditions sont en annexe 4.1.1. Le réactif résorcinol réagit avec les sucres [171] et forme un produit coloré orange. Le milieu acide à chaud de la réaction permet de libérer les sucres qui pourraient être liées à des aglycones à l'état naturel, c'est pourquoi cette méthode permet le dosage des sucres totaux. La méthode implique une mesure d'absorbance à 430 nm. La teneur en sucres neutres totaux est déterminée à 18,4 % en équivalent glucose dans l'extrait sec.

4.2.2. Contenu en composés phénoliques

Largement décrits pour la salicaire, les composés phénoliques totaux présents dans l'extrait ont été quantifiés par un dosage colorimétrique utilisant le réactif de Folin-Denis adaptée d'une méthode de Kitzberger [172] et une mesure d'absorbance à 635 nm (conditions analytiques en annexe 4.1.2). La réaction colorimétrique met en jeu une oxydation des composés phénoliques et génère une coloration bleu-vert grâce à la présence de tungstate de sodium et d'acide phospho-molybdique contenus dans le réactif de Folin-Denis. La gamme d'étalonnage a été faite en préparant des solutions d'acide chlorogénique, utilisé comme standard. L'extrait a été solubilisé à 5 $mg.mL^{-1}$ dans l'eau. La réaction colorimétrique est permise par l'ajout du réactif de Folin-Denis avec une solution saturée de carbonate de sodium à l'échantillon. Après 30 minutes d'agitation, la mesure de l'absorbance s'effectue à 635 nm.

Dans l'extrait sec, la teneur en composés phénoliques totaux est de 48,5 % en équivalent acide chlorogénique utilisé comme étalon témoin. Les composés phénoliques représentent donc une part importante de cet extrait, ceci est cohérent avec la littérature sur la plante d'une part et la polarité du mélange de macération (eau-éthanol à 70/30 v:v) d'autre part.

4.3 Caractérisation de l'extrait par HPLC-DAD-MS

Pour aller vers une analyse plus fine de l'extrait, il a été analysé en chromatographie en phase liquide couplée à la spectrométrie de masse. La Chromatographie Liquide de Haute Performance (HPLC) est couplée à un détecteur à barrettes de diodes (DAD) et une trappe d'ion dotée d'une source électrospray (ESI). Le chromatogramme UV de l'analyse HPLC-DAD-MS de l'extrait de salicaire est présenté en Figure 75 :

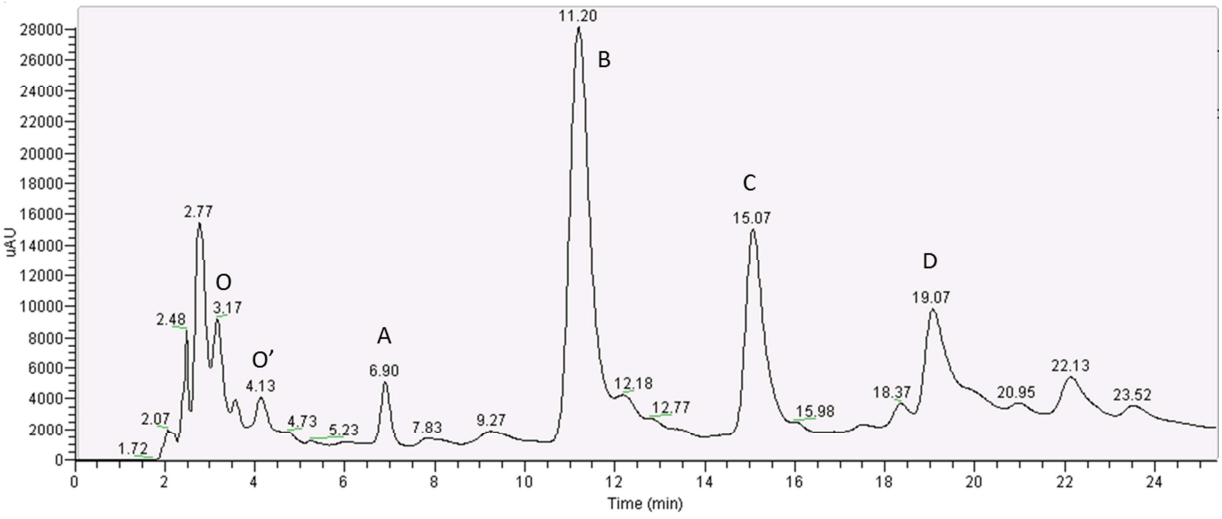


Figure 75 Chromatogramme de l'extrait de salicaire, colonne POLAR-RP (150x4.6 mm, 4µm) – Détection UV 254 nm-

On a suivi l'analyse à la longueur d'onde de 254 nm, utilisée de façon standard qui permet de détecter tout composé possédant des insaturations et des cycles aromatiques. En sortie de DAD, l'éluant est directement envoyé dans l'interface ESI sans division du flux et les composés sont analysés en mode d'ionisation négatif et positif (conditions opératoires en Annexe 3.8).

Pour étudier l'extrait de *Lythrum salicaria* plutôt polaire, la source électrospray était appropriée. L'identification des composés a été basée sur le croisement des données UV, des spectres de MS en modes positif et négatif dont les informations se complètent ainsi que grâce à la comparaison avec les données de la littérature. Il a fallu exploiter les spectres de masses pour faire correspondre des ions entre eux et ainsi proposer des masses. A partir de la base de données Greenpharma, nous avons généré la liste des composés affiliés à l'organisme, ceci suite à des recherches bibliographiques. La correspondance des masses des molécules nous a permis de rapidement déceler si des masses issues des analyses HPLC-MS correspondaient à celles attendues théoriquement. Le Tableau 26 rassemble les données LC-DAD-MS obtenu pour les pics **O** à **D**.

Pic	Temps de rétention (min)	Maximum d'absorbance (nm)	Ions moléculaires détectés	Autres ions ou adduits détectés	Masse molaire ($g.mol^{-1}$)	Proposition d'identification
O / O'	3,17 / 4,13	236 / 235	$[M-H]^-$; $[M+H]^+$	$[M+NH_4]^+$	632,44	Castaline ou Vescaline
A	6,90	269	$[M-H]^-$	$[M+HCOO]^-$; $[2M-H]^-$	170,12	Acide Gallique
B	11,20	235	$[M+H]^+$; $[M-H]^-$	$[M+Na]^+$; $[M-2H]^{2-}$	934,63	Castalagine ou Vescalagine
C	15,07	235	$[M+H]^+$; $[M-H]^-$	$[M+Na]^+$; $[M-2H]^{2-}$; $[M+HCOO]^-$	934,63	Castalagine ou Vescalagine
D	19,07	233	ND	$[M-2H]^{2-}$	1868	Une salicarinine

Tableau 26 Données LC-ESI-MS collectées par analyse de l'extrait de *Lythrum salicaria* en mode d'ionisation positif et en mode d'ionisation négatif

Une étape de caractérisation globale de l'extrait hydro-alcoolique de *Lythrum salicaria* a été menée pour collecter de premières informations sur les composés en présence. Les masses moléculaires des pics principaux sont obtenues par couplage RP-HPLC avec la spectrométrie de masse en électrospray (ESI-MS).

S'agissant du pic **A**, une masse moléculaire de 170 $g.mol^{-1}$ est retrouvée grâce à trois ions de *m/z* 169 $[M-H]^-$; 214 $[M+HCOO]^-$ et 338 $[2M-H]^-$. Le maximum d'absorbance du spectre UV est observé à 270 nm. Au vue de ces données, la littérature sur la salicaire commune suggère l'acide gallique [142]. L'injection du standard de cet acide phénolique et de l'extrait dans les mêmes conditions en HPLC nous a permis de confirmer que le composé **A** est bien l'acide gallique. En tant

qu'unité constitutive des tanins eux-mêmes décrits dans la plante, sa présence dans l'extrait n'est pas surprenante.

Pour les deux composés principaux (**B** et **C**) de l'extrait, les ions observés à *m/z* : 935, $[M+H]^+$; 933, $[M-H]^-$; 956, $[M+Na]^+$ correspondent à une masse moléculaire de 934 $g.mol^{-1}$. Les ions et les spectres UV de deux composés décrits dans la littérature pour la salicaire ont été observés : ceux de la castalagine et de la vescalagine, deux isomères d'ellagitanins C-glycosidiques [142,148]. Avec les seules données de masse, la distinction entre les deux molécules est impossible. Sans être des preuves absolues, l'ordre d'élution et la différence de polarité permettent de faire l'hypothèse que le premier des isomères élués (Pic **B**) pourrait être la vescalagine car elle est décrite comme plus polaire que la castalagine.

De plus, des ions correspondants à une masse de 632 $g.mol^{-1}$ sont détectés pour les pics **O** et **O'**. Lorsque l'extraction d'ion de *m/z* 631 est effectuée sur une analyse en mode négatif, deux pics ressortent sur le chromatogramme aux temps de rétention des pics **O** et **O'** qui nous a permis de préciser la position des pics au sein du massif de la zone de faible rétention du chromatogramme. Cela pourrait correspondre à la vescaline et la castaline (structures en Figure 76), deux isomères de la famille des tanins ayant ce poids moléculaire.

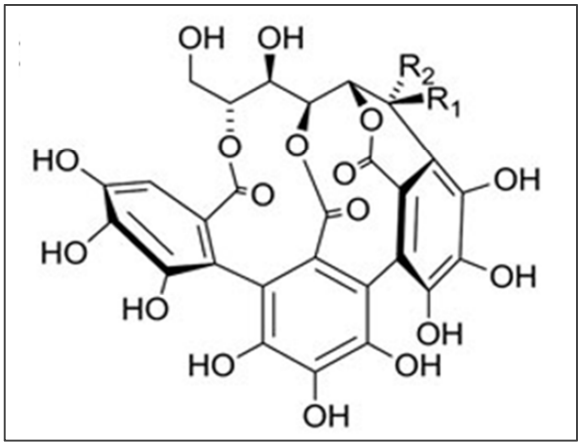


Figure 76 Structures de la vescaline (R_1 = OH, R_2 = H) et de la castaline (R_1 = H, R_2 = OH)

La vescaline et la castaline ne sont pas décrites dans la littérature sur la salicaire. En revanche, ils sont des produits de l'hydrolyse acide de la vescalagine et de la castalagine [144,147]. Leur présence dans la plante est donc probable. Le maximum UV déterminé ne va pas à l'encontre de cette hypothèse puisque les valeurs sont équivalentes à celles de la vescalagine et castalagine. Toutefois, ces composés sont présents en quantités faibles dans la plante d'après le profil HPLC et une caractérisation plus fine et plus poussée n'a pu être menée car les composés n'ont pas été isolés.

Enfin, pour le pic **D**, on détecte un ion dichargé de *m/z* 933 sans aucun autre ion pour nous aider à l'expliquer. Dans un premier temps, on a fait l'hypothèse que l'ion correspondait à l'ion

$[M-2H]^{2-}$. Dans ce cas-là, la masse moléculaire du composé est supposée être égale à 1868 $g.mol^{-1}$. Or cette masse correspond à la masse de trois molécules décrites pour la salicaire. La masse tend à indiquer la présence d'une salicarinine, dimère d'ellagitanins. D'après un article de Piwowarski et Kiss [144] sur le contenu en ellagitanins de la salicaire, trois dimères de masses molaires égales sont décrits : les salicarinines A, B et C qui sont respectivement constituées de vescalagine et strachyurine, vescalagine et casuarinine, castalagine et casuarinine. Au regard des données collectées, le composé du pic **D** devrait correspondre à l'une de ces structures.

4.4 Détermination structurale des composés majoritaires

4.4.1. Isolement par HPLC semi-préparative

Suite aux hypothèses de structures effectuées pour certains composés de l'extrait et afin d'aller plus loin dans leur description, l'isolement des composés majoritaires a été conduit en HPLC semi-préparative via l'injection en solution de l'extrait de *Lythrum salicaria* L. prétraitée par une étape de SPE qui a permis d'éliminer une partie des composés de l'extrait les moins polaires. L'optimisation de cette préparation d'échantillon sur cartouche SPE-C18 de petite échelle et la réalisation sur une plus grande cartouche à partir de 800 mg d'extrait sont détaillées en annexe 4.2.2.

L'échantillon prétraité par SPE et concentré à environ 61.5 $mg.mL^{-1}$ est donc injecté en HPLC semi-préparative. Pour cette échelle, la mise en œuvre est limitée par les colonnes à disposition, les dimensions et la phase stationnaire de la colonne sont différentes de l'échelle analytique : le débit est ainsi augmenté et le gradient est adapté pour maintenir une séparation correcte (annexe 4.4.2).

Le suivi du profil chromatographique s'est fait à 345 nm. Le choix de cette longueur d'onde résulte de son ajustement pour permettre une collecte efficace des composés. En effet, un ajustement a été nécessaire du fait que la solution d'extrait injectée était très concentrée : à la longueur d'onde plus classique de 254 nm, le signal était totalement saturé.

Le volume d'injection de 800 µL a été choisi après différents essais de volumes d'injection croissants de 50 µL à 800 µL. Le but étant d'avoir le volume d'injection le plus grand possible pour limiter le nombre d'injections et donc de collectes différentes tout en conservant une séparation permettant l'isolement des composés les plus intenses. Pour les pics **B**, **C** et **D**, chaque pic chromatographique a été collecté séparément et à la fin de l'ensemble des injections, chaque fraction a été séchée par évaporation sous pression réduite et lyophilisation. Le Tableau 27 résume les résultats de ces opérations d'isolement.

Composé	Quantité obtenue	Pureté (UV 254nm)	Rendement de purification
B	36,8 mg	99,6 %	4,5 %
C	18,6 mg	99,2 %	2,3 %
D	5,69 mg	94,0 %	0,7 %

Tableau 27 Isolement des composés majoritaires de l'extrait de *Lythrum salicaria* : quantités obtenues, pureté et rendement de purification de chaque composé à partir de l'extrait

4.4.2. Détermination structurale du composé **D**

Le composé **D** est détecté en spectrométrie de masse avec une masse moléculaire de 1868 $g.mol^{-1}$ qui indique une des trois salicarinines décrites pour la salicaire. On note un maximum d'absorbance de 233 nm, cohérent avec la description proposée. De plus, après l'avoir physiquement isolé, une analyse en HRMS en FIA a été réalisée afin de déterminer la masse monoisotopique et la formule brute du composé. En mode electrospray et ionisation négative sur une plage de 50 à 3000 uma, les ions détectés sont listés et annotés dans le Tableau 28 :

m/z	z	Formule brute	Attribution	Masse monoisotopique expérimentale ($g.mol^{-1}$)	Erreur entre masse expérimentale et théorique
933,0651	2-	$C_{82}H_{50}O_{52}$	$[M-2H]^{2-}$	1868,1448	1,2 ppm
944,0561	2-	$C_{82}H_{49}O_{52}Na$	$[M+ Na- 3H]^{2-}$	1868,1443	1,0 ppm

Tableau 28 : Résultats HRMS pour le pic D

Ainsi, on accède à la formule brute de $C_{82}H_{52}O_{52}$ de masse molaire monoisotopique égale à 1868,1425 $g.mol^{-1}$ pour le pic **D** et donc la très probable présence d'une des 3 salicarinines décrites. Les données de LC-MS ne permettent pas de définir quelle est la salicarinine en présence. Un travail en MS/MS pour fragmenter les ions n'aiderait pas facilement à déterminer le type de la salicarinine en présence car castalagine avec vescalagine et casuarinine avec strachyurine sont isomères deux à deux. Les investigations pour déterminer le type de salicarinine en présence n'ont pas été menées du fait de la trop faible quantité de produit isolé pour faire les expériences nécessaires, notamment à travers d'analyses RMN, à l'obtention de tous les éléments pour distinguer l'un ou l'autre des dimères.

4.4.3. Détermination structurale des composés majoritaires (**B** et **C**)

D'après les analyses en LC-DAD-ESI-MS, une hypothèse de structure a pu être formulée pour les pics **B** et **C**. En effet, on a pu déterminer leur maximum UV à 235 nm et leur masse molaire à 934 $g.mol^{-1}$ grâce à la correspondance de plusieurs ions dans les deux modes d'ionisation (voir paragraphe 4.3). Ainsi, **B** et **C** pourraient être de la vescalagine et de la castalagine. Les structures de ces derniers sont présentées en Figure 77.

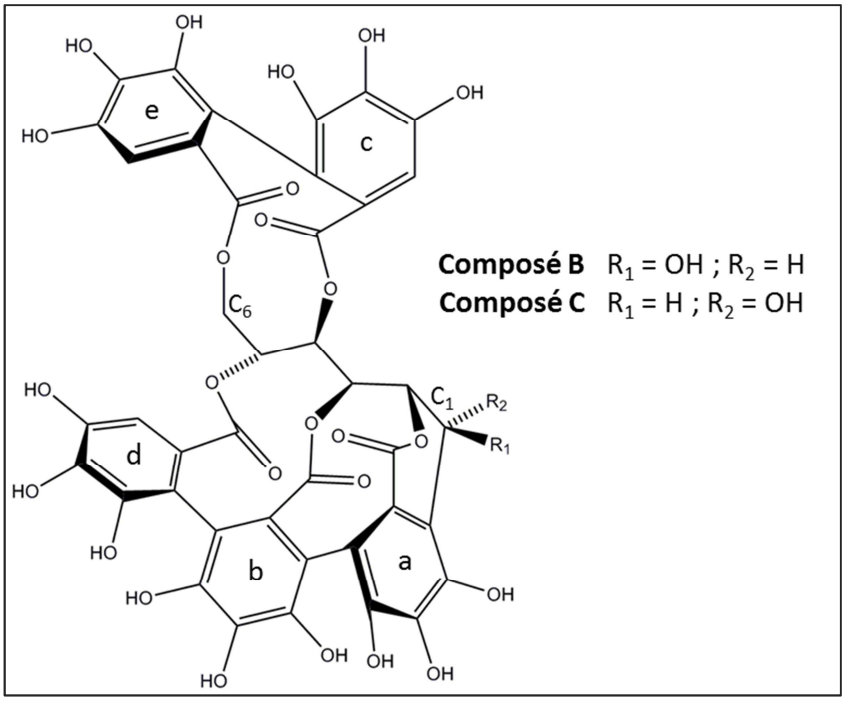


Figure 77 Structures de la vescalagine et de la castalagine

Les composés **B** et **C** ont été analysés en ESI-HRMS et MS/MS en FIA. Les composés **B** et **C** s'ionisent convenablement dans les deux modes de polarités. Les données HRMS confirment leur formule brute : $C_{41}H_{26}O_{26}$. Les deux composés sont détectés sous plusieurs formes : ions pseudo-moléculaires à la fois monochargés et dichargés ou encore sous forme d'adduit ammonium en positif. Les ions pseudo-moléculaires sous formes monochargées ont été isolés pour observer la fragmentation en MS/MS de chacune des molécules. Les expériences en MS/MS ont été gérées manuellement afin d'optimiser l'énergie de collision pour obtenir un spectre de fragmentation équilibré et informatif. Les énergies de fragmentation optimales sont de 25 eV en positif et 55 eV en négatif. Les données ont été traitées avec le logiciel *DataAnalysis* et les formules brutes de l'ensemble des ions observés ont été générées avec l'algorithme *SmartFormula* (*cf.* conditions analytiques en Annexe 3.9). Les résultats sont retranscrits dans le Tableau 29.

L'attribution d'une partie des ions a été possible grâce à la formule obtenue pour la valeur du rapport *m/z* correspondant et par comparaison avec la publication de Piwowarski and Kiss [144] qui ont publié des données de fragmentation en spectrométrie de masse basse résolution.

Les fragments observés en MS/MS étayent la proposition de tanins pour **B** et **C** puisque les pertes de neutres sont cohérentes avec les structures (perte de H_2O et de CO_2 liées respectivement à la présence de groupements hydroxy et ester). De plus, les fragments typiques d'ellagitanins sont retrouvés avec la libération de l'unité HHDP détectée sous forme protonée en positif (*m/z* 303,014) ou déprotonée en négatif (*m/z* 300,999). Lorsque cette unité a été libérée mais que la charge est conservée par le reste de la molécule, on détecte les ions $[M\text{-}HHDP+H]^+$ (*m/z* 633,072) et $[M\text{-}HHDP\text{-}H]^-$ (*m/z* 631,057) sauf pour le composé **C** où ce dernier ion n'est pas présent.

Composé	*m/z* mode positif	Formule brute	Erreur (ppm)	Données MS/MS : *m/z*, intensité relative, attribution	Formule brute	*m/z* mode négatif	Formule brute	Erreur (ppm)	Données MS/MS : *m/z*, intensité relative, attribution	Formule brute
B	**935,0783** (100) $[M+H]^+$ 952,1050 (32) $[M+NH_4]^+$	$C_{41}H_{27}O_{26}$ $C_{41}H_{30}O_{26}N$	0.3 0.1	935,0784 (71) $[M+H]^+$ 917,0671 (18) $[M-H_2O+H]^+$ 873,0776 (9) $[M-CO_2-H_2O+H]^+$ 633,0719 (100) $[M-HHDP+H]^+$ 615,0616 (75) $[M-HHDP-H_2O+H]^+$ 597,0511 (24) $[M-HHDP-2H_2O+H]^+$ 553,0608 (33) $[M-HHDP-2H_2O+-CO_2+H]^+$ 535,0506 (66) $[M-HHDP-3H_2O+-CO_2+H]^+$ 523,0507 (17) 517,0400 (23) $[M-HHDP-4H_2O+-CO_2+H]^+$ 505,0400 (50) 495,0193 (43) 469,0399 (42) 303,0137 (14) $[HHDP+H]^+$ 277,0343 (16)	$C_{41}H_{27}O_{26}$ $C_{41}H_{25}O_{25}$ $C_{40}H_{25}O_{23}$ $C_{27}H_{21}O_{18}$ $C_{27}H_{19}O_{17}$ $C_{27}H_{17}O_{16}$ $C_{26}H_{17}O_{14}$ $C_{26}H_{15}O_{13}$ $C_{25}H_{15}O_{13}$ $C_{26}H_{13}O_{12}$ $C_{25}H_{13}O_{12}$ $C_{23}H_{11}O_{13}$ $C_{22}H_{13}O_{12}$ $C_{14}H_7O_8$ $C_{13}H_9O_7$	**933,0630** (47) $[M-H]^-$ 477,0189 (10) $[M-3H+Na]^{2-}$ 466,0282 (100) $[M-2H]^{2-}$ 300,9989 (8) $[HHDP-H]^-$	$C_{41}H_{25}O_{26}$ $C_{41}H_{23}NaO_{26}$ $C_{41}H_{24}O_{26}$ $C_{14}H_5O_8$	1.0 0.9 0.4 0.3	933,0635 (57) $[M-H]^-$ 915,0521 (27) $[M-H_2O-H]^-$ 897,0449 (5) $[M-2H_2O-H]^-$ 889,0733 (9) $[M-CO_2-H]^-$ 871,0653 (13) $[M-CO_2-H_2O-H]^-$ 613,0471 (16) $[M-HHDP-H_2O-H]^-$ 569,0574 (34) $[M-HHDP-CO2_H_2O-H]^-$ 493,0052 (92) 489,0463 (32) 467,0254 (89) 425,0150 (26) 300,9992 (100) $[HHDP-H]^-$ 275,0192 (40) 249,0404 (54)	$C_{41}H_{25}O_{26}$ $C_{41}H_{23}O_{25}$ $C_{41}H_{21}O_{24}$ $C_{40}H_{25}O_{24}$ $C_{40}H_{23}O_{23}$ $C_{27}H_{17}O_{17}$ $C_{26}H_{17}O_{15}$ $C_{23}H_9O_{13}$ $C_{25}H_{13}O_{11}$ $C_{22}H_{11}O_{12}$ $C_{20}H_9O_{11}$ $C_{14}H_5O_8$ $C_{13}H_7O_7$ $C_{12}H_9O_6$
C	**935,0783** (100) $[M+H]^+$ 952,1049 (40) $[M+NH_4]^+$	$C_{41}H_{27}O_{26}$ $C_{41}H_{30}O_{26}N$	0.3 0.2	935,0781 (34) $[M+H]^+$ 917,0677 (7) $[M-H_2O+H]^+$ 873,0781 (10) $[M-CO_2-H_2O+H]^+$ 633,0721 (68) $[M-HHDP+H]^+$ 615,0615 (40) $[M-HHDP-H_2O+H]^+$ 597,0511 (20) $[M-HHDP-2H_2O+H]^+$ 553,0612 (26) $[M-HHDP-2H_2O+-CO_2+H]^+$ 535,0506 (34) $[M-HHDP-3H_2O+-CO_2+H]^+$ 523,0503 (12) 517,0402 (10) $[M-HHDP-4H_2O+-CO_2+H]^+$ 505,0400 (16) 495,0195 (81) 469,0402 (100) 439,0295 (14) 303,0136 (8) $[HHDP+H]^+$ 277,0343 (10)	$C_{41}H_{27}O_{26}$ $C_{41}H_{25}O_{25}$ $C_{40}H_{25}O_{23}$ $C_{27}H_{21}O_{18}$ $C_{27}H_{19}O_{17}$ $C_{27}H_{17}O_{16}$ $C_{26}H_{17}O_{14}$ $C_{26}H_{15}O_{13}$ $C_{25}H_{15}O_{13}$ $C_{26}H_{13}O_{12}$ $C_{25}H_{13}O_{12}$ $C_{23}H_{11}O_{13}$ $C_{22}H_{13}O_{12}$ $C_{21}H_{11}O_{11}$ $C_{14}H_7O_8$ $C_{13}H_9O_7$	**933,0633** (27) $[M-H]^-$ 477,0192 (18) $[M-3H+Na]^{2-}$ 466,0282 (100) $[M-2H]^{2-}$	$C_{41}H_{25}O_{26}$ $C_{41}H_{23}NaO_{26}$ $C_{41}H_{24}O_{26}$	0.6 0.2 0.3	933,0634 (100) $[M-H]^-$ 915,0514 (17) $[M-H_2O-H]^-$ 889,0757 (12) $[M-CO_2-H]^-$ 871,0619 (9) $[M-CO_2-H_2O-H]^-$ 631,0574 (64) $[M-HHDP-H]^-$ 569,0576 (41) $[M-HHDP-CO2_H_2O-H]^-$ 467,0257 (32) 425,0150 (74) 300,9990 (99) $[HHDP-H]^-$ 275,0199 (33) 249,0407 (28)	$C_{41}H_{25}O_{26}$ $C_{41}H_{23}O_{25}$ $C_{40}H_{25}O_{24}$ $C_{40}H_{23}O_{23}$ $C_{27}H_{19}O_{18}$ $C_{26}H_{17}O_{15}$ $C_{22}H_{11}O_{12}$ $C_{20}H_9O_{11}$ $C_{14}H_5O_8$ $C_{13}H_7O_7$ $C_{12}H_9O_6$

Tableau 29 Données HRMS et MS/MS pour les composés **B** et **C** obtenues par analyses en ESI-Q-TOF (HHDP : unité hexahydroxydiphenoyl = 302.0057 g.mol-1). Pour chaque ion : les intensités relatives apparaissent entre parenthèses, l'attribution entre crochets. Les ions isolés pour la MS/MS ont un m/z en gras. Les ions de MS/MS sans attribution (entre crochets) étaient tous monochargés.

Les deux tanins, proposés pour caractériser **B** et **C**, sont des épimères, c'est-à-dire qu'ils sont diastéréoisomères en ne différant l'un de l'autre que par la configuration d'un seul carbone asymétrique. En tant qu'isomères, la spectrométrie de masse ne permet pas différencier les composés **B** et **C** à l'inverse de la RMN. En effet, la littérature indique une différence entre la vescalagine et la castalagine au niveau du déplacement chimique et de la constante de couplage du proton anomérique (proton en C_1 du glucose de la structure). Les conditions des analyses en RMN sont en annexe 4.6.2. Les signaux obtenus et leur assignement en ^{1}H-RMN et ^{13}C-RMN sont répertoriées dans le Tableau 30.

Pour les protons H_1 de chacune des molécules, on remarque des différences notables : pour le composé **B**, le déplacement chimique (4,87 ppm) et la constante de couplage (J= 2,00 Hz) sont plus faibles que pour le composé **C** (5,71 ppm et J = 4,40 Hz). D'après l'article de Nonaka [152] sur la description de tanins C-glycosylés, une faible constante de couplage (entre 0 et 2 Hz) du proton anomérique indique une orientation en beta du groupement hydroxyl sur le carbone C_1 de la partie sucre tandis qu'une constante de couplage proche de 5 Hz indique une orientation alpha de ce substituant sur la structure.

	Position	B		C	
		δ_C	δ_H	δ_C	δ_H
Glycosyl	1	65,4	4,87 (d ; 2,00)	65,6	5,71 (d, 4,40)
	2	77,6	5,19-5,23 (m ; na)	73,4	5,03-5,05 (m ; na)
	3	68,3	4,56 (dd ; 0,80/6,80)	66,4	5,03-5,05 (m ; na)
	4	69,2	5,19-5,23 (m ; na)	68,5	5,25 (t ; 7,20)
	5	71,0	5,65 (d ; 6,40)	70,5	5,62 (dd ; 1,60/8,00)
	6	77,6	5,08 (dd ; 2,80)	64,7	5,11 (dd ; 2.40/12,80)
	6'		4,01 (d ; 12,80)		4,00 (d ; 12,80)
Aromatique	H_{ar} (c)		6,77 (s ; na)		6,81 (s ; na)
	H_{ar} (d)		6,76 (s ; na)		6,78 (s ; na)
	H_{ar} (e)		6,61 (s ; na)		6,63 (s ; na)
	C=O (a)	165,4		164,0	
	C=O (b)	165,5		165,0	
	C=O (c)	166,5		166,0	
	C=O (d)	167,2		166,6	
	C=O (e)	169,3		168,6	

Tableau 30 Données RMN ^{1}H et ^{13}C-NMR (400 et 100 MHz) des composés **B** et **C** ; déplacements chimiques (δ) exprimés en ppm et constantes de couplage (J) en Hz.

La vescalagine présente le groupement hydroxy en beta et correspond donc au composé **B**, la castalagine présente le groupement en alpha et correspond donc au composé **C**. Enfin, la comparaison aux autres données spectroscopiques disponibles dans la littérature [148,173] a permis de valider le reste des données RMN collectées à l'image des signaux des trois protons aromatiques présents sur ces structures ou

du signal de chacun des cinq carbones de type carbonyl provenant des unités d'acide gallique à l'origine de la structure.

4.5 Quantification des composés majoritaires

Une méthode de dosage des composés majoritaires par étalonnage avec un standard a été mise au point pour connaître les teneurs en composés majoritaires et avoir une méthode de contrôle de la composition des extraits de *Lythrum salicaria* qui pourraient être faits (conditions analytiques en annexe 4.1.3). La teneur en composés majoritaires identifiés dans l'extrait est quantifiée en HPLC-UV en utilisant la vescalagine comme standard.

Suite à l'analyse en HPLC-DAD de l'extrait de salicaire, le maximum d'absorbance des composés en présence a été déterminé et a permis l'ajustement de la longueur adéquate pour effectuer l'étalonnage (235 nm). Concernant le gradient d'analyse, il a été raccourci par rapport à celui utilisé au début de l'étude en HPLC-MS tout en permettant une séparation satisfaisante des composés d'intérêt pour effectuer une quantification. La Figure 78 présente le chromatogramme de l'extrait étudié obtenu pour effectuer le dosage des composés majoritaires de ce dernier.

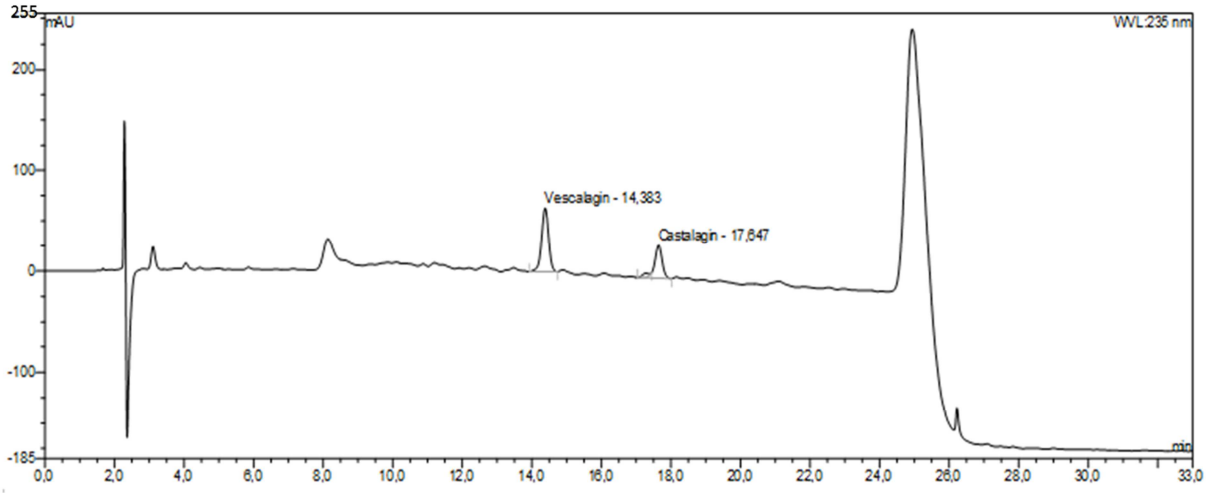


Figure 78 Chromatogramme de l'extrait de *Lythrum salicaria* L. obtenu avec le gradient utilisé pour le dosage de la vescalagine et la castalagine - UV 235 nm -

Les deux composés majoritaires de l'extrait de parties aériennes de *Lythrum salicaria* L. sont la vescalagine et la castalagine, respectivement quantifiées à hauteur de 4,6 % et 2,6 % dans l'extrait sec. La teneur totale de ces deux composés atteint ainsi 7,2 %.

5. Evaluation des activités biologiques de l'extrait

5.1 Cultures cellulaires et traitements

Les effets biologiques de l'extrait de *Lythrum salicaria* et/ou de la vescalagine (B), un de ses chémo-marqueurs sont évalués sur trois modèles distincts avec différents niveaux de complexité : des monocouches de kératinocytes (NHEK), des épidermes humains reconstruits (RHE), et des peaux reconstruites dites d'épaisseur totale (FTSK). Chacun des modèles permet d'effectuer des analyses distinctes et révélatrices d'effets à différents niveaux cellulaires, le Tableau 31 récapitule les analyses mises en œuvre dans le projet pour chaque modèle cellulaire :

Modèles	Types d'analyses effectués
Monocouches de kératinocytes (NHEK)	Expression de gènes Expression de protéines par immunofluorescence
Epidermes humains reconstruits (RHE)	Expression de protéines par histologie
Peaux reconstruites d'épaisseur totale (FTSK)	Expression de protéines par dosage ELISA Histologie

Tableau 31 Tableau de correspondance entre les modèles cellulaires utilisés pour l'étude et les types d'analyses qu'ils rendent possibles et qui ont été effectués dans l'étude qui va être présentée.

Des Kératinocytes Epidermiques Humains Normaux (Normal Human Epidermal Keratinocytes, NHEK) ont été obtenus comme décrits par Rabeony *et al.* ainsi que par Guenou *et al.* [174,175]. Les épidermes humains reconstruits (Reconstructed Human Epidermis, RHE) sont générés dans des inserts de culture à partir d'échantillons chirurgicaux comme décrits précédemment par Rabeony ou Guenou [174,175]. Le modèle de peau est composé de kératinocytes ensemencés sur une couche de collagène contenant des fibroblastes et cultivés à l'interface air/liquide pour obtenir une peau reconstruite d'épaisseur totale (Full-Thickness reconstructed Skin, FTSK)[*]**.** De plus, différents phénotypes de RHE et de FTSK permettent l'étude : un type normal de RHE/FTSK est obtenu par culture cellulaire dans un milieu complet tandis qu'un type dit ''âgé'' de RHE/FTSK obtenu par culture cellulaire dans un milieu carencé.

Les résultats qui vont être présentés proviennent de tests menés à des concentrations non cytotoxiques suivant un test MTT. C'est un test classique utilisé pour vérifier la viabilité cellulaire dans lequel les cellules sont incubées en présence de MTT (sel de tétrazolium)[†]. Les résultats ne sont pas montrés.

[*] Description de la culture des tissus : annexe 6.1.1 (NHEK), annexe 6.1.2 (RHE) et annexe 6.1.3 (FTSK)
[†] Détails sur le test en annexe 6.2.1

5.2 Effet pro-différenciant de l'extrait de *Lythrum salicaria*

5.2.1. Niveaux d'expressions de gènes impliqués dans la différenciation suivis par PCR quantitative en temps réel

Afin de mettre en évidence un effet potentiel sur la différenciation des épidermes, les monocouches de kératinocytes sont incubées avec l'extrait de *Lythrum salicaria* L. et l'effet de l'extrait est évalué au niveau de l'expression des gènes par une analyse quantitative de réaction en chaîne de polymerase et en temps réel (RT-qPCR). La méthode d'isolement et de quantification des acides ribonucléiques messagers (ARNm) est détaillée en annexe 6.2.2. La Figure 79 montre les résultats de cette expérience.

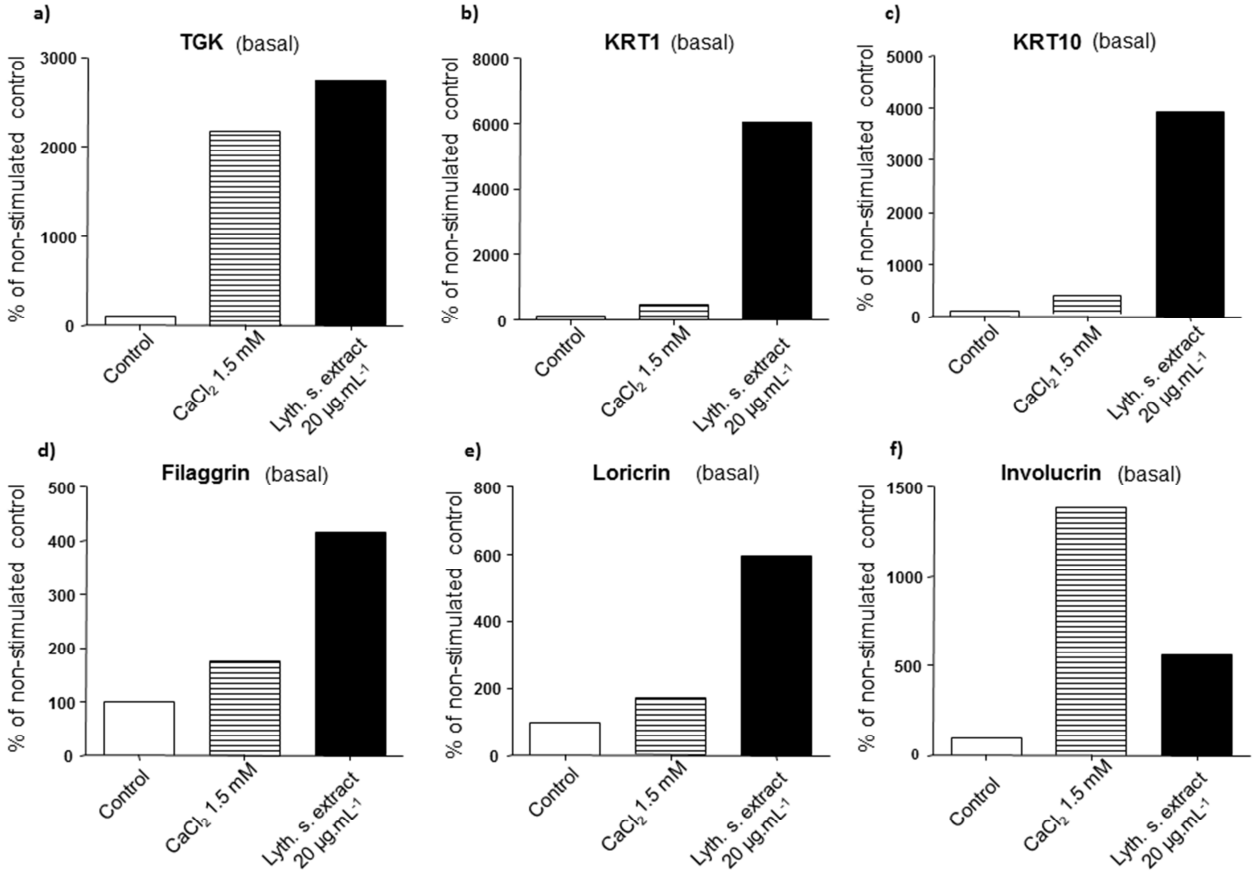


Figure 79 Effets de l'extrait de *Lythrum salicaria* sur l'expression d'ARN messagers des marqueurs de différenciation de kératinocytes. Les NHEK sont traités pendant 48 heures avec $CaCl_2$ à 1,5mM, l'extrait de *Lythrum salicaria* (extrait de Lyth. S.) à 20µg.mL^{-1} et les niveaux de ARNm sont évalués par RT-qPCR : a) TGK, b) KRT1, c) KRT10, d) Filaggrine, e) Loricrine, f) Involucrine. Ces résultats sont issus d'une expérience représentative de deux.

Un contrôle négatif indique l'expression des ARN messagers à l'état basal (absence de stimulation par agent de différenciation) qui sont normalisés pour être à 100 %. Premièrement, il faut comparer les niveaux d'expression obtenus avec le contrôle positif qui est $CaCl_2$, agent de différenciation des kératinocytes largement connu et décrit comme tel [176]. Au sein de l'épiderme, à haute concentration, le calcium génère la différenciation des kératinocytes plutôt que leur prolifération [168]. Dans notre

expérience, le niveau d'expression pour l'ensemble des ARNm suivis est supérieur à 100 % après traitement avec $CaCl_2$ à 1,5 mM. Ensuite, on observe que l'incubation des cellules en présence d'extrait a aussi induit une forte augmentation du niveau d'ARN messagers, marqueurs de différenciation, principalement pour la TGK, les kératines 1 (KRT1) et 10 (KRT10), et dans une moindre mesure pour la filaggrine, la loricrine et l'involucrine (de 400% à 6000% du contrôle non traité en fonction du marqueur). On pourra également noter que pour la plupart des marqueurs mentionnés (excepté pour l'involucrine), la stimulation induite par l'extrait de *Lythrum salicaria* L. est plus importante qu'avec $CaCl_2$.

L'observation de la présence de plusieurs protéines telles que l'involucrine, la loricrine et la filaggrine indiquent que les kératinocytes et les cornéocytes sont totalement différenciés [177]. Ces protéines structurales aident à maintenir la barrière cutanée et donne sa résistance à l'enveloppe cornée. La filaggrine permet de stabiliser le réseau de kératines, qui sert lui-même de structure de base pour les autres constituants de la couche cornée. L'involucrine et la loricrine sont réticulées par TGK1 [162]. Ce dernier joue un rôle majeur de l'assemblage des multiples composants de la couche cornée.

Les kératines spécifiques KRT1 et KRT10 sont bien plus exprimées en réponse à l'extrait de *Lythrum salicaria*. Ces kératines sont des marqueurs des couches supra-basales de l'épiderme et indiquent encore clairement que les cellules sont à un stade de développement de différenciation terminale [178].

Nous avons observé des stimulations significatives de l'expression des gènes de kératinocytes impliqués dans la différenciation et la structuration de l'épiderme grâce au traitement par l'extrait de *Lythrum salicaria*. A ce stade de l'étude, l'extrait est ainsi défini comme un agent pro-différenciant.

5.2.2. Expression de protéines caractéristiques de kératinocytes différenciés suivie par immunofluorescence

Nous avons ensuite confirmé l'effet de l'extrait de *Lythrum salicaria* au niveau protéique. Après traitement des cellules par le témoin positif ($CaCl_2$) ou par l'extrait de *Lythrum salicaria*, les monocouches de NHEK sont rincées, fixées puis incubées avec des anticorps anti-KRT10, anti-FLG ou anti-TGK. Des anticorps marqués par un produit fluorescent sont utilisés comme second anticorps et les noyaux cellulaires sont colorés avant que des photographies soient enregistrées. Le détail de la réalisation technique est en annexe 6.2.3.

L'expression des protéines TGK, KRT10 et filaggrine est évaluée sur NHEK par coloration immunofluorescente de ces protéines après 72 heures de traitement (voir Figure 80)**.** Le contrôle positif est un traitement avec $CaCl_2$ à 1,5 mM. Le contrôle négatif consiste à ne pas traiter les cellules. Par comparaison avec ces deux conditions, l'observation de l'impact du traitement par l'extrait de *Lythrum salicaria* à 20 $\mu g.mL^{-1}$ est possible.

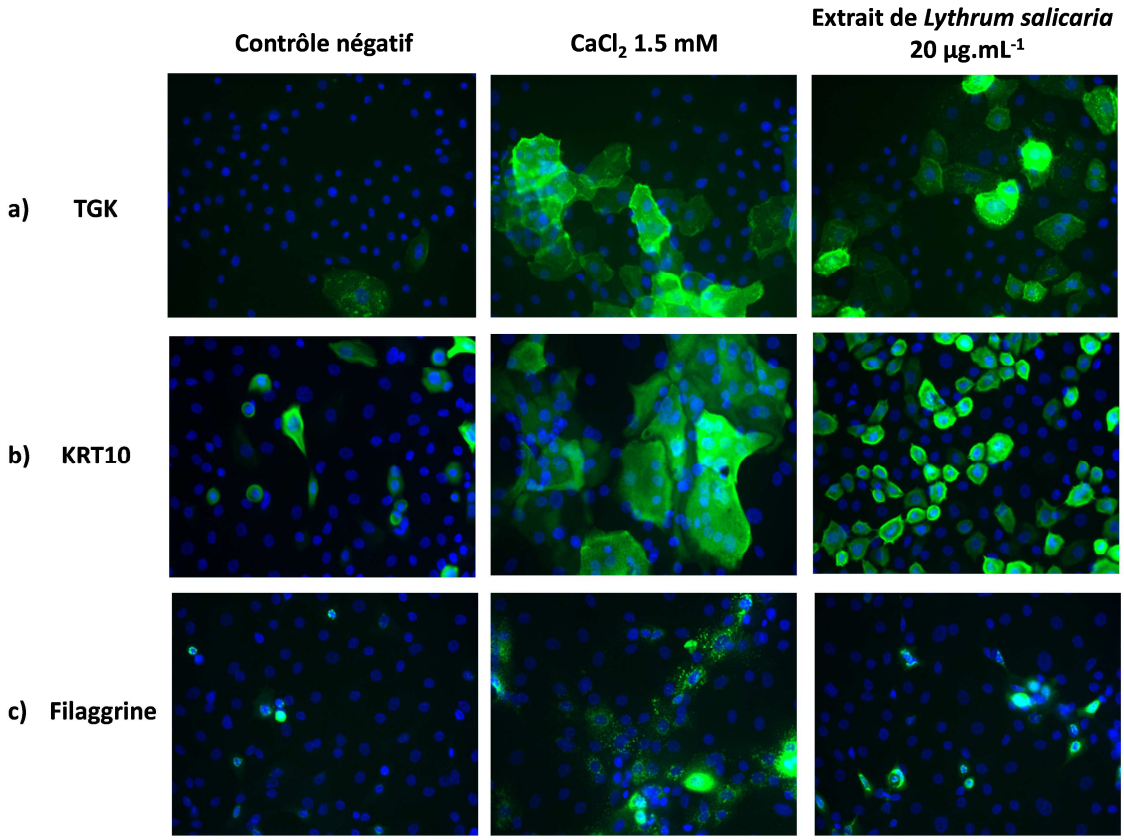


Figure 80 Analyses en immunofluorescence de l'expression de a) TGK, b) KRT10 and c) Filaggrine dans les kératinocytes épidermiques traités ou non avec $CaCl_2$ à 1,5mM ou l'extrait de Lythrum salicaria à 20 $\mu g.mL^{-1}$. Après 72 heures de traitement, les cellules sont fixées et les protéines d'intérêts sont révélées en utilisant un anticorps dirigé contre les protéines d'intérêts. Ces résultats sont issus d'une expérience représentative de trois.

Dans les cellules (noyaux cellulaires colorés en bleu) sans aucun traitement, les protéines sont peu exprimées (absence de fluorescence verte). Au contraire, l'extrait *de Lythrum salicaria* a généré un effet pro-différenciant à l'image du calcium qui, à cette concentration, joue un rôle de facteur de différenciation. L'expression des protéines montre la même tendance que précédemment à l'échelle des ARNm puisque les protéines sont largement exprimées après l'application de l'extrait. Une nouvelle fois, l'extrait démontre sa capacité à promouvoir la production de protéines permettant une différenciation satisfaisante des kératinocytes de l'épiderme pour qu'il assure efficacement ses fonctions.

5.2.3. Stimulation de l'expression de la protéine TGK dans les kératinocytes suivie par immunofluorescence

Avec l'expérience présentée au paragraphe précédent, nous avons mis en évidence l'effet de l'extrait en faveur de l'expression de protéines d'intérêt dans les monocouches de NHEK. Pour aller plus loin, nous avons voulu déterminer quelles étaient les molécules qui pouvaient être responsables de la stimulation de la différenciation des cellules de la peau. Pour cela, l'expression de TGK a été mesurée dans les monocouches de NHEK après incubation avec certaines molécules de l'extrait. Les deux composés majoritaires (vescalagine et castalagine) ont été testés ainsi que l'acide gallique, en tant qu'unité constitutive des ellagitanins cités

précédemment. La salicarinine (dimère d'ellagitanins) a été ajoutée, elle correspond au pic D (3ème pic majoritaire) qui avait été aussi purifié à partir de l'extrait.

La technique et le test utilisés sont les mêmes que dans le paragraphe précédent (après traitement, les cellules sont marquées par un premier anticorps spécifique puis par un second anticorps fluorescent). Ce test permet à la fois une observation visuelle de la fluorescence (enregistrement de photographies) et une quantification de la stimulation par traitement de ces photographies. Ici, les résultats sont représentés sous forme graphique dans la Figure 81.

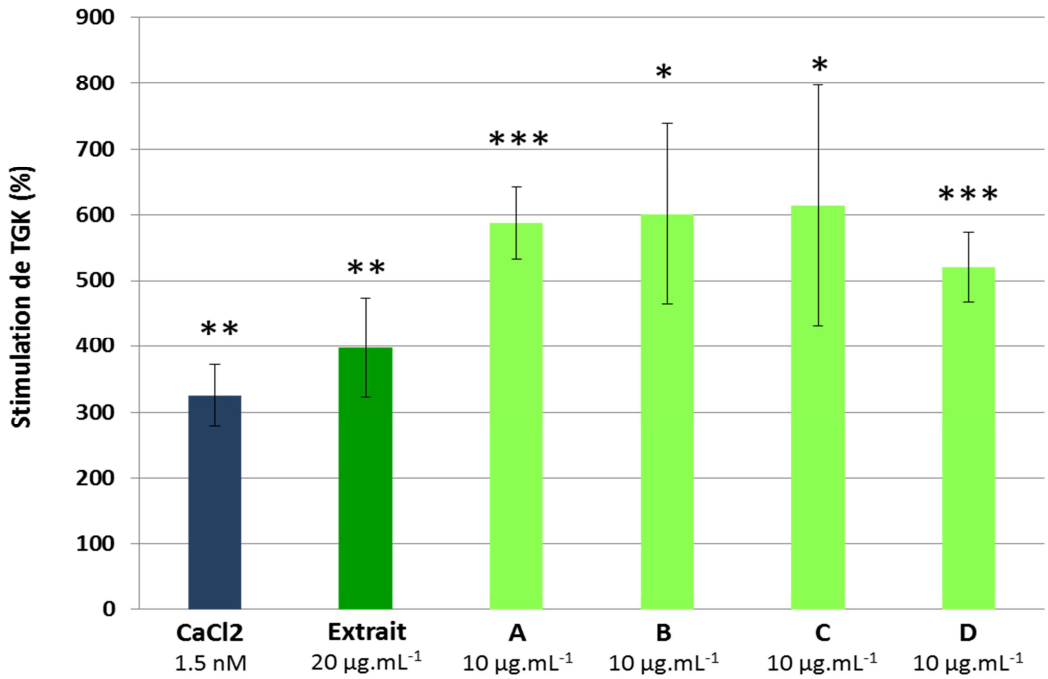


Figure 81 Expression de TGK dans les monocouches de NHEK suite à l'incubation des cellules avec le témoin positif ($CaCl_2$), l'extrait de *Lythrum salicaria*, l'acide gallique (A), la vescalagine (B), la castalagine (C) ou la salicarinine (D). (Significativités statistiques; ns : non significatif ($p > 0.05$), * : significatif ($0.01 < p < 0.05$), ** : très significatif ($0.001 < p < 0.01$), *** : extrêmement significatif ($p < 0.001$).

Les quatre molécules majoritaires de l'extrait de *Lythrum salicaria* ont induit une nette stimulation de l'expression de TGK dans les kératinocytes et ce, à un niveau équivalent. A une concentration deux fois inférieure à celle de l'extrait, les molécules ont été plus actives que ce dernier, lui-même actif à un niveau proche de celui du calcium utilisé comme témoin positif. Ainsi, l'activité de l'extrait a été expliquée par la présence des ellagitanins et de l'acide gallique : ces molécules permettent à l'épiderme d'être correctement différencié et donc de jouer son rôle de fonction barrière.

5.2.4. Caractérisation histologique d'épidermes humains reconstruits

Afin d'aller plus loin dans la caractérisation des effets bénéfiques de l'extrait de *Lythrum salicaria* au niveau de la différenciation épidermique, nous avons ensuite dirigé l'étude vers des modèles d'épidermes humains reconstruits (RHE). Des coupes transversales sont effectuées dans les épidermes humains

reconstruits pour observer l'expression de protéines après incubation avec des anticorps spécifiques et coloration histologique. La préparation de ces coupes est détaillée en annexe 6.2.4.

Dans ce modèle, la différenciation épidermique peut être observée en présence de vitamine C dans le milieu de culture. Ainsi, pour identifier un effet propre à l'extrait de *Lythrum salicaria*, ce dernier a été ajouté au milieu de culture en l'absence de vitamine C. Puis, après 7 jours de traitement, l'expression de trois marqueurs de différenciation à savoir KRT10, TGK et filaggrine a été évaluée par immunohistochimie. Les sections transversales histologiques de RHE sont montrées en Figure 82.

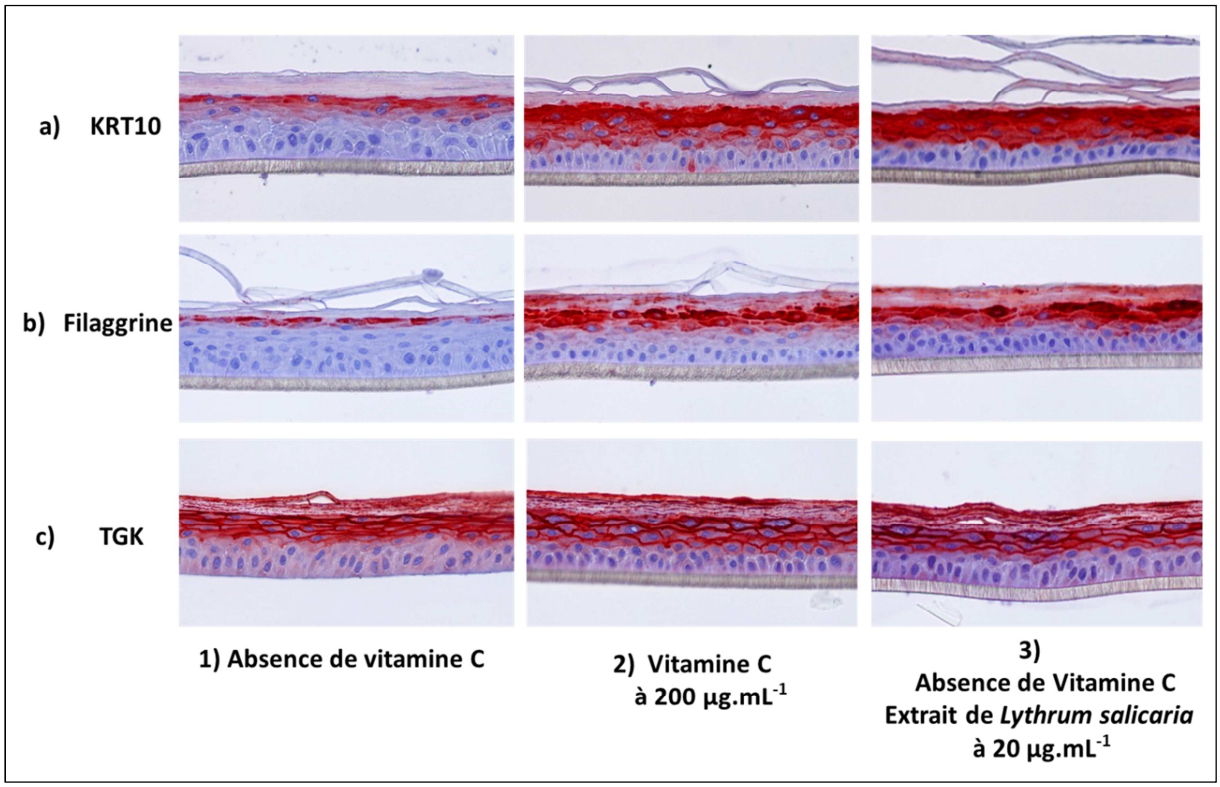


Figure 82 Différenciation des RHE : 1) sans vitamine C, 2) avec vitamine C (contrôle positif) et 3) en présence de l'extrait de *Lythrum salicaria*. Après 7 jours de traitement, l'expression des marqueurs de différenciation de l'épiderme : a) KRT10, b) filaggrine et c) TGK est évaluée par immunohistochimie. Ces résultats sont issus d'une expérience représentative de trois.

Dans notre étude, les RHE en présence de vitamine C (condition 2) expriment largement la KRT10, la filaggrine et la TGK contrairement aux RHE en absence de vitamine C (condition 1). Par ailleurs, de fortes similarités sont visibles entre les RHE traités par l'extrait de *Lythrum salicaria* (condition 3) et par la vitamine C (condition 2) qui est utilisée comme contrôle positif : de hauts niveaux des marqueurs sont exprimés comme le montre l'intensité de la coloration rouge. Les RHE traités ne sont plus du tout comparables avec ceux du contrôle négatif (absence de vitamine C). L'importante expression de marqueurs de différenciation révèle que l'extrait favorise la différenciation des kératinocytes et a donc un effet positif sur le développement des RHE.

5.3 Effet de l'extrait sur la structure de la peau

5.3.1. Amélioration de la fonction barrière de la peau sur des modèles de peaux reconstruites

Le vieillissement est caractérisé par une perte progressive de la cohésion de la peau associée à une augmentation du niveau basal d'inflammation, aussi nommé ''inflammaging'' [179]. Les résultats obtenus sur les RHE suggèrent que l'extrait de *Lythrum salicaria* peut aider à renforcer les peaux âgées ou abîmées. L'étape suivante a consisté à tester l'extrait dans un modèle plus complexe que celui des épidermes reconstruits. L'évaluation a été menée dans un modèle de peaux reconstruites d'épaisseur totale, ce qui permettait de s'approcher un peu plus de la structure de la peau. A la fin de l'incubation de 9 jours des peaux reconstruites, des coupes transversales de FTSK sont réalisées par un traitement spécifique et sont colorées (voir annexe 6.2.5). Les coupes transversales histologiques de peaux reconstruites sont présentées en Figure 83. Pour compléter l'étude, en plus de l'extrait la vescalagine, en tant que molécule majoritaire de ce dernier, a été ajoutée aux essais au moment du traitement des peaux reconstruites.

Pour commencer, on vérifie l'apparence globale des peaux reconstruites obtenues en milieu complet. Ces peaux ''normales'' sont caractérisées par la présence d'un compartiment dermique (les fibroblastes dermiques sont insérés dans une matrice de collagène) et d'une zone épidermique composée elle-même des couches cellulaires caractéristiques de l'épiderme : la couche basale, la couche de cellules épineuses, la couche granulaire et la couche cornée (ou *stratum corneum*).

En comparaison avec les peaux contrôles décrites précédemment, les peaux reconstruites de type âgé (obtenues avec le milieu de culture carencé) présentent une désorganisation caractérisée par la perte de la couche basale, le rétrécissement de la couche granulaire concomitant avec la diminution de nombre de grains de kératohyaline. Le *stratum corneum* apparait plus épais et il y a perte de la cohésion à la jonction dermo-épidermique.

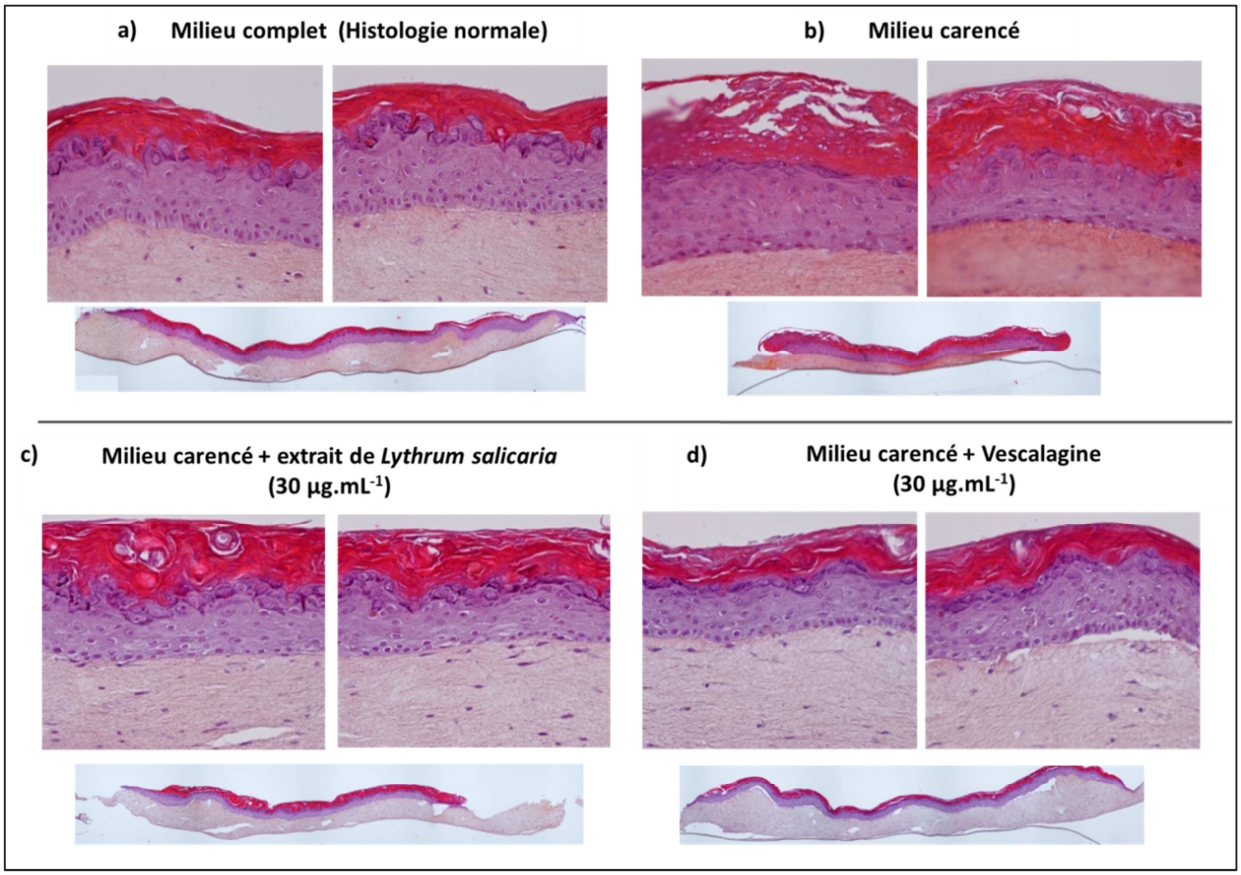


Figure 83 Evaluation de l'impact de l'extrait *de Lythrum salicaria* et de la vescalagine sur l'architecture de peaux reconstruites. Les coupes transversales histologiques des peaux sont obtenues après 9 jours d'incubation de la peau : a) en milieu complet, b) en milieu carencé et c) en présence de l'extrait de *Lythrum salicaria* ou d) de vescalagine.

Le traitement de peaux reconstruites ''âgées'' (cultivées en milieu carencé) par l'extrait ou la vescalagine conduit à une amélioration globale de leur morphologie. Elles sont presque similaires aux peaux obtenues en milieu de culture complet. Les proportions globales des trois zones de la peau (derme, épiderme, *stratum corneum*) sont similaires à celles des peaux de type normal. Au sein de l'épiderme, la couche basale et la couche granulaire sont bien définies. La cohésion entre le derme et l'épiderme est totalement retrouvée dans le cas d'un traitement avec l'extrait, et partiellement dans le cas du traitement à la vescalagine.

Ces résultats indiquent que la molécule seule possède les mêmes propriétés que l'extrait. En effet, testés à la même concentration de 30 µg.mL^{-1}, leur activité semble être équivalente. L'étude histologique de l'effet de l'extrait ou de la molécule isolée montre qu'ils sont capables de renforcer la peau. Les peaux cultivées en milieu carencé et contenant l'extrait de *Lythrum salicaria* ou la vescalagine ne sont plus abimées ou ''âgées'' mais, au contraire, présentent un aspect proche des peaux normales.

5.3.2. Effets bénéfiques sur la matrice extracellulaire

Les peaux reconstruites sont cultivées dans un insert avec une interface air-liquide (culture décrite en annexe 6.1.3). Les sous-nageants de cultures cellulaires des peaux reconstruites sont collectés pour analyses. La quantification de la libération de MMP-1, MMP-3 et du procollagène I est réalisée par dosages ELISA selon des recommandations détaillées en annexe 6.2.6.

Le derme est composé de cellules appelées fibroblastes et d'une matrice extracellulaire. Celle-ci est composée de fibres de collagène et d'élastine ainsi que de protéoglycanes et glycoprotéines. Le maillage formé structure le derme et assure des fonctions de résistance aux tensions et compressions, d'élasticité et de stockage de l'eau ou de nutriments. L'inflammation et le vieillissement sont entre autres caractérisés par l'augmentation de la libération d'enzymes dégradant la matrice extracellulaire et la baisse du taux de composés de la matrice extracellulaire tel que le collagène I. Les métalloprotéases matricielles (Matrix Metalloproteinases, MMP) sont des enzymes qui participent au remodelage de la matrice extracellulaire. MMP-1 est la collagènase prédominante qui dégrade le collagène structural (collagène de type I) tandis que MMP-3 dégrade le collagène de type III, présent notamment dans le derme juste en dessous de l'épiderme (derme papillaire), les protéoglycanes et les glycoprotéines de la matrice [180]. Par ailleurs, le procollagène I est le précurseur du collagène I. C'est le type de collagène le plus représenté dans la peau, constitutif de la matrice extracellulaire. Il assure une fonction de résistance aux tensions et d'élasticité.

Les niveaux de MMP-1, MMP-3 dans les sous-nageants de peaux reconstruites ont été suivis par dosage ELISA. Les résultats sont reportés dans les graphiques de la Figure 84 et de la Figure 85.

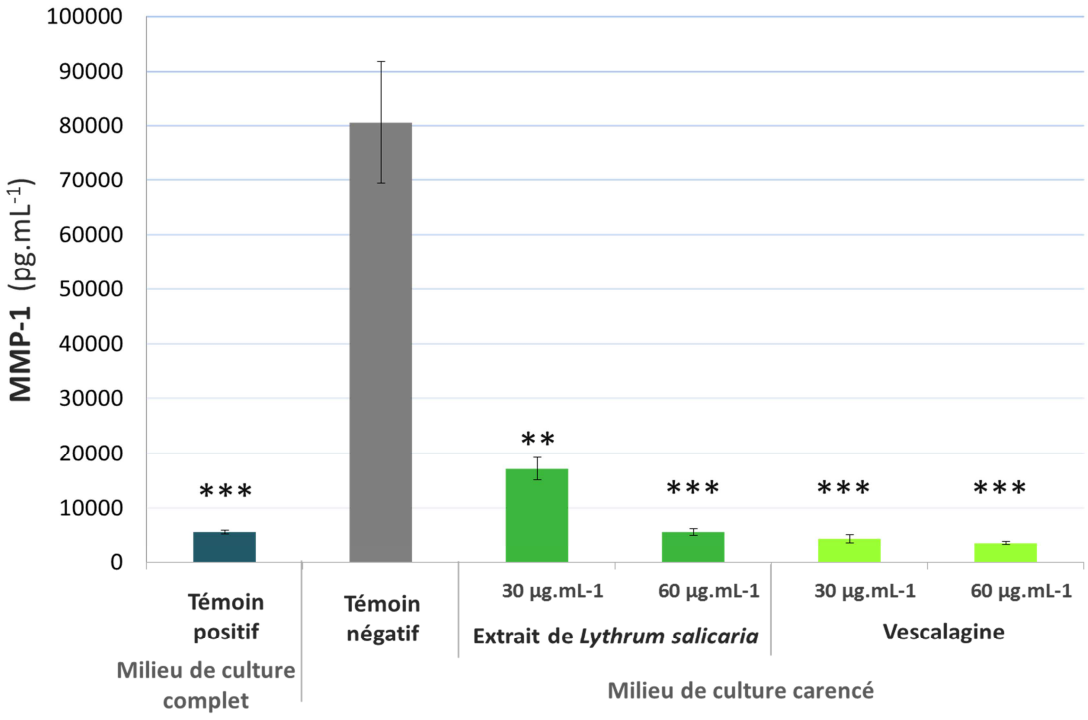


Figure 84 Effet de l'extrait de *Lythrum salicaria* et de la vescalagine sur l'expression de MMP-1 sur FTSK (Significativités statistiques; ns : non significatif ($p > 0.05$), * : significatif ($0.01 < p < 0.05$), ** : très significatif ($0.001 < p < 0.01$), *** : extrêmement significatif ($p < 0.001$)).

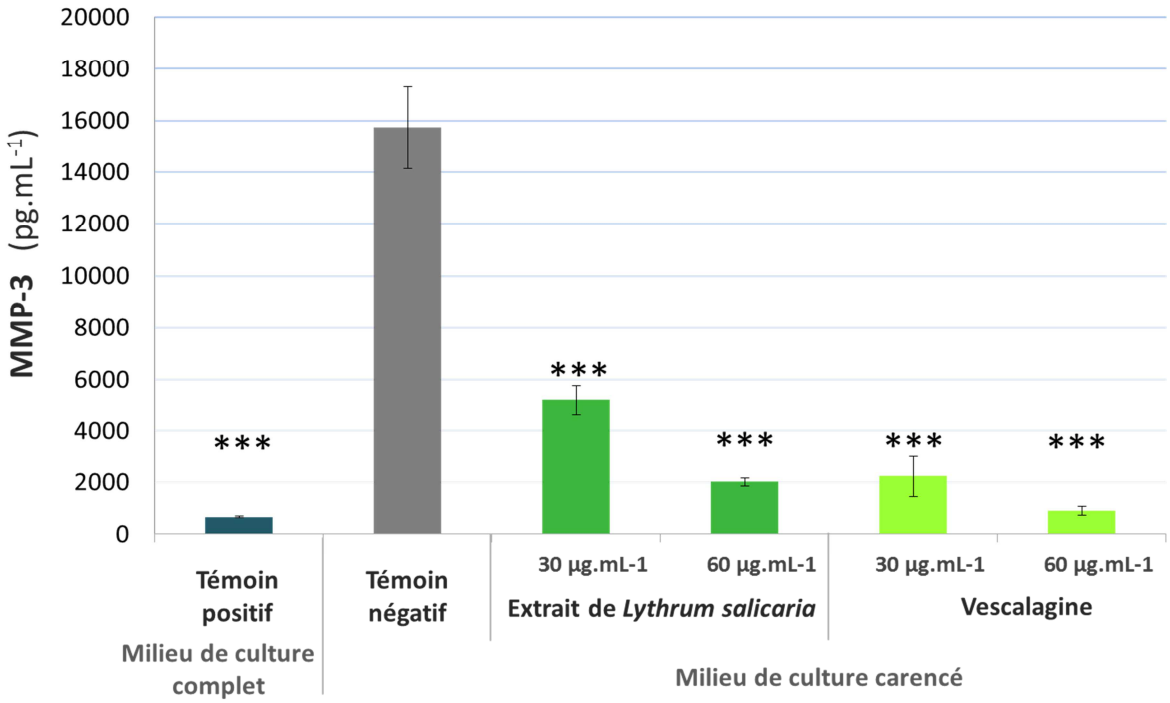


Figure 85 Effet de l'extrait de *Lythrum salicaria* et de la vescalagine sur l'expression de MMP-3 sur FTSK (Significativités statistiques ; voir détails à la Figure 84)

Comme attendu, dans les peaux cultivées en milieux carencés, il y a une augmentation considérable de la libération de MMP-1 et MMP-3 associée à une nette diminution du taux de procollagène I. Pour chaque marqueur, les différences par rapport à une peau normale (témoin reconstruit en milieu complet) varient d'un facteur supérieur à 10.

Les traitements systémiques par l'extrait et par la vescalagine impliquent des baisses très importantes de l'expression de MMP-1 et MMP-3. De plus, l'effet de l'extrait est dose-dépendant puisqu'avec une concentration de 30 µg.mL^{-1} les concentrations de MMP-1 et MMP-3 ont diminué par rapport au témoin négatif et qu'avec 60 µg.mL^{-1}, les concentrations de ces protéines sont encore plus faibles. Les produits naturels parviennent donc à contrer l'effet de la carence sur les peaux : à des concentrations de 60 µg.mL^{-1}, ils permettent un retour des niveaux de MMP équivalents à ceux des peaux normales.

Le niveau de procollagène I a parallèlement été contrôlé dans les sous-nageants de peaux reconstruites, c'est ce qu'illustre la Figure 86. Une nette augmentation du procollagène I est observée par rapport au témoin négatif mais seulement en réponse à l'extrait et non pas à la vescalagine. De plus, le niveau de procollagène I demeure plus faible que celui observé dans les peaux normales (n'ayant pas subi de carence).

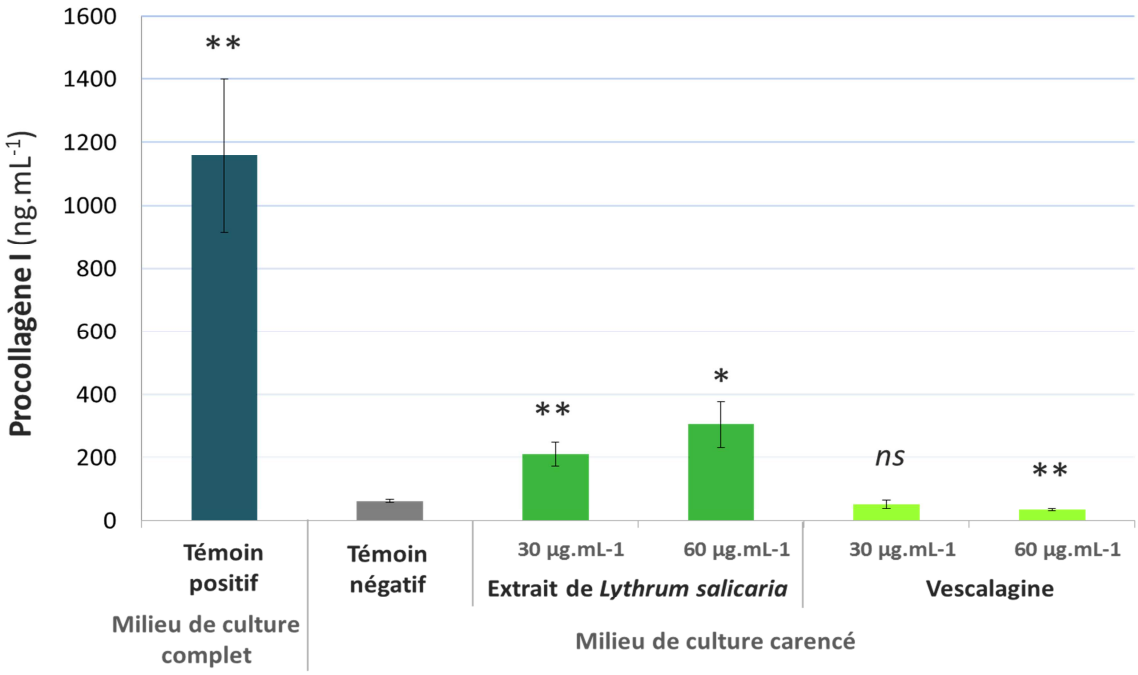


Figure 86 Effet de l'extrait de *Lythrum salicaria* et de la vescalagine sur l'expression du procollagène I sur FTSK (Significativités statistiques ; voir détails à la Figure 84)

Ainsi, les peaux traitées avec l'extrait ou la vescalagine ne sont plus comparables aux peaux obtenues en milieu carencé, mais au contraire, retrouvent un état proche de ceux de peaux normales. Ces éléments sont de forts indicateurs du fait que l'extrait de salicaire protège la peau puisque son action permet de renforcer le derme à travers la stimulation d'éléments constitutifs de sa matrice en limitant la présence d'éléments qui la dégradent. La matrice extracellulaire est une composante essentielle du derme, lui-même tissu conjonctif de la peau. Ces dosages ELISA indiquent clairement que les produits issus de la salicaire agissent d'une manière favorable sur la zone dermique, base de l'épiderme.

6. Conclusion

La salicaire est une plante commune en région Centre-Val de Loire, largement répandue dans le monde et connue depuis longtemps pour ses propriétés médicinales. Longtemps utilisée comme agent hémostatique et anti-diarrhéique, l'objectif d'amorcer une nouvelle étude phytochimique et biologique était de revisiter ce végétal en termes d'activités biologiques ciblant un usage externe sur la peau. Décrite comme riche en composés phénoliques et notamment en tanins, nous avons vérifié le contenu de la plante en nous intéressant à un extrait hydro-éthanolique à 30% d'éthanol obtenu à partir de parties aériennes en fleurs de la plante. Pour commencer, l'extrait a été analysé d'une manière globale via le dosage de familles phytochimiques telles que les sucres totaux et les composés phénoliques totaux. Ensuite, un niveau plus précis de description phytochimique a été atteint grâce à des analyses par chromatographie liquide couplée à une détection en UV et en spectrométrie de masse. Parmi les principaux résultats obtenus, on citera la mise en évidence de la présence d'acide gallique, d'une salicarinine (dimère d'ellagitanins) et de deux monomères d'ellagitanins C-glucosidiques : la vescalagine et la castalagine, qui constituent les chémo-marqueurs de l'extrait. Ces deux composés ont été isolés puis caractérisés formellement par des analyses en résonance magnétique nucléaire.

Le second axe du projet était de mettre en évidence une activité biologique pour l'extrait. Des tests d'activités biologiques ont permis d'explorer les effets tant de l'extrait hydro-éthanolique que de ses molécules constitutives et notamment du composé majoritaire, la vescalagine. L'étude a été menée sur des cellules de la peau et selon différents niveaux de complexité. En effet, les tests ont été réalisés sur des monocouches de kératinocytes, des modèles d'épidermes reconstruits et de peaux reconstruites. Cette approche a permis d'avoir une vue d'ensemble des effets biologiques via des analyses complémentaires : expression de gènes, immunofluorescence de protéines et observations histologiques des coupes transversales des tissus reconstruits.

Globalement, l'extrait et son tanin majoritaire -la vescalagine- ont agi favorablement sur la peau : leur propriété principale est une action pro-différenciante. En effet, ils ont stimulé l'expression de TGK, filaggrine et loricrine, des protéines impliquées dans le processus de différenciation des cellules au sein de l'épiderme que ce soit au niveau de l'expression génique ou protéique. Le phénomène de différenciation est essentiel pour que la peau se structure convenablement et fonctionne normalement, c'est-à-dire que ces cellules se spécialisent pour assurer des fonctions bien précises. La formation de la couche cornée est ainsi possible, ce qui rend la peau capable d'assurer sa fonction essentielle de barrière protectrice face aux dommages extérieurs.

Sur des modèles de peaux obtenus en milieu de culture carencé, l'extrait et son chémo-marqueur ont contré l'effet délétère de la carence ; les peaux ont retrouvé l'aspect de peaux normales et non plus celui

de peaux âgées. Un impact bénéfique envers la matrice extracellulaire a également été démontré par le traitement systémique avec les substances issues de la salicaire. En effet, après application de l'extrait ou de la vescalagine, on a observé une importante diminution des métalloprotéases matricielles par rapport au niveau présent dans les peaux reconstruites en milieu de culture carencé. Ces protéases sont des enzymes impliquées dans la dégradation de la matrice. Parallèlement, mais avec l'extrait uniquement, la teneur en procollagène I dans les peaux traitées était nettement supérieure à celle libérée par les peaux carencées. Ainsi, la baisse de qualité du derme impliquée par la perte de la consistance de la matrice extracellulaire est contrée par l'extrait.

A partir de ces résultats d'activités biologiques, l'extrait a pu être considéré comme un véritable actif d'intérêt pour la peau avec l'identification de molécules responsables de son activité. Ainsi, plus que simples chémo-marqueurs, les ellagitanins de l'extrait se sont avérés être des biomarqueurs.

Relier le contenu moléculaire à l'activité biologique est une étape clé pour valoriser une plante en tant qu'actif. Ce projet d'exploration de la salicaire commune (*Lythrum salicaria* L.) a donc combiné la recherche en phytochimie et chimie analytique avec des tests d'activités biologiques. Cet ensemble de travaux a permis le développement d'un ingrédient précieux pour le domaine dermo-cosmétique. Aux vues des propriétés de l'extrait, l'actif est positionné en cosmétique pour une action anti-âge permis par l'amélioration de la fonction barrière de la peau et du renforcement de la peau.

Dans l'optique de passer de l'échelle du laboratoire analytique à celle du stade pilote, nous avons par la suite mis au point un processus de fabrication d'un ingrédient. Ce dernier constitue la forme du produit fini à incorporer en tant qu'actif dans une formule cosmétique, il est standardisé pour répondre à des spécifications définies. Le développement de l'actif standardisé n'a pas été au cœur de mon travail et ne sera pas détaillé ici. L'ingrédient développé à partir de la salicaire est actuellement commercialisé par un distributeur.

Conclusion générale

Le projet doctoral de recherche au sein de l'entreprise Greenpharma et en partenariat avec l'ICOA avait pour mission principale d'améliorer les approches utilisées pour l'analyse phytochimique de plantes. La réponse donnée à cet objectif s'est traduit par le développement d'une véritable stratégie qui a pris la forme de l'outil GAINS (*Greenpharma Automated Identification of Natural Structures*). Cet outil a vu le jour grâce au travail collaboratif entre l'informaticien, qui a codé le programme informatique et moi-même en tant que phytochimiste. En accord avec le responsable du pôle phytochimie de la société, j'ai décrit le fonctionnement et la structure que devait avoir l'outil à l'informaticien et j'ai effectué les essais de tests et de validation de l'outil. Ce développement a nécessité de nombreux échanges pour avancer étapes par étapes dans la construction de l'outil. Nous avons créé un outil adapté à nos besoins en s'appuyant sur les techniques actuelles. Grâce à l'outil GAINS, l'étude rapide, pertinente et approfondie d'extraits de plantes a été possible dans le cadre de ce projet de thèse. Les extraits étudiés avaient tous vocation à être valoriser en cosmétique.

GAINS, outil de gestion et de traitement des données analytiques obtenues après analyses en chromatographie liquide couplée à la spectrométrie de masse haute résolution, fonctionne de la manière suivante. Après analyse et récupération des données, GAINS prend en charge le traitement des données afin de gérer les calculs et révéler des propositions de candidats pour les pics qui sont détectés. L'outil suit un processus d'annotation similaire à ce que ferait le phytochimiste face aux spectres de masse. Toutefois, l'outil a la capacité de le faire de façon automatisée c'est-à-dire de façon rapide, reproductible et en limitant les erreurs plus probables lorsque le travail est manuel. Les bénéfices découlant de l'utilisation d'une telle stratégie sont le gain de temps pour le laboratoire tout en conservant le caractère fiable des résultats et leur centralisation. Il reste un outil au service de l'expert sans le remplacer car celui-ci reste le garant de la validité des résultats et celui qui décide à partir de ces éléments, de mettre en place ou non les actions complémentaires nécessaires à l'étude, au cas par cas. La mise au point et l'optimisation des algorithmes créés pour GAINS ont été faites avec une évaluation sur un échantillon test de 150 molécules standards en mélange. GAINS est une interface web qui pourra évoluer et être modifiée en fonction des besoins et des développements techniques futurs. Ce travail de thèse se clôt mais les projets vont continuer. GAINS est désormais intégré en tant qu'outil support aux études phytochimiques dans la société Greenpharma. Certaines fonctions de l'outil devront être ajoutées pour rendre automatique des étapes encore manuelles ou pour répondre à des besoins particuliers (mise en évidence de composés exogènes de la plante tels que les pesticides par exemple).

Conclusion générale.

Les premières applications de GAINS pour caractériser différentes plantes nous ont permis de véritablement nous rendre compte de son intérêt. Pour Greenpharma comme pour ses collaborateurs industriels, les critères pouvant être plus ou moins exigeants en termes de niveau de caractérisation à atteindre, différentes fonctions de GAINS ont été valorisées. Trois cas d'études illustrant les différents apports de GAINS ont été présentés dans ce mémoire. Il est ainsi possible de concentrer l'étude sur les pics chromatographiques majeurs d'un extrait et de s'appuyer sur les données de fragmentation (de façon manuelle ou *via* Metfrag). Il est également possible d'obtenir la masse monoisotopique de l'ensemble des pics détectés quelle que soit leur intensité ; ainsi une vue d'ensemble des masses moléculaires des composés de l'extrait est proposée. De plus, la connexion avec la base de données de produits naturels riche d'environ 160 000 composés de Greenpharma permet : d'attribuer des structures connues à toute masse déterminée et de mettre en évidence des familles phytochimiques particulières via des requêtes adaptées. Le profilage métabolique possible avec GAINS renforce le rôle central que représente la phytochimie dans les projets de R&D mis en place au sein de Greenpharma.

Pour la tulipe noire (*Tulipa gesneriana cv. Queen of the Night*), en associant les données LC-HRMS et MS/MS à une analyse des génines des flavonoïdes présents dans l'extrait, nous avons déterminé la nature des flavonoïdes glycosylés présents (génine, sucre et substitution) dont celles des trois composés majoritaires. Puis, nous avons proposé un nombre restreint de candidats structuraux. Nous savons aussi que ces molécules n'ont pas été décrites pour la tulipe, leur masse étant absente de la liste des composés connus pour les tulipes du genre *gesneriana*. L'extrait étudié entre dans la composition d'un actif cosmétique, la Dumaflorine.

Concernant la primevère vraie (*Primula veris*), elle constitue la preuve de concept de notre outil en dehors de l'utilisation de standards pour définir ses performances. En effet, nous avons abordé l'extrait d'une manière classique en le fractionnant et purifiant le maximum de composés et parallèlement, nous avons analysé l'extrait en LC-HRMS pour traiter les résultats à l'aide de GAINS. Cette double étude de l'extrait montre que ce qui est déduit par le traitement semi-automatisé après UHPLC-HRMS en termes de contenu est plus pertinent que le travail d'isolement et d'identification par HRMS et RMN des molécules pures. En effet, en un temps bien plus court, on a pu déterminer les génines et les sucres des composés majoritaires. On a aussi obtenu des éléments sur l'ensemble des pics détectés et non pas uniquement sur les principaux. La démarche a ainsi abouti sur des propositions d'identifications fiables. Reste que la purification permet d'obtenir des molécules pures à tester biologiquement, ce qui peut être intéressant pour définir les biomarqueurs de l'extrait. Il faut mentionner que des contraintes sont rencontrées dans les deux approches. Avec GAINS, on ne peut pas décrire formellement les enchaînements des sucres des flavonoïdes qui, dans la plante étudiée, sont majoritairement diglycosylés et triglycosylés. Dans le cas de l'isolement, on n'est pas en capacité d'isoler tous les composés et l'obtention d'une grande pureté pour une molécule isolée demande

de multiples étapes. L'extrait a ensuite été travaillé au laboratoire pour valider son mode d'obtention et sa stabilité, ces aspects restent actuellement en cours. En termes d'activités biologiques, l'extrait de fleurs de *Primula veris,* riche en flavonoïdes glycosylés, a été testé et pourrait constituer un futur ingrédient d'intérêt pour la cosmétique. La propriété ciblée n'est pas encore divulguée mais les premiers résultats ont été obtenus et sont en cours de confirmation.

Au sujet de la réglisse glabre (*Glycyrrhiza glabra*), le travail supporté par GAINS nous a permis de formuler des propositions d'identifications fortes pour les pics majoritaires du profil de l'extrait étudié. Ils appartiennent à la famille des isoflavones et des chalcones. Il a aussi été entrepris de mettre en évidence l'occurrence de la famille des isoflavones dans la totalité du profil chromatographique. Grâce à l'étude des candidats associés par requête au sein de la base GPDB à l'ensemble des pics, il apparait qu'une quinzaine de pics sont potentiellement des isoflavones. Parmi elles, la daïdzeine (connue pour être présente dans la réglisse glabre) semble expliquer un des pics de l'extrait du fait notamment de la correspondance de la fragmentation du composé détecté avec celui de la daïdzeine.

En dehors de l'utilisation de GAINS, une autre étude plus classique a été menée. La salicaire (*Lythrum salicaria*), plante très répandue dans les zones humides de la région Centre-Val de Loire, a été sélectionnée. La caractérisation d'un extrait, sélectionné suite à des tests d'activités préliminaires, a été réalisée. Ses deux chémomarqueurs ont été identifiés. Leur activité biologique a démontré qu'ils étaient également des biomarqueurs de l'extrait. Ainsi, cette plante traditionnellement utilisée en tant qu'anti-diarrhéique et agent hémostatique a été repositionnée comme actif anti-âge suite à des tests d'activités biologiques sur plusieurs supports (monocouches de kératinocytes, épidermes humains reconstruits et peaux reconstruites). Nos travaux ont abouti à une publication[#] dans le journal COSMETICS, au dépôt de deux brevets[*,†] et au développement d'un ingrédient viable, aujourd'hui commercialisé par un distributeur. Si le projet sur la salicaire est terminé, l'étude de plantes locales reste d'actualité. En effet, toutes les plantes peuvent avoir un intérêt même celles qui ne sont pas rares ou exotiques et travailler des matières locales permet une démarche écologique, simple et logique.

Le travail de thèse effectué au sein de Greenpharma a eu plusieurs répercussions. Premièrement, le déploiement de GAINS pour les études phytochimiques a nécessité d'y consacrer du temps mais l'outil est devenu un pivot dans les approches de développement d'ingrédients pour l'entreprise. Il vient compléter les

[#] JOURAVEL, G.; GUÉNIN, S.; BERNARD, F.-X.; ELFAKIR, C.; BERNARD, P.; HIMBERT, F. New Biological Activities of *Lythrum salicaria* L.: Effects on Keratinocytes, Reconstructed Epidermis and Reconstructed Skins, Applications in Dermo-Cosmetic Sciences. *Cosmetics* **2017**, *4*, 52.

[*] BERNARD P., BERNARD F-X., Applications cosmétiques et pharmaceutiques de la salicaire, FR3018685, 2015

[†] BERNARD P., BERNARD F-X., HIMBERT F., Applications cosmétiques et pharmaceutiques de l'acide gallique et de dérivés de l'acide gallique, WO2015140470, 2015

outils de valorisation des données contenues dans la base de données de Greenpharma. Certaines fonctions de l'outil devront être ajoutées pour rendre automatique certaines étapes encore manuelles ou pour répondre à des besoins particuliers (mise en évidence de composés exogènes de la plante tels que les pesticides par exemple). Deuxièmement, des étapes importantes ont été réalisées à différents stades du processus de développement d'ingrédients cosmétiques. Ainsi, un ingrédient issu de la salicaire est aujourd'hui sur le marché, un second utilisant la tulipe sera présenté au secteur de la cosmétique à partir d'avril 2018, et la mise au point du troisième produit autour de la primevère vraie avance.

Pour ma part, ce projet de thèse a bénéficié du double environnement constitué par le milieu industriel et le milieu académique. J'ai acquis une vision globale des contraintes intrinsèques à chaque environnement, et notamment de celles du secteur privé qui a besoin que les projets trouvent des applications concrètes le plus rapidement possible tout en assurant un travail de recherche de pointe et de qualité. Au sein de l'entreprise, les projets pouvaient évoluer sur plusieurs aspects en parallèle et nécessitaient un suivi global et partagé entre les différents collaborateurs. Cette thèse a donc été pour moi une véritable expérience et préparation à ce qui m'attend dans le futur en termes de gestion de projets de R&D, de mobilisation de ressources multiples et d'interactions diversifiées avec les acteurs d'un projet collaboratif ou au sein même d'une entreprise.

Références bibliographiques

[1] P. André, J.-M. Aubry, S. Berteina-Raboin, C. Bouix-Peter, S. Del Bino, P. Humbert, J. Leclaire, L. Marrot, C. Masson, I. Pélisson, P. Piccerelle, P. Walter, Chimie, dermo-cosmétique et beauté, 1ère, EDP Sciences, 2017.

[2] Code la santé publique, www.legifrance.gouv.fr (accessed January 20, 2018).

[3] P.K. Mukherjee, S. Bahadur, R.K. Harwansh, S. Biswas, S. Banerjee, Paradigm shift in natural product research: traditional medicine inspired approaches, Phytochem. Rev. 16 (2017) 803–826. doi:10.1007/s11101-016-9489-6.

[4] A.L. Harvey, R. Edrada-Ebel, R.J. Quinn, The re-emergence of natural products for drug discovery in the genomics era, Nat. Rev. Drug Discov. 14 (2015) 111–129. doi:10.1038/nrd4510.

[5] TPE Crèmes hydratantes, tpe-cremes-hydratantes.jimdo.com/accueil/i-la-peau/ (accessed February 12, 2018).

[6] Santé Médecine Journal des femmes, sante-medecine.journaldesfemmes.fr/faq/17795-derme-definition (accessed February 13, 2018).

[7] J. Bruneton, Pharmacognosie, phytochimie, plantes médicinales, 5ième, Editions Lavoisier, 2016.

[8] F. Hallé, Eloge de la plante, Pour une nouvelle biologie, Editions du Seuil, 1999.

[9] A.R. Fernie, The future of metabolic phytochemistry: Larger numbers of metabolites, higher resolution, greater understanding, Phytochemistry. 68 (2007) 2861–2880. doi:10.1016/j.phytochem.2007.07.010.

[10] F. Rouessac, A. Rouessac, Analyse chimique. Méthodes et techniques instrumentales, 8ième, DUNOD, 2016.

[11] M. Halabalaki, K. Vougogiannopoulou, E. Mikros, A.L. Skaltsounis, Recent advances and new strategies in the NMR-based identification of natural products, Curr. Opin. Biotechnol. 25 (2014) 1–7. doi:10.1016/j.copbio.2013.08.005.

[12] L.W. Sumner, A. Amberg, D. Barrett, M.H. Beale, R. Beger, C.A. Daykin, T.W.-M. Fan, O. Fiehn, R. Goodacre, J.L. Griffin, T. Hankemeier, N. Hardy, J. Harnly, R. Higashi, J. Kopka, A.N. Lane, J.C. Lindon, P. Marriott, A.W. Nicholls, M.D. Reily, J.J. Thaden, M.R. Viant, Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI), Metabolomics. 3 (2007) 211–221. doi:10.1007/s11306-007-0082-2.

[13] The metabolomics society, http://metabolomicssociety.org/ (accessed February 1, 2018).

[14] E. Rathahao-Paris, S. Alves, C. Junot, J.-C. Tabet, High resolution mass spectrometry for structural identification of metabolites in metabolomics, Metabolomics. 12 (2016). doi:10.1007/s11306-015-0882-8.

[15] H.G. Gika, I.D. Wilson, G.A. Theodoridis, LC–MS-based holistic metabolic profiling. Problems, limitations, advantages, and future perspectives, J. Chromatogr. B. 966 (2014) 1–6. doi:10.1016/j.jchromb.2014.01.054.

[16] J.-L. Wolfender, G. Marti, A. Thomas, S. Bertrand, Current approaches and challenges for the metabolite profiling of complex natural extracts, J. Chromatogr. A. 1382 (2015) 136–164. doi:10.1016/j.chroma.2014.10.091.

[17] E. de Hoffmann, V. Stroobant, Spectrométrie de masse, Cours et exercices corrigés, 3ième, DUNOD, 2005.

[18] A.G. Brenton, A.R. Godfrey, Accurate mass measurement: Terminology and treatment of data, J. Am. Soc. Mass Spectrom. 21 (2010) 1821–1835.

[19] C. Deborde, A. Moing, L. Roch, D. Jacob, D. Rolin, P. Giraudeau, Plant metabolism as studied by NMR spectroscopy, Prog. Nucl. Magn. Reson. Spectrosc. 102–103 (2017) 61–97. doi:10.1016/j.pnmrs.2017.05.001.

[20] J. Hubert, J.-M. Nuzillard, J.-H. Renault, Dereplication strategies in natural product research: How many tools and methodologies behind the same concept?, Phytochem. Rev. (2015). doi:10.1007/s11101-015-9448-7.

[21] E. Werner, J.-F. Heilier, C. Ducruix, E. Ezan, C. Junot, J.-C. Tabet, Mass spectrometry for the identification of the discriminating signals from metabolomics: Current status and future trends, J. Chromatogr. B. 871 (2008) 143–163. doi:10.1016/j.jchromb.2008.07.004.

[22] S. Forcisi, F. Moritz, B. Kanawati, D. Tziotis, R. Lehmann, P. Schmitt-Kopplin, Liquid chromatography–mass spectrometry in metabolomics research: Mass analyzers in ultra high pressure liquid chromatography coupling, J. Chromatogr. A. 1292 (2013) 51–65. doi:10.1016/j.chroma.2013.04.017.

[23] C. Junot, F. Fenaille, B. Colsch, F. Bécher, High resolution mass spectrometry based techniques at the crossroads of metabolic pathways: High resolution mass spectrometry for metabolomics, Mass Spectrom. Rev. 33 (2014) 471–500. doi:10.1002/mas.21401.

[24] G.A. Theodoridis, H.G. Gika, E.J. Want, I.D. Wilson, Liquid chromatography–mass spectrometry based global metabolite profiling: A review, Anal. Chim. Acta. 711 (2012) 7–16. doi:10.1016/j.aca.2011.09.042.

[25] E. Lesellier, Extension of the carotenoid test to superficially porous C18 bonded phases, aromatic ligand types and new classical C18 bonded phases, J. Chromatogr. A. 1266 (2012) 34–42. doi:10.1016/j.chroma.2012.09.068.

[26] J. Boccard, J.-L. Veuthey, S. Rudaz, Knowledge discovery in metabolomics: An overview of MS data handling, J. Sep. Sci. 33 (2010) 290–304. doi:10.1002/jssc.200900609.

[27] J.W. Allwood, R. Goodacre, An Introduction to Liquid Chromatography– Mass Spectrometry Instrumentation Applied in Plant Metabolomic Analyses, Phytochem. Anal. 21 (2010) 33–47. doi:10.1002/pca.1187.

[28] M. Holčapek, R. Jirásko, M. Lísa, Basic rules for the interpretation of atmospheric pressure ionization mass spectra of small molecules, J. Chromatogr. A. 1217 (2010) 3908–3921. doi:10.1016/j.chroma.2010.02.049.

[29] C. Bure, Fragmentation dans une source à électronébulisation, Université d'Orléans, 2005. https://tel.archives-ouvertes.fr/tel-00010600/ (accessed August 30, 2017).

[30] P. Kebarle, A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry, J. Mass Spectrom. 35 (2000) 804–817.

[31] University of Bristol. School of Chemistry, www.chm.bris.ac.uk/ms/esi-ionisation.xhtml (accessed February 1, 2018).

[32] Bruker Daltonik, maXis user manual, (2008).

[33] P.-M. Allard, G. Genta-Jouve, J.-L. Wolfender, Deep metabolome annotation in natural products research: towards a virtuous cycle in metabolite identification, Curr. Opin. Chem. Biol. 36 (2017) 40–49. doi:10.1016/j.cbpa.2016.12.022.

[34] R. Verpoorte, Y.H. Choi, H.K. Kim, Ethnopharmacology and systems biology: A perfect holistic match, J. Ethnopharmacol. 100 (2005) 53–56. doi:10.1016/j.jep.2005.05.033.

[35] J.A. Beutler, A.B. Alvarado, D.E. Schaufelberger, P. Andrews, T.G. McCloud, Dereplication of phorbol bioactives: Lyngbya majuscula and Croton cuneatus, J. Nat. Prod. 53 (1990) 867–874.

[36] P.-M. Allard, T. Péresse, J. Bisson, K. Gindro, L. Marcourt, V.C. Pham, F. Roussi, M. Litaudon, J.-L. Wolfender, Integration of Molecular Networking and *In-Silico* MS/MS Fragmentation for Natural Products Dereplication, Anal. Chem. 88 (2016) 3317–3323. doi:10.1021/acs.analchem.5b04804.

[37] S.R. Johnson, B.M. Lange, Open-Access Metabolomics Databases for Natural Product Research: Present Capabilities and Future Potential, Front. Bioeng. Biotechnol. 3 (2015). doi:10.3389/fbioe.2015.00022.

[38] Q. Do, J. Medina-Franco, T. Scior, P. Bernard, How to Valorize Biodiversity? Let's Go Hashing, Extracting, Filtering, Mining, Fishing, Planta Med. 81 (2015) 436–449. doi:10.1055/s-0034-1396314.

[39] J. Gu, Y. Gui, L. Chen, G. Yuan, H.-Z. Lu, X. Xu, Use of Natural Products as Chemical Library for Drug Discovery and Network Pharmacology, PLoS ONE. 8 (2013) e62839. doi:10.1371/journal.pone.0062839.

[40] M. Dunkel, SuperNatural: a searchable database of available natural compounds, Nucleic Acids Res. 34 (2006) D678–D683. doi:10.1093/nar/gkj132.

[41] P. Banerjee, J. Erehman, B.-O. Gohlke, T. Wilhelm, R. Preissner, M. Dunkel, Super Natural II—a database of natural products, Nucleic Acids Res. 43 (2015) D935–D939. doi:10.1093/nar/gku886.

[42] Y. Nakamura, F. Mochamad Afendi, A. Kawsar Parvin, N. Ono, K. Tanaka, A. Hirai Morita, T. Sato, T. Sugiura, M. Altaf-Ul-Amin, S. Kanaya, KNApSAcK Metabolite Activity Database for Retrieving the Relationships Between Metabolites and Biological Activities, Plant Cell Physiol. 55 (2014) e7–e7. doi:10.1093/pcp/pct176.

[43] J.G. Jeffryes, R.L. Colastani, M. Elbadawi-Sidhu, T. Kind, T.D. Niehaus, L.J. Broadbelt, A.D. Hanson, O. Fiehn, K.E.J. Tyo, C.S. Henry, MINEs: open access databases of computationally predicted enzyme promiscuity products for untargeted metabolomics, J. Cheminformatics. 7 (2015). doi:10.1186/s13321-015-0087-1.

[44] J. González-Lergier, L.J. Broadbelt, V. Hatzimanikatis, Theoretical considerations and computational analysis of the complexity in polyketide synthesis pathways, J Am Chem Soc. 127 (2005) 9930–9938.

[45] C.S. Henry, M. Jankowski, L.J. Broadbelt, V. Hatzimanikatis, Genome- scale thermodynamic analysis of Escherichia coli metabolism, Biophys J. 90 (2006) 1453–1461.

[46] C.A. Smith, G. O'Maille, E.J. Want, C. Qin, S.A. Trauger, T.R. Brandon, D.E. Custudio, R. Abagyan, G. Siuzdak, METLIN: a metabolite mass spectral database, Ther. Drug Monit. 27 (2005) 747–751.

[47] H. Horai, M. Arita, S. Kanaya, Y. Nihei, T. Ikeda, K. Suwa, Y. Ojima, K. Tanaka, S. Tanaka, K. Aoshima, Y. Oda, Y. Kakazu, M. Kusano, T. Tohge, F. Matsuda, Y. Sawada, M.Y. Hirai, H. Nakanishi, K. Ikeda, N. Akimoto, T. Maoka, H. Takahashi, T. Ara, N. Sakurai, H. Suzuki, D. Shibata, S. Neumann, T. Iida, K. Tanaka, K. Funatsu, F. Matsuura, T. Soga, R. Taguchi, K. Saito, T. Nishioka, MassBank: a public repository for sharing mass spectral data for life sciences, J. Mass Spectrom. 45 (2010) 703–714. doi:10.1002/jms.1777.

[48] NIST, www.nist.gov/srd (accessed February 15, 2018).

[49] mzCloud, www.mzcloud.org/ (accessed February 15, 2018).

[50] Y. Sawada, R. Nakabayashi, Y. Yamada, M. Suzuki, M. Sato, A. Sakata, K. Akiyama, T. Sakurai, F. Matsuda, T. Aoki, M.Y. Hirai, K. Saito, RIKEN tandem mass spectral database (ReSpect) for phytochemicals: A plant-specific MS/MS-based data resource and database, Phytochemistry. 82 (2012) 38–45. doi:10.1016/j.phytochem.2012.07.007.

[51] M. Wang, J.J. Carver, V.V. Phelan, L.M. Sanchez, N. Garg, Y. Peng, D.D. Nguyen, J. Watrous, C.A. Kapono, T. Luzzatto-Knaan, C. Porto, A. Bouslimani, A.V. Melnik, M.J. Meehan, W.-T. Liu, M. Crüsemann, P.D. Boudreau, E. Esquenazi, M. Sandoval-Calderón, R.D. Kersten, L.A. Pace, R.A. Quinn, K.R. Duncan, C.-C. Hsu, D.J. Floros, R.G. Gavilan, K. Kleigrewe, T. Northen, R.J. Dutton, D. Parrot, E.E. Carlson, B. Aigle, C.F. Michelsen, L. Jelsbak, C. Sohlenkamp, P. Pevzner, A. Edlund, J. McLean, J. Piel, B.T. Murphy, L. Gerwick, C.-C. Liaw, Y.-L. Yang, H.-U. Humpf, M. Maansson, R.A. Keyzers, A.C. Sims, A.R. Johnson, A.M. Sidebottom, B.E. Sedio, A. Klitgaard, C.B. Larson, C.A. Boya P, D. Torres-Mendoza, D.J. Gonzalez, D.B. Silva, L.M. Marques, D.P. Demarque, E. Pociute, E.C. O'Neill, E. Briand, E.J.N. Helfrich, E.A. Granatosky, E. Glukhov, F. Ryffel, H. Houson, H. Mohimani, J.J. Kharbush, Y. Zeng, J.A. Vorholt, K.L. Kurita, P. Charusanti, K.L. McPhail, K.F. Nielsen, L. Vuong, M. Elfeki, M.F. Traxler, N. Engene, N. Koyama, O.B. Vining, R. Baric, R.R. Silva, S.J. Mascuch, S. Tomasi, S. Jenkins, V. Macherla, T. Hoffman, V. Agarwal, P.G. Williams, J. Dai, R. Neupane, J. Gurr, A.M.C. Rodríguez, A. Lamsa, C. Zhang, K. Dorrestein, B.M. Duggan, J. Almaliti, P.-M. Allard, P. Phapale, L.-F. Nothias, T. Alexandrov, M. Litaudon, J.-L. Wolfender, J.E. Kyle, T.O. Metz, T. Peryea, D.-T. Nguyen, D. VanLeer, P. Shinn, A. Jadhav, R. Müller, K.M. Waters, W. Shi, X. Liu, L. Zhang, R. Knight, P.R. Jensen, B.Ø. Palsson, K. Pogliano, R.G. Linington, M. Gutiérrez, N.P. Lopes, W.H. Gerwick, B.S. Moore, P.C. Dorrestein, N. Bandeira, Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking, Nat. Biotechnol. 34 (2016) 828–837. doi:10.1038/nbt.3597.

[52] C. Steinbeck, S. Krause, S. Kuhn, NMRShiftDB Constructing a Free Chemical Information System with Open-Source Components, J. Chem. Inf. Comput. Sci. 43 (2003) 1733–1739. doi:10.1021/ci0341363.

[53] J.L. Lopez-Perez, R. Theron, E. del Olmo, D. Diaz, NAPROC-13: a database for the dereplication of natural product mixtures in bioassay-guided protocols, Bioinformatics. 23 (2007) 3256–3257. doi:10.1093/bioinformatics/btm516.

[54] Q. Cui, I.A. Lewis, A.D. Hegeman, M.E. Anderson, J. Li, C.F. Schulte, W.M. Westler, H.R. Eghbalnia, M.R. Sussman, J.L. Markley, Metabolite identification via the Madison Metabolomics Consortium Database, Nat. Biotechnol. 26 (2008) 162.

[55] E.L. Ulrich, H. Akutsu, J.F. Doreleijers, Y. Harano, Y.E. Ioannidis, J. Lin, M. Livny, S. Mading, D. Maziuk, Z. Miller, E. Nakatani, C.F. Schulte, D.E. Tolmie, R. Kent Wenger, H. Yao, J.L. Markley, BioMagResBank, Nucleic Acids Res. 36 (2007) D402–D408. doi:10.1093/nar/gkm957.

[56] K. Dührkop, H. Shen, M. Meusel, J. Rousu, S. Böcker, Searching molecular structure databases with tandem mass spectra using CSI:FingerID, Proc. Natl. Acad. Sci. 112 (2015) 12580–12585. doi:10.1073/pnas.1509788112.

[57] F. Hufsky, K. Scheubert, S. Böcker, Computational mass spectrometry for small-molecule fragmentation, TrAC Trends Anal. Chem. 53 (2014) 41–48. doi:10.1016/j.trac.2013.09.008.

[58] M.E. Elyashberg, A.J. Williams, G.E. Martin, Computer-assisted structure verification and elucidation tools in NMR-based structure elucidation, Prog. Nucl. Magn. Reson. Spectrosc. 53 (2008) 1–104. doi:10.1016/j.pnmrs.2007.04.003.

[59] J. Hubert, S. Chollet, S. Purson, R. Reynaud, D. Harakat, A. Martinez, J.-M. Nuzillard, J.-H. Renault, Exploiting the Complementarity between Dereplication and Computer-Assisted Structure Elucidation for the Chemical Profiling of Natural Cosmetic Ingredients: *Tephrosia purpurea* as a Case Study, J. Nat. Prod. 78 (2015) 1609–1617. doi:10.1021/acs.jnatprod.5b00174.

[60] ACD/LABS NMR Predictors, www.acdlabs.com/products/adh/nmr/nmr_pred/ (accessed June 2, 2018).

[61] PERCH Solutions, http://new.perchsolutions.com/ (accessed February 6, 2018).

[62] R. Goodacre, D. Broadhurst, A.K. Smilde, B.S. Kristal, J.D. Baker, R. Beger, C. Bessant, S. Connor, G. Capuani, A. Craig, T. Ebbels, D.B. Kell, C. Manetti, J. Newton, G. Paternostro, R. Somorjai, M. Sjöström, J. Trygg, F. Wulfert, Proposed minimum reporting standards for data analysis in metabolomics, Metabolomics. 3 (2007) 231–241. doi:10.1007/s11306-007-0081-3.

[63] J. Hubert, J.-M. Nuzillard, S. Purson, M. Hamzaoui, N. Borie, R. Reynaud, J.-H. Renault, Identification of Natural Metabolites in Mixture: A Pattern Recognition Strategy Based on ^{13} C NMR, Anal. Chem. 86 (2014) 2955–2962. doi:10.1021/ac403223f.

[64] A. Bakiri, J. Hubert, R. Reynaud, S. Lanthony, D. Harakat, J.-H. Renault, J.-M. Nuzillard, Computer-Aided ^{13} C NMR Chemical Profiling of Crude Natural Extracts without Fractionation, J. Nat. Prod. 80 (2017) 1387–1396. doi:10.1021/acs.jnatprod.6b01063.

[65] C.A. Smith, E.J. Want, G. O'Maille, R. Abagyan, G. Siuzdak, XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification, Anal. Chem. 78 (2006) 779–787. doi:10.1021/ac051437y.

[66] T. Pluskal, S. Castillo, A. Villar-Briones, M. Orešič, MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data, BMC Bioinformatics. 11 (2010) 395.

[67] H.. Röst, T. Sachsenberg, S. Aiche, C. Bielow, H. Weisser, F. Aicheler, S. Andreotti, H.-C. Ehrlich, P. Gutenbrunner, E. Kenar, X. Liang, S. Nahnsen, L. Nilse, J. Pfeuffer, G. Rosenberger, M. Rurik, U. Schmitt, J. Veit, M. Walzer, D. Wojnar, W.E. Wolski, O. Schilling, J.S. Choudhary, L. Malmström, R. Aebersold, K. Reinert, O. Kohlbacher, OpenMS: A flexible open-source software platform for mass spectrometry data analysis, Nat. Methods. 13 (2016). doi:10.1038/nmeth.3959.

[68] A. Lommen, MetAlign: Interface-Driven, Versatile Metabolomics Tool for Hyphenated Full-Scan Mass Spectrometry Data Preprocessing, Anal. Chem. 81 (2009) 3079–3086. doi:10.1021/ac900036d.

[69] F. Giacomoni, G. Le Corguille, M. Monsoor, M. Landi, P. Pericard, M. Petera, C. Duperier, M. Tremblay-Franco, J.-F. Martin, D. Jacob, S. Goulitquer, E.A. Thevenot, C. Caron, Workflow4Metabolomics: a collaborative research infrastructure for computational metabolomics, Bioinformatics. 31 (2015) 1493–1495. doi:10.1093/bioinformatics/btu813.

[70] T. Kind, O. Fiehn, Metabolomic database annotations via query of elemental compositions : mass accuracy is insufficient even at less than 1 ppm, BMC Bioinformatics. 7 (2006) 234.

[71] X. Domingo-Almenara, J.R. Montenegro-Burke, H.P. Benton, G. Siuzdak, Annotation: A Computational Solution for Streamlining Metabolomics Analysis, Anal. Chem. 90 (2018) 480–489. doi:10.1021/acs.analchem.7b03929.

[72] N.G. Mahieu, J.L. Spalding, S.J. Gelman, G.J. Patti, Defining and Detecting Complex Peak Relationships in Mass Spectral Data: The Mz.unity Algorithm, Anal. Chem. 88 (2016) 9037–9046. doi:10.1021/acs.analchem.6b01702.

[73] C. Jaeger, M. Méret, C.A. Schmitt, J. Lisec, Compound annotation in liquid chromatography/high-resolution mass spectrometry based metabolomics: robust adduct ion determination as a prerequisite to structure prediction in electrospray ionization mass spectra, Rapid Commun. Mass Spectrom. 31 (2017) 1261–1266. doi:10.1002/rcm.7905.

[74] R.A. Scheltema, A. Jankevics, R.C. Jansen, M.A. Swertz, R. Breitling, PeakML/mzMatch: A File Format, Java Library, R Library, and Tool-Chain for Mass Spectrometry Data Analysis, Anal. Chem. 83 (2011) 2786–2793. doi:10.1021/ac2000994.

[75] C. Kuhl, R. Tautenhahn, C. Böttcher, T.R. Larson, S. Neumann, CAMERA: An Integrated Strategy for Compound Spectra Extraction and Annotation of Liquid Chromatography/Mass Spectrometry Data Sets, Anal. Chem. 84 (2012) 283–289. doi:10.1021/ac202450g.

[76] T.S. Lee, Y.S. Ho, H.C. Yeo, J.P.Y. Lin, D.-Y. Lee, Precursor mass prediction by clustering ionization products in LC-MS-based metabolomics, Metabolomics. 9 (2013) 1301–1310. doi:10.1007/s11306-013-0539-4.

[77] R. Daly, S. Rogers, J. Wandy, A. Jankevics, K.E. Burgess, R. Breitling, MetAssign: probabilistic annotation of metabolites from LC–MS data using a Bayesian clustering approach, Bioinformatics. 30 (2014) 2764–2771.

[78] D.P. Demarque, A.E.M. Crotti, R. Vessecchi, J.L.C. Lopes, N.P. Lopes, Fragmentation reactions using electrospray ionization mass spectrometry: an important tool for the structural elucidation and characterization of synthetic and natural products, Nat Prod Rep. 33 (2016) 432–455. doi:10.1039/C5NP00073D.

[79] B.B. Misra, New tools and resources in metabolomics: 2016-2017, Electrophoresis. (2018). doi:10.1002/elps.201700441.

[80] S. Wolf, S. Schmidt, M. Müller-Hannemann, S. Neumann, In silico fragmentation for computer assisted identification of metabolite mass spectra, BMC Bioinformatics. 11 (2010) 148.

[81] C. Ruttkies, E.L. Schymanski, S. Wolf, J. Hollender, S. Neumann, MetFrag relaunched: incorporating strategies beyond in silico fragmentation, J. Cheminformatics. 8 (2016). doi:10.1186/s13321-016-0115-9.

[82] M. Gerlich, S. Neumann, MetFusion: integration of compound identification strategies: MetFusion, J. Mass Spectrom. 48 (2013) 291–298. doi:10.1002/jms.3123.

[83] H.P. Benton, D.M. Wong, S.A. Trauger, G. Siuzdak, XCMS ^{2} : Processing Tandem Mass Spectrometry Data for Metabolite Identification and Structural Characterization, Anal. Chem. 80 (2008) 6382–6389. doi:10.1021/ac800795f.

[84] Y. Wang, G. Kora, B.P. Bowen, C. Pan, MIDAS: A Database-Searching Algorithm for Metabolite Identification in Metabolomics, Anal. Chem. 86 (2014) 9496–9503. doi:10.1021/ac5014783.

[85] Y. Wang, X. Wang, X. Zeng, MIDAS-G: a computational platform for investigating fragmentation rules of tandem mass spectrometry in metabolomics, Metabolomics. 13 (2017). doi:10.1007/s11306-017-1258-z.

[86] HighChem Ltd, Mass Frontier, Thermo Scientific, 2016.

[87] J. Zhou, R.J.M. Weber, J.W. Allwood, R. Mistrik, Z. Zhu, Z. Ji, S. Chen, W.B. Dunn, S. He, M.R. Viant, HAMMER: automated operation of mass frontier to construct in silico mass spectral fragmentation libraries, Bioinformatics. 30 (2014) 581–583. doi:10.1093/bioinformatics/btt711.

[88] F. Allen, R. Greiner, D. Wishart, Competitive fragmentation modeling of ESI-MS/MS spectra for putative metabolite identification, Metabolomics. 11 (2015) 98–110. doi:10.1007/s11306-014-0676-4.

[89] A. Aguilar-Mogas, M. Sales-Pardo, M. Navarro, R. Guimerà, O. Yanes, iMet: A Network-Based Computational Tool To Assist in the Annotation of Metabolites from Tandem Mass Spectra, Anal. Chem. 89 (2017) 3474–3482. doi:10.1021/acs.analchem.6b04512.

[90] F. Hufsky, S. Böcker, Mining molecular structure databases: Identification of small molecules based on fragmentation mass spectrometry data: Mining molecular structure databases, Mass Spectrom. Rev. 36 (2017) 624–633. doi:10.1002/mas.21489.

[91] C. Guijas, J.R. Montenegro-Burke, X. Domingo-Almenara, A. Palermo, B. Warth, G. Hermann, G. Koellensperger, T. Huan, W. Uritboonthai, A.E. Aisporna, D.W. Wolan, M.E. Spilker, H.P. Benton, G. Siuzdak, METLIN: A Technology Platform for Identifying Knowns and Unknowns, Anal. Chem. (2018). doi:10.1021/acs.analchem.7b04424.

[92] F. Hufsky, K. Scheubert, S. Böcker, New kids on the block: novel informatics methods for natural product discovery, Nat. Prod. Rep. 31 (2014) 807. doi:10.1039/c3np70101h.

[93] L.-F. Nothias-Scaglia, M. Esposito, J. Costa, J. Paolini, D. Touboul, M. Litaudon, Les réseaux moléculaires, une approche bio-informatique globale pour interpréter les données de spectrométrie de masse tandem, Spectra Anal. 307 (2015) 73–78.

[94] F. Olivon, G. Grelier, F. Roussi, M. Litaudon, D. Touboul, MZmine 2 Data-Preprocessing To Enhance Molecular Networking Reliability, Anal. Chem. 89 (2017) 7836–7840. doi:10.1021/acs.analchem.7b01563.

[95] J. Bruneton, Pharmacognosie, phytochimie, plantes médicinales, 4ième, Editions Lavoisier, 2009.

[96] M. Zhang, J. Sun, P. Chen, FlavonQ: An Automated Data Processing Tool for Profiling Flavone and Flavonol Glycosides with Ultra-High-Performance Liquid Chromatography–Diode Array Detection–High Resolution Accurate Mass–Mass Spectrometry, Anal. Chem. 87 (2015) 9974–9981. doi:10.1021/acs.analchem.5b02624.

[97] P. Kachlicki, J. Einhorn, D. Muth, L. Kerhoas, M. Stobiecki, Evaluation of glycosylation and malonylation patterns in flavonoid glycosides during LC/MS/MS metabolite profiling, J. Mass Spectrom. 43 (2008) 572–586.

[98] W.-Z. Yang, T. Bo, S. Ji, X. Qiao, D.-A. Guo, M. Ye, Rapid chemical profiling of saponins in the flower buds of Panax notoginseng by integrating MCI gel column chromatography and liquid chromatography/mass spectrometry analysis, Food Chem. 139 (2013) 762–769. doi:10.1016/j.foodchem.2013.01.051.

[99] W. Yang, Y. Li, C. Kang, H. Zhao, L. Xiang, C. Li, Q. Wang, Sodiation-based in-source collision for profiling of pyranocoumarins in Radix Peucedani (Qianhu): utility of sodium adducts' stability with in-source collision: Sodiation-based in-source collision, J. Mass Spectrom. 52 (2017) 152–164. doi:10.1002/jms.3913.

[100] X.-F. Li, M. Ma, K. Scherban, Y.K. Tam, Liquid chromatography-electrospray mass spectrometric studies of ginkgolides and bilobalide using simultaneous monitoring of proton, ammonium and sodium adducts, Analyst. 127 (2002) 641–646. doi:10.1039/B200849A.

[101] A.S.P. Moreira, M.A. Coimbra, F.M. Nunes, J. Simões, M.R.M. Domingues, Evaluation of the Effect of Roasting on the Structure of Coffee Galactomannans Using Model Oligosaccharides, J. Agric. Food Chem. 59 (2011) 10078–10087. doi:10.1021/jf2021072.

[102] H.T. Pham, A.T. Maccarone, M.C. Thomas, J.L. Campbell, T.W. Mitchell, S.J. Blanksby, Structural characterization of glycerophospholipids by combinations of ozone- and collision-induced dissociation mass spectrometry: the next step towards "top-down" lipidomics, Analyst. 139 (2013) 204–214. doi:10.1039/C3AN01712E.

[103] L. Abrankó, J.F. García-Reyes, A. Molina-Díaz, In-source fragmentation and accurate mass analysis of multiclass flavonoid conjugates by electrospray ionization time-of-flight mass spectrometry, J. Mass Spectrom. 46 (2011) 478–488. doi:10.1002/jms.1914.

[104] B. Abad-García, L.A. Berrueta, S. Garmón-Lobato, B. Gallo, F. Vicente, A general analytical strategy for the characterization of phenolic compounds in fruit juices by high-performance liquid chromatography with diode array detection coupled to electrospray ionization and triple quadrupole mass spectrometry, J. Chromatogr. A. 1216 (2009) 5398–5415. doi:10.1016/j.chroma.2009.05.039.

[105] K.A. Mortier, G.-F. Zhang, C.H. Peteghem, W.E. Lambert, Adduct formation in quantitative bioanalysis: Effect of ionization conditions on paclitaxel, J. Am. Soc. Mass Spectrom. 15 (2004) 585–592. doi:10.1016/j.jasms.2003.12.013.

[106] J.D. Chapman, D.R. Goodlett, C.D. Masselon, Multiplexed and data-independent tandem mass spectrometry for global proteome profiling: MULTIPLEXED AND DATA-INDEPENDENT TANDEM MS, Mass Spectrom. Rev. 33 (2014) 452–470. doi:10.1002/mas.21400.

[107] Inventaire National du Patrimoine Naturel, (n.d.). inpn.mnhn.fr (accessed February 27, 2018).

[108] G. Fico, G. Rodondi, G. Flamini, D. Passarella, F. Tomé, Comparative phytochemical and morphological analyses of three Italian Primula species, Phytochemistry. 68 (2007) 1683–1691. doi:10.1016/j.phytochem.2007.04.019.

[109] A. Müller, M. Ganzera, H. Stuppner, Analysis of phenolic glycosides and saponins in Primula elatior and Primula veris (primula root) by liquid chromatography, evaporative light scattering detection and mass spectrometry, J. Chromatogr. A. 1112 (2006) 218–223. doi:10.1016/j.chroma.2005.10.067.

[110] ANSM, (n.d.). ansm.sante.fr (accessed March 14, 2018).

[111] Personal Care Products Council (USA), INCI Name Application System, (2018). https://inci.personalcarecouncil.org/inci-app/.

[112] C. Ngo Ngoc Dong, Les plantes provoquant des dermatites de contact chez l'Homme: intérêt en médecine vétérinaire, PhD Thesis, 2013.

[113] K. Aslam, I.A. Nawchoo, M.A. Bhat, A.H. Ganie, N. Aslam, Ethno-pharmacological review of genus Primula, Int. J. Adv. Res. 2 (2014) 29–34.

[114] V. Okršlar, I. Plaper, M. Kovač, A. Erjavec, T. Obermajer, A. Rebec, M. Ravnikar, J. Žel, Saponins in tissue culture of Primula veris L., Vitro Cell. Dev. Biol. - Plant. 43 (2007) 644–651. doi:10.1007/s11627-007-9072-3.

[115] L. Apel, D. Kammerer, F. Stintzing, O. Spring, Comparative Metabolite Profiling of Triterpenoid Saponins and Flavonoids in Flower Color Mutations of Primula veris L., Int. J. Mol. Sci. 18 (2017) 153. doi:10.3390/ijms18010153.

[116] I. Çalis, A. Yürüker, H. Rüegger, A.D. Wright, O. Sticher, Triterpene saponins from Primula veris subsp. macrocalyx and Primula elatior subsp. meyeri, J. Nat. Prod. 55 (1992) 1299–1306.

[117] K. Kazuma, N. Noda, M. Suzuki, Malonylated flavonol glycosides from the petals of Clitoria ternatea, Phytochemistry. 62 (2003) 229–237. doi:10.1016/S0031-9422(02)00486-7.

[118] W.K. Andersen, A.A. Omar, S.B. Christensen, Isorhamnetin 3-(2, 6-dirhamnosylgalactoside)-7-rhamnoside and 3-(6-rhamnosylgalactoside)-7-rhamnoside from Rhazya stricta, Phytochemistry. 26 (1987) 291–294.

[119] C. Karl, G. Müller, P.A. Pederson, Flavonoids in the flowers of Primula officinalis, Planta Med. 41 (1981) 96–99. doi:10.1055/s-2007-971683.

[120] M.-A. Beck, H. Häberlain, Flavonol glycosides from Eschscholtzia californica, Phytochemistry. 50 (1999) 329–332. doi:10.1016/S0031-9422(98)00503-2.

[121] E.H. Lee, H.J. Kim, Y.S. Song, C. Jin, K.-T. Lee, J. Cho, Y.S. Lee, Constituents of the stems and fruits of Opuntia ficus-indica var. saboten, Arch. Pharm. Res. 26 (2003) 1018–1023. doi:10.1007/BF02994752.

[122] Personal Care Products Council (USA), Cosmetics Info, (2018). www.cosmeticsinfo.org/ingredient/glycyrrhiza-glabra-licorice-root-extract.

[123] S. Ji, Z. Li, W. Song, Y. Wang, W. Liang, K. Li, S. Tang, Q. Wang, X. Qiao, D. Zhou, S. Yu, M. Ye, Bioactive Constituents of *Glycyrrhiza uralensis* (Licorice): Discovery of the Effective Components of a Traditional Herbal Medicine, J. Nat. Prod. 79 (2016) 281–292. doi:10.1021/acs.jnatprod.5b00877.

[124] X. Wang, H. Zhang, L. Chen, L. Shan, G. Fan, X. Gao, Liquorice, a unique "guide drug" of traditional Chinese medicine: A review of its role in drug interactions, J. Ethnopharmacol. 150 (2013) 781–790. doi:10.1016/j.jep.2013.09.055.

[125] A.I.R.N.A. Barros, F.M. Nunes, C. Barros, A.M.S. Silva, M.R.M. Domingues, Structural Characterization of Nitrated 2'-Hydroxychalcones by Electrospray Ionization Tandem Mass Spectrometry, Eur. J. Mass Spectrom. 15 (2009) 605–616. doi:10.1255/ejms.1020.

[126] J.C.J.M.D.S. Menezes, J.A.S. Cavaleiro, S.P. Kamat, C.M.R.F. Barros, M.R.M. Domingues, Electrospray tandem mass spectrometry analysis of methylenedioxy chalcones, flavanones and flavones: ESI-MS of methylenedioxy chalcones, flavanones and flavones, Rapid Commun. Mass Spectrom. 27 (2013) 1303–1310. doi:10.1002/rcm.6577.

[127] J. Zhang, J. Lu, Q. Zhang, L. Dai, Y. Liu, P. Tu, Y. Qiao, Simultaneous Screening and Identifying Four Categories of Particular Flavonoids in the Leaves of Murraya exotica L. by HPLC-DAD-ESI-MS-MS, J. Chromatogr. Sci. 52 (2014) 103–114. doi:10.1093/chromsci/bms253.

[128] Y. Tai, S. Pei, J. Wan, X. Cao, Y. Pan, Fragmentation study of protonated chalcones by atmospheric pressure chemical ionization and tandem mass spectrometry, Rapid Commun. Mass Spectrom. 20 (2006) 994–1000. doi:10.1002/rcm.2404.

[129] N. Gampe, A. Darcsi, S. Lohner, S. Béni, L. Kursinszki, Characterization and identification of isoflavonoid glycosides in the root of Spiny restharrow (Ononis spinosa L.) by HPLC-QTOF-MS, HPLC–MS/MS and NMR, J. Pharm. Biomed. Anal. 123 (2016) 74–81. doi:10.1016/j.jpba.2016.01.058.

[130] F. Kuhn, M. Oehme, F. Romero, E. Abou-Mansour, R. Tabacchi, Differentiation of isomeric flavone/isoflavone aglycones by MS2 ion trap mass spectrometry and a double neutral loss of CO, Rapid Commun. Mass Spectrom. 17 (2003) 1941–1949. doi:10.1002/rcm.1138.

[131] A. de Villiers, P. Venter, H. Pasch, Recent advances and trends in the liquid-chromatography–mass spectrometry analysis of flavonoids, J. Chromatogr. A. 1430 (2016) 16–78. doi:10.1016/j.chroma.2015.11.077.

[132] J. Kang, Larry Hick A., W.E. Price, A fragmentation study of isoflavones in negative electrospray ionization by MSn ion trap mass spectrometry and triple quadrupole mass spectrometry, Rapid Commun. Mass Spectrom. 21 (2007) 857–868.

[133] J. Vacek, B. Klejdus, L. Lojková, V. Kubán, Current trends in isolation, separation, determination and identification of isoflavones: A review, J. Sep. Sci. 31 (2008) 2054–2067. doi:10.1002/jssc.200700569.

[134] Y.L. Ma, Q.M. Li, H. Van den Heuvel, M. Claeys, Characterization of flavone and flavonol aglycones by collision-induced dissociation tandem mass spectrometry, Rapid Commun. Mass Spectrom. 11 (1997) 1357–1361.

[135] Tela Botanica, www.tela-botanica.org (accessed September 15, 2017).

[136] M. Wichtl, Herbal Drugs and Phytopharmaceuticals, 3rd ed., CRC Press, 2004.

[137] Conservation de la nature Canada, (n.d.). www.natureconservancy.ca (accessed September 15, 2017).

[138] Ontario's Invading Species Awareness Program, (n.d.). www.invadingspecies.com (accessed September 15, 2017).

[139] Invasive Species Specialist Group, Global invasive species database, (2008). www.issg.org/worst100_species.html (accessed September 15, 2017).

[140] C. Lavoie, Should we care about purple loosestrife? The history of an invasive plant in North America, Biol. Invasions. 12 (2010) 1967–1999. doi:10.1007/s10530-009-9600-7.

[141] X.-J. Ma, C.-R. Ji, Y.-M. Wang, G. Zhang, Y. Liu, New tannins from Lythrum salicaria L., J. Chin. Pharm. Sci. 5 (1996) 225.

[142] J.-P. Rauha, J.-L. Wolfender, H. Vuorela, Characterization of the polyphenolic composition of purple loosestrife (Lythrum salicaria), Z. Für Naturforschung C. 56c (2000) 13–20.

[143] H. Becker, J.M. Scher, J.-B. Speakman, J. Zapp, Bioactivity guided isolation of antimicrobial compounds from Lythrum salicaria, Fitoterapia. 76 (2005) 580–584. doi:10.1016/j.fitote.2005.04.011.

[144] J.P. Piwowarski, A.K. Kiss, C-glucosidic Ellagitannins from Lythri herba (European Pharmacopoeia): Chromatographic Profile and Structure Determination, Phytochem. Anal. 24 (2013) 336–348. doi:10.1002/pca.2415.

[145] S. Granica, J.P. Piwowarski, A.K. Kiss, Determination of C-glucosidic Ellagitannins in Lythri salicaria herba by Ultra-High Performance Liquid Chromatography Coupled with Charged Aerosol Detector: Method Development and Validation, Phytochem. Anal. 25 (2014) 201–206. doi:10.1002/pca.2492.

[146] J.P. Piwowarski, A.K. Kiss, Contribution of C-glucosidic ellagitannins to Lythrum salicaria L. influence on pro-inflammatory functions of human neutrophils, J. Nat. Med. 69 (2015) 100–110. doi:10.1007/s11418-014-0873-5.

[147] E. Petit, Réactivité de polyphénols du vin sous conditions oxydantes: hémisynthèse des mongolicaïnes, et d'adduits entre polyphénols et thiols odorants, Université Sciences et Technologies-Bordeaux I, 2013. https://tel.archives-ouvertes.fr/tel-01086493/ (accessed September 12, 2017).

[148] J.P. Piwowarski, A.K. Kiss, C-glucosidic Ellagitannins from Lythri herba (European Pharmacopoeia): Chromatographic Profile and Structure Determination, Phytochem. Anal. 24 (2013) 336–348. doi:10.1002/pca.2415.

[149] T. Bencsik, L. Bartho, V. Sandor, N. Papp, G. Horvath, Phytochemical evaluation of Lythrum salicaria extracts and their effects on guinea-pig ileum, Nat. Prod. Commun. 8 (2013) 1247–1250.

[150] T. Bencsik, Comparative Histological, Phytochemical, Microbiological and Pharmacological Characterization of Some Lythrum salicaria L. Populations, (2014). http://pea.lib.pte.hu/bitstream/handle/pea/16405/bencsik-timea-phd-2014.pdf?sequence=1&isAllowed=y (accessed August 9, 2017).

[151] Z. Tunalier, M. Koşar, E. Küpeli, İ. Çaliş, K.H.C. Başer, Antioxidant, anti-inflammatory, anti-nociceptive activities and composition of Lythrum salicaria L. extracts, J. Ethnopharmacol. 110 (2007) 539–547. doi:10.1016/j.jep.2006.10.024.

[152] G. Nonaka, I. Nishioka, Tannins and Related Compounds. XCVII. Structure revision of C-glycosidic ellagitannins, Castalagin, Vescalagin, Casuarinin and Stachyurin, and related hydrolyzable tannins, Chem. Pharm. Bull. (Tokyo). 38 (1990) 2151–2156.

[153] P.K. Mukherjee, N. Maity, N.K. Nema, B.K. Sarkar, Bioactive compounds from natural resources against skin aging, Phytomedicine. 19 (2011) 64–73. doi:10.1016/j.phymed.2011.10.003.

[154] Y. Brun, X.-P. Wang, J. Willemot, T. Sevenet, P. Demenge, Experimental study of antidiarrheal activity of Salicairine®, Fundam. Clin. Pharmacol. 12 (1998) 30–36. doi:10.1111/j.1472-8206.1998.tb00920.x.

[155] EurekaSante-Vidal, (n.d.). eurekasante.vidal.fr (accessed September 16, 2017).

[156] D. Vincent, G. Segonzac, Quelques données nouvelles sur la salicaire (Lythrum salicaria L.), Acta Phytother. 1 (1954) 1–12.

[157] I. Pawlaczyk, P. Capek, L. Czerchawski, J. Bijak, M. Lewik-Tsirigotis, A. Pliszczak-Król, R. Gancarz, An anticoagulant effect and chemical characterization of Lythrum salicaria L. glycoconjugates, Carbohydr. Polym. 86 (2011) 277–284. doi:10.1016/j.carbpol.2011.04.048.

[158] J.P. Piwowarski, S. Granica, A.K. Kiss, Lythrum salicaria L.—Underestimated medicinal plant from European traditional medicine. A review, J. Ethnopharmacol. 170 (2015) 226–250. doi:10.1016/j.jep.2015.05.017.

[159] E. Miraldi, S. Ferri, V. Mostaghimi, Botanical drugs and preparations in the traditional medicine of West Azerbaijan (Iran), J. Ethnopharmacol. 75 (2001) 77–87.

[160] R. Di Novella, N. Di Novella, L. De Martino, E. Mancini, V. De Feo, Traditional plant use in the National Park of Cilento and Vallo di Diano, Campania, Southern, Italy, J. Ethnopharmacol. 145 (2013) 328–342. doi:10.1016/j.jep.2012.10.065.

[161] J.P. Piwowarski, A.K. Kiss, M. Kozłowska-Wojciechowska, Anti-hyaluronidase and anti-elastase activity screening of tannin-rich plant materials used in traditional Polish medicine for external treatment of diseases with inflammatory background, J. Ethnopharmacol. 137 (2011) 937–941. doi:10.1016/j.jep.2011.05.039.

[162] C.L. Simpson, D.M. Patel, K.J. Green, Deconstructing the skin: cytoarchitectural determinants of epidermal morphogenesis, Nat. Rev. Mol. Cell Biol. 12 (2011) 565–580. doi:10.1038/nrm3175.

[163] E. Candi, R. Schmidt, G. Melino, The cornified envelope: a model of cell death in the skin, Nat. Rev. Mol. Cell Biol. 6 (2005) 328–340. doi:10.1038/nrm1619.

[164] W.M. Holleran, Y. Takagi, G.K. Menon, G. Legler, K.R. Feingold, P.M. Elias, Processing of epidermal glucosylceramides is required for optimal mammalian cutaneous permeability barrier function., J. Clin. Invest. 91 (1993) 1656–1664. doi:10.1172/JCI116374.

[165] A.V. Rawlings, Trends in stratum corneum research and the management of dry skin conditions, Int. J. Cosmet. Sci. 25 (2003) 63–95.

[166] E. Fuchs, S. Raghavan, Getting under the skin of epidermal morphogenesis, Nat. Rev. Genet. 3 (2002) 199–209. doi:10.1038/nrg758.

[167] B. Marte, J. Finkelstein, L. Anson, Skin Biology, Nature. 445 (2007) 833–833. doi:10.1038/445833a.

[168] M. Rinnerthaler, M.K. Streubel, J. Bischof, K. Richter, Skin aging, gene expression and calcium, Exp. Gerontol. 68 (2015) 59–65. doi:10.1016/j.exger.2014.09.015.

[169] W.M. Holleran, Y. Takagi, G.K. Menon, G. Legler, K.R. Feingold, P.M. Elias, Processing of epidermal glucosylceramides is required for optimal mammalian cutaneous permeability barrier function., J. Clin. Invest. 91 (1993) 1656–1664. doi:10.1172/JCI116374.

[170] M. Monsigny, C. Petit, A.-C. Roche, Colorimetric determination of neutral sugars by a resorcinol sulfuric acid micromethod, Anal. Biochem. 175 (1988) 525–530. doi:10.1016/0003-2697(88)90578-7.

[171] S.S. Nielsen, Phenol-Sulfuric Acid Method for Total Carbohydrates, in: Food Anal. Lab. Man., Springer, Boston, MA, 2010: pp. 47–53. doi:10.1007/978-1-4419-1463-7_6.

[172] C.S.G. Kitzberger, A. Smânia, R.C. Pedrosa, S.R.S. Ferreira, Antioxidant and antimicrobial activities of shiitake (Lentinula edodes) extracts obtained by organic solvents and supercritical fluids, J. Food Eng. 80 (2007) 631–638. doi:10.1016/j.jfoodeng.2006.06.013.

[173] T. Tanaka, I. Kouno, Four new C-glycosidic ellagitannins, Castacrenins D-G, from Japanese Chesnut Wood (Castanea crenata SIEB. et ZUCC.), Chem. Pharm. Bull. (Tokyo). 45 (1997) 1751–1755.

[174] H. Rabeony, I. Petit-Paris, J. Garnier, C. Barrault, N. Pedretti, K. Guilloteau, J.-F. Jegou, G. Guillet, V. Huguier, J.-C. Lecron, F.-X. Bernard, F. Morel, Inhibition of Keratinocyte Differentiation by the Synergistic Effect of IL-17A, IL-22, IL-1α, TNFα and Oncostatin M, PLoS ONE. 9 (2014) e101937. doi:10.1371/journal.pone.0101937.

[175] H. Guenou, X. Nissan, F. Larcher, J. Feteira, G. Lemaitre, M. Saidani, M. Del Rio, C.C. Barrault, F.-X. Bernard, M. Peschanski, others, Human embryonic stem-cell derivatives for full reconstruction of the pluristratified epidermis: a preclinical study, The Lancet. 374 (2009) 1745–1753.

[176] J. Vičanová, E. Boelsma, A.M. Mommaas, J.A. Kempenaar, B. Forslind, J. Pallon, T. Egelrud, H.K. Koerten, M. Ponec, Normalization of epidermal calcium distribution profile in reconstructed human epidermis is related to improvement of terminal differentiation and stratum corneum barrier formation, J. Invest. Dermatol. 111 (1998) 97–106.

[177] W.J. Lee, K.H. Park, H.W. Cha, M.Y. Sohn, K.D. Park, S.-J. Lee, D.W. Kim, The Expression of Involucrin, Loricrin, and Filaggrin in Cultured Sebocytes, Ann. Dermatol. 26 (2014) 134. doi:10.5021/ad.2014.26.1.134.

[178] E. Fuchs, Scratching the surface of skin development, Nature. 445 (2007) 834–842. doi:10.1038/nature05659.

[179] Y. Zhuang, J. Lyga, Inflammaging in Skin and other Tissues - The Roles of Complement System and Macrophage, Inflamm. Allergy - Drug Targets. 13 (2014) 153–161. doi:10.2174/1871528113666140522112003.

[180] N. Philips, S. Auler, R. Hugo, S. Gonzalez, Beneficial Regulation of Matrix Metalloproteinases for Skin Health, Enzyme Res. 2011 (2011) 1–4. doi:10.4061/2011/427285.

ANNEXES

Annexe 1 ..... SOLVANTS, REACTIFS ET STANDARDS

1.1.1 Solvants

A partir de l'eau courante, nous avons deux étapes successives de purification : la première est gérée par un purificateur d'eau ELGA de modèle DV-35 (La marque ELGA appartient à Veolia) qui produit une eau de qualité supérieure à celle de l'eau déminéralisée avec une résistivité de 10 à 15 MΩ.cm. La seconde étape consiste en l'envoi de cette eau pré-purifiée sur un purificateur Milli-Q Gradient A10 (marque Millipore groupe Merck) muni de résines échangeuses d'ions qui fournit de l'eau pure à 18MΩ.cm

Pour les extractions, l'eau utilisée est celle issue de la première étape de purification. Pour toutes les préparations d'échantillons et les analyses (phases mobiles), l'eau utilisée est celle de plus grande pureté.

L'éthanol absolu utilisé est de qualité analytique et fourni par VWR. Le méthanol utilisé est de qualité LC-MS et provient de chez Honeywell. Le DMSO provient de chez Honeywell.

L'acétone deutérée (Acétone-d_6) et l'eau deutérée (D_2O) utilisées en RMN sont pures à plus de 99,9 % et proviennent de chez Sigma-Aldrich* (* groupe Merck).

1.1.2 Réactifs et standards

Réactifs	Fournisseur	Réactifs	Fournisseur
Acide chlorhydrique	Sigma Aldrich*	**Charbon actif**	Sigma Aldrich*
Acide formique	Honeywell	**KH_2PO_4**	Sigma Aldrich*
Acide phosphorique	Sigma Aldrich*	**Réactif de Folin-Denis**	Sigma Aldrich*
Acide sulfurique	Honeywell	**Résorcinol**	Acros Organics (Thermo Fisher Scientific)
BHA (butylated hydroxyanisole)	Sigma Aldrich*	**Vitamine C**	Sigma Aldrich*
Carbonate de sodium	Fisher scientific (Thermo Fisher Scientific)	**Soude**	Thermo-Fisher Scientific
Chlorure de calcium	VWR		

Standards	Fournisseur	Standards	Fournisseur
Apigénine	Sigma Aldrich*	**Lutéoline**	TCI Europe
Acide chlorogénique	Sigma Aldrich*	**Mannose**	Sigma-Aldrich*
Acide gallique	Sigma Aldrich*	**Myricétine**	TCI Europe
Arabinose	Sigma Aldrich*	**Quercétine**	Sigma Aldrich*
Fucose	Sigma Aldrich*	**Rhamnose**	Sigma Aldrich*
Galactose	Sigma Aldrich*	**Rutine**	Sigma Aldrich*
Glucose	Sigma Aldrich*	**Vescalagine**	Sigma Aldrich*
Isorhamnétine	Biopurify	**Xylose**	Sigma Aldrich*
Kaempférol	TCI Europe		

Annexe 2 ..... MISE AU POINT DE L'OUTIL GAINS

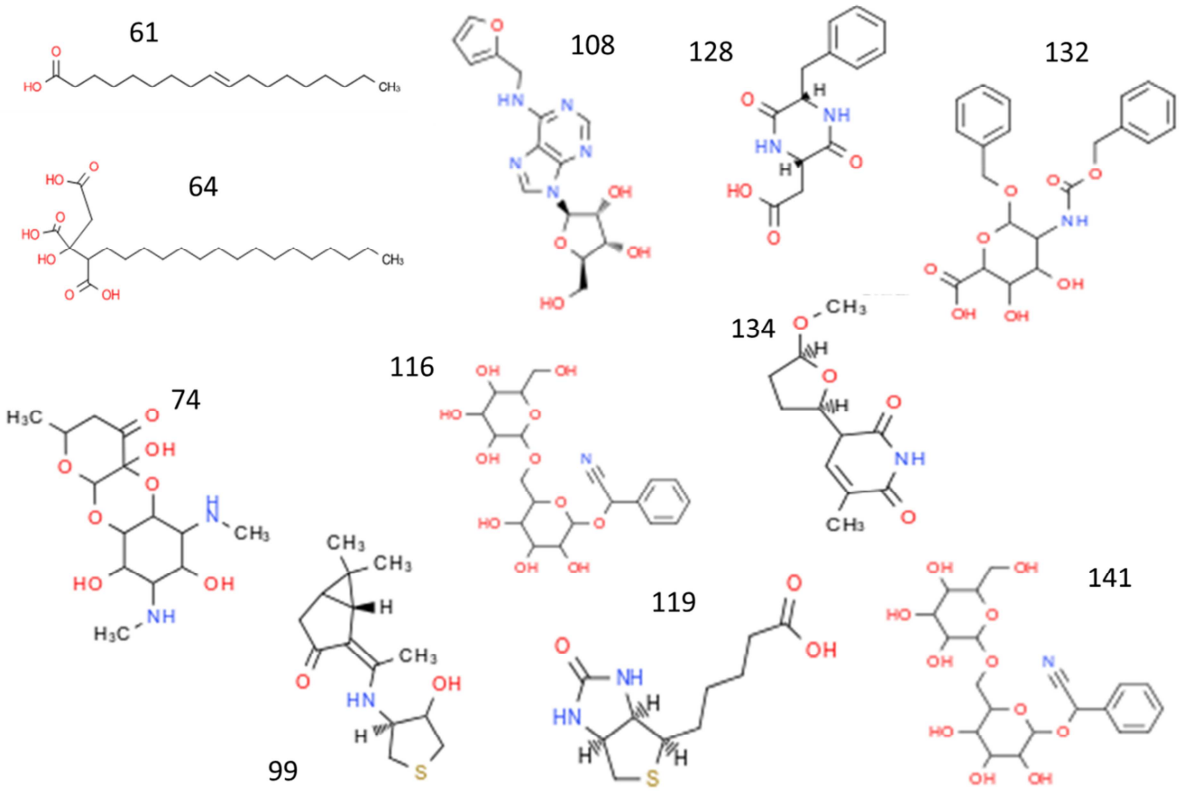
ANNEXE 2.1 VALEURS DES ADDUITS, CLUSTERS ET PERTES DE NEUTRES UTILISEES PAR L'ALGORITHME DE GAINS

Types d'ions	Mode positif	Mode négatif
Ions pseudo-moléculaires	H^+ = + 1,007276	$-H^+$ = - 1,007276
Adduits	Na^+ = + 22,989221 NH_4^+= + 18,033825 K^+= + 38,963158 Li^+ = + 7,015455	$HCOO^-$ = + 44,998204 Cl^- = + 34,969402 (Na^+-2H^+) = + 20,974669
Clusters	ACN = + 41,026549 MeOH = + 32,026215 HCOOH = + 46,005480 H_2O= + 18,010565	
Pertes de neutres	$(+H^+ - H_2O)$ = - 17,003289 $(+H^+ - 2H_2O)$ = -35,013305 CO_2 = - 43,989830	CO_2 = - 43,989830

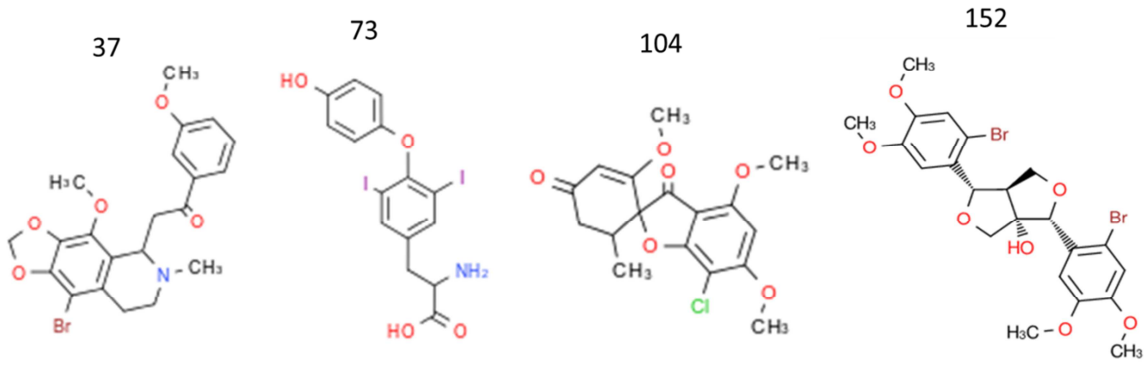
ANNEXE 2.2 STRUCTURES MOLECULAIRES DES STANDARDS

2.2.1 Structures du mélange standard utilisé pour valider l'outil GAINS

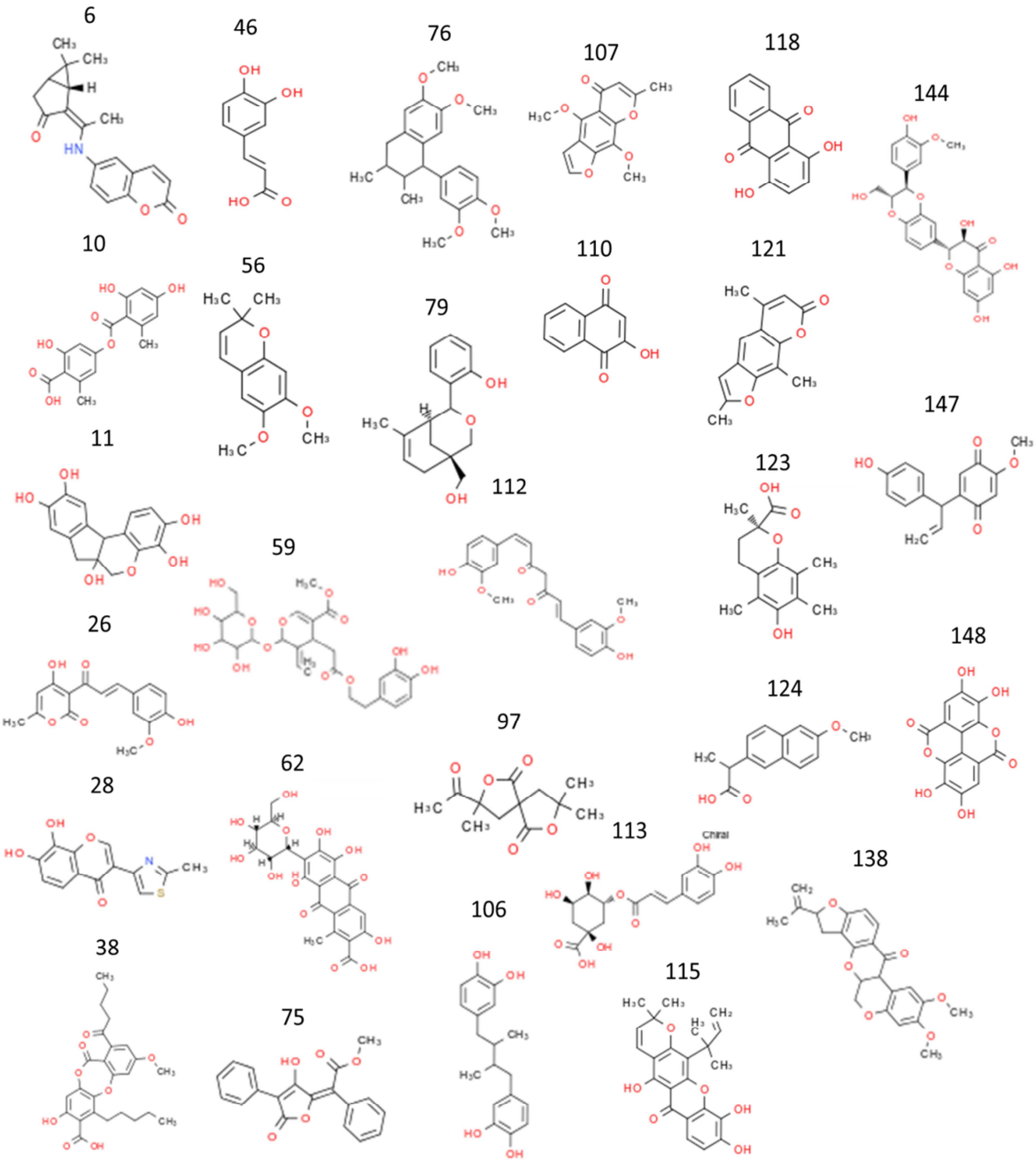
Composés du métabolisme primaire



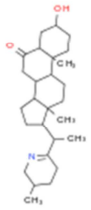
Composés halogénés

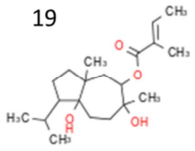


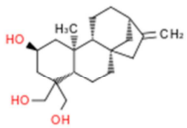
Composés phénoliques

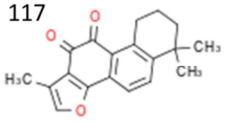


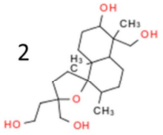
Terpènes et stéroïdes

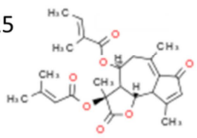
1


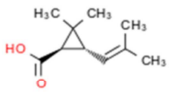
19


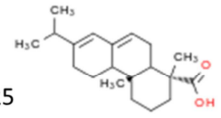
55


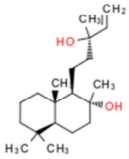
117


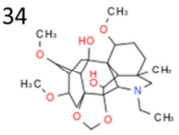
2


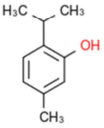
25


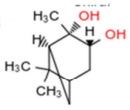
58


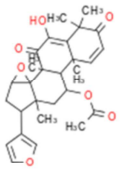
125


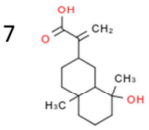
4


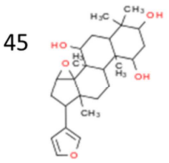
34


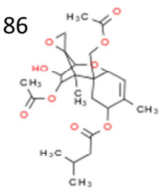
60


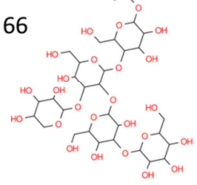
126


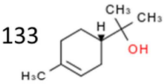
130


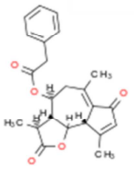
7


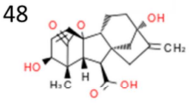
45


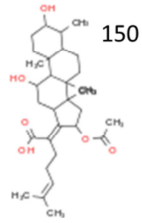
86


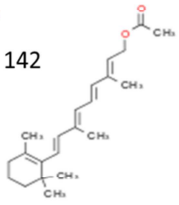
66


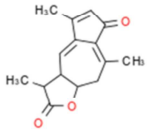
133


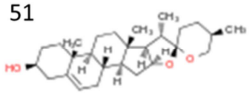
8


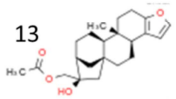
48


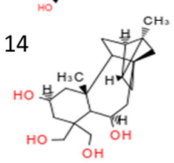
150


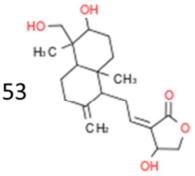
142


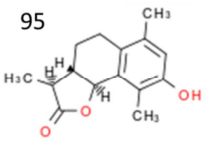
90


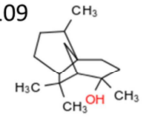
51


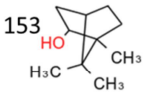
13


14


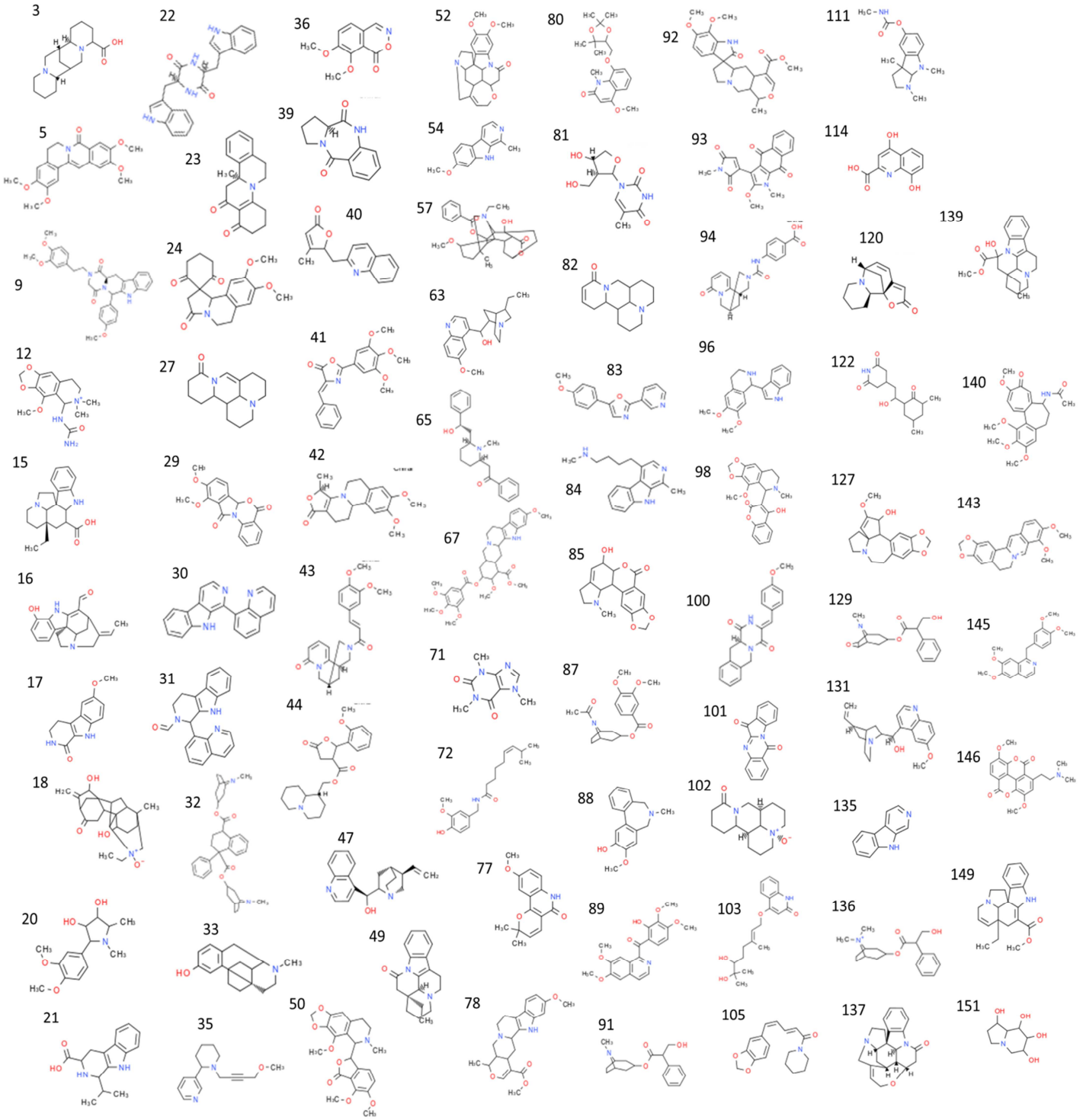
53


95


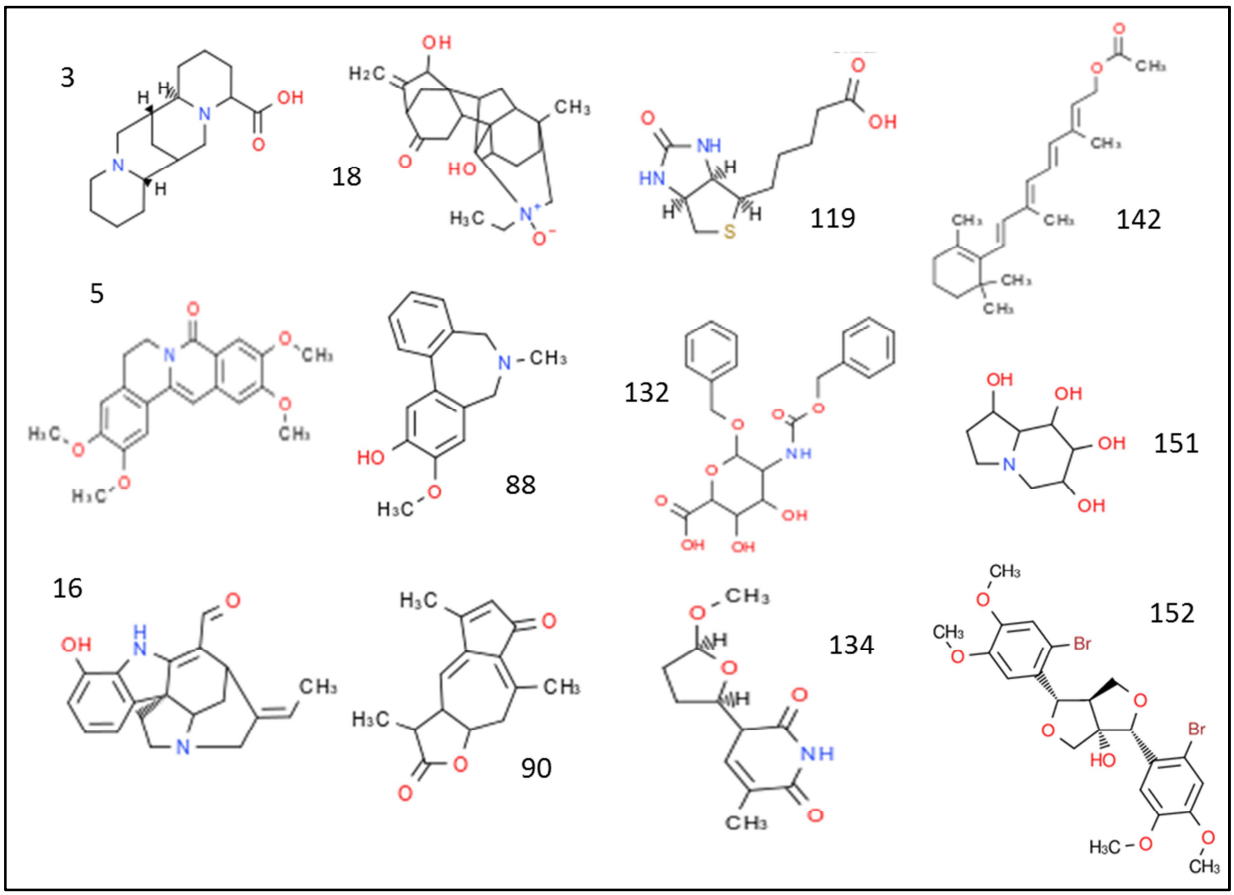
109


153


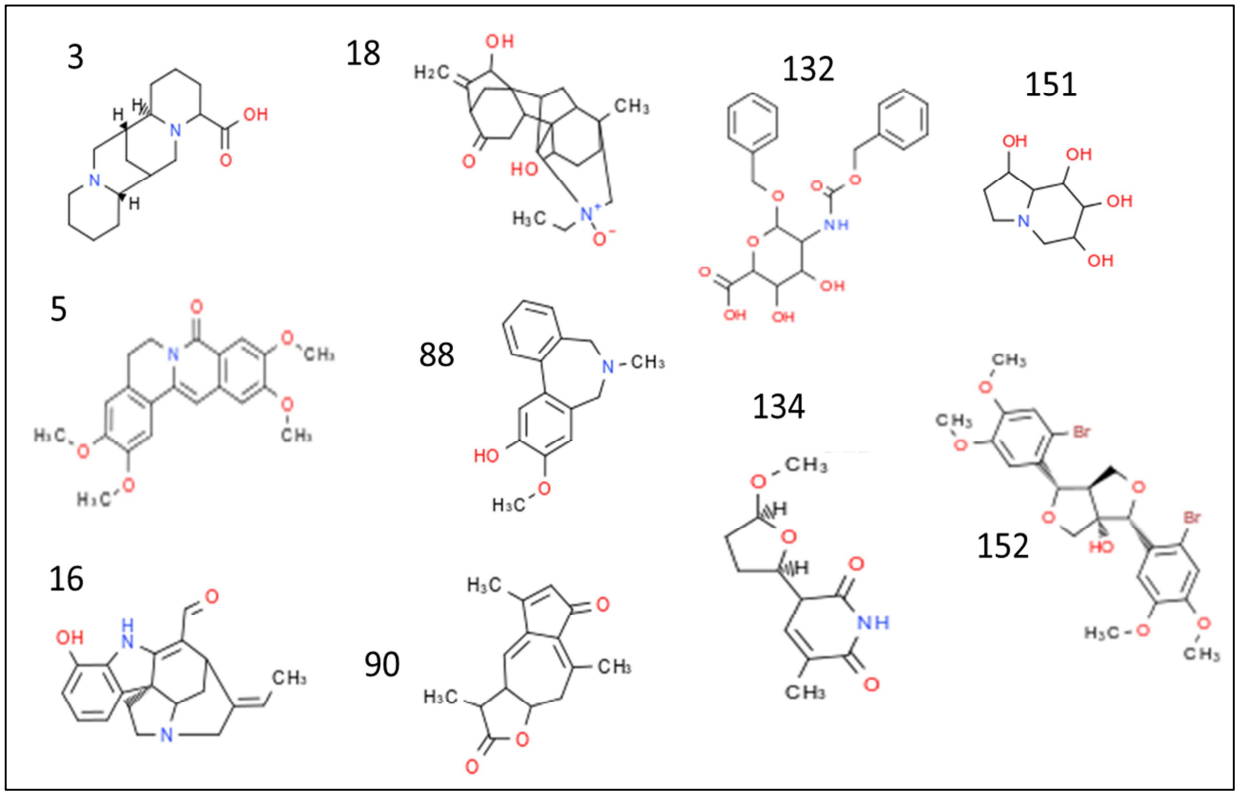
Alcaloïdes


3
22
36
52
80
92
111
5
39
54
81
93
114
23
57
40
94
120
139
9
24
63
82
12
27
41
83
96
122
140
65
15
29
42
84
98
127
143
67
16
30
43
85
100
129
145
17
31
44
71
87
101
131
146
18
32
72
88
102
135
47
77
89
103
136
149
20
33
49
50
78
91
105
137
151
21
35

2.2.2 Structures non détectées lors du 1er réplicat



2.2.3 Structures non détectées lors du 2nd réplicat



ANNEXE 2.3 MODULE METFRAG : DETAILS DES LIGNES DE COMMANDE UTILISEES

Metfrag est intégré à notre processus automatique GAINS directement dans le *script* sous forme de lignes de commande qui permettent de le faire fonctionner. Elles gèrent l'exécution de l'algorithme de Metfrag et son paramétrage. Les lignes de commande utilisées et leur réglage (équivalent à un mode d'utilisation) sont listés dans le tableau ci-dessous. Ces lignes de commande correspondent au fonctionnement de Metfrag dans sa dernière version dont les améliorations ont été décrites en 2016 dans une nouvelle publication intitulée *Metfrag relaunched* [81]. L'outil a en effet été amélioré et de nouveaux paramètres expérimentaux ont été intégrés pour gérer l'identification des composés. De nouvelles composantes des scores (prise en compte du temps de rétention ou application de filtres tels que la présence ou absence d'éléments chimiques particuliers) sont donc disponibles.

Lignes de commandes	Réglages
DatabaseSearchRelativeMassDeviation	**Homogène** (égale) **à la valeur de tolérance de correspondance d'ions** choisie pour les calculs de masses
FragmentPeakMatchRelativeMassDeviation	**Homogène** (égale) **à la valeur de tolérance de correspondance d'ions** choisie pour les calculs de masses
IsPositiveIonMode	Défini la charge d'ionisation (**True = positive, False = negative**)
LocalDatabasePath	**LocalSDF** et indication du chemin d'accès
MaximumTreeDepth	Valeur par défaut : **2**
MetFragCandidateWriter	**csv**
MetFragDatabaseType	**LocalSDF**
NeutralPrecursorMass	Masse monoisotopique neutre du précurseur
MetFragScoreTypes	**FragmenterScore**
PeakListPath	Chemin d'accès au fichier contenant la liste des pics MS
PrecursorIonMode	Type de l'ion adduit : **1 = $[M+H]^+$, 0 = $[M]^{+-}$, 18 = $[M+NH_4]^+$, 23 = $[M+Na]^+$, 39 = $[M+K]^+$, -1 = $[M-H]^-$, 35 = $[M+Cl]^-$, 45 = $[M+HCOO]^-$, 59 = $[M+CH_3COO]^-$**

Annexe 3 ..... ANALYSES PAR SPECTROMETRIE DE MASSE

ANNEXE 3.1 APPAREILLAGE POUR LES ANALYSES EN HRMS

Une chaîne Ultimate 3000 RSLC (Dionex), équipée d'un passeur automatique, d'une pompe binaire, d'un détecteur à barrette de diodes (DAD) et d'un four à colonne a été utilisée pour réaliser le couplage à l'HRMS. L'ensemble des analyses a été réalisé avec une source d'ionisation de type électrospray (ESI). Le spectromètre de masse haute résolution (HRMS) utilisé est un maXis Q-TOF (Bruker, Bremen, Allemagne) dont le schéma instrumental est présenté dans la figure ci-dessous. Les acquisitions ont été réalisées avec le logiciel Hystar (Bruker) et le retraitement des analyses a été effectué avec le logiciel *Data Analysis* version 4.4 (Bruker).

L'analyseur de géométrie Q-TOF (association d'un quadripôle avec un temps de vol) est présenté ci-dessous avec un schéma de l'instrument :

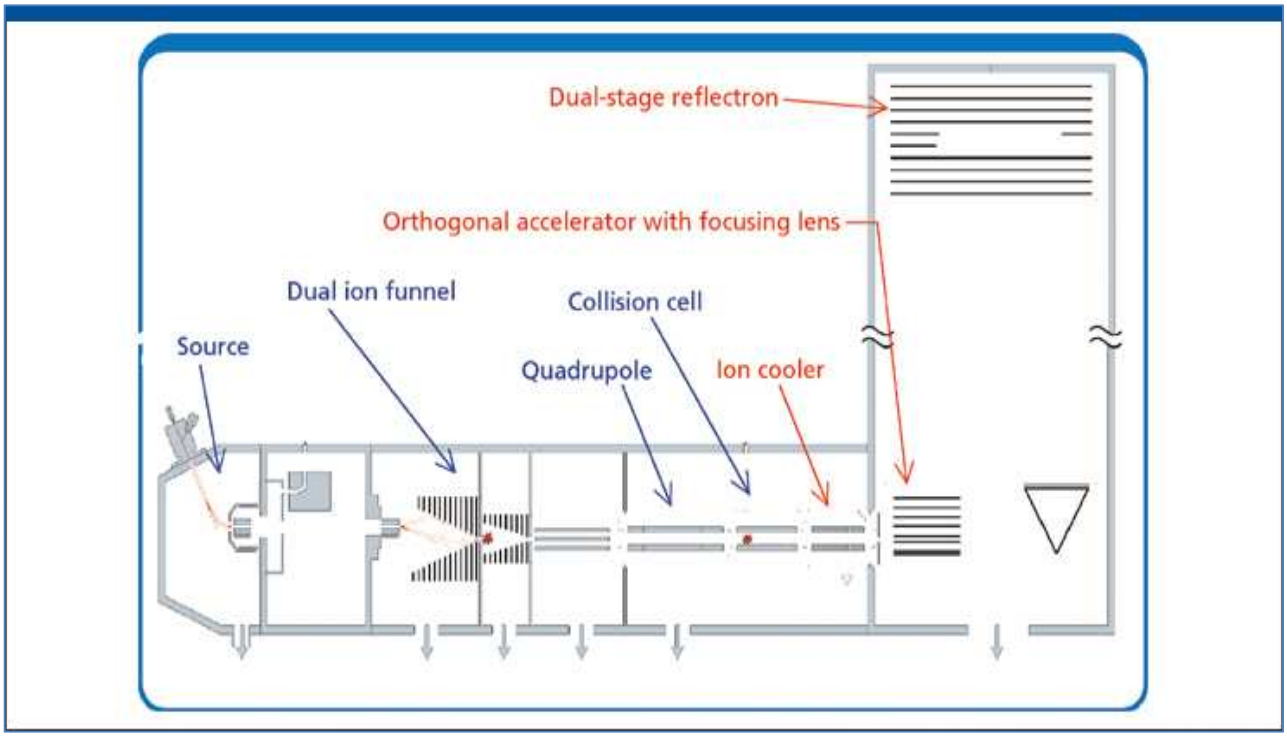


Schéma instrumental du spectromètre de masse maXis HR-Q-TOF (Bruker). Image provenant du site LCGC chromatography online, issue spéciale novembre 2008: *The Impact of Fast Chromatography and Its Effects on Mass Spectrometry*.

ANNEXE 3.2 CONDITIONS DE FIA-HRMS/MS : MISE AU POINT DE GAINS

Pour chacun des 150 composés :

Solution entre 1,5 et 2,5 µg.mL^{-1} dans MeOH
Injection de 0,2 µL en positif et 0,3 µL en négatif
Phase mobile : ACN à 0.1% HCOOH
Débit de la phase mobile : 0,2 mL.min^{-1}
Analyse en mode positif d'ionisation et analyse en mode négatif d'ionisation
Paramétrages de la source :
Gaz de séchage (N_2) : 7,0 L.min^{-1}
Température du gaz de séchage : 200°C
Gaz nébuliseur : 0,6 bar
Gamme de *m/z* : 50-1650

Méthode d'acquisition : DDA (*AutoMS/MS*)
Temps d'accumulation des ions : 1 seconde en MS et 1 seconde en MS/MS
Energie de fragmentation en positif et négatif : 25 eV
Prétraitement des données avec *Data Analysis* 4.4 : DISSECT (S/N = 25) ; SNAP
Importation des données de chaque composé dans des fichiers XML individuels
Fusion des fichiers XML en un unique fichier pour chaque mode d'ionisation
Traitement des données par le module de calculs GAINS : Tolérance à 10 ppm

Adduits, clusters ou pertes de neutres considérés pour les calculs	Mode positif	H^+ ; Na^+ ; NH_4^+ ; H_2O ; $2H_2O$
	Mode négatif	$-H^+$; $HCOO^-$; Cl^- ; (Na^+-2H^+)
	Modes positif et négatif	ACN ; HCOOH ; H_2O / CO_2

ANNEXE 3.3 CONDITIONS DE LC-HRMS : MISE AU POINT DE GAINS

Pour la solution des 150 composés en mélange :

Chaque composé en solution entre 3 et 10 µg.mL^{-1} dans MeOH
Injection de 1,5 µL en positif et 0,3 µL en négatif
Phases mobiles : HCOOH 0,1% et ACN à 0,1% HCOOH
Colonne : Acquity UPLC BEH Shield RP18 150 x 2,1 mm ; 1,7 µm (Waters)
Température de l'enceinte thermostatée : 35°C
Débit : 0,4 mL.min^{-1}
Gradient :

Temps (min)	0	16	20	32	35
HCOOH 0,1% (%)	97	60	60	5	5
ACN à 0,1% HCOOH (%)	3	40	40	95	95

Paramétrages de la source :
Gaz de séchage : 8,5 L.min^{-1}
Température du gaz de séchage : 200°C
Gaz nébuliseur : 2,0 bars en positif ; 1,2 bars en négatif
Gamme de *m/z* : 50-1650

Analyse en mode positif d'ionisation, analyse en mode négatif d'ionisation
Méthode d'acquisition : DDA (*AutoMS/MS)*

Energie de fragmentation : 25 eV en positif et 30 eV en négatif
Temps d'accumulation des ions : 0,4 seconde en MS et 2 x 0,6 secondes en MS/MS
Prétraitement des données avec *Data Analysis* 4.4 : DISSECT (S/N = 25) ; SNAP
Importation des données dans fichier XML
Traitement des données par le module de calculs GAINS
Tolérance à 5 ppm

Adduits, clusters ou pertes de neutres considérés pour les calculs	Mode positif	H^+ ; Na^+ ; NH_4^+ ; H_2O ; $2H_2O$
	Mode négatif	$-H^+$; $HCOO^-$; Cl^- ; (Na^+-2H^+)
	Modes positif et négatif	ACN ; HCOOH ; H_2O / CO_2

ANNEXE 3.4 CONDITIONS DE FIA-HRMS POUR LES MOLECULES PURIFIEES DE L'EXTRAIT DE PRIMEVERE VRAIE

Chaque molécule a été analysée en FIA-HRMS en mode positif et en mode négatif d'ionisation. Pour ces analyses, la phase mobile utilisée est : ACN 65 % à 0,1 % HCOOH. Le débit est de 0,2 $mL.min^{-1}$.

Les réglages de l'acquisition en MS sont les suivants :

Gaz de séchage (N_2) : 7,0 $L.min^{-1}$
Température du gaz de séchage : 200°C
Gaz nébuliseur : 0,6 bar
Gamme de *m/z* : 50-3000

Composé injecté	**Concentration** avant dilution par 10 pour injection en FIA ($mg.mL^{-1}$)	**Solvant de solubilisation**♦	**Solvant de dilution**	**Volume injecté** (µL)	
				positif	**négatif**
F2-02GJ3022	1,262	D_2O	H_2O	0,4	0,2
F3-01GJ3121	0,685	D_2O	H_2O	0,2	0,2
F3-02GJ3121	2,600	D_2O	H_2O	0,2	0,2
F3-03GJ3121	3,005	D_2O	H_2O	0,2	0,2
F5-02GJ3100	7,768	H_2O	H_2O	0,2	0,2
F6-05GJ3056	2,035	D_2O	H_2O	0,2	0,2
F8-01GJ3087	1,545	MeOH-d_4	MeOH	0,2	0,2

Pour les composés F3-02GJ3121, F3-03GJ3121 et F5-02GJ3100, des analyses supplémentaires ont été réalisées afin de déterminer si les oses de la molécule sont sous forme continue. Pour les deux premiers composés, une analyse de isCID MS (= *in source Collision Induced Dissociation*) a été réalisée en mode positif et pour le dernier composé, une analyse MS/MS manuelle a été effectuée. L'énergie de fragmentation

♦ Lorsque les échantillons sont solubilisés dans les solvants deutérés, cela correspond à la prépration de l'échantillon en vue de la RMN. Pour les analyses FIA-HRMS, un aliquot de la solution destinée aux analyses RMN est prélevé et dilué par 10 dans le solvant équivalent non deutéré.

appliquée est de : 8,0 eV pour F3-02GJ3100 et F3-03GJ3100 et de 50 eV pour F5-02GJ3100. Les réglages pour effectuer l'ionisation et l'acquisition sont équivalents à ceux utilisés pour la FIA-MS. Pour chaque composé, le volume injecté est de 0,2 µL et la concentration de 0,1 $mg.mL^{-1}$.

ANNEXE 3.5 CONDITIONS EN UHPLC-HRMS : EXTRAIT DE PRIMEVERE VRAIE

Une solution de l'extrait sec EtOH 60% de primevère vraie a été analysée en UHPLC-HRMS dans les conditions suivantes :

Colonne : Acquity UPLC BEH Shield RP18 (Waters) 150 x 2,1 mm ; 1,7 µm
Température de l'enceinte thermostatée : 30°C
Débit : 0,2 $mL.min^{-1}$
Volume injecté : 0,8 µL
Concentration de la solution d'extrait : 1 $mg.mL^{-1}$ solubilisé dans MeOH/H_2O 50/50
Phases mobiles : HCOOH 0,1% et MeOH à 0,1% HCOOH
Gradient :

Temps (min)	0	3	35	40	43
HCOOH 0,1 % (%)	90	90	30	30	90
MeOH à 0,1 % HCOOH (%)	10	10	70	70	10

Paramétrages de la source :

Gaz de séchage : 7,0 $L.min^{-1}$
Température du gaz de séchage : 200°C
Gaz nébuliseur : 1,0 bar

Gamme de scan des *m/z* : 50-1650
Analyse en mode positif d'ionisation, analyse en mode négatif d'ionisation
Méthode d'acquisition : DDA (*AutoMS/MS)*
Energie de fragmentation : 35 eV en positif et 35 eV en négatif
Temps d'accumulation des ions : 0,4 seconde en MS et 2 x 0,4 secondes en MS/MS

Prétraitement des données avec *Data Analysis* 4.4 : DISSECT (S/N = 35 pour ESI^+ ; 50 pour ESI^-) ; SNAP
Importation des données dans fichier XML
Traitement des données par le module de calculs GAINS
Tolérance à 5 ppm

Adduits, clusters ou pertes de neutres considérés pour les calculs	Mode positif	H^+ ; Na^+ ; NH_4^+ ; H_2O ; $2H_2O$
	Mode négatif	$-H^+$; $HCOO^-$; Cl^- ; (Na^+-2H^+)
	Modes positif et négatif	MeOH ; HCOOH ; H_2O / CO_2

ANNEXE 3.6 CONDITIONS EN UHPLC-HRMS : EXTRAIT DE TULIPE NOIRE

Une solution de l'extrait de Tulipe noire est analysé en chromatographie liquide en phase inverse. La chaîne HPLC Ultimate 3000 RSLC (Thermo Fisher Scientific) est équipée d'un passeur automatique, d'une pompe binaire et d'un four à colonne. Le signal est suivi avec un détecteur à barrette de diode (DAD-UV). Cet

ensemble est couplé à un spectromètre de masse haute résolution maXis Q-TOF (Bruker France, Champs-sur Marne, France) après ionisation en électrospray. Les conditions d'analyses sont indiquées ci-dessous :

Colonne WATERS Acquity BEH C18 –150 x 2,1mm – 1,7µm ;
Débit : 0,2 mL.min^{-1} ;
Température : 30°C ;
Concentration injectée : 1 mg.mL^{-1} solubilisé dans H_2O
Solvant d'injection : MeOH ;
Volume injecté : 0,75 µL (mode positif) et 0,25 µL (mode négatif) ;
Gradient :

Temps (min)	% (HCOOH 0,1 %)	% (MeOH à HCOOH 0,1 %)
0	90	10
3	90	10
30	20	80
35	20	80

Paramétrages de la source :

Gaz de séchage : 7,0 L.min^{-1}
Température du gaz de séchage : 200°C
Gaz nébuliseur : 1,0 bar

Gamme de scan des *m/z* : 50-1650
Analyse en mode positif d'ionisation, analyse en mode négatif d'ionisation
Méthodes d'acquisition : DDA (*AutoMS/MS)*
Energie de fragmentation : 35 eV en positif et 35 eV en négatif
Temps d'accumulation des ions : 0,4 seconde en MS et 2 x 0,4 secondes en MS/MS

Prétraitement des données avec *Data Analysis* 4.4 : DISSECT (S/N = 10 pour ESI$^+$; 10 pour ESI$^-$) ; SNAP
Importation des données dans fichier XML
Traitement des données par le module de calculs GAINS
Tolérance à 2 ppm

Adduits, clusters ou pertes de neutres considérés pour les calculs	Mode positif	H^+ ; Na^+ ; NH_4^+ ; H_2O ; $2H_2O$
	Mode négatif	$-H^+$; $HCOO^-$; (Na^+-2H^+)
	Modes positif et négatif	MeOH ; HCOOH ; H_2O / CO_2

ANNEXE 3.7 CONDITIONS EN UHPLC-HRMS : EXTRAIT DE REGLISSE GLABRE

L'extrait est analysé en chromatographie liquide en phase inverse et détecté par une barrette de diode DAD-UV ainsi que par un spectromètre de masse haute résolution (HRMS) après ionisation en électrospray. Les conditions HPLC sont les suivantes :

Colonne Phenomenex Kinetex C18 100A 100 x 4,6mm – 2,6µm ;
Débit : 1,0 mL.min^{-1} ;
Température : 30°C ;
Concentration injectée : 1 mg.mL^{-1} (extrait liquide de racines de réglisse à 0,2 % de matière sèche dilué par 2 dans l'eau avant injection)
Volume d'injection : 4 µL

Gradient d'élution :

Temps (min)	% (HCOOH 0,1 %)	% (ACN à HCOOH 0,1 %)
0	95	5
5	95	5
38	20	80
43	20	80

Les réglages du spectromètre de masse sont les suivants :

Un premier segment d'analyse n'est pas envoyé dans la source électrospray pour éviter de saturer la source avec la glycérine et la bétaïne présentes dans l'extrait du fait de l'extraction. Aussi de 0 à 3 minutes, il n'y a pas de données de spectrométrie de masse.

Paramétrages de la source :

Gaz de séchage : 9,0 $L.min^{-1}$
Température du gaz de séchage : 200°C
Gaz nébuliseur : 2,0 bars

Gamme de scan des *m/z* : 50-1650
Analyse en mode positif d'ionisation, analyse en mode négatif d'ionisation
Méthodes d'acquisition : mode DDA (*AutoMS/MS)*
Energie de fragmentation : 35 eV en positif et 35 eV en négatif
Temps d'accumulation des ions : 0,4 seconde en MS et 2 x 0,4 secondes en MS/MS

Prétraitement des données avec *Data Analysis* 4.4 : DISSECT (S/N = 40 pour ESI^+ ; 35 pour ESI^-) ; SNAP
Importation des données dans fichier XML
Traitement des données par le module de calculs GAINS
Tolérance à 10 ppm

Adduits, clusters ou pertes de neutres considérés pour les calculs	Mode positif	H^+ ; Na^+ ; NH_4^+ ; H_2O ; $2H_2O$
	Mode négatif	$-H^+$; $HCOO^-$; (Na^+-2H^+)
	Modes positif et négatif	ACN ; HCOOH ; H_2O / CO_2

ANNEXE 3.8 CONDITIONS EN HPLC-DAD-MS : EXTRAIT DE SALICAIRE

3.8.1 Appareillage

L'appareillage de la Chromatographie Liquide de Haute Performance (HPLC Thermo) est couplé avec un détecteur à barrettes de diodes (DAD) et une trappe d'ions LCQ Finnigan dotée d'une source électrospray (ESI) (Thermo Fisher Scientific, Courtaboeuf, France).

3.8.2 Conditions chromatographiques et détection DAD

HPLC : Finnigan MAT avec pompe P4000, Autosampler AS3000
Colonne : Synergi Polar -RP de 150 mm x 4.6 mm; 4µm (Phenomenex, Le Pecq, France)
Température : 30°C (Enceinte thermostatée GECKO 2000)

Débit : 1 $mL.min^{-1}$
Concentration : extrait de *Lythrum salicaria* à 1 $mg.mL^{-1}$ dans H_2O
Volume injecté : 20 µL
Phases mobiles : eau à 0,1% d'acide formique, méthanol à 0,1% d'acide formique
Gradient :

Temps (min)	0	5	30	42
HCOOH 0,1% (%)	100	100	30	30
MeOH à 0,1% HCOOH (%)	0	0	70	70

Détection : DAD-UV avec suivi à 254 nm (Détecteur UV6000LP)

Réglages pour la spectrométrie de masse

Tension de la source : 4.5kV en positif ; - 3.5kV en négatif
Gaz de nébulisation (*Sheath gas*) (N_2) : 80 (unités arbitraires)
Gaz auxiliaire (N_2) = 20 (unités arbitraires)
Température du capillaire : 200°C
Plage de masse : 50 à 2000 *m/z*
Mode d'acquisition du signal : *Full scan*

ANNEXE 3.9 CONDITIONS DE FIA-HRMS : MOLECULES PURIFIEES DE L'EXTRAIT DE SALICAIRE

Les analyses FIA-HRMS et MS/MS ont été réalisées en ESI-Q-TOF. Les composés isolés ont été solubilisés à 40 $\mu g.mL^{-1}$ dans le méthanol. Le volume d'injection a été de 0,5µL. Les analyses ont été effectuées en mode positif d'ionisation et en mode négatif. Au niveau de la source ESI, le débit du gaz de séchage est réglé à 7,0 $L.min^{-1}$ et celui du gaz nébuliseur à 0,6 bar. La température du gaz de séchage est de 200°C. La tension du capillaire est de 4,5 kV en positif et 4 kV en négatif.

Dans un premier temps, les ions observés sans application d'une énergie de collision ont été enregistrés sur une plage de *m/z* de 50 à 2500. Dans un second temps, l'ion monochargé le plus intense a été isolé avec une largeur de fenêtre d'isolement de 8 *m/z*. L'ion isolé est fragmenté par application de l'énergie de collision optimale déterminée manuellement (faible intensité de l'ion précurseur et intensité importante de plusieurs ions). Les énergies de fragmentation appliquées sont 25 eV en positif et 55 eV en négatif.

Les données pour la détermination des masses monoisotopiques des ions de MS et MS/MS ont été traitées avec le logiciel *Data Analysis* 4.4 et les formules brutes possibles ont été déterminées en utilisant l'algorithme *SmartFormula*.

Annexe 4 ..... ANALYSES PHYTOCHIMIQUES COMPLEMENTAIRES A LA SPECTROMETRIE DE MASSE. METHODES DE FRACTIONNEMENT ET D'ISOLEMENT DE COMPOSES

ANNEXE 4.1 DOSAGES DE FAMILLES PHYTOCHIMIQUES : EXTRAIT DE SALICAIRE

4.1.1 Dosages des oses neutres

L'extrait a été solubilisé à 1 $mg.mL^{-1}$ dans l'eau. Des solutions standards de glucose ont aussi été préparées entre 10 $\mu g.mL^{-1}$ et 200 $\mu g.mL^{-1}$. L'échantillon et les solutions standards ont ensuite été mis en contact avec le réactif résorcinol à 6 $mg.mL^{-1}$ et de l'acide sulfurique 75 %. Après agitation, les échantillons sont incubés 30 minutes à 90°C puis laissés à refroidir 30 minutes. Les mesures d'absorbance sont effectuées à 430 nm. La droite d'étalonnage obtenue avec la gamme de standards permet de déterminer la teneur en sucres de l'échantillon. Les mesures pour chaque échantillon ou solution standard ont été effectuées en duplicat (n=2).

4.1.2 Dosages des composés phénoliques

La gamme d'étalonnage a été faite en préparant 5 solutions de 20 $\mu g.mL^{-1}$ à 2 $mg.mL^{-1}$ d'acide chlorogénique. L'extrait a été solubilisé à 5 $mg.mL^{-1}$ dans l'eau. La réaction colorimétrique est permise par l'ajout du réactif de Folin-Denis avec une solution saturée de carbonate de sodium à l'échantillon. Après 30 minutes d'agitation, la mesure de l'absorbance s'effectue à 635 nm.

4.1.3 Dosages des composés majoritaires (vescalagine et castalagine)

La teneur en composés majoritaires identifiés dans l'extrait est quantifiée en HPLC-UV en utilisant la vescalagine (Sigma-Aldrich, Saint-Quentin Fallavier, France) comme standard. Les solutions étalons sont préparées dans un mélange eau/méthanol en proportion 50/50 (v:v). L'extrait est solubilisé dans le même solvant et centrifugé pendant 5 minutes à 18 000 g.

Les conditions chromatographiques, d'injection et de détection sont les suivantes :

HPLC : U3000 (Thermo Fisher Scientific, Courtaboeuf, France)
Colonne : Synergi Polar-RP de 150 mm x 4.6 mm; 4µm (Phenomenex, Le Pecq, France)
Température : 30°C
Débit : 1 $mL.min^{-1}$
Volume injecté : 20 µL
Phases mobiles : eau à 0,1 % d'acide formique, méthanol à 0,1 % d'acide formique
Gradient :

Temps (min)	0	5	22	23	33
HCOOH 0,1 % (%)	100	100	87	30	30
MeOH à 0,1 % HCOOH (%)	0	0	13	70	70

Détection : DAD-UV avec enregistrement à 235 nm

ANNEXE 4.2 TRAITEMENT DES EXTRAITS PAR EXTRACTION SUR PHASE SOLIDE (SPE)

4.2.1 Traitement SPE : extrait de primevère vraie

Le traitement par SPE a été optimisé sur une cartouche SPE Strata C18-E 500 mg/3 mL (Phenomenex, Le Pecq, France) avec 40 mg d'extrait en solution (concentration de 13,5 $mg.mL^{-1}$). On a déterminé le rendement du traitement SPE, c'est-à-dire la quantité de composés obtenus après traitement SPE par rapport à la quantité d'extrait traité, ce rendement est d'environ 20 %. Ensuite, le traitement optimisé a été réalisé sur une cartouche SPE Strata C18-E 20 g/60 mL. L'extrait sec (4,84 g) de primevère vraie est solubilisé dans 360 mL MeOH 30 %. La solution jaune marron obtenue est trouble ; elle est donc passée aux ultra-sons. On a effectué le dépôt de 120 mL de solution d'extrait à 40 $mg.mL^{-1}$ environ (correspondant à une quantité de 1,6 g d'extrait). Ainsi, la totalité de la solution a été traitée par SPE par le biais de trois répétitions successives du traitement SPE. Une étape de rinçage au MeOH a eu lieu à la fin de traitement d'un dépôt et avant de commencer le traitement du suivant.

Après l'étape de dépôt, l'élution est réalisée avec une solution d'un mélange H_2O/MeOH 70/30 (v:v). Après le prétraitement SPE, on concentre la fraction éluée par évaporation rotative du méthanol et d'une partie de l'eau, ceci a résulté sur un volume à injecter de 70 mL dont la concentration est estimée à 13,8 $mg.mL^{-1}$ (20 % de la quantité initiale d'extrait / volume final). L'échantillon est rouge orangé et n'a pas précipité.

4.2.2 Traitement SPE : extrait de salicaire

Le traitement par SPE a été optimisé sur une cartouche SPE Strata C18-E 500mg/3mL (Phenomenex, Le Pecq, France) avec 10 mg d'extrait en solution (10 $mg.mL^{-1}$). Ensuite, le traitement a été réalisé sur une cartouche SPE Strata C18-E 20g/60mL sur laquelle deux dépôts de 400 mg d'extrait en solution (10 $mg.mL^{-1}$ environ) dans un mélange eau/méthanol 80/20 (v:v) avec 0.5% d'acide formique successifs ont été effectués. Une étape de rinçage et séchage au MeOH a eu lieu entre les deux traitements. Le prétraitement par SPE a donc été fait sur 800 mg d'extrait au total. Après dépôt, l'élution est réalisée avec H_2O-MeOH 80:20 à 0,5 % HCOOH. La solution éluée a été concentrée par évaporation du méthanol et d'une partie de l'eau sous pression réduite. La concentration finale en matière sèche a été estimée à 61,5 $mg.mL^{-1}$.

ANNEXE 4.3 HPLC PREPARATIVE : FRACTIONNEMENT DE L'EXTRAIT DE PRIMEVERE

Afin de fractionner l'extrait de primevère vraie en plusieurs fractions, une solution concentrée de l'extrait après traitement SPE a été injectée en HPLC préparative. Pour ce faire, nune HPLC préparative (NOVAPREP 800) est utilisée avec une colonne (NOVAPREP) de 25 mm de diamètre interne remplie avec une phase stationnaire Kromasil 10 µm (Thermo Fisher Scientific, Courtabœuf, France) et un détecteur monochromatique (Varian, modèle 320 – Agilent Technologies, Santa Clara, Etats-Unis).

La phase mobile est un mélange H_2O/MeOH en proportion 62/38 % (v:v). La séparation HPLC est effectuée en mode isocratique et dure 50 minutes. Le volume d'injection est de 10 mL. On a donc réalisé 7 injections successives. Le suivi est fait en UV à 254 nm.

ANNEXE 4.4 HPLC SEMI-PREPARATIVE

Pour effectuer l'isolement des composés en HPLC à l'échelle semi-préparative, on dispose d'un système HPLC (Dionex 1, Thermo Fisher Scientific, Courtabœuf, France) munie d'une pompe binaire P680, d'un autosampler GINA 50. Une enceinte thermostatée IGLOO-CIL (Cluzeau Info Lab, Ste Foy-La-Grande, France) régule la température. Un détecteur DAD-UV PDA 100 enregistre le signal (Thermo Fisher Scientific).

4.4.1 Isolement des composés majoritaires à partir de fractions de l'extrait de primevère

Les conditions opératoires utilisées en HPLC semi-préparative pour séparer les composés de chaque fraction issue de l'extrait de *Primula veris* et obtenues après la purification par HPLC préparative sont tabulées ci-après :

Fraction	Condition d'élution	Colonne utilisée	Température colonne	Durée par analyse	Volume d'injection	Concentration
F1GJ2196	Isocratique : 28 % MeOH	Synergi Fusion	30°C	52 min	100 µL	50 $mg.mL^{-1}$
F2GJ2196	Isocratique : 28 % MeOH	Synergi Fusion	30°C	52 min	100 µL	50 $mg.mL^{-1}$
F3GJ2196	Isocratique : 13 % ACN	Kromasil C_{18}	30°C	55 min	50 µL	50 $mg.mL^{-1}$
F5GJ2196	Isocratique : 14 % ACN	Kromasil C_{18}	30°C	45 min	50 µL	20 $mg.mL^{-1}$
F6GJ2196	Isocratique : 16 % ACN	Kromasil C_{18}	30°C	27 min	100 µL	10 $mg.mL^{-1}$
F7GJ2196	Isocratique : 40 % MeOH	Kromasil C_{18}	25°C	30 min	50 µL	50 $mg.mL^{-1}$
F8GJ2196	Isocratique : 17 % ACN	Kromasil C_{18}	30°C	35 min	75 µL	20 $mg.mL^{-1}$
F9GJ2196	Isocratique : 17 % ACN	Kromasil C_{18}	30°C	45 min	75 µL	20 $mg.mL^{-1}$

La colonne Synergi FUSION est de la marque Phenomenex et la colonne Kromasil C_{18} provient de chez Thermo Fisher Scientific. Les deux colonnes ont une longueur de 250 mm et un diamètre interne de 10 mm. La porosité est de 4 µm pour la Synergi et de 5µm pour la Kromasil C_{18}. Le débit utilisé était de 4 $mL.min^{-1}$ pour l'ensemble des séparations effectuées.

Les solutions ont été injectées aux concentrations les plus importantes possibles en fonction de leur solubilité. Les volumes injectés ont été optimisés pour être les plus importants possibles : une optimisation du volume a été effectuée pour chaque solution de façon à éviter les effets de surcharge de la colonne. Le nombre d'injections a été fonction du volume total de solution à traiter et du volume d'injection par analyse. Pour chaque solution, la totalité de la solution concentrée a été purifiée par HPLC semi-préparative. Les

collectes ont été gérées selon une collecte automatique programmée pour séparer les collectes de chaque pic du chromatogramme.

4.4.2 Isolement des composés majoritaires de l'extrait de salicaire

A partir de l'extrait de salicaire traité par SPE, nous avons pu réaliser l'isolement des composés majoritaires. Les conditions utilisées sont les suivantes :

Colonne : Synergi Fusion-RP de 250 mm x 10 mm; 4µm (Phenomenex, Le Pecq, France)
Température : 30°C
Débit : 4 $mL.min^{-1}$
Volume injecté : 800 µL
Concentration de la solution d'extrait : 61,5 $mg.mL^{-1}$ environ dans H_2O
Gradient :

Temps (min)	0	15	23,5	26	31
H_2O à 0,5% HCOOH (%)	100	90	90	30	30
MeOH à 0,1% HCOOH (%)	0	10	10	70	70

Détection : DAD-UV avec suivi à 345 nm

ANNEXE 4.5 HYDROLYSE ACIDE

4.5.1 Hydrolyse acide des extraits (Primevère vraie et Tulipe noire)

L'extrait de plante solubilisé à 10 $mg.mL^{-1}$ dans MeOH/HCl 0,02N (40/60) a été hydrolysé en milieu acide (HCl 6M) en présence d'hydroxyanisole butylé (BHA) à 2 $g.L^{-1}$ pendant deux heures à 90°C. L'échantillon hydrolysé a été dilué dans le MeOH avant analyse HPLC.

Pour l'extrait de primevère vraie, une solution standard de myricétine est ajoutée à l'échantillon et constitue le standard interne pour le dosage qui suivra l'étape d'hydrolyse. Pour l'extrait de Tulipe, ce standard interne n'a pas été ajouté puisque l'extrait en contenait.

4.5.2 Hydrolyse acide des molécules purifiées de l'extrait de primevère vraie

Pour hydrolyser les molécules purifiées, le protocole suivant a été mis en place. On solubilise 1 à 2 mg du composé dans un faible volume de H_2O. Un volume 10 µL de la solution concentrée est ajoutée à un volume de HCl 150 mM. Le volume de solution acide est adapté afin que le composé à hydrolyser soit à 0,25 mM en solution. L'hydrolyse acide est réalisée à 80°C (étuve) pendant 1h30 et avec agitation magnétique. Après ce délai, l'échantillon est séché au SpeedVac puis repris dans 200 µL d'H_2O.

La solution de l'hydrolysat est analysée en chromatographie ionique et en RP-HPLC.

4.5.3 Analyses RP-HPLC après hydrolyse acide (extraits et molécules purifiées)

Après hydrolyse, les extraits ou molécules sont analysés en chromatographie liquide en phase inverse et détecté en UV à 365 nm. La chaîne HPLC (U3000, Thermo Fisher Scientific) est composée d'une pompe binaire, d'un injecteur automatique, d'une enceinte thermostatée et d'un détecteur DAD-UV. Les conditions d'analyses sont indiquées ci-dessous :

Colonne MAX-RP C12 –250 x 4,6mm – 4µm (Phenomenex);
Débit : 1 $mL.min^{-1}$;
Température : 25°C ;
Solvant d'injection : MeOH ;
Concentration de solution d'extrait injectée : 625 $\mu g.mL^{-1}$;
Volume d'injection : 20 µL ;
Gradient d'élution :

Temps (min)	0	13	17	20	30
ACN (%)	20	40	40	55	55
KH_2PO_4 25 mM (%)	80	60	60	45	45

Des solutions standards de flavonols (apigénine, isorhamnétine, kaempférol, quercétine) sont préparées par solubilisation dans MeOH et passage aux ultra-sons. Des gammes d'étalonnage sont réalisées : on mélange les quatre flavonols et réalise des dilutions par ajout de MeOH permettant d'avoir des gammes de concentrations entre 0,5 et 100 $\mu g.mL^{-1}$ pour chaque génine. La myricétine est utilisée comme standard interne. Pour l'extrait de Tulipe noire, ce standard n'a pas été ajouté puisque l'extrait en contient naturellement et de ce fait, le dosage est semi-quantitatif.

4.5.4 Analyses en chromatographie ionique après hydrolyse acide (molécules purifiées)

Le système HPLC utilisée est une pompe (GP40), un injecteur automatique (U3000), un four à colonne, un détecteur électrochimique (Dionex ED50) munie d'une électrode d'or. La séparation est assurée avec une colonne CarboPak PA1 (4 x 250 mm) associée à une colonne aminotrap (4 x 50 mm). Tout le système HPLC (tubulures et vanne d'injection) doit être en peek.

Les conditions chromatographiques et de détection sont les suivantes :

Débit : 1$mL.min^{-1}$
Température four : 20°C
Volume d'injection : 25µL
Détection : Electrode d'or
Phases mobiles : Soude 20 mM dégazée à l'hélium ; H_2O dégazée à l'hélium

Gradient :

Temps (min)	0	17	24	30	30,1
NaOH 20mM (%)	10	10	50	50	10
H2O (%)	90	90	50	50	90

Des solutions standards de sept monosaccharides (arabinose, fucose, galactose, glucose, mannose, rhamnose et xylose) sont préparées par solubilisation des composés dans H_2O. L'hydrolysat est injecté après centrifugation et dilution du surnageant dans H_2O.

ANNEXE 4.6 ANALYSES RMN

Les spectres de Résonance Magnétique Nucléaire (RMN) du proton (^{1}H) et du carbone (^{13}C) ont été enregistrés sur l'appareil Bruker Avance II NMR (BrukerDaltonics, Billerica, MA, USA).

4.6.1 Conditions RMN pour les molécules purifiées de l'extrait de primevère vraie

Les conditions sont rappelées dans la fiche récapitulative des données analytiques pour chaque composé analysé (voir annexe 5).

4.6.2 Conditions RMN pour les molécules purifiées de l'extrait de salicaire

Les expériences étaient effectuées à 298,1K. Les molécules purifiées ont été solubilisées dans l'acétone-d_6 et l'acétone-d_6 mélangé à de l'eau deutérée respectivement pour les analyses en RMN ^{1}H (400 MHz) et RMN ^{13}C (100 MHz).

Annexe 5 ..... RECAPITULATIF DES DONNEES ANALYTIQUES OBTENUES POUR LES MOLECULES PURIFIEES A PARTIR DE L'EXTRAIT DE PRIMEVERE VRAIE

ANNEXE 5.1 DONNEES ANALYTIQUES DU COMPOSE F2-02GJ3022 (A)

Provenance : Fraction F2GJ2196 de l'extrait EtOH 60 % de *Primula veris* traité sur SPE C_{18}
Masse isolée : 18 mg
Aspect : poudre amorphe jaune foncé

Formule brute : $C_{33}H_{40}O_{22}$
UV λ_{max} : 256, 267, 359 nm

Hydrolyse acide et HPLC : Quercétine
Hydrolyse acide et chromatographie ionique : Glucose

Données HRMS-ESI-QTOF

Mode	m/z	Attribution	Masse monoisotopique moyenne
ESI $^+$	789,208091	$[M+H]^+$	788,200744 g.mol^{-1}
ESI $^+$	811,189798	$[M+Na]^+$	
ESI $^+$	627,154693	$[M + H - hexose]^+$	
ESI $^+$	465,102271	$[M + H - 2\ hexoses]^+$	
ESI $^+$	303,049505	$[M + H - 3\ hexoses]^+$	
ESI $^-$	787,194206	$[M-H]^-$	

Données RMN

RMN ^{1}H (D_2O ; 400 MHz ; 299,4 K) : signal mal résolu (les multiplicités et les intégrations ne sont pas exploitables).

RMN ^{13}C (D_2O ; 300 MHz ; 300,1 K) : enregistrement d'un signal bruiteux : aucun signal de carbone n'est détecté.

Eléments structuraux déterminés

Famille	Flavonoïde triglycosylé
Génine	Quercétine
Sucres	3 Glucoses
Glycosylation	Forme continue non déterminée
Arrangement des sucres	Non déterminé

ANNEXE 5.2 DONNEES ANALYTIQUES DU COMPOSE F3-01GJ3121 (B)

Provenance : Fraction F3GJ2196 de l'extrait EtOH 60 % de *Primula veris* traité sur SPE C_{18}
Masse isolée : 20,5 mg
Aspect : poudre amorphe jaune vif

Formule brute : $C_{27}H_{30}O_{17}$
UV λ_{max} : 257, 265, 367 nm

Hydrolyse acide et HPLC : Quercétine
Hydrolyse acide et chromatographie ionique : Glucose

Données HRMS-ESI-QTOF

Mode	m/z	Attribution	Masse monoisotopique moyenne
ESI $^+$	**627,155391**	**$[M+H]^+$**	**626,148170 g.mol^{-1}**
ESI $^+$	**649,137101**	**$[M+Na]^+$**	
ESI $^+$	**465,102246**	**$[M + H - hexose]^+$**	
ESI $^+$	**303,049675**	**$[M + H - 2 hexoses]^+$**	
ESI $^-$	**625,141734**	**$[M-H]^-$**	

Données RMN

Pas de données RMN : le composé ne s'est pas solubilisé dans les solvants testés suivants : D_2O, D_2O/MeOH-d_4 50/50 et MeOH-d_4. Cela aurait dû être le cas.

Eléments structuraux déterminés

Famille	Flavonoïde diglycosylé
Génine	Quercétine
Sucres	2 Glucoses
Glycosylation	Forme continue non déterminée
Arrangement des sucres	Non déterminé

ANNEXE 5.3 DONNEES ANALYTIQUES DU COMPOSE F3-02GJ3121 (C)

Provenance : Fraction F3GJ2196 de l'extrait EtOH 60 % de *Primula veris* traité sur SPE C_{18}
Masse isolée : 19,5 mg
Aspect : poudre amorphe jaune vif

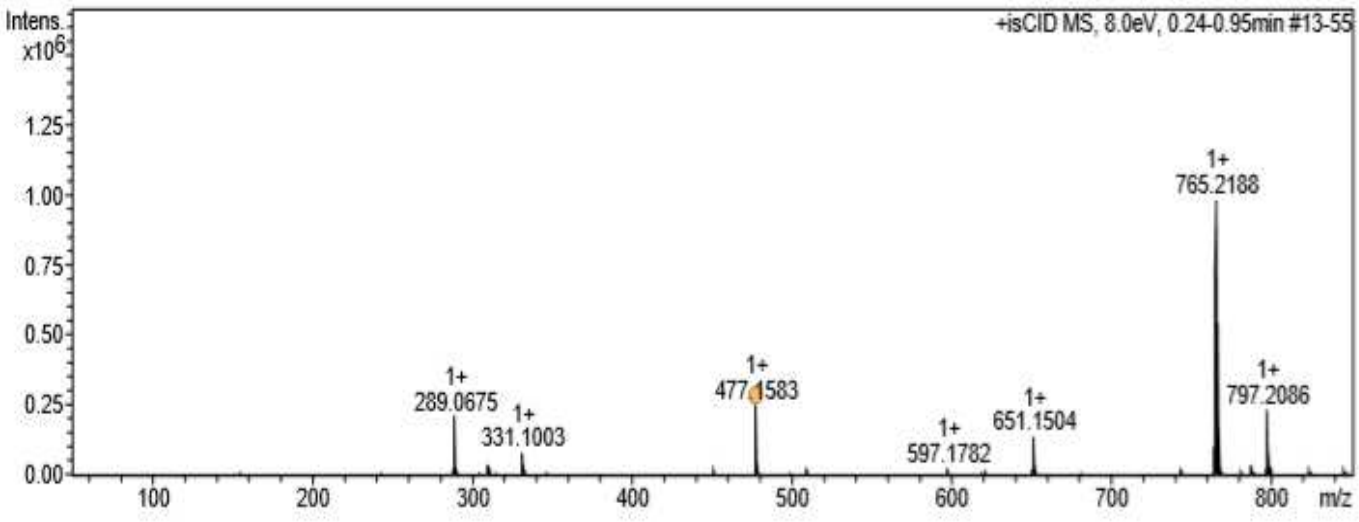
Formule brute : $C_{33}H_{40}O_{19}$
UV λ_{max} : 266, 349 nm

Hydrolyse acide et HPLC [¥]: **Kaempférol** / Quercétine
Hydrolyse acide et chromatographie ionique : **Rhamnose** / Glucose / (Galactose)

Données HRMS-ESI-QTOF

Mode	m/z	Attribution	Masse monoisotopique moyenne
ESI $^+$	**741,223574**	**$[M+H]^+$**	**740,216734 g.mol^{-1}**
ESI $^+$	**763,206074**	**$[M+Na]^+$**	
ESI $^+$	**595,165755**	**[M + H - desoxyhexose]$^+$**	
ESI $^+$	**449,107894**	**[M + H - 2 desoxyhexoses]$^+$**	
ESI $^+$	**287,055198**	**[M + H - 2 desoxyhexoses - 1 hexose]$^+$**	
ESI $^-$	**739,210247**	**$[M-H]^-$**	
ESI $^-$	**369,101258**	**$[2M-H]^{2-}$**	

FIA-MS/MS (mode isCID) de l'adduit sodé de la molécule

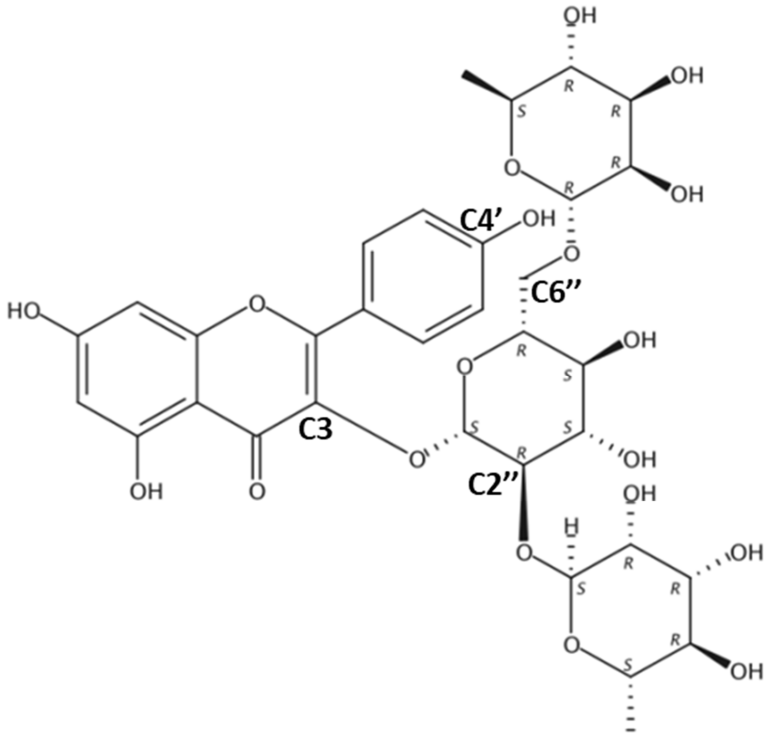


L'adduit sodé de la molécule (portant deux atomes de deutériums) est observé *à m/z* 765,2188. L'ion fragment source de *m/z* 477,1583 correspond à la masse d'un hexose avec de deux desoxyhexoses complexé à un ion sodium à 0,9 ppm près. Les sucres qui avaient capté le sodium (Na^+) le conservent. Les 3 sucres sont donc sous forme continue dans la structure.

[¥] Note à propos de la nomenclature : les génines ou sucres indiqués **en gras** sont ceux qui sont majoritaires lors de l'analyse après hydrolyse acide et ceux indiqués entre parenthèse sont détectés en très faible proportion.

Données RMN

La référence bibliographique utilisée est ‘’Malonylated flavonol glycosides from the petals of *Clitoria ternatea*’’ [117] qui décrit un kaempférol 3-O-(2’’-O-α-rhamnosyl-6’’-O-β-glucosyl)-β-glucoside (*cf*. structure ci-dessous). Le solvant utilisé pour les analyses dans la publication est le méthanol deutéré.



RMN ^{1}H (MeOH-d_4 + HCOOH ; 250 MHz ; 298,1 K)

δ (ppm) [I, multiplicité, J en Hz]	**Type de proton**	**Attribution d’après la comparaison avec la référence**
0,98 [3 ; d ; 6,25Hz]	CH_3	C6 du Rha lié au C2’’
1,17 [3 ; d ; 6,00Hz]	CH_3	C6 du Rha lié au C6’’
1,99 [0,56 ; s]	na	na
3,27-4,01 [19 ; massif]	H sucres non résolus	H sucres
4,53 [1 ; s]	H1 anomère en α	H1 Rha lié à C2''Glu
5,22 [1 ; s]	H1 anomère en α	H1 Rha lié à C6'' Glu
5,60 [1 ; d ; 7,75Hz]	H1 anomère en β	H1 Glu
6,19 [0,3 ; s]	H aromatique	H6
6,38 [0,3 ; s]	H aromatique	H8
6,90 [2 ; d ; 8,75Hz]	2H aromatiques	H3’ et H5’
8,06 [2 ; d ; 8,75Hz]	2H aromatiques	H2’ et H6’

On observe un problème de détection : les intégrations des protons aromatiques H6 et H8 sont faibles et très inférieures à 1.

RMN ^{13}C ($MeOH\text{-}d_4$ + HCOOH ; 100 MHz ; 298,1 K)

δ en ppm	**Type de carbone**	**Attribution d'après comparaison avec la référence**	**Données de la référence (MeOH-d4) Différence > 1 ppm avec expérience**
17,5	CH3	C6 Rhamnose	17,7
17,9	CH3	C6 Rhamnose	**18,0**
67,1	CH2	C6'' du Glucose	68,4
69,7	C sucre	C2'' ou C3''	79,9 ou 78,9
69,8	C sucre	C5 Rhamnose	**69,9**
70,7	C sucre	C5 Rhamnose	70,0
72,1	C sucre	C4''	72,1
72,3	C sucre	C2 Rhamnose	**72,3**
72,3	C sucre	C3 Rhamnose	72,4
72,4	C sucre	C3 Rhamnose	**72,4**
73,9	C sucre	C2 Rhamnose	72,5
74,1	C sucre	C4 Rhamnose	**74,0**
75,3	C sucre	C4 Rhamnose	74,2
75,7	C sucre	C2'' ou C3''	79,9 ou 78,9
77,5	C sucre	C5''	77,3
100,9	C anomère	C1 Glucose (corrèle avec H beta)	100,6
101,8	C anomère	C1 Rhamnose (corrèle avec H alpha)	**102,4**
102,6	C anomère	C1 Rhamnose (corrèle avec H alpha)	102,8
105,9		C10	106,1
116,2	C aromatique	C3' et C5'	116,3
123,0		C1'	123,3
132,2	C aromatique	C2' et C6'	132,2
134,4		C3	134,5
158,3		C2	159,2
158,7		C9	158,6
161,3		C4'	161,4
163,1		C5	163,3
165,5		C7	165,8
179,4		C4	179,5

On ne détecte pas de signal expérimental pour C6 et C8 qui sont deux carbones aromatiques.

On n'a pas les mêmes signaux pour C2'' et C3'' qui sont deux carbones appartenant au glucose.

On observe un signal de carbone à 165,092 ppm provenant de l'acide ajouté à l'analyse (HCOOH).

Eléments structuraux déterminés

Famille	**Flavonoïde triglycosylé**
Génine	**Kaempférol**
Sucres	**2 Rhamnoses + 1 hexose (Glucose ou Galactose)**
Glycosylation	**Forme continue**
Arrangement des sucres	**Non déterminé**

ANNEXE 5.4 DONNEES ANALYTIQUES DU COMPOSE F3-03GJ3121 (D)

Provenance : Fraction F3GJ2196 de l'extrait EtOH 60 % de *Primula veris* traité sur SPE C_{18}
Masse isolée : 22,5 mg
Aspect : poudre amorphe jaune vif

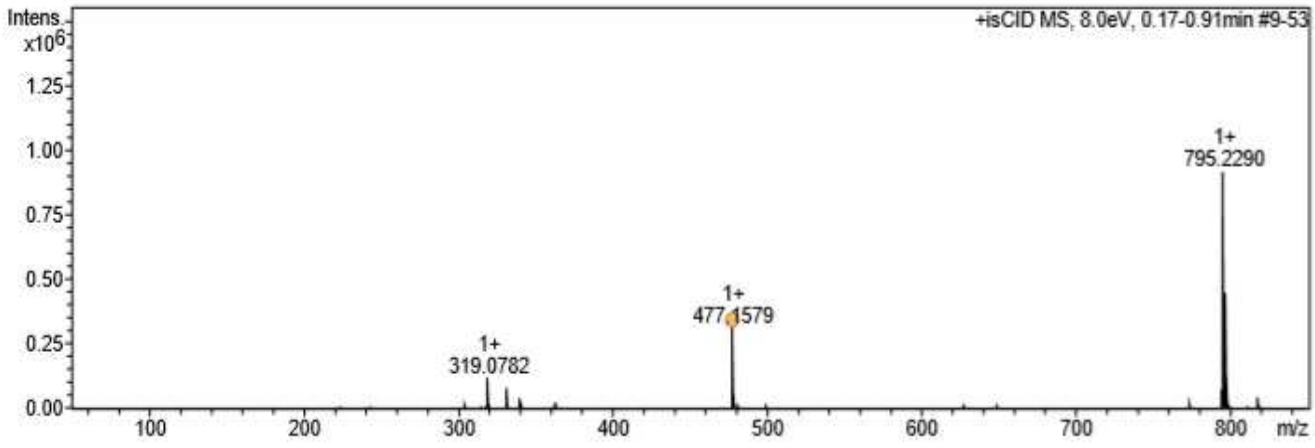
Formule brute : $C_{34}H_{42}O_{20}$
UV λ_{max} : 255, 268, 357 nm

Hydrolyse acide et HPLC : Isorhamnétine
Hydrolyse acide et chromatographie ionique : **Rhamnose** / Glucose / Galactose

Données HRMS-ESI-QTOF

Mode	m/z	Attribution	Masse monoisotopique moyenne
ESI $^+$	771,234368	$[M+H]^+$	770,227297 g.mol^{-1}
ESI $^+$	793,216786	$[M+Na]^+$	
ESI $^+$	625,176580	$[M + H - desoxyhexose]^+$	
ESI $^+$	479,119236	$[M + H - 2\ desoxyhexoses]^+$	
ESI $^+$	317,065143	$[M + H - 2\ desoxyhexose - 1\ hexose]^+$	
ESI $^-$	769,220825	$[M-H]^-$	
ESI $^-$	384,106682	$[2M-H]^{2-}$	

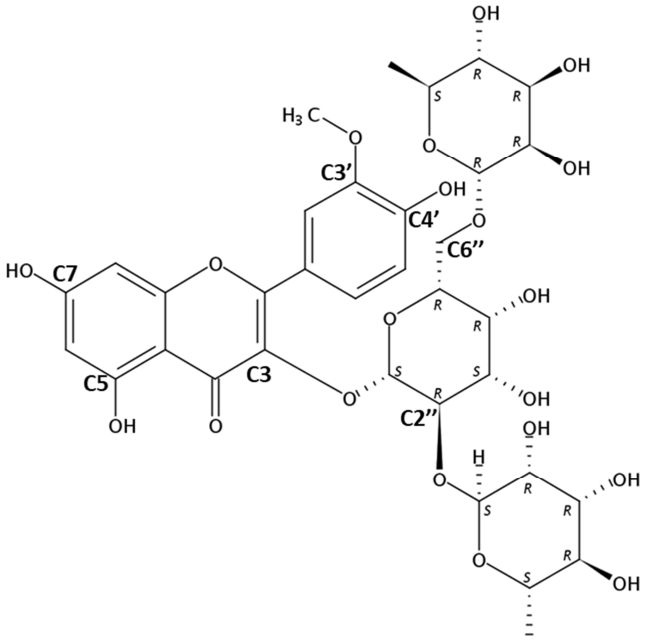
FIA-MS/MS (en isCID) de l'adduit sodé



On observe l'adduit sodé de la molécule et portant deux deutériums $[M_{d2}+Na]^+$ à *m/z* 795,2290. Lors de la fragmentation, les sucres conservent le sodium et donc la charge ; l'ion *m/z* 477,1579 correspond à 0,2 ppm à la masse d'un hexose et deux desoxyhexoses complexés à un atome de sodium. Les 3 sucres sont donc sous forme continue dans la structure.

Données RMN

La publication ayant servi de base à l'annotation des signaux de RMN est le papier de Phytochemistry, 26 : 291-294, 1987 [118] afin de comparer les données du composé de structure présentée à la page suivante :



RMN ^{1}H (D_2O ; 400 MHz ; 298,1 K)

δ en ppm [I ; multiplicité ; J en Hz]	**Attribution d'après comparaison avec la référence**
1,09 - 1,11 [3 ; d ; 5,60 Hz]	CH_3 de Rhamnose
1,25 [3 ; br s]	CH_3 de Rhamnose
1,94 [1 ; s]	na
3,28 [1 ; t ; 9,20/18,80 Hz]	na
3,45	H sucre
3,47	H sucre
3,49	H sucre
3,59	H sucre
3,63	H sucre
3,65	H sucre
3,71	H sucre
3,81	H sucre
3,84	H sucre
3,88	H sucre
3,96	H sucre
3,98	H sucre
4,15 [1 ; s]	H sucre
4,30 [1 ; s]	H sucre
4,50 [1 ; s]	H anomère α
5,16 [1 ; s]	H anomère α
5,25 [1 ; d ; 6,40 Hz]	H anomère β
6,85 [1 ; *br* s]	H aromatique
7,43 [1 ; *br* s]	H aromatique
7,58 [1 ; s]	H aromatique

Il manque 2 signaux de H aromatique.

RMN ^{13}C (D_2O ; 100 MHz ; 298,1 K)

δ en ppm	**Type de carbone**	**Attribution d'après comparaison avec la référence**
16,6	$-CH_3$	C6 de Rhamnose
16,5	$-CH_3$	C6 de Rhamnose
56,1	$-CH_3$	CH_3-O (génine)
67,3	$-CH_2$	C6'' du Galactose
68,6	C sucre	C5 de Rhamnose
68,9	C sucre	C5 de Rhamnose
69,1	C sucre	C4'' du Galactose
70,0	C sucre	C2 de Rhamnose
70,1	C sucre	C2 de Rhamnose
70,3	C sucre	C3 de Rhamnose
70,4	C sucre	C3 de Rhamnose
71,8	C sucre	C4 de Rhamnose
72,1	C sucre	C4 de Rhamnose
73,2	C sucre	C3'' ou C5'' du Galactose
73,5	C sucre	C3'' ou C5'' du Galactose
76,9	C sucre	C2'' du Galactose
99,7	C anomère (corrèle avec H en α)	C1 de Rhamnose
100,6	C anomère (corrèle avec H en α)	C1 de Rhamnose
101,3	C anomère (corrèle avec H en β)	C1 du Galactose
102,7	C quaternaire	C10
112,9	CH aromatique	C2'
115,2	CH aromatique	C5'
121,6	CH aromatique	C6'
123,0	C quaternaire	C1'
132,5	C quaternaire	C3
147,0	C quaternaire	C4'
149,3	C quaternaire	C3'
157,0	C quaternaire	C2
157,2	C quaternaire	C9
159,7	C quaternaire	C5
176,6	C=O	C4

Les signaux de 3 carbones quartenaires sont manquants : celui de C_7 (vers 162 ppm) ; C_8 (vers 94,5 ppm) et C_6 (vers 99 ppm).

Tous les signaux détectés sont en concordance avec les signaux référencés dans la publication de référence : de manière générale, de légers décalages de déplacements chimiques sont observés (souvent <0,5 ppm) sauf pour C5, C10, C1', ainsi que les C2'', C4'' et C6'' du galactose avec des différences entre 1 et 2,5 ppm).

En complément, l'analyse RMN du produit a été faite dans le DMSO-d_6. Les signaux des carbones de Galactose de la publication sont retrouvés avec les mêmes déplacements chimiques. Par contre, il manque 5

signaux de carbone (C3', C7, C10 et C6, C8 comme carbones aromatiques). De plus, des décalages de maximum 2 ppm de déplacements existent pur C3,C4, C9 et C1'.

Eléments structuraux déterminés

Famille	**Flavonoïde triglycosylé**
Génine	**Isorhamnétine**
Sucres	**2 Rhamnoses, 1 hexose (Glucose ou Galactose)**
Glycosylation	**Forme continue**
Arrangement des sucres	**O-α-rhamnosyl-(1→2)-O-[α-rhamnosyl-(1→6)-β-Galactose]**

ANNEXE 5.5 DONNEES ANALYTIQUES DU COMPOSE F5-02GJ3100 (E)

Provenance : Fraction F5GJ2196 de l'extrait EtOH 60 % de *Primula veris* traité sur SPE C_{18}
Masse isolée : 31 mg
Aspect : poudre amorphe jaune vif

Formule brute : $C_{34}H_{42}O_{22}$
UV λ_{max} : 253, 266, 357 nm

Hydrolyse acide et HPLC : **Isorhamnétine** / (Kaempférol / Quercétine)
Hydrolyse acide et chromatographie ionique : Glucose

Données HRMS-ESI-QTOF

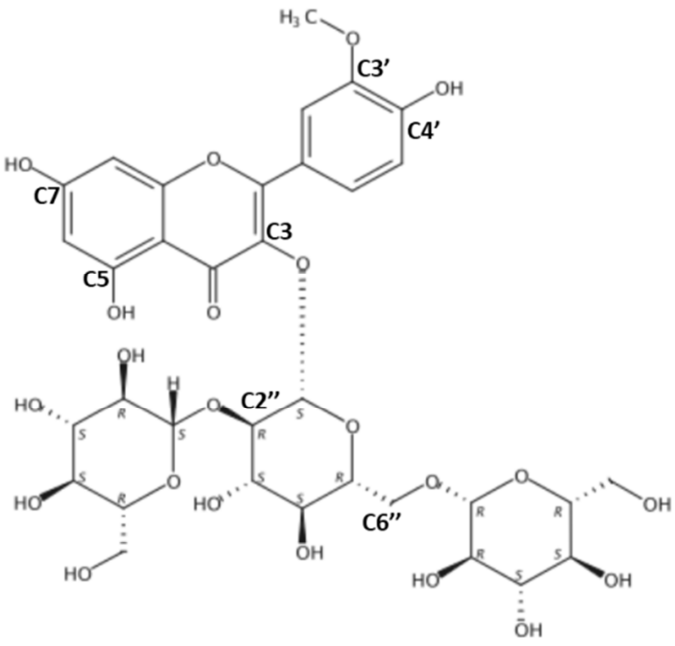
Mode	m/z	Attribution	Masse monoisotopique moyenne
ESI $^+$	803,224156	$[M+H]^+$	802,217175 g.mol^{-1}
ESI $^+$	825,206114	$[M+Na]^+$	
ESI $^+$	641,171764	$[M + H - hexose]^+$	
ESI $^+$	479,118345	$[M + H - 2\ hexoses]^+$	
ESI $^+$	317,065648	$[M + H - 3\ hexoses]^+$	
ESI $^+$	801,210345	$[M-H]^-$	
ESI $^+$	400,101701	$[2M-H]^{2-}$	

Spectre MS/MS de l'ion *m/z* = 509,121483 ($C_{18}H_{30}NaO_{15}$) qui correspond à 3 hexoses sous forme sodée se fragmente en *m/z* = 365,086827 ($C_{12}H_{22}NaO_{11}$) et m/z = 347,077090 ($C_{12}H_{20}NaO_{10}$)

Données RMN

Comparaison à la référence bibliographique ''Comparative phytochemical and morphological analyses of three Italian *Primula species*'' [108] dont les analyses RMN ont été effectuées dans le DMSO-d_6.

La molécule décrite qui pourrait correspondre à notre composé est l'isorhamnetin-3O-(2,6-di-O-β-D-glucopyranosyl-β-D-glucopyranoside) dont la structure est rappelée ci-contre :



RMN ^{1}H (D_2O ; 400 MHz ; 298,1 K)

δ en ppm [I ; multiplicité ; J en Hz]	**Type de proton**	**Attribution d'après comparaison avec la référence :** (**+** = concordance et **-** = pas de concordance)
1,87 [1 ; br s]	Non déterminé	**-**
1,94 [1 ; s]	Non déterminé	**-**
1,99 [1 ; t ; 8,40 Hz]	Non déterminé	**-**
2,90 à 3,11 [4 ; m]	4 H de sucre	**+**
3,19 à 3,27 [3 ; m]	3 H de sucre	**+**
3,41 [1 ; dd ; 5,20/12,00 Hz]	H sucre	**+**
3,54 à 3,64 [6 ; m]	6 H de sucre	**+**
3,73 [3 ; br s]	O-C**H_3**	**+**
3,79 à 3,92 [4 ; m]	4 H de sucre	**+**
3,98 [1 ; d ; 7,60 Hz]	H anomère en β	Sucre terminal en 2''-Glc
4,14 [1 ; d ; 7,60 Hz]	H anomère en β	Sucre terminal en 6''-Glc
5,24 [1 ; d ; 6,40 Hz]	H anomère en β	Sucre primaire lié à la génine
6,13 [2 ; br s]	2 H aromatique	H6 et H8
6,89 [1 ; d ; 7,60 Hz]	H aromatique	H5'
7,55 [1 ; d ; 7,60 Hz]	H aromatique	H6'
7,89 [1 ; s]	H aromatique	H2'

RMN ^{13}C (D_2O ; 100 MHz ; 298,2 K)

δ en ppm	**Type de carbone**	**Attribution d'après comparaison avec la littérature**
55,6	-CH_3	**C**H3-O
60,1	-CH_2	-CH_2 du Glu attaché en C6''
61,5	-CH_2	-CH_2 du Glu attaché au C2''
67,6	CH_2	C6''
69,1	CH	C sucre
69,6	CH	C sucre
70,5	CH	C sucre
74,2	CH	C sucre
74,4	CH	C sucre
74,6	CH	C sucre
74,7	CH	C sucre
74,9	CH	C sucre
75,3	CH	C sucre
76,4	CH	C sucre
76,5	CH	C sucre
83,0	CH	C sucre
95,9	CH aromatique	C8
101,0	CH aromatique	C6
101,3	C anomère	C1''
101,8	C anomère	C1 du 6'' Glu
103,5	C quaternaire	C10
104,6	C anomère	C1 du 2'' Glu
112,4	CH aromatique	C5'
115,3	CH aromatique	C2'

δ en ppm	Type de carbone	Attribution d'après comparaison avec la littérature
121,1	C quaternaire	C6'
122,6	CH aromatique	C1'
134,1	O=C-**C**-O-Glu	C3
146,6	C quaternaire	C3'
149,0	C quaternaire	C4'
154,6	C quaternaire	C2
157,0	C quaternaire	C9
159,3	C quaternaire	C5
177,4	C=O	C4

Les spectres RMN du carbone et du proton de la molécule isolée correspondent à ceux décrits dans la littérature même si les deux analyses ne sont pas faites dans le même solvant (D_2O *versus* DMSO-d_6) : les déplacements chimiques observés sont similaires. Il manque un signal de carbone quaternaire (C_7). De plus, les signaux des protons sont superposés dans le spectre RMN ^1H expérimental, ce qui n'est pas le cas pour l'analyse publiée et réalisée dans le DMSO-d_6.

Eléments structuraux déterminés

Famille	**Flavonoïde triglycosylé**
Génine	**Isorhamnétine**
Sucres	**3 Glucoses**
Glycosylation	**Forme continue**
Arrangement des sucres	**O-β-Glucosyl-(1→2)-O-[β-glucosyl-(1→6)-Glucose]**

ANNEXE 5.6 DONNEES ANALYTIQUES DU COMPOSE F6-05GJ3056 (F)

Provenance : Fraction F6GJ2196 de l'extrait EtOH 60 % de *Primula veris* traité sur SPE C_{18}
Masse isolée : 15 mg
Aspect : poudre amorphe jaune marron

Formule brute : $C_{28}H_{32}O_{17}$ (composé majoritaire) et $C_{29}H_{34}O_{18}$
UV λ_{max} : 256, 265, 365 nm

Hydrolyse acide et HPLC : **Isorhamnétine** / Kaempférol / Quercétine
Hydrolyse acide et chromatographie ionique : **Glucose** / Galactose / Rhamnose

Données HRMS-ESI-QTOF du composé majoritaire

Mode	m/z	Attribution (Indice $_d$: deutérium)	Masse monoisotopique moyenne
ESI $^+$	643,183821	$[M_{d2}+H]^+$	640,163998 g.mol^{-1}
ESI $^+$	641,171399	$[M+H]^+$	
ESI $^+$	665,165843	$[M_{d2}+Na]^+$	
ESI $^+$	481,130222	$[M_{d2}+H-\text{hexose}]^+$	
ESI $^+$	319,078022	$[M_{d2}+H-2\ \text{hexoses}]^+$	
ESI $^-$	641,170022	$[M_{d2}-H]^-$	

Données RMN

Pas de référence bibliographique pouvant correspondre aux éléments déterminés et présentant des données de RMN.

RMN ^{1}H (D_2O ; 400 MHz ; 296,2 K)

δ en ppm [I ; multiplicité ; J en Hz]	Type de proton
1,92 [2 ; s]	
2,48 [1 ; br s]	
2,78 [1 ; t ; na]	
2,92 [1 ; t ; na]	
3,11 [1 ; t ; na]	
3,30 à 3,76 [13 ; massif]	H des sucres
3,89 [2 ; d ; 12,4 Hz]	
4,08 [1 ; d ; 7,20 Hz]	H anomère (β)
5,07 [1 ; d ; 5,20 Hz]	H anomère
6,80 [1 ; br s]	H aromatique
7,45 [1 ; d ; 6,8 Hz]	H aromatique
7,68 [1 ; s]	H aromatique
8,46 [0,4 ; s]	H aromatique

RMN ^{13}C (D_2O ; 300 MHz ; 297,0 K)

δ en ppm	Attribution
23,2	na (pourrait être CH3 mais ne corrèle pas avec 3 protons. Corrélation avec 2 protons)
55,6	C sucre
60,3	CH2 d'un sucre
66,9	CH2 d'un sucre
69,2	C sucre
69,9	C sucre
73,0	C sucre
73,9	C sucre
75,2	C sucre
75,5	C sucre
76,7	C sucre
101,4	C anomère (corrélé à H anomère à 5,073 ppm)
101,9	C anomère (corrélé à un H β à 4,082 ppm)
112,3	C aromatique
115,8	C aromatique
119,8	C aromatique
123,0	C aromatique
132,8	C quaternaire
147,4	C quaternaire
155,6	C quaternaire
157,4	C quaternaire
159,3	C quaternaire
171,0	C quaternaire
173,9	C quaternaire
176,8	C quaternaire
181,5	C quaternaire

Eléments structuraux déterminés pour le composé majoritaire (présent à 85 %)

Famille	**Flavonoïde diglycosylé**
Génine	**Isorhamnetin (non confirmée par RMN)**
Sucres	**2 Glucoses (non confirmé par RMN)**
Glycosylation	**Forme continue non déterminée**
Arrangement des sucres	**Non déterminé**

Problèmes

Les analyses HRMS indiquent clairement que la fraction contient 2 composés dont le majoritaire est diglycosylé. Cela est cohérent avec la faible pureté du produit majoritaire déterminée à 85 % par HPLC. L'hydrolyse acide a mis en évidence la présence de 3 génines et de 3 oses : l'isorhamnétine et le glucose sont majoritaires.

ANNEXE 5.7 DONNEES ANALYTIQUES DU COMPOSE F8-01GJ3087 (G)

Provenance : Fraction F8GJ2196 de l'extrait EtOH 60 % de *Primula veris* traité sur SPE C_{18}
Masse isolée : 23 mg
Aspect : poudre amorphe jaune vif

Formule brute : $C_{28}H_{32}O_{16}$
UV λ_{max} : 256, 265, 365 nm

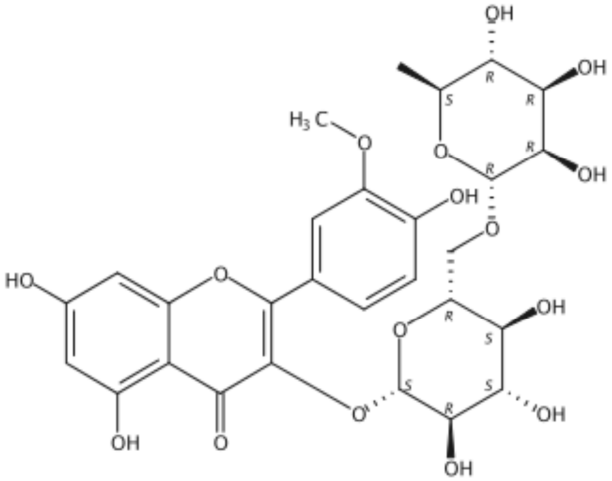
Hydrolyse acide et HPLC : Isorhamnétine / Kaempférol
Hydrolyse acide et chromatographie ionique : Glucose / Galactose / Rhamnose

Données HRMS-ESI-QTOF

Mode	m/z	Attribution	Masse monoisotopique moyenne
ESI $^+$	625,176362	$[M+H]^+$	624,168982 g.mol^{-1}
ESI $^+$	647,158186	$[M+Na]^+$	
ESI $^+$	463,123219	$[M + H$ - hexose$]^+$	
ESI $^+$	479,118548	$[M + H$ - desoxyheose$]^+$	
ESI $^+$	317,065451	$[M + H$ – hexose - desoxyheose$]^+$	

Données RMN

Les références bibliographiques utilisées présentent les données obtenues pour l'**isorhamnetin-3O-α-L-rhamnosyl (1→6) β-D-glucoside** dont la structure est rappelée ci-dessous :



Les références sont pour les données ^{13}C : Flavonol Glycosides from *Eschschotlzia californica* [120] et pour les données ^{1}H : Constituents of the stems and fruits of *Opuntia ficus-indica* var. saboten [121]. Les analyses RMN ont été effectuées dans le même solvant (MeOH-$_{d4}$) que celui des publications.

RMN ^{1}H (MeOH-d_4 ; 250 MHz ; 298,1 K)

δ en ppm [I ; multiplicité ; J en Hz]	**Données référence** (Hexose = Glucose)	**Attribution d'après comparaison avec la référence**	**Remarques**
1,12 [3 ; d ; 6,25Hz]	1,11 [3 ; d ; 6,2Hz]	Rha-**H6**	
3,28 [1 ; br s]	na	na	
3,30 à 3,79 [11 ; m]	3,26 à 3,84 [nd, m]	H sucres	Massif avec signaux superposés
3,97 [3 ; s]	3,96 [3 ; s]	O-C**H_3**	
4,53 [1 ; br s]	4,54 [1 ; d ; 1,3Hz]	Rha-**H1'''**	Large singulet = α-Rhamnose
5,17 [1 ; d ; 7,75Hz]	5,24 [1 ; d ; 7,5Hz]	Glu-**H1''**	7 < J < 8 Hz = β-Hexose
6,16 [1 ; t ; 1,00/2,25Hz]	6,22 [1 ; d ; 2,1Hz]	**H6**	
6,35 [1 ; t ; 1,00/2,25Hz]	6,42 [1 ; d ; 2,1Hz]	**H8**	
6,90 [1 ; d ; 8,5Hz]	6,92 [1 ; d ; 8,5Hz]	**H5'**	
7,60 [1 ; dd ; 8,5/2,0Hz]	7,64 [1 ; dd ; 8.5/2,1Hz]	**H6'**	
8,03 [1 ; d ; 2,0Hz]	7,95 [1 ; d ; 2,0Hz]	**H2'**	

RMN ^{13}C (MeOH-d_4 ; 63 MHz ; 298,1 K)

δ en ppm (différence avec la référence > 1 ppm = case en rouge)	**δ en ppm (Référence)**	**Attribution d'après comparaison avec la référence**
18,0	17,9	C6''' du Rhamnose
56,9	56,8	CH3-O de la génine
67,5	68,6	C6'' du Glucose
69,7	69,8	C5''' du Rhamnose
70,0	71,6	C4'' du Glucose
72,1	72,1	C2''' du Rhamnose
72,3	72,3	C3''' du Rhamnose
73,1	73,9	C4''' du Rhamnose
73,9	75,9	C2'' du Glucose
75,1	77,4	C5'' du Glucose
75,5	78,2	C3'' du Glucose
95,4	95,0	C8
100,8	100,1	C6
101,9	102,5	C1''' du Rhamnose
105,0	104,5	C1'' du Glucose
105,2	105,7	C10
114,6	114,6	C2'
116,0	116,1	C5'
123,0	123,0	C6'
123,8	124,0	C1'
135,4	135,5	C3
148,4	148,3	C4'
150,9	150,9	C3'
158,5	158,5	C2

158,7	158,9	C9
162,9	163,0	C5
168,5	166,2	C7
179,2	179,3	C4

Incohérence : on observe un décalage des déplacements chimiques entre les signaux expérimentaux et ceux indiqués pour le glucose dans la référence bibliographique. Les différences sont de l'ordre de 1 à 2,5 ppm.

Eléments structuraux déterminés

Famille	**Flavonoïde diglycosylé**
Génine	**Isorhamnetin**
Sucres	**Glucose ou Galactose - Rhamnose**
Glycosylation	**Forme continue non détectée/ non déterminée**
Arrangement des sucres	**Non déterminé**

Annexe 6 ..... TESTS D'ACTIVITÉS BIOLOGIQUES

ANNEXE 6.1 CULTURES CELLULAIRES ET TRAITEMENTS

6.1.1 Cultures et traitements de monocouches de kératinocytes épidermiques normaux humains

Les cellules sont ensemencées dans des plaques 96 puits et laissées croître jusqu'à confluence. Ensuite, elles sont traitées ou non (contrôle) par l'extrait de *Lythrum salicaria* à 20 $\mu g.mL^{-1}$ ou par $CaCl_2$ à 1,5 mM, et incubées pendant des temps variables selon la finalité : 24 heures (expression des gènes), 72 heures (expression de la filaggrine et TGK) ou encore 96 heures (expression de KRT10). Les expériences sont effectuées en triplicats.

6.1.2 Culture et traitements d'épidermes normaux humains

Les épidermes humains reconstruits (RHE) sont générés dans des inserts de culture en polycarbonate à partir d'échantillons chirurgicaux provenant de prépuces pédiatriques. Après le passage à l'interface air/liquide, les RHE ont été cultivés pendant 7 jours sans vitamine C, avec vitamine C à 200 $\mu g.mL^{-1}$ et enfin sans vitamine C mais avec l'extrait de *Lythrum salicaria* à 20 $\mu g.mL^{-1}$. Les traitements des RHE ont été renouvelés tous les 2 ou 3 jours.

6.1.3 Culture et traitements de peaux reconstruites

Le modèle de peau est composé de kératinocytes ensemencés sur une couche de collagène contenant des fibroblastes et cultivés à l'interface air/liquide pour obtenir une peau reconstruite d'épaisseur totale. Un équivalent de derme est préparé en mixant une solution de collagène de queue de rat à une suspension de fibroblastes (1,3 $mg.mL^{-1}$ et 0,5 x 10^6 cellules/mL, concentration/contenu final(e), dans un volume de 400 µL/insert). Après 2 heures de polymérisation/solidification, la suspension de cellules épidermiques est ajoutée et les FTSK sont incubés dans un milieu complet Epilife (Cascade Biologics, Portland, Etats-Unis) supplémenté par 5 $\mu g.mL^{-1}$ d'insuline et 1,5 mM $CaCl_2$ pour deux jours. Les inserts sont ensuite placés à l'interface air-milieu sur gazes stériles. Les FTSK non traitées sont incubées dans un milieu complet Epilife (supplémenté avec 5 $\mu g.mL^{-1}$ d'insuline et 1,5 mM de $CaCl_2$, 50 $\mu g.mL^{-1}$ vitamine C et 3 $ng.mL^{-1}$ KGF). Pour les conditions déplétées, les FTSK sont cultivées dans un milieu incomplet consistant en un milieu Epilife supplémenté sans complément et contenant ou non (FTSK carencées et non traitées) les composés à tester. L'extrait de *Lythrum salicaria* ou la vescalagine sont testés à 30 $\mu g.mL^{-1}$. Toutes les différentes peaux ont ensuite été incubées pendant 9 jours avec un renouvellement du milieu et des traitements tous les deux ou trois jours. Toutes les conditions expérimentales sont effectuées en duplicat.

ANNEXE 6.2 TESTS D'ACTIVITES BIOLOGIQUES

6.2.1 Test MTT (détermination de la cytotoxicité)

Après traitement par les produits dont la cytotoxicité est à déterminer, les cellules ont été incubées en présence de MTT (sel de tétrazolium) dont la transformation en cristaux bleus de formazan est proportionnelle à l'activité de la succinate deshydrogénase (enzyme mitochondriale). Après dissociation des cellules et solubilisation du formazan par ajout de DMSO, la densité optique (DO), représentative du nombre de cellules vivantes et de leur activité métabolique, a été mesurée avec un lecteur de microplaques à 540 nm.

6.2.2 Analyse par RT-qPCR

Les acides ribonucléiques (ARN) totaux des monocouches de kératinocytes épidermiques humains normaux sont isolés grâce au kit NucleoSpin RNA II kit (Macherey-Nagel, Duren, Allemagne) et transcrits inversement avec SuperScript II Reverse Transcriptase (Invitrogen Life Technologies) selon les instructions du fabricant. Des PCR quantitatives en temps réel (Real-Time quantitative Polymerase Chain Reaction, RT-qPCR) sont exécutées en utilisant le kit LightCycler-FastStart DNA MasterPlus SYBR Green I sur LightCycler 480 (Roche Diagnostics, Meylan, France). Les composés de réaction sont 1X DNA Master Mix, et 0,5 µM d'oligonucléotides sens et anti-sens purifiés en HPLC et achetés chez Eurogentec (Eurogentec France, Angers, France), conçus en utilisant Primer3 software. L'expression des gènes dans les cellules est normalisée avec le gène ménager GAPDH (Glyceraldehyde-3-phosphate dehydrogenase) et reportée selon la méthode des delta delta Ct ($\Delta\Delta C_T$) avec une augmentation des ARN définie par l'équation suivante : $2^{\Delta\Delta CT} = 2^{\Delta CT\ sample\ -\Delta CT\ reference}$.

6.2.3 Analyse par immunofluorescence pour le suivi de l'expression de protéines caractéristiques de kératinocytes

Après traitement par le témoin positif, l'extrait de salicaire ou les molécules A-D de l'extrait (voir traitements des monocouches de NHEK en annexe 6.1.1), les monocouches de NHEK sont rincées avec du PBS et fixées. Les cellules sont incubées avec les anticorps anti-KRT10 (SC-23877, SantaCruz Biotechnology, Heidelberg, Allemagne), anti-FLG (SC-66192, SantaCruz Biotechnology, Heidelberg, Allemagne) ou anti-TGK (0175003, Tebubio, Le Perray-en-Yvelines, France). Des anticorps de chèvre-anti souris marqués avec Alexa Fluor®488 sont utilisés comme second anticorps et les noyaux cellulaires sont colorés avec Hoechst 33258.

L'acquisition des images a été réalisée avec un système d'imagerie à haute résolution, INCell Analyzer™1000 (GE Healthcare, Aulnay-sous-Bois, France). Les marquages ont été quantifiés par mesure de l'intensité de fluorescence des protéines rapportée au nombre de noyaux identifiés par le Hoechst (intégration des données numériques par le logiciel Developer Toolbox 1.5, GE Healthcare).

6.2.4 Analyses histologiques des épidermes reconstruits : préparation des coupes transversales et coloration

Après culture et traitement (voir annexe 6.1.2), les RHE sont fixés avec une solution de formaldéhyde, déshydratés par de multiples bains avec des concentrations croissantes d'éthanol puis inclus dans de la paraffine. Des coupes transversales sont réalisées au microtome (5 µm d'épaisseur). Les coupes sont déparaffinées, lavées et incubées avec du peroxyde d'hydrogène pour inhiber la peroxydase endogène. Les coupes sont incubées avec les anticorps anti-KRT10, anti-FLG ou anti-TGK et ensuite un second anticorps botine-conjuguée permet la détection (Dako, Kit LSAB+, K0690, Agilent Technologies, Santa Clara, Etats-Unis). Après l'addition de la streptavidine peroxydase conjuguée (Dako, Kit LSAB+, K0690) et du substrat peroxydase (Dako, Substrate hyper-sensitive AEC+), les noyaux sont contre-colorés grâce à une solution d'hématoxyline. Les coupes sont observées grâce au microscope NIKON E400. La saisie des photographiques est effectuée par la caméra NIKON DS-Ri1 et pilotée avec le logiciel NIS-Elements 3.10.

6.2.5 Analyses histologiques des peaux reconstruites : préparation de coupes transversales et coloration

A la fin de l'incubation de 9 jours des peaux reconstruites (détails en annexe 6.1.3), les FTSK fixées dans une solution de formaldéhyde sont déshydratées par des bains d'éthanol successifs et croissants puis inclus en paraffine. Des coupes transversales sont réalisées au microtome (5 µm d'épaisseur). La paraffine est ensuite retirée et les coupes sont colorées avec des solutions d'hématoxyline, d'éosine et de safran. Après avoir été montées dans "CV Ultra Mounting Medium", les coupes sont observées à l'aide d'un microscope NIKON E400 (Objectif x40). Les images sont enregistrées avec la caméra numérique NIKON DS-Ri1 pilotée par le logiciel NIS-Elements 4.13.04.

6.2.6 Dosage ELISA dans les sous-nageants de peaux reconstruites

Les sous-nageants de cultures cellulaires des peaux reconstruites (culture décrite en annexe 6.1.3) de chaque condition sont collectés et stockés à -20°C jusqu'au moment des analyses. La quantification en ELISA de la libération de MMP-1 (R&D Systems, Lille, France), MMP-3 (R&D Systems) et du procollagène I (Takara, Saint-Germain-en-Laye, France) est réalisée selon des recommandations du fournisseur.

Glorianne JOURAVEL

Stratégies innovantes pour une valorisation d'extraits de plantes en cosmétique : Mise en œuvre d'un outil de profilage métabolique et recherche de nouvelles activités biologiques

Résumé : La cosmétique valorise des extraits de plantes en les intégrant dans des produits de soin. Ces extraits constituent les ingrédients actifs de la formule cosmétique. Les plantes sont diverses, riches et contiennent de nombreux composés d'intérêt biologique. La phytochimie s'intéresse à décrire le contenu métabolique d'extraits de plantes. Mais la caractérisation moléculaire de ces matrices complexes reste un défi majeur aujourd'hui. En effet, les étapes de traitements de données sont longues et laborieuses. Un outil de profilage métabolique, GAINS, a ainsi été développé afin de traiter de façon automatisée les données issues d'analyses en chromatographie liquide couplée à la spectrométrie de masse haute résolution. Il constitue un support au phytochimiste car le traitement automatisé des données permet de gagner du temps par rapport à un traitement manuel. Cet outil associé à une vaste base de données de produits naturels permet d'attribuer des candidats potentiels aux pics détectés. GAINS fait aussi appel à un module de fragmentation *in silico* afin d'étayer les propositions de candidats. Cela permet de comparer le spectre de fragmentation modélisé des candidats avec le spectre de fragmentation expérimental.

L'ensemble des études phytochimiques réalisées pour identifier ou isoler des composés va de pair avec l'étude du potentiel d'effet biologique des extraits sur la peau, organe ciblé par les cosmétiques. Cela sert à découvrir les actions bénéfiques que peut avoir l'extrait. En connaissant le contenu phytochimique, on peut expliquer et rationaliser les tests d'activités biologiques. Le développement d'un actif anti-âge à partir de la salicaire, une plante de la région Centre-Val de Loire, en est l'exemple.

Mots clés : Phytochimie, UHPLC-ESI-HRMS, traitement automatisé de données, déréplication, extrait

Innovative strategies for a valorization of plant extracts in cosmetic: development of a tool for metabolic profiling and search for new biological activities

Summary: The cosmetic field valorizes plant extracts by integrating them in care products. These extracts constitute the active ingredients of the cosmetic formulation. Plants are diverse, rich and contain numerous compounds of biological interest. Phytochemistry is a way to describe the metabolic content of plant extracts. But molecular characterization of these complex matrices remains a major challenge nowadays. Indeed, steps of data treatment are time-consuming and laborious. In this way, a tool of metabolic profiling, GAINS, has been developed in order to treat in an automatized way data from analyses performed in liquid chromatography coupled with high-resolution mass spectrometry. It constitutes a real support for phytochemists because automatized data treatment allows gaining time compared to manual treatment. This tool, associated with a large database of natural compounds make possible to assign potential candidates to detected peaks. GAINS appeals a module of *in silico* fragmentation for holding candidates assignments up. This permits to compare modeled spectrum of fragmentation of candidates with experimental spectrum of fragmentation.

The whole set of phytochemical studies realized to identify or isolate compounds goes hand in hand with the study of potential biological effects of extracts to the skin, targeted organ by skin-care products. This allows the discovery of beneficial actions that the extract could have. By knowing the phytochemical content, it is possible to explain and rationalize assays about biological activities. The development of an anti-aging ingredient from purple loosestrife, a plant occurring in the region Centre-Val de Loire, is an example of it.

Keywords: Phytochemistry, UHPLC-ESI-HRMS, automated data treatment, dereplication, extract

Institut de Chimie Organique et Analytique
Université d'Orléans, UMR 7311, rue de Chartres,
BP 67059, 45067 Orléans Cedex 2, France