AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10
http://www.cfcopies.com/V2/leg/leg_droi.php
http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm

UNIVERSITE DE LORRAINE

Ecole Doctorale 410 Sciences et Ingénierie des Ressources, Procédés, Produits, Environnement

UMR 1121 Université de Lorraine-INRA, Laboratoire Agronomie et Environnement, 2 avenue Forêt de Haye, TSA 40602, 54518 Vandœuvre-lès-Nancy Cedex

THESE

Présentée pour l'obtention du titre de

Docteur de l'Université de Lorraine

En Sciences agronomiques

Par

Célia KRIEGER

Identification moléculaire et caractérisation fonctionnelle d'une nouvelle sous-famille de cytochromes P450, CYP71AZ, impliquée dans la synthèse de furanocoumarines et de coumarines chez *Pastinaca sativa*.

Thèse soutenue publiquement le 16 décembre 2014, devant le jury composé de :

Jean-Louis Hilbert	Professeur, Université de Lille, Lille	Rapporteur
Hubert Schaller	Directeur de Recherche, CNRS-IBMP, Strasbourg	Rapporteur
Eric Gontier	Professeur, Université de Picardie Jules Verne, Amiens	Examinateur
Alain Hehn	Maître de Conférences, Université de Lorraine, Nancy	Directeur de Thèse
Frédéric Bourgaud	Professeur, Université de Lorraine, Nancy	Co-directeur de Thèse
Alexandre Olry	Ingénieur de Recherche, Université de Lorraine, Nancy	Co-directeur de Thèse

A Michel et Paul

Remerciements

J'aimerai exprimer toute ma reconnaissance à M. Jean-Louis Hilbert, professeur à l'université de Lille, et M. Hubert Schaller, directeur de recherche au CNRS à l'IBMP de l'université de Strasbourg, pour avoir accepté d'être les rapporteurs de cette thèse et du temps qu'ils ont consacré à juger mon travail. Je remercie vivement M. Eric Gontier, professeur à l'université de Picardie Jules Verne à Amiens, d'avoir accepté d'être examinateur de cette thèse.

Je tiens à remercier le Pr. Frédéric Bourgaud, directeur du Laboratoire Agronomie et Environnement, de m'avoir accueillie au sein de son équipe. Je remercie mon directeur de thèse, Alain Hehn, qui m'a accordé toute sa confiance en me proposant ce projet de thèse. Merci pour tes conseils, ta disponibilité, ton écoute, ta capacité à me canaliser, ton soutien et encore merci de m'avoir permis de découvrir les joies de l'enseignement. Je remercie également le Dr. Alexandre Olry et le Pr. Frédéric Bourgaud, mes deux co-directeurs de thèse, pour leurs conseils avisés et leur disponibilté. Merci Alex pour tous tes conseils avisés et ton expertise en enzymologie. Je tiens également à exprimer ma gratitude à mes encadrants pour m'avoir impliqué dans le projet Sakura qui m'a permis d'effectuer un mois de travaux de recherche au Japon. A ce titre, je remercie le Pr. Yazaki Kazufumi et le Dr. Sugiyama Akifumi pour leur accueil dans leur structure de recherche, leur disponibilité et leur bienveillance. Je tiens également à remercier tous les stagiaires de leur équipe pour leur gentillesse, leur disponibilité pour gérer les petits tracas de la vie quotidienne et les nombreuses visites que j'ai pu effectuer avec eux.

Je tiens également à remercier toutes les personnes qui ont contribué à ces recherches : le Pr. Ulrich Matern de l'université de Marburg, le Pr. Henrik Toft Simonsen, le Dr. Corinna Weitzel et Bjørn Dueholm de l'université de Copenhague, le Dr. Dave Richie du Loria de Nancy et le Dr. Bernd Schneider de l'institut Max Planck à Jena (Allemagne).

Je souhaite également remercier les différents analystes qui ont travaillé sur ce projet, à savoir Cédric Paris du laboratoire commun d'analyse de l'ENSAIA, Thibault Duval et Jérémy Grosjean. Merci pour votre travail, vos conseils, votre disponibilité et pour le temps que vous m'avez accordé dans l'éclaircissement du fonctionnement et démontage/remontage de ces machines.

Je remercie également tout le personnel technique du laboratoire, Aude, Elodie, Joffrey et Claude, pour leur aide précieuse dans la préparation du matériel au laboratoire, la mise en place de la qualité, de l'hygiène et sécurité, de la gestion des déchets. Un grand merci à Elodie Warisse pour son soutien et son aide dans l'échantillonage des expériences d'élicitation qui restera mémorable, car ce n'est pas tous les jours qu'on effectue des pesées à la lampe torche au beau milieu de la nuit !

Je tiens également à montrer ma gratitude à mes anciens encadrants, le directeur de recherche Didier Merdinoglu et le Dr. Paule Blasi de l'équipe Génétique et Amélioration de la Vigne de l'INRA de Colmar, le Pr. Moussa Sié et le Dr. Moukoumbi Yonnelle Déa de l'ADRAO à Cotonou (Bénin), la directrice de l'équipe Gene Mapping de Plant and Food Research (Nouvelle-Zélande) Suzanne Gardiner, le Dr. David Chagné et le Dr. Tony McGhie. Merci pour votre soutien, vos conseils, votre

accompagnement, votre disponibilité et votre bonne humeur qui m'ont donné l'envie de poursuivre vers une thèse. Merci du fond du cœur à toutes ces équipes pour leur convivialité et leur bonne humeur.

J'aimerai remercier les stagiaires que j'ai eu le plaisir d'encadrer au cours de ce doctorat et qui ont contribué à l'avancée de ces travaux, Cécile Befve, Livia Réa, Ann-Kathrin Leonhard et Mohamed Mouloud Iddir.

Je tiens également à remercier l'ensemble des personnes que j'ai pu côtoyer au cours de cette thèse, notamment les permanents de notre laboratoire, Christophe, Romain, Jean, Nadia, Sylvain, Françoise, Chhoy, Armand, Benoît et Béatrice, mais également Cédric, Manu, Xavier, Arnaud, Annelore et Julien pour nos cessions « gâteaux » du lundi matin. Je remercie Thamara pour ses conseils avisés concernant les procédures administratives plus qu'obscures.

J'adresse un grand merci aux jeunes chercheurs fraîchement ou prochaînement diplômés à savoir, Guilhem, Fazeelat, Mathilde, Nicolas et Victoire avec qui j'ai partagé notre magnifique bureau, mais également Sissi, Estelle, Flore, Corinna et Sandy pour nos nombreux échanges scientifiques, et Aline, Simon, Abdelhak, Linde, Frédéric et Charles à qui j'adresse tous mes vœux de réussite pour la suite de leur carrière.

Un grand merci tout particulièrement à Aline, Natalia, Thibault, Victoire, Simon et Nicolas, plus que des collègues, vous êtes devenus des amis très chers. Je vous remercie pour votre écoute, votre réconfort, votre présence, votre gentillesse et pour tous ces moments de complicité et de bonne humeur que nous avons partagé. Merci également à Cédric et Yohann de La Flambée, QG de thèse, pour votre accueil, votre bonne humeur et tout le réconfort culinaire que vous nous avez apporté ! Je remercie également tous mes amis de Strasbourg et d'ailleurs qui ont toujours été présents au cours de ces nombreuses années.

J'exprime enfin, mes plus tendres remerciements à ma famille et plus particulièrement à mes parents pour leur soutien constant, leurs encouragements et leur générosité. Merci de m'avoir permis d'aller aussi loin en dépit de votre réticence à mon inscription en faculté après le baccalauréat ! A Damien, je te remercie de tout cœur d'avoir été à mes côtés dans les moments les plus difficiles. Merci pour ton soutien, ta patience et tes encouragements. Nous nous sommes rencontrés dans une période qui fut certainement la plus éprouvante de mon existence et nous tenons bon. C'est une page qui se tourne pour notre plus grand bonheur à tous les deux.

Sommaire

Remerciements 5
Sommaire 7
Liste des figures 15
Liste des tableaux 18
Liste des photographies 19
Liste des abréviations 20
Noms vernaculaires des espèces citées 23

PARTIE 1 : Synthèse bibliographique et présentation des objectifs de cette thèse 25
1 Phénylpropanoïdes 27
1.1 Phénylpropanoïdes 27
1.1.1 Généralités 27
1.1.2 Voie de biosynthèse 28
1.2 Coumarines 30
1.2.1 Généralités 30
1.2.2 Rôle physiologique des coumarines 31
1.2.3 Voie de biosynthèse des coumarines 32
1.3 Furanocoumarines 35
1.3.1 Structure et propriétés chimiques 35
1.3.2 Distribution dans le règne végétal 35
1.3.3 Localisation au sein de la plante et stockage dans la cellule 37
1.3.4 Rôles des furanocoumarines 37
1.3.5 Propriétés biologiques des furanocoumarines 39
1.3.5.1 Réaction de photooxydation et photolyse 39
1.3.5.2 Réaction de photodimérisation et photocycloaddition 40
1.3.5.3 Interactions avec les acides nucléiques 41
1.3.5.4 Interactions avec les acides gras insaturés 43
1.3.5.5 Interactions avec les protéines 44
1.3.6 Voie de biosynthèse des furanocoumarines 44
1.3.7 Utilisations thérapeutiques 47
1.3.7.1 La photochimiothérapie 47
1.3.7.2 La photophérèse 48

1.3.7.3 Traitement antidépresseur 48
1.3.7.4 Activités bactériostatiques, fongicides et antimycobactérienne 48
1.3.7.5 Activité antiproliférative et apoptotique 49
1.3.8 Toxicité des furanocoumarines 50
1.3.8.1 Photodermatite 50
1.3.8.2 Mélanome 50
1.3.8.3 Par ingestion 50
2 *Pastinaca sativa*, *Ruta graveolens* et *Cullen cinereum* : trois plantes productrices de furanocoumarines 51
2.1 Description botanique 51
2.1.1 *P. sativa* 51
2.1.2 *R. graveolens* 53
2.1.3 *C. cinereum* 54
2.2 Intérêt de ces trois plantes pour l'étude de la voie de biosynthèse des furanocoumarines 55
3 Les cytochromes P450 56
3.1 Description 56
3.2 Nomenclature 58
3.3 Classification et origine du pouvoir réducteur 58
3.4 Structures secondaire et tertiaire et domaines conservés 60
3.5 Cycle catalytique des P450s 63
3.6 P450s et inactivation autocatalytique 65
3.7 Rôles et intérêts des P450s 65
3.7.1 Fonctions des P450s chez les végétaux 66
3.7.1.1 Biosynthèse de métabolites secondaires 66
3.7.1.2 Catabolisme de xénobiotiques 67
3.7.2 Les cytochromes P450 à l'origine de l'adaptation entre les plantes et les insectes 68
3.7.2.1 Adaptation et synthèse de composés allélopathiques par les P450s de plantes 68
3.7.2.2 Adaptation et détoxication de toxines par les P450s d'insectes 70
3.7.2.3 P450s et co-évolution 71
4 Analyse de transcriptomes 72
4.1 Méthode de séquencage « traditionnelle » de Sanger 72
4.2 Séquençage haut débit 73
4.2.1 Technologie 454 (Roche) 73
4.2.2 Technologie Illumina Solexa 75
4.2.3 Technologie Helicos tSMS 75

4.2.4 Technologie Applied Biosystems SOLiD 76
4.2.5 Et dans le futur ? 76
4.3 Assemblage de séquences 76
4.4 Annotations de séquences 77
4.4.1 Annotation structurale 78
4.4.2 Annotation fonctionnelle 79
4.4.2.1 Annotation de définition fonctionnelle 79
4.4.2.2 Annotation de différents signaux 80
4.4.2.3 Annotation de domaines protéiques 80
4.4.2.4 Annotation Enzyme Commission 81
4.4.2.5 Annotation Gene Ontology 81
5 Objectifs de la thèse 82

PARTIE 2 : Matériel et Méthodes 85
1 Matériel 87
1.1 Matériel végétal 87
1.1.1 *P. sativa* 87
1.1.2 *R. graveolens* 87
1.1.3 *C. cinereum* 87
1.1.4 *Nicotiana benthamiana* 87
1.2 Souches bactériennes 87
1.2.1 *Escherichia coli* Top 10® (Invitrogen) 87
1.2.2 *E. coli* GenHogs® (Invitrogen) 88
1.2.3 *E. coli* XL1 Blue (Stratagene) 88
1.2.4 *E. coli* MC1022 88
1.2.5 *E. coli ccd*B Survival™ (Invitrogen) 88
1.2.6 *Agrobacterium tumefaciens* LBA4404 (pAL4404) 89
1.3 Souches de levures 89
1.3.1 *S. cerevisiae* WAT11 89
1.3.2 *S. cerevisiae* WAT21 89
1.4 Vecteurs et plasmides recombinants 89
1.4.1 pCR8®/GW/TOPO® (Invitrogen) 89
1.4.2 p*Ye*DP60 91
1.4.3 p*Ye*DP60_GW® 92
1.4.4 pBIN_GW® 92

2 Milieux de culture 93
2.1 Milieux de culture pour bactéries 93
2.1.1 LB liquide et solide 93
2.1.2 YEB 94
2.1.3 Antibiotiques utilisés 94
2.2 Milieux de culture pour levures 95
2.2.1 YPGA 95
2.2.2 SGI 95
2.2.3 Milieu YPGE 95
2.2.4 Milieu YPL 95
2.3 Culture de plantes en terre 96
2.4 Culture de panais en hydroponie 96
2.4.1 Induction aux UV 97
2.4.2 Induction au méthyljasmonate 98
3 Méthodes de bioinformatique 98
3.1 Utilisation de la suite logicielle blast2GO 98
3.2 Utilisation du logiciel BioEdit 100
3.3 Génération d'arbres phylogénétiques avec le logiciel MEGA6 100
4 Méthodes de biologie moléculaire 100
4.1 Extraction d'ADN génomique de plante 100
4.2 Extraction d'ARN totaux de plante 100
4.3 Synthèse des ADNc simple brin par transcription inverse (RT) 101
4.4 Amplification d'un fragment d'ADN par PCR 101
4.4.1 Amplification classique 101
4.4.2 TAIL-PCR 103
4.5 Electrophorèse d'ADN sur gel d'agarose 106
4.6 Extraction d'ADN à partir d'un gel d'agarose 107
4.7 Digestion par des enzymes de restriction 107
4.8 Ligations 107
4.8.1 Dans le vecteur pCR8®/GW/TOPO® 107
4.8.2 Dans les autres vecteurs 108
4.9 Construction d'un plasmide recombinant pour la transformation génétique de levures en utilisant la technologie Gateway® 108
4.10 Obtention des plasmides recombinants p*Ye*PD60_GW®_X et pBIN_GW®_X *via* la technologie Gateway® 109

4.11 Préparation de bactéries compétentes 110
4.11.1 Bactéries *E. coli* électrocompétentes 110
4.11.2 Agrobactéries *A. tumefaciens* chimiocompétentes 111
4.12 Transformation génétique de bactéries compétentes 111
4.12.1 Par électroporation 111
4.12.2 Par choc thermique 111
4.12.2.1 *E. coli* 111
4.12.2.2 *A. tumefaciens* 112
4.13 Extraction d'ADN plasmidique 112
4.14 Construction des trois banques d'ADNc normalisées 112
4.15 Séquençage 113
4.16 Electrophorèse sur gel polyacrylamide en conditions dénaturantes ou SDS PAGE 113
4.17 Western blot 114
4.18 Quantification du niveau d'expression des gènes par PCR en temps réel 114
4.18.1 Quantification relative 114
4.18.2 Préparation du matériel 115
4.18.3 Mélange réactionnel 115
5 Systèmes d'expression hétérologue 115
5.1 Système d'expression hétérologue dans les cellules eucaryotes de type levure 115
5.1.1 Préparation de levures compétentes 115
5.1.2 Transformation de levures 116
5.1.3 Expression des P450s dans les levures 116
5.1.3.1 Conditions de culture 116
5.1.3.2 Préparation de microsomes 117
5.2 Système d'expression hétérologue dans les plantes 118
5.2.1 Principe de l'expression transitoire 118
5.2.2 Préparation de l'inoculum bactérien 118
5.2.3 Agro-infiltration de *N. benthamiana* 118
5.2.4 Préparation de microsomes de plantes 119
6 Méthodes d'analyses biochimiques 119
6.1 Extraction de composés polyphénoliques 119
6.1.1 Protocole général 119
6.1.2 Protocole utilisé lors de la deuxième expérience d'élicitation 120
6.2 Mesure d'activités enzymatiques 120
6.2.1 Substrats 120

6.2.2 Détermination des constantes cinétiques 120
6.2.3 Test de l'inhibition par des furanocoumarines 121
6.3 Détection des produits formés par HPLC 121
6.4 Identification des produits formés par spectrométrie de masse 122
6.5 Dosage de furanocoumarines par LC-MS/MS 122
7 Traitement statistique des données 122

PARTIE 3 : Résultats et discussion 123

CHAPITRE 1 : CONSTRUCTION ET ANALYSE DE 3 BANQUES TRANSCRIPTOMIQUES DE PLANTES PRODUCTRICES DE FURANOCOUMARINES EN VUE D'IDENTIFIER DE NOUVEAUX GENES CANDIDATS IMPLIQUES DANS LA SYNTHESE DE FURANOCOUMARINES 125

1 Exploitation de trois banques normalisées d'ADNc de plantes produisant des furanocoumarines 125
1.1 Recherche de gènes candidats 125
1.1.1 Approche bibliographique 125
1.1.2 Approche expérimentale 126
1.1.3 Approche bioinformatique 128
1.1.3.1 Recherche globale : utilisation de la suite logicielle blast2GO 131
1.1.3.2 Identification des séquences nucléotidiques codant pour des cytochromes P450 136
1.1.3.3 Classification des séquences codant pour des P450s 140
1.1.3.4 Identification des séquences codant pour des cytochromes P450 NADPH réductase 142
1.2 Discussion et conclusion 146

CHAPITRE 2 : ETUDE ET CARACTERISATION FONCTIONNELLE D'ENZYMES DE LA SOUS-FAMILLE CYP71AZ CHEZ *PASTINACA SATIVA* 149

2.1 Identification des séquences correspondant à des CYP71AZ à partir de la banque d'ADNc de *P. sativa* 149
2.2 CYP71AZ3 et CYP71AZ4 : analyse fine des séquences protéiques 151
2.3 Analyse du profil d'expression de ces gènes en fonction des organes de la plante 155
2.3.1 Analyse du niveau d'expression de CYP71AZ3 et 4 dans différents organes de la plante . 156
2.3.2 Tentative d'analyse du niveau d'expression de CYP71AZ3 et 4 en réponse à un stress éliciteur de la synthèse de furanocoumarines 158
2.4 Expression hétérologue de P450s de *P. sativa* dans un système eucaryote : la levure 163

2.4.1 « *Amélioration* » du système d'expression de protéines développé par le Dr. Pompon..... 164
2.4.1.1 Introduction de la technologie Gateway® dans le vecteur p*Ye*DP60..... 164
2.4.1.2 Validation de la fonctionnalité du plasmide p*Ye*DP60_GW® 166
2.4.1.3 Utilisation du système pour l'expression des CYP71AZ de *P. sativa* 167
2.4.2 Caractérisation fonctionnelle de CYP71AZ3 et CYP71AZ4 *in vitro* 170
2.4.2.1 Criblage fonctionnel..... 170
2.4.2.2 Détermination des constantes cinétiques de CYP71AZ3 et CYP71AZ4..... 183
2.4.2.2.1 Détermination du pH optimal d'activité..... 183
2.4.2.2.2 Détermination de la température optimale d'activité 184
2.4.2.2.3 Linéarité de la réaction..... 186
2.4.2.2.4 Constantes cinétiques 189
2.4.2.2.5 Recherche d'une inhibition compétitive de CYP71AZ3 et CYP71AZ4 par des coumarines et furanocoumarines 192
2.5 Caractérisation fonctionnelle de CYP71AZ3 et CYP71AZ4 *in vivo* 194
2.5.1 Expression transitoire de CYP71AZ3 et CYP71AZ4 dans un système d'expression *in planta* : *N. benthamiana*..... 194
2.5.1.1 Test préliminaire avec CYP71AJ1 et l'ATR1 d'*A. thaliana*..... 194
2.5.1.2 Infiltration de CYP71AZ3 et CYP71AZ4..... 195
2.5.2 Extraction des polyphénols et identification par spectrométrie de masse 197
2.6 Discussion 197
2.7 Conclusion..... 201

CHAPITRE 3 : IDENTIFICATION DE VARIANTS DE CYP71AJ CHEZ *PASTINACA SATIVA* 203

3.1 Identification de nouveaux P450s appartenant à la sous-famille CYP71AJ 203
3.2 Recherche de gènes complets correspondant aux séquences partielles présentes dans la banque de *P. sativa* : approche TAIL-PCR 209
3.3 Recherche de la fonction de CYP71AJ13 212
3.3.1 Clonage et expression chez la levure..... 213
3.3.2 Criblage fonctionnel..... 215
3.3.3 Tentative de compréhension de la fonction par une approche d'analyse bio-informatique 216
3.4 Discussion 220
3.5 Conclusion..... 226

CONCLUSION GENERALE ET PERSPECTIVES 227

REFERENCES BIBLIOGRAPHIQUES 237
ANNEXES 271

Liste des figures

Figure 1 : Squelette carboné en C6-C3 à l'origine des phénylpropanoïdes simples. 27
Figure 2 : Représentation schématique simplifiée de la voie de biosynthèse des phénylpropanoïdes et de ses principaux dérivés..... 29
Figure 3 : Représentation de la structure chimique développée de base des coumarines et de quelques dérivés. 30
Figure 4 : Classification des différents types de coumarines. 31
Figure 5 : Voie de biosynthèse des coumarines, formations de l'umbelliférone, de l'esculétine et de la scopolétine..... 34
Figure 6 : Exemples de structures chimiques de furanocoumarines linéaires et angulaires..... 35
Figure 7 : Exemple de plantes appartenant aux quatre familles botaniques produisant des furanocoumarines en grande quantité..... 36
Figure 8 : Réactions de photolyse et de photooxydation causées par les furanocoumarines. 39
Figure 9 : Réactions de photodimérisation de composés coumariniques..... 40
Figure 10 : Réactions de photocycloaddition du psoralène avec un groupement éthylène substitué... 41
Figure 11 : Représentation schématique de la photocycloaddition du 8-méthoxypsoralène (8-MOP) sur les thymines des deux brins d'ADN. 42
Figure 12 : Représentation schématique de l'implication des photocycloadditions de furanocoumarines avec des acides gras insaturés dans l'activation de la mélanogénèse. 43
Figure 13 : Voie de biosynthèse hypothétique des furanocoumarines linéaires et angulaires.. 46
Figure 14 : Structure de la paradisine A..... 51
Figure 15 : A. Planche botanique de *P. sativa* L. B. Photographie de *P. sativa*. 52
Figure 16 : A. Planche botanique de *R. graveolens* L. B. Photographie de *R. graveolens* L..... 53
Figure 17 : Photographies de *C. cinereum*. 55
Figure 18 : Exemples de réactions catalysées par les P450s..... 57
Figure 19 : Equation bilan de la réaction de monooxygénation catalysée par les P450s. 58
Figure 20 : Représentation schématique des différents partenaires redox des P450s. 59
Figure 21 : Représentation schématique d'un cytochrome P450 eucaryote microsomal associé à la cytochrome P450 réductase constituée d'un domaine FAD et d'un domaine FMN. 60
Figure 22 : Représentation schématique de la structure primaire d'un P450 membranaire eucaryote. 61
Figure 23 : Représentation de la structure tridimensionnelle de CYP2D6 complexé avec du dextrométhorphane (en vert) obtenues à partir des modèles cristallographiques de CYP101, 102, 107A, 108 et 2C5. 62
Figure 24 : Structure chimique développée d'un hème de type protoporphyrine IX. 63
Figure 25 : Représentation schématique du cycle catalytique d'un P450..... 64
Figure 26 : Représentation schématique des différentes étapes du pyroséquençage 454 de Roche..... 74
Figure 27 : Représentation schématique du processus d'assemblage de séquences. 77
Figure 28 : Représentation schématique du vecteur pCR8®/GW/TOPO®. 90
Figure 29 : Polylinker du vecteur pCR8®/GW/TOPO®..... 90
Figure 30 : Représentation schématique du vecteur d'expression dans la levure pYeDP60..... 91
Figure 31 : Représentation schématique du vecteur pBIN_GW®. 93
Figure 32 : Schéma illustrant le principe de la TAIL-PCR..... 103
Figure 33 : Représentation schématique des trois cassettes Gateway®..... 109
Figure 34 : Représentation schématique de la LR recombinaison via la technologie Gateway®..... 110
Figure 35 : Comparaison des mécanismes réactionnels impliqués dans la synthèse du menthofurane (A), de la marmésine (B) et de la sécologanine (C). 127
Figure 36 : Analyse sur gel d'agarose à 1 % des ARN totaux extraits de *R. graveolens* 129

Figure 37 : Analyse GO de la banque d'ADNc de *P. sativa* (A), *R. graveolens* (B) et *C. cinereum* (C). 132
Figure 38 : Analyse des fonctions moléculaires associées aux différents contigs de *P. sativa* (A), *R. graveolens* (B) et *C. cinereum* (C). 134
Figure 39 : Analyse des processus biologiques associés aux différents contigs de *P. sativa* (A), *R. graveolens* (B) et *C. cinereum* (C). 135
Figure 40 : Alignement des structures primaires partielles des CYP71AZ identifiés à partir de la banque d'ADNc de *P. sativa* avec celle de CYP71AZ1 d'*A. majus* 150
Figure 41 : Arbre phylogénétique des différents membres de CYP71AJ et CYP71AZ.. 152
Figure 42 : Alignement des séquences peptidiques de CYP71AZ1 (*A. majus*), CYP71AZ3 (*P. sativa*) et CYP71AZ4 (*P. sativa*) 153
Figure 43 : Comparaison de l'expression relative de CYP71AJ1 par rapport au profil d'accumulation de psoralène après induction dans le milieu de culture de cellules d'*A. majus* par des extraits de *Pmg*. . 155
Figure 44 : Quantification des niveaux d'expression de CYP71AJ3-4, CYP71AZ3-4, d'une prényltransférase et d'une dioxygénase réalisée sur les feuilles, les tiges, les racines et les graines de *P. sativa*. 158
Figure 45 : Traitement des feuilles de *P. sativa* aux UVB. 159
Figure 46 : Traitement des racines de *P. sativa* aux UVB. 161
Figure 47 : Traitement des feuilles de *P. sativa* au méthyl jasmonate. 160
Figure 48 : Traitement des racines de *P. sativa* au méthyljasmontate 162
Figure 49 : (A) Profils de digestions théoriques des plasmides pYeDP60_GW® ayant insérés la cassette de clonage RfA dans la bonne orientation (directe) et dans la mauvaise orientation (inverse) digérés par les enzymes de restriction EcoRI et MluI, comparés au marqueur de taille MassRuler DNA Ladder Mix, ready-to-use (Fermentas). (B) Profils de digestions de six plasmides supposés pYeDP60_GW® digérés par les enzymes de restriction EcoRI et MluI (puits 1 à 6) comparés au maqueur de taille MassRuler DNA Ladder Mix, ready-to-use (Fermentas). 165
Figure 50 : Comparaison des chromatogrammes extraits à 300 nm, issus du test de métabolisation de CYP73A1 de topinambour avec de l'acide cinnamique en présence de NADPH. 167
Figure 51 : Détection de la production des protéines CYP71AZ4 (A) et CYP71AZ3 (B) dans le système levure WAT11 par western blot après séparation en SDS-PAGE en comparaison avec le marqueur de taille Prestained Protein Molecular Weight Marker (ready-to-use) (Fermentas). 170
Figure 52 : Liste des furanocoumarines linéaires (A) et angulaires (B) utilisées pour le premier criblage métabolique réalisé avec CYP71AZ3 et CYP71AZ4. 172
Figure 53 : Détection de la formation de xanthotoxol à partir du psoralène catalysée par CYP71AZ4. 173
Figure 54 : Liste des substrats utilisés au cours du deuxième criblage métabolique réalisé avec CYP71AZ3 et CYP71AZ4 174
Figure 55 : Détection de la formation de fraxétine à partir de la scopolétine catalysée par CYP71AZ4. 176
Figure 56 : Détection de la formation de daphnétine 7-méthyléther à partir de la 7-méthoxycoumarine catalysée par CYP71AZ4. 177
Figure 57 : Détection de la formation de la scopolétine à partir de la 6-méthoxycoumarine catalysée par CYP71AZ4. 178
Figure 58 : Détection du deuxième produit de la conversion de la 6-méthoxycoumarine catalysée par CYP71AZ4. 179
Figure 59 : Détection du produit de la conversion de la 7-méthoxy 3-méthylcoumarine catalysée par CYP71AZ4. 181
Figure 60 : Détection du produit de la conversion de l'esculétine catalysée par CYP71AZ3. 182
Figure 61 : Détermination du pH optimum pour l'activité de la psoralène 8-monooxygénase. 184
Figure 62 : Détermination du pH optimum d'activité de CYP71AZ3 en présence d'esculétine. 184

Figure 63 : Détermination de la température optimale d'activité de CYP71AZ4 en présence de psoralène..... 185
Figure 64 : Détermination de la température optimale d'activité de CYP71AZ3 en présence d'esculétine. 185
Figure 65 : Accumulation de xanthotoxol à partir de psoralène par CYP71AZ4 en fonction du temps. 186
Figure 66 : Accumulation de fraxétine à partir de scopolétine par CYP71AZ4 en fonction du temps. 187
Figure 67 : Accumulation de daphnétine 7-méthyléther à partir de 7-méthoxycoumarine par CYP71AZ4 en fonction du temps. 187
Figure 68 : Accumulation de scopolétine à partir de 6-méthoxycoumarine par CYP71AZ4 en fonction du temps. 187
Figure 69 : Accumulation du deuxième produit formé par CYP71AZ4 à partir de la 6-méthoxycoumarine en fonction du temps..... 188
Figure 70 : Accumulation du produit formé par CYP71AZ4 à partir de la 7-méthoxy 3-méthylcoumarine en fonction du temps. 188
Figure 71 : Accumulation du produit formé par CYP71AZ3 à partir de l'esculétine en fonction du temps. 188
Figure 72 : Quantité de xanthotoxol formé par CYP71AZ4 en fonction de la concentration en psoralène. 190
Figure 73 : Quantité de fraxétine formée par CYP71AZ4 en fonction de la concentration en scopolétine. 190
Figure 74 : Quantité de daphnétine 7-méthyléther formée par CYP71AZ4 en fonction de la concentration en 7-méthoxycoumarine. 190
Figure 75 : Quantité de scopolétine formée par CYP71AZ4 en fonction de la concentration en 6-méthoxycoumarine. 191
Figure 76 : Quantité de produit inconnu formé par CYP71AZ4 en fonction de la concentration en 6-méthoxycoumarine. 191
Figure 77 : Quantité de produit inconnu formé par CYP71AZ4 en fonction de la concentration en 7-méthoxy 3-méthylcoumarine..... 191
Figure 78 : Quantité de produit inconnu formé par CYP71AZ3 en fonction de la concentration en esculétine. 192
Figure 79 : Alignement des séquences peptidiques de CYP71AZ1 (*A. majus*), CYP71AZ3 (*P. sativa*), CYP71AZ4 (*P. sativa*), CYP71AZ5 (*P. sativa*) et CYP71AZ6 (*P. sativa*). 199
Figure 80 : Alignement des séquences peptidiques partielles des CYP71AJ identifiés à partir de la banque d'ADNc de *P. sativa* avec celles de CYP71AJ3 et CYP71AJ4 de *P. sativa*. 205
Figure 81 : Représentation schématique des séquences partielles de CYP71AJ identifiées dans la banque d'ADNc de *P. sativa*. 207
Figure 82 : Analyse sur gel d'agarose à 1 % des produits de PCR réalisés avec les différents couples d'amorces spécifiques des fragments CYP71AJ isolés à partir de la banque d'ADNc de *P. sativa*... 208
Figure 83 : Vérification de l'intégrité des ARNm extraits à partir des feuilles de *P. sativa* à l'aide du kit RNeasy® Plant Mini (Qiagen). 213
Figure 84 : Amplification de la séquence codante de CYP71AJ13 isolé à partir de *P. sativa.*..... 214
Figure 85 : Détection de la production de la protéine CYP71AJ13 dans le système levure WAT11 par western blot après séparation en SDS-PAGE en comparaison avec le marqueur de taille Prestained Protein Molecular Weight Marker (ready-to-use) (Fermentas)..... 215
Figure 86 : Alignement des séquences peptidiques de CYP71AJ1 (*Ammi majus*), CYP71AJ2 (*Apium graveolens*), CYP71AJ3 (*P. sativa*), CYP71AJ4 (*P. sativa*) et CYP71AJ13 (*P. sativa*)..... 219
Figure 87 : Arbre phylogénétique des différents membres de CYP71AJ. 223
Figure 88 : Arbre phylogénétique de différentes Apioidées et occurrence de la sous-famille CYP71AJ. 225

Liste des tableaux

Tableau 1 : Composition du milieu YEB. 94
Tableau 2 : Concentrations des solutions stocks et des solutions finales d'antibiotiques et solvants utilisés. 94
Tableau 3 : Composition des différents milieux de culture pour la levure. 96
Tableau 4 : Termes IPR utilisés pour identifier les séquences correspondantes à des cytochromes P450 et des cytochromes P450 réductases (CPRs) contenues dans les banques d'ADNc de *P. sativa*, *R. graveolens* et *C. cinereum*. 99
Tableau 5 : Programmes utilisés pour les réactions de TAIL-PCR. 104
Tableau 6 : Amorces utilisées dans les réactions de TAIL-PCR. 105
Tableau 7 : Composition des tampons utilisés pour la préparation des microsomes de levures. 117
Tableau 8 : Bilan du séquençage et de l'assemblage des trois banques d'ADNc. 130
Tableau 9 : Caractéristiques des séquences identifiées à partir des trois banques d'ADNc codant pour des P450s. 137
Tableau 10 : Annotations des termes IPR obtenus pour les cytochromes P450 identifiés à partir des trois banques d'ADNc. 139
Tableau 11 : Caractéristiques des séquences identifiées à partir des trois banques d'ADNc codant pour des CPRs. 143
Tableau 12 : Annotations des champs de Gene Ontology obtenus pour les séquences correspondant à des CPRs identifiées à partir des trois banques d'ADNc. 143
Tableau 13 : Annotations des termes IPR obtenus pour les séquences codant pour des CPRs identifiées à partir des trois banques d'ADNc. 144
Tableau 14 : Comparaison des pourcentages d'identité (en vert) et d'homologie (en violet) nucléotidiques (A) et peptidiques (B) entre CYP71AZ3 et CYP71AZ4 identifiés dans la banque d'ADNc de *P. sativa* et CYP71AZ1 isolé chez *A. majus*. 151
Tableau 15 : Pourcentages d'identité des séquences peptidiques codant pour les potentiels SRSs et les séquences complètes des membres de CYP71AJ et CYP71AZ. 154
Tableau 16 : Amorces utilisées pour l'amplification des séquences codantes de CYP71AZ3 et CYP71AZ4 de *P. sativa*. 168
Tableau 17 : Constantes d'affinités de CYP71AZ3 et de CYP71AZ4 calculées à l'aide du logiciel Sigmaplot. 189
Tableau 18 : Recherche d'inhibition compétitive des coumarines et furanocoumarines sur CYP71AZ3 et CYP71AZ4. 193
Tableau 19 : Contigs correspondant à des P450s appartenant à la sous-famille CYP71AJ identifiés chez *P. sativa* par *tblastn* en utilisant comme sondes les séquences peptidiques de CYP71AJ3 et CYP71AJ4 de *P. sativa*. 203
Tableau 20 : Pourcentages d'identité (en noir) et d'homologie (en rouge) peptidiques entre CYP71AJ1 d'*A. majus*, CYP71AJ2 d'*Apium graveolens*, CYP71AJ3 et 4 isolés chez *P. sativa* par Larbat et ses collaborateurs (Larbat et al., 2007, 2009), et les séquences candidates identifiées dans la banque d'ADNc de *P. sativa*. 204
Tableau 21 : Amorces utilisées dans les réactions de PCR pour déterminer si les différents fragments de CYP71AJ identifiés dans la banque d'ADNc de *P. sativa* appartiennent au même gène ou non. . 207
Tableau 22 : Comparaison des pourcentages d'identité (en vert) et d'homologie (en violet) nucléotidiques (A) et peptidiques (B) entre CYP71AJ13 identifié dans la banque d'ADNc de *P. sativa* et CYP71AJ1 (*A. majus*), CYP71AJ2 (*Apium graveolens*), CYP71AJ3 et CYP71AJ4 (*P. sativa*).... 216
Tableau 23 : Pourcentages d'identité des séquences peptidiques codant pour les potentiels SRSs des membres de CYP71AJ. 217

Liste des photographies

Photographie 1 : Dispositif de culture hydroponique pour *Pastinaca sativa*. 97
Photographie 2 : Dispositif d'induction des plants de *P. sativa* aux UVB. 98
Photographie 3 : Expression transitoire de P450s chez *N. benthamiana* avec co-infiltration de substrats. 119

Liste des abréviations

4CL : *p*-Coumarate:CoA-Ligase
5-MOP : Bergaptène
8-MOP : Xanthotoxine
aa : Acides Aminés
AD : Amorce Dégénérée
ADN : Acide DésoxyriboNucléique
ADNc : ADN complémentaires
ADNg : ADN génomique
ARNm : ARN messagers
ARNr : ARN ribosomaux
ARNt : ARN de transfert
ATP : Adénosine Tri Phosphate
BAC : Bacterial Artificial Chromosome
BCIP : 5-bromo-4-chloro-3-indole-phosphate
Bcl : B-cell lymphoma
BEt : Bromure d'Ethidium
Blast : Basic Local Alignment Search Tool
BOMT : Bergaptol-*O*-MéthylTransférase
C2'H : *p*-Coumaroyl CoA 2'-Hydroxylase
C3'H : *p*-Coumaroyl shikimate/quinate 3'-Hydroxylase
C4H ou CA4H : Cinnamate-4-Hydroxylase
Caco-2 : Colorectal AdenoCarcinoma cells
CCD : Charged-Coupled Divice = Dispositif à Transfert de Charge
CCoAMT1 : Caféoyl CoA *O*-MéthylTransférase 1
CDS : Coding Sequence
cHH : *cis* head-to-head
cHT : *cis*-head-to-tail
CIAP : phosphatase alcaline de veau
CNRGV : Centre National de Ressources Génétiques Végétales
CNRS : Centre National de la Recherche Scientifique
CPR : Cytochrome P450 Réductase
CYP : CYtochrome P450
DAG : Directed Acyclic Graph
DD : Degré de Dégénérescence
ddNTP : (ddATP, ddGTP, ddTTP, ddCTP) didésoxiribonucléotides triphosphates
DIMBOA : 2,4-dihydroxy-7-méthoxy-2H-1,4-benzoxazin-3(4H)-one
DMAPP : diméthylallyl pyrophosphate
DMS : déméthylsubérosine
DMSO : diméthyl sulfoxide
dNTP : (dATP, dGTP, dTTP, dCTP) désoxyribonucléotides triphosphates
DO : Densité Optique
DTT : dithiothréitol
EC : Enzyme Commission
EDTA : Ethylène Diamine Tétra Acétique
emPCR : PCR en émulsion
ENSAIA : Ecole Nationale Supérieure d'Agronomie et des Industries Alimentaires
EORs : Espèces Oxygénées Réactives
E-value : Expected Value
F6'H : Féruloyl CoA 6'-Hydroxylase

FAD : Flavine Adénine Dinucléotide
FMN : Flavine MonoNucléotide
Fx : Ferrédoxine
FxR : Ferrédoxine réductase
GO : Gene Ontology
GOA : Gene Ontology Anotation
GW : Gateway
HCT : HydroxyCinnamoyl CoA shikimate/quinate hydroxycinnamoyl Transférase
HL-60 : Human promyelocytic Leukemia cells
HPA : hypothalamo-pituitaro-surrénalien
HPLC : High Performance Liquid Chromatography
HQT : hydroxycinnamoyl CoA quinate hydroxycinnamoyl transférase
INRA : Institut National de la Recherche Agronomique
IPR : InterPRo
ITS : Internal Transcribed Spacer
K_m : constante d'affinité
LAE : Laboratoire Agronomie et Environnement
LB : Luria-Bertani
LC : Liquid Chromatography
Loria : Laboratoire lorrain de Recherche en Informatique et ses Applications
LTQ : Linear Trap Quadrupole
MeJa : MéthylJasmonate
MIB : Mechanism-Based Inactivors
MS : Mass Spectrometry
NADH : Nicotiamide Adénine Dinucléotide
NADPH : Nicotiamide Adénine Dinucléotide Phosphate
NaPi : phosphate de sodium
NBT : 4-nitroblue tétrazolium chloride
NCBI nr : National Center for Biotechnology Information Non-Redundant
Nt : Nucléotide
ORF : Open Reading Frame
P450 : cytochromes P450
P450nor : Nitrite Oxyde Réductase
PACBIO : Pacific Biosciences
PAL : phénylalanine ammonia lyase
pb : Paire de Bases
PCR : Réaction en Chaîne par Polymérase
PKC : Protéine Kinase C
Pmg : *Phytophtora megasperma*
PPi : PyroPhosphate Inorganique
PPO : PolyPhénol Oxidase
P-UVA thérapie : photochimiothérapie
RACE-PCR : Rapid Amplification of cDNA Ends
RMN : Résonnance Magnétique Nucléaire
rpm : Rotation par minute
RT : Reverse Transcription = Transcription Inverse
Sad : Succinate semiAldehyde Dehydrogenase
SDS : dodécylsulfate de sodium
SLN : Signaux de Localisation Nucléaire
SOLiD : Supported Oligo Ligation Detection
SP : amorce spécifique
SRS : Substrate Recognition Site
TAE : Tris 40 mM acétate, EDTA 1 mM
TAIL-PCR : Thermal Asymmetric InterLaced PCR

TAL : Tyrosine Ammonia Lyase
TE : Tris-HCl 10 mM pH 8 ; EDTA 1 mM
*t*HH *: trans*-head-to-head
tHT : *trans*-head-to-tail
Tm : température de fusion
tSMS : true Single Molecule Sequencing
UMR : Unité Mixte de Recherche
UTR : UnTranslated Region
UV : Ultra-Violet
XOMT : Xanthotoxol-*O*-Méthyltransférase

Noms vernaculaires des espèces citées

Espèces végétales :

Ammi majus : Queen Anne's Lace
Angelica archangelica : Angélique vraie, archangélique, angélique officinale
Anthriscus sylvestris : Anthrisque sauvage, cerfeuil des bois
Apium graveolens : Céleri, ache des marais, persil des marais
Arabidopsis thaliana : Arabette des dames, arabette de Thalius, arabette rameuse, fausse arabette
Bituminaria bituminosa :
Bupleurum chinense : Buplèvre chinois
Chenopodium album : Chénopode blanc, ansérine blanche, poule grasse, drageline, senousse, blé-blanc, herbe aux vendangeurs, chou gras
Citrus bergamia : Bergamotier
Citrus jambhiri : Citrus de Floride
Coronilla emerus : Coronille des jardins, coronille faux-séné, séné-bâtard, faux baguenaudier
Cullen cinereum anciennement *Psoralea cinerea* : Psoralée
Daucus carota : Carotte, faux chervis, gironille, pastenade, moulette
Dorstenia gigas : Figuier de Socotra
Ferula communis : Férule commune
Ficus carica : Figuier, figuier comestible, figuier commun
Glehnia littorasis : American glehnia, beach sylvertop, american sylvertop
Helianthus tuberosus : Topinambour, artichaut de Jérusalem, truffe du Canada, soleil vivace
Heracleum lanatum : Berce laineuse, berce très grande
Heracleum mantegazziannum : Berce du Caucase, berce de Mantegazzi
Hierichloë odorata : Avoine odorante, hiérichloé odorant, herbe aux bisons
Ipomea batatas : Patate douce, kumara
Laserpitium siler : Laser siler, Sermontain
Lavandula officinalis : Lavande officinale, lavande vraie, lavande à feuilles étroites
Melilotus alba : Mélilot blanc
Melilotus officinalis : Mélilot officinal, mélilot jaune
Mentha pulegium : Menthe pouliot
Nicotiana benthamiana : Tabac indigène d'Australie, tjuntiwari, muntju
Nicotiana tabacum : Tabac cultivé
Opopanax chironium : Opopanax de Chiron, opopanax de Somalie
Pastinaca sativa : Panais, pastenade, racine-blanche, grand chervis
Petroselinum crispum ou *Petroselinum hortense* : Persil, persil cultivé, persil odorant, persin
Petunia hybrida : Pétunia
Peucedanum cervaria : Peucedan herbes aux cerfs
Phaseolus aureus ou *Vigna radiata* : Haricot mungo, ambérique verte, soja vert
Psoralea bituminosa : Psoralée bitumineuse
Psoralea corylifolia : Bogi-vittulu, vakuchi, babchi
Ruta graveolens : Rue officinale, rue fétide, rue commune, herbe de Grâce
Seseli montanum : Séséli des montagnes
Sium suave : Berle douce, berle suave
Tetradium daniellii : Arbre à miel, arbre aux abeilles
Thapsia garganica : Thapsie
Thapsia laciniata anciennement *Thapsia villosa type 2* : Thapsia velue
Zea mays : Maïs, blé d'Inde

Bactéries et Micro-organismes :

Agrobacterium rhizogenes ou *Rhizobium rhizogenes* : bactérie à coloration de Gram négative responsable du hairy root
Agrobacterium tumefaciens : bactérie à coloration de Gram négative responsable de la galle du collet
Aspergillus niger : Aspergille noire
Candida albicans : Levure, agent du muguet
Cryptococcus neoformans : Chryptochrome responsable de la méningite de Torula
Escherichia coli : Colibacille
Fusarium oxysporum : Jaunisse fusarienne
Mycobacterium fortuitum : Mycobactérie
Mycobacterium phlei : Mycobactérie
Mycobacterium smegmatis : Mycobactérie
Pectobacterium astrosepticum : Bactérie responsable de la pourriture molle bactérienne et de la jambe noire chez la pomme de terre
Phomopsis viticola : Excoriose de la vigne
Phytophtora megasperma f.s.p. glycinea : oomycète provoquant des symptômes de pourriture des racines et des tiges chez de nombreux végétaux
Pseudomonas aeruginosa : Bacille pyocyanique, bacille du pus bleu
Rhizoctonia solani : Rhizoctone brun, rhizoctone noir, variole des tubercules, maladie des manchettes
Saccharomyces cerevisiae : Levure de boulanger
Salmonella typhimurium : Bacille d'Aertrycke, bacille de Breslau, bacille de Nocard

Insectes :

Aedes aegypti : moustique vecteur de la dengue et de la fièvre jaune
Anopheles gambia : Anophèle
Depressaria pastinacella : Mouche du panais
Helicoverpa zea : Chenille des épis du maïs
Papilio glaucus : Papillon tigré de l'Est
Papilio multicaudatus : Two-tailed Tiger Swallowtail
Papilio polyxenes : Papillon du céleri
Trichopulsia ni : Ni, plusie ni

PARTIE 1 :

Synthèse bibliographique et présentation des objectifs de cette thèse

PARTIE 1 :
Synthèse bibliographique et présentation des objectifs de cette thèse

1 Phénylpropanoïdes

1.1 Phénylpropanoïdes

1.1.1 Généralités

Les phénylpropanoïdes sont des métabolites secondaires majoritairement présents dans le règne végétal. Ils constituent une grande classe contenant plusieurs dizaines de milliers de molécules. Les dérivés phénylpropanes possèdent tous le même squelette carboné issu de la phénylalanine, qui est constitué d'un noyau aromatique et d'une chaîne latérale de trois carbones (structure dite en C6-C3) (Figure 1, p. 27). Le squelette carboné subit des réactions d'hydroxylation, méthylation, glycosylation, acylation, prénylation ou encore sulfatation ce qui conduit à une très grande diversité de molécules. Les phénylpropanoïdes simples sont produits à partir de l'acide *t*-cinnamique qui est hydroxylé, méthylé et/ou réduit pour former les acides *p*-coumarique, caféique, férulique, sinapiques et leurs dérivés ainsi que les coumarines simples (Figure 2, p. 29). Ils sont en général présents dans la plante sous forme de conjugués (esters de glucose, esters d'acides organiques, polymères structuraux). La structure dite en C6-C3 des phénylpropanoïdes simples peut être additionnée de 3 groupements acétyl-CoA pour former la structure dite en C6-C3-C6 commune à de nombreux composés comme les flavonoïdes qui comprend plus de 8000 molécules, les isoflavonoïdes ou encore les stilbènes (Dixon and Paiva, 1995).

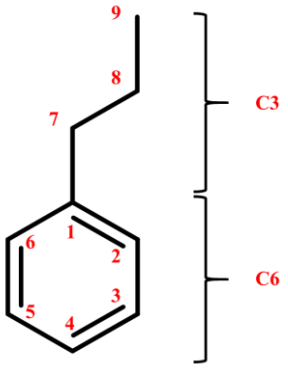


Figure 1 : Squelette carboné en C6-C3 à l'origine des phénylpropanoïdes simples.

Les phénylpropanoïdes sont impliqués dans diverses fonctions chez les plantes telles que la fertilité (Taylor and Grotewold, 2005), la pigmentation (Grotewold *et al*., 1994; Koes *et al*., 2005), la lignification, la signalisation entre plantes et microorganismes (allélopathie) (La Camera *et al*., 2004) et

les protections contre les stress biotiques et abiotiques (synthèse de phytoalexines par exemple) (Agati and Tattini, 2010; La Camera *et al*., 2004; Nuringtyas *et al*., 2012). Beaucoup de ces molécules présentent un intérêt pharmaceutique de par leurs propriétés antioxydantes (Rice-Evans *et al*., 1996), anticancéreuses (Ferrazzano *et al*., 2011), anti-inflammatoires (Chang *et al*., 2010), antibactériennes (Daglia, 2012) et protectrice contre les rayonnements UV (Heinrich *et al*., 2011; Korkina, 2007).

1.1.2 Voie de biosynthèse

La voie de biosynthèse des phénylpropanoïdes est commune à tous les végétaux supérieurs. Ces molécules sont synthétisées à partir d'un acide aminé, la phénylalanine ou la tyrosine dans le cas des monocotylédones. Elle est initiée par la conversion de la phénylalanine en acide *t*-cinnamique par l'action de la phénylalanine ammonia lyase (PAL) qui catalyse la rupture de liaison carbone-azote (Rosler *et al*., 1997) (Figure 2, p. 29). Cette enzyme constitue le point d'embranchement entre le métabolisme primaire du shikimate qui conduit aux acides aminés aromatiques, et le métabolisme secondaire des phénylpropanoïdes (Bauer *et al*., 2011). Les enzymes PAL sont généralement codées par plusieurs gènes chez les végétaux, et le nombre de gènes impliqués varie selon l'organisme considéré (Hamberger *et al*., 2007). Pour les plantes haploïdes, le nombre de gènes varie en général de 1 à 5, alors que chez la pomme de terre ce nombre peut atteindre la cinquantaine (Joos and Hahlbrock, 1992; Kumar and Ellis, 2001; Logemann *et al*., 1995). Des études ont montré que les différentes isoformes de la PAL peuvent avoir différentes localisations cellulaires et peuvent être impliquées dans la synthèse spécifique de certains métabolites secondaires (Jørgensen *et al*., 2005; Kao *et al*., 2002; Olsen *et al*., 2008).

La deuxième étape de cette voie permet l'hydroxylation de l'acide *t*-cinnamique par la cinnamate-4-hydroxylase (C4H ou CA4H) pour former l'acide *p*-coumarique (Figure 2, p. 29). Chez les monocotylédones, l'acide *p*-coumarique est directement formé à partir de la tyrosine grâce à la tyrosine ammonia lyase (TAL) (Figure 2, p. 29) (Nishiyama *et al*., 2010). La C4H a été identifiée pour la première fois à partir de tubercules d'*Helianthus tuberosus* (Gabriac *et al*., 1991) et de plantules de *Phaseolus aureus* (Mizutani *et al*., 1993). Cette enzyme qui est localisée au niveau du réticulum endoplasmique, fait partie de la sous-famille CYP73A des cytochromes P450 (P450s) (cf. 3.2. Nomenclature, p. 58) qui compte plus d'une centaine de membres à ce jour (http://drnelson.uthsc.edu/CytochromeP450.html). La C4H est exprimée de manière constitutive dans les tissus produisant des phénylpropanoïdes. L'expression de la C4H peut néanmoins être induite par des stress biotiques et abiotiques (Fahrendorf and Dixon, 1993; Hübner *et al*., 2003; Xu *et al*., 2010). Elle peut également être régulée au niveau enzymatique par inhibition compétitive ou par inhibition autocatalytique (Gravot *et al*., 2004; Schalk *et al*., 1998; Tohge *et al*., 2013).

L'acide *p*-coumarique est ensuite estérifié grâce à la *p*-coumarate:CoA-ligase (4CL) pour former le *p*-coumaroyl-CoA (Costa *et al*., 2005; Endler *et al*., 2008) qui est le précurseur de nombreux métabolites secondaires tels que les flavonoïdes, les monomères de lignines, les coumarines ou encore les furanocoumarines (Figure 2, p. 29) (Boudet, 2007; Bourgaud *et al*., 2006; Rani *et al*., 2009; Vogt, 2010).

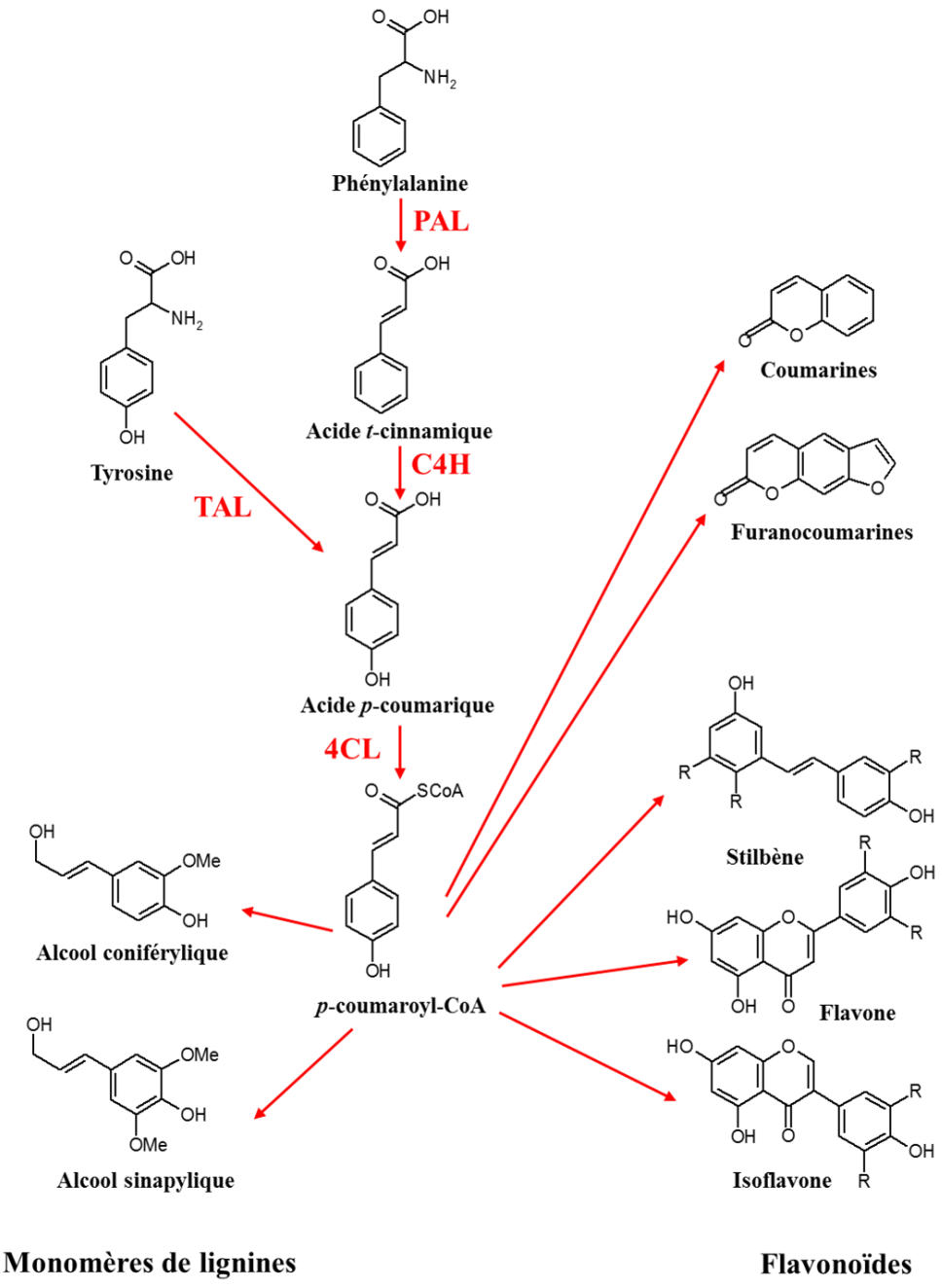
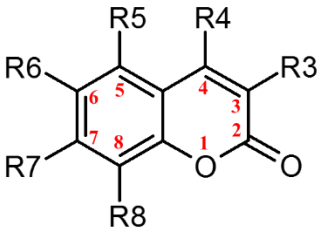


Figure 2 : Représentation schématique simplifiée de la voie de biosynthèse des phénylpropanoïdes et de ses principaux dérivés. TAL : tyrosine ammonia lyase ; PAL phénylalanine ammonia lyase ; C4H : cinnamate-4-hydroxylase ; 4CL : *p*-coumarate:CoA-ligase.

1.2 Coumarines

1.2.1 Généralités

Les coumarines représentent une classe de composés phénoliques simples qui dérivent tous de la même structure de base, la 2H-1-benzopyran-2-one (Figure 3, p. 30) (Bourgaud *et al*., 2006). La coumarine tire son nom de *kumarù*, le nom local du gaïac de Cayenne dont elle a été pour la première fois isolée en 1820 par Vogel à partir de fèves tonka (Clark, 1995). La structure de base peut subir de nombreuses modifications comme des hydroxylations, des *o*-méthylations, des géranylations ou encore des glycosylations ce qui conduit à la synthèse de nombreux dérivés dont plus d'un millier a été caractérisé à ce jour dans les plantes (Figure 3, p. 30).



Coumarine : R3-R8 = H
Esculétine : R3-R5 = H, R6-R7 = OH, R8 = H
Umbelliférone : R3-R6 = H, R7 = OH, R8 = H
Scopolétine : R3-R5 = H, R6 = OCH_3, R7 = OH, R8 = H
Scopoline : R3-R5 = H, R6 = OCH_3, R7 = *O*-glucose, R8 = H
Herniarine : R3-R6 = H, R7 = OCH_3, R8 = H
Auraptène : R3-R6 = H, R7 = *O*-géranyl, R8 = H

Figure 3 : Représentation de la structure chimique développée de base des coumarines et de quelques dérivés.

Les coumarines peuvent être classées en quatre catégories (Bourgaud *et al*., 2006) (Figure 4, p. 31) :

- les coumarines simples (benzo-α-pyrones) dont la structure diffère de la 2H-1-benzopyran-2-one par des hydroxylations, méthylations, géranylations et glycosylations,
- les furanocoumarines (furobenzo-α-pyrones) qui présentent l'ajout d'un noyau furane à l'hétérocycle coumarine. La position du noyau furane en C6-C7 ou en C7-C8 au niveau de la coumarine permet de distinguer les formes linéaires et angulaires,
- les pyranocoumarines (benzodipyran-2-ones) dont la structure est formée par l'addition d'un hétérocycle à six carbones sur la coumarine. Comme pour les furanocoumarines, la position de l'hétérocycle sur la coumarine définit les formes linéaires et angulaires,

- et les phénylcoumarines (benzo-benzopyrones) qui sont caractérisées par l'addition d'un cycle phényl sur la 2H-1-benzopyran-2-one.

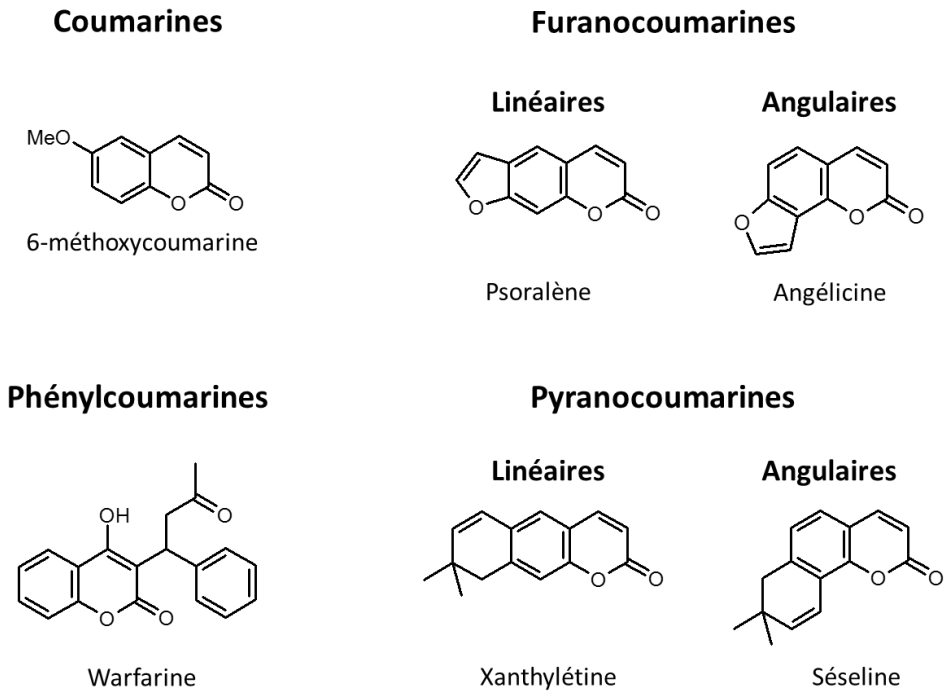


Figure 4 : Classification des différents types de coumarines.

1.2.2 Rôle physiologique des coumarines

Les coumarines sont synthétisées chez tous les végétaux supérieurs, mais leurs profils métaboliques varient selon l'espèce botanique considérée (Gliszczyńska and Brodelius, 2012; Gray and Waterman, 1978). Leur présence a été décrite notamment chez les Rutacées, Apiacées, Fabacées, Solanacées, Rubiacées, Saxifragacées, Astéracées et Rosacées (Ojala, 2001). Les coumarines sont principalement synthétisées au niveau des parties aériennes (de Castro *et al*., 2007; Tal and Robeson, 1985, 1986), mais elles peuvent également être produites dans les parties racinaires comme c'est le cas des furanocoumarines chez le panais. Ces molécules peuvent être stockées dans la plante dans les tissus où elles sont produites ou acheminées vers d'autres organes (Gorz and Haskins, 1962), mais elles peuvent également être excrétées dans le milieu (Gutiérrez-Mellado *et al*., 1996). La plupart des coumarines sont des phytoalexines impliquées dans des réactions de défense des plantes contre des attaques par des organismes pathogènes (Afek *et al*., 1995; Churngchow and Rattarasarn, 2001; Stange Jr. *et al*., 1999; Tal and Robeson, 1985, 1986). Certaines de ces molécules peuvent également être accumulées à la suite

d'un traitement au méthyljasmonate ou à l'acide 2,4-dichlorophénoxyacétique (Kai *et al*., 2006; Sharan *et al*., 1998). Les coumarines sont utilisées dans l'industrie pharmaceutique grâce à leur large spectre d'activités. En effet, elles présentent des propriétés antibactériennes (phytoalexines) (Cowan, 1999; Ojala *et al*., 2000), antifongiques (Carpinella *et al*., 2005; Valle *et al*., 1997), anti-inflammatoires (Chen *et al*., 2008), anticoagulantes (Cravotto *et al*., 2001), antidépresseurs (Sashidhara *et al*., 2011), antivirales et seraient impliquées dans la réduction du stress oxydatif (Chong *et al*., 2002). L'usage le plus connu est l'extrait de mélilot, riche en coumarines, utilisé dans le traitement de l'insuffisance veineuse et des phlébites et commercialisé sous le nom Esberiven. Certaines coumarines sont également valorisées dans les industries agroalimentaire, cosmétique et chimique (Egan *et al*., 1990). A titre d'exemple, la coumarine qui est caractérisée par une odeur de foin fraîchement coupé fut l'une des premières molécules synthétisée chimiquement au XIXème siècle pour la création de parfums et de nombreux produits cosmétiques (Ehlers *et al*., 1995), mais elle est aussi utilisée comme arôme dans l'élaboration de tabac et de cigarettes (Stanfill *et al*., 2006), ou comme neutralisateur d'odeurs dans l'élaboration de colles et de peintures.

1.2.3 Voie de biosynthèse des coumarines

La voie de biosynthèse des coumarines dérive de la voie de biosynthèse des phénylpropanoïdes (Figure 2, p. 29) (Brown, 1963; Kai *et al*., 2006). Le noyau phénol de l'acide *t*-cinnamique et de ses dérivés peut être hydroxylé en différentes positions. Les hydroxylations en position 4 de l'acide *t*-cinnamique, en position 5 de l'acide férulique et en position 3 de l'acide *p*-coumarique sont des réactions dépendantes de P450s (Gabriac *et al*., 1991; Meyer *et al*., 1996; Nair *et al*., 2002). De la même manière, les esters *p*-coumaroyl quinate et *p*-coumaroyl shikimate formés à partir du *p*-coumaroyl CoA grâce à l'HCT (hydroxycinnamoyl CoA shikimate/quinate hydroxycinnamoyl transférase) et à l'HQT (hydroxycinnamoyl CoA quinate hydroxycinnamoyl transférase) (Comino *et al*., 2007, 2009; Niggeweg *et al*., 2004) peuvent être hydroxylés en position 3' par un cytochrome P450 de la sous-famille CYP98A et former respectivement le caféoyl shikimate et le caféoyl quinate (acide chlorogénique) (Schoch *et al*., 2001) (Figure 5, p. 34). Les hydroxylations en position 2 de l'acide *t*-cinnamique, de l'acide *p*-coumarique, du férulate et du caféate ainsi que les hydroxylations en position 2' si ces molécules sont estérifiées par un coenzyme A, l'acide shikimique ou quinique orientent cette voie vers la biosynthèse de coumarines. Cette étape d'hydroxylation fut longtemps supposée être catalysée par un P450, mais récemment des études ont montré que l'étape d'*ortho*-hydroxylation de l'acide *p*-coumarique et de ses dérivés est catalysée par une dioxygénase (FeII)/α-cétoglutarate dépendante. Kai et ses collaborateurs ont créé différents mutants d'*Arabidopsis thaliana* n'exprimant pas la *p*-coumaroyl shikimate/quinate 3'-hydroxylase (CYP98A3) ou pas la caféoyl CoA *O*-méthyltransférase 1 (CCoAOMT1). La

comparaison des profils métaboliques des mutants par rapport aux plantes sauvages indique que ces mutations conduisent toutes les deux à une diminution importante de la synthèse de scopolétine et de sa forme glycosylée, la scopoline dans les racines (Kai *et al*., 2008). Ils en concluent que la scopolétine est synthétisée à partir du féruloyl CoA et non pas à partir de l'acide férulique comme les études précédentes semblaient le démontrer (Cabello-Hurtado *et al*., 1998a; Fritig *et al*., 1970). L'enzyme capable de réaliser l'*ortho*-hydroxylation du féruloyl CoA est appelée la féruloyl CoA 6'-hydroxylase (F6'H) (Figure 5, p. 34). Une autre étude réalisée par notre groupe de recherche décrit l'isolement et la caractérisation fonctionnelle d'une *p*-coumaroyl CoA 2'-hydroxylase de *Ruta graveolens* (Figure 5, p. 34) (Annexe 1, p. 273-281) (Vialart *et al*., 2012).

Cette dioxygénase (FeII)/α-cétoglutarate dépendante est capable de former de la scopolétine à partir du féruloyl CoA et de l'umbelliférone à partir du *p*-coumaroyl CoA (Vialart *et al*., 2012) (Figure 5, p. 34). Des résultats similaires ont été obtenus chez *Ipomea batatas* avec l'isolement de deux C2'H, l'une étant également capable de former de la scopolétine et de l'umbelliférone respectivement à partir du féruloyl CoA et du *p*-coumaroyl CoA alors que la deuxième ne produit que de la scopolétine à partir du féruloyl CoA (Matsumoto *et al*., 2012). L'umbelliférone peut alors être hydroxylée par une polyphénol oxidase en position 6 pour former l'esculétine (Garcia-Molina *et al*., 2013) ou être prénylée et ainsi conduire à la formation de furanocoumarines (Figure 5, p. 34).

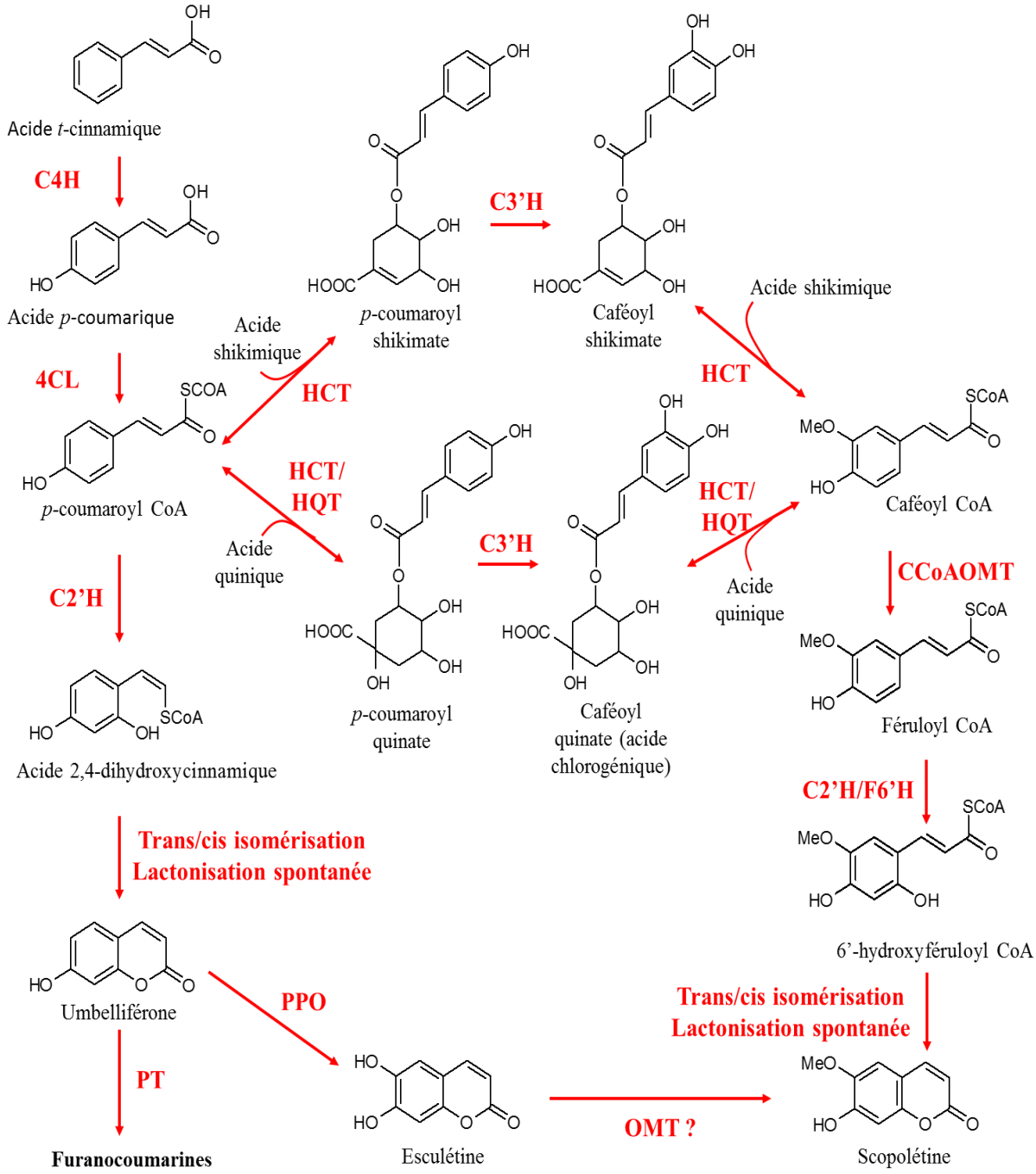
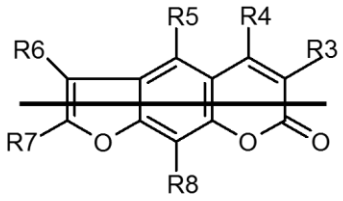


Figure 5 : Voie de biosynthèse des coumarines, formations de l'umbelliférone, de l'esculétine et de la scopolétine. C4H : cinnamate-4-hydroxylase ; 4CL : *p*-coumarate :CoA-ligase ; C2'H : *p*-coumaroyl CoA 2'-hydroxylase ; HCT : hydroxycinnamoyl CoA shikimate/quinate hydroxycinnamoyl transférase ; HQT : hydroxycinnamoyl CoA quinate hydroxycinnamoyl transférase ; C3'H : p-coumaroyl shikimate/quinate 3'-hydroxylase ; CCoAOMT : caféoyl CoA *O*-méthyltransférase ; F6'H : féruloyl CoA 6'-hydroxylase ; PPO : polyphénol oxidase ; PT : prényltransférase.

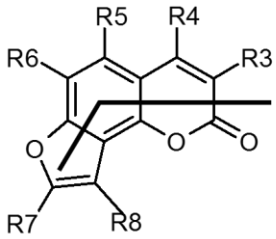
1.3 Furanocoumarines

1.3.1 Structure et propriétés chimiques

Les furanocoumarines sont des métabolites secondaires qui dérivent de la voie des phénylpropanoïdes. Ce sont des molécules tricycliques produites par la condensation de deux hétérocycles : coumarine et furane. La position du cycle furane permet de distinguer deux types de furanocoumarines (Figure 6, p. 35) qui définissent deux voies de biosynthèse différentes (Matern *et al*., 1988) : les furanocoumarines linéaires et les angulaires. Ces dernières sont probablement apparues à la suite de phénomènes de co-évolution entre les plantes et des insectes phytophages (Li *et al*., 2003) (cf. 3.7.2.3. P450s et co-évolution, p. 71). Les furanocoumarines constituent une classe de molécules très diversifiée grâce à la présence d'une grande variété de groupements greffés sur les carbones 2, 5 et 8 (hydroxyle, alkyle, méthoxyle, prényle, géranyle…) (Bourgaud *et al*., 2006; Dugrand *et al*., 2013; Matern *et al*., 1988).



Psoralène : R3 - R8 = H
Xanthotoxol : R3 - R7 = H, R8 = OCH_3
Isopimpinelline : R3, R4, R6 et R7 = H, R5 et R8 = OCH_3
Cnidiline : R3, R4, R6 - R8 = H, R5 = *O*-isopentényl
8-Géranylpsoralène : R3 - R7 = H, R8 = *O*-géranyl



Angélicine : R3 - R8 = H
Sphondine : R3 - R5 et R7 = H, R6 = OCH_3, R8 = OH
Pimpinelline : R3 - R5 et R7 = H, R6 et R8 = OCH_3
Hératomine : R3 - R5, R7 et R8 = H, R6 = *O*-isopentényl

Figure 6 : Exemples de structures chimiques de furanocoumarines linéaires et angulaires.

Les cycles aromatiques de ces molécules leur confèrent un maximum d'absorbance pour des longueurs d'onde comprises entre 250 et 320 nm. Les furanocoumarines présentent une forte solubilité dans les solvants aprotiques apolaires (chloroforme) et dans certains solvants aprotiques polaires tels que le DMSO, néanmoins elles sont très peu solubles dans les solvants protiques polaires (eau, éthanol, méthanol).

1.3.2 Distribution dans le règne végétal

Les furanocoumarines sont distribuées principalement chez quatre familles botaniques à savoir les Apiacées, les Rutacées, les Fabacées et les Moracées (Pathak *et al*., 1962) (Figure 7, p. 36).

Les concentrations et profils métaboliques varient d'une famille, d'une espèce et d'un écotype à un autre, mais également selon le tissu considéré ainsi que l'âge de la plante (Pathak *et al*., 1962). Certaines familles botaniques ne synthétisent que des formes linéaires. C'est le cas des Rutacées (*R. graveolens*, *Citrus bergamia*) qui synthétisent principalement du bergaptène (5-MOP), de la xanthotoxine (8-MOP) et de l'isopimpinelline (Abyshev *et al*., 1992; Gattuso *et al*., 2007; Schimmer and Kühne, 1990), de la majorité des Moracées (*Ficus carica*) (Marrelli *et al*., 2012) bien que récemment des formes angulaires aient été mises en évidence chez *Dorstenia gigas* (Heinke *et al*., 2012), et de certaines Apiacées (*Ammi majus*) (Matern *et al*., 1988; Pokrovskii *et al*., 2009). D'autres familles botaniques synthétisent à la fois des furanocoumarines linéaires et angulaires. C'est le cas des genres *Psoralea*, *Cullen* et *Coronilla* de la famille des Fabacées qui synthétisent uniquement des formes simples telles que le psoralène et son isomère angulaire, l'angélicine (Boardley *et al*., 1986). Certaines plantes de la famille des Apiacées (*Apium graveolens*, *Pastinaca sativa*) produisent également les deux formes linéaires et angulaires (Ekiert and Kisiel, 2000; Larson *et al*., 2013). A ce jour, il n'existe pas d'espèce décrite ne produisant que des formes angulaires. Quelques autres familles présentent en faibles quantités des furanocoumarines. Il s'agit des Astéracées, des Pittosporacées, des Rosacées, des Solanacées et des Thymélacées (Murray *et al*., 1982).

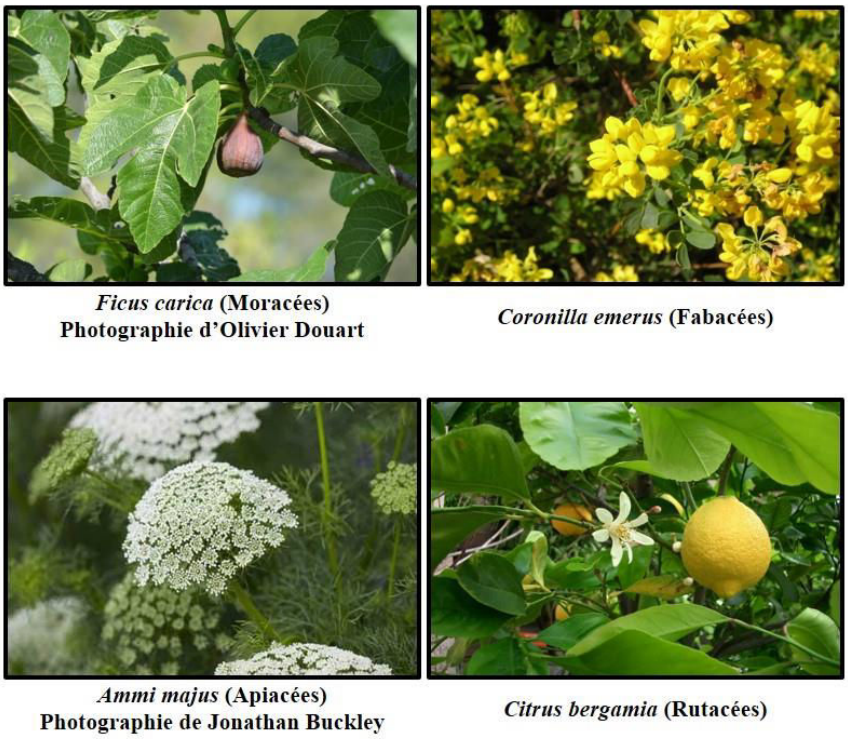


Figure 7 : Exemple de plantes appartenant aux quatre familles botaniques produisant des furanocoumarines en grande quantité.

1.3.3 Localisation au sein de la plante et stockage dans la cellule

Les furanocoumarines sont majoritairement localisées dans et à la surface des jeunes feuilles, dans les fruits et dans les graines (Pathak *et al*., 1962; Zobel and Brown, 1990a; Zobel *et al*., 1990, 1991), mais on les trouve également en plus faible quantité dans les tiges, les racines et les rhizomes (Page *et al*., 2006; Pathak *et al*., 1962). Cette localisation préférentielle des furanocoumarines dans les organes reproducteurs et sujets à la prédation (Nitao and Zangerl, 1987; Zangerl and Rutledge, 1996) s'inscrit dans le cadre de la théorie de défense optimale (McKey, 1974) qui repose sur trois facteurs : le coût, le bénéfice et la probabilité d'attaque.

Une étude récente menée sur des cellules en culture de *R. graveolens* a montré que le taux d'accumulation de furanocoumarines dépendait du niveau de différenciation cellulaire et que les cellules accumulant du psoralène, du 5-MOP et du 8-MOP étaient différentes (Diwan and Malpathak, 2010).

Les furanocoumarines sont des composés toxiques pour la cellule, car elles ont la capacité de former des liaisons covalentes entre deux brins d'ADN et ainsi d'inhiber la réplication cellulaire. Afin de se protéger de ces propriétés mutagènes, les plantes ont mis en place des mécanismes de stockage spécifiques. Chez certaines plantes comme les Rutacées, les furanocoumarines sont glycosylées par des glycosyl-transférases présentes dans le cytosol et sont ensuite transférées dans la vacuole (Taguchi *et al*., 2000; Zobel and Brown, 1990a). Dans le cas d'une attaque par un bioagresseur, les molécules ainsi stockées sont immédiatement remobilisées et transformées en forme aglycone par des β-glycosidases cytosoliques (Wittstock and Gershenzon, 2002). Cette stratégie permet de libérer des molécules de défense toxiques de manière rapide et localisée. De plus, certaines plantes stockent les furanocoumarines sous forme libre à l'extérieur de la cellule, dans les espaces intercellulaires, à la surface des feuilles de *R. graveolens* (Zobel and Brown, 1990a), à la surface des fruits de *Sium suave* et à la surface des trichomes de *Daucus carota* et *Psoralea bituminosa* (Zobel and March, 1993).

1.3.4 Rôles des furanocoumarines

Les furanocoumarines sont des phytoalexines (Beier and Oertli, 1983) qui protègent la plante des attaques d'herbivores ou de micro-organismes pathogènes. De nombreuses activités antimicrobiennes ont été mises en évidence à ce jour. Des extraits de *Petroselinum crispum* (persil) et de *R. graveolens* (rue) sont toxiques pour *Rhizoctonia solani* (Ojala *et al*., 2000), des extraits de *Citrus jambhiri* sont toxiques pour *E. coli* (McCloud *et al*., 1992), ou encore des molécules extraites de feuilles de rue sont nocives pour *Fusarium oxysporum* et *Phomopsis viticola* (Oliva *et al*., 2003). De nombreux effets insecticides ont également été démontrés sur le développement de larves d'insectes en utilisant des extraits de plantes, d'huiles essentielles ou des solutions contenant uniquement des furanocoumarines

(Berenbaum *et al*., 1991; Guo *et al*., 2012; Hadaček *et al*., 1994; Pavela and Vrchotová, 2013; Stevenson *et al*., 2003). Les furanocoumarines sont également capables d'inhiber les P450s d'insectes qui sont impliqués dans la détoxication de composés toxiques (Neal and Wu, 1994). Ces molécules présentent des propriétés phototoxiques grâce à leur activité mutagène (Dardalhon *et al*., 1998). En effet, elles sont capables après photoactivation par un rayonnement UV de former des liaisons covalentes avec les thymines situées de part et d'autre du double brin d'ADN et ainsi bloquer la réplication et la transcription de l'ADN (cf. 1.3.5.3. Interactions avec les acides nucléiques, p. 41). Les furanocoumarines présentent également des activités allélopathiques avantageuses pour les plantes en inhibant la germination et la croissance des plantes avoisinantes et concurrentes. Par exemple, les furanocoumarines extraites d'Apiacées et de Rutacées sont des inhibiteurs de germination et de croissance de la laitue, de la tomate, de l'amarante ou encore de l'ail (Anaya *et al*., 2005; Baskin *et al*., 1967; Friedman *et al*., 1982; Hale *et al*., 2004; Shilling *et al*., 1992).

Les furanocoumarines sont considérées comme des molécules de défense chez les plantes qui en produisent. Leur synthèse peut être induite en réponse à des stress biotiques et abiotiques. Zangerl et Berenbaum ont montré qu'une blessure des feuilles de *P. sativa* à l'aide d'une aiguille mimant l'attaque par des insectes phytophages activait la synthèse de métabolites primaires et secondaires et notamment des furanocoumarines (Zangerl and Berenbaum, 1998). Ces résultats ont été confirmés lorsque des feuilles de panais sont mises en contact avec *Trichopulsia ni* (Zangerl *et al*., 2002). D'autres études menées sur l'ajout d'extraits de levures dans des cultures cellulaires de *R. graveolens* et de *Glehnia littoratis* ont démontré une corrélation entre l'activité de la PAL, de la 4CL et de la xanthotoxol-*O*-méthyltransférase (XOMT) et le contenu en furanocoumarines (Bohlmann *et al*., 1995; Kitamura *et al*., 1998). Chez certaines plantes, il est possible de modifier la teneur en furanocoumarines par l'ajout d'extraits de pathogènes dans le milieu de culture cellulaire (élicitation). A titre d'exemples, on peut citer les ajouts d'extraits de *Pectobacterium astrosepticum* et de divers *Bacillus* chez *R. graveolens* (Orlita *et al*., 2008a), d'extraits de *Phytophtora megasperma* f.sp. *glycinea* chez *A. majus* (Hamerski and Matern, 1988a) et chez *Petroselinum hortense* (Tietjen *et al*., 1983). Ce résultat peut également être obtenu grâce à d'autres molécules comme la benzothiadiazole et la saccharine qui entraînent une augmentation de la teneur en 5-MOP, d'isopimpinelline, de psoralène et de 8-MOP chez la rue (Orlita *et al*., 2008b). De même, la concentration de nutriments dans le sol (N, P, K) (Zangerl and Berenbaum, 1987), la teneur de l'air en ozone (Eckey-Kaltenbach *et al*., 1994), la pollution atmosphérique ou encore l'exposition aux rayonnements UV ($\lambda = 366$ nm) (Zobel and Brown, 1993) ont un impact significatif sur la synthèse de ces molécules. Enfin, une étude a démontré que le contenu en furanocoumarines dans les feuilles de *Heracleum lanatum* variait en fonction des saisons avec une concentration maximale fin avril pour décroitre avec l'agrandissement des feuilles (Zobel and Brown, 1990b).

1.3.5 Propriétés biologiques des furanocoumarines

Les furanocoumarines sont des molécules capables d'interagir avec un large spectre de macromolécules telles que les acides nucléiques, les protéines et les lipides notamment lorsqu'elles sont soumises à un rayonnement UV (λ = 345 à 410 nm) (Schoonderwoerd *et al*., 1991). Ces molécules sont particulièrement photoréactives et peuvent former des adduits avec les macromolécules (Henegouwen *et al*., 1989).

1.3.5.1 Réaction de photooxydation et photolyse

Les furanocoumarines ont la capacité de former des produits de photolyse lorsqu'elles sont soumises à un rayonnement UV. Suivant la nature des molécules et les conditions expérimentales, ce processus conduit à la formation de nombreux produits qui peuvent être classés selon leur mécanisme de production dans deux groupes : les produits formés par des mécanismes anoxiques (produits formés par cyclodimérisation ou addition de solvent) et les produits formés par des mécanismes oxiques (produits oxydés formés par l'addition de solvant ou par cassure de molécules) (Figure 8, p. 39) (Caffieri, 2002). Une furanoquinolinone de synthèse (4,6,9-triméthylfuro[2,3-*h*]quinolin-2(1*H*)-one) est capable de former des monoadduits et des liaisons covalentes avec l'ADN en présence d'UVA. Il en résulte la formation de nombreux produits de photolyse (Marzano *et al*., 2002).

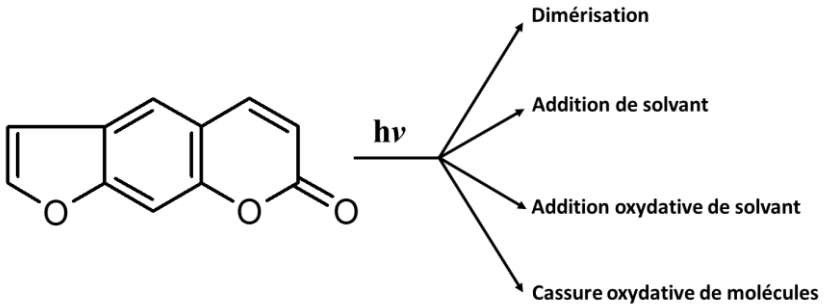


Figure 8 : Réactions de photolyse et de photooxydation causées par les furanocoumarines.

Les furanocoumarines excitées par les UVA peuvent également générer la formation au niveau des mitochondries d'espèces oxygénées réactives ($^1O_2$, H_2O_2, $H_20^{\bullet}$) qui sont impliquées dans le processus d'apoptose (Mignotte and Vayssiere, 1998).

1.3.5.2 Réaction de photodimérisation et photocycloaddition

Les coumarines et les furanocoumarines sont capables de former des dimères ce qui aboutit à la formation de nombreux isomères (Rojas-Lima *et al*., 1999; Zdero *et al*., 1990). La réaction de dimérisation s'effectue après une induction aux UV sur les doubles liaisons C3=C4 aboutissant à la formation de 4 isomères chez les coumarines (*trans*-head-to-head (tHH), *trans*-head-to-tail (tHT), *cis* head-to-head (cHH) et *cis*-head-to-tail (cHT)) (Figure 9, p. 40) (Schonberg *et al.*, 1968) et à la formation de principalement deux isomères dans le cas du psoralène (*trans*-head-to-head (tHH), *trans*-head-to-tail (tHT)) (Zdero *et al*., 1990). Ces réactions de photocycloaddition permettent de générer un grand nombre de nouvelles molécules qui peuvent présenter un intérêt pour l'industrie pharmaceutique (Lin *et al*., 2010; Sinkel *et al*., 2011).

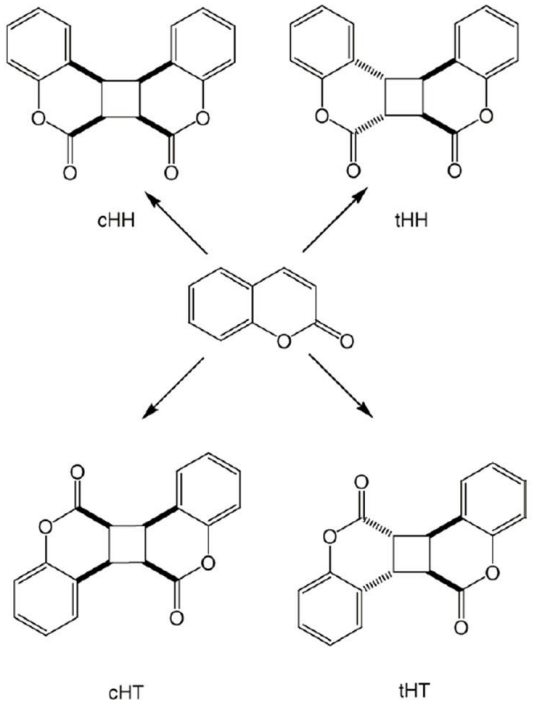


Figure 9 : Réactions de photodimérisation de composés coumariniques (Kitamura *et al*., 2005). cHH : *cis*-head-to-head ; tHH : *trans*-head-to-head ; cHT : *cis*-head-to-tail ; tHT : *trans*-head-to-tail.

Le psoralène peut également interagir avec les carbones C3=C4 et/ou C4'=C5' de molécules insaturées possédant une ou plusieurs doubles liaisons. Dans le cas de mono-adduit, 8 isomères peuvent être générés à partir du psoralène (Figure 10, p. 41). Des di-adduits peuvent être générés sur les furanocoumarines et dans le cas du psoralène ils conduisent à la formation de 16 isomères. Cette

propriété biologique est à l'origine de la photoaddition des furanocoumarines sur les acides gras insaturés.

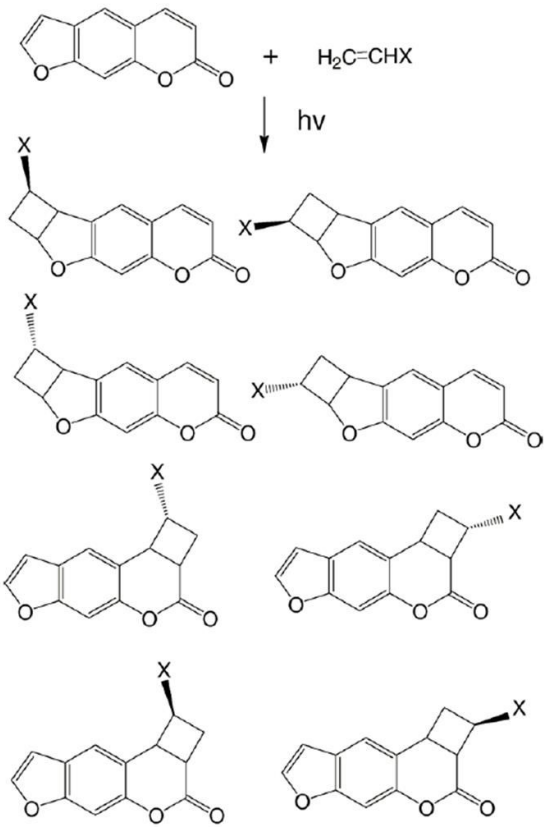


Figure 10 : Réactions de photocycloaddition du psoralène avec un groupement éthylène substitué (Kitamura *et al*., 2005).

1.3.5.3 Interactions avec les acides nucléiques

Musajo fut le premier à décrire la photoaddition des furanocoumarines avec l'ADN (Musajo *et al*., 1965). La réaction de photocycloaddition du psoralène et de ses dérivés sur les bases pyrimidiques de l'ADN se déroule dans le noyau de la cellule (Cadet *et al*., 1992; Sasaki *et al*., 1988). Cette réaction comprend trois étapes. Dans un premier temps, les furanocoumarines s'intercalent entre les doubles brins d'ADN par la formation de liaisons hydrogènes de faible énergie (Figure 11-A, p. 42). Dans un second temps, une exposition à un rayonnement UV (λ = 315 - 400 nm) et/ou visible va activer la furanocoumarine qui va former des liaisons covalentes entre les carbones C3=C4 du noyau pyrane et/ou C4'=C5' du noyau furane avec une base de l'ADN (Figure 11-B, p. 42). Enfin, si la conformation de

l'ADN le permet, une deuxième réaction de photocycloaddition peut avoir lieu sur l'autre brin d'ADN (Figure 11-C, p. 42) (Serrano-Pérez *et al.*, 2008).

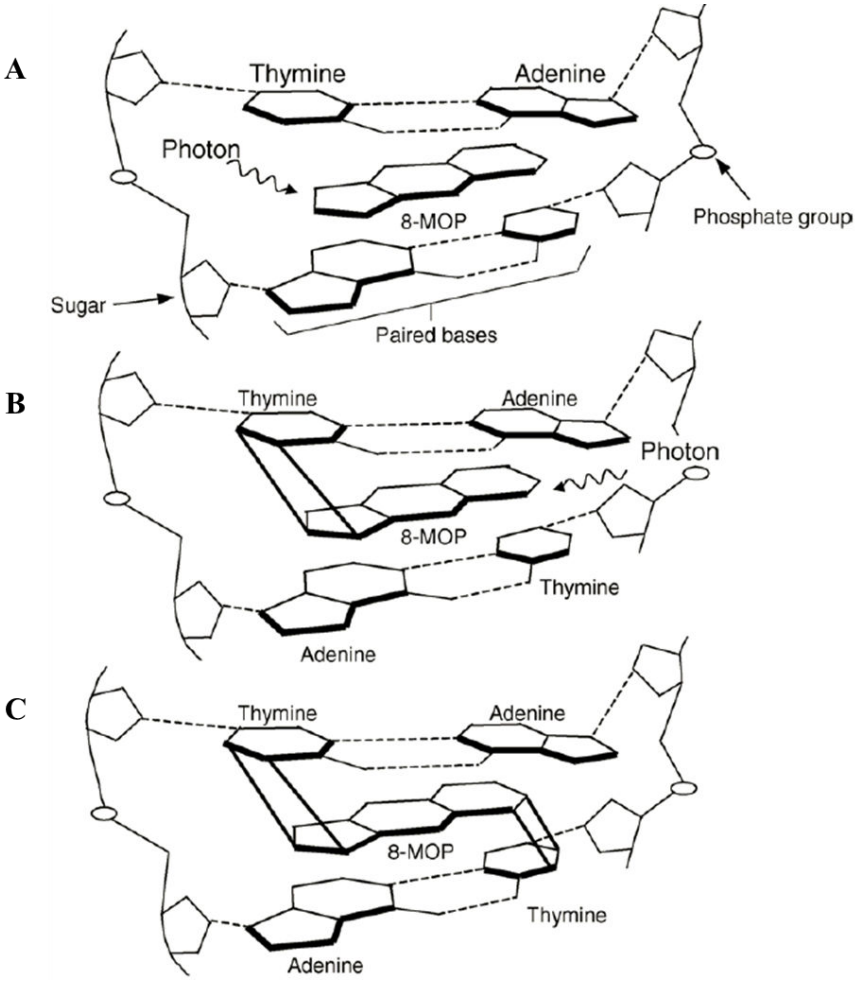


Figure 11 : Représentation schématique de la photocycloaddition du 8-méthoxypsoralène (8-MOP) sur les thymines des deux brins d'ADN (Kitamura *et al*., 2005) .

Le complexe ainsi formé empêche la fixation de polymérases et de transcriptases et bloque ainsi la réplication et la transcription de l'ADN ce qui explique le caractère antimitotique et mutagène des furanocoumarines (Dardalhon *et al*., 1998; Kevekordes *et al*., 1999). Les furanocoumarines forment préférentiellement des liaisons covalentes avec la thymine plutôt qu'avec l'adénine. Les régions de l'ADN riches en séquences poly[dA-dT].poly[dA-dT] sont les plus favorables à l'établissement de liaisons avec les furanocoumarines (Kang *et al*., 1992).

1.3.5.4 Interactions avec les acides gras insaturés

Les furanocoumarines soumises à un rayonnement UV peuvent endommager les membranes cellulaires en interagissant avec les lipides dont elles sont majoritairement constituées. Il existe deux types de réactions entre les furanocoumarines et les lipides membranaires : 1) une réaction induite par des formes oxygénées réactives qui conduit à la peroxydation des acides gras insaturés, 2) une réaction de photocycloaddition des furanocoumarines avec les acides gras insaturés qui pourrait stimuler la mélanogénèse (Dall'Acqua and Martelli, 1991; Zarebska *et al*., 2000).

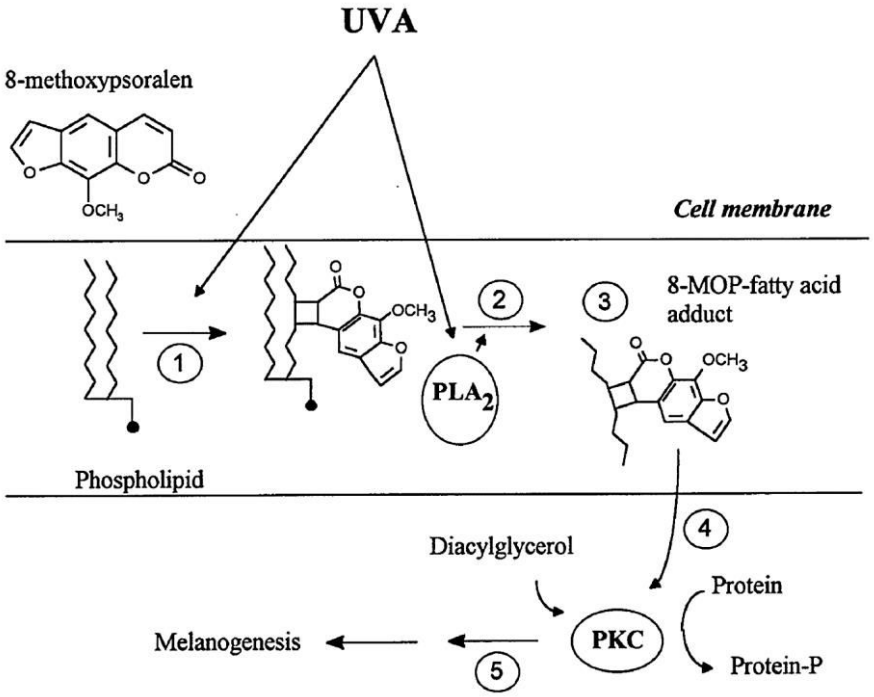


Figure 12 : Représentation schématique de l'implication des photocycloadditions de furanocoumarines avec des acides gras insaturés dans l'activation de la mélanogénèse (Zarebska *et al*., 2000).

L'établissement de liaisons covalentes entre les carbones C3=C4 du cycle pyrane et/ou du cycle furane avec un acide gras s'effectue selon les mêmes mécanismes réactionnels mis en jeu avec les acides nucléiques (Dall'Acqua and Martelli, 1991; Frank *et al.*, 1998). La photocycloaddition des furanocoumarines sur les phospholipides active la phospholipase A2 (PLA2) qui va dégrader le complexe formé et générer l'accumulation d'acides gras liés aux furanocoumarines dans le cytosol. Ces photoadduits vont mimer l'action du diacylglycérol (DAG) et activer la protéine kinase C (PKC) permettant la phosphorylation de nombreuses protéines et conduisant à une activation de la mélanogénèse (Figure 12, p. 43) (Anthony *et al*., 1997; Gordon and Gilchrest, 1989).

1.3.5.5 Interactions avec les protéines

Des expériences menées sur des cellules de rats traitées successivement avec du 8-MOP et un rayonnement UV montrent que 57 % du 8-MOP est lié aux protéines contre 26 % lié aux lipides et 17 % lié à l'ADN (Henegouwen *et al*., 1989). La photoaddition de furanocoumarines se fait sur une large gamme de protéines, telles que les ribonucléases, les lyzozymes, les histones, l'albumine bovine, les glutamate déshydrogénases ou encore les ADN polymérases I d'*E. coli* (Schmitt *et al*., 1994). Des études complémentaires ont montré que l'oxygène était nécessaire dans toutes ces réactions sauf pour la photoaddition aux ADN polymérases (Schmitt *et al*., 1995). L'isolement et la caractérisation partielle d'un photoadduit formé par le psoralène et la tyrosine furent réalisées pour la première fois par Sastry (Sastry, 1997). Plus récemment, une étude a montré que les furanocoumarines pouvaient inhiber l'activité des topoisomérases I en se photoadditionnant avec le complexe ADN/topoisomérase (Diwan and Malpathak, 2009). Les réactions de photoaddition des furanocoumarines avec un large éventail de protéines conduisent à des perturbations cellulaires importantes comme la modification de signaux cellulaires, l'induction et/ou la répression de facteurs de transcription, ou encore l'inactivation d'enzymes. Enfin des interactions avec des enzymes de la famille des P450s ont été décrites. Ceci sera développé ultérieurement dans une partie dédiée à ce point (cf. 3.6. P450s et inactivation autocatalytique, p. 65).

1.3.6 Voie de biosynthèse des furanocoumarines

La première étape spécifique de la voie de biosynthèse des furanocoumarines correspond à la prénylation de l'umbelliférone (7-hydroxycoumarine) par l'ajout d'une unité isoprène (diméthylallyl pyrophosphate, DMAPP) sur son carbone 6 ou son carbone 8 conduisant respectivement à la formation de déméthylsubérosine (DMS), précurseur des furanocoumarines linéaires, ou d'osthénol, précurseur des composés angulaires (Figure 13, p. 46). L'étape de prénylation est donc une étape importante puisqu'elle permet d'orienter la synthèse vers les furanocoumarines linéaires ou angulaires. Des premiers travaux menés sur des cellules de *R. graveolens* ont mis en évidence une enzyme capable de former de la DMS à partir d'umbelliférone (Ellis and Brown, 1974). Cette enzyme nécessite la présence de Mg^{2+} pour atteindre son activité optimale et serait chloroplastique (Dhillon and Brown, 1976). Ce n'est que très récemment que le premier gène codant pour une prényltransférase capable de former à la fois de la DMS et de l'osthénol à partir de l'umbelliférone a été identifié chez le persil, dans le cadre d'une collaboration de notre groupe de recherche avec l'équipe du professeur Yazaki de l'université de Kyoto (Karamat *et al*., 2014). La DMS et l'osthénol sont ensuite convertis respectivement en marmésine et en columbianétine au cours d'une étape de cyclisation du groupement prényl sur le groupement

hydroxyle en C7 (Figure 13, p. 46). Des expériences menées avec des précurseurs radiomarqués ont mis en évidence la conversion de la DMS en marmésine dans des cellules d'*A. majus* (Hamerski and Matern, 1988b). Cette étape serait catalysée par un P450, mais à ce jour ces résultats n'ont pu être confirmés au niveau moléculaire. L'étape suivante de cette voie est la conversion de la marmésine en psoralène pour les furanocoumarines linéaires et la conversion de la columbianétine en angélicine pour les furanocoumarines angulaires (Figure 13, p. 46). Cette étape est réalisée par des P450s localisés dans le réticulum endoplasmique et libèrent une molécule d'acétone (Hamerski and Matern, 1988b; Steck and Brown, 1970). Ces deux enzymes sont appelées psoralène synthase et angélicine synthase et ont été caractérisées chez *A. majus*, *P. sativa* et *A. graveolens* (Larbat *et al.*, 2007, 2009). Le psoralène et l'angélicine peuvent être hydroxylés en position 5 ou 8. Seule l'activité de la psoralène-5-monoxygénase qui conduit à la formation de bergaptol a pu être mise en évidence dans des cellules d'*A. majus* (Hamerski and Matern, 1988a). Les produits ainsi obtenus sont ensuite *O*-méthylés pour former le 8-MOP ou le 5-MOP dans le cas des furanocoumarines linéaires ou encore la sphondine dans le cas des furanocoumarines angulaires (Figure 13, p. 46). Seules les bergaptol – et xanthotoxol-*O*-méthyl transférases ont été purifiées à partir de cellules de *R. graveolens* (Thompson *et al.*, 1978). Plus récemment, la séquence codante de la bergaptol-*O*-méthyl transférase a été identifiée à partir de cellules d'*A. majus* (Hehmann *et al.*, 2004). Le 8-MOP et le 5-MOP peuvent être hydroxylés en 5 ou en 8 puis *O*-méthylés pour former l'isopimpinelline (Figure 13, p. 46). Le 8-MOP semble être un précurseur préférentiel dans la formation d'isopimpinelline (Brown and Sampathkumar, 1977).

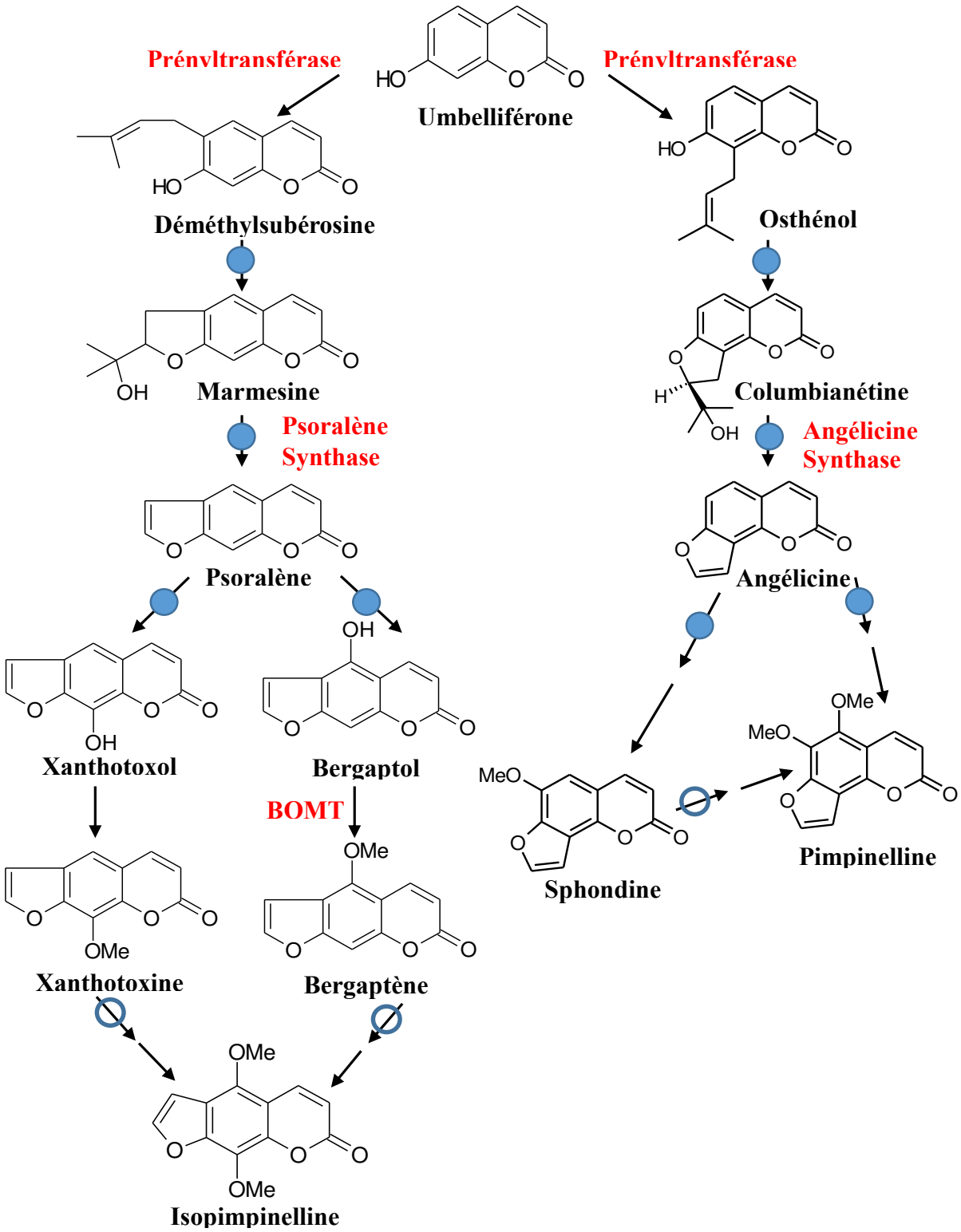


Figure 13 : Voie de biosynthèse hypothétique des furanocoumarines linéaires et angulaires. Les étapes identifiées comme catalysées par des cytochromes P450 sont marquées d'un point bleu. Les étapes potentiellement catalysées par des cytochromes P450 sont indiquées par un cercle bleu. Les séquences codantes pour les enzymes caractérisées à ce jour sont indiquées en rouge. BOMT = bergaptol-*O*-méthyl transférase.

1.3.7 Utilisations thérapeutiques

Depuis des millénaires, les plantes produisant des furanocoumarines sont utilisées dans les médecines traditionnelles pour le traitement de maladies de la peau. *A. majus* et *Psoralea corylifolia* étaient utilisées respectivement par les égyptiens et les indiens sous forme de décoctions dans le traitement du vitiligo (Pathak and Fitzpatrick, 1992). Mais ce n'est qu'en 1946 que le rôle du 8-MOP isolé à partir d'*A. majus* a été démontré dans la repigmentation de la peau pour le traitement du vitiligo (Fahmy and Abu-Shady, 1946). Depuis, les études sur les propriétés pharmacologiques des furanocoumarines se sont multipliées et elles sont utilisées actuellement pour traiter différentes pathologies.

1.3.7.1 La photochimiothérapie

La photochimiothérapie (ou P-UVA thérapie) est un traitement qui combine l'association de médicaments contenant des furanocoumarines avec une exposition aux rayonnements UV (320-340 nm) et ceci dans le but de traiter certaines maladies de peau liées à une hyperprolifération cellulaire (Parrish *et al*., 1974). Le 8-MOP, le 5-MOP, le 4,5',8-triméthylpsoralène et quelques autres furanocoumarines de synthèse sont communément utilisées dans ce traitement (Marzano *et al*., 2002). Les furanocoumarines peuvent être administrées par voie orale, par application locale et en bain (Danno, 1999; Giacomo Calzavara-Pinton *et al*., 1997). Les furanocoumarines associées à un rayonnement sont utilisées pour leur effet antiprolifératif grâce à leur capacité à former des adduits avec l'ADN, mais elles induisent également l'apoptose et sont impliquées dans des réactions d'immunosuppression (Godar, 1999; Moor and Gasparro, 1996; Sasaki *et al*., 1988). La P-UVA thérapie utilisant des furanocoumarines est utilisée pour le traitement de nombreuses affections de la peau comme le psoriasis, le vitiligo, le lichen plantaire ou encore l'eczéma (Domínguez *et al*., 2004; Pathak and Fitzpatrick, 1992; Poiesz *et al*., 1980). Le 5-MOP fut commercialisé sous la formulation Psoraderm-5® pour le traitement du psoriasis et le 8-MOP fut commercialisé sous les noms Meladinine®, Oxsoralène®, Deltasoralène® et Uvadex® pour le traitement du vitiligo, du psoriasis, de dermatites atopiques ou encore de dermatoses. Malgré des études prometteuses sur de nouvelles furanocoumarines de synthèse (Marzano *et al*., 2002; Serrano-Pérez *et al*., 2008), leur utilisation dans la P-UVA thérapie pour le traitement d'affections de la peau décroît progressivement notamment à cause de la découverte de nouveaux immunodépresseurs ayant des effets thérapeutiques plus importants et plus ciblés (Krueger, 2002; Nograles *et al*., 2010).

1.3.7.2 La photophérèse

La photophérèse est une photochimiothérapie extracorporelle des leucocytes utilisée pour le traitement des lymphomes cutanés à cellules T qui est une prolifération maligne des cellules helper T (Poiesz *et al.*, 1980). Les lymphocytes T mutés sont accumulés au niveau de l'épiderme et génèrent des lésions cutanées. Après séparation des leucocytes des globules rouges par dialyse, la photophérèse permet de traiter spécifiquement les leucocytes avec du 8-MOP et de les soumettre à un rayonnement UVA avant de les réinjecter au patient (Chiesa-Fuxench and González-Chávez, 2010). Ce traitement induit la formation de photoadduits avec les bases pyrimidiques de l'ADN qui peut induire l'apoptose chez un grand nombre de cellules traitées (Edelson *et al*., 1987). La photophérèse présente peu d'effets secondaires comparée aux traitements classiques qui incluent immunosuppresseurs et corticoïdes. Actuellement, cette approche est utilisée pour prévenir les rejets de greffes d'organes (Hivelin *et al*., 2009; Jagasia *et al*., 2009), mais également pour traiter les formes réfractaires de la sclérodermie, du lupus, de la sclérose en plaques, de la polyarthrite rhumatoïde (Guyot *et al*., 2007) ou encore la maladie de Crohn (Abreu *et al*., 2009).

1.3.7.3 Traitement antidépresseur

La dépression est une maladie qui peut avoir pour origine une hyperactivation de l'axe hypothalamo-pituitaro-surrénalien (HPA, A pour adrenal gland), une augmentation du catabolisme des monoamines et des dérèglements des systèmes oxydatifs et de la peroxydation des lipides (Bilici *et al*., 2001). De nombreux patients dépressifs présentent également une augmentation de l'activité de la superoxide dismutase et de la glutathion peroxydase (Abdalla *et al*., 1986). Récemment, Chen et ses collaborateurs ont montré qu'un extrait de furanocoumarines de graines de *Psoralea corylifolia* diminuait les symptômes de stress chroniques (Chen *et al.*, 2007). Ces résultats suggèrent que les furanocoumarines sont des antidépresseurs potentiels qui régulent la monoamine oxydase, la peroxydation des lipides et agit sur l'axe HPA (Chen *et al*., 2007).

1.3.7.4 Activités bactériostatiques, fongicides et antimycobactérienne

Comme cela a été discuté précédemment, les furanocoumarines présentent des propriétés bactériostatiques. Une récente étude a montré que la bergamottine et la dihydroxybergamottine inhibaient les activités des *N*-acylhomosérine lactones et des diesters de boronate impliqués dans la communication intra et interspécifiques entre les bactéries Gram-positf et Gram-négatif (Girennavar *et al*., 2008). De plus, ces deux furanocoumarines empêchaient la formation des biofilms associés à la

pathogénicité des *E. coli*, *Salmonella typhimurium* et *Pseudomonas aeruginosa* (Girennavar *et al*., 2008). Enfin, les furanocoumarines angulaires possèdent également des propriétés antifongiques. Sardari et ses collaborateurs ont testé la toxicité de 36 coumarines et furanocoumarines sur *Candida albicans*, *Cryptococcus neoformans*, *Saccharomyces cerevisiae* et *Aspergillus niger* et ont démontré que l'hydroxylation sur le carbone C6 des furanocoumarines angulaires était déterminante pour l'activité antifongique (Sardari *et al*., 1999). Les furanocoumarines géranylées isolées à partir de fruits de *Tetradium daniellii* présentent aussi une activité antimycobactérienne contre *Mycobacterium smegmatis*, *M. fortuitum* et *M. phlei* (Adams *et al*., 2006).

1.3.7.5 Activité antiproliférative et apoptotique

De nombreux travaux ont porté sur l'activité antiproliférative et apoptotiques des furanocoumarines sur des cellules cancéreuses humaines. Les mécanismes réactionnels mis en jeu sont en cours d'étude et permettront potentiellement la découverte de nouveaux anticancéreux. Les hypothèses sont nombreuses. Certains articles tendent à dire que c'est la présence d'un groupement méthyl sur le carbone C5 ainsi qu'une conformation linéaire qui pourraient être responsables de leur activité antiproliférative (Sumiyoshi *et al*., 2014). D'autres travaux comme ceux de Sigurdsson et de ses collaborateurs ont montré que les propriétés antiprolifératives, d'une teinture de fruits d'*Angelica archangelica* sur des cellules cancéreuses pancréatiques, provenaient majoritairement du 8-MOP (groupement méthoxy en C8) et de l'impératorine (Sigurdsson *et al*., 2004). Ces propriétés sont également démontrées sur des cellules cancéreuses de poumon, d'ovaire, de mélanome, du système nerveux central et du colon (Kim *et al*., 2007). Appendino, quant à lui, a montré que l'impératorine et l'héraclénine isolées à partir de racines et de graines d'*Opoparax chironium* bloquaient les cellules dans le cycle cellulaire, respectivement dans les phases G1/S et G2/M (Appendino *et al*., 2004). Un premier mécanisme réactionnel est décrit pour l'implication de l'impératorine dans l'induction de l'apoptose dans des cellules humaines HL-60 *via* la voie du cytochrome c/caspase-9. L'impératorine induirait la dépolarisation des membranes des mitochondries qui conduit ainsi à la dérégulation de la protéine Bcl-2 et la libération du cytochrome c des mitochondries vers le cytosol. Les caspases-9 et -3 sont ensuite activées par le cytochrome c et clivent de nombreuses protéines notamment la poly(ADP-ribose) polymérase, conduisant à l'apoptose (Pae *et al*., 2002).

1.3.8 Toxicité des furanocoumarines

Les furanocoumarines présentent de nombreuses propriétés thérapeutiques, mais sont également des molécules toxiques qui peuvent causer des photodermatites de contact, provoquer des mélanomes et être toxiques par ingestion.

1.3.8.1 Photodermatite

Les photodermatites sont des affections de la peau causées par l'action conjointe de molécules photosensibilisantes et d'une exposition aux rayonnements UV (soleil par exemple). Les lésions cutanées causées apparaissent 4 à 6 jours après l'exposition et peuvent se caractériser par un érythème cutané (Moreira *et al*., 2010), une hyperpigmentation de la peau et l'apparition de vésicules bulleuses (Kung *et al*., 2009). Les plantes pouvant causer des réactions cutanées phototoxiques appartiennent principalement aux familles des Apiacées, Rutacées, et Moracées (Lovell, 1997), mais ont également été observées chez *Chenopodium album* une plante de la famille des Chénopodiacées produisant des furanocoumarines (Bilgili *et al*., 2011).

1.3.8.2 Mélanome

Le mélanome cutané est le plus grave des cancers de la peau. Il existe une prédisposition génétique à cette maladie, mais il peut également être causé lors de traitements associant le 8-MOP avec un rayonnement UVA (Stern *et al*., 1997). Deux études récentes ont montré que les personnes qui consommaient très régulièrement du jus d'orange, avaient plus de chance de développer un mélanome cutané (Feskanich *et al.*, 2003; Sayre and Dowdy, 2008). L'absorption régulière de denrées alimentaires contenant des furanocoumarines augmente donc la probabilité de développer un mélanome cutané.

1.3.8.3 Par ingestion

De nombreuses études ont montré que 60 % des enzymes impliquées dans le catabolisme de médicaments et de molécules xénobiotiques en règle générale se faisait par des enzymes de la famille des P450s. Ces enzymes jouent donc un rôle essentiel dans l'élimination de molécules toxiques pour l'organisme. Un certain nombre d'études mettent en évidence que les furanocoumarines ont un effet inhibiteur sur ces enzymes engendrant de graves conséquences sur la santé (Guo *et al*., 2000). Par exemple, Ohnishi et ses collaborateurs ont démontré que les furanocoumarines contenues dans du jus de pamplemousse inhibaient CYP3A4 conduisant à une non dégradation de la vinblastine dans des

cellules Caco-2 (Ohnishi *et al*., 2000). Girennavar et ses collaborateurs ont montré que 5 furanocoumarines et dérivés de furanocoumarines contenues dans les jus de pamplemousse et de pomelo inhibaient les cytochromes CYP3A4, CYP2D6 et CYP2C9, la molécule la plus efficace étant la paradisine A (Figure 14, p. 51) (Girennavar *et al*., 2007).

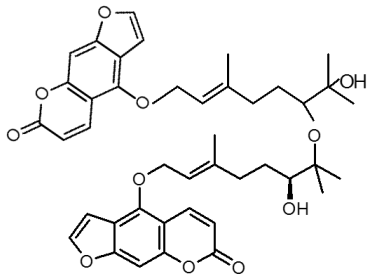


Figure 14 : Structure de la paradisine A.

Les furanocoumarines contenues dans les *Citrus* peuvent également inhiber CYP1A2, CYP2A6, CYP2C19 et CYP2E1 (Tassaneeyakul *et al*., 2000). Ce phénomène de surdosage médicamenteux induit par les furanocoumarines a été largement décrit sous le vocable « effet Pomelo » (Dahan and Altman, 2004). Mais cet effet n'est pas restreint aux agrumes. En effet, Lampe a pu montrer que l'ingestion de légumes de la famille des Apiacées (céleri, panais, persil, carotte et aneth) pouvait également conduire à l'inhibition de CYP1A2 impliqué dans le métabolisme de molécules procarcinogènes (hétérocycles amines, nitrosamines et aflatoxines B1) (Lampe *et al*., 2000).

2 *Pastinaca sativa*, *Ruta graveolens* et *Cullen cinereum* : trois plantes productrices de furanocoumarines

2.1 Description botanique

2.1.1 *P. sativa*

P. sativa communément appelée panais, pastenade, racine-blanche ou encore grand chervis est une plante herbacée bisannuelle de 50 cm à plus d'un mètre de haut appartenant à la famille des Apiacées qui fut autrefois largement cultivée comme légume et comme plante fourragère. L'espèce est représentée par trois sous-espèces : *P. sativa* L. subsp. *sylvestris* (panais sauvage), *P. sativa* L. subsp. *urens* (panais urticant) et *P. sativa* L. subsp. *sativa* (panais cultivé dont la variété 'Demi-Long de Guernesey' est la plus représentée actuellement).

Le panais sauvage est une plante érigée à feuilles alternes poilues qui porte des inflorescences en ombelles composées portant 6 à 15 rayons inégaux de fleurs à pétales jaunes portées par des tiges creuses et sillonnées (Figure 15-A, p. 52). La floraison s'étend de juin à octobre. Les fruits secs sont des diakènes aplatis qui se scindent en deux à maturité, chaque partie contenant une graine (Figure 15-A, p. 52). Les feuilles sont grandes, simples et pennées composées de 5 à 11 folioles entières ou tridentées (Figure 15-B, p. 52). La racine principale pivotante de la plante sauvage est amère et ligneuse alors que celles des variétés cultivées est douce et tendre. Le panais sauvage est natif d'Eurasie entre la région ouest de la Méditerranée et les montagnes caucasiennes. Il a été naturalisé dans de nombreuses régions tempérées du monde notamment en Amérique du Nord, en Australie, en Chine, au Japon, en Nouvelle-Zélande, en Afrique du Sud et en Amérique du Sud.

Le panais est un légume ancien déjà utilisé en médecine traditionnelle par les Grecs et les Romains qui fut longtemps assimilé et confondu avec la carotte. Dioscoride, dans son ouvrage *De Materia Medica*, note que l'espèce sauvage a des propriétés plus efficaces que celle comestible cultivée. Selon lui, les graines pouvaient être utilisées pour soigner la dysurie, l'hydropisie, les pleurésies, les morsures de bêtes venimeuses, les blessures provoquées par les coups ainsi que pour protéger des serpents et favoriser la conception. La racine quant à elle était utilisée pour ses propriétés diurétiques, aphrodisiaques et pour soigner les ulcères cancéreux. Actuellement, *P. sativa* est utilisé en aromathérapie pour le traitement de sclérophtalmies catarrhales, de glaucomes et d'entérocolites spasmodiques.

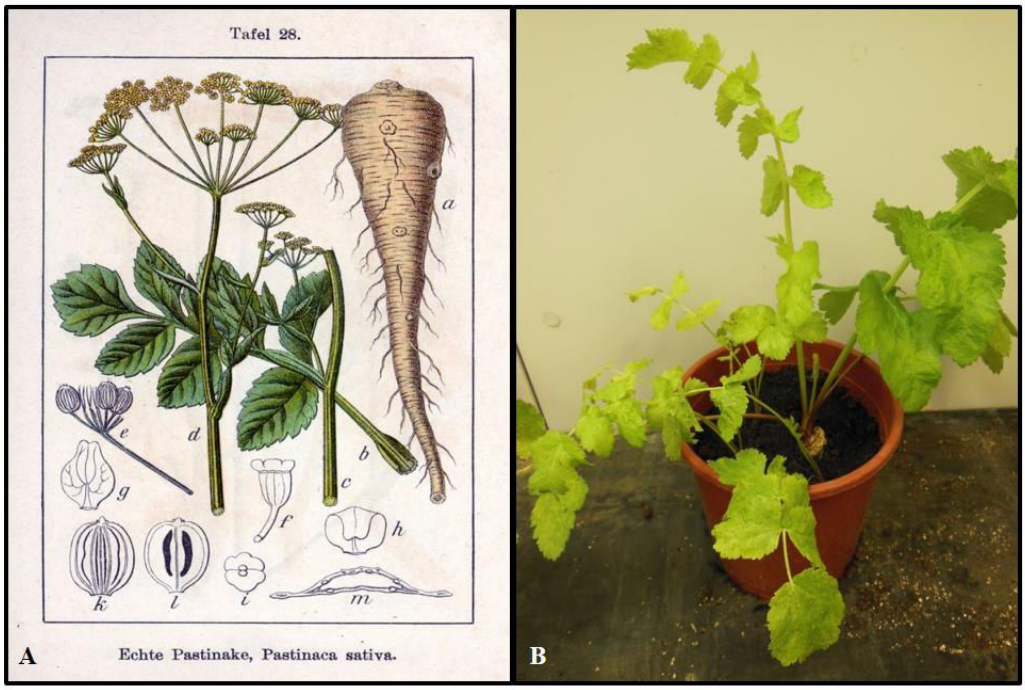


Figure 15 : A. Planche botanique de *P. sativa* L. (illustration de Johann Georg Sturm, 1796 Deutschlands Flora in Abbildungen) (Sturm and Sturm, 1796). B. Photographie de *P. sativa*.

2.1.2 *R. graveolens*

R. graveolens communément appelée rue des jardiniers, rue fétide, rue odorante, rue officinale ou encore herbe de grâce est un arbrisseau sous-frutescent de la famille des Rutacées. Cette plante méditerranéenne semi-arbustive de 70 cm à un mètre de haut est très ramifiée et ligneuse à la base. Ses feuilles glauques semi-persistantes sont alternes, pennatiséquées (souvent trilobées) et de consistance peu charnue (Figure 16-B, p. 53). Ses fleurs en cymes corymbiformes sont composées de 4 à 5 pétales jaunes verdâtres soudés à la base (Figure 16-A, p. 53). La floraison s'étend de mai à août. Les fruits sont des capsules lobées déhiscentes par 4 fentes libérant à maturité de petites graines noires (Figure 16-A, p. 53). *R. graveolens* dégage une forte odeur âcre et pénétrante émise par les huiles contenues dans les poches schizolysogènes à la surface de ses feuilles. La rue est une plante originaire du Sud-Est de l'Europe (Ukraine, Bulgarie, Albanie, ex-Yougoslavie) qui s'est ensuite naturalisée à toute l'Europe et à l'Afrique du Nord. Elle est également distribuée en Amérique du Sud où on la cultive pour des raisons ornementales et médicinales mais aussi à cause de croyances locales car elle protégerait du mal.

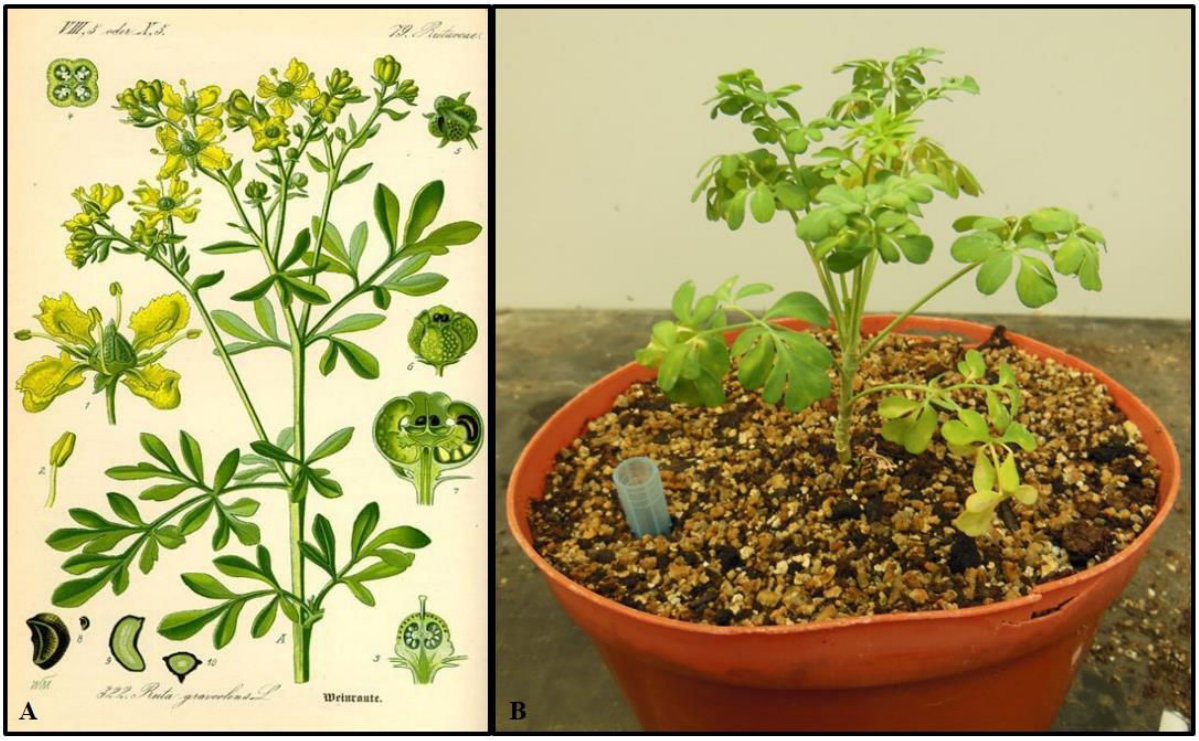


Figure 16 : A. Planche botanique de *R. graveolens* L. (illustration d'Otto Wilhelm Thomé, 1885 Flora von Deutschland, Österreich und der Schweiz) (Thomé, 1885). B. Photographie de *R. graveolens* L.

R. graveolens était utilisée dès l'Antiquité en médecine traditionnelle par les Romains et les Grecs pour ses propriétés antispasmodiques, emménagogues et abortives. Son utilisation s'est par la suite diversifiée. Les fruits et les parties aériennes, administrés sous forme d'infusion, de décoction ou de poudre étaient utilisés dans le traitement de maux de tête, de vertiges et de rhumatismes (De Feo and Senatore, 1993). Au Moyen-Âge, la rue entrait dans la composition du « vinaigre des quatre voleurs » sensé protégé de la peste et du typhus. Dans la Renaissance, de nombreux artistes comme Michelangelo l'utilisaient en usage externe pour les douleurs oculaires de surmenage. *R. graveolens* constitue également un excellent antiparasitaire et vermifuge (Guarrera, 1999, 2005). Actuellement la rue est utilisée principalement en homéopathie et en aromathérapie pour le traitement de douleurs musculaires et de la fatigue oculaire.

2.1.3 *C. cinereum*

C. cinereum Lindl. est une plnte herbacée pouvant atteindre une hauteur de 1 à 1,5 mètres appartenant à la famille des Fabacées, tribu des *Psoraleaceae* (Boardley *et al.*, 1986). Cette plante aussi connue sous le nom de *Psoralea cinerea* Lindl., appartenait anciennement au genre *Psoralea* qui comprenait approximativement 130 espèces réparties sur les cinq continents. Les règles de classification étant établies sur les végétaux européens, elles s'appliquent difficilement aux plantes australiennes. De plus, la grande variabilité morphologique au sein d'une même espèce, ainsi que les fécondations interspécifiques possibles engendrent des confusions taxonomiques. Une nouvelle classification taxonomique des Psoralées australiennes a été proposée par James Grimes et place 25 de ces espèces dans le genre *Cullen* (Grimes, 1997).

C. cinereum Lindl. présente des feuilles alternes trifoliées, chacune comportant trois folioles elliptiques et pétiolées avec un bord denté (Figure 17-C, p. 55). Les plantes présentent des inflorescences de couleur rose violet organisées en épis composés dont les bourgeons latéraux donnent chacun trois fleurs et portées par un pédoncule assez long (Figure 17-A, p. 55). La floraison de *C. cinereum* a lieu d'avril à mai. Les fruits sont des gousses glabres et ovales d'environ 5 mm ne contenant qu'une seule graine (Figure 17-B, p. 55). *C. cinereum* est une plante endémique d'Australie qui se développe dans les zones semi-arides (250 à 500 mm d'eau par an) et dans les régions côtières plus humides. Elle est associée aux plantes de cycle annuel mais peut survivre plusieurs années si les conditions sont favorables. Cette plante se développe sur des sols alluviaux, mais aussi sur des sols de type sableux à latéritique. *C. cinereum* est capable de s'adapter aux conditions climatiques et édaphiques de la région à blé de l'ouest de l'Australie et pourrait être intégrée comme plante fourragère dans les rotations de culture pour le développement de pâturages pérennes (Bennett *et al.*, 2011; Nicol *et al.*, 2013).

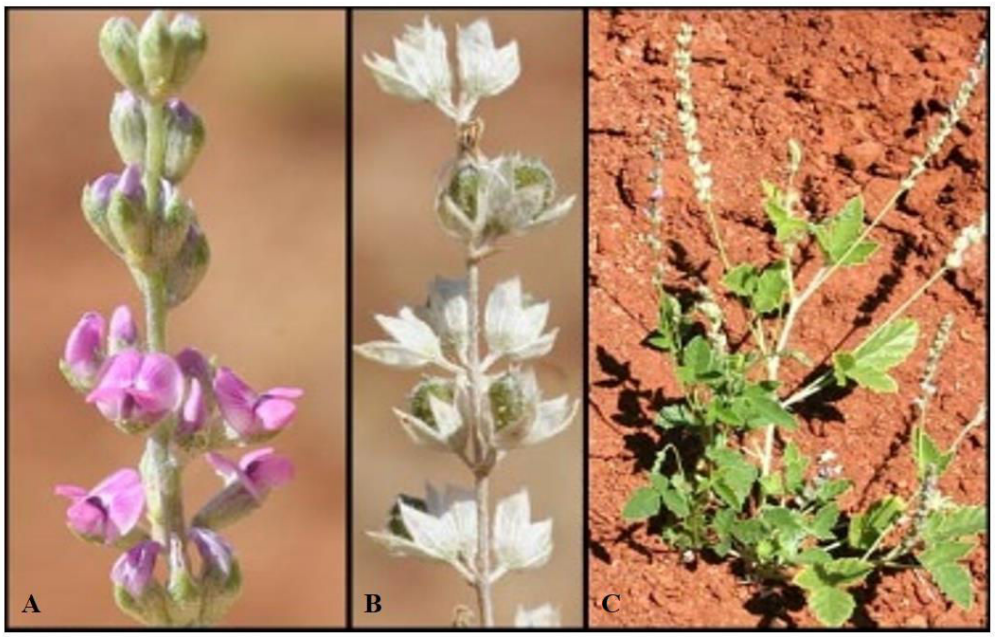


Figure 17 : Photographies de *C. cinereum*. Source : http://florabase.dpaw.wa.gov.au

2.2 Intérêt de ces trois plantes pour l'étude de la voie de biosynthèse des furanocoumarines

P. sativa, *R. graveolens* et *C. cinereum* appartiennent tous à la classe des *Magnoliopsida*, mais à des familles botaniques distinctes, respectivement les Apiacées, les Rutacées et les Fabacées. Ces trois plantes présentent une grande variété et teneur en métabolites secondaires dont certaines présentent un intérêt pour l'industrie pharmaceutique.

Des analyses effectuées sur les organes reproducteurs de panais à différents stades de développement ont permis de quantifier 24 métabolites primaires et secondaires dont des esters d'acides gras, des terpènes (dont deux nouvellement décrites chez le panais : le cubebène et le bergamotène), du palmitolactone, des furanocoumarines et de la myristicine (Zangerl *et al*., 1997). D'autres analyses effectuées sur des extraits réalisés au CO_2 supercritique ont permis de mettre en évidence chez *R. graveolens* 39 quinolines et alcaloïdes (dictamnine, ɣ-fagarine, skimmianine ou encore kokusagarine), 34 coumarines (rutacultine, rutamarine, chalepensine, chalepine, ou encore gravéoline) et furanocoumarines et plus d'une trentaine d'autres composés tels que des cétones et des aldéhydes (Stashenko *et al.*, 2000). Plus récemment, des études réalisées sur des racines transformées avec *Agrobacterium rhizogenes* cultivées en hairy roots ont démontré la présence d'autres molécules telles que l'osthole, l'osthénol ou encore la rybalinine (Sidwa-Gorycka *et al.*, 2009). Des cultures *in vitro* de rue sous différentes conditions de lumière ont mis en évidence la présence de cinq acides phénoliques,

d'acide cinnamique (Szopa *et al*., 2012). *C. cinereum* synthétise quant à lui, des flavonoïdes tels que la daidzéine, la génistéine et le coumestrol (Bourgaud *et al*., 1999).

Ces trois plantes produisent des furanocoumarines, mais qui sont présentent sous des formes chimiques différentes. *P. sativa* synthétise à la fois des furanocoumarines linéaires et angulaires, principalement du 8-MOP, de l'impératorine, du 5-MOP, de la sphondine et de l'isopimpinelline (Berenbaum *et al.*, 1984), mais la présence de psoralène, d'angélicine, d'isobergaptène et de phelloptérine (Waksmundzka-Hajnos *et al*., 2004) a également été démontrée (Peroutka *et al.*, 2007). *R. graveolens* synthétise uniquement des furanocoumarines linéaires, majoritairement du 5-MOP, mais également du psoralène, du 8-MOP, de l'isopimpinelline et de l'impératorine (Ekiert and Gomólka, 1999). Récemment de la marmésine a été caractérisée dans des cultures *in vitro* de *R. graveolens* (Szopa *et al*., 2012). Enfin, *C. cinereum* synthétise uniquement des furanocoumarines simples telles que le psoralène et son isomère angulaire, l'angélicine (Cappelletti *et al.*, 1984).

3 Les cytochromes P450

3.1 Description

Les P450s représentent une superfamille multigénique d'hémoprotéines enzymatiques (EC 1.14.14.1) à propriété redox. C'est en 1958 que Klingenberg décrit pour la première fois l'apparition d'un pigment jaune-orangé en présence de monoxyde de carbone dans des fractions microsomales de cellules de foie de rat et de cochon (Klingenberg, 1958). Ce pigment fut d'abord considéré comme étant une hémoprotéine de la famille des cytochromes *b*, mais il sera finalement désigné sous le terme de P450 (Omura and Sato, 1964). Cette appellation s'explique par la capacité de ces enzymes à présenter une forte absorption à 450 nm lorsque le fer de l'hème sous sa forme réduite (ion ferreux Fe(II)) se lie au monoxyde de carbone en présence de dithionite de sodium (Omura and Sato, 1964). Cette particularité d'absorption lui confère sa dénomination actuelle ; P pour pigment et 450 correspondant à la longueur d'onde maximale d'absorption. Depuis la découverte de cette famille d'enzymes, les études biochimique et moléculaire ce sont multipliées et les P450s sont actuellement un des sujets de recherche les plus en vogue. La superfamille des P450s a été mise en évidence chez presque tous les organismes vivants : archaebactéries, procaryotes, eucaryotes, végétaux et animaux (Werck-Reichhart and Feyereisen, 2000). Depuis l'essor de la biologie moléculaire et des techniques de séquençage massif, plus de 21 000 séquences de P450s ont été identifiées dont plus de 7 000 séquences chez les plantes (http://drnelson.uthsc.edu/CytochromeP450.html). Les P450s sont des monooxygénases capables d'activer des molécules de dioxygène en des espèces oxygénées réactives (EORs) et ainsi catalyser l'addition d'un atome d'oxygène à un atome de carbone, d'azote ou de soufre. Ces enzymes sont

principalement impliquées dans des réactions d'hydroxylations, mais elles peuvent également réaliser des déalkylations, des oxydations, des dimérisations, des désaminations et des époxydations (Mansuy, 1994, 1998) (Figure 18, p. 57). Cette grande diversité de réactions catalysées explique la présence de cette famille d'enzymes dans de nombreuses voies de biosynthèse. Lors de la réaction de monooxygénation (Figure 19, p. 58), l'activation et le clivage du dioxygène requiert dans la plupart des cas un pouvoir réducteur fourni par du NADPH ou du NADP *via* une cytochrome P450 réductase (CPR) (EC 1.6.2.4) (Guengerich, 2007; Matsubara *et al*., 1976). Les relations phylogénétiques entre les membres de cette superfamille restent à déterminer. Néanmoins, Wickramasinghe émet l'hypothèse que l'apparition des P450s daterait de plusieurs milliards d'années et qu'à l'époque où l'oxygène atmosphérique était rare, ils étaient impliqués dans des processus de détoxication de cette molécule (Wickramasinghe and Villee, 1975). Ces enzymes se seraient ensuite multipliées dans les organismes aérobiques pour générer de nombreuses voies de biosynthèse et seraient impliquées dans de nombreux processus de dégradation de molécules (Wickramasinghe and Villee, 1975). Nelson, quant à lui, déduit à partir d'analyses phylogénétiques, que le gène ancestral serait apparu il y a 3,5 milliards d'années chez la bactérie (Nelson, 1999; Nelson *et al*., 1993). Ce gène ancestral aurait une très forte identité avec CYP51, un P450 ubiquitaire à tous les organismes impliqué nottament dans la 14α-déméthylation du lanostérol chez les bactéries et de l'obtusifoliol chez les plantes, des intermédiaires dans la synthèse des stérols.

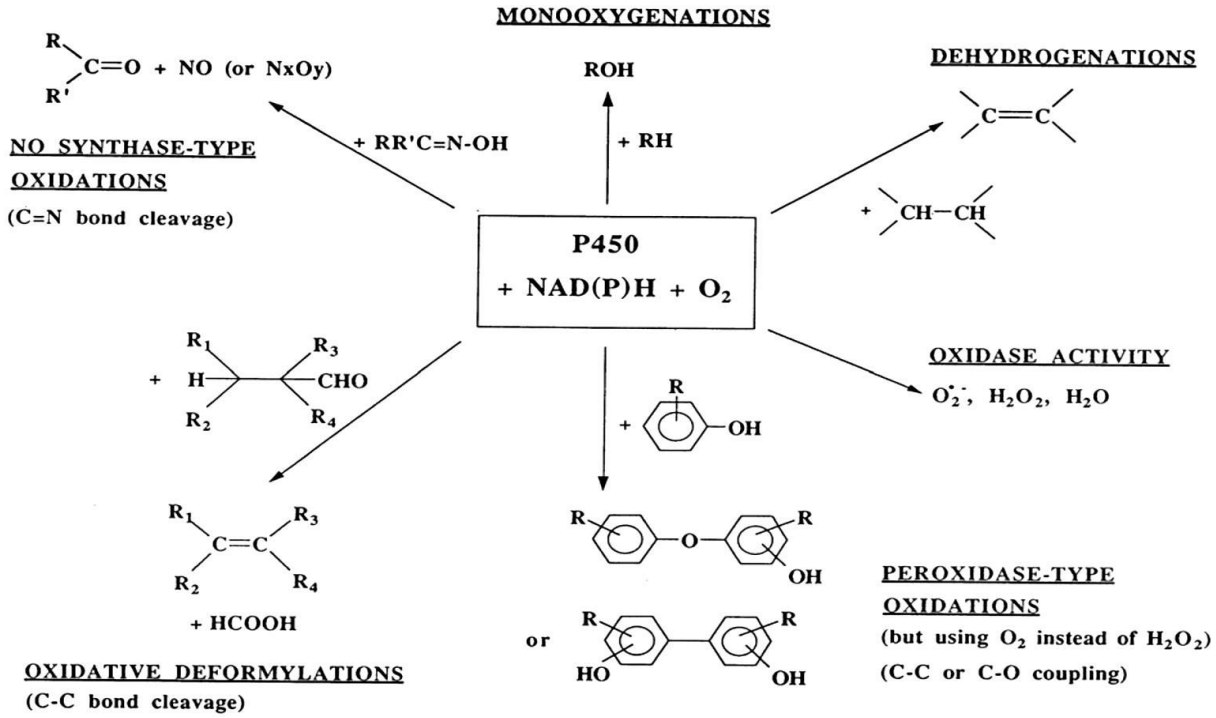


Figure 18 : Exemples de réactions catalysées par les P450s (Mansuy, 1998).

$$RH + O_2 + NADPH, H^+ \xrightarrow{\text{Cytochrome P450}} ROH + H_20 + NADP^+$$

Figure 19 : Equation bilan de la réaction de monooxygénation catalysée par les P450s.

3.2 Nomenclature

En raison de la multitude de gènes isolés, il fut rapidement nécessaire d'établir une nomenclature et une classification. La nomenclature pour l'ensemble des P450s est basée sur l'alignement des structures primaires des protéines. Une base de données établie par le Prof. David Nelson (http://drnelson.uthsc.edu/CytochromeP450.html) a été adoptée par la communauté scientifique. Les P450s sont décrits par les trois lettres CYP (pour **CY**tochrome **P**450) suivies d'un chiffre qui correspond à la famille. Lorsque deux protéines partagent au moins 40 % d'identité, elles appartiennent à la même famille. Suit ensuite une ou deux lettres qui définissent la sous-famille du P450. Les enzymes appartenant à la même sous-famille partagent au moins 55 % d'identité. Enfin, un dernier chiffre désigne l'ordre chronologique de la découverte des gènes dans la sous-famille considérée. Les séquences qui présentent plus de 97 % d'identité sont considérées comme des variants alléliques. Les familles de P450s de plantes sont les familles 51 et celles numérotées de 71 à 99 et de 701 à 736. Vu le nombre important de familles identifiées, un niveau de classification supplémentaire a été mis en place. En se basant sur des arbres phylogéniques, les familles proches sont alors réunies en clans (Nelson, 1999; Nelson *et al*., 2004). Les clans rassemblent les membres qui seraient issus d'un même gène ancestral. Ils sont désignés par le plus petit numéro des familles représentées ou par le chiffre de la famille majoritaire. A titre d'exemple, le clan 71, qui est le plus grand, regroupe les membres des familles CYP71, 73, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 89, 92, 93, 98, 99, 701, 703, 705, 706, 712, 719, 723, 726 et 736.

3.3 Classification et origine du pouvoir réducteur

Les P450s nécessitent l'apport d'un pouvoir réducteur fourni de manière générale par le NADPH ou le NADH afin de réaliser les réactions de monooxygénations. Ces enzymes sont réparties en quatre classes selon leur partenaire redox (Werck-Reichhart and Feyereisen, 2000). Les classes I et II sont majoritaires, les classes III et IV sont beaucoup moins fréquentes.

La classe I contient des P450s qui fonctionnent avec deux partenaires redox, une NAD(P)H réductase constituée d'un domaine flavine adénine dinucléotide (FAD) ou d'un domaine flavine mononucléotide

(FMN), associée à une ferrédoxine fer-soufre (Figure 20-A, B et C, p. 59). Les P450s de cette classe sont majoritairement des enzymes bactériennes, mais on trouve également des P450s mitochondriaux.

La majorité des P450s de plantes appartiennent à la classe II qui comprend des enzymes associées à une cytochrome P450 réductase NADPH dépendante (CPR) qui comprend un domaine FAD et un domaine FMN responsables du transfert d'électrons vers le site catalytique du P450 (Figure 20-D, p. 59 et Figure 21, p. 60). Cette classe est majoritaire chez les eucaryotes et contient principalement des enzymes localisées au niveau du réticulum endoplasmique, mais elle comprend également un système soluble présent chez les bactéries (Figure 20-E, p. 59). Les P450s de la classe II peuvent également recevoir des électrons *via* un cytochrome *b5* et la NADH cytochrome *b5* réductase.

Les P450s appartenant à la classe III ne nécessitent pas d'apport extérieur d'électrons ou d'oxygène moléculaire. Ces enzymes sont exclusivement eucaryotes et catalysent des réactions d'isomérisation et de déshydratation. Les électrons sont apportés par les substrats de ces enzymes.

La classe IV ne comprend à ce jour qu'un seul membre, la Nitrite Oxyde Réductase (P450nor). Cette enzyme réduit le monoxyde d'azote en N_2O et utilise directement le NADH comme source d'électrons sans recourir à des protéines intermédiaires.

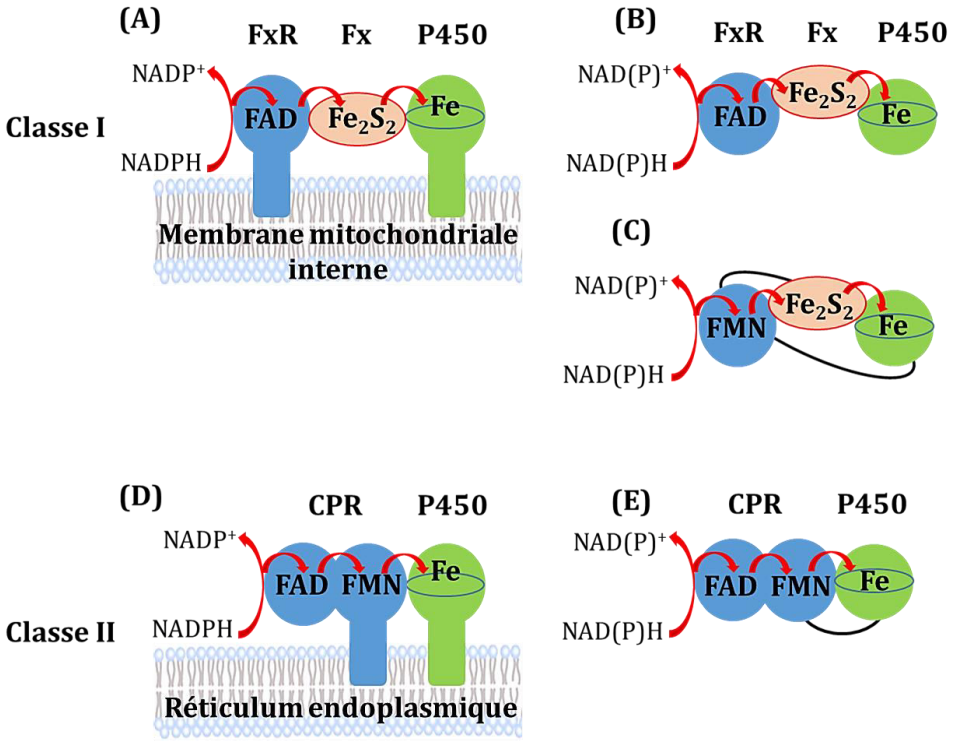


Figure 20 : Représentation schématique des différents partenaires redox des P450s (adapté de De Montellano, 2005). Fx = ferrédoxine ; FxR = ferrédoxine réductase ; CPR = Cytochrome P450 réductase.

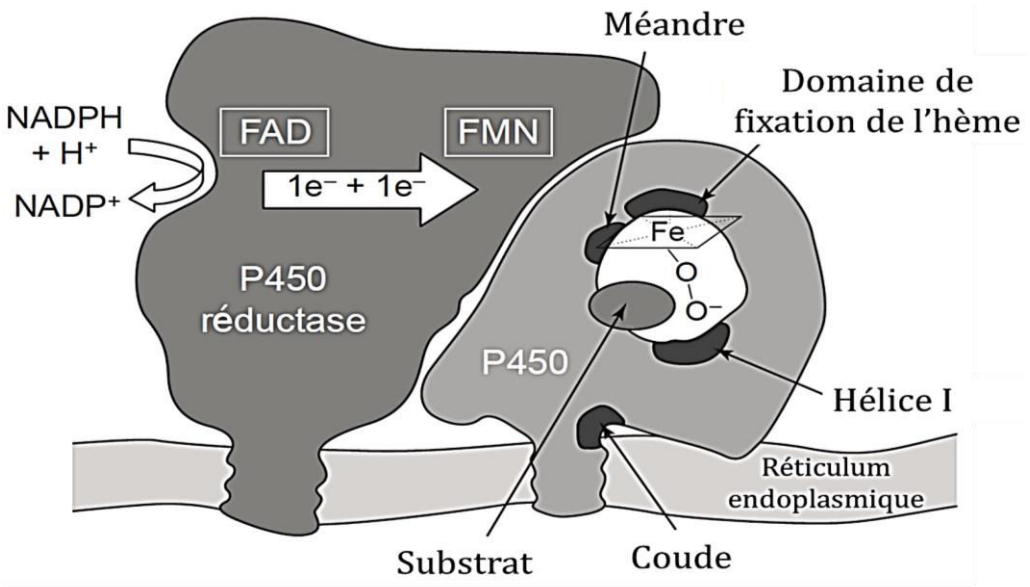


Figure 21 : Représentation schématique d'un cytochrome P450 eucaryote microsomal associé à la cytochrome P450 réductase constituée d'un domaine FAD et d'un domaine FMN (Werck-Reichhart *et al*., 2000).

Les CPRs présentent peu de sélectivité d'interaction avec les cytochromes P450. D'ailleurs, la diversité des CPRs est très faible par rapport au nombre de P450s présents dans chaque génome. A titre d'exemples, l'Homme et la levure n'ont qu'une seule CPR, les plantes quant à elles, ont généralement deux ou trois CPRs (Benveniste *et al*., 1991; Mizutani and Ohta, 1998). Ces isoformes pourraient avoir des rôles physiologiques différents chez les plantes supérieures (Ohta and Mizutani, 2004). Chez *A. thaliana*, *Ath*R1 est exprimée de manière constitutive, alors que la synthèse d'*Ath*R2 peut être induite par la lumière ou par des blessures (Mizutani and Ohta, 1998).

3.4 Structures secondaire et tertiaire et domaines conservés

Bien que les identités de séquences peptidiques entre P450s procaryotes et eucaryotes soient relativement faibles (16 % d'identité au minimum), cette famille d'enzymes présente une forte conservation de structure tertiaire. Ces résultats ont été obtenus à partir de données structurales cristallographiques de P450s cytosoliques et membranaires (Schoch *et al*., 2004; Scott *et al*., 2003; Wester *et al*., 2004; Williams *et al*., 2000a, 2000b, 2004; Yano *et al*., 2004, 2005). Ces protéines sont constituées par plusieurs hélices α et de quelques feuillets β qui s'associent en deux domaines globulaires α et β (Figure 22, p. 61).

Malgré l'importante diversité génétique des P450s, cette famille est caractérisée par plusieurs séquences protéiques consensus :

- La séquence FxxGxRxCxG située du côté proximal de l'hème en amont de l'hélice L est un domaine très conservé appelé « Cys-pocket » (Figure 22, p. 61). La cystéine de cette séquence forme une liaison thio-ester avec l'atome de fer de l'hème contribuant ainsi à la fixation et au maintien de l'hème au sein de l'enzyme.
- La séquence (G/A)Gx(E/D)T située du côté proximal de l'hème au niveau de l'hélice I est moins bien conservée au sein des P450s (Figure 22, p. 61). Cette séquence est impliquée dans l'activation de l'oxygène moléculaire lors de la réaction enzymatique.
- La séquence ExxR localisée au niveau de l'hélice K est également susceptible de stabiliser la structure du noyau de la protéine en maintenant la poche de l'hème dans sa position.

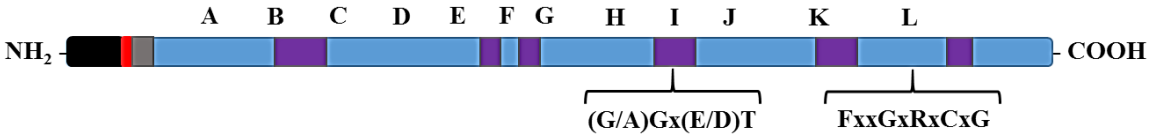


Figure 22 : Représentation schématique de la structure primaire d'un P450 membranaire eucaryote. En noir : ancre membranaire ; en rouge : acides aminés basiques ; en gris : coude proline ; en violet : séquences SRSs. Les hélices α sont annotées de A à L.

Les cytochromes P450 sont localisés à différents endroits au niveau cellulaire. Chez les procaryotes, les P450s sont des enzymes solubles dans le cytosol, alors que chez les eucaryotes ce sont des protéines membranaires distribuées dans différents compartiments cellulaires. Chez les plantes, les P450s sont le plus souvent localisés sur la face cytosolique de la membrane du réticulum endoplasmique, mais ils peuvent également être ancrés dans la membrane des mitochondries ou plus rarement au niveau de la membrane des chloroplastes (Guirimand *et al*., 2011; Quinlan *et al*., 2012).

Les P450s associés à la membrane du réticulum endoplasmique disposent d'une hélice hydrophobe d'une vingtaine d'acides aminés au niveau de l'extrémité N-terminale de l'enzyme (Figure 22, p. 61). Cette région est suivie d'une région constituée de résidus basiques qui délimite l'ancre membranaire et d'une région riche en prolines qui définit le coude de l'enzyme et permet le bon repliement de la protéine (Chen and Kemper, 1996). Cette région permet également à la partie globulaire de l'enzyme de s'accoler à la membranaire du réticulum endoplasmique et de se placer correctement par rapport à la CPR qui est également membranaire (Figure 22, p. 61) (Kemper, 2004).

Les régions de l'enzyme responsables de l'entrée et du positionnement du substrat au niveau du site actif sont appelées Substrate Recognition Site (SRSs). Ces séquences ont été pour la première fois mises en

évidence chez la famille CYP2 en comparant 51 des membres de cette famille avec 8 séquences de P450s bactériens (Gotoh, 1992). Ces SRSs, au nombre de 6, correspondent à ces régions spécifiques conservées chez tous les P450s, surtout pour les SRSs 1, 4, 5 et 6 (Figure 23, p. 62) (Williams *et al*., 2000a). L'obtention de structures cristallographiques et de modéles moléculaires ont permis de mettre en évidence plusieurs acides aminés déterminants pour l'activité de l'enzyme. L'arginine 104 du SRS 1 de CYP71AJ1 semble interagir avec l'arginine 434 pour stabiliser l'hème du site actif (Larbat *et al*., 2007).

Suivant les P450s, l'accès du substrat au site actif s'effectue par deux canaux différents. Le premier est définit par la boucle F-G qui fait intervenir les SRSs 2 et 3 (Chen and Kemper, 1996) (Wilmouth *et al*., 2002), le deuxième implique la boucle B-C. Une étude effectuée dans notre laboratoire a montré que la séquence HPP du SRS 5 est conservée chez de nombreux membres de la famille CYP71, mais elle est remplacée par YFT chez la psoralène synthase (CYP71AJ1-3) et par YIT chez l'angélicine synthase (CYP71AJ4) et semble déterminante pour l'activité de ces deux enzymes (Larbat *et al.*, 2007, 2009). De plus, la modélisation de CYP71AJ1 semble montrer que la thréonine 479 du SRS 6 est essentielle à l'activité catalytique de l'enzyme (Larbat *et al.*, 2007). Les fonctions de ces gènes sont uniquement connues pour CYP71AJ1-4 et correspondent à la psoralène synthase et à l'angélicine synthase.

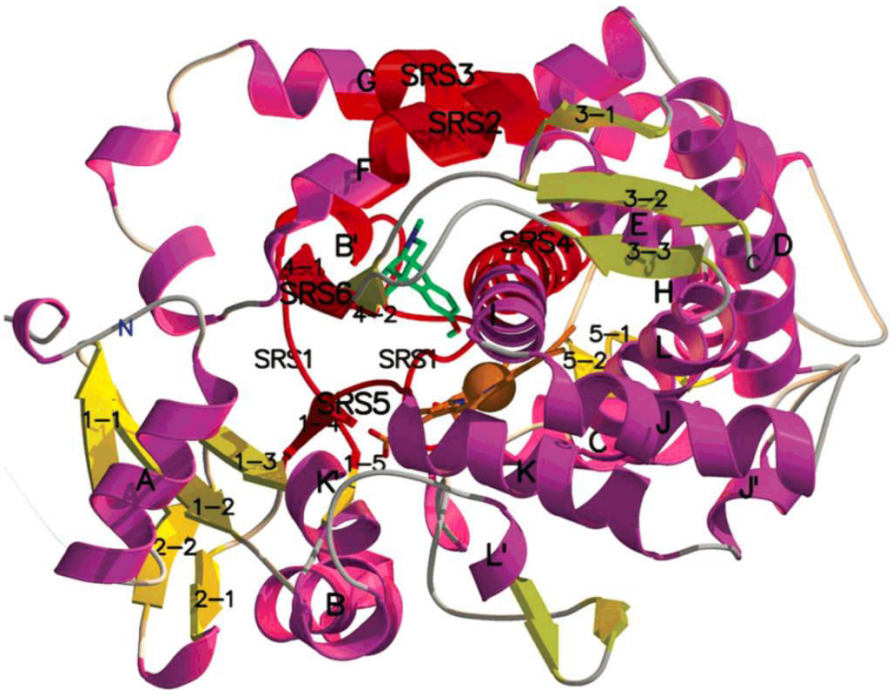


Figure 23 : Représentation de la structure tridimensionnelle de CYP2D6 complexé avec du dextrométhorphane (en vert) obtenues à partir des modèles cristallographiques de CYP101, 102, 107A, 108 et 2C5 (de Graaf *et al*., 2005).

A ce jour, tous les P450s identifiés sont des hémoprotéines d'un poids moléculaire compris entre 45 et 60 kDa. Ces hémoprotéines sont caractérisées par un groupe prosthétique central, l'hème, qui catalyse la réaction enzymatique. Le plan de l'hème définit la face proximale et la face distale contenant le site actif du P450. L'hème est relié au reste de la protéine au niveau d'une cystéine située dans la région C terminale de l'enzyme. Cette liaison thio-ester est non-covalente et complexe l'atome de fer du groupe prosthétique. L'hème est constitué d'un noyau protoporphyrine IX identique chez tous les P450s (Figure 24, p. 63). Ce noyau est caractérisé par 4 cycles pyroles qui renferment à leur centre un atome de fer capable d'effectuer 6 liaisons de coordinence. Quatre d'entre elles relient cet atome avec le tétrapyrole, une cinquième liaison permet de stabiliser le noyau protoporphyrine dans le site actif *via* la liaison thio-ester décrite précédemment et la dernière liaison peut être établie avec une molécule d'H_2O, d'O_2 ou de CO (Figure 24, p. 63).

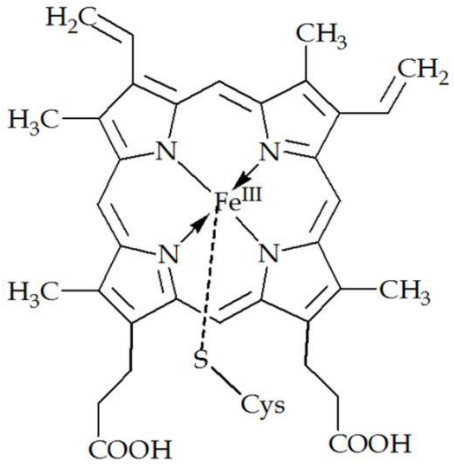


Figure 24 : Structure chimique développée d'un hème de type protoporphyrine IX.

3.5 Cycle catalytique des P450s

En l'absence de substrat, le fer du groupe prosthétique est à l'état ferrique (Fe^{3+}) et forme une liaison avec une molécule d'eau (Figure 25-A, p. 64). Le fer est à l'état de spin bas et présente une symétrie maximale par rapport au plan de l'hème. La fixation du substrat provoque un changement de conformation de la protéine. L'atome de fer passe alors vers un état de haut spin pentacoordiné et se déplace par rapport au plan de l'hème (Figure 25-B, p. 64). L'arrivée d'un électron fourni par la CPR réduit le fer ferrique (Fe^{3+}) en fer ferreux (Fe^{2+}) pentacoordiné (Figure 25-C, p. 64). L'atome de fer peut alors former une liaison avec une molécule de dioxygène pour former un intermédiaire Fe^{3+}-0_2 (Figure 25-D, p. 64) qui est à son tour réduit par un électron pour former un complexe peroxo-ferrique (Figure

25-E, p. 64). L'apport de deux protons aboutit à la formation d'un complexe ferryl-oxo et d'une molécule d'eau (Figure 25-F, p. 64). L'oxygène est alors transféré au substrat, le produit formé est relargué et l'atome de fer retourne à l'état initial. Les P450s peuvent également fonctionner en cycle court (passage de l'état B à l'état E, Figure 25, p. 64) lorsque le NADPH et l'oxygène sont remplacés par des molécules de type alkyle, hydroperoxyde, peroxyacides ou peroxyde d'hydrogène.

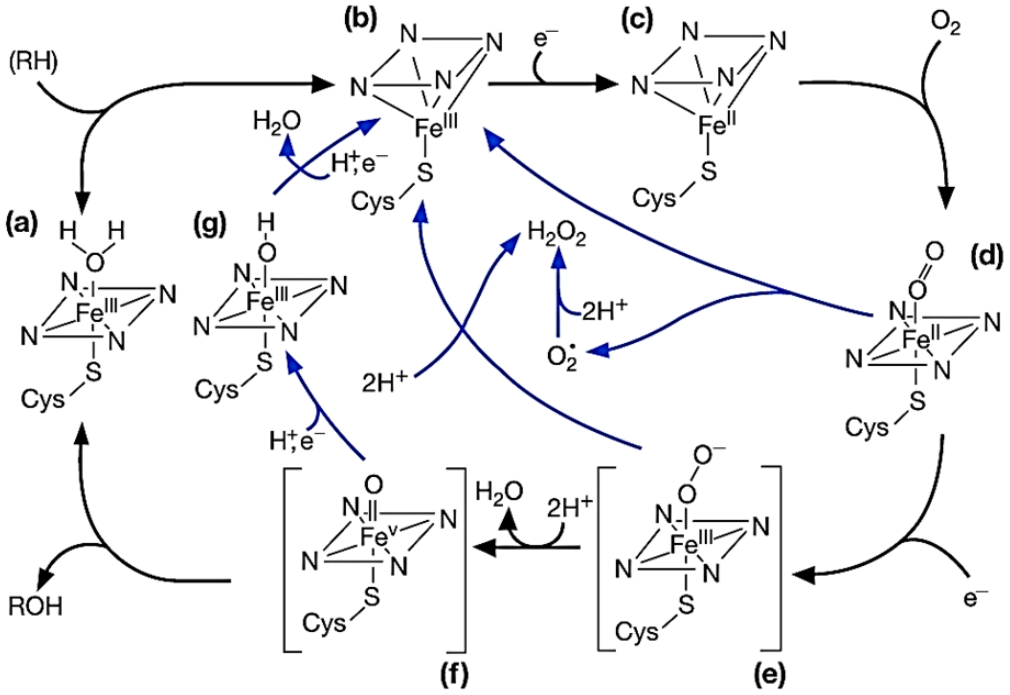


Figure 25 : Représentation schématique du cycle catalytique d'un P450 (Zangar *et al*., 2004).

De manière générale, le cycle catalytique complet s'accompagne de la consommation d'une molécule d'oxygène et de deux électrons fournis par le NADPH. Néanmoins, il existe des réactions de découplage entre la consommation de NADPH et la formation du produit qui génèrent la formation d'EORs (Figure 25, p. 64, réactions représentées par des flèches bleues). Ces molécules hautement réactives et très toxiques pour la cellule sont impliquées dans la signalisation de l'apoptose (Mignotte and Vayssiere, 1998). L'abondance des P450s chez les eucaryotes ainsi que le fort degré de découplage font que le cycle catalytique des P450s est une des source principale de stress oxydant (Zangar *et al.*, 2004). La production d'EORs est étroitement contrôlée dans la cellule par la régulation directe de l'activité des P450s ainsi qu'au niveau transcriptionnel (répression/induction d'enzymes antioxydantes telles que la catalase) (Barouki and Morel, 2001; Marí and Cederbaum, 2001).

3.6 P450s et inactivation autocatalytique

Les inactivateurs autocatalytiques, les inhibiteurs suicides ou les Mechanism-Based Inactivators (MIB) sont des analogues de substrats capables de se positionner dans le site actif de l'enzyme et d'être métabolisés. Ils ont été décrits pour la première fois en 1968 avec l'étude de la *β*-hydroxydécanoyl thioester déhydratase d'*E. coli* (Helmkamp *et al*., 1968). Ces molécules sont métabolisées en intermédiaires réactifs capables d'interagir de manière irréversible ou quasi irréversible avec l'enzyme et la rendant ainsi inactive. Les inhibiteurs suicides semblent plus efficaces et plus sensibles que les inhibiteurs classiques, car les enzymes ainsi inactivées irréversiblement sont alors retirées du pool catalytique (Gravot *et al*., 2004).

Il existe quatre types de MIBs chez les P450s selon le type de liaisons qu'ils forment avec l'enzyme. On distingue les MIBs capables de se fixer irréversiblement à l'enzyme, ceux qui se fixent de manière irréversible à l'hème, ceux qui se fixent de manière quasi irréversible à l'atome de fer de l'hème et enfin ceux qui provoquent la dégradation de l'hème. Certaines molécules comme le 17α-éthynylestradiol, peuvent alkyler l'hème et provoquer sa destruction mais elles peuvent également former des liaisons covalentes irréversibles avec l'apoprotéine au niveau du site actif (Lin *et al*., 2002). De nombreux inhibiteurs autocatalytiques de P450s ont déjà été identifiés et ces molécules présentent, en général, un groupement amine, méthylène dioxy, hydrazine ou acétylène. Des molécules contenant un noyau furane, comme certaines furanocoumarines sont également des inhibiteurs suicides efficaces de P450s (Fouin-Fortunet *et al*., 1986; Kharasch *et al*., 2000; Koenigs *et al*., 1997; Sahali-Sahly *et al*., 1996). La plupart des études d'inactivation autocatalytique ont été menées sur des P450s de mammifères appartenant aux familles CYP1, CYP2 et CYP3 (Obach *et al*., 2007; Roberts *et al*., 1993, 1997), mais certains P450s végétaux peuvent également être inactivés par des furanocoumarines. C'est le cas de la C4H qui est inactivée par l'établissement de liaison covalente avec le psoralène ou le 8-MOP. Les plantes non productrices de furanocoumarines sont beaucoup plus sensibles à l'inactivation de la C4H que celles qui en produisent (Gravot, 2002; Gravot *et al.*, 2004). Ces résultats suggèrent qu'il y aurait eu une co-évolution entre l'apparition de P450s résistants aux MIB et l'apparition des furanocoumarines afin d'assurer l'activité de la C4H en présence de MIB, ou de réguler la synthèse de ces molécules toxiques par un effet seuil (Gravot *et al*., 2004).

3.7 Rôles et intérêts des P450s

Les P450s étant capables de réaliser une grande variété de réactions, il est logique que cette famille d'enzymes soit impliquée dans de nombreuses voies métaboliques. Chez les procaryotes, les P450s sont principalement impliqués dans la biosynthèse de molécules de défense comme les antibiotiques et les

stérols, alors que chez les eucaryotes ils participent à la synthèse de molécules endogènes (hormones stéroïdiennes, métabolites primaires et secondaires) et à la dégradation de composés exogènes comme les xénobiotiques, les polluants et les pesticides. Bien que l'ingénierie des P450s soit prometteuse pour des applications biotechnologiques, il n'existe à ce jour que quelques rares applications industrielles comme la production de certaines plantes transgéniques (œillet, roses bleues) ou la production d'hydrocortisone chez la levure (Bernhardt, 2006).

3.7.1 Fonctions des P450s chez les végétaux

C'est chez les végétaux que l'on retrouve la plus grande diversité de P450s avec plus de 7000 séquences identifiées à ce jour. En raison de l'accélération du séquençage des génomes ce chiffre ne cesse d'augmenter, mais les fonctions de tous ces gènes ont été relativement peu documentées jusque là (enzyme orphelines). Cette famille d'enzyme intervient dans le métabolisme de molécules endogènes et dans le catabolisme de molécules exogènes.

3.7.1.1 Biosynthèse de métabolites secondaires

Les P450s sont impliqués dans la biosynthèse de nombreux métabolites assurant des diverses fonctions. Ils participent au maintien de la plante (lignines), à son interaction avec l'environnement (anthocyanes, flavonoïdes, …), à sa protection (glucosinolates, alcaloïdes, furanocoumarines, glucoside cyanogéniques, …) et également à sa régulation hormonale (auxines, gibbérélines, …). Dans la voie de biosynthèse des phénylpropanoïdes et de ses dérivés, plusieurs P450s ont déjà été caractérisés. Les membres de la famille CYP73A correspondent à la C4H (Park *et al*., 2010; Ro *et al*., 2001) et ceux de la famille CYP98A réalisent l'hydroxylation en 3' des esters *p*-coumaroyl CoA comme nottament le *p*-coumaroyl quinate et shikimate pour former respectivement le caffeoyl quinate et l'acide chlorogénique (Morant *et al*., 2002; Schoch *et al*., 2001). La voie de biosynthèse des flavonoïdes fait intervenir de nombreux membres de la famille CYP75 présentant une activité flavonoïde 3'-hydroxylase (CYP75B) (Seitz *et al*., 2007), flavonoïde 3'5'-hydroxylase (CYP75A) (Olsen *et al*., 2010) ou encore flavonoïde 6'-hydroxylase (CYP75D) (Latunde-Dada *et al*., 2001). La voie de biosynthèse des furanocoumarines comporte également de nombreuses étapes catalytiques P450 dépendantes, pour l'instant seule la famille des CYP71AJ a été caractérisée et catalyse la formation de psoralène et d'angélicine (Larbat *et al*., 2007, 2009). Deux autres enzymes, CYP82C2 et CYP82C4, identifiées chez *Arabidopsis*, sont capables de réaliser l'hydroxylation du 8-MOP, bien que cette plante ne produise pas de furanocoumarine (Kruse *et al*., 2008). Plusieurs enzymes ont été caractérisées dans la synthèse des terpènoïdes. C'est le cas des CYP85 et CYP90 impliqués dans la synthèse de brassinostéroïdes (Kim *et al*., 2005; Shimada *et al.*,

2001) des CYP701 et CYP88 impliqués dans la synthèse de gibbérellines (Helliwell *et al*., 2001; Wang *et al*., 2012) ou encore des CYP707 qui participent à la dégradation de l'acide abscissique (Saito *et al.*, 2004). De nombreuses familles de P450s interviennent dans la biosynthèse des alcaloïdes : CYP71D pour la tabersonine 16-hydroxylase, CYP71BJ pour la tabersonine 19-hydroxylase, CYP72A est une sécologanine synthase, ou encore CYP76B pour la géraniol/nérol 10-hydroxylase (Collu *et al*., 2001; Giddings *et al*., 2011; Schröder *et al.*, 1999). Quatre membre de la famille CYP71C catalysent des réactions successives dans la synthèse du 2,4-dihydroxy-7-méthoxy-2H-1,4-benzoxazin-3(4H)-one (DIMBOA) (Frey *et al*., 1997, 2003). Deux familles de P450s ont été démontrées à ce jour comme étant impliquées dans la synthèse des glucosinolates. CYP79A et CYP79B catalysent la première étape de cette voie qui conduit à la formation d'un oxyme qui est pris en charge par des membres de CYP83A et CYP83B pour donner un intermédiaire réactif qui sera ensuite conjugué avec une cystéine (Bak *et al*., 1998; Naur *et al*., 2003; Sibbesen *et al*., 1995).

3.7.1.2 Catabolisme de xénobiotiques

Bien que le rôle joué par les P450s dans de nombreuses voies de biosynthèse soit bien étudié, leur implication dans les processus de détoxication est moins documentée. La majorité des études de la fonction des P450s dans le catabolisme de molécules endogènes ou exogènes a été effectuée sur des P450s de mammifères, néanmoins un certain nombre d'études sont réalisées sur des P450s végétaux. Les P450s végétaux sont capables de cataboliser de nombreux composés exogènes toxiques comme de nombreuses classes d'herbicides (phénylurées, sulfonylurées, chloroacétamines, …), certains insecticides et certains polluants (Schuler and Werck-Reichhart, 2003). A titre d'exemple, CYP71B1 d'*H. tuberosus* est capable de détoxifier de nombreux herbicides de la classe des phénylurées (chlortoluron, isoproturon, …) par deux réactions d'alkylation (Didierjean *et al*., 2002; Robineau *et al*., 1998). L'ingénierie de ces P450s végétaux pourrait permettre le développement de plantes transgéniques utilisables en biotechnologie pour leur résistance aux pesticides ou en phytoremédiation pour décontaminer des sols pollués (Santosh *et al*., 2012; Werck-Reichhart *et al*., 2000). De par l'importante diversité de xénobiotiques catabolisés, il est raisonnable de penser qu'au sein d'une même plante plusieurs P450s sont impliqués dans le processus de dégradation. Certains P450s sont donc capables d'intervenir à la fois dans des processus de synthèse et de dégradation (Pierrel *et al*., 1994; Wen-Sheng *et al*., 2006) mettant ainsi en évidence le caractère opportuniste de ces enzymes et, dans certains cas, une spécificité de substrat peu importante.

Le processus de détoxication se déroule en trois étapes : transformation, conjugaison et compartimentation (Sandermann Jr, 1992). Les réactions de transformation et de conjugaison sont couplées et catalysent en général une hydroxylation ou une alkylation du xénobiotique qui le rende ainsi

plus polaire et plus soluble, suivie d'une étape de conjugaison avec des composés tels que des sucres, des groupements glutathion, ou encore des acides aminés (Kent *et al*., 2006a, 2006b). Les xénobiotiques ainsi dépourvus de toxicité sont, soit transportés et stockés dans les vacuoles, soit acheminés au niveau de la paroi cellulaire (Schuler and Werck-Reichhart, 2003). Les P450s interviennent généralement dans la première étape.

3.7.2 Les cytochromes P450 à l'origine de l'adaptation entre les plantes et les insectes

Les plantes et les insectes coexistent depuis plusieurs centaines de millions d'années et sont en étroite interaction. Les insectes phytophages ingèrent des acides aminés, des vitamines, des stérols et des minéraux végétaux essentiels à leur développement (Fraenkel, 1959) et les plantes utilisent les insectes pour la dissémination de leur pollen ou des graines. Des phénomènes adaptatifs et évolutifs ont émergé chez les plantes et les insectes qui subissent des modifications de leur métabolisme. Les plantes synthétisent de nouveaux composés allélopathiques et les insectes mettent en place de nouvelles stratégies de détoxication de ces toxines végétales. De nombreux métabolites secondaires, comme les terpénoïdes, acaloïdes, tannins et saponines présentent des propriétés allélopathiques. Ces différentes voies métaboliques contiennent toutes des étapes P450s dépendantes. La capacité d'un organisme à s'adapter à la présence de composés toxiques dans son environnement, est directement corrélée à la diversité de cytochromes P450 contenus dans leur génome (Schuler, 2011). Ces premiers résultats sont basés sur diverses observations, de nombreuses études tentent aujourd'hui d'apporter des éléments de réponse sur les phénomènes évolutifs et adaptatifs du point de vue moléculaire (Berenbaum, 2002; Bono *et al*., 2008; David *et al*., 2006; Després *et al*., 2007; Poupardin *et al*., 2008; Zumwalt and Neal, 1993).

3.7.2.1 Adaptation et synthèse de composés allélopathiques par les P450s de plantes

Au cours de l'évolution, les plantes se sont adaptées à leur environnement souvent hostile en synthétisant de nombreux métabolites secondaires aux propriétés multiples. Les pressions de sélection exercées au cours de l'évolution ont ainsi conduit à la synthèse de composés innombrables répartis sur l'ensemble du règne végétal dont seulement 200 000 ont été identifiés à ce jour (Epifano *et al*., 2007; Wu and Chappell, 2008). Plusieurs familles de molécules sont impliquées dans les processus de défense de la plante : les terpénoïdes dont plus de 40 000 membres sont décrits dans la littérature (Zulak and Bohlmann, 2010), les alcaloïdes composés de plus de 12 000 structures identifiées (Facchini, 2001; Ziegler and Facchini, 2008), les phénylpropanoïdes et leur dérivés comme les flavonoïdes, les anthocyanes (Dixon and Paiva, 1995; Murray, 2000; Ververidis *et al*., 2007) et les 200 structures

identifiées chez les furanocoumarines (Curini *et al.*, 2006). D'autres composés moins bien représentés dans le règne végétal comme certains glucosinolates, iridoïdes et glucosides d'iridoïdes, pipéramides et benzoazinoïdes seraient également impliqués dans les processus de défense de la plante et dans son adaptation à son environnement (Gierl and Frey, 2001; Halkier and Gershenzon, 2006; Richards *et al*., 2012; Whitehead and Bowers, 2013; Wilmouth *et al*., 2002). Les profils métaboliques varient considérablement d'une plante à l'autre, certaines vont synthétiser un large panel de molécules peu toxiques, alors que d'autres produisent majoritairement une classe de composés beaucoup plus toxiques capables d'effets synergiques (Cianfrogna *et al*., 2002a, 2002b).

De nombreuses étapes de ces voies de biosynthèse sont catalysées par des P450s. Comme cela a déjà été précisé plus tôt, les génomes de plantes renferment un très grand nombre de gènes permettant de coder pour ces enzymes. On en dénombre plusieurs centaines chez *A. thaliana*, la vigne ou encore le riz. Le nombre total de P450s présents dans le génome des plantes représenterait plus d'un pour cent de la totalité des gènes. Cependant, la fonction de la plupart d'entre elles reste encore à découvrir (Nelson *et al*., 2004). A titre d'exemple, la fonction d'une soixante de protéines seulement a été identifiée sur les 276 gènes identifiés chez *A. thaliana* (Chapple, 1998; Schuler and Werck-Reichhart, 2003). Il a été montré que certains de ces gènes, parfois redondants, sont regroupés physiquement sur le génome et sont régulés de la même manière que les opérons bactériens (Blanco *et al*., 2008; Michalak, 2008; Osbourn and Field, 2009). Ces clusters de P450s concernent des voies de biosynthèse particulières comme celle de la synthèse du thalianol chez *Arabidopsis* (Wilmouth *et al*., 2002), des benzoazinoïdes chez *Zea mays* (Frey *et al*., 1997), de l'avénacine chez l'avoine (Qi *et al*., 2004) et des momilactones chez le riz (Okada *et al*., 2009; Shimura *et al.*, 2007). Cette multitude de P450s chez les végétaux résulterait de différents événements génétiques tels que la duplication, les mutations, la néo- et la sub-fonctionnalisation (Bak and Hamberger, 2013; Schuler, 2011). La présence d'introns localisés à des positions identiques dans certains clusters indique que ces clusters de gènes résulteraient de phénomènes de duplication génique par enjambement (ou crossing over) (Paquette *et al*., 2000). Certaines familles de P450s ont conservé au cours de l'évolution la même fonction dans les plantes où elles sont exprimées. C'est le cas par exemple, des familles CYP51 et CYP710 impliqués dans la synthèse des stérols chez les plantes supérieures (Bak *et al*., 1997; Morikawa *et al*., 2006), de la famille CYP97 impliquée dans la synthèse des xanthophylles (Morikawa *et al*., 2006), et des familles CYP86, CYP94 et CYP704 qui interviennent dans la formation des acides gras chez la plupart des plantes vasculaires en réalisant leur *ortho*-hydroxylation afin de former les éléments de base nécéssaires à la biosynthèse des polymères de type cutine ou subérine (Le Bouquin *et al*., 2001; Höfer *et al*., 2008; Li *et al*., 2010; Mizutani and Ohta, 2010).

La majorité des autres familles de P450s sont spécifiques d'une ou de plusieurs plantes de la même espèce. C'est le cas de la sous-famille CYP79F présente uniquement chez les Brassicacées et qui

intervient dans la synthèse de composés aliphatiques glucosinolés (Chen *et al*., 2003), de la sous-famille CYP79D qui est impliquée dans la synthèse de linamarine chez certaines Fabacées (Forslund *et al*., 2004) et de la sous-famille CYP71AJ qui intervient dans la voie de biosynthèse des furanocoumarines linéaires et angulaires chez les Apiacées (Larbat *et al*., 2007, 2009).

3.7.2.2 Adaptation et détoxication de toxines par les P450s d'insectes

En dépit des nombreuses voies de biosynthèse de molécules de défense mis en œuvre par les plantes, certains insectes phytophages ont réussi à s'adapter à la présence de ces molécules toxiques. Nous l'avons vu précédemment, les P450s peuvent être impliqués dans la détoxication de xénobiotiques chez les plantes. Il en va de même pour certains P450s d'insectes qui sont capables d'inactiver et/ou d'éliminer ces toxines en réalisant une hydroxylation ou une époxydation (Schuler, 1996). Une grande diversité de P450s d'insectes ont été isolés à ce jour, néanmoins leur nombre est largement inférieur à celui chez les plantes (Tijet *et al*., 2001). A titre d'exemples, *Anopheles gambia* et *Aedes aegypti*, vecteurs de la malaria, de la dengue et de la fièvre jaune, possèdent respectivement 105 et 153 gènes codant pour des P450s (Schuler, 2011). Afin d'élucider les phénomènes évolutifs qui ont permis aux insectes de s'adapter à la présence de toxines, de nombreuses études ont porté sur l'adaptation d'insectes phytophages s'attaquant à des plantes productrices de furanocoumarines (Berenbaum, 1990; Schuler, 1996). Huit insectes aux régimes alimentaires différents sont démontrés comme étant capables de dégrader les furanocoumarines grâce à des P450s majoritairement à la sous-famille CYP6B. Plusieurs membres des CYP6B ont été mis en évidence chez des insectes polyphages, comme *Papillio glaucus*, *P. multicaudatus* ou *Helicoverpa zea*, et chacun de ces P450s catalyse la dégradation de nombreux composés (furanocoumarines linéaires et angulaires, flavonoïdes, insecticides, …) (Schuler, 2011). CYP6B1 est mis en évidence chez *Papilio polyxenes*, un insecte oligophage consommant quelques plantes productrices de furanocoumarines. CYP6B1 métabolise quelques furanocoumarines linéaires et lorsque ce P450 est co-exprimé avec une CPR d'insecte, il peut également dégrader des formes angulaires (Wen *et al*., 2003). Chez *Depressaria pastinacella*, un parasite monophage spécifique du panais, CYP6AB3 ne métabolise que l'impératorine et la myristicine et serait le P450 le plus spécialisé dans la dégradation des furanocoumarines (Schuler, 2011). Cette grande diversité de gènes présentant des activités diverses plus ou moins spécifiques sont à l'origine des différents types de régime alimentaire chez les insectes (Mao *et al*., 2007; Wen *et al*., 2006). L'expression de certains P450s d'insectes est induite par des éliciteurs et permet une réponse rapide en cas d'ingestion. C'est le cas de CYP6B1 et de CYP6B3 chez *P. polyxenes* qui sont fortement exprimés en présence de 8-MOP et faiblement exprimés en présence d'angélicine (Hung *et al*., 1995; Petersen *et al*., 2001). Chez *H. zea*, CYP6B8 métabolise une grande variété de molécules (large panel de furanocoumarines et quelques

insecticides) et sa synthèse peut être induite par de nombreux dérivés phénylpropanoïdes dont les furanocoumarines (Wen *et al*., 2009; Zeng *et al*., 2009). De manière générale, les insectes polyphages possèdent des P450s dont l'expression est inductible par une plus grande diversité de molécules que celle observée chez les insectes monophages. Des approches de mutagénèse dirigée ou de modélisation moléculaire chez les CYP6B ont permis de mettre en évidence l'importance de certains acides aminés pour la fonction de l'enzyme. Les résidus phénylalanine 116 et histidine 117 du SRS1 et phénylalanine 371 et phénylalanine 484 du SRS6 sont essentiels pour le positionnement du substrat au sein du site actif (Baudry *et al*., 2003; Wen *et al*., 2006). Les sites SRSs 2, 5 et 6 de CYP6B1 contiennent de multiples résidus aromatiques qui sont remplacés par des résidus non polaires chez CYP6B8 ce qui entraîne un changement de polarité au niveau du site actif. La nature de ces acides aminés déterminerait le panel de molécules métabolisées : plus le nombre de résidus polaires est élevé, plus la variabilité de réactions catalysées sera grande (Mao *et al*., 2007; Schuler, 2011).

3.7.2.3 P450s et co-évolution

Le phénomène de co-évolution décrit pour la première fois en 1964 défini l'apparition d'un trait évolutif dans une espèce en réponse à un trait d'une autre espèce qui elle-même va évoluer (Ehrlich and Raven, 1964; Janzen and Muller, 1969; Wheat *et al*., 2007). Les plantes ont co-évolué avec les insectes, elles produisent de nouvelles molécules de défense que les insectes seront capables par la suite de détoxifier. Certaines plantes de la famille des Apiacées synthétisent des furanocoumarines linéaires et d'autres des linéaires et des angulaires. L'apparition des formes angulaires serait ultérieure à celle des linéaires et résulterait d'une adaptation aux insectes devenus résistants aux formes linéaires (Berenbaum and Feeny, 1981). Les P450s sont des enzymes très importantes pour l'évolution, car sous la pression de sélection de nouvelles fonctions apparaissent. De récentes études sont conduites sur la co-évolution plante-insecte entre le genre *Bursera* et son prédateur *Blepharida*, un coléoptère. Cet arbrisseau synthétise des molécules de la famille des terpènes (Becerra *et al*., 2001). Les profils terpéniques varient considérablement au sein des différentes espèces, certaines plantes ne produisant qu'un composé alors que d'autres en synthétisent plus d'une trentaine (Becerra, 2007; Evans and Becerra, 2006; Evans *et al.*, 2000). Certains *Blepharida* sont monophages alors que d'autres attaquent plusieurs espèces de *Bursera* différentes (Becerra, 2007). Il existe une forte corrélation entre l'apparition de nouvelles espèces de *Blepharida* et l'apparition de nouvelles molécules de défense chez cette plante (Becerra *et al.*, 2009). Toutefois, l'implication des P450s dans la modification de la synthèse des terpènes n'a pas été investiguée à ce jour dans la co-évolution *Bursera*/*Blepharida*. Une rare étude décrit l'implication des P450s dans la co-évolution plante/insecte. Les membres CYP51 sont impliqués dans la synthèse des stérols chez les animaux, les plantes et les champignons. Deux gènes, *Sad1* et *Sad2*, correspondant à des

P450s de la sous-famille CYP51H, ont acquis une nouvelle fonction par des phénomènes de duplication et de divergence et interviennent dans deux étapes successives de la production d'avenacines, des molécules ayant des propriétés antimicrobiennes (Qi *et al*., 2006). Il y a donc co-évolution entre la voie de biosynthèse des stérols et celle des avenacines, et ces dernières dériveraient des stérols.

4 Analyse de transcriptomes

Les plantes produisant des furanocoumarines ne sont pas des plantes modèles et les données génomiques les concernant sont rares dans les bases de données (seulement 137 séquences nucléotidiques disponibles dans Genbank pour *R. graveolens* et 40 pour *P. sativa*). Pour réaliser une analyse du transcriptome de ces plantes, et ainsi faciliter l'identification de gènes impliqués dans la synthèse de certaines molécules, il est essentiel de compléter les données disponibles dans les bases de données par de nouvelles séquences. Depuis quelques années de nouvelles méthodes de séquençage ont émergé qui ont permis, à ce jour, de réaliser le séquençage de plus de 1000 organismes. Les approches tradionnelles basées sur l'utilisation de la méthode de Sanger couplée à la séparation des fragments d'ADN par électrophorèse (sur gel ou en tube capilaire) ont été détrônées par des nouvelles technologies transversales ne faisant pas seulement intervenir des méthodes de séquençage améliorées mais également des techniques analytiques plus poussées avec l'utilisation de fluorochomes, de PCR en émulsion, de capteurs CCD de haute précision etc….

4.1 Méthode de séquencage « traditionnelle » de Sanger

Cette méthode est basée sur la polymérisation d'un brin d'ADN complémentaire à un brin matrice, à partir d'un oligonucléotide (ou amorce) grâce à l'ADN polymérase ADN dépendante (Sanger *et al*., 1977). Le milieu réactionnel contient le fragment d'ADN à séquencer, la polymérase, des désoxyribonucléotides triphosphates (dNTPs : dATP, dCTP, dGTP, dTTP) ainsi qu'une faible quantité de didésoxiribonucléotides triphosphates (ddNTPs : ddATP, ddCTP, ddGTP, ddTTP). L'incorporation aléatoire des ddNTPs dans le nouveau brin synthétisé entraîne l'arrêt de l'élongation, car l'absence de leur groupement 3'-OH empêche la formation d'une liaison 5'-3' phosphodiester nécessaire à la polymérisation. Afin de séquencer entièrement un fragment d'ADN, cette réaction est réalisée quatre fois en parallèle, chacun des mélanges réactionnels ne comportant qu'un seul ddNTP. Les fragments d'ADN ainsi obtenus sont séparés par électrophorèse sur gel de polyacrylamide et l'analyse de la séquence s'effectue de la même manière qu'avec la méthode de séquençage chimique. La détection des fragments synthétisés est réalisée grâce à l'incorporation d'un marqueur dans le brin synthétisé. Les premiers travaux de séquençage utilisaient des radiomarqueurs qui ont laissé place progressivement aux

fluorochromes. Il est à noter que cette méthode n'est plus réalisée en laboratoire de nos jours, car elle est trop longue et très coûteuse.

4.2 Séquençage haut débit

Les améliorations de la méthode de séquençage de Sanger ont conduit cette technique à atteindre son rendement maximum. Les seules possibilités d'accélération sont la multiplication des gels et capillaires, des séquenceurs, car si la vitesse d'électrophorèse est augmentée cela causerait une diminution de la qualité du séquençage. De nouvelles technologies ont été créées et elles permettent le séquençage base par base en temps réel. Il existe actuellement principalement quatre technologies haut débit, une cinquième est en cours de développement depuis 1995 (Rusk, 2009). Celle-ci utilise des nanopores, mais elle doit encore faire face à de nombreux défis technologiques avant d'être mise sur le marché.

4.2.1 Technologie 454 (Roche)

Le séquenceur 454 mis au point par Roche repose sur une technique de pyroséquençage (Figure 26-E, p. 74) (Ronaghi *et al*., 1998). Au cours de l'élongation, à chaque fois qu'un dNTP est utilisé par l'ADN polymérase, une molécule de pyrophosphate inorganique (PPi) est relâchée. La mesure de la quantité de PPi produite est réalisée à l'aide d'ATP sulfurylase, de luciférase, d'adénosine 5'phosphosulfate et de luciférine. Lorsqu'une molécule de PPi est relâchée, l'ATP sulfurylase le converti en ATP à partir de l'adénosine 5'phosphosulfate. La luciférase utilise alors cet ATP pour convertir la luciférine en oxyluciférine. Cette dernière réaction émet un signal lumineux proportionnel à la quantité d'ATP utilisée et donc de dNTPs incorporés.

La technologie 454 nécessite la préparation d'une librairie d'ADN simple brin en vue du séquençage. Cette approche peut être utilisée pour séquencer soit de l'ADN génomique (ADNg), soit des ARN messagers (ARNm) au préalablement rétrotranscrits en ADN complémentaires (ADNc). L'ADNg est au préalable fractionné par des traitements physiques (nébullisation). Les ARNm étant généralement de petite taille, ce fractionnement n'est pas nécessaire lorsqu'on souhaite séquencer un transcriptome. Les fragments d'ADNg ou d'ADNc obtenus sont ligués à des séquences adaptatrices (ou adaptateurs) à leurs extrémités (Figure 26–A, p. 74). Un de ces adaptateurs comporte une biotine en 5' qui va servir à fixer les fragments sur les billes grâce à un système streptavidine/biotine. Les fragments d'ADN sont ensuite dénaturés, seules les séquences simple brin possédant les deux adaptateurs sont conservées en vue d'être séquencées. Les fragments d'ADN simple brin sont ensuite fixés sur des microbilles à raison d'un fragment par bille (Figure 26-B, p. 74) puis sont amplifiés grâce à une PCR par émulsion (emPCR) : les

billes sont émulsionnées dans un mélange eau/huile afin d'inclure un seul couple bille-ADN simple brin par gouttelette d'eau. Chaque gouttelette fonctionne comme un puits pour la réaction de PCR. A la fin de la réaction de PCR, chaque gouttelette contient une bille sur laquelle sont fixés tous les fragments amplifiés (Figure 26-C, p. 74).

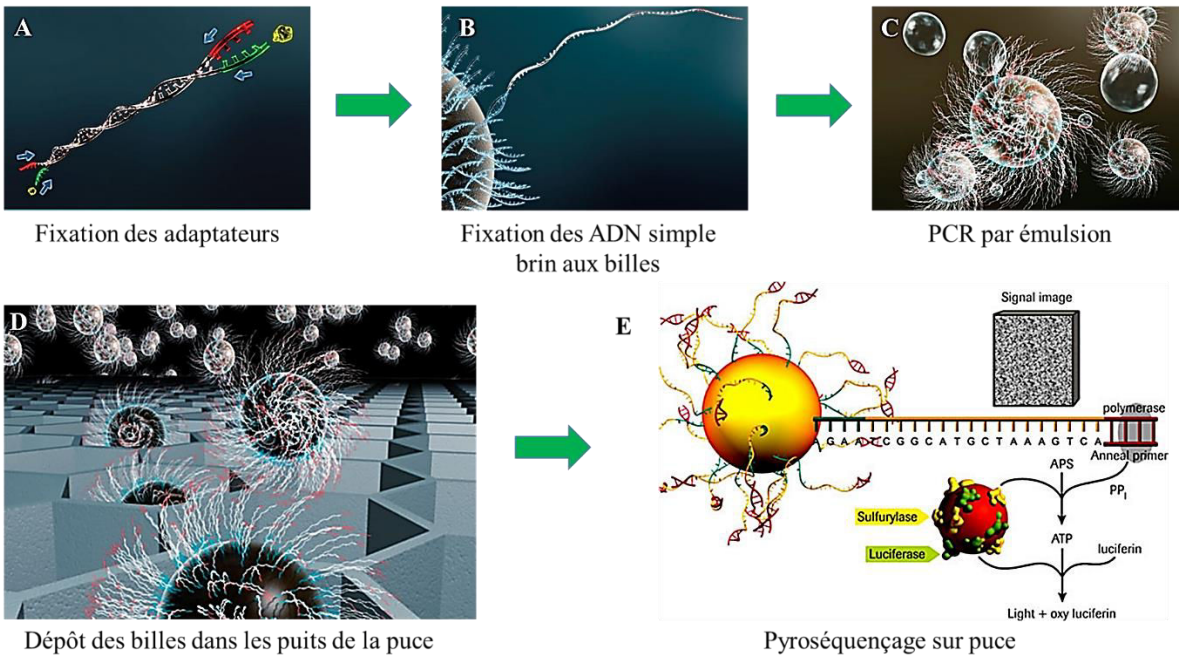


Figure 26 : Représentation schématique des différentes étapes du pyroséquençage 454 de Roche (source : http://www.454.com/).

Les billes ainsi couvertes de clones d'ADN simple brin sont alors additionnées au mélange réactionnel contenant l'ADN polymérase et étalées sur une microplaque comprenant plusieurs centaines de milliers de puits (le diamètre d'un puits est inférieur à 45µm) (Figure 26-D, p. 74). La puce est conçue de manière à ce qu'un puits ne puisse contenir qu'une seule bille. Les puits sont ensuit recouverts par des billes ayant à leur surface la sulfurylase et la luciférase. Le pyroséquençage est effectué avec l'ajout de chacune des bases de manière séquentielle (Figure 26-E, p. 74). Les intensités de lumière émises sont enregistrées sur chaque puits grâce à un capteur adapté après l'ajout successif de chaque base. Les bases non utilisées sont finalement dégradées par une solution contenant une apyrase. Les cycles successifs contenant les différentes bases sont généralement réalisés 42 fois et conduisent à l'obtention de séquences ayant une longueur moyenne de 100 pb avec un total de 20 millions de bases séquencées en 4h30 (sur l'ensemble de la microplaque). Le séquenceur 454 GS20 utilise une puce qui contient 200 000 puits d'environ 44 µm de diamètre, ce qui permet d'effectuer des réactions de séquençage en parallèle de manière massive. Le séquenceur 454 FLX Titanium utilise quant à lui des puces ayant une capacité de plus d'un million

de puits. Les séquences ainsi obtenues ont une longueur moyenne de 500 pb avec un total de 400 à 600 millions de bases séquencées en 10 h.

4.2.2 Technologie Illumina Solexa

La technologie Illumina est fondée sur des principes identiques à ceux de la technologie 454. La parallélisation du séquençage repose également sur l'utilisation d'une matrice de clones attachés à une surface solide, et le séquençage s'effectue toujours en temps réel. Néanmoins, avec la technologie Illumina, le séquençage s'effectue grâce à des fluorochromes fixés aux dNTPs. Comme pour la technologie 454, il est nécessaire de préparer une librairie d'ADN simple brin pour le séquenceur Solexa. Des adaptateurs sont fixés aux extrémités des fragments d'ADN (ADNg fragmenté par nébullisation ou ADNc correspondant aux ARNm présents dans la cellule). Ces morceaux d'ADN sont ensuites dénaturés et hybridés à des amorces complémentaires aux adaptateurs. Ces amorces auront au préalable été fixées sur une lame. Après une phase d'élongation et de dénaturation, les fragments simple brin obtenus forment un pont au niveau du support grâce à une autre amorce fixée à proximité. Ces fragments sont alors réamplifiés afin de former une structure « en pont » double brin. L'ADN est à nouveau dénaturé et les fragments simple brin obtenus forment également des ponts sur le support. Cette étape est répétée 35 fois pour former des clusters de 2 000 fragments. Les fragments ainsi obtenus sont ensuite séquencés en ajoutant un mélange des quatre dNTPs terminateurs marqués. Ces dNTPs terminateurs ont la particularité d'agir comme des ddNTPs, ils bloquent la polymérase après avoir été incorporés. Néanmoins, ces dNTPs terminateurs ont la capacité d'être réversibles. Lorsque la première série de dNTPs terminateurs est incorporée, les intensités lumineuses des différents fluorochromes sont enregistrées. Les dNTPs terminateurs sont alors inactivés et un nouveau cycle peut commencer.

4.2.3 Technologie Helicos tSMS

La technologie du séquenceur Helicos tSMS (True Single Molecule Sequencing) est une amélioration de la technologie Illumina. Le détecteur de l'Helicos tSMS est extrêmement performant et est capable d'enregistrer la fluorescence émise par un seul brin d'ADN. Aucune amplification préalable n'est donc nécessaire. Les fragments d'ADN à séquencer sont fixés sur une surface plane avec une densité très élevée. Les fragments séquencés ont une longueur moyenne de 50 pb et ce séquenceur a la capacité de séquencer plus de 3 milliards de bases par jour.

4.2.4 Technologie Applied Biosystems SOLiD

La technologie Supported Oligo Ligation Detection (SOLiD) a été créée par la société Applied Biosystems. L'étape de préparation de la librairie est comme pour la technologie 454, réalisée par une PCR en émulsion. A la fin de cette étape, les extrémités 3' des ADN simple brin sont modifiées avec un adaptateur pour permettre leur fixation sur une plaque de verre. Chaque brin d'ADN est alors associé à une amorce à proximité de lui pour permettre sa réplication. La particularité de cette technologie réside dans le fait qu'elle n'utilise pas de polymérase au cours du séquençage, mais une ADN ligase. Le séquençage est basé sur un mélange réactionnel d'oligonucléotides marqués par des fluorochromes. Cette technologie génère des séquences de 50 pb environ et peut séquencer plus de 2 milliards de bases par jour.

4.2.5 Et dans le futur ?

Avec l'essor de l'automatisation et des nouvelles technologies de séquençage, le séquençage est devenu de plus en rapide et abordable. Alors qu'il a fallu près de 10 ans pour entièrement séquencer le génome humain, il est possible avec les technologies actuelles de le faire en une seule journée. D'ailleurs certaines célébrités ont récemment fait séquencer leur génome (Levy *et al.*, 2007; Wheeler *et al*., 2008). Le séquençage complet de génomes pourrait permettre l'étude de variations génétiques sur une grande population, mais permettrait également de traiter spécifiquement certains patients. Même si les capacités journalières de séquençage des séquenceurs haut-débit sont 300 à 1100 fois plus élevées que celle d'un séquenceur classique, ces technologies présentent en général un taux d'erreur de 1 à 2 % (Schröder *et al*., 2009) alors que les séquenceurs classiques ont un taux d'erreur de 1 pour 10 millions (Mandel *et al*., 2008). De plus, ces nouveaux séquenceurs produisent des séquences de 50 à 500 pb contre 1200 pb pour un séquençage classique. Ces séquenceurs haut-débit nécessitent donc de lourds traitements d'assemblage de données. Des méthodes alternatives se sont également développées ces dernières années pour réaliser des séquençages de très long fragments d'ADN (PACBIO, www.pacificbiosciences.com) et deviendront vraisemblablement les méthodes de séquençage de génome de référence dans les années à venir. La course est lancée à celui qui proposera en premier le séquençage du génome humain en une journée pour 1000 dollars.

4.3 Assemblage de séquences

La majorité des techniques de séquençage haut débit utilisés actuellement génère des fragments ayant une taille d'environ 500 à 800 nucléotides. Il est donc nécessaire d'effectuer un assemblage des

fragments séquencés afin d'obtenir les séquences complètes des gènes transcrits dans le cas d'une banque d'ADNc voire des chromosomes entiers dans le cas d'un séquençage entier de génome.

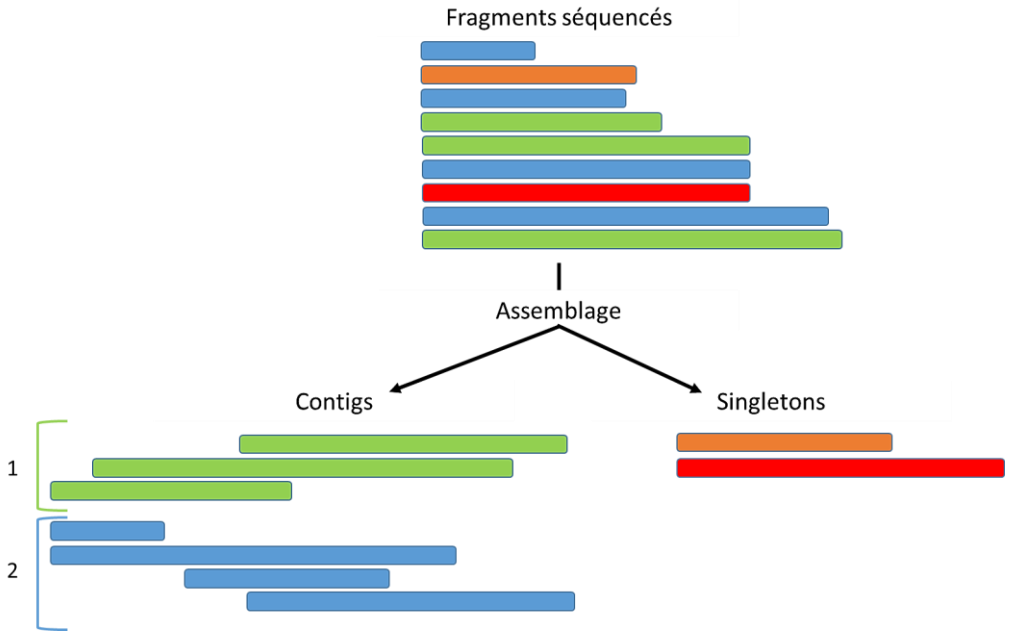


Figure 27 : Représentation schématique du processus d'assemblage de séquences. Les fragments de couleur identique se chevauchent et peuvent être rassemblées dans un même contig.

Le processus d'assemblage des fragments d'ADNg ou d'ADNc séquencés consiste dans un premier temps à identifier les parties qui se recouvrent partiellement dites chevauchantes des différents fragments. Si deux fragments se recouvrent, ils sont alors réunis pour former un « contig » (Figure 27, p. 77). Puis, ce contig sera comparé aux autres fragments de la banque afin de prolonger sa séquence. Ces comparaisons doivent tenir compte des contraintes du sens de séquençage de chaque fragment. La méthode d'assemblage permet donc d'obtenir des séquences de plus grande taille et dont la qualité est améliorée. Les séquences qui ne peuvent être assemblées en contig sont appelées « singletons » (Figure 27, p. 77). Il existe plusieurs programmes informatiques permettant cette opération d'assemblage de séquences nucléotidiques comme par exemple Cap3, Phrap ou encore la suite logicielle Staden (Bastide and McCombie, 2007; Huang *et al*., 2008; Staden *et al*., 1999).

4.4 Annotations de séquences

En raison de l'essor des nouvelles technologies de séquençage, une quantité massive de séquences ont été identifiées à ce jour. Le travail d'annotation consiste à identifier la structure et la fonction des séquences identifiées. L'annotation structurale consiste à identifier et à localiser les éléments génétiques

qui composent la séquence. L'annotation fonctionnelle, quant à elle, consiste à déterminer la fonction de ces éléments génétiques.

4.4.1 Annotation structurale

L'annotation structurale permet d'identifier et de localiser les éléments génétiques qui composent la séquence. Pour les séquences d'ADNg, il s'agit principalement des gènes, des introns et exons, des régions codantes (CDS : Coding Sequence), mais également des régions promotrices et non-traduites Untranslated Region (UTR). Alors que dans le cas des séquences d'ADNc, on cherche majoritairement à localiser la région codante.

La prédiction de séquences nucléiques codant pour la synthèse de protéines consiste à identifier la région d'ADN transcrite qui peut conduire à la formation de plusieurs variants d'épissage. Il existe de nombreux programmes de prédiction de gènes spécialisés pour les procaryotes comme GeneMarkS, GeneMark.hmm et Glimmer (Besemer *et al*., 2001; Delcher *et al*., 2007; Lukashin and Borodovsky, 1998) ou pour les eucaryotes tels que GENSCAN, GeneZilla, TigrScan ou encore CONTRAST (Burge and Karlin, 1997; Gross *et al*., 2007; Majoros *et al*., 2004). L'architecture des gènes protéiques procaryotes est plus simple et mieux connue que celle des gènes protéiques eucaryotes. Les gènes procaryotes sont caractérisés par l'absence d'intron dans leur séquence. Les programmes spécialisés dans la recherche de gènes procaryotes sont capables de prédire la structure de ces gènes avec une sensibilité et une spécificité de plus de 90% (Besemer *et al*., 2001; Delcher *et al*., 2007). Les gènes eucaryotes quant à eux, possèdent des régions promotrices beaucoup plus variables et peuvent contenir des introns ce qui rend la tâche de prédiction beaucoup plus difficile. Néanmoins, de récents programmes ont été créés et prennent en compte des motifs spécifiques comme les îlots CpG, les sites d'épissage ou encore les sites de polyadénylation et permettent de prédire les gènes eucaryotes avec une spécificité allant de 40 à 60 % (Gross *et al*., 2007; Majoros *et al*., 2004). Il existe également des logiciels comme tRNAscan-SE qui permettent de localiser les motifs et appariements caractéristiques des ARN de transfert (ARNt) (Lowe and Eddy, 1997). Les ARN ribosomaux (ARNr) quant à eux, peuvent être facilement identifiés par comparaison de séquences avec des bases de données d'ARNr comme SILVA ou Comparative RNA Web en raison de la forte conservation qu'il existe entre eux (Cannone *et al*., 2002; Pruesse *et al.*, 2007). Mais il existe également des logiciels permettant leur localisation comme par exemple le programme RNAmmer (Lagesen *et al*., 2007). Enfin, la banque Rfam regroupe les différentes familles d'ARN non codants (Gardner *et al*., 2009).

La détection de la région codante des ARN messagers est effectuée par des logiciels qui modélisent le biais d'utilisation en héxanucléotides dans les régions codantes et dans les UTRs et prennent en compte les sites d'initiation et de terminaison de la traduction (ATG et codon STOP). Ces logiciels sont dans la plupart des cas capables de corriger les erreurs de séquençage en identifiant les décalages de cadre de lecture qui engendrent de faux codons STOP. A titre d'exemples, des logiciels comme ESTScan2, FrameDP, DIANA-EST ou encore DECODER peuvent être utilisés pour prédire les CDS (Fukunishi and Hayashizaki, 2001; Gouzy *et al*., 2009; Hatzigeorgiou *et al*., 2001; Lottaz *et al*., 2003). Néanmoins, ces logiciels ne seraient capables de prédire qu'avec une spécificité de 60 %. Plus récemment, le logiciel prot4EST a été développé (Wasmuth and Blaxter, 2004). Il s'appuie sur les méthodes de prédictions d'ESTScan2 et de DECODER, mais il repose également sur une méthode par similarité de séquence par BLASTX.

4.4.2 Annotation fonctionnelle

Dans le contexte de l'annotation, une séquence peut présenter plusieurs fonctions. Ces fonctions sont souvent moléculaires, mais elles peuvent également désigner la localisation subcellulaire, les domaines fonctionnels, les signaux de signalisation ou encore l'appartenance à une voie métabolique.

4.4.2.1 Annotation de définition fonctionnelle

L'attribution d'une annotation fonctionnelle est généralement réalisée à l'aide des programmes BLAST ou PSI-BLAST par recherche de similarité avec des séquences dont la fonction a déjà été caractérisée (Altschul *et al*., 1997). Ces programmes comparent la séquence recherchée avec celles d'une ou de plusieurs banques en réalisant des alignements deux-à-deux, locaux et non optimaux. Afin d'estimer la fiabilité des résultats de recherches ceux-ci sont dotés d'un score et d'une valeur d'espérance mathématique (*E-value* : Expect value). Plus l'*E-value* est proche de 0, plus l'alignement est significatif. Une séquence présentant une *E-value* inférieure ou égale à 0,001 est en général considérée comme significative avec la séquence d'intérêt. Cette méthode de recherche est très rapide, mais peut générer des erreurs d'annotation (Sjölander, 2004). Il est également possible d'attribuer une annotation fonctionnelle à l'aide de banques comme par exemple la banque TIGRFAMs qui regroupe de nombreuses séquences de familles ou sous-familles protéiques ayant la même fonction moléculaire (Selengut *et al*., 2007), ou la banque HAMAP qui elle contient des séquences bactériennes (Gattuso *et al*., 2007).

Lorsqu'il n'est pas possible d'effectuer l'annotation par similarité de séquences, d'autres approches de génomique comparative peuvent donner des indications quant à la fonction potentielle d'un gène. Chez

les procaryotes, les gènes appartenant à un même opéron sont co-transcrits et codent le plus souvent pour des protéines interagissant physiquement et/ou fonctionnellement ensembles au sein d'une voie métabolique. Il est donc possible d'annoter fonctionnellement les différents gènes de l'opéron (Overbeek *et al.*, 1999). De la même manière, la méthode de la « Pierre de Rosette » utilise la variation d'organisation des gènes entre organismes et la modularité des protéines afin de réaliser une annotation fonctionnelle (Marcotte and Marcotte, 2002). Si dans un organisme, des protéines interagissant ensemble sont codées par des gènes distincts, et si ces gènes sont fusionnés dans un autre organisme et codent pour une seule protéine, alors il est fort probable que ces gènes aient une fonction proche. Enfin, la méthode des profils phylogénétiques peut également mettre en évidence des liens fonctionnels entre gènes (Ranea *et al*., 2007). Cette méthode utilise les corrélations de variation absence/présence de gènes orthologues entre différents organismes. Elle est basée sur le fait que les protéines présentes et absentes chez les mêmes espèces sont souvent fonctionnellement liées : un organisme nécessitant une voie métabolique particulière possède les gènes nécessaires à la synthèse de cette voie, alors que les gènes d'une voie métabolique ne servant plus à un organisme peuvent être éliminés de son génome.

4.4.2.2 Annotation de différents signaux

Certains signaux de localisation et d'adressage présentent une structure protéique très conservée au sein des différents organismes et peuvent donc être annotés. Il existe de nombreuses méthodes et de logiciels permettant la reconnaissance de ces motifs. A titre d'exemples, on peut citer le logiciel TMHMM qui peut être utilisé pour prédire avec précision les hélices transmembranaires des protéines (Krogh *et al*., 2001), le logiciel SignalP qui permet de prédire les sites de coupures des peptides signaux de protéines potentiellement solubles (Bendtsen *et al*., 2004), predictNLS qui prédit les signaux de localisation nucléaire (SLN) (Cokol *et al*., 2000) ou encore PSORTb qui permet de prédire la localisation subcellulaire des protéines bactériennes (Gardy *et al*., 2005).

4.4.2.3 Annotation de domaines protéiques

Les domaines fonctionnels peuvent également être annotés. La recherche de domaines fonctionnels est plus précise que la recherche de définition fonctionnelle, car si des protéines ont une fonction identique, les gènes qui les codent peuvent avoir des séquences très différentes, mais certains motifs de séquence ou de structure nécessaires à leur fonction sont très bien conservés. Les banques PROSITE (Hulo *et al*., 2008) et Pfam-A (Finn *et al*., 2008) contiennent de nombreuses familles et domaines protéiques et ont l'avantage d'avoir été construites manuellement par des chercheurs. Ces deux banques sont intégrées à la banque InterPro (Hunter *et al*., 2009) qui regroupe également d'autres banques de motifs comme par

exemple TIGRFAMs. Le logiciel InterProScan peut être utilisé pour effectuer une recherche rapide dans un grand nombre de séquences afin d'identifier les domaines intégrées par InterPro (Mulder and Apweiler, 2007).

4.4.2.4 Annotation Enzyme Commission

La nomenclature Enzyme Commission (EC) (http://www.chem.qmul.ac.uk/iubmb/enzyme/) est une classification hiérarchique des enzymes basée sur le type de réaction chimique qu'elles catalysent. Chaque code d'enzyme est composé des lettres majuscules EC suivies de quatre nombres séparés par des points. Chaque nombre représente un niveau de précision : le premier indique le type de réaction catalysée (1 : oxydo-réduction ; 2 : transfert d'un groupement fonctionnel ; 3 : hydrolyse ; 4 : clivages de liaison autres que l'hydrolyse ou l'oxydation ; 5 : isomérisation ; 6 : établissement de liaisons covalents entre deux molécules), le deuxième indique le type général de substrat, le troisième correspond au substrat spécifique et le quatrième nombre correspond au numéro de série de l'enzyme et représente le niveau le plus précis de la classification. A titre d'exemples, certaines banques comme KEGG PATHWAY, MetaCyc ou encore Reactome utilisent les numéros d'EC pour annoter les enzymes de différentes voies métaboliques décrites dans la littérature (Caspi *et al*., 2008; Kanehisa *et al*., 2008; Matthews *et al*., 2009). L'attribution des numéros d'EC peut également être réalisée par transfert d'annotation (Tian and Skolnick, 2003), par comparaison de profils avec la banque PRIAM (Claudel-Renard *et al*., 2003) qui contient des domaines conservés des familles d'enzymes de la banque ENZYME (Bairoch, 2000) ou en utilisant le logiciel EFICAz^2 (Arakaki *et al*., 2009) qui utilise des modèles provenant des domaines Pfam et PROSITE.

4.4.2.5 Annotation Gene Ontology

La Gene Ontology (GO) regroupe un ensemble de termes qui sont reliés entre eux par des relations bien définies (Ashburner *et al*., 2000). Cette ontologie modélise les informations d'un domaine particulier sous la forme d'un graphique orienté acyclique (DAG : Directed Acyclic Graph). Cette annotation est composée de trois vocabulaires différents et permet de décrire les fonctions moléculaires, les processus biologiques et les localisations subcellulaires des gènes et de leurs produits. Elle a été créée afin de standardiser les données, car la plupart des annotations de définition fonctionnelle sont effectuées librement et une séquence peut alors être annotée de différentes manières. Cette ontologie est l'une des plus utilisée en biologie actuellement. Elle est intégrée dans de nombreuses banques de données biologiques. A titre d'exemple, elle est utilisée manuellement et automatiquement dans le cadre du consortium GOA (GO Annotation) afin d'identifier les membres de la banque UniprotKB (Barrell *et*

al., 2009; Boeckmann *et al*., 2003). L'annotation GO fournit un code d'évidence qui indique la source de l'annotation afin d'évaluer la pertinence de l'annotation. A ce jour, l'annotation GO peut être réalisée à partir de mots clefs de la banque UniprotKB, de termes EC (EC2GO), de domaines InterPro (InterPro2GO), de domaines Pfam (Pfam2GO), ou encore de familles HAMAP (HAMAP2GO). Cette annotation peut être réalisée de manière automatique à l'aide de logiciels comme Blast2GO (Conesa and Götz, 2008; Conesa *et al*., 2005; Götz *et al*., 2008) ou encore GOAnno (Chalmel *et al*., 2005). Blast2GO est une suite logicielle qui permet d'analyser et de réaliser une annotation fonctionnelle de gènes ou de protéines. Cet outil était dédié initialement à l'enrichissement de l'annotation GO, mais il contient à ce jour d'autres annotations comme l'identifiant ENZYME, la voie métabolique KEGG, et les motifs InterPro. Cette suite réalise une recherche par BLAST de la séquence d'intérêt sur des banques publiques avant de récupérer l'annotation GO des hits sélectionnés en fonction d'une valeur seuil (*E-value*) et d'une longueur d'alignement. Pour chaque terme GO récupéré, un score est calculé afin de déterminer la pertinence de l'annotation.

5 Objectifs de la thèse

Les métabolites secondaires végétaux constituent une classe extrêmement large de substances naturelles qui participent de manière déterminante à l'adaptation des plantes à leur environnement. Par ailleurs, certaines de ces molécules, comme les furanocoumarines, représentent également une source importante de substances naturelles valorisables dans les domaines pharmaceutiques, cosmétiques et agro-alimentaires. Les intermédiaires réactionnels de la voie de biosynthèse des furanocoumarines ont été mis en évidence dans les années 1960 par des approches biochimiques couplées à l'utilisation de cultures cellulaires et de traceurs radiomarqués. Plusieurs familles d'enzymes interviennent dans cette voie de biosynthèse comme les prényltransférases, les dioxygénases, les méthyltransférases et les cytochromes P450. Néanmoins, peu d'enzymes de cette voie ont été caractérisées au niveau moléculaire. Afin de mieux comprendre les mécanismes moléculaires régulant cette voie de biosynthèse, le laboratoire Agronomie et Environnement de l'UMR 1121 Université de Lorraine - INRA Nancy/Colmar a mis en place une thématique « Furanocoumarines » en 1998. Les premières études basées sur une approche agronomique ont permis de déterminer les plantes qui produisaient le plus de furanocoumarines et d'optimiser les conditions de culture pour la production de ces métabolites d'intérêt (Massot *et al*., 2000; Milesi *et al*., 2001). Par la suite, le laboratoire a acquis des techniques de clonage ce qui a permis d'orienter les recherches sur une composante moléculaire. La première enzyme caractérisée fonctionnellement fut la C4H de *R. graveolens* (Gravot *et al*., 2004). Puis, l'identification moléculaire et la caractérisation fonctionnelle de la psoralène synthase et de l'angélicine synthase furent réalisées par Romain Larbat (Larbat *et al*., 2007, 2009). La compréhension moléculaire de cette voie de

biosynthèse s'est encore améliorée récemment avec la caractérisation de la *p*-coumaroyl coA 2'-hydroxylase qui convertit le *p*-coumaroyl coA en umbelliférone et la caractérisation d'une prényltransférase capable de synthétiser de la DMS et de l'osthénol à partir de l'umbelliférone (Karamat *et al*., 2014; Vialart *et al*., 2012). Néanmoins, il reste encore de nombreuses étapes de cette voie à identifier d'un point de vue moléculaire et à caractériser de manière fonctionnelle.

L'objectif majeur de ma thèse est de réaliser l'identification moléculaire et la caractérisation fonctionnelle d'enzymes impliquées dans la voie de biosynthèse des furanocoumarines.

Le premier objectif de ma thèse est d'identifier des séquences nucléotidiques codant pour des enzymes potentiellement impliquées dans la voie de biosynthèse des furanocoumarines. Pour cela, des approches bibliographiques et bio-informatiques sont utilisées pour identifier *in silico* des séquences candidates à partir de banques d'ADNc de *P. sativa*, *R. graveolens* et *C. cinereum*. Ces résultats sont présentés dans la première partie du chapitre III « Résultats et Discussion ».

Dans un second temps, certaines de ces séquences candidates ont fait l'objet d'analyses génétiques et enzymatiques plus fines en vue de caractériser les fonctions biochimiques et physiologiques des protéines correspondantes. Les résultats de la caractérisation fonctionnelle des enzymes CYP71AZ3 et CYP71AZ4 dans le système hétérologue « levure » sont exposés dans la deuxième partie du chapitre III.

Enfin, dans la dernière partie du chapitre « Résultats et Discussion », je présente les résultats concernant l'étude d'une autre sous-famille de cytochromes P450, les CYP71AJ dont certains membres interviennent dans la biosynthèse des furanocoumarines.

PARTIE 2 :

Matériel et Méthodes

PARTIE 2 : Matériel et Méthodes

1 Matériel

1.1 Matériel végétal

1.1.1 *P. sativa*

Les plants de panais utilisés au cours de cette étude sont issus de la germination de graines commerciales de *P. sativa* L. subsp *sativa* variété 'Demi-Long de Guernesey'.

1.1.2 *R. graveolens*

Les plantes utilisées pour la construction de la banque d'ADNc normalisée sont issues de la germination de graines de *R. graveolens* L. provenant de la société Conrad Appel, Samen und Planzen GmbH, Bismarckstraβe 59, 64293 Darmstadt, Deutschland.

1.1.3 *C. cinereum*

Les plantes utilisées pour la construction de la banque d'ADNc normalisée sont issues de la germination de graines de *C. cinereum* Lindl. anciennement appelée *P. cinerea* provenant du jardin botanique de Perth (Australie).

1.1.4 *Nicotiana benthamiana*

Les plantes utilisées pour l'agro-infiltration transitoire proviennent de la germination de graines de *N. benthamiana* fourni par le Dr. Etienne Herrbach de l'équipe Virologie et Vection de l'INRA de Colmar.

1.2 Souches bactériennes

1.2.1 *Escherichia coli* Top 10® (Invitrogen)

Cette souche de bactéries est utilisée pour le clonage de fragments d'ADN dans le vecteur pCR8/GW®/TOPO. Elle permet également la réplication de manière stable des plasmides.

Génotype : [F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G]

1.2.2 *E. coli* GenHogs® (Invitrogen)

Cette souche de bactéries dérive de la souche *E. coli* DH10™. Elle permet d'amplifier efficacement les plasmides de grande taille. Elle est également utilisée comme souche bactérienne de clonage dans divers vecteurs plasmidiques (dans notre cas : pCR8/GW®/TOPO, p*Ye*DP60 et pBIN).
Génotype : [F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *deo*R *rec*A1 *end*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K 1 *rps*L *nup*G]

1.2.3 *E. coli* XL1 Blue (Stratagene)

Cette souche est utilisée pour le clonage d'inserts dans divers vecteurs plasmidiques et également pour l'amplification des constructions plasmidiques. Elle est déficiente pour l'activité endonucléase (*end*A) ce qui augmente la qualité de l'ADN lors de l'extraction plasmidique et pour l'activité de recombinaison (*rec*A) ce qui permet d'augmenter la stabilité de notre insert. Elle est également dépourvue de l'activité endonucléase ($hsdR^-$) afin de prévenir le clivage de l'insert par le système endonucléase *EcoK*.
Génotype : *end*A1 *gyr*A96 *hsd*R17($r_k^- m_k^+$) *lac rec*A1 *rel*A1 *sup*E44 *thi*-1 [F´ *pro*AB *lac*I^qZΔM15 Tn10 (Tet^R)]

1.2.4 *E. coli* MC1022

Cette souche est utilisée pour les clonages d'inserts dans des vecteurs plasmidiques et pour l'amplification de plasmides (Casadaban and Cohen, 1980). Elle peut permettre la sélection du type blanc/bleu des clones recombinants par la couleur en présence d'IPTG et de X-gal, un substrat chromogène de la *β*-galactosidase.
Génotype : [*ara*D139 Δ(*ara-leu*)7697 Δ(lacZ)M15 *gal*U *gal*K *rps*L]

1.2.5 *E. coli ccd*B Survival™ (Invitrogen)

Cette souche est résistante aux effets toxiques du gène *ccdB* et est utilisée pour l'amplification et le maintien de vecteurs plasmidiques contenant le gène *ccdB* (p*Ye*DP60_GW® vide et pBIN_GW® vide).
Génotype : [F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G *ton*A::P_{trc}-*ccd*A]

1.2.6 *Agrobacterium tumefaciens* LBA4404 (pAL4404)

La souche LBA4404 a été créée par le Dr. P.J.J. Hooykaas de l'université de Leiden, Pays-Bas et est largement utilisée comme vecteur de transformation génétique de nombreuses plantes comme par exemple *A. thaliana*, *N. benthamiana* ou *R. graveolens*. Cette souche contient un plasmide Ti désarmé, le plasmide pAL4404 qui ne contient que les gènes responsables de l'induction des gènes *vir* et du transfert de l'ADN-T. Cette souche est résistante à la rifampicine et la présence du plasmide pAL4404 lui confère également la résistance à la streptomycine.

1.3 Souches de levures

1.3.1 *S. cerevisiae* WAT11

La souche de levure WAT11 a été construite en 1996 par le Dr Denis Pompon (Laboratoire d'Ingénierie des Protéines Membranaires LIPM, Gif-sur-Yvette, France) pour l'expression spécifique de P450s de plantes. Elle dérive de la souche W303 et le gène endogène de la P450 réductase a été remplacé par recombinaison homologue par la P450 réductase *AthR1* d'*A. thaliana* (Urban *et al*., 1997). Ce gène *AthR1* est sous contrôle du promoteur *GAL10-CYC1* et est réprimé par le glucose et induit par le galactose. Cette souche est ADE2 et est donc incapable de pousser sur un milieu dépourvu d'adénine.

1.3.2 *S. cerevisiae* WAT21

La souche *S. cerevisiae* WAT21 diffère de la souche *S. cerevisiae* WAT11 par le remplacement de la cytochrome P450 réductase *AthR1* par l'*AthR2* d'*A. thaliana*.

1.4 Vecteurs et plasmides recombinants

1.4.1 pCR8®/GW/TOPO® (Invitrogen)

Le pCR8®/GW/TOPO® dans sa version kit de clonage est un vecteur d'entrée de type Gateway® utilisé pour cloner rapidement des produits de PCR avec une efficacité de plus de 95%. Une topoisomérase de type I (Figure 28, p. 90) issue du virus *Vaccinia* est fixée de manière covalente sur un site spécifique (CCCTT) permettant une ligation rapide du produit de PCR dans le vecteur linéarisé. Ce vecteur possède deux extrémités sortantes 3'-T empêchant sa recircularisation et permettant la ligation de produits de PCR dont les extrémités 3' ont été au préalable additionnées d'une adénine grâce à certaines *Taq* DNA

polymérases. De plus, ce vecteur possède deux sites *Eco*RI (Figure 29, p. 90) flanquant le site d'insertion de l'amplicon permettant une vérification rapide de la présence de l'insert par digestion enzymatique. Les séquences complémentaires aux amorces universelles M13 et T7 permettent de vérifier par PCR le sens d'insertion du gène d'intérêt dans le plasmide et de réaliser des séquençages des fragments insérés. Les sites *att*L1 et *att*L2 issus du bactériophage λ sont des séquences situées à moins de 55 paires de bases du site d'insertion qui permettent une recombinaison rapide grâce à l'enzyme LR clonase (Invitrogen) vers un vecteur de destination de type Gateway®. Le vecteur pCR8®/GW/TOPO® contient un gène de résistance à la spectinomycine et une origine de réplication du plasmide pUC pour l'obtention d'un grand nombre de copies du plasmide et le maintien dans *E. coli*.

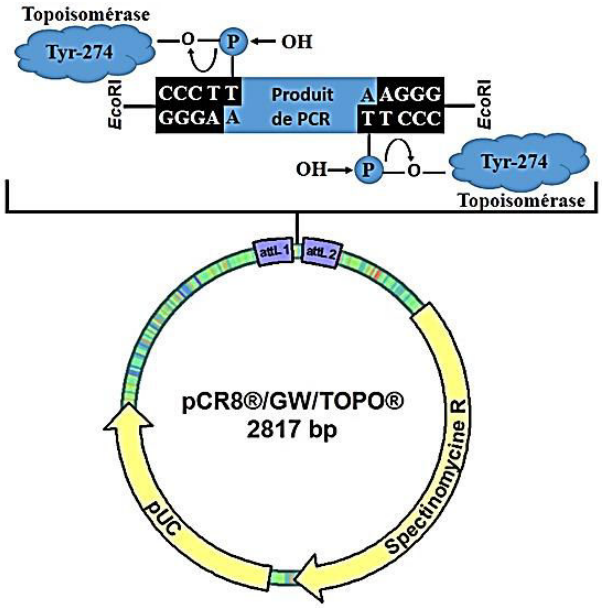


Figure 28 : Représentation schématique du vecteur pCR8®/GW/TOPO® (Invitrogen).

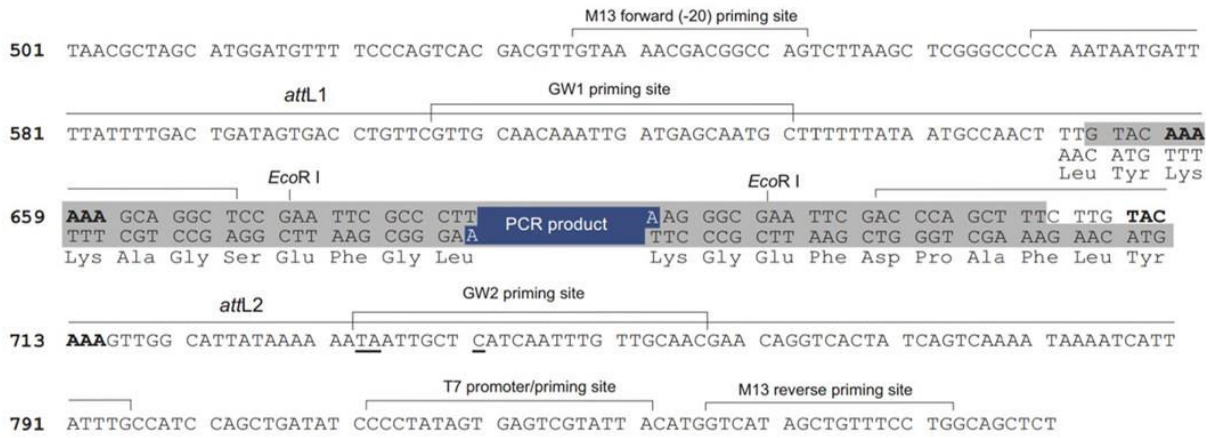


Figure 29 : Polylinker du vecteur pCR8®/GW/TOPO® (Invitrogen).

1.4.2 p*Ye*DP60

Le plasmide p*Ye*DP60 a été fourni par le Dr Denis Pompon du Laboratoire d'Ingénierie des Protéines Membranaires (LIPM) du CNRS à Gif-sur-Yvette, France. Ce plasmide est un vecteur d'expression réplicatif doté d'une origine de réplication *2µ* de levure (Figure 30, p. 91), d'un polylinker *Bam*HI/*Sma*I/*Kpn*I/*Sac*I/*Eco*RI encadré en amont par le promoteur *GAL10-CYC1* inductible par le galactose et réprimé par le glucose, et en aval par un terminateur *PGK*. Une origine de réplication *E. coli* (ColE1) ainsi qu'un gène de résistance à l'ampicilline permettent la sélection et l'amplification du plasmide. Ce plasmide présente également deux gènes *URA3* et *ADE2* qui assurent la complémentation des levures pour l'auxotrophie à l'uracile et à l'adénine. Seules les levures transformées pourront se développer dans un milieu dépourvu en adénine. Le p*Ye*DP60 est un plasmide très stable dans les milieux riches, à de fortes densités cellulaires. Ce plasmide a été spécialement conçu pour l'expression de P450s dans une série de souches *S. cerevisiae* (WAT11 et WAT21).

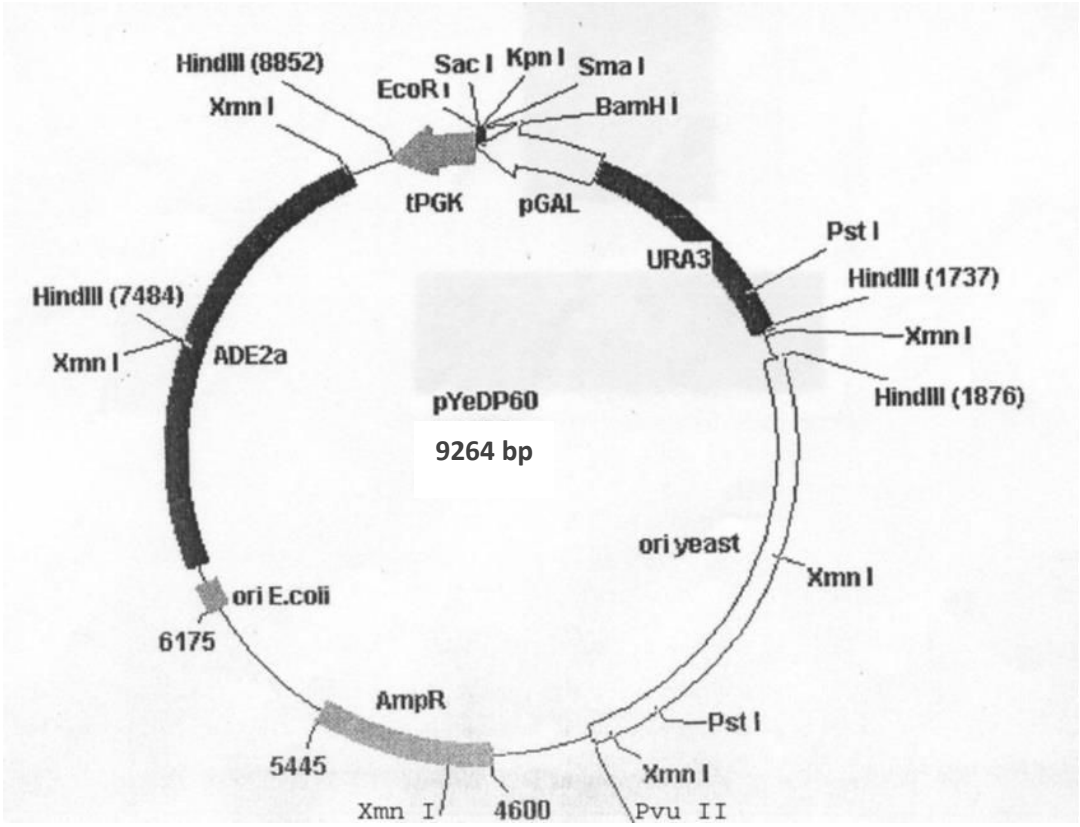


Figure 30 : Représentation schématique du vecteur d'expression dans la levure p*Ye*DP60 (communication personnelle du Dr. Denis Pompon).

1.4.3 p*Ye*DP60_GW®

Ce vecteur est un plasmide de destination Gateway® qui dérive du plasmide p*Ye*DP60. La cassette de recombinaison RfA clonée au niveau du site de restriction *Bam*HI (rendu franc par l'utilisation du fragment de Klenow de la DNA *Taq* Polymérase, puis déphosphorylé) du polylinker, est encadrée en amont par le promoteur *GAL10-CYC1* et en aval par le terminateur *PGK*. La cassette de recombinaison RfA flanquée par les sites *att*R1 et *att*R2 comporte le gène de résistance au chloramphénicol et le gène *ccdB*. L'amplification de ce plasmide est réalisée par la transformation de bactéries *E. coli* ccdB survival™. Le gène de résistance à l'ampicilline permet de sélectionner les bactéries recombinantes. Les différents plasmides recombinants issus de ce vecteur seront nommés p*Ye*DP60_GW®_X.

1.4.4 pBIN_GW®

Le plasmide pBIN_GW® a été construit par le Dr. Sébastien Doerper durant sa thèse au laboratoire Agronomie et Environnement. Ce vecteur dérive du plasmide pBIN *m-gfp5-ER* fourni par le Pr. Jim Haselhoff (Division of Cell Biology, MRC Laboratory of Molecular Biology, Addenbrookes Hospital, Hills Road, Cambridge, CB2 2QH, United Kingdom). Le pBIN_GW® est un plasmide de destination Gateway® à faible nombre de copies. Entre les bordures droite et gauche délimitant l'ADN-T de ce plasmide se situent le gène de sélection *npt II* et la cassette de recombinaison RfB flanquée par les sites *att*R1 et *att*R2 (Figure 31, p. 93). Le gène *npt II* qui confère la résistance à la kanamycine est sous le contrôle du promoteur et du terminateur de la nopaline synthase (Pnos et 3'nos). La cassette de recombinaison RfB qui comporte le gène de résistance au chloramphénicol et le gène *ccdB* est encadrée en amont par le promoteur 35S et en aval par le terminateur 3'nos. L'amplification de ce plasmide est réalisée par la transformation de bactéries *E. coli* ccdB survival™. Le gène de résistance à la kanamycine permet de sélectionner les bactéries recombinantes.

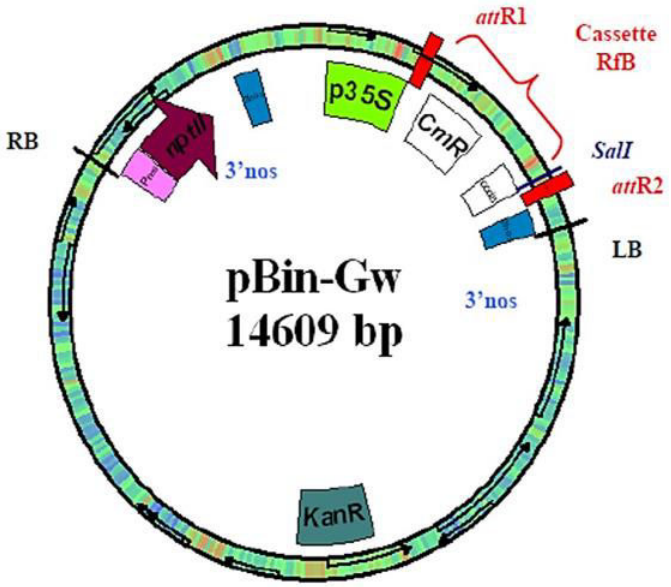


Figure 31 : Représentation schématique du vecteur pBIN_GW® (Biteau, 2009).

2 Milieux de culture

2.1 Milieux de culture pour bactéries

2.1.1 LB liquide et solide

Le milieu LB (Luria-Bertani) est utilisé pour la croissance de toutes nos souches de bactéries *E. coli*. Le milieu liquide est préparé en dissolvant 25 g de poudre de LB (Sigma Aldrich) dans un litre d'eau distillée. La solution est ensuite autoclavée à chaleur humide (20 min, 1 bar), puis stockée à température ambiante. La composition pour un litre de ce milieu est la suivante : 10 g de bactotryptone, 5 g d'extrait de levure, 10 g de NaCl.

Afin de préparer le milieu solide, de l'agar à 16 g/l est ajouté au milieu LB liquide avant autoclavage. Ce milieu est ensuite réparti dans des boîtes de Pétri avant de se solidifier à température ambiante. Les boîtes sont alors stockées à 4°C.

2.1.2 YEB

Ce milieu dont la composition est décrite dans le Tableau 1 (p. 94), est utilisé pour la croissance d'*A. tumefaciens*. La solution est ensuite autoclavée à chaleur humide (20 min, 1 bar), puis stockée à température ambiante.

Le milieu YEB solide est obtenu en additionnant au milieu YEB liquide 16 g/l d'agar (Sigma Aldrich). Ce milieu est ensuite réparti dans des boîtes de Pétri avant de se solidifier à température ambiante. Les boîtes sont alors stockées à 4°C.

Composition	**Bacto Beef Extract**	**Bacto Yeast Extract**	**Peptone**	**Saccharose**	**H_20**
Concentration (g/l)	5	1	5	5	qsp 1l

Tableau 1 : Composition du milieu YEB.

2.1.3 Antibiotiques utilisés

Afin de sélectionner une souche de bactérie recombinante particulière, un ou plusieurs antibiotiques sont additionnés au milieu de culture. Les solutions stock d'antibiotiques sont préparées à partir de poudre et sont dilués dans de l'eau, de l'éthanol ou du DMSO (Tableau 2, p. 94). Les antibiotiques sont utilisés à la concentration finale indiquée dans le Tableau 2 (p. 94) et sont toujours ajoutés au milieu après autoclavage. Dans le cas de milieux solides, ils sont ajoutés lorsque le milieu est en surfusion (environ 50°C) avant de le couler dans des boîtes de Pétri.

Antibiotiques	**Solvant**	**Concentration stock**	**Concentration finale**
Ampicilline	H_2O	100 mg/ml	100 µg/ml
Chloramphénicol	Ethanol	30 mg/ml	30 µg/ml
Kanamycine	H_2O	50 mg/ml	50 µg/ml
Rifampicine	DMSO	100 mg/ml	100 µg/ml
Spectinomycine	H_2O	100 mg/ml	100 µg/ml
Streptomycine	H_2O	100 mg/ml	100 µg/ml
Tétracycline	Ethanol	100 mg/ml	100 µg/ml

Tableau 2 : Concentrations des solutions stocks et des solutions finales d'antibiotiques et solvants utilisés.

2.2 Milieux de culture pour levures

2.2.1 YPGA

Le milieu YPGA (Tableau 3, p. 96) est un milieu riche, complémenté en adénine, utilisé pour la croissance des souches de *S. cerevisiae* WAT11 et WAT21 non transformées. Le milieu YPGA est utilisé sous forme liquide, mais également sous forme solide, coulée dans des boîtes de Pétri par l'ajout d'agar à 16 g/l avant autoclavage.

2.2.2 SGI

Le milieu SGI (Tableau 3, p. 96) est un milieu sélectif dépourvu en adénine qui permet de sélectionner les levures transformées. Le plasmide p*Ye*DP60 ou le p*Ye*DP60_GW® introduit dans la levure rend la souche prototrophe pour l'adénine et les transformants sont ainsi sélectionnés sur milieu SGI. Le milieu SGI est utilisé sous forme liquide, mais également sous forme solide, coulée dans des boîtes de Pétri par l'ajout d'agar à 16 g/l avant autoclavage.

2.2.3 Milieu YPGE

Le milieu YPGE (Tableau 3, p. 96) permet la croissance des levures transformées tout en inhibant la synthèse de la protéine d'intérêt et de la P450 réductase d'*A. thaliana*. L'absence de pression de sélection de ce milieu permet aux levures de s'approvisionner en source carbonée *via* le glucose.

2.2.4 Milieu YPL

Le milieu YPL (Tableau 3, p. 96) est un milieu inducteur de la synthèse de la protéine d'intérêt et de la CPR. Le galactose active le promoteur *GAL10-CYC1* qui induit simultanément la synthèse de ces deux protéines.

	YPGA	SGI	YPGE	YPL
Yeast Extract (g/l)	10		10	10
Bactopeptone (g/l)	10		10	10
Bactocasaminoacids (g/l)		1		
Yeast Nitrogen Base sans acides aminés (g/l)		7		
Tryptophane (mg/l)		20		
Adénine (solution à 20 mg/ml)	10 ml			
Glucose (g/l)	20	20	5	
Ethanol après autoclavage (%, v/v)			3	
Galactose (g/l)				20

Tableau 3 : Composition des différents milieux de culture pour la levure.

2.3 Culture de plantes en terre

Les plantes sont cultivées 1) sous serre au laboratoire Agronomie et Environnement (latitude 48.65, longitude 6.14) d'avril à octobre et 2) en conditions contrôlées en phytotron (température de 20°C, photopériode de 16 h) de novembre à mars avec un arrosage tri hebdomadaire.
Les graines sont mises à germer dans un même pot, puis les plantes âgées de deux à trois semaines sont transplantées dans des pots individuels. Le semis ainsi que le rempotage des différentes plantes sont effectués dans un terreau correspondant à la norme NF U 44-551 contenant 60 % d'un mélange de tourbes blonde et noire, des écorces de bois et des fibres de pin compostées pour alléger le terreau, de la fibre de coco pour une bonne porosité du milieu, de la vermiculite et de la perlite pour favoriser le drainage.

2.4 Culture de panais en hydroponie

Les jeunes plantules âgées d'une à deux semaines sont mises dans des godets d'un centimètre de diamètre (Photographie 1-A, p. 97) et les godets sont ensuite regroupés par 24 dans un bac métallique, étanche et opaque d'une contenance de 4L contenant la solution nutritive (cf. ci-dessous) (Photographie 1-B, p. 97). Les plantes alors âgées d'un mois sont placées dans des contenants en plastique d'une capacité d'un litre à raison de quatre godets par bac plastique (Photographie 1-C, p. 97).

Les plantes ont été cultivées en phytotron en conditions contrôlées (température de 20°C avec une photopériode de 16 h) sous lampe T5 Superplant 96W (4 x 24W) (Cisproduct). Afin d'éviter l'hypoxie des racines, une pompe à air délivre une aération continue dans la solution de culture.

Les plantes sont cultivées dans une solution nutritive composée de 3 solutions : la solution FloraGro™ (7 ml pour 10 L), de solution FloraMicro™ (7 ml pour 10 L) et de solution FloraBloom™ (7 ml pour 10 L) de la marque General Hydroponics. La solution FloraGro™ fournit l'azote et le potassium à la plante, la solution FloraBloom™ contient de la silice pour combattre les moisissures et le pourrissement des racines et la solution FloraMicro™ apporte les oligoéléments et les microéléments nécessaires à la plante sous forme de chélates. Le pH de la solution nutritive est ajusté entre 6 et 6,5 et la solution est renouvelée de manière hebdomadaire.

Photographie 1 : Dispositif de culture hydroponique pour *Pastinaca sativa*.

2.4.1 Induction aux UV

Les plantes de *P. sativa* âgées de 2 à 3 mois sont élicitées dans une chambre à l'aide d'une lampe à UVB (λ=312 nm) placée à 70 cm des parties aériennes pendant 24 h à 48 h. Les plantes, placées par quatre dans des pots en plastique laissant passer les UVB, sont exposées aux UV soit uniquement au niveau des racines soit uniquement au niveau des parties aériennes (Photographie 2, p. 98).

Photographie 2 : Dispositif d'induction des plants de *P. sativa* aux UVB.

2.4.2 Induction au méthyljasmonate

Les plantes de *P. sativa* âgées de 2 à 3 mois sont élicitées dans une chambre séparée des autres plantes par une solution à 200 µM de méthyljasmonate (MeJa). Une solution mère 1000 fois concentrée de méthyljasmonate est préparée dans de l'éthanol.

Dans le cas d'une élicitation des parties aériennes, la solution de méthyljasmonate est directement administrée sur les parties aériennes par pulvérisation. Dans le cas d'une élicitation des racines, celles-ci sont directement plongées dans la solution nutritive supplémentée de méthyljasmonate. Les plantes contrôles sont cultivées dans une solution nutritive additionnée de la même quantité d'éthanol, mais dépourvue de méthyljasmonate.

3 Méthodes de bioinformatique

3.1 Utilisation de la suite logicielle blast2GO

La suite logicielle Blast2GO (http://www.blast2go.com/b2ghome) est utilisée pour réaliser l'annotation fonctionnelle des séquences contenues dans les banques normalisées d'ADNc de *P. sativa*, de *R. graveolens* et de *C. cinereum*. Afin de réaliser ces annotations, une recherche par comparaison de séquences (blastx ayant pour *E-value* minimum 1^{E-03}) est réalisée sur la base de données NCBI nr. La recherche de séquences codantes pour des cytochromes P450 et des cytochromes P450 réductases est

réalisée à l'aide des annotations de Gene Ontology ainsi qu'avec les termes InterPro (IPR) listés dans le Tableau 4 (p. 99).

Les termes GO obtenus se subdivisent en trois catégories : fonction moléculaire, processus biologique et compartiment cellulaire. Les classifications utilisées pour l'obtention des diagrammes circulaires sont celles des niveaux 2 (sur 9 niveaux) de la classification.

Termes IPR	Description	Séquence ciblée
IPR001094	Flavodoxin	CPR
IPR001128	Cytochrome P450	P450
IPR002397	Cytochrome P450, B-class	P450
IPR002399	Cytochrome P450, mitochondrial	P450
IPR002401	Cytochrome P450, E-class group I	P450
IPR002402	Cytochrome P450, E-class group II	P450
IPR002403	Cytochrome P450, E-class group IV	P450
IPR002949	Cytochrome P450, E-class CYP24A mitochondrial	P450
IPR002974	Cytochrome P450, E-class CYP52	P450
IPR008066	Cytochrome P450 group I, CYP1	P450
IPR008067	Cytochrome P450 group I, CYP2A-like	P450
IPR008068	Cytochrome P450 group I, CYP2B-like	P450
IPR008069	Cytochrome P450 group I, CYP2D-like	P450
IPR008070	Cytochrome P450, E-class group I CYP2E-like	P450
IPR008071	Cytochrome P450 group I, CYP2J-like	P450
IPR008072	Cytochrome P450, E-class CYP3A	P450
IPR023173	NADPH-cytochrome P450 reductase, FAD-binding, alpha-helical domain-3	CPR
IPR023206	Bifunctional cytochrome P450/NADPH-Cytochrome P450 reductase	P450/CPR
IPR023208	NADPH-cytochrome P450 reductase	CPR
IPR024204	Cytochrome P450, cholesterol 7-alpha-monooxygenase-type	P450

Tableau 4 : Termes IPR utilisés pour identifier les séquences correspondantes à des cytochromes P450 et des cytochromes P450 réductases (CPRs) contenues dans les banques d'ADNc de *P. sativa*, *R. graveolens* et *C. cinereum.*

3.2 Utilisation du logiciel BioEdit

Les alignements de séquences complètes (ClustalW), l'analyse de résultats de séquençage, la recherche par homologie de séquences (*blastn*, *tblastx*, *blastx*, *tblastn* et *blastp* à l'aide de l'algorythme BLOSUM62) ainsi que l'obtention de cartes de restriction de plasmides sont effectués avec le logiciel BioEdit.

3.3 Génération d'arbres phylogénétiques avec le logiciel MEGA6

Les arbres phylogénétiques construits dans cette étude sont réalisés à l'aide du logiciel MEGA6 en utilisant l'algorythme Maximum Likelihood après avoir généré un alignement de séquences avec ClustalX (algorythmes Gonnet pour les séquences complètes ou PAM dans le cas de séquences partielles).

4 Méthodes de biologie moléculaire

4.1 Extraction d'ADN génomique de plante

L'extraction d'ADN génomique de *P. sativa* est réalisée après un broyage à l'azote liquide à l'aide d'un mortier et d'un pilon en céramique de 100 mg de feuilles fraîches. La poudre ainsi obtenue est utilisée pour extraire l'ADN génomique à l'aide du kit « DNeasy® Plant Mini » de Qiagen selon les recommandations du fournisseur. L'élution de l'ADN est réalisée dans un volume de deux fois 50 µl de tampon AE. L'ADN génomique extrait est analysé sur gel et dosé à l'aide du BioPhotometer (Eppendorf) avant d'être conservé à -20 °C.

4.2 Extraction d'ARN totaux de plante

L'extraction d'ARN totaux est une opération délicate à cause de l'importante présence d'ARNases dans l'environnement. Afin de palier à cette contrainte, une solution de neutralisation et de décontamination (RNaseZap®, Ambion) est utilisée pour nettoyer la paillasse, les pipetmans ainsi que les spatules.
L'extraction d'ARN totaux est réalisée après un broyage à l'azote liquide à l'aide d'un mortier et d'un pilon en céramique de 100 mg de feuilles, de tiges et de graines de *P. sativa*, *R. graveolens* et *C. cinereum* ou de 50 mg de racines de *P. sativa*. La poudre ainsi obtenue est utilisée pour extraire les ARN totaux à l'aide du kit « RNeasy® Plant Mini » de Qiagen selon les recommandations du fournisseur. Une digestion

de l'ADN génomique par la DNase est réalisée en cours d'extraction selon les recommandations du fournisseur. L'élution des ARN totaux est réalisée dans deux fois 40 µl de tampon d'élution RNase free. Les ARN ainsi extraits sont immédiatement dosés à l'aide du BioPhotometer (Eppendorf), analysés sur gel (l'intégrité des ARN est vérifiée par la visualisation de bandes majoritaires correspondantes aux ARN ribosomiques cytoplasmiques et chloroplastiques) et conservés à -20 °C.

4.3 Synthèse des ADNc simple brin par transcription inverse (RT)

La synthèse des ADNc complémentaires aux ARN totaux extraits précédemment est réalisée à l'aide du kit « High Capacity RNA-to-cDNA » d'Applied Biosystems selon les recommandations du fournisseur. Les réactions de transcription inverse sont effectuées dans un volume final de 20 µl et contiennent 10 µl de tampon RT 2X, 1 µl d'ADN polymérase ARN dépendante, jusqu'à 9 µl de solution contenant les ARN (quantité ≤ 2 µg) et de l'eau ultrapure stérile pour ajuster le volume à 20 µl. La réaction est réalisée à 37 °C pendant une heure, suivie d'une étape à 95 °C pendant cinq minutes afin d'inactiver la réverse transcriptase. Les ADNc peuvent être utilisés directement ou sont conservés à -20 °C.

4.4 Amplification d'un fragment d'ADN par PCR

4.4.1 Amplification classique

Toutes les amplifications PCR sont conduites dans un thermocycler Gradient Cycler PTC-200 (MJ Research) ou C1000 Touch^{TM} Thermal Cycler (Biorad).

La réaction en chaîne par polymérase ou PCR est une méthode d'amplification d'ADN *in vitro* qui débute par une étape de dénaturation de l'ADN (94 °C - 98 °C pendant 15 secondes - 2 minutes selon les kits utilisés), suivie de 25 à 35 cycles comprenant une étape de dénaturation (94 °C – 98 °C pendant 15 - 30 sec selon les kits), une étape d'hybridation des amorces oligonucléotidiques à l'ADN matriciel (54 °C – 62 °C pendant 30 sec) et une étape d'élongation (68°C ou 72°C selon les kits et en comptant 1 minute pour la synthèse de 1000 pb).
La température d'hybridation détermine le degré de spécificité de l'hybridation et dépend de la température de fusion (Tm) de chaque dimère amorce/matrice. La température de fusion correspond à la température à laquelle la moitié de l'ADN est sous forme monobrin et l'autre moitié sous forme double brin. Elle est déterminée en fonction de la longueur des amorces et de leur teneur en bases puriques. Afin d'assurer une hybridation optimale des amorces sens et antisens, les températures de

fusion de celles-ci doivent être proches. La température d'hybridation correspond généralement à Tm – 5 °C.

Pour une amorce de 14 à 20 paires de bases, le Tm est défini par l'équation de Wallace-Ikatura:

$$Tm = 2 \times (A + T) + 4 \times (G + C)$$

Pour une amorce supérieure à 20 paires de base, le Tm est défini par l'équation :

$$Tm = 81,5 + \log10\ [0,05] + 0,41 \times (\%\ G + \%\ C) - (675/N)$$

où N est la longueur de l'amorce en paire de base

Le kit PCR Master Mix (2X) de Fermentas est utilisé pour déterminer la présence et/ou l'insertion d'un fragment d'ADN dans un plasmide. Il est également utilisable en vue de réaliser un clonage TA et est donc utilisable en vue d'effectuer une ligation dans le pCR8®/GW/TOPO®. Il est utilisé selon les recommandations du fournisseur. Les réactions sont réalisées dans un volume de 20 µl contenant : 0,4 µM d'amorces sens et antisens, 10 µl de Master Mix (2X) (contenant le tampon de réaction, 0,4 mM de dNTPs, 4mM de $MgCl_2$ et 0,05 U/µl de *Taq* DNA polymérase), 1 à 2 µl d'ADN (10-100 ng de matrice) et de l'eau ultrapure stérile. La dénaturation initiale se fait à 94 °C pendant 2 min, suivie de 35 cycles (94 °C pendant 30 sec, 54°C - 62°C pendant 30 sec, 72°C pendant 2 min), l'extension finale se fait à 72°C pendant 10 min.

Le kit Platinium® *Taq* DNA Polymerase High Fidelity d'Invitrogen est utilisé afin d'amplifier des fragments d'ADN avec une haute fidélité, elle permet également l'ajout d'une adénine en 3' à chaque extrémité du produit de PCR, ce qui permet un clonage efficace dans le vecteur pCR8®/GW/TOPO®. Ce kit contient une polymérase issue de *Pyrococcus species* GB-D, une *Taq* DNA polymérase et un anticorps anti-*Taq* DNA Polymerase qui inactive l'activité polymérase à température ambiante. La polymérase GB-D de *Pyrococcus species* présente la capacité de corriger les erreurs grâce à son activité 3'→5'exonucléasique.
Ce kit est utilisé selon les recommandations du fournisseur. Les réactions sont réalisées dans un volume de 50 µl contenant : 0,2 µM d'amorces sens et antisens dans le cas d'une matrice ADN génomique ou ADNc ou 0,4 µM d'amorces sens et antisens dans le cas d'une matrice plasmidique, 5 µl de 10X High Fidelity PCR Buffer, 0,2 mM de dNTPs, 2mM de $MgSO_4$, 1U de Platinium® *Taq* DNA polymérase, 1 à 4 µl d'ADN (10-100 ng de matrice) et de l'eau ultrapure stérile. La dénaturation initiale ainsi que l'activation de la Platinium® *Taq* DNA polymérase se fait à 94 °C pendant 2 min, suivie de 30 cycles (94 °C pendant 30 sec, 55 °C pendant 30 sec, 68 °C pendant 2 min), l'extension finale se fait à 68 °C pendant 10 min.

4.4.2 TAIL-PCR

La TAIL-PCR (Thermal Asymmetric InterLaced PCR) est une méthode efficace pour l'identification de fragments d'ADN adjacents à des séquences nucléotidiques connues (Liu *et al*., 2004; Meng *et al*., 2005). La TAIL-PCR utilise trois amorces nichées spécifiques (SP1, SP2, SP3) du fragment d'ADN connu dans des réactions successives en association avec une amorce arbitraire présentant un certain degré de dégénérescence (amorce dégénérée AD) (Figure 32, p. 103). L'amorce dégénérée devra avoir une Tm basse et les amorces spécifiques devront avoir une Tm élevée pour que les fréquences d'amplification des produits spécifiques et non spécifiques puissent être contrôlées thermiquement. Le but est d'obtenir un amplicon final spécifique d'une séquence d'ADN dont seule une extrémité est connue en vue de séquencer la partie inconnue.

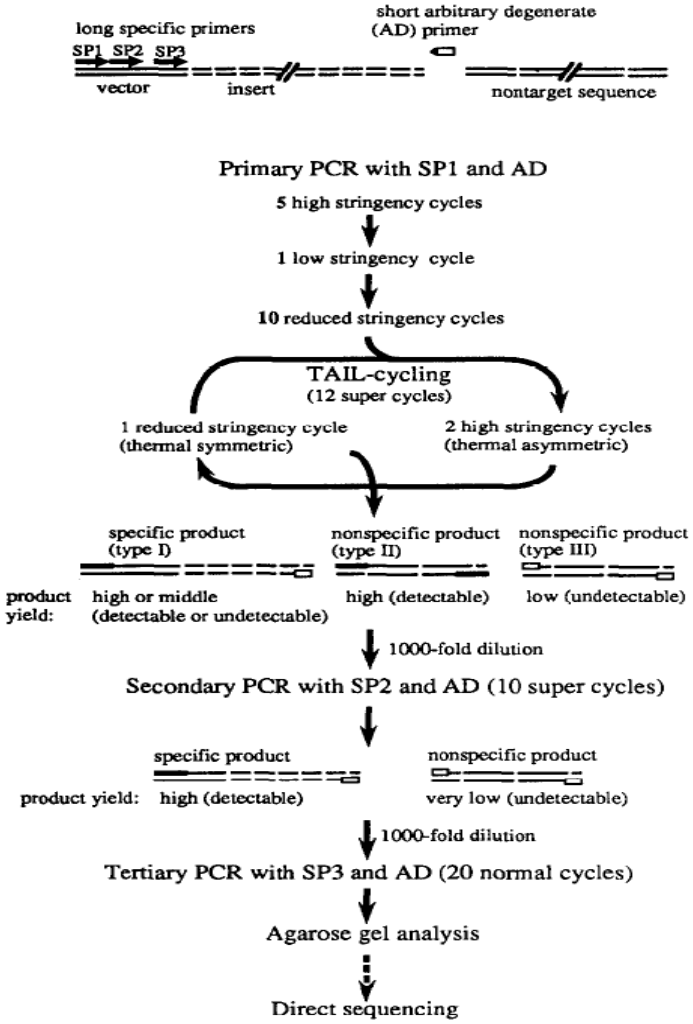


Figure 32 : Schéma illustrant le principe de la TAIL-PCR (Liu and Whittier, 1995).

Dans la première réaction de PCR, un cycle de faible spécificité réalisé à 25°C permet d'hybrider l'amorce dégénérée sur un ou plusieurs sites de la séquence recherchée (Tableau 5, p. 104, PCR 1 étape 3). Les fragments spécifiques de type I sont ensuite sélectionnés par rapport aux fragments aspécifiques de type II et III (Figure 32, p. 103) par des cycles de PCR présentant de hautes températures d'hybridation (Tableau 5, p. 104, PCR 1 étape 4), car seule l'amorce spécifique SP1 peut s'hybrider correctement à ces températures. La deuxième réaction de PCR est réalisée avec de hautes températures d'hybridation en utilisant les amorces SP2 et AD. Cette réaction permet de sélectionner les fragments spécifiques de type I (Tableau 5, p. 104, PCR 2). La dernière réaction de PCR est réalisée à basse température d'hybridation avec les amorces SP3 et AD favorise l'hybridation de l'amorce dégénérée et continue la sélection des fragments spécifiques de type I (Tableau 5, p. 104, PCR 3).

PCR n° / amorces	**Etape n°**	**Nombre de cycles**	**Programme**
1 Amorce SP1 et une des AD	1	1	94 °C, 5 min
	2	5	94 °C, 30 sec ; 58 °C, 1 min ; 72 °C, 1 min
	3	1	94 °C, 30 sec ; 25 °C, 1 min ; incrémentation de 25 °C à 72°C en 2 min ; 72 °C, 2 min
	4	15	94 °C, 30 sec ; 58 °C, 1 min ; 72 °C, 2 min ; 94 °C, 30 sec ; 58 °C, 1 min ; 72 °C, 2 min ; 94 °C, 30 sec ; 45 °C, 1 min ; 72 °C, 2 min
	5	1	72 °C, 1 min
2 Amorce SP2 et une des AD	1	1	94 °C, 5 min
	2	12	94 °C, 30 sec ; 58 °C, 1 min ; 72 °C, 2 min ; 94 °C, 30 sec ; 58 °C, 1 min ; 72 °C, 2 min ; 94 °C, 30 sec ; 45 °C, 1 min ; 72 °C, 2 min
	3	1	72 °C, 5 min
3 Amorce SP3 et une des AD	1	1	94 °C, 5 min
	2	20	94 °C, 1 min ; 45 °C, 1 min ; 72 °C, 2 min
	3	1	72 °C, 5 min

Tableau 5 : Programmes utilisés pour les réactions de TAIL-PCR.

En se basant sur les séquences nucléotidiques des contigs c5642, c29131, c14511 et c64662 de la banque normalisée d'ADNc de *P. sativa*, trois amorces antisens successives spécifiques du début de la séquence de chaque contig sont synthétisées. Ces amorces sont décrites dans le Tableau 6 (p. 105). Plusieurs amorces présentant un degré de dégénérescence variable sont également synthétisées (Tableau 6, p. 105).

Contig	Nom de l'amorce	Séquences des amorces 5'→3'	DD	Tm (°C)
Amorces spécifiques				
c5642	c5642SP1	GGGTCCCTTCCGATTCCCCAAGCG		59
	c5642SP2	CGTAGCCCATAACTTTAACATCCTGTCTTGCAGTTC		59
	c5642SP3	GGGACTAATAACGGAAGTGGTGGATGTAACCGT		59
c29131	c29131SP1	TAGGCATTAGTAGTGGAGCTGTAAAGTACAATCGC		58
	c29131SP2	CTGATCTCATTCTGCAATTTAGTCATGACACTCGG		58
	c29131SP3	GGGCTGCTATTGTCCAGTTTAAAGTTGTAGAGGATG		59
c14511	c14511SP1	GTTGCAGGGTCTCTTCCTATTGCCCAGG		59
	c14511SP2	CGGAATTGGTGGATGCAGACGGAGAGTTTC		59
	c14511SP3	GTCTCAAGAGTTCCGACATTGCCCATTCTAGG		59
c64662	c64662SP1	CAAGAGATCCTTCTCTGTGGGAAAACCC		56
	c64662SP2	GCCTGGGCAATTGACAAGAGATCCTTC		56
	c64662SP3	GCTAGGACACAAGTACTTATAAATGCCTGG		55
Amorces dégénérées				
	AD1	NGTCGASWGANAWGAA	128	46
	AD2	WGTGNAGWANCANAGA	256	48
	AD3	NTCGASTWTSGWGTT	64	46
	AD4	TGWGNAGSANCASAGA	128	48
	AD5	AGWGNAGWANCAWAGG	128	46
	AD6	WGCNAGTNAGWANAAG	256	48
	AD7	AWGCANGNCWGANATA	256	48
	AD8	TGWGNAGWANCASAGA	128	46

Tableau 6 : Amorces utilisées dans les réactions de TAIL-PCR. DD : Degré de dégénérescence.

Les différentes réactions de TAIL-PCR sont réalisées à l'aide du Master Mix (2X) de Fermentas dans un volume réactionnel de 50 µl de la manière suivante :

- PCR n°1 : 25 µl de Master Mix (2X), 2,5 µM d'AD présentant un degré de dégénérescence de 64 et 128 (ou 5 µM d'AD présentant un degré de dégénérescence de 256), 0,5 µM de l'amorce spécifique SP1, 1 µl d'ADN génomique de panais (1-10 ng) et de l'eau ultrapure stérile.
- PCR n°2 : 25 µl de Master Mix (2X), 2,5 µM d'AD présentant un degré de dégénérescence de 64 et 128 (ou 5 µM d'AD présentant un degré de dégénérescence de 256), 0,5 µM de l'amorce spécifique SP1, 5 µl du produit de réaction de la PCR n°1 diluée 40 fois et de l'eau ultrapure stérile.
- PCR n°3 : 25 µl de Master Mix (2X), 2,5 µM d'AD présentant un degré de dégénérescence de 64 et 128 (ou 5 µM d'AD présentant un degré de dégénérescence de 256), 0,5 µM de l'amorce spécifique SP1, 19 µl du produit de réaction de la PCR n°2 diluée 40 fois et de l'eau ultrapure stérile.

Les concentrations des amorces dégénérées varient en fonction de leur degré de dégénérescence. Plus une amorce est dégénérée, plus elle aura de possibilité de s'hybrider à l'ADN génomique, par conséquent plus une amorce est dégénérée plus sa concentration dans le milieu réactionnel sera élevée. Les produits réactionnels de la PCR n°3 sont analysés sur gel d'agarose à 1%. Les fragments d'intérêts sont prélevés à partir du gel, élués et clonés dans le plasmide pCR8®/GW/TOPO® et séquencés par la société Eurofins MWG Operon.

4.5 Electrophorèse d'ADN sur gel d'agarose

Les fragments d'ADN et d'ARN sont séparés en fonction de leur poids moléculaire de leur charge électrique par électrophorèse sur gel d'agarose. Plus le gel est concentré en agarose et plus la taille des mailles est fine. Généralement, les gels sont préparés à 1% d'agarose dans du tampon Tris 40 mM acétate, EDTA 1 mM (TAE) ce qui permet de séparer un large éventail de fragments, mais le pourcentage d'agarose peut être augmenté à 2% pour la séparation de très petits fragments. Le gel est déposé dans une cuve de migration contenant du tampon TAE. Les échantillons à analyser son supplémentés de 4 µl de tampon de charge (Tris 50 mM pH 7,5 ; EDTA 100 mM pH 7,5 ; 10% glycérol ; 0,05% bleu de bromophénol). Le gel est ensuite soumis à un champ électrique de 120 V (12 V/cm) pendant 25 min. Un marqueur de taille (MassRuler™ DNA Ladder Mix, ready-to-use #SM0403 de Fermentas utilisé pour les fragments allant de 500 pb à 10 000 pb ou GeneRuler™ 50 pb DNA Ladder #SM0371 de Fermentas utilisé pour les fragments de 50 à 1 000 pb) migre en même temps que les

échantillons et permet d'identifier la taille approximative des différentes bandes obtenues. La révélation est effectuée avec du bromure d'éthidium (BEt) qui est un agent intercalent employé comme marqueur d'acides nucléiques en biologie moléculaire. Le BEt permet la visualisation d'ADN et d'ARN en présence d'un rayonnement ultra-violet ($\lambda = 254$ nm). A la fin de la migration, le gel est baigné pendant 10 minutes dans un bain de BEt à 1 mg/l puis placé sous rayonnement UV.

4.6 Extraction d'ADN à partir d'un gel d'agarose

Après migration et révélation sous rayonnement ultra-violet, les bandes de gel correspondant aux fragments d'ADN d'intérêt sont prélevés à l'aide d'une lame de scalpel propre et placés dans un tube de 2 ml. L'extraction de l'ADN est réalisée à l'aide du kit « MinElute® Gel Extraction » de Qiagen selon les recommandations du fournisseur.

4.7 Digestion par des enzymes de restriction

La technologie FastDigest® de Fermentas permet de réduire les temps d'incubation d'un facteur 10 par rapport aux enzymes de restriction standards. De par son tampon d'activité FastDigest® universel, ce système permet de réaliser aisément des digestions par plusieurs enzymes simultanément. La réaction de digestion est en général réalisée dans un volume réactionnel de 10 µl contenant 1 µl de tampon d'activité (10X), 1 µl d'enzyme de restriction et entre 0,1et 1 µg d'ADN à digérer. La réaction est incubée à 37 °C pendant 15 min. Il est possible d'inactiver l'enzyme en incubant le mélange réactionnel 5 min à 80 °C.

4.8 Ligations

4.8.1 Dans le vecteur pCR8®/GW/TOPO®

Les ligations dans le plasmide pCR8®/GW/TOPO® se font rapidement et efficacement à l'aide du kit « pCR8®/GW/TOPO® TA Cloning® » d'Invitrogen. Les réactions de ligations sont réalisées dans un volume final de 3 µl contenant 10 à 20 ng de vecteur pCR8®/GW/TOPO® linéarisé, 0,6 µM de NaCl et 0,03 µM de $MgCl_2$ et 2 µl de produits de PCR (100 à 200 ng). Le mélange réactionnel est incubé pendant 3 h à température ambiante afin d'assurer une efficacité de ligation de 95 %.

4.8.2 Dans les autres vecteurs

Afin d'effectuer une ligation dans des vecteurs standards au niveau de sites de restriction, il est nécessaire au préalable d'amplifier le fragment d'ADN d'intérêt par PCR en utilisant des amorces permettant l'insertion d'un site de restriction excédentaire en 5' et en 3'. Dans un deuxième temps, le plasmide sera digéré par ces mêmes enzymes de restriction.
La ligation dans les vecteurs standards s'effectue à l'aide de l'enzyme T4 DNA Polymérase (Invitrogen) qui permet de liguer des fragments d'intérêt ayant des extrémités cohésives ou franches dans des vecteurs d'expression. La réaction s'effectue en général dans un volume réactionnel de 20 µl contenant 4 µl de tampon d'activité (5X), 0,1 U de T4 DNA ligase, 3 à 30 fmol de vecteur et 9 à 90 fmol d'insert en conservant un ratio insert/vecteur de 3:1. Le milieu réactionnel est mis à incuber pendant 2 h à température ambiante ou pendant la nuit à 16 °C avant d'être utilisé pour la transformation de bactéries *E. coli*.

4.9 Construction d'un plasmide recombinant pour la transformation génétique de levures en utilisant la technologie Gateway®

La technologie Gateway® est une technique de clonage propriétaire basée sur les propriétés de recombinaison site spécifique du bactériophage λ. Cette méthode de clonage est rapide et efficace et permet de transférer notre gène d'intérêt vers de multiples vecteurs de destination.
Afin de réaliser des clonages dans le p*Ye*DP60 plus rapidement que par digestion par des enzymes de restriction, le système de conversion de vecteur Gateway® est utilisé afin de transformer le p*Ye*DP60 en vecteur de destination Gateway®. La cassette de clonage RfA (898 pb) flanquée des sites *att*R contenant le gène *ccd*B et le gène de résistance au chloramphénicol (Figure 33, p. 109) est insérée au niveau du site *Bam*HI de notre plasmide afin de conserver le cadre de lecture initial.

Le plasmide p*Ye*DP60 (5 µg) est digéré par *Bam*HI pendant 1 h à 37 °C. Les extrémités sortante du vecteur sont converties en extrémités franches en utilisant la T4 DNA polymérase ou fragments de Klenow selon les recommandations du fournisseur (Invitrogen). Les extrémités 5' phosphates sont ensuite déphosphorylées grâce à la phosphatase alcaline de veau (CIAP) selon les recommandations du fournisseur (Invitrogen). La réaction de ligation s'effectue dans un volume réactionnel de 10 µl contenant 25 à 50 ng de vecteur linéarisé déphosphorylé, 2 µl de tampon de ligation (5X), 2 µl de cassette RfA à 5 ng/µl et 1 U de ligase selon les recommandations du fournisseur (Invitrogen). La réaction se déroule durant la nuit à 16 °C et le lendemain 1 µl du mélange réactionnel est utilisé pour transformer des bactéries *E. coli ccd*B survival™.

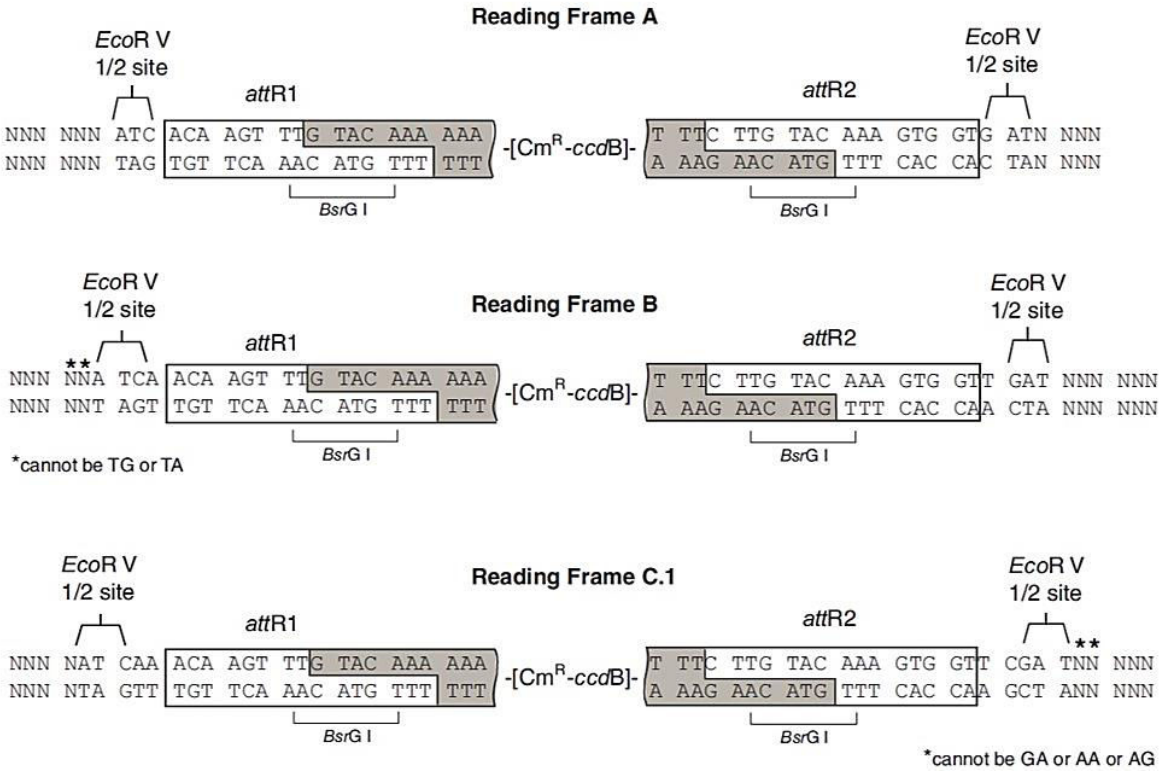


Figure 33 : Représentation schématique des trois cassettes Gateway®.

4.10 Obtention des plasmides recombinants p*Ye*PD60_GW®_X et pBIN_GW®_X *via* la technologie Gateway®

La réaction de LR recombinaison s'effectue entre les sites *att*L d'un plasmide d'entrée de type Gateway® (pCR8®/GW/TOPO®) et les sites *att*R d'un vecteur de destination de type Gateway® p*Ye*DP60_GW® ou pBIN_GW® (Figure 34, p. 110). Des bactéries *E. coli* électrocompétentes sont ensuite transformées avec le produit de recombinaison. Seules les bactéries ayant intégré les vecteurs recombinants pourront se développer sur milieu sélectif. Les plasmides recombinants sont vérifiés par digestion de restriction, par PCR ou par séquençage avant de transformer les levures ou avant d'être intégrés dans les agrobactéries.

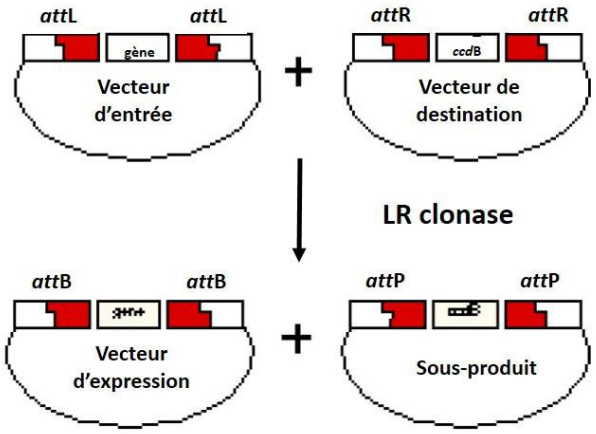


Figure 34 : Représentation schématique de la LR recombinaison via la technologie Gateway®.

La réaction de LR recombinaison se fait à l'aide du kit Gateway® LR Clonase® II Enzyme mix d'Invitrogen selon les recommandations du fournisseur de la manière suivante dans un volume réactionnel de 10 µl : 150 ng de de plasmide de destination (p*Ye*DP60_GW® ou pBIN_GW®) sont incubés pendant 1 h à 25 °C en présence de 150 ng de plasmide d'entrée (pCR8®/GW_X) et de 2 µl d'enzyme LR clonase™ (Invitrogen). 1 µl de protéinase K est ensuite ajouté au milieu réactionnel et celui-ci est incubé à 37 °C pendant 10 min.

4.11 Préparation de bactéries compétentes

4.11.1 Bactéries *E. coli* électrocompétentes

Ce protocole est utilisé pour les bactéries *E. coli* GenHogs®, XL1 Blue, MC1022 et *ccd*B survival™. 10 ml de LB liquide sont ensemencés avec des colonies de bactéries fraîchement étalées sur boîte LB. Après une nuit d'incubation à 37 °C sous agitation, 1 ml de cette préculture est utilisé pour ensemencer 100 ml de milieu LB. La culture est incubée sous agitation à 37 °C jusqu'à obtenir une DO_{700} de 0,7. La culture est ensuite centrifugée à 5000 g sans frein à 4 °C. Le culot est lavé avec de l'eau stérile deux fois, puis lavé deux fois dans 20 ml de glycérol 10%. Les bactéries ainsi rendues compétentes sont resuspendues dans 1 ml de glycérol 10 %, puis aliquotées par 40 µl et congelés dans de l'azote liquide avant d'être conservées à -80 °C.

4.11.2 Agrobactéries *A. tumefaciens* chimiocompétentes

Ce protocole est utilisé pour rendre compétentes les agrobactéries *A. tumefaciens* LBA 4404 afin de permettre leur transformation avec le plasmide binaire pBIN. Une préculture d'une colonie d'*A. tumefaciens* préalablement étalée sur une boîte YEB est mise en culture dans 2 ml de YEB additionnés des antibiotiques de sélection (rifampicine et streptomycine). Cette culture est incubée à 28 °C pendant la nuit sous agitation. Le lendemain, 2 ml de cette préculture sont utilisés pour ensemencer 50 ml de milieu qui est incubé à 28 °C pendant 4 heures. La culture est centrifugée à 4 °C à 3000 g pendant 20 minutes. Le culot est lavé deux fois avec du tampon Tris-HCl 10 mM pH 8 ; EDTA 1 mM (TE) à 4 °C. Les bactéries ainsi rendues compétentes sont resuspendues dans 2 ml de YEB, puis aliquotées par 500 µl et directement utilisées pour la transformation.

4.12 Transformation génétique de bactéries compétentes

4.12.1 Par électroporation

Ce protocole est utilisé pour les souches de bactéries *E. coli* GenHogs®, XL1 Blue, MC1022 et *ccd*B survival™. La transformation par électroporation est une technique qui consiste à introduire une construction plasmidique dans des bactéries par l'application du choc électrique de courte durée et de haut voltage qui va déstabiliser ponctuellement les membranes et les rendre poreuses et perméables. Un aliquot de 40 µl de bactéries électrocompétentes est mélangé avec 1 à 10 ng de plasmide ou produit de ligation ou produit de recombinaison. Le mélange est introduit dans une cuve d'électroporation (électrodes séparées de 2 mm) qui est placée dans le Biorad Micropulser™. L'échantillon est soumis à un choc électrique de 2,5 kV ; 200 Ω ; 25 µF puis directement resuspendu dans 700 µl de milieu LB. La suspension est incubée pendant 1 h à 37 °C avant d'être étalée sur une boîte contenant le ou les antibiotiques de sélection. Ce temps de régénération permet aux bactéries transformées d'acquérir la résistance à l'antibiotique rapportée par le plasmide. Les boîtes sont incubées pendant une nuit à 37 °C.

4.12.2 Par choc thermique

4.12.2.1 *E. coli*

Ce protocole est utilisé pour transformer les bactéries *E. coli* One Shot® TOP10 chimiocompétentes suivant les recommandations du fournisseur (Invitrogen). 1 µl de plasmide est ajouté à un aliquot de 100 µl de bactéries TOP10. Le mélange est incubé sur de la glace pendant 30 min, puis un choc thermique

est réalisé à 42 °C sans agitation pendant 30 sec. Le mélange réactionnel est immédiatement remis sur de la glace pendant 5 min. 500 µl de milieu LB sont ajoutés et la solution est incubée à 37 °C pendant 1 h avant d'être étalée sur des boîtes LB contenant le ou les antibiotiques de sélection. Les boîtes sont incubées pendant une nuit à 37 °C.

4.12.2.2 *A. tumefaciens*

500 µl d'agrobactéries chimiocompétentes sont incubées avec 0,5 à 1 µg de plasmide pendant 5 min sur la glace, puis congelées pendant 5 min dans de l'azote liquide avant d'être transférées dans un bain-marie à 37 °C pendant 5 min. 1 ml de milieu YEB est ajouté à la solution de bactéries transformées et celle-ci est incubée à 28 °C pendant 2 h. 500 µl de cette solution sont étalés sur une boîte de milieu YEB contenant les antibiotiques de sélection (rifampicine, kanamycine et streptomycine) puis incubés pendant 48 h à 28 °C.

4.13 Extraction d'ADN plasmidique

L'amplification d'un plasmide est réalisée en cultivant une colonie bactérienne transformée à 37 °C pendant une nuit dans une solution de 2 ml de milieu LB contenant l'antibiotique de sélection. Les bactéries sont ensuite sédimentées par centrifugation à 10 000 g pendant 2 min. L'extraction plasmidique est réalisée à l'aide du kit GeneJET™ Plasmid Miniprep (Fermentas) selon les recommandations du fournisseur sur le principe d'une lyse alcaline des bactéries, suivie d'une purification des plasmides sur colonne d'affinité. Les plasmides élués dans 50 µl de tampon d'élution sont conservés à -20 °C.

4.14 Construction des trois banques d'ADNc normalisées

Afin de réaliser les banques d'ADNc normalisées de *P. sativa*, *R. graveolens* et *C. cinereum*, les ARN totaux de ces plantes sont extraits à partir de feuilles de ces trois plantes. 100 à 300 µg d'ARN totaux de ces plantes ont été envoyés à la société Eurofins MWG Operon en Allemagne. Les séquences d'ADN complémentaires normalisées des ARN extraits sont séquencées par des approches de pyroséquençage grâce à la technologie GS FLX Titanium type 454 (Roche). Cette technologie permet l'obtention de 500 000 séquences par cycle de production et génère des longueurs moyennes supérieures à 350 bases. Les résultats de ce séquençage massif sont renvoyés sous forme de fichier informatique et sont analysés à l'aide de la suite logicielle blast2GO ou grâce au logiciel BioEdit.

4.15 Séquençage

1 µg de plasmide contenant notre séquence d'intérêt est envoyé à la société Eurofins MWG Operon en Allemagne pour effectuer le séquençage. Les résultats du séquençage en 3' et en 5' nous sont envoyés sous deux jours ouvrés sous forme de fichier informatique et analysés à l'aide du logiciel BioEdit.

4.16 Electrophorèse sur gel polyacrylamide en conditions dénaturantes ou SDS PAGE

L'électrophorèse en gel de polyacrylamide contenant du dodécylsulfate de sodium (SDS) permet de séparer les protéines selon leur poids moléculaire apparent. La concentration du gel en acrylamide détermine la porosité de la matrice et influence donc la vitesse de migration des protéines. Les protéines sont d'abord concentrées dans un gel de concentration à 5 % d'acrylamide en fonction du pH du tampon, puis séparées dans un gel de séparation à 10 % d'acrylamide qui permet d'obtenir un domaine de fractionnement de 15 à 75 kDa. Le SDS est un détergent qui se lie aux protéines et permet de les dénaturer totalement. Le SDS masque entièrement les charges naturelles de la protéine en le recouvrant de manière uniforme. Chaque protéine acquiert une densité de charge nette négative identique ce qui permet de séparer les protéines en fonction de leur taille et non de leur charge. 40 µl d'échantillon protéique au préalable additionnés de 10 µl de tampon de charge (Tris HCl 125 mM pH 6,8, SDS 5 %, glycérol 20 %, β-mercaptoéthanol 5 %, bleu de bromophénol 0,05 %) contenant du dithiothréitol (DTT) à 400 mM sont chauffés 5 min à 95 °C. Le DTT est un agent réducteur fort qui permet de réduire les ponts disulfures des protéines. Les échantillons ainsi que 5 µl de marqueur de taille (Prestained Protein Molecular Weight Marker ready-to-use, #SM0441, Fermentas) sont directement déposés dans les différents puits du gel. Le gel est placé dans une cuve de migration contenant le tampon de migration (Tris/glycine 1X (à partir de la solution mère 10X contenant 30,4 g de Tris, 144 g de glycine pour 1 litre pH 8,3), SDS 0,1 %) et une tension électrique de 100 V est appliquée pendant 90 à 120 min.

Une fois l'électrophorèse achevée, les protéines sont séparées mais restent incluses dans le gel et sont donc inaccessibles aux révélateurs. Les différentes protéines contenues dans le gel sont transférées sur une membrane de nitrocellulose (GE Healthcare Amsterham PVDF Hybond^{TM}-P) immobilisante. Le principe du transfert sur membrane est le même que pour l'électrophorèse ; les protéines après séparation sont toujours liées au SDS et peuvent donc migrer de l'anode à la cathode sous l'influence d'un champ électrique. Le gel de polyacrylamide est mis en contact avec la membrane de nitrocellulose, le tout est entouré de papier filtre afin de combler les espaces vides entre le gel et la membrane. La migration pour réaliser le transfert s'effectue à 80 V pendant 45 min dans un tampon Tris/glycine 1X, éthanol 20 %. Un bloc froid est ajouté dans la cuve de migration afin d'éviter l'échauffement de la solution tampon. Le marqueur de taille initialement présent sur le gel permet d'estimer l'état du transfert sur membrane.

4.17 Western blot

Le western blot est une technique d'immunodétection qui permet la visualisation des protéines d'intérêt dans un échantillon donné. Après avoir séparé les protéines sur gel de polyacrylamide en conditions dénaturantes et les avoir transféré sur membrane, celle-ci est rincée dans une solution de PBS X1 – Tween (à partir de la solution PBS mère 10X contenant 87,66g de NaCl, 2,04 g de KH_2PO_4, 14,24 g de Na_2HPO_4, 2,01 g de KCl pour 1 litre à pH 7,4 ; Tween 20 à 1 %). La membrane est ensuite immergée pendant 90 min dans une solution PBS X1 – Tween contenant 5 % de lait en poudre afin de bloquer les sites aspécifiques. Puis elle est immergée pendant 90 min dans une solution (PBS X1 – Tween contenant 5 % de lait en poudre) contenant un anticorps primaire spécifique de la protéine d'intérêt (Anti-His (C-term) Antibody, Invitrogen). La membrane est ensuite rincée trois fois 10 min à l'aide d'une solution PBS X1 – Tween et mise en présence de l'anticorps secondaire (Pierce Goat Anti-Mouse IgG, (H+L), (min x BvHsHn Sr Prot) Alkaline Phosphatase Conjugated, Thermo Scientific) contenu dans une solution PBS X1 – Tween contenant 5 % de lait en poudre pendant 90 minutes. La membrane est ensuite rincée trois fois 10 minutes à l'aide d'une solution PBS X1 – Tween, puis incubée deux fois dans 10 ml d'une solution de Tris-NaCl (Tris 0,1 M pH 9,6 et NaCl 0,1M). Enfin, les protéines sont révélées par réaction colorimétrique (NBT/BCIP). Le 5-bromo-4-chloro-3-indole-phosphate (BCIP) est un substrat incolore de la phosphatase alcaline. En sa présence, l'enzyme phosphoryle BCIP et forme un produit instable. Deux de ces molécules vont se lier pour former le 5'-5'-dibromo-4-4'-dichloro-indigo qui va réduire le 4-nitroblue tétrazolium chloride (NBT). Le NBT ainsi réduit change de couleur et devient le bleu de formazan visualisable. La révélation est réalisée en ajoutant 37,5 µl de BCIP (100 mg/ml, Roche) et 50 µl de NBT (100 mg/ml, Roche) à 10 ml de tampon Tris-NaCl. La réaction s'effectue à l'obscurité et sa durée dépend de la quantité de protéine d'intérêt présente sur la membrane. Celle-ci est ensuite rincée à l'eau distillée et photographiée.

4.18 Quantification du niveau d'expression des gènes par PCR en temps réel

4.18.1 Quantification relative

Afin de quantifier le niveau d'expression des P450s chez le panais, un gène de référence est choisi (ARN 28S) afin de réaliser une quantification relative. Le niveau d'expression du gène d'intérêt est exprimé en fonction d'un gène de ménage considéré comme stable, obtenu à partir du même échantillon. Dans notre cas, la quantification relative s'appuie sur les efficacités respectives des gènes de référence et cible, et les Ct obtenus pour les échantillons et le calibrateur sans courbe de calibration, c'est la méthode des ΔΔCt. Les résultats sont exprimés sous forme de ratio gène cible/référence de chaque échantillon et

normalisés par le rapport cible/référence d'un échantillon appelé calibrateur (ARN 28S, échantillon où le gène est le moins bien exprimé). Les résultats sont analysés à l'aide du logiciel PikoReal™ Software 2.2 (Thermo Scientific).

4.18.2 Préparation du matériel

Les extractions d'ARN totaux des différentes plantes sont réalisées à l'aide du kit RNeasy® Plant Mini de Quiagen comme décrit précédemment. Les ARNm sont convertis en ADNc à l'aide du kit High Capacity RNA-to-cDNA comme décrit précédemment afin d'obtenir 75 ng à 100 ng d'ADNc pour chaque échantillon. Les ADNc sont ensuite dilués au 1/100ème avant d'être congelés à -20 °C.

4.18.3 Mélange réactionnel

Les amorces du gène de référence ainsi que les amorces de nos gènes d'intérêt sont conçues à l'aide du logiciel Primer Express 3.0. La spécificité de chaque couple d'amorces est vérifiée en réalisant une série de 5 dilutions en cascade au 1/5ème d'une solution d'ADNg de panais. Les couples d'amorces retenus présentent une efficacité de 97 à 103 % et les PCR réalisées sur les gènes cibles et référence ont donc une efficacité très proche.
Le mélange réactionnel contient 10 µl de SYBR Premix Ex Taq™ (2X) (Tli RNaseH Plus), Bulk de Takara, 0,4 µM d'amorces sens et antisens, 0,4 µl de ROX Reference Dye (50X), 5 µl d'ADNc dilués au 1/100ème et de l'eau ultrapure stérile pour arriver à un volume réactionnel final de 20 µl. Les réactions de PCR quantitative sont réalisées à l'aide du thermocycler Pikoreal™ (Thermo Scientific) suivant le protocole suivant : 95 °C pendant 10 min ; puis 40 cycles : 95 °C 5 sec, 60 °C 30 sec ; suivi d'une étape de courbe de fusion. Les résultats sont analysés à l'aide du logiciel PikoReal™ Software 2.2 (Thermo Scientific).

5 Systèmes d'expression hétérologue

5.1 Système d'expression hétérologue dans les cellules eucaryotes de type levure

5.1.1 Préparation de levures compétentes

La veille de la transformation, des colonies de levures (WAT11 ou WAT21) sont resuspendues dans 10 ml de milieu YPGA et incubées pendant la nuit à 30 °C. Le lendemain, cette préculture est diluée dans 50 ml de milieu YPGA pour obtenir une DO_{700} de 0,2. La culture est incubée à 30 °C sous agitation

jusqu'à obtention d'une DO_{700} de 0,8. Les levures sont alors sédimentées à 4000 g pendant 10 min. Le culot est repris dans 1 ml d'eau ultrapure stérile, transféré dans un tube de 2 ml avant d'être centrifugé à 4000 g pendant 2 min. Le culot est repris dans 1 ml d'acétate de lithium 100 mM/TE 1X (0,1 mM Tris HCl pH 7,5 ; 10 µM EDTA). Une dernière étape de centrifugation à 4000 g permet de concentrer les levures dans 200 µl d'acétate de lithium 100 mM/TE 1X. Les levures sont ensuite directement transformées.

5.1.2 Transformation de levures

Afin de réaliser la transformation de nos levures, 50 µl de levures compétentes sont incubées à 30 °C pendant 30 min en présence de 500 µl d'acétate de lithium 100 mM/TE 1X dilué dans du PEG 4000 (50%), de 10 µl d'ADN de sperme de saumon dénaturé (10 mg/ml) et de 100 à 200 ng de plasmide. Un choc thermique est réalisé pour transformer les levures en incubant le mélange à 42 °C pendant 15 min, puis en le plaçant sur de la glace pendant 2 min. Le mélange est ensuite centrifugé pendant 2 min à 4000g et le culot est lavé deux fois avec 500 µl de milieu YPGA. Les levures sont ensuite régénérées à 30 °C pendant 2 h avant d'être étalées sur des boîtes contenant le milieu sélectif SGI. Les boîtes sont placées à 28 °C à l'étuve et l'apparition de colonies de levures transformées peut être observable à partir de 48 h d'incubation.

5.1.3 Expression des P450s dans les levures

L'expression hétérologue des P450s dans la levure ainsi que la préparation de microsomes sont des protocoles adaptés des méthodes mises au point par Diesperger et Pompon (Diesperger *et al*., 1974; Pompon *et al*., 1996).

5.1.3.1 Conditions de culture

Une colonie de levure transformée est utilisée pour ensemencer 10 ml de milieu sélectif SGI. La culture est incubée pendant 24 h à 28 °C sous agitation orbitale (170 rpm). Cette préculture est utilisée pour ensemencer 200 ml de milieu YPGE contenant du glucose en faible quantité et de l'éthanol. La culture est incubée sous agitation orbitale pendant 24 h à 28 °C. Le but de cette étape est d'augmenter la biomasse des levures transformées sans induire l'expression de la protéine d'intérêt. Au bout de 24 h de culture, les levures ont entièrement utilisé le glucose et utilisent l'éthanol comme source carbonée. L'expression de la protéine d'intérêt est obtenue en remplaçant le milieu YPGE par 200 ml de milieu

YPL contenant du galactose (20 g/l). Les levures sont incubées pendant 16 h à 18 °C avant de procéder à la préparation de microsomes.

5.1.3.2 Préparation de microsomes

Afin de limiter la dégradation de notre protéine d'intérêt la totalité de la préparation de microsomes s'effectue à 4 °C. Les levures en cultures sont sédimentées pendant 10 minutes à 6 000 g, puis lavées dans 35 ml de tampon TEK froid (Tableau 7, p. 117). Après une centrifugation de 10 min à 6 000 g, le tampon TEK est remplacé par 2 ml de tampon d'extraction TES (Tableau 7, p. 117). Des billes de verre de 425 à 600 µm de diamètre (Sigma Life Science) sont ajoutées dans le tube et les levures sont broyées par 5 agitations manuelles d'une minute en replaçant le tube sur la glace 1 min entre deux agitations. Les billes de verres sont ensuite lavées plusieurs fois avec du tampon TES jusqu'à obtenir un volume final de 25 ml. La précipitation des microsomes se réalise par ultracentrifugation (Thermo Scientific Sorvall WX ultra 80) à 100 000 g pendant 1 h. Les microsomes culottés au fond du tube sont lavés avec 500 µl de tampon TEG (Tableau 7, p. 117), puis resuspendus dans 2 ml de tampon TEG et homogénéisés à l'aide d'un Potter. Les microsomes sont aliquotés par fraction de 200 µl, congelés dans de l'azote liquide et stockés à – 20 °C.

	TEK	TES	TEG
Tris HCl pH 7,5	50 mM	50 mM	50 mM
EDTA	1 mM	1 mM	1 mM
KCl	100 mM		
Sorbitol*		0,6 M	
BSA*		1 % (p/v)	
B-mercaptoéthanol*		20 mM	
Glycérol			30 % (v/v)

Tableau 7 : Composition des tampons utilisés pour la préparation des microsomes de levures.
* produits ajoutés le jour même de la préparation de microsomes.

5.2 Système d'expression hétérologue dans les plantes

5.2.1 Principe de l'expression transitoire

L'expression transitoire par agro-infiltration consiste à exprimer un ou plusieurs gènes d'intérêt dans le noyau des cellules du mésophylle et de l'épiderme foliaire, sans que le gène soit intégré dans le génome de la cellule. L'expression des protéines d'intérêt n'est possible que pendant quelques jours (5 à 10 jours en fonction de la protéine).

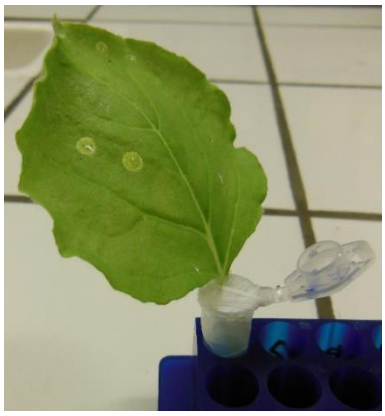
5.2.2 Préparation de l'inoculum bactérien

Une colonie d'agrobactéries transformées avec le plasmide pBIN_GW®_X contenant le gène d'intérêt, une colonie transformée avec le plasmide pBIN_GW®_p19 (fourni par le Pr. David Baulcombe de l'université de Cambridge) contenant le suppresseur de « gene silencing » viral p19 ainsi qu'une colonie transformée avec le plasmide pBIN_GW®_ATR1 contenant le gène codant pour la CPR 1 d'*A. thaliana* sont mises en culture séparément dans 50 ml de milieu YEB contenant les antibiotiques (rifampycine à 20 mg/l et kanamycine à 50 mg/l). La culture est incubée à 28 °C pendant 24 heures. Les agrobactéries sont sédimentées par centrifugation à 4 000 g pendant 15 minutes. Les culots obtenus sont lavés trois fois avec 30 ml d'eau déminéralisée stérile et le culot est suspendu dans 1 ml d'eau déminéralisée stérile afin d'éliminer toute trace d'antibiotique (nocif pour la plante). L'inoculum bactérien est préparé de manière à obtenir une DO_{600} de 0,2 pour les agrobactéries contenant le gène d'intérêt ainsi que le gène *ATR1* et le vecteur vide, et une DO_{600} de 0,4 pour les agrobactéries contenant le gène codant pour la protéine p19. De la même manière, une solution d'agrobactéries contenant le pBIN_GW® vide sera préparée afin de servir de témoin négatif. La solution est conservée sur la glace jusqu'à infiltration des feuilles de tabac. Il est possible juste avant de réaliser l'infiltration, d'ajouter de l'acétosyringone à 100 µM dans l'inoculum bactérien afin d'activer les gènes de virulence des agrobactéries.

5.2.3 Agro-infiltration de *N. benthamiana*

Deux à trois feuilles par plant de *N. benthamiana* âgés d'un à deux mois sont sélectionnées pour l'agro-infiltration. Les feuilles sont lésées sur leur face abaxiale à l'aide d'une aiguille, puis la solution d'agrobactéries est directement injectée dans le tissu à l'aide d'une seringue dépourvue d'aiguille (Photographie 3, p. 119). Après infiltration, les plantes sont placées dans des compartiments spécifiques confinés au phytotron pendant 4 jours afin d'éviter la contamination des autres plantes par les agrobactéries. Les feuilles infiltrées par les agrobactéries sont utilisées (au bout de 96 h) pour effectuer

une deuxième infiltration avec des mélanges de 5 à 6 substrats à 100 µM ou utilisées pour effectuer une préparation de microsomes. Après la deuxième infiltration, les feuilles sont soient coupées au niveau du pétiole et placées dans le reste de la solution de substrats (Photographie 3, p. 119) pendant 3 h avant de réaliser une extraction de phénylpropanoïdes, soit laissées directement sur la plante et broyées au bout de trois heures.



Photographie 3 : Expression transitoire de P450s chez *N. benthamiana* avec co-infiltration de substrats.

5.2.4 Préparation de microsomes de plantes

Quatre jours après l'infiltration par des agrobactéries, les feuilles sont pesées afin d'avoir le même poids total pour chaque mélange infiltré. Les feuilles sont placées dans un tube de 50 ml contenant 30 ml de tampon Kpi 0,1 M à pH 7 et une tablette d'inhibiteur de protéases (complete Mini, EDTA-free, Protease inhibitor cocktail tablets, Roche). Les feuilles sont broyées à l'ultraturax (Polytron® PT 2100, agrégats PT-DA 2112/2EC et PT-DA 2105/EC, Kinematica) jusqu'à obtention d'une solution homogène. 0,1 % de poly(vinylpolypyrrolidone) (p/p) par gramme de feuille ainsi que 1 ml de DTT sont ajoutés à la solution. Le mélange est centrifugé 30 min à 7 000 g et le surnagent est filtré sur du papier Whatman® (grade 2V, 8 µm) et réparti dans deux tubes prévus pour l'ultracentrifugeuse. Les microsomes sont culottés à 100 000 g pendant 1 h à 4 °C. Les microsomes sont repris dans 600 µl de tampon Kpi, aliquotés par 200 µl, congelés dans de l'azote liquide et stockés à -20 °C.

6 Méthodes d'analyses biochimiques

6.1 Extraction de composés polyphénoliques

6.1.1 Protocole général

Le même protocole est utilisé pour réaliser toutes les extractions de composés polyphénoliques à partir des feuilles, de tiges et de racines issues des différentes plantes de cette étude. 50 à 100 mg de matériel végétal frais sont congelés et broyés dans de l'azote liquide à l'aide d'un mortier en céramique et d'un pilon au préalablement refroidis. 1 ml d'une solution de méthanol à 80 % contenant ou non le standard interne à 5 µM (pimpinelline) est ajouté à la poudre ainsi obtenue. L'ajout de l'étalon interne permet de s'affranchir des pertes engendrées par la préparation de l'échantillon, mais également des volumes injectés en chromatographie liquide. Le mélange est homogénéisé à l'aide d'un vortex pendant 1 min, puis centrifugé pendant 30 min à 10 000 g. Le surnageant est prélevé et évaporé dans un nouveau tube en utilisant une centrifugeuse sous vide (Concentrator Plus, Eppendorf). Les résidus sont resuspendus dans 300 µl de méthanol, puis la solution est centrifugée pendant 30 min à 10 000 g avant d'être placée dans des vials. Les échantillons sont conservés à -20 °C avant d'être analysés par chromatographie.

6.1.2 Protocole utilisé lors de la deuxième expérience d'élicitation

50 à 100 mg de racines fraîches sont lyophilisés puis broyés pendant 2 min à l'aide d'un broyeur à bille (MM 400, Retsch®). 500 µl d'une solution de méthanol à 80 % contenant le standard interne à 5 µM (pimpinelline) est ajouté à la poudre ainsi obtenue. Le mélange est homogénéisé à l'aide d'un vortex pendant 1 min, puis centrifugé pendant 30 min à 10 000 g. Le surnagent est prélevé et mis à évaporer dans un nouveau tube en utilisant une centrifugeuse sous vide (Concentrator Plus, Eppendorf). Les résidus sont resuspendus dans 150 µl de méthanol, puis la solution est centrifugée pendant 30 min à 10 000 g avant d'être placée dans des vials. Les échantillons sont conservés à -20 °C avant d'être analysés par chromatographie.

6.2 Mesure d'activités enzymatiques

6.2.1 Substrats

La liste de la totalité des molécules utilisées pour le criblage métabolique comportant les informations telles que le poids moléculaire ou le numéro CAS se trouve en Annexe 2 (p. 284-287). Les solutions stocks sont préparées à 10 mM ou à 1 mg/ml dans du DMSO ou de l'éthanol et stockées à -20 °C.

6.2.2 Détermination des constantes cinétiques

Les réactions sont conduites dans un volume réactionnel de 100 µl comprenant 20 µl de microsomes de tabac ou de levure, 200 µM de substrat, 1 mM de NADPH (Sigma) et du tampon phosphate de sodium

NaPi 0,1 M pH 7,5. De la même manière, les microsomes sont incubés en absence de cofacteur afin de constituer un contrôle négatif. L'incubation est réalisée pendant 1 min 30 à 30 min à 28 °C sous agitation (160 rpm) à l'aide d'un Thermomixer (Eppendorf). La réaction est stoppée par l'ajout de 37,5 µl d'acétonitrile/HCl (1%). La solution est ensuite centrifugée à 8 000 g pendant 30 min. Le surnagent est prélevé afin d'être analysé par HPLC ou LC-MS/MS.

Afin de déterminer les constantes cinétiques, l'incubation est réalisée avec des concentrations croissantes en substrat (de 0,1 µM à 200 µM). Les expériences sont répétées trois fois et le temps de la réaction est adapté pour que la conversion du substrat n'excède pas 50 % pour les concentrations les plus faibles. Les constantes cinétiques apparentes sont déterminées par le logiciel Sigmaplot (Systat Software Inc., http://www.sigmaplot.com/). La représentation en double inverse de Lineweaver-Burk qui exprime l'inverse de l'activité enzymatique en fonction de l'inverse de la concentration en substrat permet d'obtenir les constantes d'affinités K_m de l'enzyme.

6.2.3 Test de l'inhibition par des furanocoumarines

Les réactions enzymatiques effectuées pour les tests d'inhibitions des P450s sont réalisées exactement dans les mêmes conditions que pour la détermination des constantes cinétiques. L'inhibiteur est ajouté au milieu réactionnel à la concentration de 100 µM.

6.3 Détection des produits formés par HPLC

Les échantillons issus du criblage métabolique sont analysés sur une colonne C18 en phase inverse (LiChrospher RP-18, 250x4.6, 5 µm) équipée d'une précolonne (LiChroCART® 250-4) par HPLC (Shimadzu Corp., Kyoto, Japan). 50 µl de l'échantillon sont injectés. La séparation est réalisée à un débit de 0,7 ml/min à l'aide d'un solvant A (eau ultrapure, acide formique 0,1 %) et d'un solvant B (méthanol, acide formique 0,1 %). Les composés sont séparés à l'aide d'un gradient de phase mobile (A : B ; v/v), (90 : 10) jusqu'à 3 min, (1 : 99) à 34 min, (1 : 99) à 37 min, (90 : 10) à 38 min, (90 : 10) à 43 min. Une détection de longueur d'onde entre 200 et 600 nm est réalisée et les chromatogrammes sont extraits à la longueur d'onde maximale du composé (en général entre 280 nm et 350 nm). L'identification des produits formés se fait par recoupement des temps de rétention et des spectres d'absorbance par rapport à ceux des molécules standards. Les résultats sont confirmés grâce aux spectres de fragmentations obtenus par LC-MS/MS. Lorsque les standards ne sont pas disponibles dans le commerce, les fractions correspondantes aux produits formés sont collectées et les molécules sont analysées par RMN du proton.

6.4 Identification des produits formés par spectrométrie de masse

La spectrométrie de masse (MS) est une technique de détection performante qui permet d'identifier les molécules d'intérêts par comparaison des spectres de fragmentation avec une grande précision par mesure des intensités associées à leurs masses. Les molécules chargées sont séparées suivant leur rapport masse/charge (m/z).

Le système LC-MS/MS disponible au Laboratoire Commun d'Analyses de l'ENSAIA (Vandœuvre-lès-Nancy) est équipé d'une pompe binaire et d'un spectromètre de masse à trappe ionique (LTQ-MS, Thermo Scientific, San Jose, CA, USA). La séparation des molécules se réalise par chromatographie liquide (Thermo Fischer Scientific, San Jose, CA, USA) avec une colonne C18 en phase inverse (C18 Alltima revrese phase column 150 x 2.1mm, 5µm, Grace/Alltech, Darmstadt, Germany). La méthode de séparation utilisée dure 36 minutes avec un débit de 0,2 ml/min et utilise les mêmes mélanges de solvants que ceux décrits précédemment. La méthode de séparation est la suivante : (100 : 0) à 0 min, (90 : 10) à 25 min, (100 : 0) à 30 min, (90 : 10) à 36 min. 20 µl de l'échantillon à analyser sont injectés. L'ionisation des molécules est réalisée par électronébulisation (ESI) en mode positif. Le champ électrique de nébulisation est de 4,5 kV et la température du capillaire maintenue à 300 °C. Les débits du gaz coaxial, du gaz auxiliaire et du gaz barrière sont paramétrés respectivement à 40, 10 et 10 (unité arbitraire.min^{-1}). Le voltage du capillaire de transfert, du split lens et du front lens sont respectivemet de 36 V, -44V et -3,5V. L'identification des composés se fait grâce à la réalisation d'un scan complet entre 50 et 1000 m/z et d'une recherche spécifique de la masse du substrat et de ses produits attendus. L'analyse des données est réalisée à l'aide du logiciel Xcalibur (version 2.1).

6.5 Dosage de furanocoumarines par LC-MS/MS

Le dosage des furanocoumarines dans les extraits de panais est réalisé par rapport à un étalon interne (pimpinelline) lors des expériences d'élicitation. L'absence de pimpinelline dans les extraits réalisés à partir de feuilles, de tiges et de racines de panais a été vérifiée au préalable et a permis d'utiliser cette molécule comme étalon interne dans les expériences d'élicitation.

7 Traitement statistique des données

Le test de Kruskal-Wallis, test non paramétrique, a été utilisé avec une *p-value* < 0,05 afin de rejeter l'hypothèse d'égalité entre le groupe de plantes témoins et les groupes de plantes traitées avec différents éliciteurs grâce au logiciel R 3.0.0 (R core team 2013).

PARTIE 3 : Résultats et discussion

CHAPITRE 1 : CONSTRUCTION ET ANALYSE DE 3 BANQUES TRANSCRIPTOMIQUES DE PLANTES PRODUCTRICES DE FURANOCOUMARINES EN VUE D'IDENTIFIER DE NOUVEAUX GENES CANDIDATS IMPLIQUES DANS LA SYNTHESE DE FURANOCOUMARINES

1 Exploitation de trois banques normalisées d'ADNc de plantes produisant des furanocoumarines

1.1 Recherche de gènes candidats

Des avancées ont récemment été réalisées sur le déterminisme moléculaire de la voie de biosynthèse des furanocoumarines. Des gènes codants pour une méthyltransférase et pour une prényltransférase ont été identifiés chez les Apiacées et une enzyme codant pour une dioxygénase a été isolée chez les Rutacées. Parmi les différentes enzymes supposées intervenir dans la voie de biosynthèse, seul 2 gènes de P450s ont été caractérisés. Le premier code pour une enzyme qui catalyse la synthèse de psoralène (CYP71AJ1, AJ2, AJ3) et le second pour une enzyme intervenant dans la synthèse de l'angélicine (CYP71AJ4). Malgré ces percées, une part importante de cette voie de biosynthèse reste à l'heure actuelle encore inconnue dans la mesure où la majorité de ces étapes enzymatiques est catalysée par des enzymes de la famille des cytochromes P450s (Hamerski and Matern, 1988a; Hamerski and Matern, 1988b; Hamerski *et al*., 1990; Matern *et al*., 1988). Dans ce premier chapitre, je détaillerai les approches mises en œuvre pour identifier d'autres gènes codant pour des enzymes intervenant dans cette voie de biosynthèse, en particulier les enzymes de la famille des P450s et ce chez différentes plantes : *P. sativa*, *R. graveolens* et *C. cinereum.*

1.1.1 Approche bibliographique

Les cytochromes P450 représentent l'une des familles enzymatiques les plus diversifiées chez les végétaux supérieurs. Il existe des exemples de voies de biosynthèse qui impliquent une seule famille de P450s. Ainsi, des études effectuées chez le maïs ont montré que la totalité des enzymes de type P450 intervenant dans la voie de biosynthèse du DIMBOA appartiennent tous à la même sous-famille de CYP71C (Frey *et al*., 1997; Glawischnig *et al*., 1999). Si ce genre de duplications géniques suivies de mutations et néo-fonctionnalisation a été observé dans certains cas, dans d'autres cas, une voie de biosynthèse peut faire intervenir plusieurs P450s différents. C'est le cas, par exemple, de la voie de

biosynthèse de la dhurrine chez le sorgho ou de la voie de biosynthèse des momilactones chez le riz (Bak *et al*., 2000; Okada *et al*., 2009; Shimura *et al*., 2007). Les résultats obtenus au laboratoire en 2007 et en 2009 ont permis de montrer que la sous-famille CYP71AJ est impliquée dans la synthèse des furanocoumarines puisque 2 enzymes de cette même sous-famille catalysent 2 réactions différentes. La première hypothèse que nous avons donc émise est que, à l'image de la synthèse du DIMBOA, d'autres gènes de cette même sous-famille jouent un rôle dans la synthèse de ces molécules. La seconde hypothèse serait que des P450s de familles différentes sont impliqués dans la synthèse de ces molécules. La diversité des cytochromes P450 est telle que d'autres familles peuvent également être envisagées comme étant impliquées dans la synthèse des furanocoumarines. Des expériences menées sur *Catharanthus roseus* ont mis en évidence que CYP72A1 était capable de réaliser une réaction de décyclisation de la loganine en sécologanine (Irmler *et al*., 2000). Cette réaction est une réaction inverse de celle de la marmésine synthase (Figure 35-B et C, p. 127). Bien que cette réaction s'effectue sur la loganine qui est une molécule glycosylée, la sous-famille CYP72A pourrait potentiellement être impliquée dans la synthèse des furanocoumarines.

1.1.2 Approche expérimentale

La seconde hypothèse émise ci-dessus est renforcée par des données expérimentales préliminaires obtenues en 2007 dans le cadre de travaux réalisés en collaboration avec l'équipe du Pr. Matern de l'université de Marburg en Allemagne. Ces travaux réalisés sur des cultures cellulaires d'*A. majus* ont montré que l'ajout d'extraits de *Phytophtora megasperma* (*Pmg*) dans le milieu de culture entraînait après 3 à 5 h une accumulation de coumarines et de furanocoumarines (Hamerski *et al*., 1990). Pour identifier les gènes spécifiquement impliqués dans la synthèse des furanocoumarines, les ARN totaux exprimés dans les cellules *d'A. majus* traitées ou non avec des extraits de champignons ont été extraits et les ARNm convertis en ADNc par une étape de transcription inverse. Les séquences nucléotidiques correspondant à des cytochromes P450 ont été amplifiées spécifiquement à l'aide d'amorces dégénérées qui ciblent les régions peptidiques consensus PERF et PFG conservés chez la plupart des P450s végétaux (Akashi *et al*., 1997; Schopfer and Ebel, 1998). Les mélanges de produits de PCR ont ensuite été analysés et les profils électrophorétiques comparés. Cette stratégie a permis d'identifier huit P450s différents présents majoritairement dans les cellules élicitées. L'obtention des séquences complètes de ces fragments a été ensuite réalisée par des techniques de 5' et/ou 3' RACE-PCR. Certains de ces gènes ont été caractérisés fonctionnellement et notamment CYP71AJ1 (psoralène synthase), CYP73A41 (C4H) et CYP98A21 (C3'H). Les cinq autres séquences isolées correspondent à CYP71D97, CYP71D98, CYP71AZ1, CYP76B8 et CYP82H1. La famille des CYP71 est la plus grande famille de P450s chez les plantes et présente un panel de fonctions extrêmement large. En dehors de ces travaux

réalisés par les collègues de Marburg, il semble très probable que la famille de gènes impliqués dans la synthèse des furanocoumarines appartienne au clade 71. Au sein de ce clade, les CYP71 proprement dit présentent une très grande plasticité et sont impliqués dans un grand nombre de réactions différentes (Mizutani and Ohta, 2010). Il est vraisemblable que l'apparition de la voie de biosynthèse des furanocoumarines soit passée par un recrutement de certaines enzymes impliquées dans d'autres voies de biosynthèse. En plus de CYP71AJ1-4, des travaux effectués chez *Mentha pulegium* ont permis de mettre en évidence la menthofurane synthase qui catalyse la réaction de cyclisation du pulégone en menthofurane (Bertea *et al*., 2001). Cette réaction est catalysée par un P450 appartenant à la sous-famille des CYP71A, et le mécanisme réactionnel mis en jeu présente une certaine analogie avec celui catalysé par la marmesine synthase (Figure 35-A et B, p. 127). De ce fait, CYP71D et CYP71AZ, isolé par l'équipe de Marburg, ont été identifiés comme étant des cibles privilégiées pour notre recherche d'enzymes impliquées dans la synthèse des furanocoumarines.

Les autres familles pointés par les travaux de l'équipe allemande, CYP76 et CYP82 et plus particulièrement les sous-familles CYP76B et CYP82H sont également potentiellement impliquées dans la voie de biosynthèse des furanocoumarines. Des travaux réalisés sur CYP82C2 et 4 isolés à partir d'*A. thaliana* montrent que ces enzymes sont capables de réaliser l'hydroxylation de la xanthotoxine (Kruse *et al*., 2008) et ce, bien que cette plante ne produise pas ce type de molécules. Ces résultats appuient l'hypothèse d'une implication de la famille CYP82 dans la synthèse des furanocoumarines.

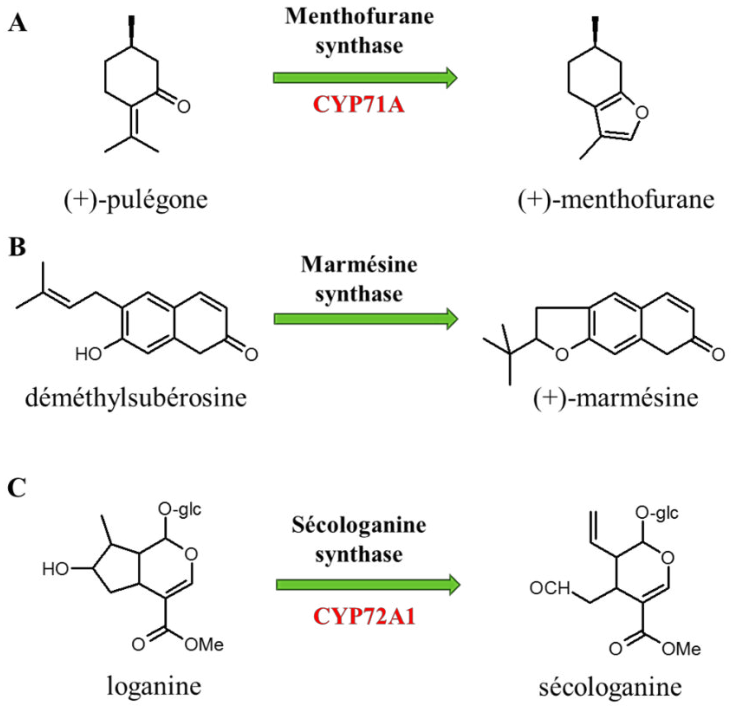


Figure 35 : Comparaison des mécanismes réactionnels impliqués dans la synthèse du menthofurane (A), de la marmésine (B) et de la sécologanine (C).

1.1.3 Approche bioinformatique

Envisager des pistes de recherche et cibler des familles de P450s est essentiel pour trouver de nouveaux candidats impliqués dans la synthèse des furanocoumarines. La mise en œuvre de la recherche de ces gènes s'est avérée plus complexe devant l'absence de données informatiques concernant la plupart des plantes produisant des furanocoumarines. Il n'existe actuellement dans les bases de données publiques, comme GenBank que 137 séquences pour *R. graveolens*, 46 séquences pour *P. sativa* et seulement 17 pour *C. cinereum* (dont 2 pour *P. cinerea*) et parmi ces séquences très peu sont relatives à des cytochromes P450. Pour pouvoir avancer dans ce projet il a donc été nécessaire, dans un premier temps, de générer de nouvelles banques de données. Le développement de nouvelles techniques de séquençage à haut débit a permis de lever ce verrou. Pour construire de telles banques de données, la première étape a consisté à préparer le matériel génétique qui sera séquencé. Nous avons choisi de faire séquencer les ARN totaux de plantes productrices de furanocoumarines. Nous nous sommes restreints à 3 plantes modèles particulières. *R. graveolens* correspond à la plante « historique » du laboratoire. En effet, le LAE dispose d'une expertise très large sur la culture de cette plante aussi bien *in vitro* qu'en terre. Par ailleurs des travaux réalisés par le passé ont mené à la mise au point d'un système permettant de la transformer génétiquement (Lievre *et al*., 2005). Cet outil s'est d'ores et déjà avéré être très utile pour l'étude *in planta* de la fonction de certains gènes. La seconde plante qui a été utilisée est *P. sativa*. Cette plante a fait l'objet d'études au laboratoire puisque c'est elle qui a permis d'isoler et d'identifier les enzymes impliquées dans la synthèse du psoralène et de l'angélicine. Cette plante a la particularité de produire à la fois les furanocoumarines linéaires et angulaires, contrairement à *R. graveolens* qui ne produit que des formes linéaires. Nous avons choisi de travailler sur une troisième famille de plantes : les Fabacées produisant également les 2 formes linéaires et angulaires. Parmi cette famille de plante, nous avons choisi de nous focaliser sur *C. cinereum*. A l'heure actuelle nous n'avons pas encore entamé d'étude sur la dernière famille de plantes produisant des furanocoumarines, à savoir les Moracées.

Pour réaliser la construction des banques de données, la première étape consiste à extraire des ARNs totaux de bonne qualité et en quantité suffisante. Dans le cas de ces plantes cette extraction s'est avérée relativement facile à faire puisqu'aucune adaptation particulière des kits commerciaux n'a été nécessaire. Les ARNs totaux ont donc été extraits comme décrit dans le matériel et méthode sur de jeunes plantules. Une analyse préliminaire a cependant été menée pour s'assurer de la présence de furanocoumarines. Afin d'en augmenter la concentration et donc *a priori* également la concentration d'ARNm qui sont à l'origine de leur synthèse, nous avons au préalable stimulé ce métabolisme. Les approches utilisées sont légèrement différentes en fonction des plantes : méthyljasmonate (MeJa) pour le panais, UV pour la rue et pulvérisation de $CuSO_4$ pour *C. cinereum*. Le traitement au MeJa a été

réalisé par pulvérisation d'une solution à 100 µM préparée dans de l'eau sur des plantules de 8 semaines. Après 48 h, les feuilles ont été récoltées et congelées. Des plantules de rue ont été exposées pendant 24 h à un rayonnement UV comme cela a été décrit dans la thèse de Romain Larbat. Enfin, le traitement au $CuSO_4$ a été effectué en suivant le protocole élaboré par Bourgaud et ses collaborateurs (Bourgaud *et al.*, 1992). Les feuilles ont été pulvérisées avec une solution à 10 mM et récoltées après 72 h. Dans les 3 cas, les ARNs ont été extraits en utilisant un kit RNAeasy commercialisé par Qiagen. L'ADN génomique a été éliminé en suivant les recommandations du fournisseur par une étape de digestion sur colonne impliquant de la DNAse I. L'extraction d'ARNs de plantules présentant un métabolisme secondaire actif est souvent un défi. Afin de vérifier que l'ARN produit est de bonne qualité, une analyse sur gel d'agarose a été effectuée (Figure 36, p. 129).

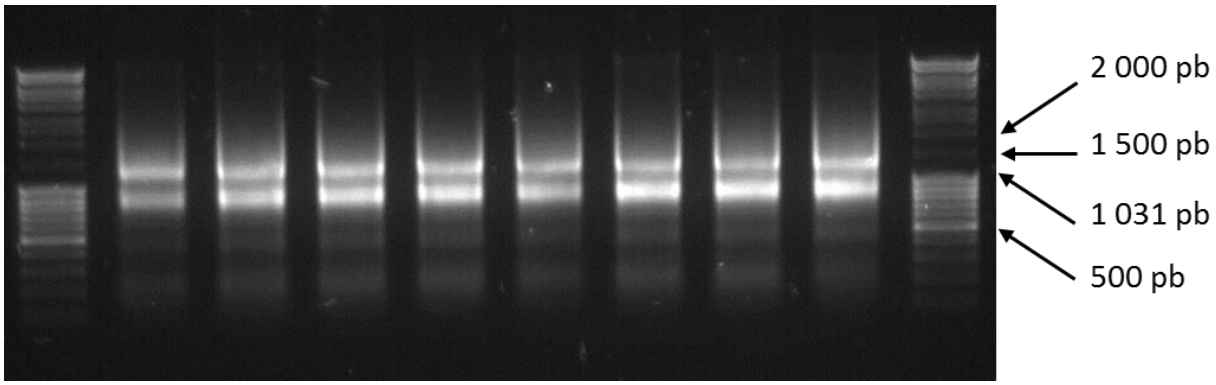


Figure 36 : Analyse sur gel d'agarose à 1 % des ARN totaux extraits de *R. graveolens*. Puits 1 et 10 : marqueur de taille SM00403. Puits 2 -9 : ARN totaux de R. graveolens.

Le travail de séquençage sur les ARNs extraits (29 µg chez *P. sativa* ; 67,4 µg chez *R. graveolens* et 113,4 µg chez *C. cinereum*) a été réalisé par un prestataire de service, Operon MWG Biotech. Pour chaque plantule, le cahier des charges prévoyait de générer entre 400 000 et 500 000 séquences d'une longueur moyenne de 350 à 450 paires de bases avec la technologie GS FLX Titanium de type 454 (Roche). L'ensemble des séquences générées par le prestataire de service ont été rassemblées dans un fichier global. Il existe peu de gènes de petite taille ne faisant que 350 à 450 paires de bases. Par exemple, les séquences codant pour un cytochrome P450 sont généralement de l'ordre de 1500 paires de bases. Pour faciliter l'analyse de la banque de données et disposer de séquences complètes de gènes candidats, un assemblage a été effectué par le prestataire de service. Ce travail de « contigage » a consisté à faire une recherche dans la banque de données pour retrouver les régions chevauchantes entre les différentes séquences brutes et ainsi reconstruire certaines séquences codantes. Ce travail a restreint le nombre de séquences à analyser puisque la majorité des séquences brutes ont pu être assemblées dans un fichier nommé « All contigs » (Tableau 8, p. 130). Un niveau supérieur d'assemblage en « Large contigs » a

également été réalisé et réduit encore le nombre de séquences à analyser (Tableau 8, p. 130). Les séquences qui n'ont pu être assemblées en contigs sont appelées singletons (Tableau 8, p. 130).

	P. sativa	*R. graveolens*	*C. cinereum*
Séquences brutes	703 719	688 686	664 422
Longueur moyenne (en nt)	311	346	437
Séquences assemblées	489 108	581 921	560 523
All contigs	78 117	62 212	50 175
Large contigs	3 047	7 486	12 423
Singletons	195 997	80 583	98 033

Tableau 8 : Bilan du séquençage et de l'assemblage des trois banques d'ADNc.

Il est également nécessaire de préciser ici un dernier point concernant la construction de cette banque. Par définition, le métabolisme secondaire correspond à des molécules présentes en faibles concentrations chez les plantes. Cette faible concentration est souvent accompagnée d'une faible représentation des ARNs messagers codant les enzymes qui en catalysent la synthèse en comparaison avec des ARNs codant des protéines comme l'actine, la tubuline ou autre. Afin de pallier ce problème nous avons demandé à ce que les banques générées soient normalisées. Ceci présente l'avantage d'avoir une représentation plus ou moins équivalente de tous les messagers. Cependant, l'inconvénient est que par ce biais, il n'est plus possible de faire une analyse quantitative du transcriptome. Bien qu'une analyse stricte de l'aspect quantitatif ne puisse être faite, nous verrons qu'il est tout de même possible de faire une légère distinction entre des gènes qui sont fortement exprimés et des gènes qui sont exprimés de manière très faible. Cette différence va être visualisable par le nombre de séquences codantes complètes pour lesquelles on dispose au minimum d'une séquence allant de l'ATG initiateur au codon stop. Dans certains cas, on disposera également des séquences non codantes 5' et 3' terminales. Par contre, dans d'autres cas, lorsque le messager est faiblement présent, on n'aura accès qu'à une partie de la séquence codante. Leur analyse sera alors plus compliquée à réaliser de manière fiable.

Afin d'identifier les différentes séquences codantes dans cette banque de données, nous en avons réalisé une analyse globale sans *a priori* de sorte à identifier les différents gènes et familles de gènes présents. Cette analyse a été faite grâce à la suite logicielle Blast2GO accessible en ligne. Les analyses faites par ce biais consistent à comparer chacune des séquences inclues dans la banque à l'ensemble des séquences disponibles dans les bases de données publiques. Cette première étape fait appel à un algorithme qui se nomme Basic Local Alignment Search Tool (BLAST). Il va permettre de rechercher les séquences

présentant le taux d'homologie le plus proche dans l'ensemble des séquences déjà décrites. Cette suite logicielle va également permettre d'identifier un certain nombre d'enzymes à partir des termes d'InterPro (IPR) et des annotations des champs de Gene Ontology (GO).

1.1.3.1 Recherche globale : utilisation de la suite logicielle blast2GO

Afin de restreindre le nombre de séquences à analyser, nous avons choisi de travailler sur le fichier All contigs. En effet, ce fichier réduit d'un facteur 10 le nombre de séquences par rapport au nombre de séquences brutes et reste représentatif des gènes exprimés au moment de l'extraction d'ARN totaux (Tableau 8, p. 130). L'ensemble des séquences assemblées en All contigs a été confronté aux banques de données publiques disponibles. Le nombre d'informations générées étant relativement important et nécessitant un temps de travail considérable, nous avons fragmenté les fichiers All contigs des différentes banques en sous parties de 2000 séquences. Ces batchs de 2000 séquences ont étés soumis à l'analyse en suivant le protocole ci-dessous.

Dans un premier temps, l'ensemble des séquences a été comparé à la base de données NCBInr (nr : non-redundant) en utilisant l'algorithme blastx afin d'y rechercher des séquences homologues. Pour éviter de constituer des fichiers trop volumineux et accélérer le temps d'analyse, nous avons restreint les résultats aux 5 meilleurs alignements. Les données obtenues ont, dans un second temps été utilisées pour effectuer un GO-mapping. Cette étape consiste à associer des descriptions de gènes (Gene Ontology) aux séquences homologues. Ces descriptions sont classées en trois catégories : fonction moléculaire, processus biologique et compartiment cellulaire. Enfin, une étape d'annotation est réalisée dans un troisième temps et permet d'assigner des annotations de motifs InterPro (IPR), Enzyme Code (EC) ou encore KEGG Maps aux différentes séquences.
Une analyse globale des données générées indique que les 3 banques construites sont comparables en termes de qualité même si les séquences contenues dans la banque d'ADNc de panais ont reçu sensiblement moins d'annotations GO. Dans tous les cas, on constate qu'il y a, en moyenne entre 2 et 8 annotations GO par séquences analysées (Figure 37, p. 132).

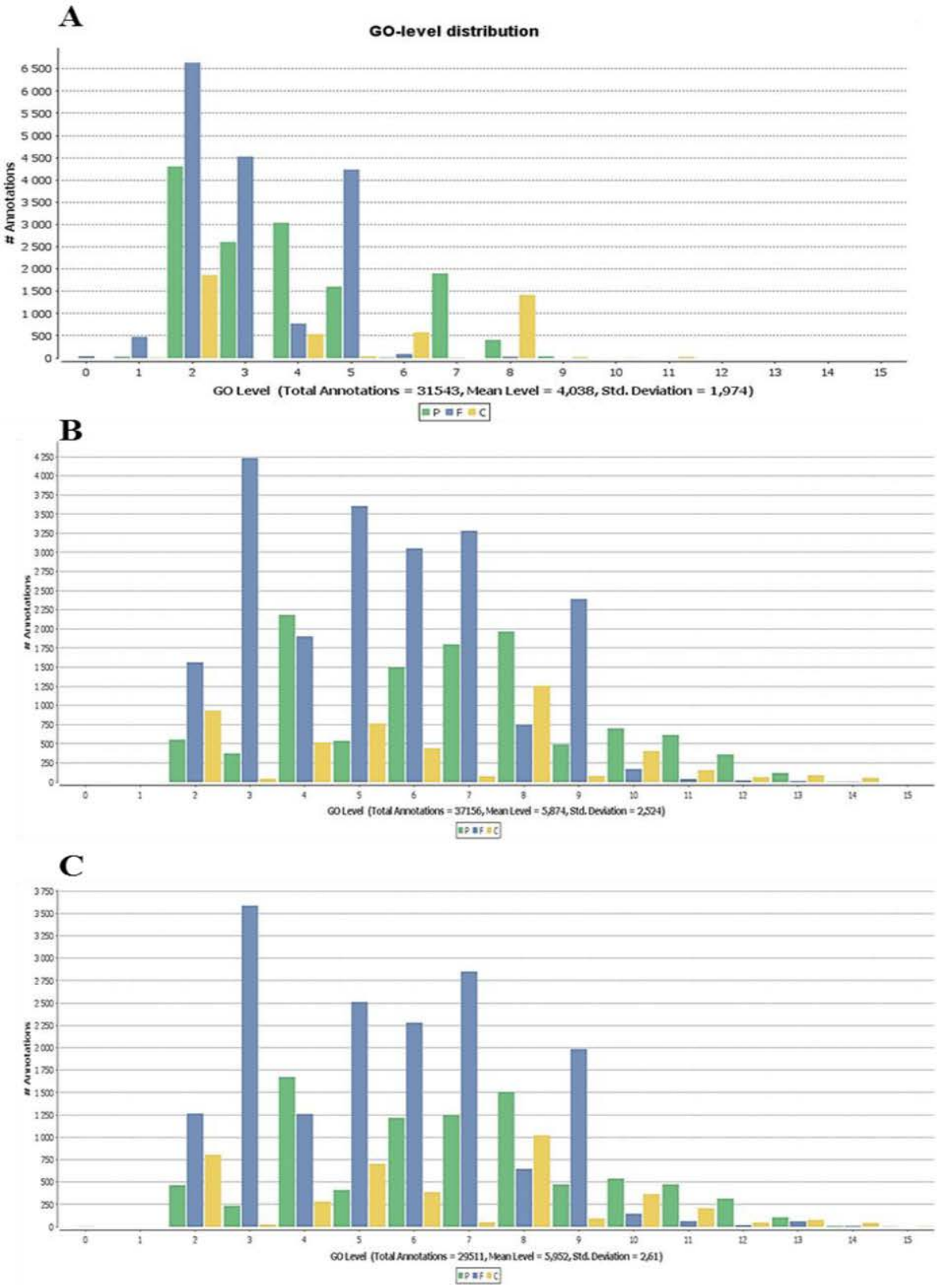


Figure 37 : Analyse GO de la banque d'ADNc de *P. sativa* (A), *R. graveolens* (B) et *C. cinereum* (C). La plupart des séquences ont entre 2 et 8 annotations GO (P = processus biologique, F = fonction moléculaire, C= compartiment cellulaire).

Une analyse plus détaillée des fonctions moléculaires indique que de nombreuses similitudes sont observées chez le panais et la rue essentiellement avec des gènes potentiels identifiés dans le génome de la vigne, du peuplier, du soja et du ricin. Chez *C. cinereum*, cette analyse montre de fortes similitudes avec des gènes identifiés chez le soja, la luzerne tronquée, la vigne et le pois chiche. Les potentielles fonctions moléculaires identifiées sont très similaires entre les différentes banques (Figure 38, p. 134). Parmi ces différentes fonctions identifiées, un peu moins de la moitié des séquences assemblées est associé à des activités catalytiques (Figure 38, p. 134). Certaines fonctions ont été uniquement identifiées chez la rue et *C. cinereum.* Il s'agit des fonctions 'antioxydant activity', 'protein binding transcription factor activity', 'nutrient reservoir activity' et 'metallochaperone activity' (Figure 38, p. 134). Ces résultats peuvent s'expliquer par le fait que la qualité de la banque de panais est inférieure à celle des deux autres banques. De la même manière, les processus biologiques dans lesquels sont impliqués les différents gènes des banques sont très similaires entre les trois plantes (Figure 39, p. 135). Un peu moins de la moitié des séquences assemblées en All contigs est identifié comme intervenant dans des processus métaboliques (Figure 39, p. 135). Certains processus ne sont représentés que chez *C. cinereum* (Figure 39, p. 135). Il s'agit des processus 'locomotion', 'growth' et 'immune system process'. La proportion de gènes impliqués dans des processus biologiques incluant un seul organisme ('single organism process', GO:0044699) est moins importante chez le panais que chez la rue et *C. cinereum* (Figure 39, p. 135). Dans la suite de ce travail, notre attention va se focaliser sur les catégories de séquences associées à des activités catalytiques et impliquées dans des processus métaboliques qui renferment les séquences codant pour les enzymes de la famille des cytochromes P450.

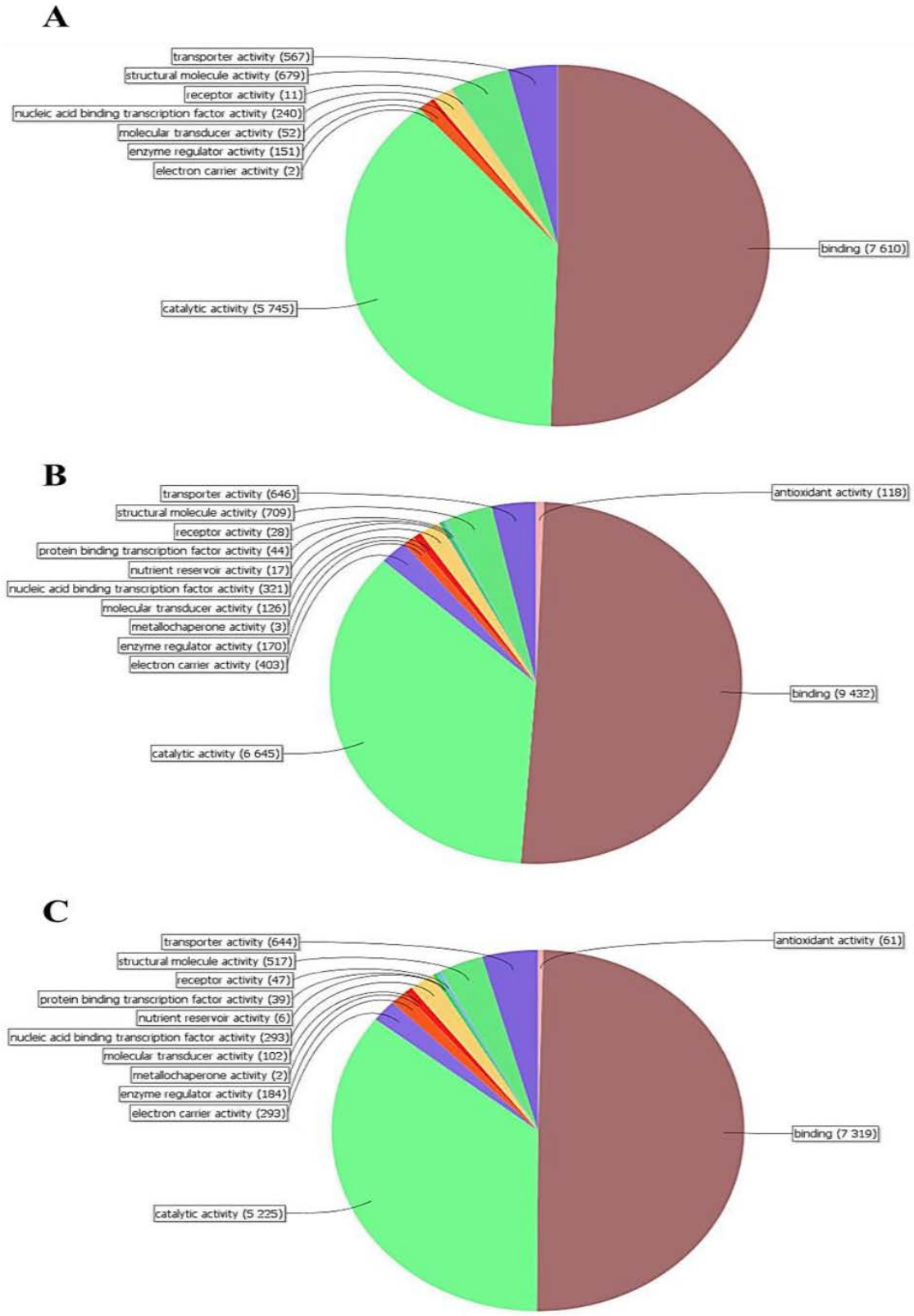


Figure 38 : Analyse des fonctions moléculaires associées aux différents contigs de *P. sativa* (A), *R. graveolens* (B) et *C. cinereum* (C).

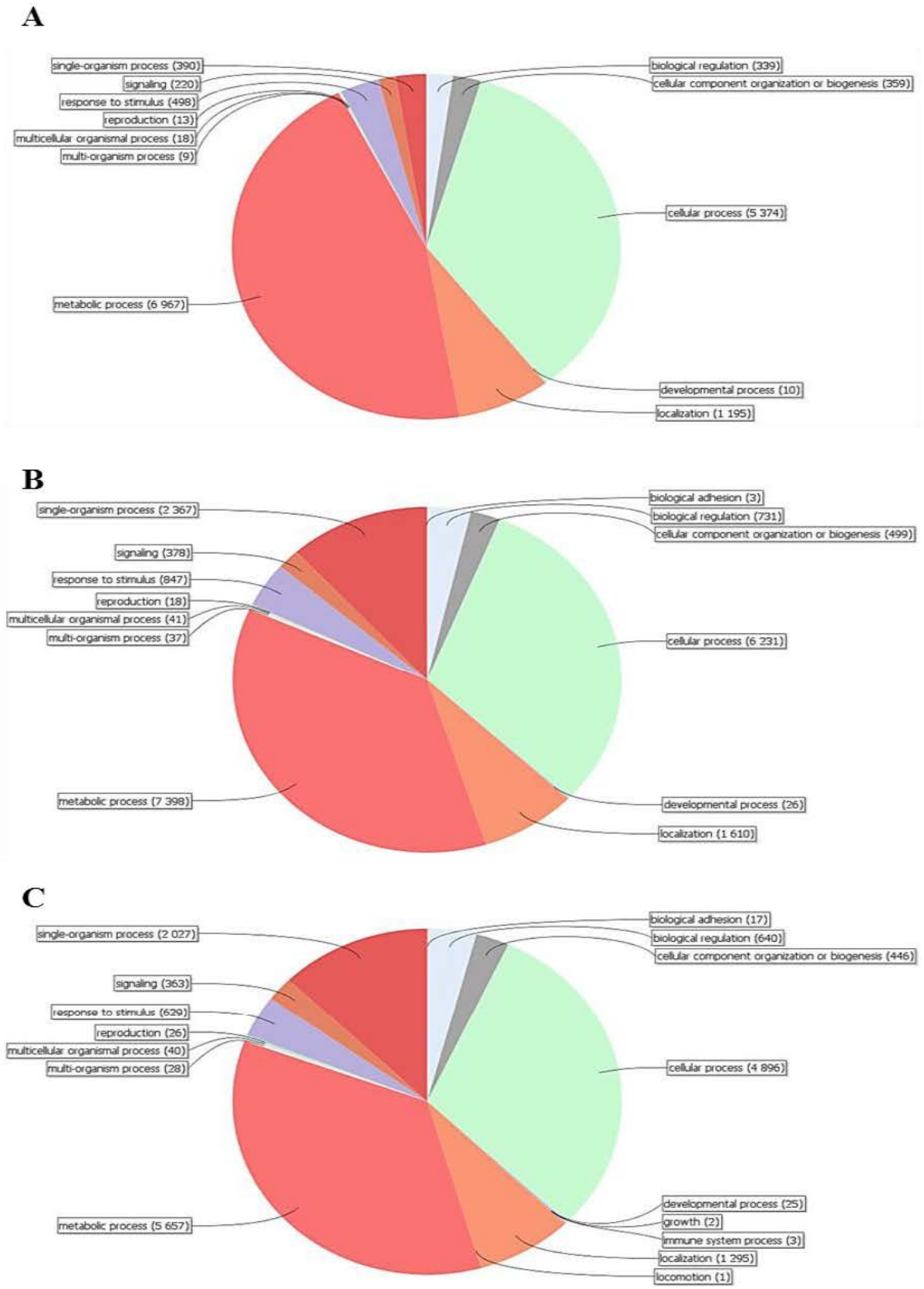


Figure 39 : Analyse des processus biologiques associés aux différents contigs de *P. sativa* (A), *R. graveolens* (B) et *C. cinereum* (C).

1.1.3.2 Identification des séquences nucléotidiques codant pour des cytochromes P450

L'identification des séquences codant pour des cytochromes P450 dans les trois banques d'ADNc est réalisée à partir des termes d'InterPro (IPR) et des annotations des champs de Gene Ontology (GO). Cette extraction réalisée a permis de mettre en évidence un total de 336 contigs chez *P. sativa*, 279 chez *R. graveolens* et 221 chez *C. cinereum* présentant des homologies de séquence avec cette famille multienzymatique. Ces séquences présentent un minimum d'identité de 44,4 - 50,2 %, un maximum d'identité de 97,6 - 98,4 % et une moyenne de 79,8 - 82,9 % d'identité avec la séquence la plus proche contenue dans la banque publique NCBI nr (Tableau 9, p. 137). Parmi toutes les séquences codant pour des P450s, la majorité est identifiée avec un score très élevé ($E < 10^{-35}$) dans les trois banques (Tableau 9, p. 137). C'est chez *P. sativa* que l'on retrouve le plus de séquences identifiées codant pour des P450s avec un score moyen ($10^{-35} < E < 10^{-10}$) ce qui représente 27 % du total des séquences, contre seulement 19 % et 11 % respectivement chez *R. graveolens* et *C. cinereum* (Tableau 9, p. 137). Cette première analyse démontre bien l'importance de cette famille d'enzymes dans le métabolisme des végétaux supérieurs, et ce visiblement, quelle que soit l'espèce étudiée. Enfin, dans ces trois banques d'ADNc, très peu de séquences sont identifiées avec un score faible ($E < 10^{-3}$) (Tableau 9, p. 137). La taille moyenne de ces séquences est de 619 nucléotides (nt) chez *P. sativa*, 760 nt chez *R. graveolens* et 1042 nt chez *C. cinereum* (Tableau 9, p. 137). Chez le panais et la rue, les séquences les plus courtes ne mesurent que 155 et 165 nt alors que la séquence la plus courte chez *C. cinereum* est de 383 nt. De plus, la longueur moyenne des contigs identifiés est de 619 nt chez le panais, de 760 nt chez la rue et de 1042 nt chez *C. cinereum* (Tableau 9, p. 137). Ces résultats indiquent que la qualité de la banque de *C. cinereum* est supérieure à celle de *R. graveolens* qui est elle-même de meilleure qualité que celle de *P. sativa*. Les quantités d'ARN fournies au prestataire de service ne sont pas identiques pour les trois banques. Il semble que la qualité des banques soit liée à la quantité d'ARN fournie, mais étant donné que ces banques ont été réalisées à des temps diffrents, il est également possible que ces résultats soient dus à l'amélioration de la technologie. La fréquence de ces séquences de petites tailles reflète l'abondance de chaque enzyme dans les tissus analysés. En effet, plus un ARN messager est représenté dans la banque et plus de séquences correspondantes sont disponibles ce qui impacte directement l'étape d'assemblage. Les séquences les plus longues assemblées codants pour des P450s mesurent entre 1999 nt et 2222 nt et devraient contenir la totalité de la séquence codante (Tableau 9, p. 137).

		P. sativa	*R. graveolens*	*C. cinereum*
Pourcentage d'identité (%)	Maximum	98,4	97,6	97,8
	Minimum	44,4	46,2	50,2
	Moyenne	79,8	80,1	82,9
E-value	$E < 10^{-35}$	223	216	192
	$10^{-35} < E < 10^{-10}$	91	55	25
	$E < 10^{-3}$	17	7	2
Longueur (en nt)	Maximum	2069	1999	2222
	Minimum	155	165	383
	Moyenne	619	760	1042

Tableau 9 : Caractéristiques des séquences identifiées à partir des trois banques d'ADNc codant pour des P450s.

Parmi toutes les séquences identifiées comme codant pour des P450s, la majorité d'entre elles ont pu obtenir des annotations de définition fonctionnelle avec Gene Ontology (GO) ce qui signifie que ces gènes sont associés à des propriétés comparables pour des gènes identifiés chez d'autres espèces (végétales dans notre cas). Seules 5 séquences chez le panais, 1 chez la rue et 2 chez *C. cinereum* n'ont pas reçu d'annotation GO. Cette analyse permet d'avoir une image plus ou moins fidèle du nombre et des différentes fonctions associées aux P450s présents dans les différentes banques. En effet, il est nécessaire de préciser que les séquences courtes disponibles dans les banques peuvent provenir de différentes zones de la séquence codante (extrémités 5'/3' ou région centrale) impliquant de fait que certaines de ces séquences distinctes peuvent appartenir au même P450. Cette analyse permet donc de relativiser le nombre de P450s présent dans chacune des 3 banques. Ainsi, au total, 174 champs GO différents sont identifiés et c'est chez *P. sativa* qu'on en retrouve le plus avec 96 annotations différentes (Annexe 3, p. 290-293), alors que 336 contigs avaient été identifiés chez cette plante. Chez *C. cinereum*, 75 différentes annotations (contre 221 séquences correspondant à des P450s identifiées) sont utilisées alors qu'il y en a seulement 68 chez la rue (contre 229 séquences identifiées) (Annexe 3, p. 290-293).

Dans certains cas, cette approche a permis d'obtenir des annotations très précises concernant la fonction de la protéine correspondante (par exemple : cinnamate 4-hydroxylase) et a permis de classer un certain nombre de séquences dans une sous-famille de P450s (Annexe 3, p. 290-293). Cette analyse reste néanmoins très partielle et relativement peu informative puisqu'une bonne partie de ces séquences a obtenu des annotations très généralistes de type 'cytochrome', 'cytochrome p450', 'cytochrome p450 like_tbp', 'cytochrome p450 like protein', 'cytochrome p450 monooxygenase', 'cytochrome p450 a',

ou encore 'cytochrome p450 chloroplastic' (Annexe 3, p. 290-293). Une analyse comparative des différentes banques permet néanmoins de constater que certaines annotations GO sont présentes pour les P450s présents dans les 3 banques tandis que d'autres ne sont mises en évidence que chez une ou deux plantes. A titre d'exemples, l'annotation 'cytochrome p450 82c4' est utilisée dans les trois banques, alors que l'annotation 'aba 8-oxidase' n'est utilisée que chez *P. sativa*. Cette information est importante dans la mesure où la synthèse des furanocoumarines n'a été décrite que chez 4 familles de plantes supérieures phylogénétiquement très distinctes. Il est donc fortement possible que l'apparition de ces mêmes molécules soit le résultat d'une évolution convergente faisant intervenir des enzymes différentes. La mise en évidence de P450s présentant des ontologies différentes permet donc d'identifier des pistes de recherche pour étudier de nouveaux gènes candidats.

La majorité des séquences identifiées comme correspondant à des P450s se sont vues assignées une ou plusieurs annotations de domaines fonctionnels avec les termes InterPro (IPR). L'algorithme InterProScan recherche dans un jeu de séquences des signatures protéiques en les comparant à plusieurs bases de données de motifs ou de domaines protéiques telles que PROSITE, PRINTS, Pfam, ou encore TIGRFAMs. L'analyse de ces domaines conservés qui peuvent correspondre à des domaines fonctionnels ou à des sites actifs peut permettre d'extrapoler la fonction potentielle des protéines. Dans les trois banques d'ADNc, la majorité de ces séquences est identifiée comme 'cytochrome P450' (Tableau 10, p. 139) et certains de ces contigs sont reconnus grâce à des domaines conservés (IPR017972) (Tableau 10, p. 139). Cette approche a permis de mettre en évidence que les P450s identifiés dans les trois banques appartiennent principalement à la classe E qui regroupe les P450s ayant un seul partenaire redox et qui est le groupe le plus riche (Tableau 10, p. 139). Une majeure partie classée dans le groupe I qui regroupe CYP1, 2, 17, 21 et 71 (Tableau 10, p. 139). Une minorité est classée dans le groupe IV qui inclut CYP7, 8 et 51 (Tableau 10, p. 139). Quelques rares membres sont identifiés comme appartenant au groupe II qui regroupe les familles CYP3, 4, 5, 6, 52, 53, 56, 102 et 110 (Tableau 10, p. 139). Seules 2 séquences chez *P. sativa* et *C. cinereum* appartiennent à la classe B qui contient des P450s ayant deux partenaires redox, contre 5 chez la rue (Tableau 10, p. 139) et correspondraient à des P450s mitochondriaux. Enfin, 36 séquences chez le panais, 50 séquences chez la rue et 41 séquences chez *C. cinereum* n'ont pas obtenu de terme IPR ce qui est vraisemblablement lié au fait que ces séquences sont partielles et ne renferment pas certaines des séquences signatures conservées. De plus, certaines séquences des banques de panais et de rue sont annotées 'allene oxide synthase' ou 'cytochrome p450 allene oxide synthase', or ces P450s ne nécessitent pas de partenaire redox, car les électrons sont directement apportés par le substrat ce qui peut également expliquer que ces séquences n'ont pas obtenu de terme IPR.

Terme IPR	Désignation terme IPR	*P. sativa*	*R. graveolens*	*C. cinereum*
IPR001128	Cytochrome P450	260	203	163
IPR002397	Cytochrome P450, B-class	2	5	2
IPR002401	Cytochrome P450, E-class group I	117	116	98
IPR002402	Cytochrome P450, E-class group II	2	3	4
IPR002403	Cytochrome P450, E-class group IV	14	12	11
IPR017972	Cytochrome P450, conserved site	112	74	75

Tableau 10 : Annotations des termes IPR obtenus pour les cytochromes P450 identifiés à partir des trois banques d'ADNc.

Comme cela a été déjà discuté précédemment, une analyse plus fine de chacune de ces séquences montre que parmi toutes les séquences protéiques correspondant à des P450s, peu d'entre elles sont prédites complètes, c'est-à-dire qu'elles vont du codon initiateur ATG au codon stop qui délimitent le cadre de lecture ouvert (ORF pour Open Reading Frame). Sept séquences sont identifiées chez *P. sativa* (Annexe 4, p. 296), 26 chez *R. graveolens* (Annexe 5, p. 297) et 50 chez *C. cinereum* (Annexe 6, p. 298-297).

Le séquençage avec la technologie 454 est reconnu comme générant un certain nombre d'erreur (1 à 2 %) (Schröder *et al*., 2009). Dans notre cas, la traduction de ces séquences nucléotidiques dans les six cadres de lecture montre que parmi les séquences prédites complètes, la majorité d'entre elles ne possèdent pas de changement de cadre de lecture indiquant par-là que les erreurs liées à l'approche de séquençage sont *a priori* tolérables et renforcent la solidité des assemblages effectués. En effet, le « contigage » se base sur une complémentarité des séquences qui se chevauchent. Dans certains cas, une tolérance de quelques nucléotides différents est acceptée. Dans la mesure où des P450s de la même famille ont des séquences très proches, cet assemblage peut mener à des constructions *in silico* de gènes chimères. Cet aspect doit être pris en compte lorsqu'est venu le moment de cloner une séquence codante. En effet, étant donné que le prestataire de service n'a pas fourni de précisions sur la réalisation de l'assemblage, il est possible que certains gènes ne puissent être amplifiés car ils ne sont que le résultat d'un assemblage erroné de deux (ou plus) morceaux de gènes appartenant à des enzymes différentes. Dans le cas de séquences présentant des changements de cadre de lecture, l'amplification des fragments d'ADN suivi d'un séquençage par la technique classique de Sanger permettra de déterminer si ces séquences correspondent à un pseudogène ou si, comme évoqué ci-dessus, ils proviennent d'erreur de séquençage lors de la construction de la banque. Six séquences peptidiques complètes sont identifiées chez le panais (Annexe 7, p. 300), 18 chez la rue (Annexe 8, p. 301-301) et 31 chez *C. cinereum* (Annexe

9, p. 304-306). A noter que certaines de ces séquences possèdent plusieurs méthionines (indiquées en rouge, Annexe 7, Annexe 8 et Annexe 9, p. 300-306) en début de séquence.

Le résultat de cette première analyse nous amène donc à constater que le nombre de séquences complètes directement exploitable est relativement faible. En effet, seuls pour ces quelques cas il est possible de générer des amorces et d'amplifier les séquences codantes en utilisant une approche de PCR simple sur des ADNc produits à partir des différentes plantes. Pour les autres cas, il sera nécessaire de réaliser des expérimentations complémentaires pour compléter les séquences codantes (approches par 5' ou 3' RACE-PCR, TAIL PCR). Les banques construites ne sont donc pas optimales et auraient nécessité une couverture plus importante pour avoir une meilleure qualité.

1.1.3.3 Classification des séquences codant pour des P450s

La seconde étape de ce travail va maintenant consister à réaliser un tri au sein de toutes les séquences correspondant à des P450s de sorte à identifier lesquelles codent pour des enzymes appartenant aux familles ciblées. Pour réaliser ce travail, l'ensemble des séquences nucléotidiques identifiées codant pour des cytochromes P450 dans les trois banques a été comparé à une banque de données fournie par le Dr. David Nelson. Cette banque de données confidentielle correspond à un fichier informatique contenant 476 séquences peptidiques de P450s appartenant à des sous-familles différentes. Cette analyse est plus précise et plus rapide qu'une analyse générale sur une banque de données publiques dans la mesure où 1) elle est restreinte à des séquences correspondant à des P450s et 2) elle renferme un représentant de toutes les familles de P450s (y compris des séquences qui ne sont actuellement pas encore rendues disponibles pour la communauté scientifique). L'analyse réalisée grâce à l'algorithme blast x passe d'une part par la traduction de toutes les séquences codantes dans les 6 phases de lectures possible et d'autre part par une comparaison avec les séquences de protéines disponibles. Les résultats donnent une information relativement précise quant à l'affiliation d'une enzyme à une famille voir une sous-famille de P450s mais nécessitent dans tous les cas une validation finale par le comité international de nomenclature des P450s.

La recherche par similarité de séquences a permis d'assigner une sous-famille à la majorité des séquences identifiées comme codant pour des P450s, seules 6 séquences chez le panais, 2 chez la rue et 4 chez *C. cinereum* n'ont pu être identifiées plus précisément. Chez les trois plantes, les résultats indiquent sans surprise que le clan le plus important est le clan 71 avec 217 membres identifiés chez le panais, 157 chez la rue et 138 chez *C. cinereum* (Annexe 10, p. 309). Le clan 71 est représenté par 19 familles chez *P. sativa* et par 16 familles chez *R. graveolens* et *C. cinereum* (Annexe 10, p. 309). Chez

les trois plantes, la famille la plus importante de ce clan est CYP71 avec 66 séquences chez *P. sativa* (dont 18 appartenant à CYP71AJ, 10 à CYP71AZ et 4 à CYP71D), 55 séquences chez *R. graveolens* (dont 3 appartenant à CYP71D) et 41 chez *C. cinereum* (dont 4 appartenant à CYP71D) (Annexe 10, p. 309). **Il semble qu'aucune des séquences candidates de la banque de rue et de *C. cinereum* n'appartienne aux sous-familles CYP71AJ, CYP71AZ et CYP71A**. Le clan 71 contient également de nombreux membres des familles CYP73, 75, 76, 79, 81, 82, 83, 84, 89, 92 et 706 qui sont présentes en proportions différentes dans les trois banques (Annexe 10, p. 309). Les autres familles qui composent également le clan 71 sont les familles CYP77, 78, 80, 98, 701, 712 et 736 chez *P. sativa*, CYP98, 701, 703 et 712 chez *R. graveolens* et CYP 77, 78, 93 et 736 chez *C. cinereum* (Annexe 10, p. 309).

Chez les trois plantes, trois autres clans sont relativement bien représentés. Il s'agit des clans 72, 85 et 86 qui sont distribués de manière différente entre les trois plantes (Annexe 11, p. 310). Chez *R. graveolens* et *C. cinereum*, le clan 85 est le deuxième le plus important suivi par le clan 72 et le clan 86, alors que chez le panais le deuxième clan le plus représenté est le 86, suivi du 72 et du 85 (Annexe 11, p. 310). Trois autres clans moins bien représentés sont présents chez ces trois plantes, il s'agit des clans 51, 74 et 97 (Annexe 11, p. 310). Enfin certains clans sont spécifiques de la plante étudiée. A titre d'exemples, le clan 710 n'est présent que chez le panais et *C. cinereum*, alors que le clan 711 est spécifique du panais et le clan 85 n'est retrouvé que chez *C. cinereum* (Annexe 11, p. 310). Enfin, d'autres familles qui n'appartiennent pas à un clan sont identifiées dans chacune des banques. Ces familles sont à chaque fois uniquement représentées par un seul membre. Il s'agit des familles CYP741, 749 et 771 chez le panais, de la famille CYP749 chez *R. graveolens* et des familles CYP745, 746, 758 et 779 chez *C. cinereum* (Annexe 11, p. 310).

Ces analyses faites sur l'ensemble des fragments d'ADN assemblés des 3 banques sont néanmoins à prendre avec précaution. En effet, l'attribution d'une sous-famille par similarité de séquences pour les séquences prédites complètes ne pose pas de problème, mais pour des séquences incomplètes il est possible que celles-ci soient classées dans une autre sous-famille que celle à laquelle elles appartiennent en réalité. Ainsi, si on restreint cette analyse aux séquences complètes et ne possédant pas de modification du cadre de lecture on apprend que (Annexe 12, p. 311-310) :

- Les six séquences complètes isolées chez le panais appartiennent au clan 71 (sous-familles CYP71AJ, 71AZ, 71D et CYP73A) et au clan 86 (CYP86A).
- Chez *R. graveolens*, les 18 séquences complètes appartiennent à 4 clans différents : le clan 71 (sous-familles CYP71BE, 81Q, 82H, 83F, 92A, 701A et 706G), le clan 72 (CYP72A), le clan 90 (CYP90A) et le clan 96 (CYP96K).

- Chez *C. cinereum*, les 31 séquences complètes représentent 6 clans différents. Le clan 71 est largement représenté avec les sous-familles CYP71AN, BE, BF, CYP73A, CYP75B, CYP76B, E et X, CYP81E, CYP82A, CYP83E, CYP89A, CYP92A, CYP93B et C, CYP706J. Le clan 85 est le deuxième caractérisé par les sous-familles CYP85A, CYP88A et CYP716D. Le clan 86 est quant à lui représenté par des membres des sous-familles CYP86A et CYP94A. Enfin, les clans 51, 72, 74 et 07 ne sont représentés que par une seule séquence appartenant respectivement à CYP51G, CYP72A, CYP74B et CYP97A.

Cette approche de recherche par similarité de séquences classe de nombreuses séquences dans des familles voire des sous-familles différentes de celles obtenues avec la recherche des champs GO. Ces observations ne sont pas surprenantes dans la mesure où de nombreuses séquences disponibles dans la banque de données privée de David Nelson n'ont pas encore été rendues publiques. A titre d'exemple, chez le panais, l'annotation avec les termes GO plaçait les contigs P_sat_rep_c588 et c729 dans la sous-famille des CYP83B, alors que l'identification par similarité de séquences les placent dans la sous-famille des CYP71AZ (Annexe 12, p. 311-310). Ces deux séquences ont été soumises au Dr. Nelson afin de les classer dans la nomenclature officielle et correspondent aujourd'hui respectivement à CYP71AZ4 et CYP71AZ3. Dix autres séquences complètes isolées chez la rue sont également classées dans la nomenclature officielle : R_gra_c149 (CYP71BE20), c578 (CYP92A53), c643 (CYP71BE21), c1542 (CYP72A279), c1672 (CYP82D46), c1737 (CYP71AT56), c2181 (CYP71AT57), c3135 (CYP92A52), c3823 (CYP72A278) et c7149 (CYP72A277).

Enfin plusieurs raisons peuvent expliquer que certaines séquences candidates n'ont pu être classées avec l'approche par similarité de séquences : ces séquences sont soit trop courtes, soit comportent des séquences correspondant aux régions 5' et 3'-UTR, soit contiennent un ou plusieurs changements de cadre de lecture. L'utilisation de la suite logicielle blast2GO n'a pas non plus permis d'obtenir plus d'informations sur ces séquences vu qu'elles sont annotées avec les champs GO 'cytochrome p450' ou ' cytochrome p450 like_tbp'.

1.1.3.4 Identification des séquences codant pour des cytochromes P450 NADPH réductase

Les cytochromes P450 sont des enzymes membranaires qui, chez les végétaux du moins, sont étroitement associées à des NADPH cytochromes P450 réductases (CPR). Ces enzymes permettent le transfert d'électrons vers les P450s et sont donc essentielles à leur bon fonctionnement. Les données disponibles dans la littérature indiquent que, contrairement aux cytochromes P450, ces enzymes sont

très peu variables. Ainsi il est généralement admis que pour une plante donnée il n'y a qu'une poignée de CPRs qui fonctionnent de manière unilatérale avec l'ensemble des P450s présents ce qui sous-entend que cette enzyme doit être exprimée en grande quantité pour subvenir aux besoins de tous les P450s présents dans une plante. L'interaction entre le P450 et la CPR est un élément essentiel au bon fonctionnement des enzymes. Certains travaux démontrent que dans le cadre d'expression hétérologue, les P450s peuvent être plus actifs lorsqu'ils sont associés avec une CPR issue de son organisme d'origine. Pour disposer d'un dispositif de caractérisation fonctionnelle optimisé associant un P450 et une CPR d'une plante donnée, nous avons effectué une recherche de gène codant pour les CPRs au sein des trois banques qui ont été générées au laboratoire. La recherche des termes IPR et des annotations des champs GO correspondant à des CPRs a permis de mettre en évidence un total de 17 contigs chez *P. sativa*, 13 chez *R. graveolens* et 9 chez *C. cinereum.*

		P. sativa	*R. graveolens*	*C. cinereum*
Pourcentage d'identité (%)	Maximum	95,8	98,8	96,2
	Minimum	72,8	81,4	57,6
	Moyenne	85,5	89,3	83,4
E-value	$E < 10^{-35}$	13	10	9
	$10^{-35} < E < 10^{-10}$	4	3	0
Longueur (en nt)	Maximum	2701	2421	2548
	Minimum	327	442	569
	Moyenne	654	886	1190

Tableau 11 : Caractéristiques des séquences identifiées à partir des trois banques d'ADNc codant pour des CPRs.

Annotations GO	*P.sativa*	*R. graveolens*	*C. cinereum*
nadph cytochrome p450 reductase	15	5	3
cytochrome p450 reductase	1		
nadph-ferrihemoprotein reductase	1	1	
nadph-cytochrome p450 oxydoreductase		4	
nadph--cytochrome p450 reductase-like		3	4
nadph:p450 redutase			1
nadph-dependent cytochrome p450 reductase			1

Tableau 12 : Annotations des champs de Gene Ontology obtenus pour les séquences correspondant à des CPRs identifiées à partir des trois banques d'ADNc.

Ces séquences présentent un minimum d'identité de 57,6 – 81,4 %, un maximum d'identité de 95,8 - 98,8 % et une moyenne de 83,4 – 89,3 % d'identité avec la séquence la plus proche contenue dans la banque publique NCBI nr (Tableau 11, p. 143). Parmi toutes les séquences codant pour des CPRs, la majorité voire la totalité d'entre elles sont identifiées avec un score très élevé ($E < 10^{-35}$) démontrant ainsi une conservation importante de ces protéines entre les différentes espèces végétales (Tableau 11, p. 143). Chez *P. sativa* et la rue, on retrouve quelques séquences identifiées avec un score moyen ($10^{-35} < E < 10^{-10}$) (Tableau 11, p. 143). La taille moyenne de ces séquences est de 654 nucléotides (nt) chez *P. sativa*, 886 nt chez *R. graveolens* et 1190 nt chez *C. cinereum* (Tableau 11, p. 143). Chez le panais et la rue, les séquences les plus courtes ne mesurent que 327 et 442 nt alors que la séquence la plus courte chez *C. cinereum* est de 569 nt (Tableau 11, p. 143). Les séquences les plus longues codants pour des CPRs mesurent entre 2421 nt et 2701 nt et devraient contenir la totalité de la séquence codante, car la longueur moyenne d'un gène codant pour une CPR est d'environ 2100 nt (Tableau 11, p. 143).

Toutes les séquences identifiées comme codant pour des CPRs ont obtenu des annotations de champs de Gene Ontology. Au total, 7 champs GO différents sont identifiés dont 3 chez *P. sativa*, 4 chez la rue et chez *C. cinereum* (Tableau 12, p. 143).

Terme IPR	Désignation terme IPR	*P. sativa*	*R. graveolens*	*C. cinereum*
IPR001094	flavodoxin	4	3	2
IPR001433	Oxidoreductase FAD/NAD(P)-binding	3	4	2
IPR001709	Flavoprotein pyridine nucleotide cytochrome reductase	3	4	3
IPR003097	FAD-binding, type 1	6	3	4
IPR008254	Flavodoxin/nitric oxide synthase	4	5	2
IPR017927	Ferredoxin reductase-type FAD-binding domain		2	1
IPR017938	Riboflavin synthase-like beta-barrel	5	3	3
IPR023173	NADPH-cytochrome p450 reductase, FAD-binding, alpha-helical domain-3	2	3	2

Tableau 13 : Annotations des termes IPR obtenus pour les séquences codant pour des CPRs identifiées à partir des trois banques d'ADNc.

La majorité des séquences identifiées comme correspondant à des CPRs se sont vues assignées un ou des termes IPR (Tableau 13, p. 144). La recherche des termes IPR001094 et IPR023173 a permis de mettre en évidence un certain nombre de séquences (Tableau 13, p. 144). Par contre, aucune séquence

n'est identifiée avec les termes IPR023206 ('Bifunctional cytochrome P450/NADPH--cytochrome P450 reductase') ou IPR023208 ('NADPH-cytochrome P450 reductase') utilisés pour la recherche de séquences codant pour des CPRs. Par contre, certaines de ces séquences ont également reçu des termes IPR qui n'avaient pas été utilisés pour la recherche de gènes candidats (IPR001433, IPR001709, IPR003097, IPR008254, IPR017938, IPR017927) (Tableau 13, p. 144).

Parmi toutes les séquences identifiées comme des CPRs, une seule est prédite complète dans chacune des trois banques d'ADNc. Chez le panais, il s'agit du contig P_sat_rep_c249. Malheureusement, cette séquence comporte un changement de cadre de lecture : le début de la séquence codante est dans le cadre de lecture 3, puis au niveau des nucléotides en position 1125 à 1129 la lecture s'effectue dans le cadre 1. De plus, deux méthionines (indiquées en rouge dans l'Annexe 13, p. 313) en début de séquence peuvent correspondre au codon initiateur, la longueur de la séquence protéique peut donc être de 696 ou 709 aa. La séquence peptidique de 709 aa est représentée dans l'Annexe 13 (p. 313). L'acide aminé en position 316 au niveau du changement de cadre de lecture est indiqué en vert dans l'Annexe 13 (p. 313) et correspond soit à une phénylalanine (F, cadre de lecture 3) soit à une leucine (L, cadre de lecture 1). Cette séquence de 709 aa présente respectivement 60,8 % et 72,1 % d'identité avec les séquences des CPRs ATR1 et ATR2 d'*A. thaliana*. Néanmoins, il reste nécessaire de confirmer par des approches expérimentales la nature exacte de cette séquence codante.

Chez la rue, il s'agit du contig R_gra_c3549 composé de 2421 nt. Deux méthionines (indiquées en rouge dans l'Annexe 13, p. 313) sont présentes en début de séquence et peuvent correspondre au codon intiateur. La séquence protéique peut donc être de 703 ou de 713 acides aminés. La séquence peptidique de 713 aa est représentée dans l'Annexe 13 (p. 313). Elle présente respectivement 62 %, 72 % et 64,5 % d'identité avec les séquences des CPRs ATR1 et ATR2 d'*A. thaliana* et celle identifiée chez *P. sativa*.

Enfin chez *C. cinereum* le contig P_cin_c6685 de 2548 nt est prédit complet. Deux méthionines (indiquées en rouge dans l'Annexe 13, p. 313) sont présentes en début de séquence et peuvent correspondre au codon intiateur. La séquence protéique peut donc être de 690, mais au vu d'alignement de séquences il est plus probable qu'elle soit de 704 acides aminés. La séquence peptidique de 704 aa est représentée dans l'Annexe 13 (p. 313). Elle présente respectivement 61,7 %, 71,3 %, 74 % et 76,3 % d'identité avec les séquences des CPRs ATR1 et ATR2 d'*A. thaliana* et celles identifiées chez *P. sativa* et chez *R. graveolens*.

1.2 Discussion et conclusion

Les plantes produisant des furanocoumarines ne sont pas des plantes modèles pour la communauté scientifique et de fait, elles ne font pas l'objet de nombreuses études génomiques. En effet, les ressources moléculaires disponibles n'étaient pas très nombreuses ce qui constituait un handicap majeur pour réaliser la caractérisation fine de la voie de biosynthèse de ces molécules. Pour pallier ce problème nous avons choisi de nous appuyer sur les percées technologiques dans le domaine du séquençage haut débit pour générer un grand nombre de séquences correspondant aux transcriptomes de différentes plantes. Les furanocoumarines ont été décrites dans 4 familles botaniques différentes : les Rutacées, les Moracées, les Apiacées et les Fabacées. Dans le cadre de ce travail, nous avons réalisé une analyse transcriptomique sur des plantules issues de 3 de ces familles. Notre choix s'est porté sur *P. sativa*, *R. graveolens* et *C. cinereum*. Chacune de ces plantes dispose d'arguments en sa faveur, argument qui ont déjà été évoqués précédemment. Si l'objectif majeur de ce travail était de générer des données moléculaires pour chacune de ces plantes, un objectif sous-jacent était de réaliser une analyse comparative qualitative de ces différents individus. En effet, ces plantes sont phylogénétiquement très différentes mais sont en mesure de synthétiser les mêmes familles de molécules. Il est fort probable que ces plantes fassent appel à des enzymes différentes pour synthétiser les différents intermédiaires de la synthèse des furanocoumarines. Cette hypothèse semble maintenant renforcée par l'analyse des différents transcriptomes et notamment par la mise en évidence de l'absence notamment de membres de la famille CYP71AJ chez *R. graveolens* et chez *C. cinereum*. Or, cette sous-famille d'enzyme a été caractérisée comme étant impliquée dans la synthèse de psoralène et d'angélicine chez *P. sativa*. Ces résultats confirment des résultats expérimentaux non publiés au laboratoire par Romain Larbat, résultats qui indiquent qu'aucun gène correspondant à un P450 CYP71AJ n'a pu être amplifié par une approche de PCR sur les ARNs extraits de ces 2 plantes.

L'analyse du transcriptome de ces différentes plantes a permis d'identifier un grand nombre de nouveaux P450s candidats. Les séquences accessibles dans les banques générées sont de qualité très hétérogène dans le sens ou certaines correspondent à des séquences codantes complètes (contenant un ATG initiateur et un codon stop) alors que d'autres ne sont que des séquences partielles. Ainsi, si de nombreuses nouvelles données sont maintenant accessibles, elles ne peuvent malheureusement pas toutes être exploitées en l'état. Les séquences codantes complètes vont pouvoir très facilement être clonées et les protéines correspondantes exprimées dans des systèmes d'expression hétérologue. Pour les séquences partielles, un travail supplémentaire est nécessaire et va consister à rechercher les séquences manquantes par des approches moléculaires telles que la 5' ou 3'RACE-PCR ou encore la TAIL-PCR. Une autre possibilité pour identifier les séquences complètes peut également consister à

générer des banques d'ADN génomique, projet qui a été initié au laboratoire au travers d'un projet porté par le Dr. Sandro Roselli sur *P. sativa*.

Au travers d'un travail préliminaire basé sur une étude bibliographique, sur une analyse de mécanismes réactionnels et sur des données expérimentales disponibles au laboratoire, nous avions pointé un certain nombre de familles de cytochromes P450 intéressantes : CYP71, CYP72, CYP76 et CYP82. Les analyses réalisées sur les 3 banques générées au laboratoire ont permis de mettre en évidence que ces différentes familles de gènes étaient représentées chez chacune des plantes modèles étudiées. Ainsi, on trouve 66 séquences codant des CYP71 chez *P. sativa*, 55 chez *R. graveolens* et 41 chez *C. cinereum* (Annexe 10, p. 309). Ce nombre très élevé de gènes appartenant à cette famille n'est vraisemblablement pas le reflet de la réalité. Même si cette famille de P450s est la plus grande décrite chez les plantes, il est vraisemblable que parmi ces 162 séquences, une grande partie corresponde à des fragments d'une même séquence codante réduisant ainsi le nombre de gènes candidats. Ce constat est le même pour les 3 autres familles ciblées dans l'étude préliminaire. Ainsi CYP82 dispose d'une vingtaine de représentants chez *P. sativa* (21), chez *R. graveolens* (21) et *C. cinereum* (20), et CYP 72 est représenté par 24 contigs chez le panais, 31 chez la rue et 21 chez *C. cinereum* (Annexe 11, p. 310). De plus, CYP76 dispose de nombreux représentants chez *P. sativa* (23) et chez *C. cinereum* (15), alors que l'on en dénombre que 5 chez *R. graveolen*s (Annexe 10, p. 309). L'analyse du transcriptome de ces plantes ne peut de fait pas être réalisée de manière exhaustive. En effet, si l'affectation d'une séquence partielle d'un P450 peut être effectuée avec des chances raisonnables de ne pas se tromper de famille d'enzyme, il est très difficile de valider à ce stade leur appartenance à une sous-famille. Or, il a été démontré que, dans le cas de la famille CYP71, les fonctions étaient extrêmement diversifiées en fonction de leur appartenance à une famille plutôt qu'à une autre. Par exemple, les CYP71AV et CYP71BL ont été démontrés comme intervenant dans la synthèse de costunolides qui sont des molécules appartenant aux terpénoïdes (Eljounaidi *et al*., 2014). Un autre exemple concerne la synthèse de la camalexine, une phytoalexine décrite chez *Arabidopsis* et qui nécessite la présence de CYP71A et CYP71B (Nafisi *et al*., 2007; Zhou *et al*., 1999). Afin d'avoir des informations plus robustes, il sera donc nécessaire, le cas échéant, de réaliser de nouvelles banques d'ADNc avec une couverture plus importante.

Si seule une partie minime des données générées par cette analyse transcriptomique est exploitée dans la suite de ce document, ces banques constituent une source de données incontournable et d'une valeur inestimable pour poursuivre la compréhension de cette voie de biosynthèse. De nouveaux projets ont déjà vu le jour, voire aboutit au laboratoire. Ainsi, une partie des travaux de thèse de Guilhem Vialart se sont appuyés sur l'identification de gènes codants pour des dioxygénases α cétoglutarate dépendantes dans la banque de *R. graveolens*. Par ailleurs des travaux ont été initiés dans le cadre de la thèse d'Audray Dugrand pour identifier des P450s et des prényltransférases impliqués dans la synthèse des

furanocoumarines chez cette même plante. Enfin un projet collaboratif a été initié avec un doctorant de l'équipe du professeur Del Rio de l'université de Murcia (Espagne) pour identifier les gènes impliqués dans la synthèse de ces molécules chez diverses plantes de la famille des Fabacées (*Bituminaria bituminosa*). Dans tous les cas, des pistes sérieuses ont été mises en évidence et sont actuellement en cours d'investigation.

Devant le nombre important de gènes différents et le nombre de pistes possibles à explorer nous avons fait le choix de poursuivre ce travail sur une plante et sur 2 sous-familles de gènes différentes. Nous nous sommes focalisés sur les Apiacées qui, à ce stade, sont les seules ayant fait l'objet de recherche approfondies sur le détérminisme moléculaire de la synthèse des furanocoumarines (*A. majus*, *P. crispum*, *P. sativa*). Par ailleurs, grâce aux travaux réalisés par Hamerski et Matern dans les années 80 on sait que des P450s sont impliqués dans la synthèse de plusieurs intermédiaires réactionnels de la voie de biosynthèse des furanocoumarines (Hamerski and Matern, 1988a; Hamerski and Matern, 1988b; Hamerski *et al*., 1990; Matern *et al*., 1988). Pour ce qui concerne les sous-familles étudiées dans la suite de ces travaux, nous avons porté notre choix sur les sous-familles **CYP71AZ et CYP71AJ**. Ce choix est justifié pour plusieurs raisons. Ces deux enzymes ont été mises en évidence dans les travaux préliminaires réalisés par Matern et ses collaborateurs sur des cellules d'*A. majus* élicitées par des extraits de *Pmg*. Le nombre de représentants (complet ou partiels) pour chacune de ces deux familles est important dans la banque de *P. sativa* : 18 pour CYP71AJ et 10 pour CYP71AZ laissant ainsi supposer une surexpression par rapport à d'autres enzymes de la même famille. En ce qui concerne CYP71AJ, 2 gènes ont déjà été identifiés comme étant impliqués dans la synthèse des furanocoumarines (Larbat *et al*., 2007, 2009). Il est possible que les autres représentants de cette sous-famille le soient aussi. Malheureusement, les séquences correspondant aux autres familles ne sont que partielles et leur étude va donc nécessiter un travail préliminaire pour obtenir les séquences codantes complètes. Pour CYP71AZ, 2 séquences complètes ainsi que plusieurs séquences partielles sont présentes dans la banque d'ADNc.

CHAPITRE 2 : ETUDE ET CARACTERISATION FONCTIONNELLE D'ENZYMES DE LA SOUS-FAMILLE CYP71AZ CHEZ *PASTINACA SATIVA*

2.1 Identification des séquences correspondant à des CYP71AZ à partir de la banque d'ADNc de *P. sativa*

Pour rappel, les études préliminaires réalisées dans le cadre d'une collaboration avec l'équipe du professeur Matern avaient permis, par des techniques de tri d'ARNm (differential display RT-PCR) menées sur des cellules d'*A. majus* élicitées, d'identifier une sous-famille de gènes dont l'expression semblait en lien avec la production de furanocoumarines. Le gène identifié a été classé par le comité international de nomenclature des P450s comme étant le premier représentant de la sous-famille CYP71AZ (Genbank ID ABO32529).

L'analyse du transcriptome de panais réalisée dans la première partie de ce travail met en évidence la présence d'une dizaine de « contigs » présentant de très fortes homologies de séquence avec CYP71AZ1. L'analyse de ces 10 séquences indique que seulement 2 d'entre elles semblent contenir l'intégralité de la séquence codante pour un P450. Elles commencent par un ATG initiateur et se terminent par un codon stop. Les 8 séquences restantes sont visiblement incomplètes comme on peut le voir sur la Figure 40 (p. 150). Parmi ces 8 séquences, une (P_sat_rep_c12500) présente 100% d'identité avec la première séquence complète (P_sat_rep_c729) et une autre (P_sar_rep_c18527) a 100% d'identité avec la seconde séquence complète (P_sat_rep_c588). Les 6 autres fragments de P450s orthologues à CYP71AZ1 divergent légèrement puisque, sur les fragments disponibles, l'homologie de séquence peut aller de 84 % à 94 % indiquant par là une diversité vraisemblablement importante de cette sous-famille de P450s.

Comme cela a déjà été mentionné précédemment, ces observations doivent cependant être prises avec précaution pour deux raisons. La première raison est inhérente à la technique de séquençage utilisée pour générer la banque et qui peut être à l'origine d'erreurs. De fait, certaines des variabilités observées d'une séquence à l'autre pourraient être virtuelles. La seconde raison est que les contigs observés sont le résultat d'assemblages réalisés *in silico* et pourraient donc être artéfactuels.

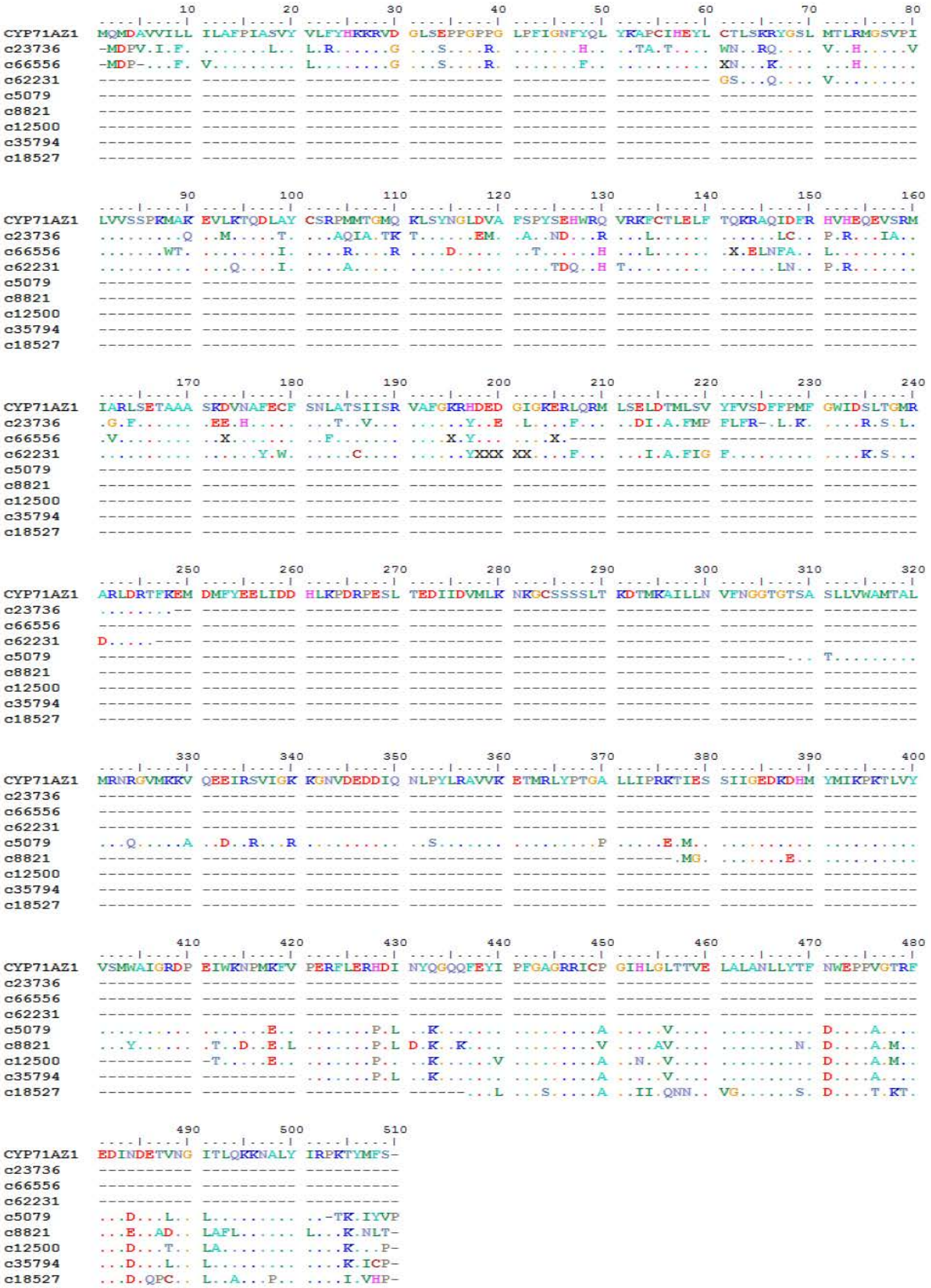


Figure 40 : Alignement des structures primaires partielles des CYP71AZ identifiés à partir de la banque d'ADNc de *P. sativa* avec celle de CYP71AZ1 d'*A. majus*. X : inconnu

2.2 CYP71AZ3 et CYP71AZ4 : analyse fine des séquences protéiques

Les 2 séquences codantes complètes identifiées dans la banque de panais ont été soumises au comité de nomenclature qui leur a attribué les codes CYP71AZ3 (P_sat_rep_c729) et CYP71AZ4 (P_sat_rep_c588). La comparaison des pourcentages d'identité et d'homologie de séquences peptidiques entre les différents membres de cette sous-famille indique que CYP71AZ1 est plus proche de CYP71AZ3 (81 % d'identité) que de CYP71AZ4 (71 %) (Tableau 14, p. 151). Il en est de même lorsque l'on compare les pourcentages d'identités nucléotidiques entre les différents membres de CYP71AZ (Tableau 14, p. 151). Ces résultats suggèrent que les différents membres de CYP71AZ dérivent d'un même gène ancestral qui pourrait être CYP71AZ1. Il est de fait possible que les enzymes correspondantes aient acquis de nouvelles fonctions par des phénomènes de duplications géniques et de mutations.

A	CYP71AZ1	CYP71AZ3	CYP71AZ4
CYP71AZ1		87 %	81 %
CYP71AZ3	87 %		83 %
CYP71AZ4	81 %	83 %	

B	CYP71AZ1	CYP71AZ3	CYP71AZ4
CYP71AZ1		81 %	71 %
CYP71AZ3	96 %		73 %
CYP71AZ4	84 %	85 %	

Tableau 14 : Comparaison des pourcentages d'identité (en vert) et d'homologie (en violet) nucléotidiques (A) et peptidiques (B) entre CYP71AZ3 et CYP71AZ4 identifiés dans la banque d'ADNc de *P. sativa* et CYP71AZ1 isolé chez *A. majus*.

Nous avons généré un arbre phylogénétique à partir d'un alignement réalisé par ClustalX et en utilisant la méthode du maximum de vraisemblance (maximum likelihood). Nous avons intégré dans cet arbre les différentes séquences peptidiques des P450s de la sous-famille CYP71AZ ainsi que les P450s de la sous-famille CYP71AJ dont le rôle dans la synthèse des furanocoumarines a déjà été démontré. Cet arbre indique que les deux gènes CYP71AZ1 et CYP71AZ3 se regroupent (ce qui est cohérent avec les pourcentages d'identité décrit précédemment) et que CYP71AZ4 est un peu plus éloigné (Figure 41, p. 152). Si on compare les résultats obtenus avec ceux de la sous-famille CYP71AJ on peut constater une certaine analogie de structure de l'arbre. Pour rappel, CYP71AJ1-3 ont été identifiées comme ayant des activités psoralène synthases tandis que CYP71AJ4 est une angélicine synthase (Figure 41, p. 152). Basé sur cette observation, on ne peut s'empêcher de faire le parallèle et d'émettre l'hypothèse (qui n'est

qu'une hypothèse de travail à ce stade) que certains CYP71AZ pourraient jouer un rôle dans la synthèse de furanocoumarines linéaires et que d'autres seraient plutôt dédiés à la synthèse de furanocoumarines angulaires. Comme CYP71AZ1 a été identifié chez *A. majus*, une plante qui ne produit que des furanocoumarines linéaires, nous pouvons alors imaginer que CYP71AZ1 et 3 pourraient être liés à la synthèse de ce type de molécules.

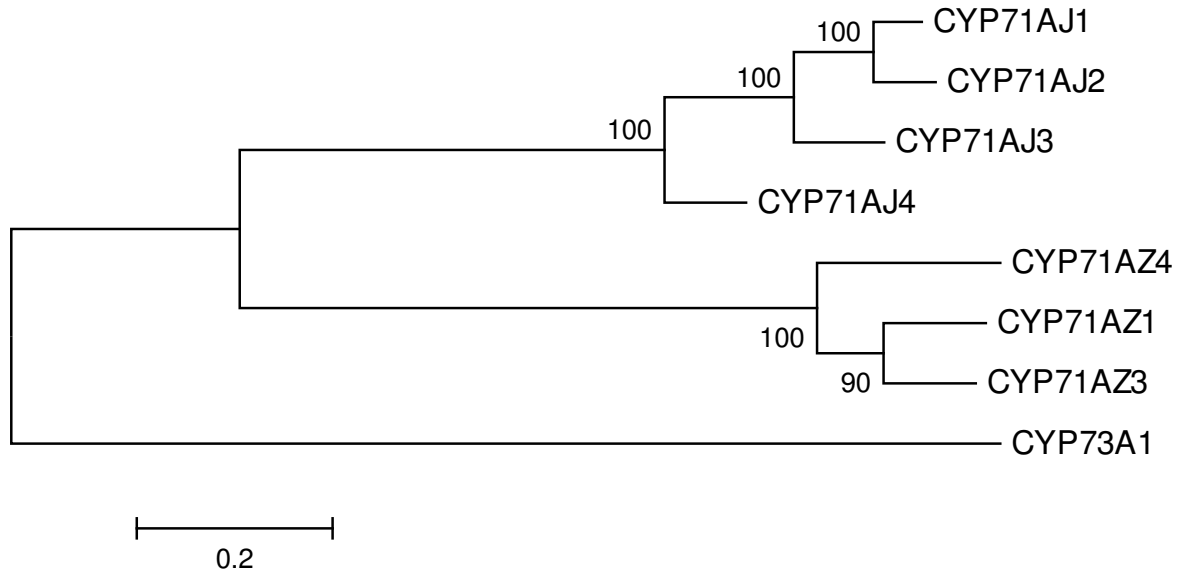


Figure 41 : Arbre phylogénétique des différents membres de CYP71AJ et CYP71AZ. Alignement de séquences peptidiques réalisé avec ClustalX et arbre phylogénétique construit avec MEGA6 en utilisant la méthode de maximum likelihood. Les numéros d'accession sur GenBank/EMBL/DDBJ des gènes caractérisés sont les suivants : CYP71AJ1 (psoralène synthase) *A. majus* (Q6QNI4), CYP71AJ2 (psoralène synthase) *A. graveolens* (C0SJS4), CYP71AJ3 (psoralène synthase) *P. sativa* (C0SJS2), CYP71AJ4 (angélicine synthase) *P. sativa* (C0SJS3).

En 1992, Gotoh et ses collaborateurs ont identifié des régions importantes chez les cytochromes P450, régions qui, dans la structure tertiaire de la protéine, interagissent directement avec le substrat. Une modification des acides aminés présents dans ces séquences SRSs peut générer une altération ou un changement de spécificité de l'enzyme vis-à-vis d'un substrat. En effet, l'encombrement stérique des chaînes latérales ainsi que les propriétés physico-chimiques des acides aminés peuvent être totalement différentes. Nous avons prêté attention aux différences existant entre CYP71AZ1, 3 et 4 au niveau des 6 SRSs potentiels (Gotoh, 1992; Larbat *et al.*, 2007). Dans ce cas, nous considérons que les structures tertiaires des protéines sont superposables et que, de fait, les SRSs sont localisés dans les mêmes régions de la structure primaire des 3 P450s (Figure 42, p. 153).

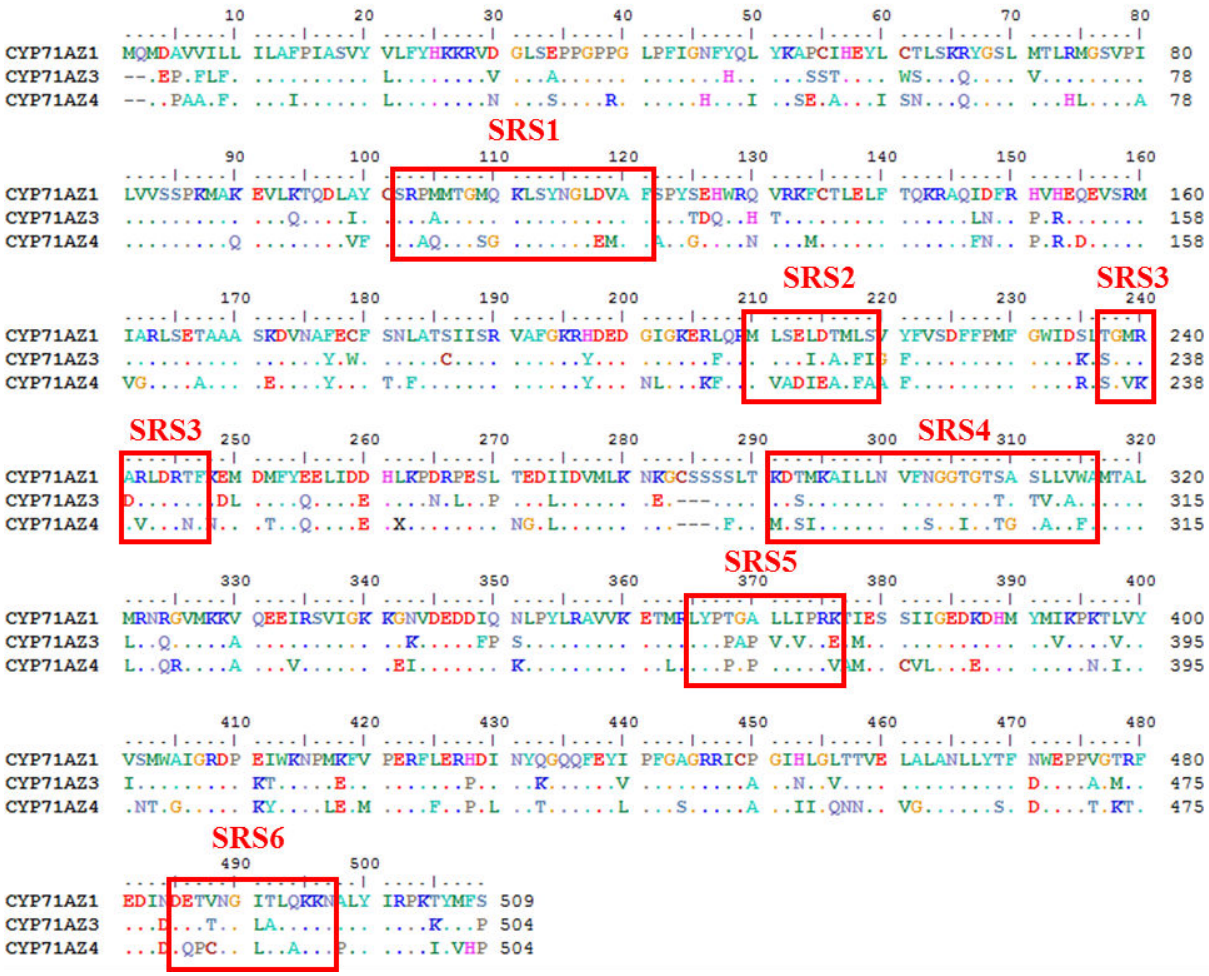


Figure 42 : Alignement des séquences peptidiques de CYP71AZ1 (*A. majus*), CYP71AZ3 (*P. sativa)* et CYP71AZ4 (*P. sativa)*. Les six séquences potentielles de reconnaissance du substrat (SRSs) sont encadrées en rouge.

L'analyse comparative des différents SRSs des membres de CYP71AZ, indique qu'*a priori*, CYP71AZ1 est plus proche de CYP71AZ3 (entre 50 et 95 % d'identité) que de CYP71AZ4 (entre 20 et 75 % d'identité) (Tableau 15-A, p. 154). Ces résultats vont dans le sens des comparaisons de séquence globale des séquences protéiques décrites précédemment. Cependant, une analyse plus détaillée des identités de séquences restreintes à l'un ou l'autre SRS donnent des informations nouvelles. Les travaux de Larbat et de ses collaborateurs indiquent que les SRS2 et SRS3 pourraient former un canal permettant l'entrée du substrat dans l'enzyme (Larbat *et al*., 2007). Au niveau du SRS2, CYP71AZ3 et CYP71AZ4 présentent une grande différence par rapport à CYP71AZ1 (60 % d'identité avec CYP71AZ3, 20 % d'identité avec CYP71AZ4) (Tableau 15-A, p. 154). Par ailleurs, l'analyse détaillée de ce SRS2 montre que la leucine 218 de CYP71AZ1 est remplacée par une phénylalanine chez CYP71AZ3 et CYP71AZ4 (Figure 42, p. 153). Ces 2 types d'enzymes ont des encombrements stériques très différents. Il est donc

très probable que cet accès au site actif laisse passer des molécules très différentes. De même, la thréonine 368 et l'alanine 370 présentes chez CYP71AZ1 sont remplacés par des prolines chez CYP71AZ3 et CYP71AZ4 (Figure 42, p. 153). Or, ces deux acides aminés sont connus pour introduire des modifications de structure secondaire comme l'interruption d'une hélice ou la destabilisation d'un feuillet β. Au niveau du SRS3, l'arginine 242 de CYP71AZ1 est remplacée par une valine ayant un encombrement stérique beaucoup plus faible chez CYP71AZ4 (Figure 42, p. 153). Enfin, au niveau du SRS5, la thréonine 368 et l'alanine 370 de CYP71AZ1 sont remplacés par des prolines chez CYP71AZ3 et 4, ce qui peut également provoquer d'importantes modifications de structure secondaire des protéines (Figure 42, p. 153). L'ensemble de ces résultats semble donc jouer en faveur d'une fonction différente de CYP71AZ1 par rapport à CYP71AZ3 ou 4.

A	**SRS 1**	**SRS 2**	**SRS 3**	**SRS 4**	**SRS 5**	**SRS 6**	**Complet**
CYP71AZ3	95	60	82	81	50	77	81
CYP71AZ4	70	20	55	65	75	62	71

B	**SRS 1**	**SRS 2**	**SRS 3**	**SRS 4**	**SRS 5**	**SRS 6**	**Complet**
CYP71AZ1	35	20	18	31	58	38	36
CYP71AZ3	35	20	9	35	75	23	34
CYP71AZ4	30	0	18	23	58	31	32

C	**SRS 1**	**SRS 2**	**SRS 3**	**SRS 4**	**SRS 5**	**SRS 6**	**Complet**
CYP71AZ1	50	0	27	35	50	38	36
CYP71AZ3	50	0	27	42	58	23	36
CYP71AZ4	35	10	36	35	42	23	34

Tableau 15 : Pourcentages d'identité des séquences peptidiques codant pour les potentiels SRSs et les séquences complètes des membres de CYP71AJ et CYP71AZ. Les séquences SRSs ont été identifiées sur la Figure 42. Ce tableau présente les pourcentages d'identité entre CYP71AZ3 et 4 et CYP71AZ1 d'*A. majus* (A). Les tableaux B et C représentent respectivement le pourcentage d'identité entre les membres de CYP71AZ et la psoralène synthase (CYP71AJ3), et l'angélicine synthase (CYP71AJ4).

Enfin, pour compléter cette étude, nous avons comparé les SRSs et les séquences complètes des CYP71AZ avec ceux qui ont été décrits pour CYP71AJ3 (Tableau 15-B, p. 154) et CYP71AJ4 (Tableau 15-C, p. 154). Dans ce cas, les différences se révèlent être très importantes (25 à 100 %). Ces analyses *in silico* semblent donc indiquer très clairement que malgré des identités de séquences assez importantes, il est fort probable que ces enzymes ne soient pas impliquées dans la synthèse des mêmes types de molécules.

2.3 Analyse du profil d'expression de ces gènes en fonction des organes de la plante

Une recherche ciblée de la fonction de CYP71AZ1 isolé à partir d'*A. majus* a été initiée en 2007 par Sandra Kellner (Kellner, 2008). Afin de voir s'il existait une corrélation entre l'expression du gène et la synthèse de furanocoumarines, des analyses cinétiques ont été réalisées sur des cellules en culture. L'ajout d'un éliciteur à base d'extrait de champignon (*Pmg*) dans la culture cellulaire influe sur la synthèse de furanocoumarines. Une analyse du niveau d'expression indique que CYP71AJ1 précède l'accumulation de psoralène (Figure 43, p. 155) (Larbat *et al*., 2007). Des résultats similaires ont été obtenus pour le gène CYP71AZ1 (communication personnelle du Pr. Ulrich Matern).

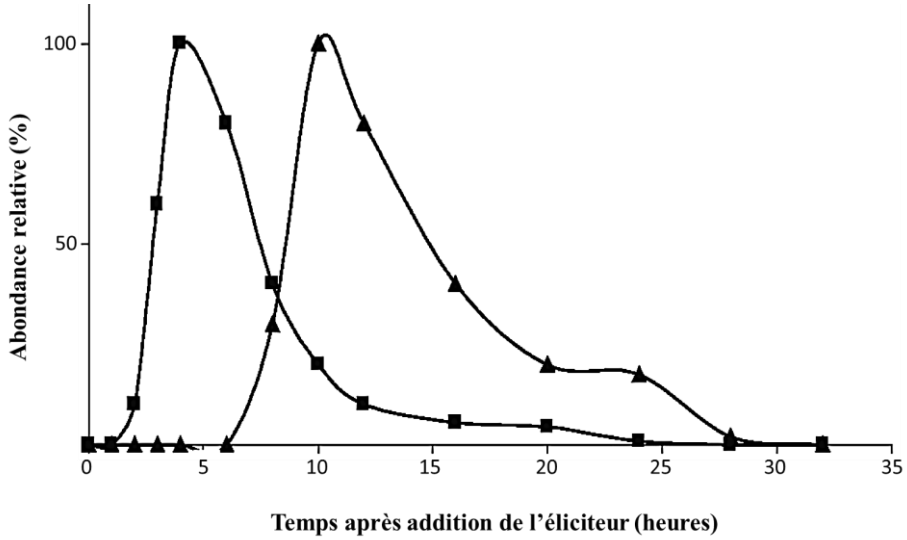


Figure 43 : Comparaison de l'expression relative de *CYP71AJ1* par rapport au profil d'accumulation de psoralène après induction dans le milieu de culture de cellules d'*A. majus* par des extraits de *Pmg* (d'après Larbat *et al*., 2007).

Afin de déterminer si CYP71AZ3 et CYP71AZ4 jouent un rôle dans la voie de biosynthèse des furanocoumarines, j'ai choisi de mettre en œuvre une stratégie proche de celle décrite précédemment. J'ai ainsi opté pour une analyse du profil d'expression de ces gènes au sein de la plante dont ils sont issus avant de voir s'il existait une corrélation avec la synthèse de furanocoumarines. Afin d'avoir les résultats les plus homogènes possible, les analyses décrites ci-dessous ont été réalisées sur des plantes, cultivées dans des conditions similaires et en parallèle. Néanmoins, il est important de noter pour la suite des expérimentations que la variété de panais 'Demi-Long de Guernesey' utilisée dans le cadre de ce travail n'est pas une variété lignée fixée, mais une variété population, et que, de fait, tous les individus ne disposent pas du même génotype. Le risque qui en découle est une variabilité phénotypique assez importante.

Cette analyse *in planta* est subdivisée en 2 parties. La première partie concerne l'étude du spectre d'expression spatial des gènes dans les différents organes de la plante, à savoir dans les feuilles, tiges, racines et graines. Dans la seconde partie j'ai souhaité voir s'il existait une corrélation entre le niveau d'expression des gènes candidats et les concentrations en furanocoumarines en réponse à un stress éliciteur chez *P. sativa*. Pour cela, j'ai d'abord essayé de déterminer les conditions d'élicitation de la production de furanocoumarines dans les feuilles et les racines de panais en essayant différents traitements : exposition à un rayonnement UV et au méthyljasmonate. Puis dans un deuxième temps, je me suis intéressée à l'impact d'un rayonnement UV sur la quantité de furanocoumarines dans les racines de panais en comparaison avec des plantes témoins.

2.3.1 Analyse du niveau d'expression de CYP71AZ3 et 4 dans différents organes de la plante

Une manière de tester l'hypothèse d'un rôle dans la voie de biosynthèse des furanocoumarines consiste à vérifier qu'il existe une corrélation entre l'expression des gènes candidats avec des gènes dont l'implication dans cette voie de biosynthèse a déjà été vérifiée. Chez le panais des études antérieures ont clairement démontré que CYP71AJ3 et CYP71AJ4 intervenaient dans la synthèse de psoralène et d'angélicine. Des études réalisées en parallèle de ce travail de thèse ont également permis d'identifier un gène codant pour une dioxygénase et un gène codant pour une prényltransférase impliqués respectivement dans la synthèse d'umbelliférone et de DMS/osthénol (Annexe 1, p. 273-281) (Karamat *et al*., 2014; Vialart *et al*., 2012). Ces 4 gènes seront donc utilisés comme gènes de références auxquels seront comparés les gènes CYP71AZ3 et CYP71AZ4.

Les ARN totaux sont extraits à partir de feuilles, de tiges, de racines de 4 plantes de *P. sativa*, ainsi que de graines. Les plantes ont été cultivées en terre pendant un an. Pour amplifier spécifiquement les différents gènes ciblés, des couples d'amorces ont été générés à l'aide du logiciel Primer Express 3.0 (CYP71AJ3 Dir : AACAATAGCGACGGCTTTGG ; CYP71AJ3 Rev : CATGGCGTCCGGACTTTTAA ; CYP71AJ4 Dir : CATTGCTCGGTTGGGTTGA ; CYP71AJ4 Rev : TGGCCGTTTTCTCCACCTT ; CYP71AZ3 Dir : TGACGACCGTAGAGCTTGCA ; CYP71AZ3 Rev : CCAGCAGGTGGTTCCCAAT ; CYP71AZ4 Dir : GGAAAAGAGAAATTCCAACGAATG ; CYP71AZ4 Rev : GAAACAAAAAAGGCAGCAAACA ; Dioxygénase Dir : GTGCCCAAACCCTGAGCTT ; Dioxygénase Rev : TGTGAGTGTGGAGACGTCAGAAT ; Prényltransférase Dir : CGAAGCACCATGGCAGAATA ; Prényltransférase Rev :

GACCAAACGTGCAAAATGCA ; 28S Dir : GAGGCCGGCAGGCATAC ; 28S Rev : CGCCCCTATACCCAAGTC). Les conditions de qPCR sont décrites dans le chapitre « Matériel et Méthodes ».

Les résultats obtenus après traitement des données par le logiciel PikoReal software 2.2 indiquent que les différents gènes sont très peu exprimés dans les graines (Figure 44, p. 158). Ce résultat n'est pas surprenant étant donné que le métabolisme secondaire y est moins actif dans la mesure où il s'agit d'un organe à l'état de dormance. Inversement, l'analyse des RT-PCR quantitatives réalisées sur les autres organes permet de mettre en évidence un niveau d'expression élevé des différents gènes. Les résultats obtenus (Figure 44, p. 158) mettent en exergue des écart-types importants dans certains cas qui peuvent s'expliquer, comme cela a été précisé plus tôt, par le fait que les plantes utilisées dans cette étude possèdent des génotypes différents. Ces différences génétiques impliquent vraisemblablement des variations alléliques des gènes impliqués dans la voie de biosynthèse des furanocoumarines et/ou des gènes régulant cette voie de biosynthèse. On constate que les 4 gènes impliqués dans la synthèse des furanocoumarines sont fortement exprimés dans tous les tissus de la plante (feuilles, tiges, racines) (Figure 44, p. 158) de même que les gènes codant pour CYP71AZ3 et 4. Les profils d'expression des gènes déjà caractérisés dans la voie de biosynthèse des furanocoumarines sont relativement similaires : c'est dans les feuilles que ces gènes sont le plus exprimés, puis dans les racines, les tiges et leur expression est la plus faible dans les graines de la plante (Figure 44, p. 158). Le même type de profil est observable pour CYP71AZ4 même s'il semblerait que le niveau d'expression soit globalement plus faible que pour les autres gènes. Pour CYP71AZ3, par contre, on constate que le profil diffère légèrement dans les tiges analysées en comparaison avec les différents gènes testés. Cette différence de profil pourrait être un indice et laisser supposer que son rôle soit différent dans le métabolisme de la plante.

Ces résultats restent néanmoins difficilement exploitables et il est hasardeux d'en tirer des conclusions. Il a été démontré dans le cadre d'autres études que, même en travaillant sur des plantes ayant un génome identique et cultivées dans des conditions similaires, la variabilité du métabolisme secondaire peut être importante (Gontier *et al*., 1993). Toutefois, ces résultats montrent qu'au niveau des feuilles, où le métabolisme est le plus actif, les profils de CYP71AZ3 et 4 sont bien corrélés avec ceux de CYP71AJ3 et 4 déjà démontrés comme étant impliqués dans la synthèse de furanocoumarines. Ces résultats permettent d'émettre l'hypothèse que CYP71AZ3 et 4 soient galement impliqués dans la synthèse de furanocoumarines.

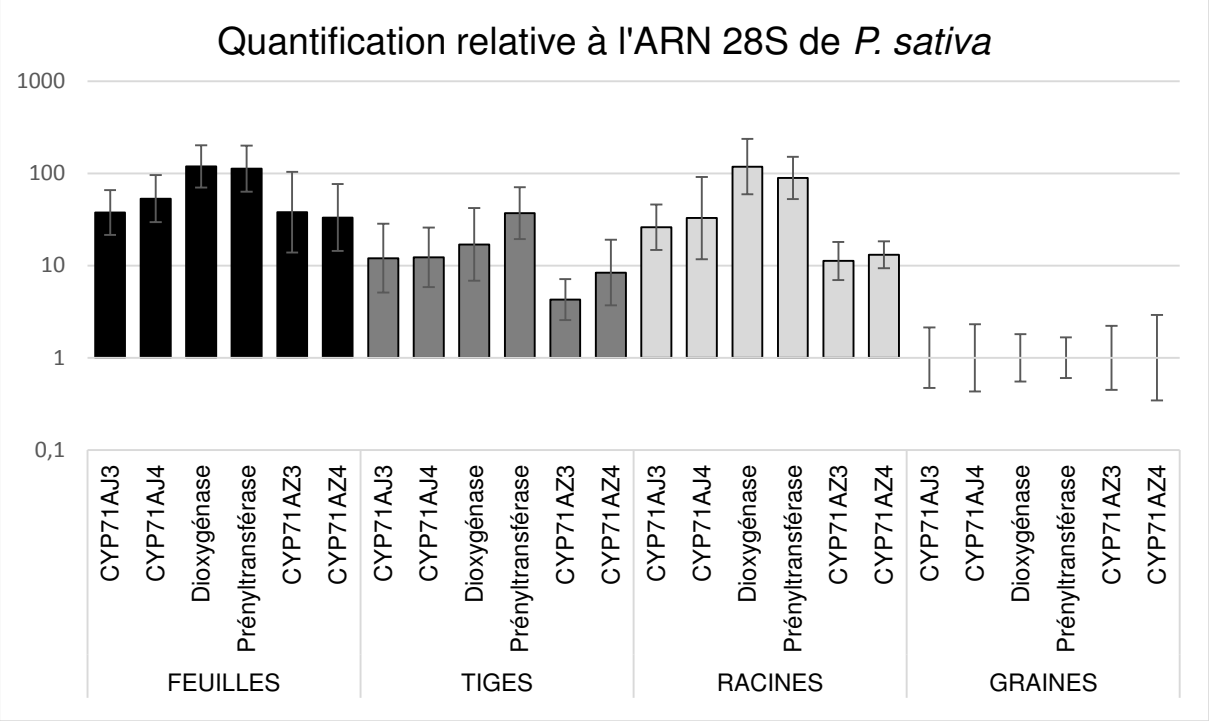


Figure 44 : Quantification des niveaux d'expression de CYP71AJ3-4, CYP71AZ3-4, d'une prényltransférase et d'une dioxygénase réalisée sur les feuilles, les tiges, les racines et les graines de *P. sativa*.

2.3.2 Tentative d'analyse du niveau d'expression de CYP71AZ3 et 4 en réponse à un stress éliciteur de la synthèse de furanocoumarines

Afin de déterminer s'il existe chez le panais une corrélation entre le niveau d'expression des gènes candidats et les concentrations en furanocoumarines en réponse à un stress éliciteur, il a d'abord fallu déterminer les conditions d'élicitation de la production de furanocoumarines chez *P. sativa*. Pour cela, d'après des données bibliographiques, nous avons choisi de tester deux éliciteurs différents, à savoir le méthyljasmonate (Ali *et al*., 2006; Lee-Parsons *et al*., 2004) et les UVB (Vialart *et al*., 2012). Dans le cas des UVB, nous avons restreint l'analyse à une longueur d'onde de 312nm qui est connue pour sa capacité à altérer la croissance, le développement, la production de biomasse et la période de floraison des plantes. Nous avons choisi de travailler avec des plantes cultivées en hydroponie afin d'accéder facilement aux racines de la plante. Les plantes ont ainsi été placées les unes à côté des autres sur un support et les racines plongées dans une solution nutritive. Les plantes ont été déplacées sur le dispositif expérimental tous les 3 jours afin de limiter les hétérogénéités du dispositif de culture qui pourraient causer des différences métaboliques de par la position des plantes par rapport à une source de lumière.

Une semaine avant l'expérimentation d'élicitation, les plantes ont été placées dans des cuves en plastique d'un litre.

Trois plantes de *P. sativa* âgées de deux mois, cultivées en hydroponie ont subi différents traitements : 1) rayonnement UVB sur les feuilles, 2) rayonnement UVB sur les racines au travers de la cuve en plastique (la perméabilité des cuves plastiques au rayonnement UV a été vérifiée en plaçant un gel d'agarose contenant de l'ADN traité au bromure d'éthidium sous un rayonnement UV), 3) pulvérisation de MeJa sur les feuilles et 4) ajout de MeJa dans la solution nutritive où baignent les racines. Un dernier lot de plantules, cultivé de la même manière que les quatre autres lots, composé également de trois plantes a servi de contrôle négatif. Des prélèvements de 100 mg de feuilles fraîches (prélèvements de folioles sur les deuxième et troisième feuilles les plus jeunes) et de 100 mg de racines fraîches sont réalisés sur la totalité des plantes à 0 h, 4 h, 8 h, 12 h et 24 h et placés dans des tubes de 1,5 ml. Dans chaque cas, le matériel végétal prélevé est broyé directement dans le tube à l'aide d'un pilon en plastique et d'azote liquide. Les polyphénols présents dans ces échantillons sont ensuite extraits et certains quantifiés par des analyses LC-MS/MS. Dans tous les cas, j'ai réalisé une analyse des composés présents dans les feuilles et dans les racines quel que soit l'organe élicité. En fonction des échantillons, nous avons analysé jusqu'à 11 composés différents : umbelliférone, marmésine, psoralène, angélicine, 8-MOP, sphondine, 5-MOP, isobergaptène, isopimpinelline, osthénol et déméthylsubérosine. Les résultats sont exprimés en $mg.g^{-1}$ de matière fraîche.

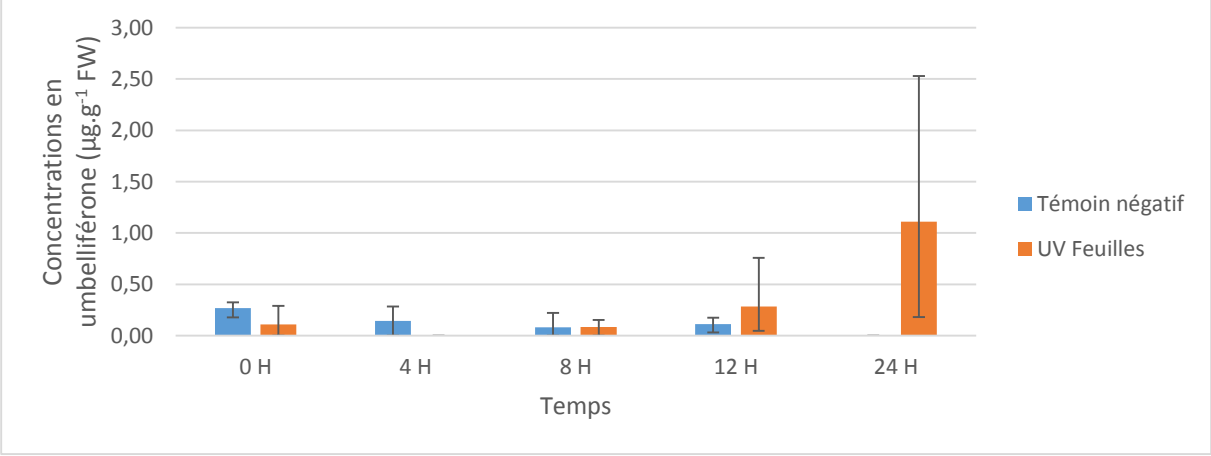


Figure 45 : Traitement des feuilles de *P. sativa* aux UVB. Analyse quantitative de la teneur en umbelliférone dans les feuilles de *P. sativa* élicitées.

Les écart-types obtenus dans le cadre de cette expérience préliminaire sont très importants et rendent les résultats relativement fragiles. Cette variabilité importante entre les différents échantillons prélevés sur

des plantes différentes est, comme cela a été évoqué précédemment, probablement en lien avec la nature génétique de la population d'individus testé. En prenant en considération cette absence de reproductibilité, ces résultats nous permettent tout de même d'envisager un impact sur l'accumulation de certaines furanocoumarines chez certains individus en fonction des différents éliciteurs et modalités d'élicitation. Ainsi l'irradiation aux UV peut provoquer une augmentation jusqu'à un facteur 6 de la teneur en umbelliférone ou en angélicine dans les feuilles après 24h de traitement et ce, si les feuilles sont traitées (Figure 45, p. 159) ou si ce sont les racines qui sont traitées (Figure 47-A, p. 161) et en angélicine à 24 h dans les feuilles (Figure 47-B, p. 161).

L'analyse des racines traitées par les UV indique même une accumulation d'umbelliférone, de psoralène, de marmésine, d'angélicine et d'isobergaptène pouvant atteindre un facteur 30 chez certains individus (Figure 47-C à G, p. 161). L'augmentation de la concentration de certaines furanocoumarines dans les feuilles lorsque les racines sont traitées aux UV semblent indiquer une transduction d'un signal au sein de la plante. Ce signal reste à être caractérisé. A l'inverse, l'exposition des feuilles aux UV ne semble par contre pas induire d'augmentation de la teneur en furanocoumarines dans les racines.

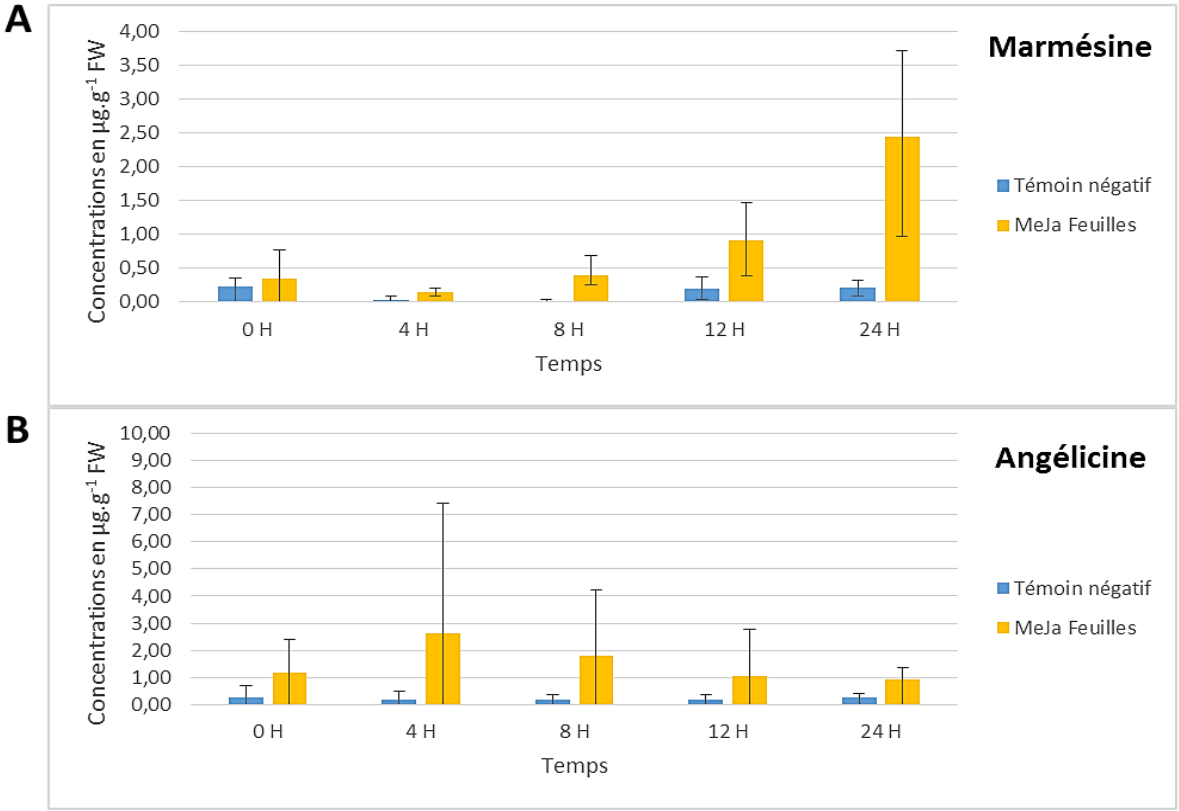


Figure 46 : Traitement des feuilles de *P. sativa* au méthyl jasmonate. A) Analyse quantitative de la teneur en marmésine dans les feuilles de panais élicitées ou non. B) Analyse quantitative de la teneur en angélicine dans les racines de *P. sativa* élicitées ou non.

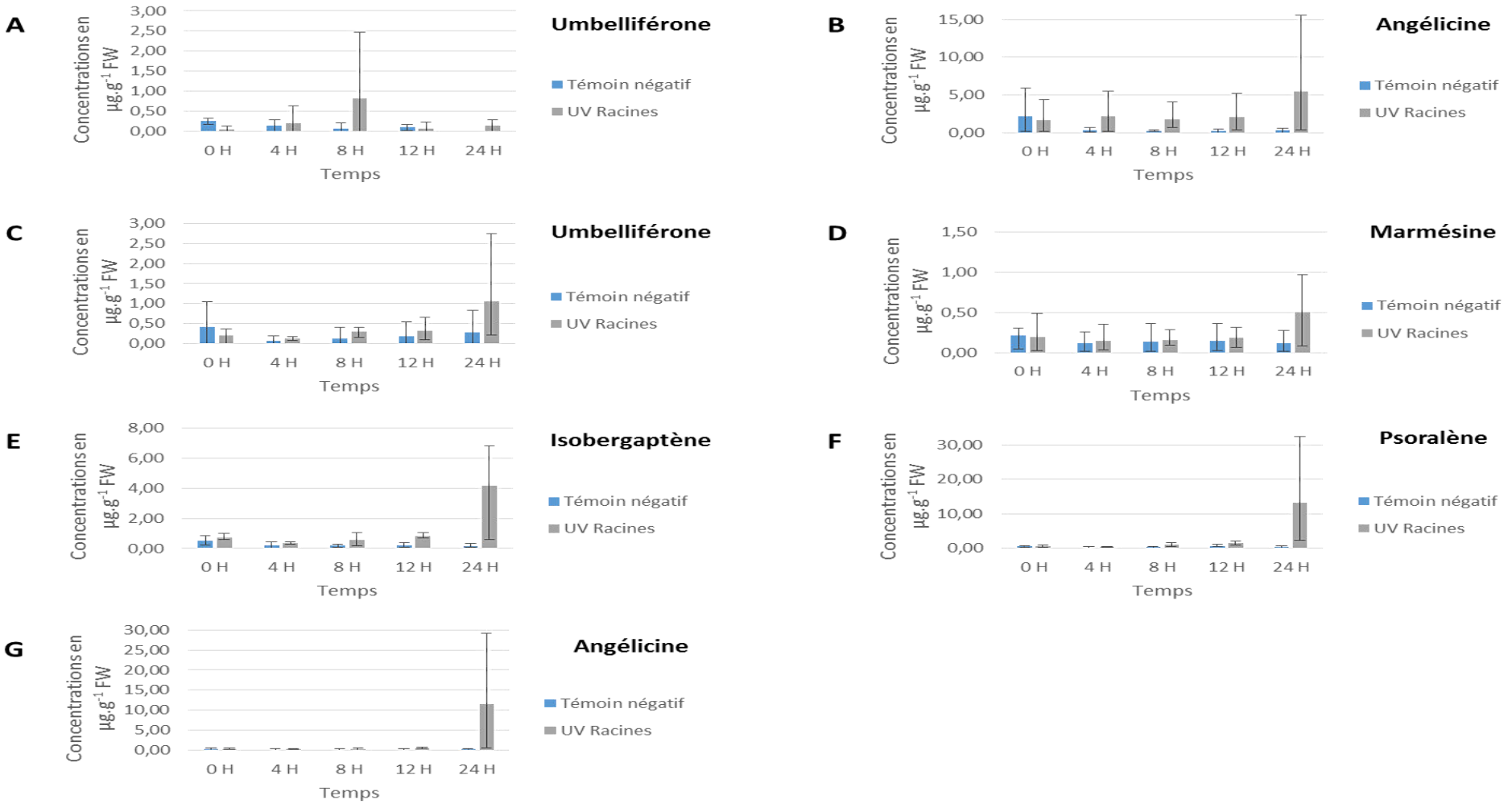


Figure 47 : Traitement des racines de *P. sativa* aux UVB. A et B : Analyses quantitatives des teneurs en umbelliférone (A) et en angélicine (B) dans les feuilles de *P. sativa* élicitées ou non. C-G : Analyse quantitative de la teneur en umbelliférone (C), marmésine (D), isobergaptène (E), psoralène (F), angélicine (G) dans les racines de *P. sativa* élicitées ou non, après une exposition des racines aux UV.

Le méthyljasmonate, qu'il soit appliqué sur les feuilles ou inclut dans la solution nutritive, provoque après 24 h l'accumulation de marmésine dans les feuilles (Figure 46-A, p. 160 et Figure 48-A, p. 162). Cette stimulation peut atteindre un facteur 3,5. Les analyses réalisées au niveau des racines sont difficilement exploitables.

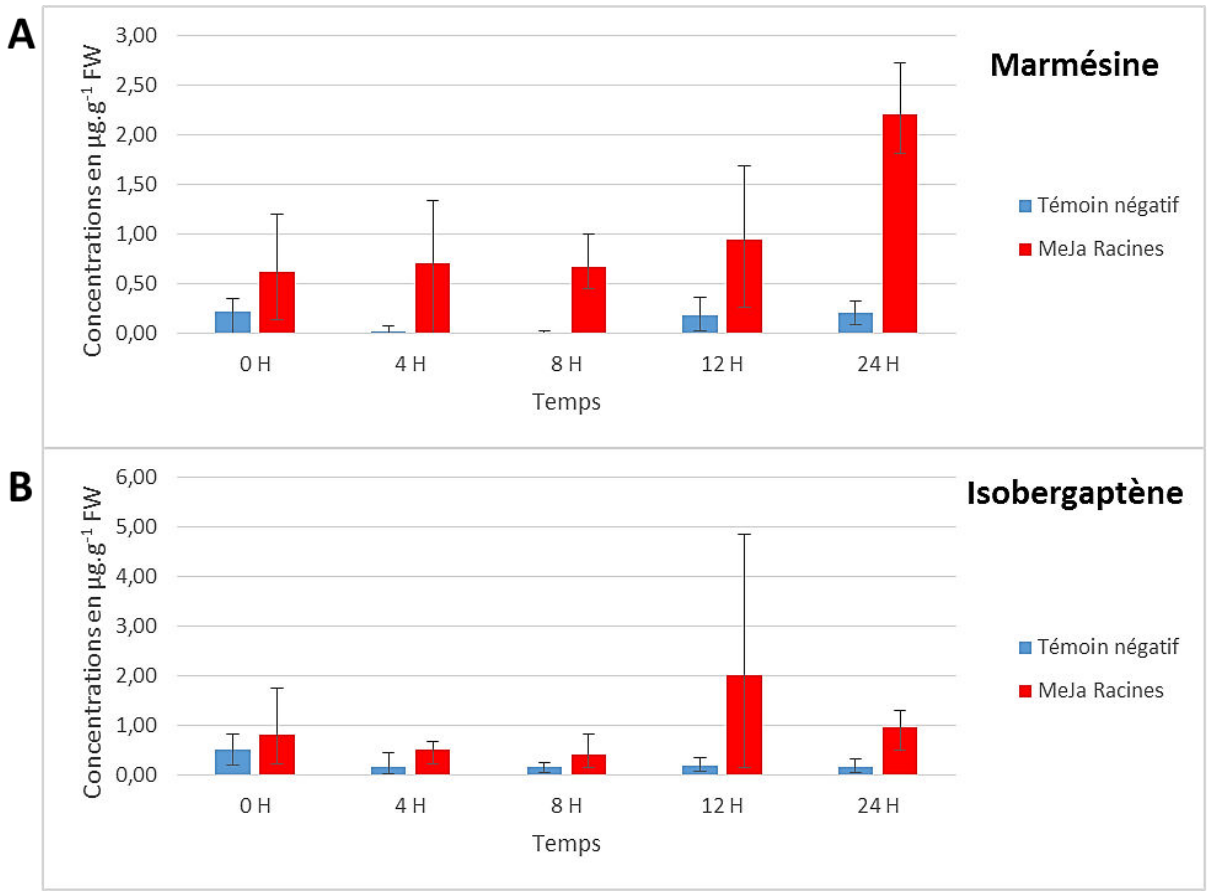


Figure 48 : Traitement des racines de *P. sativa* au méthyl jasmontate. A) Analyse quantitative de la teneur en marmésine dans les feuilles de panais après une élicitation ou non avec un traitement des racines au MeJa. B) Analyse quantitative de la teneur en isobergaptène dans les racines de *P. sativa* élicitées ou non par addition de MeJa dans la solution nutritive.

Au vue de l'hétérogénéité de ces premiers résultats, nous avons décidé de modifier les conditions expérimentales afin d'obtenir des résultats plus robustes. Pour cela, nous avons d'une part restreint le nombre de modalités et par ailleurs nous avons modifié les conditions expérimentales. Les résultats préliminaires nous indiquaient que parmi l'ensemble des conditions testées, c'est le traitement des racines aux UV qui semblait impacter le plus la synthèse des furanocoumarines dans les plantes (facteur 30 chez certains individus lorsque les racines sont irradiées pendant 24h). Nous avons donc choisi de reproduire cette modalité en doublant le nombre de plantes pour chaque traitement afin d'augmenter la

puissance statistique de l'essai. Par ailleurs, les premiers résultats ayant clairement indiqué qu'après 4 h de traitement, il n'y avait aucune modification de la teneur en furanocoumarines dans les tissus étudiés, nous avons choisi de ne plus effectuer de prélèvement à ce temps. L'accumulation de furanocoumarines commence dans certains cas à 12 h et dans d'autres à 24 h. Afin de déterminer plus précisément à partir de combien de temps commence l'accumulation et à quel temps la teneur en furanocoumarines est maximale, nous avons choisi d'ajouter les temps 18 h et 48 h. Ainsi, pour la seconde expérience d'élicitation, nous avons donc choisi d'effectuer les prélèvements à six temps (t = 0 h, 8 h, 12 h, 18 h, 24 h et 48 h). Afin de disposer de suffisamment de matériel végétal, des plantes âgées de quatre mois sont utilisées dans cette deuxième expérience (elles n'avaient que 2 mois pour la première expérimentation).

Si la variabilité des mesures obtenues peut être liée au matériel génétique et donc au métabolisme des plantes, il est également possible que cette variabilité provienne de la teneur en eau des tissus testés. En effet, lors de la première expérience, nous avons travaillé sur du matériel frais. Pour pallier ce problème potentiel, nous avons, dans cette deuxième expérience, prélevé 50 mg de racines fraîches qui ont ensuite été lyophilisées. L'analyse a ensuite été réalisée en fonction du poids sec des tissus prélevés.

Les analyses obtenues dans le cadre de cette seconde expérimentation ne sont pas présentés dans ce manuscrit de thèse. En effet malgré l'augmentation du nombre de prélèvements et malgré toutes les précautions prises, **les résultats obtenus ont de nouveau montré une hétérogénéité trop importante les rendant inexploitables** pour tenter de mettre en évidence une corrélation entre le niveau d'expression de certains gènes et la synthèse de furanocoumarines. Ces travaux devrons être réitérés par la suite sous réserve de prendre certaines nouvelles précautions. Il est envisageable de travailler sur un même génotype, mais il reste possible dans ce cas que l'hétérogénéité de conditions de culture génèrent des réponses métaboliques différentes, ou sur des cultures cellulaires qui devraient premettre de s'affranchir des conditions de culture. Il est également possible que les traitements choisis dans le cadre de l'étude présentée ici ne soient pas les mieux adaptés. Même si les UV semblent prometteurs, d'autres éliciteurs biotiques ou abiotiques pourraient permettre de mettre en évidence une accumulation plus significative des furanocoumarines.

2.4 Expression hétérologue de P450s de *P. sativa* dans un système eucaryote : la levure

Les protéines appartenant à la famille des P450s sont des protéines membranaires exprimées au niveau de la membrane externe du réticulum endoplasmique. Dans la littérature, de nombreux P450s ont été

caractérisés à l'aide d'un système d'expression hétérologue construit sur mesure par l'équipe du Dr Pompon du Laboratoire d'Ingénierie des Protéines Membranaires du CNRS (Gif/Yvette) (Cabello-Hurtado *et al*., 1998b; Humphreys *et al*., 1999; Kandel *et al*., 2007). Ce système a déjà été utilisé avec succès au laboratoire et a notamment permis de caractériser la C4H de rue (CYP73A32), la psoralène synthase (CYP71AJ1-3) et l'angélicine synthase (CYP71AJ4) (Larbat *et al*., 2007, 2009).

2.4.1 « *Amélioration* » du système d'expression de protéines développé par le Dr. Pompon

Le système d'expression hétérologue conçu par le Dr. Pompon utilise le plasmide p*Ye*DP60 comme vecteur d'expression du gène d'intérêt. Le p*Ye*DP60 est un plasmide non intégratif qui nécessite une pression de sélection pour subsister dans une levure hôte. Ce plasmide dispose d'une taille relativement importante (> 9Kb) et les étapes de clonage de gènes dans ce vecteur sont souvent longues et fastidieuses. Un des objectifs de ce travail de thèse était de cloner et d'exprimer de nombreux P450s dans le système levure. Afin d'être le plus efficace possible dans les étapes de clonage et ainsi de perdre le moins de temps possible, nous avons décidé de construire un plasmide dérivé du p*Ye*DP60 permettant ensuite des clonages facilités. Dans ce plasmide dérivé de p*Ye*DP60 nous avons introduit une cassette de clonage Gateway® (voir chapitre « Matériel et Méthodes ») commercialisée par la société Invitrogen, cassette qui devait permettre des étapes de clonage simplifiées et optimisées. Ce nouveau plasmide, appelé p*Ye*DP60_GW®, a été utilisé pour réaliser toutes les constructions ultérieures.

2.4.1.1 Introduction de la technologie Gateway® dans le vecteur p*Ye*DP60

Le vecteur p*Ye*DP60_GW® est un plasmide de destination Gateway® que nous avons obtenu en digérant 1 µg de plasmide p*Ye*DP60 par l'enzyme de restriction *Bam*HI. Les extrémités de ce plasmide ont été rendues franches par l'utilisation du fragment de Klenow de l'ADN *Taq* polymérase. Les extrémités ont ensuite été déphosphorylées afin d'éviter que le plasmide ne se referme sur lui-même. La cassette de recombinaison RfA a finalement été liguée aux extrémités franches ainsi générées à l'aide de l'ADN T4 ligase. Après insertion de la cassette de clonage dans le plasmide, des bactéries *E. coli* ccdB survival™ ont été transformées par électroporation avec le mélange de ligation. Les colonies ainsi sélectionnées sur milieu LB contenant du chloramphénicol et de l'ampicilline, ont été mises en culture en milieu liquide afin d'amplifier le plasmide. La validation de l'insertion et de l'orientation de l'insert de la cassette de clonage dans le vecteur p*Ye*DP60 est réalisée dans un premier temps en établissant une carte de restriction par une digestion des constructions obtenues aves les enzymes de restriction *Eco*RI et

*Mlu*I. Les profils de digestions obtenus sont comparés avec les profils théoriques attendus (Figure 49, p. 165).

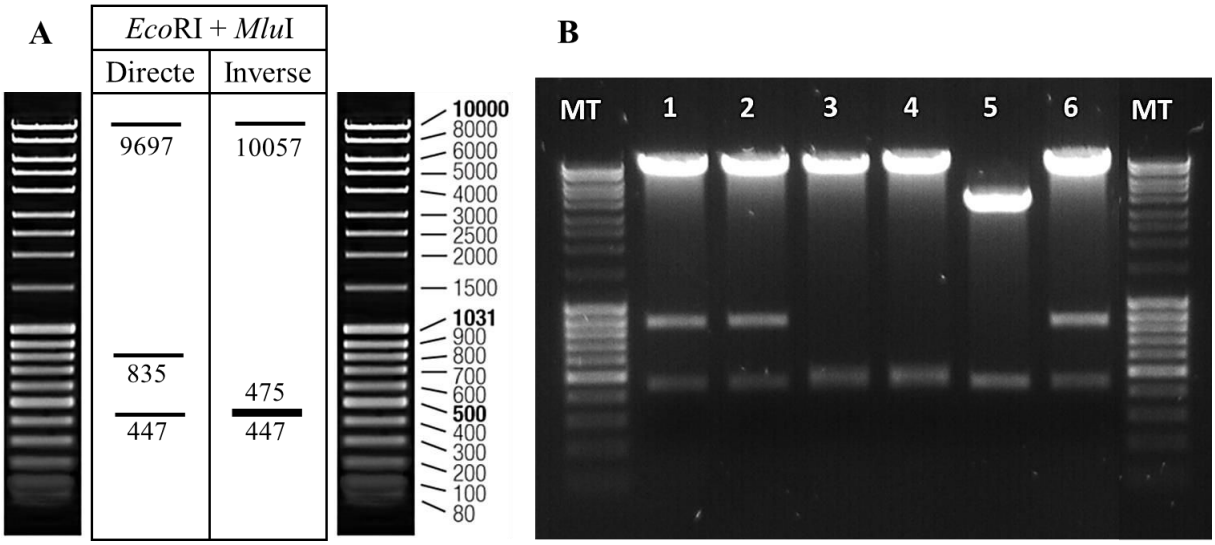


Figure 49 : (A) Profils de digestions théoriques des plasmides p*Ye*DP60_GW® ayant insérés la cassette de clonage RfA dans la bonne orientation (directe) et dans la mauvaise orientation (inverse) digérés par les enzymes de restriction *Eco*RI et *Mlu*I, comparés au marqueur de taille MassRuler DNA Ladder Mix, ready-to-use (Fermentas). (B) Profils de digestions de six plasmides supposés p*Ye*DP60_GW® digérés par les enzymes de restriction *Eco*RI et *Mlu*I (puits 1 à 6) comparés au maqueur de taille MassRuler DNA Ladder Mix, ready-to-use (Fermentas).

Les profils obtenus pour les 6 constructions testées indiquent que, pour les constructions 1, 2 et 6, le clonage de la cassette RfA est réalisé dans le bon sens, alors que pour les constructions 3 et 4, le clonage s'est effectué dans la mauvaise orientation. Enfin, pour la construction 5, il semble qu'il y ait eu une perte d'un fragment du plasmide p*Ye*DP60 lors de la digestion ou du clonage. Les constructions 1 et 2 ont ensuite été réamplifiées à l'aide de bactéries *E. coli* ccdB survival™, et envoyées à la société Eurofins MWG Operon afin de séquencer les régions contenant l'insert à l'aide d'une amorce directe (TCTTTCCTTATACATTAGGTCC) située en amont du site *Bam*HI du p*Ye*DP60 et une amorce réverse (CAACACCTGGCAATTCCTTACC) située en aval du site *Eco*RI du plasmide p*Ye*DP60. Les résultats obtenus indiquent que ces deux constructions sont fidèles à ce que nous attendions et que la séquence de l'insert est identique à celle de la cassette de clonage RfA (Annexe 14, p. 314).

Une fois que la construction a été validée par le séquençage, il a ensuite été nécessaire de vérifier si le plasmide reste fonctionnel. En effet, le clonage de cette cassette GATEWAY® au niveau du site *Bam*HI du plasmide p*Ye*DP60 introduit des séquences supplémentaires entre le promoteur GAL10/CYC1 et le

gène qui sera exprimé dans les levures. Cette distance peut affecter de manière significative le niveau d'expression des protéines. Il est donc nécessaire de valider cette construction en réalisant un clonage et l'expression fonctionnelle d'une protéine dont la fonction est connue.

2.4.1.2 Validation de la fonctionnalité du plasmide p*Ye*DP60_GW®

Afin de vérifier l'intégrité et la fonctionnalité du p*Ye*DP60_GW®, nous avons choisi de cloner et d'exprimer dans des levures WAT11 un gène dont la fonction est connue. Notre choix c'est porté sur CYP73A1, isolé à partir d'*H. tuberosus* qui code pour une C4H et qui catalyse la formation d'acide *p*-coumarique à partir d'acide cinnamique. La séquence codante de CYP73A1 a été clonée dans le plasmide d'entrée pCR8™/GW par Antoine Gravot durant sa thèse au laboratoire. Après m'être assurée, par séquençage, que le plasmide était bien le plasmide attendu, j'ai introduit la séquence codante dans le plasmide p*Ye*DP60_GW® par LR recombinaison. Le plasmide recombinant obtenu a été introduit dans des levures. Afin de m'assurer du bon fonctionnement du plasmide, j'ai transformé des levures WAT11 avec trois constructions différentes : p*Ye*DP60 vide (témoin négatif), p*Ye*DP60_CYP73A1 (témoin positif) et p*Ye*DP60_GW®_CYP73A1. Pour chaque construction plasmidique, une colonie de levure WAT11 a été isolée et mise en culture afin d'exprimer les protéines correspondantes (cf. chapitre « Matériel et Méthodes »). Une préparation de microsomes a ensuite été réalisée pour les trois constructions testées. Des incubations ont été réalisées dans 146 µl de tampon de phosphate de sodium (NaPi) 0,1 M pH 7,5 contenant 200 µM d'acide cinnamique en présence et en absence de NADPH (1 mM). Les réactions sont initiées par l'ajout des différents microsomes. Les réactions sont incubées pendant 30 minutes à 28 °C sous agitation (200 rpm) et sont stoppées par l'ajout de 37,5 µl d'acétonitrile/HCl (99/1). Les milieux réactionnels sont ensuite centrifugés à 10 000 g pendant 30 min, puis les surnageants sont prélevés et mis dans des vials pour être analysés en HPLC.

Dans le cas du vecteur vide, que ce soit en présence ou en absence de cofacteur, aucune métabolisation n'est observée (Figure 50, p. 167, en bleu). Aucune métabolisation n'est observée pour les deux constructions exprimant CYP73A1 en absence de cofacteur. L'incubation de microsomes exprimant CYP73A1 contenu dans le vecteur p*Ye*DP60 en présence d'acide cinnamique et de NADPH donne un produit qui correspond à l'acide *p*-coumarique (Figure 50, p. 167, en rouge). De la même manière, les microsomes exprimant CYP73A1 contenu dans le p*Ye*DP60_GW® incubés en présence d'acide cinnamique et de NADPH forment également de l'acide *p*-coumarique (Figure 50, p. 167, en noir). La détection du produit de CYP73A1 (acide *p*-coumarique) pour les deux constructions exprimant CYP73A1 nous permet donc de valider la fonctionnalité de la protéine et donc de notre plasmide

p*Ye*DP60_GW®. Ce plasmide de destination sera utilisé ultérieurement pour tous les clonages de P450s à réaliser en vue de les exprimer chez la levure.

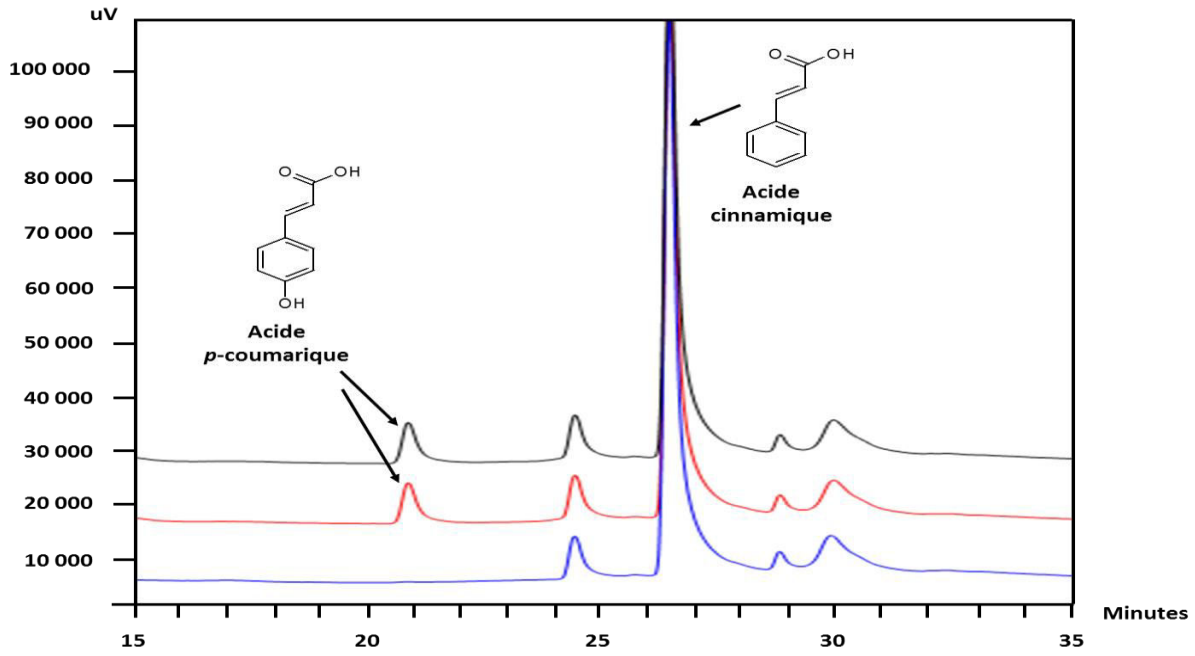


Figure 50 : Comparaison des chromatogrammes extraits à 300 nm, issus du test de métabolisation de CYP73A1 de topinambour avec de l'acide cinnamique en présence de NADPH. En bleu : vecteur vide, en rouge : p*Ye*DP60_CYP73A1, en noir : p*Ye*DP60_GW®_CYP73A1. L'acide *p*-coumarique apparait uniquement en présence de NADPH pour les levures transformées avec les plasmides p*Ye*DP60_CYP73A1 et p*Ye*DP60_GW®_CYP73A1.

2.4.1.3 Utilisation du système pour l'expression des CYP71AZ de *P. sativa*

La caractérisation fonctionnelle dans un système hétérologue des CYP71AZ isolés chez *P. sativa* nécessite d'avoir ces gènes clonés dans un vecteur d'entrée en vue de les transférer ultérieurement dans un vecteur de destination adapté à l'organisme hétérologue choisi. Pour cela, les ARN totaux de feuilles de panais ont été extraits à l'aide du kit RNeasy® Plant Mini (Qiagen) selon les recommandations du fournisseur. L'intégrité de ces ARN a été vérifiée sur gel d'agarose 1 %. Puis, une réaction de transcription inverse a été réalisée selon les conditions citées dans le chapitre « matériel et méthodes ». Les ADNc correspondant à CYP71AZ3 et CYP71AZ4 sont amplifiés par PCR avec les amorces décrites le Tableau 16 (p. 168).

Amorces	Séquences 5' → 3'	Site de restriction
CYP71AZ3 dir	GGATCCATGGAGCCAGTATTTCTCTTTC	*Bam*HI
CYP71AZ3 rev	GAATTCTTA**GTGGTGATGGTGATGATG**TGGA AACATATATTTTTTAGG	*Eco*RI
CYP71AZ4 dir	GGTACCATGGATCCAGCAGCTATC	*Kpn*I
CYP71AZ4 rev	GAATTCTTA**GTGGTGATGGTGATGATGTG**GA TGTACATATATTTTAGG	*Eco*RI

Tableau 16 : Amorces utilisées pour l'amplification des séquences codantes de CYP71AZ3 et CYP71AZ4 de *P. sativa*. Les sites de restriction contenus dans les amorces sont soulignés. Les séquences correspondantes à une queue poly histidine (6 histidines) sont indiquées en rouge.

L'amplification a été réalisée en effectuant 35 cycles (94 °C 30 s, 55 °C 30 s, 68 °C 2 minutes) grâce à la Platinium® *Taq* DNA Polymerase High Fidelity (Invitrogen). Les produits de PCR ont été déposés sur gel d'agarose 1 % et les fragments ayant une taille d'environ 1 500 pb ont été prélevés puis purifiés à l'aide du kit « MinElute® Gel Extraction » (Qiagen) selon les recommandations du fournisseur. Ces produits ont ensuite été clonés dans le vecteur pCR8®/GW/TOPO® et envoyés au séquençage afin de vérifier la séquence des inserts ainsi que l'orientation des inserts dans le vecteur. Les séquences codantes de CYP71AZ3 et CYP71AZ4 ont ensuite été transférées dans le plasmide de destination p*Ye*DP60_GW® grâce à une réaction de LR recombinaison (voir chapitre « Matériel et Méthodes »). Les milieux réactionnels ont ensuite été utilisés pour transformer des bactéries GeneHogs étalées sur milieu LB + ampicilline afin de sélectionner les bactéries transformées. Une colonie isolée par construction a été mise en culture afin de multiplier nos plasmides recombinants. Les plasmides ont ensuite été extraits et envoyés à la société Eurofins MWG Operon pour être séquencés afin de s'assurer qu'il n'y ait pas eu de modification de séquences des inserts.

Les deux constructions plasmidiques validées par le séquençage ont été introduites dans la souche de levure WAT11 selon la méthode appelée « à l'acétate de lithium » décrite dans le chapitre « Matériel et Méthodes ». Pour chaque construction plasmidique, quatre colonies de levures transformées sont isolées, puis mises en culture dans un milieu SGI, aliquotées par 1 ml et stockées à – 80 °C dans un milieu additionné de 30 % de glycérol.

Dans notre laboratoire, l'expression hétérologue de P450s végétaux dans le système levure est une technique bien maîtrisée et de routine. Le principe de cette méthode est décrit dans le chapitre « Matériel et Méthodes ». L'expression hétérologue de CYP71AZ3 et CYP71AZ4 a été réalisée avec une période de croissance de 24 h à 28 °C suivie d'une période d'induction de 16 h à 18 °C. A l'issue de l'induction,

les levures ont été lysées par agitation mécanique et les microsomes ont été préparés comme cela est décrit dans le chapitre « Matériel et Méthodes ». Des aliquots de 200 µl ont été préparés et stockés à – 20 °C en vue de réaliser ultérieurement un criblage métabolique.

Avant d'initier un criblage fonctionnel il est nécessaire de s'assurer que la protéine a été correctement synthétisée dans le milieu de culture. Pour cela plusieurs approches ont été décrites. La première technique historiquement décrite par Omura et Sato (Omura and Sato, 1964) concerne l'établissement de spectres différentiels au CO en incubant les microsomes avec un agent réducteur (dithionite de sodium) et en présence/absence de monoxyde de carbone. Si un pic d'absorbance est détecté à 450 nm, la protéine est considérée comme exprimée et fonctionnelle, alors que si un pic d'absorbance est détecté à 420 nm, la protéine est considérée comme en cours de dégradation voire déjà dégradée.

Maheureusement, le niveau d'expression des protéines P450s est souvent relativement faible. L'établissement de spectres différentiels au CO devient souvent une gageure et dans certains cas, même si aucun pic d'absorbance n'est détecté à 420 ou 450 nm, la protéine peut tout de même être exprimée et active. Nous avons donc décidé de vérifier la production de nos protéines d'intérêt dans le système levure en utilisant une approche plus sensible par une technique de western blot en utilisant des anticorps spécifiques d'un motif poly histidine. Ce motif constitué de 6 histidines n'est pas présent sur les séquences de P450s sauvages. Nous l'avons rajouté artificiellement à l'extrémité C terminale du gène en amont du codon STOP en utilisant des amorces de PCR spécifiques. Les protéines contenues dans nos préparations de microsomes ont été séparées selon leur poids moléculaire apparent par une électrophorèse en gel de polyacrylamide en conditions dénaturantes comme cela est décrit dans le chapitre « Matériel et Méthodes ». Deux colonies ont été testées par construction et les microsomes de levure transformées avec le vecteur vide ont servi de contrôle négatif. Le contenu du gel a ensuite été transféré sur une membrane afin de réaliser un western blot avec des anticorps spécifiques du motif 6 X Histidines. La révélation a été réalisée avec un mélange de NBT/BCIP.

Aucune bande n'est révélée pour les levures transformées avec le vecteur vide (Figure 51-A et B, p. 170, puits 3). Les microsomes préparés à partir de deux colonies de levure WAT11 exprimant CYP71AZ4 produisent la protéine d'intérêt (Figure 51-A, p. 170, puits 1 et 2), mais il semble que la protéine soit plus exprimée dans la première colonie (Figure 51-A, p. 170, puits 1) que dans la deuxième (Figure 51-A, p. 170, puits 1). De la même manière, CYP71AZ3 est exprimée dans les deux colonies de levure testées, mais c'est dans la deuxième colonie (Figure 51-A, p. 170, puits 2) que son expression est la plus forte. Les deux protéines ont une taille légèrement supérieure à 49 kDa ce qui correspond au poids

moléculaire théorique attendu. Pour la suite des expérimentations, nous avons choisi de partir des colonies de levures présentant le signal le plus intense (la colonie n°1 de levure WAT11 exprimant CYP71AZ4 et la colonie n°2 de levure WAT11 exprimant CYP71AZ3).

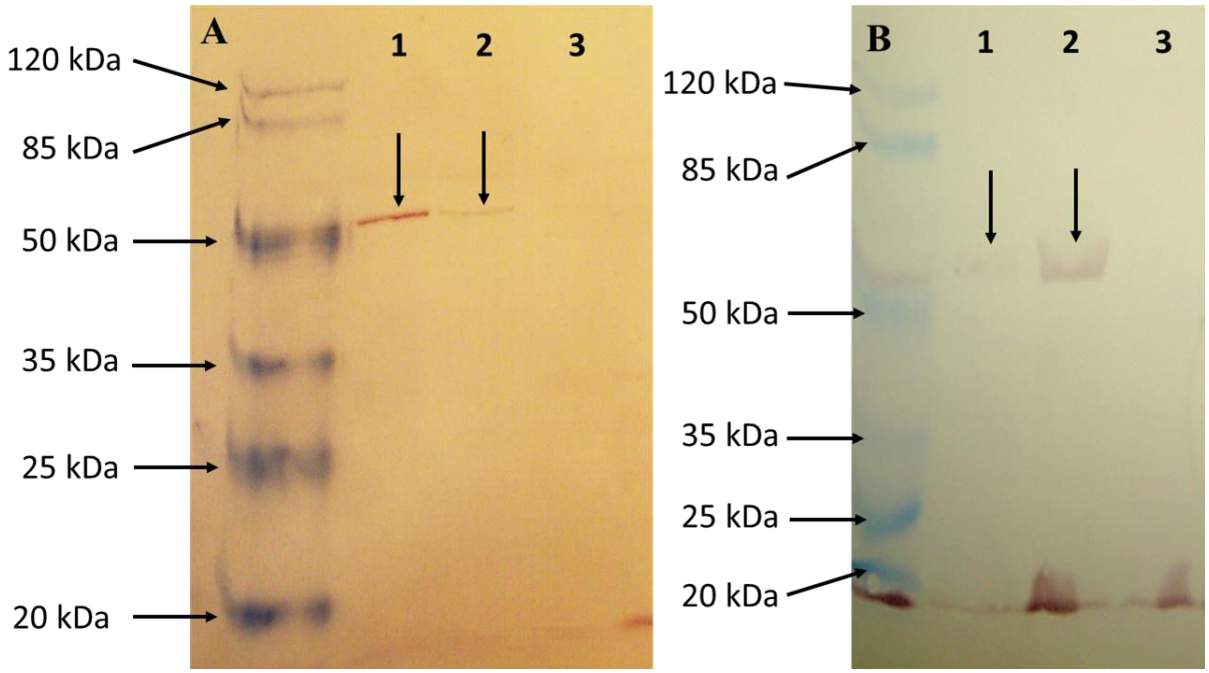


Figure 51 : Détection de la production des protéines CYP71AZ4 (A) et CYP71AZ3 (B) dans le système levure WAT11 par western blot après séparation en SDS-PAGE en comparaison avec le marqueur de taille Prestained Protein Molecular Weight Marker (ready-to-use) (Fermentas).

2.4.2 Caractérisation fonctionnelle de CYP71AZ3 et CYP71AZ4 *in vitro*

2.4.2.1 Criblage fonctionnel

En vue de déterminer la fonction de CYP71AZ3 et CYP71AZ4 de panais, un criblage métabolique a été réalisé sur un large panel de molécules disponibles au laboratoire. Une première série de tests de métabolisation a été réalisée sur des furanocoumarines linéaires et angulaires (30 composés au total) incluant majoritairement des intermédiaires de la synthèse de furanocoumarines, mais également des dérivés (Figure 52, p. 172).

Les incubations ont été réalisées dans un volume final de 100 µl conteant 76 µl de tampon NaPi 0,1 M pH 7,5, 200 µM de substrat, les microsomes et 1 mM de NADPH. Les mêmes incubations ont été effectuées en absence de cofacteur. Afin de vérifier qu'une métabolisation de molécule n'est pas

effectuée par des protéines endogènes des levures, nous avons en parallèle réalisé les mêmes incubations avec des microsomes préparés à partir de levures transformées avec le vecteur p*Ye*DP60 vide. Les réactions ont été effectuées à 28 °C pendant 30 min sous agitation à 200 rpm puis stoppées par l'ajout de 37,5 µl d'acétonitrile/HCl (99/1). Les milieux réactionnels ont ensuite été centrifugés à 10 000 g pendant 30 min et les surnageants ont été prélevés et analysés par HPLC.

Le criblage métabolique réalisé avec des furanocoumarines a permis de détecter la formation d'**un produit** lorsque des microsomes de levures exprimant **CYP71AZ4** sont incubés en présence de psoralène et de NADPH (Figure 53-A, p. 173). Ce produit formé n'apparaît qu'en présence de NADPH (Figure 53-A, p. 173) et n'est pas produit lorsque l'on incube des microsomes de levure transformées avec le vecteur p*Ye*DP60 vide (en absence ou en présence de cofacteur). **Le produit ainsi formé possède le même temps de rétention que le xanthotoxol** (23,52 min) et le même spectre d'absorption (250 nm – 380 nm) caractérisé par deux longueurs d'onde maximale à 267 nm et 308 nm (Figure 53-B et C, p. 173). Le produit formé a ensuite été analysé en LC-MS/MS en mode ionisation positive $(M-H)^+$ afin de confirmer son identité selon la méthode décrite dans le chapitre « Matériel et Méthodes ». Cette analyse a montré que le produit formé possède le même temps de rétention en LC que le standard xanthotoxol. De plus, le spectre de fragmentation du produit formé est parfaitement identique à celui de la molécule standard xanthotoxol (masse totale de 202 g/mol avec des pics majoritaires de fragmentation à 130, 146, 158, 174 et 184) (Figure 53-B et C, p. 173). **Ces résultats permettent donc de confirmer que le produit formé lorsque l'on incube des microsomes de levure exprimant CYP71AZ4 avec du psoralène en présence de cofacteur est bien du xanthotoxol (8-hydroxypsoralène).**

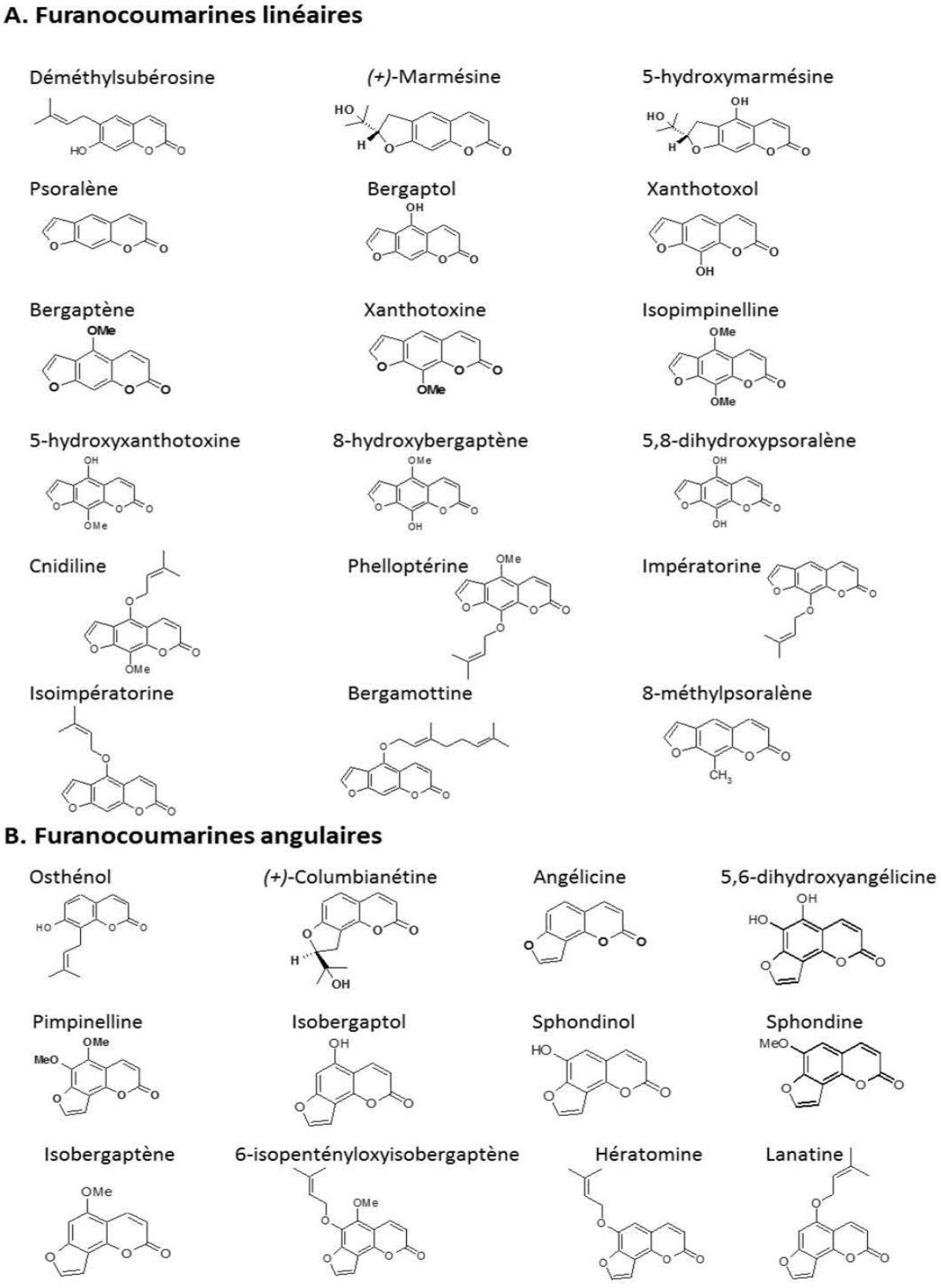


Figure 52 : Liste des furanocoumarines linéaires (A) et angulaires (B) utilisées pour le premier criblage métabolique réalisé avec CYP71AZ3 et CYP71AZ4.

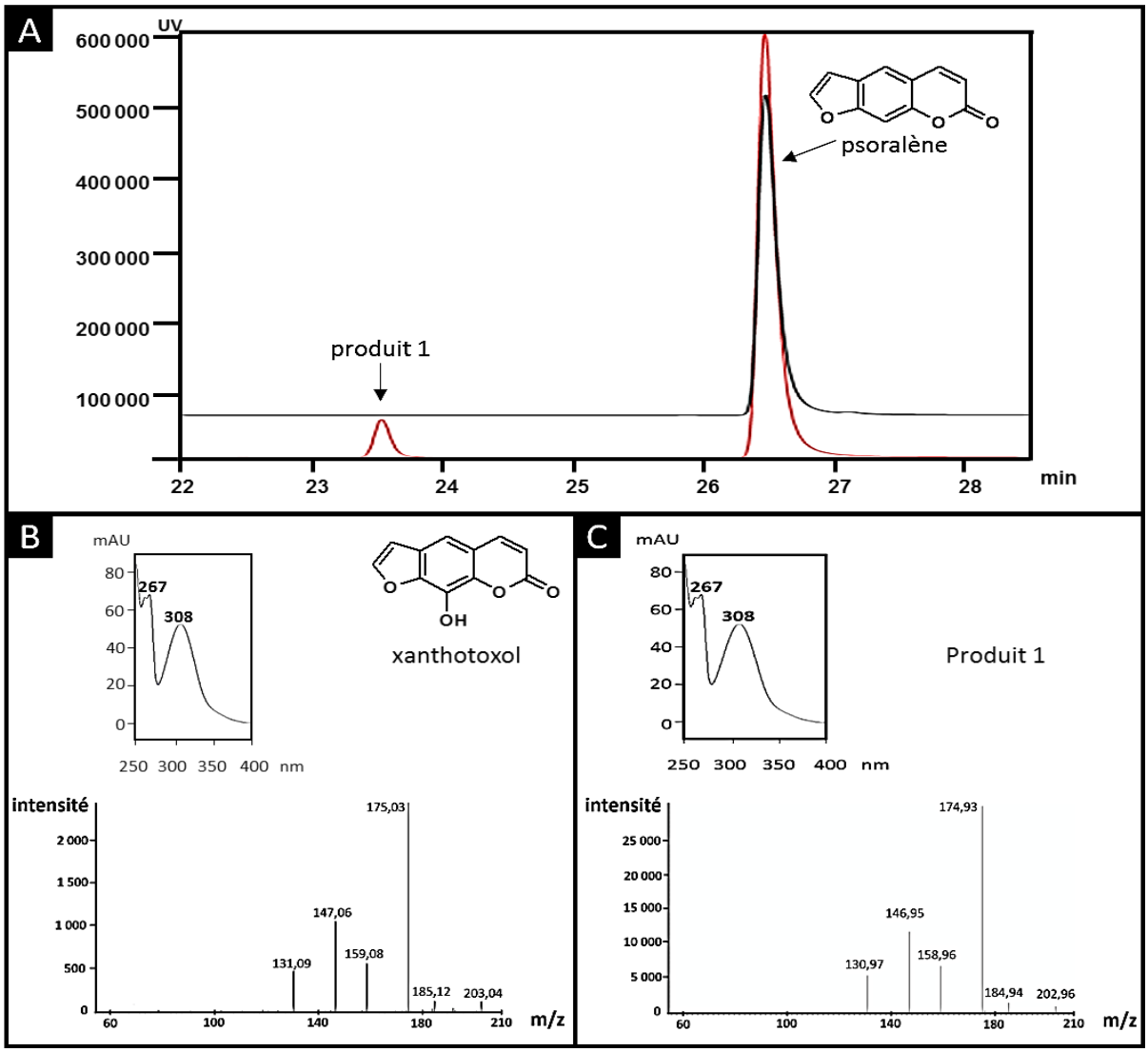


Figure 53 : Détection de la formation de xanthotoxol à partir du psoralène catalysée par CYP71AZ4. A : Détection en HPLC du produit de la métabolisation du psoralène en xanthotoxol. En rouge : avec NADPH, en noir : même incubation sans NADPH. B : Spectre d'absorbance et spectre de masse en mode ionisation positive $(M-H)^+$ du standard xanthotoxol. C : Spectre d'absorbance et spectre de masse en mode ionisation positive $(M-H)^+$ du produit formé par CYP71AZ4 en présence de psoralène.

Malheureusement, le criblage métabolique effectué avec les furanocoumarines n'a pas permis de détecter de produit formé dans le cas de microsomes exprimant CYP71AZ3. Nous avons donc décidé d'élargir le panel des molécules testées et nous avons choisi de réaliser dans un deuxième temps des tests de métabolisation avec toutes les coumarines disponibles au laboratoire (22 composés), l'acide cinnamique et quelques un de ses dérivés (3 composés), deux flavonoïdes et une furoquinoline alcaloïde (Figure 54, p. 174). Les incubations ont été réalisées dans les mêmes conditions que lors des tests de métabolisation réalisés avec des furanocoumarines.

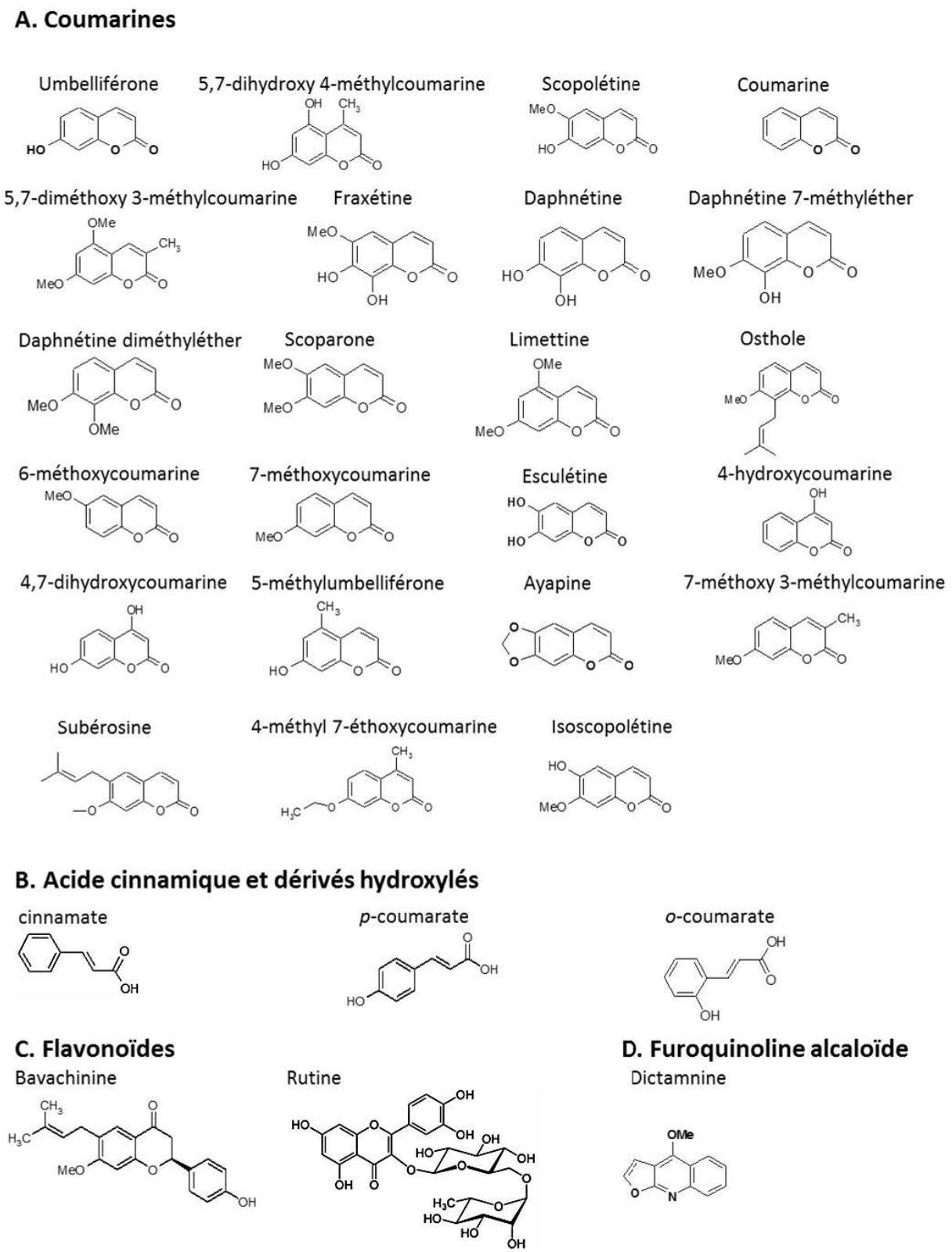


Figure 54 : Liste des substrats utilisés au cours du deuxième criblage métabolique réalisé avec CYP71AZ3 et CYP71AZ4. (A) : Coumarines, (B) : Acide cinnamique et ses dérivés hydroxylés, (C) : Flavonoïdes, (D) : Furoquinoline alcaloïdes.

Ce deuxième criblage métabolique a permis de mettre en évidence la **métabolisation de 4 substrats** qui conduisent à la formation de 5 produits lorsque des microsomes exprimant **CYP71AZ4** sont incubés en présence de NADPH et de 4 coumarines : **scopolétine, 7-méthoxycoumarine, 6-méthoxycoumarine et 7-méthoxy 3-méthylcoumarine**. Ces métabolisations n'apparaissent qu'en présence de NADPH et lorsque l'on incube des microsomes de levure transformées avec le vecteur p*Ye*DP60 vide en présence ou en absence de cofacteur, aucun produit n'est formé.

CYP71AZ4 et métabolisation de coumarines

Lorsque des microsomes de levures exprimant CYP71AZ4 sont incubées en présence de scopolétine et de NADPH, le produit ainsi formé possède le même temps de rétention que la fraxétine (19,01 min) (Figure 55-A, p. 176). De plus, le produit formé possède le même spectre d'absorption (270 nm – 390 nm) que la fraxétine avec un maximum d'absorption détecté à 338 nm (Figure 55-B et C, p. 176). Le produit a été analysé en LC-MS/MS et les résultats indiquent qu'il possède le même de temps de rétention que le standard fraxétine et que son spectre de fragmentation est identique à celui de la molécule standard fraxétine avec une masse totale de 208 g/mol avec des pics de fragmentation majoritaires à 148, 162, 180 et 193 (Figure 55-B et C, p. 176).

De la même manière, lorsque des microsomes de levure exprimant CYP71AZ4 sont incubés en présence de 7-méthoxycoumarine (herniarine) et de NADPH, un produit est formé (Figure 56-A, p. 177). Ce produit possède le même temps de rétention (21,59 min) (Figure 56-A, p. 177) et le même spectre d'absorption que la daphnétine 7-méthyléther (250 nm – 375 nm, avec deux maxima d'absorption détectés à 260 nm et 320 nm) (Figure 56-B et C, p. 177). Le produit formé à partir de la 7-méthoxycoumarine a également été analysé en LC-MS/MS et les résultats obtenus montrent que ce produit possède les mêmes caractéristiques que le standard daphnétine 7-méthyléther avec un temps de rétention (16,33 min) et un spectre de fragmentation identiques (masse totale de 192 g/mol avec des pics de fragmentation majoritaires à 133, 148, 160 et 177) (Figure 56-B et C, p. 177).

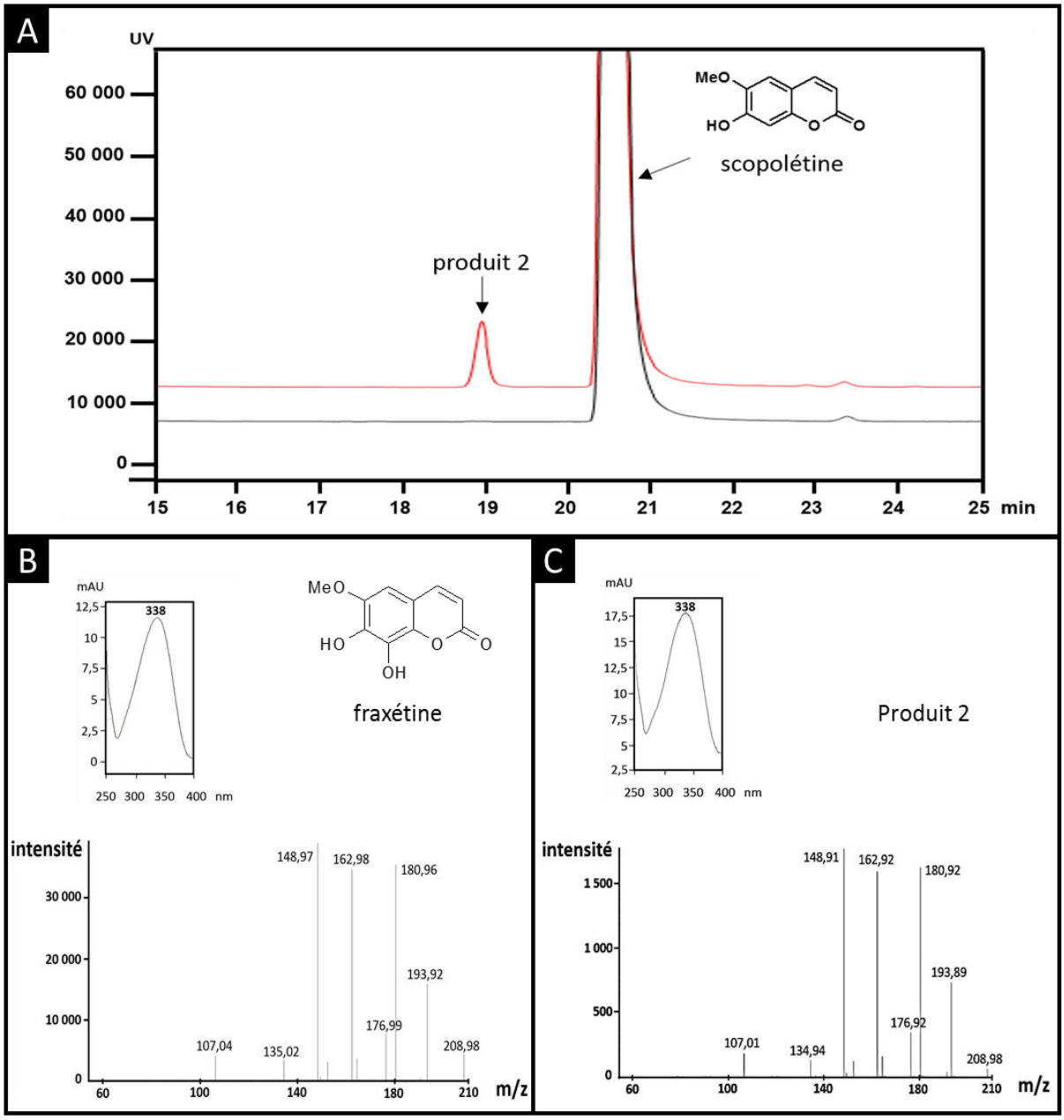


Figure 55 : Détection de la formation de fraxétine à partir de la scopolétine catalysée par CYP71AZ4. A : Détection en HPLC du produit de la métabolisation de la scopolétine en fraxétine. En rouge : avec NADPH, en noir : même incubation sans NADPH. B : Spectre d'absorbance et spectre de masse en mode ionisation positive $(M-H)^+$ du standard fraxétine. C : Spectre d'absorbance et spectre de masse en mode ionisation positive $(M-H)^+$ du produit formé par CYP71AZ4 en présence de scopolétine.

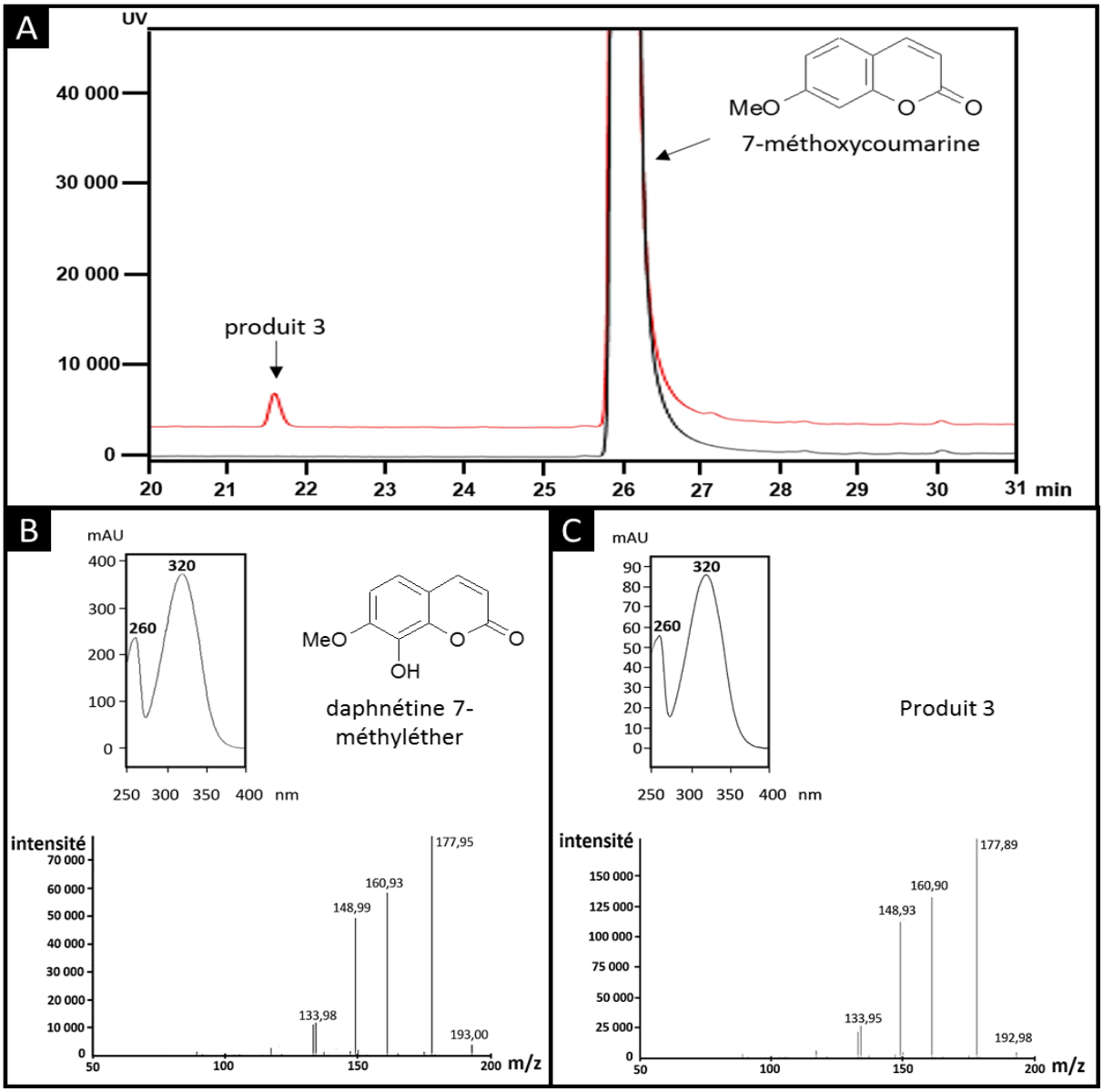


Figure 56 : Détection de la formation de daphnétine 7-méthyléther à partir de la 7-méthoxycoumarine catalysée par CYP71AZ4. A : Détection en HPLC du produit de la métabolisation de la 7-méthoxycoumarine en daphnétine 7-méthyléther. En rouge : avec NADPH, en noir : même incubation sans NADPH. B : Spectre d'absorbance et spectre de masse en mode ionisation positive $(M-H)^+$ du standard daphnétine 7-méthyléther. C : Spectre d'absorbance et spectre de masse en mode ionisation positive $(M-H)^+$ du produit formé par CYP71AZ4 en présence de 7-méthoxycoumarine.

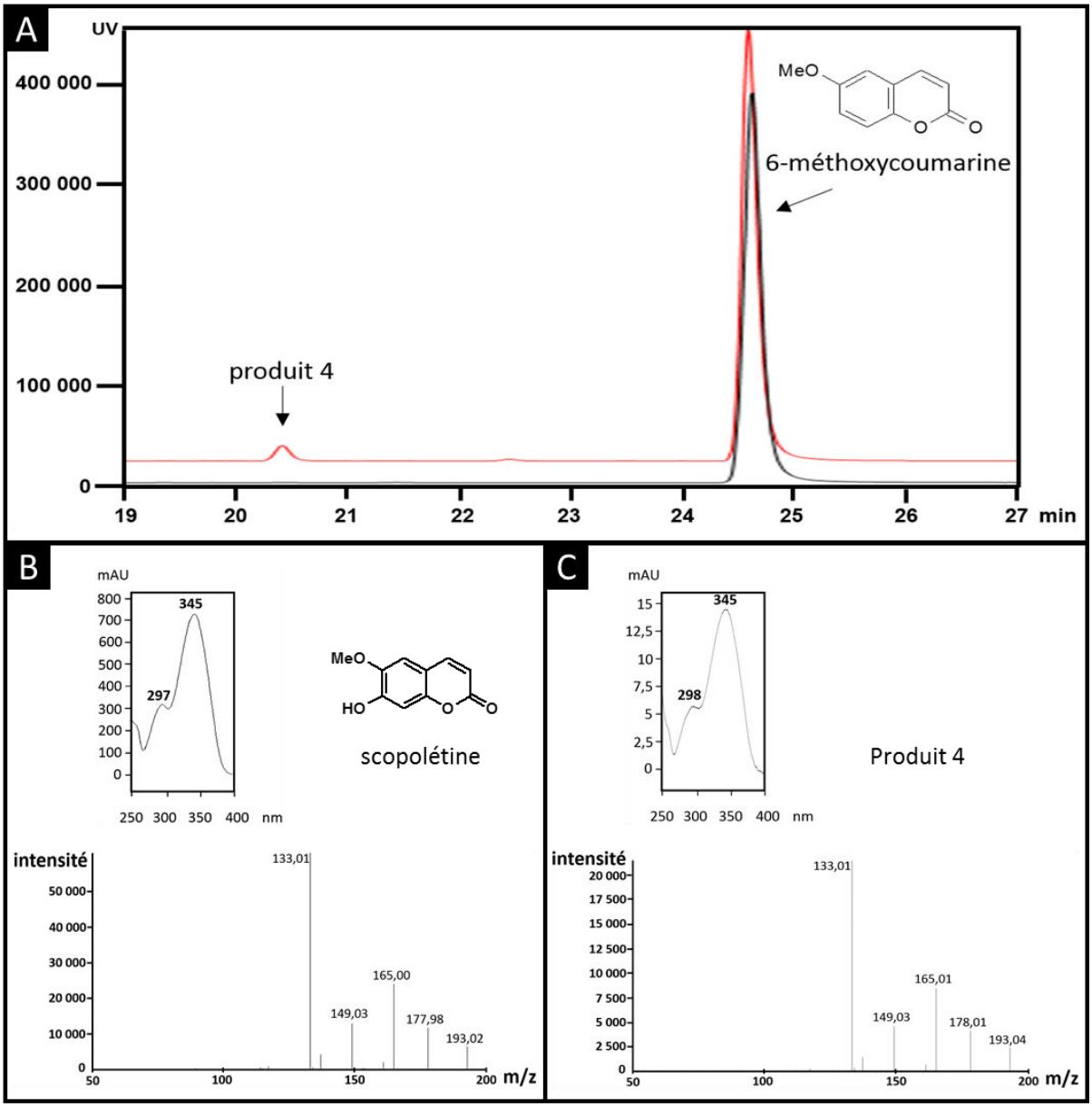


Figure 57 : Détection de la formation de la scopolétine à partir de la 6-méthoxycoumarine catalysée par CYP71AZ4. A : Détection en HPLC du produit de la métabolisation de la 6-méthoxycoumarine en scopolétine. En rouge : avec NADPH, en noir : même incubation sans NADPH. B : Spectre d'absorbance et spectre de masse en mode ionisation positive $(M-H)^+$ du standard scopolétine. C : Spectre d'absorbance et spectre de masse en mode ionisation positive $(M-H)^+$ du premier produit formé par CYP71AZ4 en présence de 6-méthoxycoumarine.

L'incubation de microsomes de levure exprimant CYP71AZ4 en présence de NADPH avec la 6-méthoxycoumarine conduit à l'apparition de deux produits ayant des temps de rétention différents.

Le premier composé formé possède le même temps de rétention (20,52 min) (Figure 57-A, p. 178) ainsi que le même spectre d'absorption que la scopolétine (269 nm – 390 nm, avec deux maxima d'absorption

détectés à 297 nm et 345 nm) (Figure 57-B et C, p. 178). L'analyse de ce composé en LC-MS/MS a permis de confirmer la nature exacte du produit formé qui possède un temps de rétention (16,16 min) et un spectre de fragmentation (masse totale de 192 g/mol avec des pics de fragmentation majoritaires à 132, 148, 164 et 177) identiques à celui du standard commercial de la scopolétine (Figure 57-B et C, p. 178). Le premier produit formé par CYP71AZ4 en présence de NADPH et de 6-méthoxycoumarine est donc de la scopolétine.

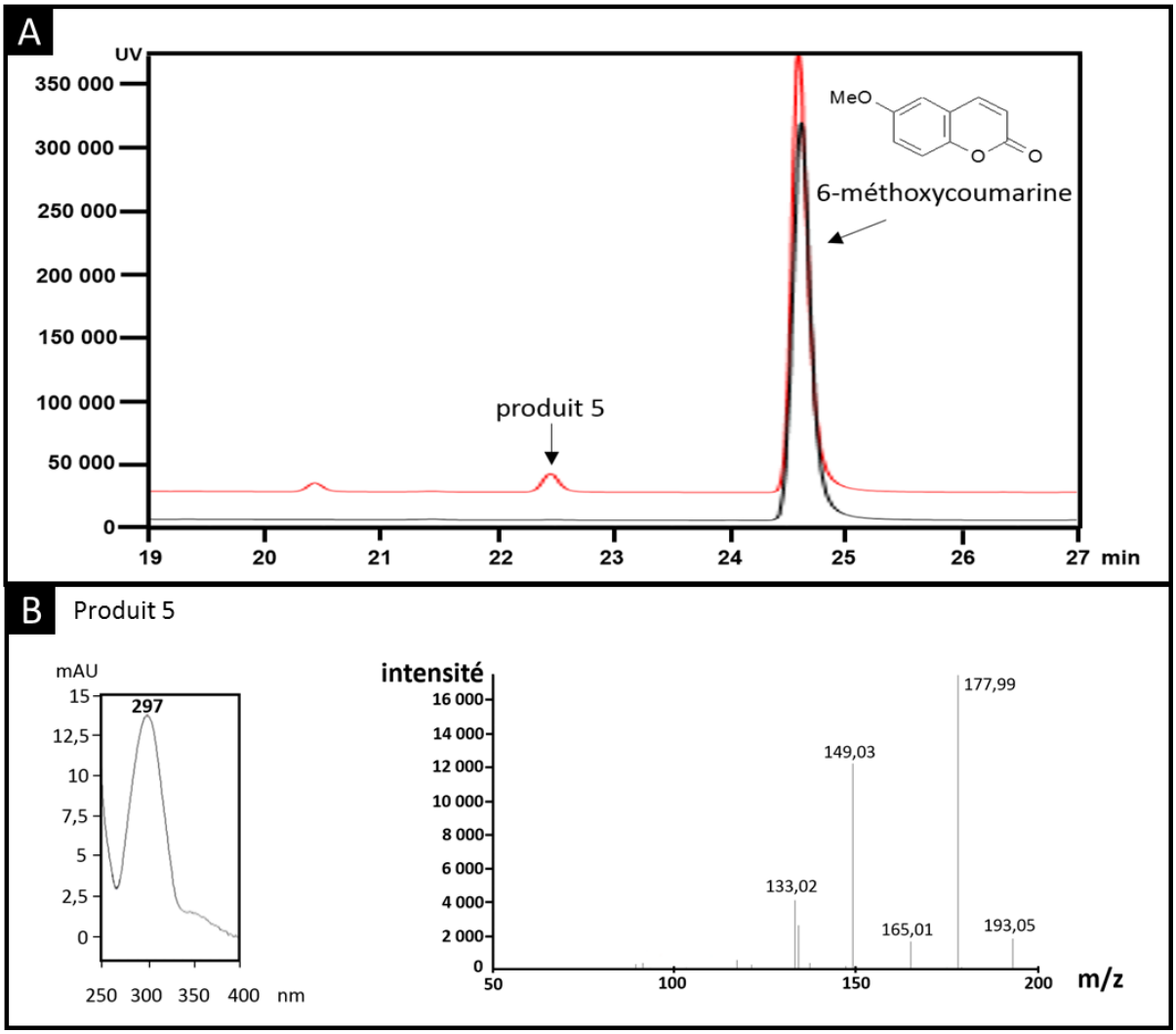


Figure 58 : Détection du deuxième produit de la conversion de la 6-méthoxycoumarine catalysée par CYP71AZ4. A : Détection en HPLC du deuxième produit de la métabolisation de la 6-méthoxycoumarine. En rouge : avec NADPH, en noir : même incubation sans NADPH. B : Spectre d'absorbance et spectre de masse en mode ionisation positive $(M-H)^+$ du deuxième produit formé par CYP71AZ4 en présence de 6-méthoxycoumarine.

En ce qui concerne le deuxième produit formé à partir de la 6-méthoxycoumarine, il possède un temps de rétention de 22,56 min lors de l'analyse par HPLC (Figure 58-A, p. 179). Ce produit possède un

spectre d'absorption de 250 nm à 395 nm avec un maximum d'absorption détecté à 297 nm (Figure 58-B, p. 179). Néanmoins, ces temps de rétention et spectre d'absorption ne correspondent à aucun des substrats testés lors de notre criblage. Nous avons donc décidé d'analyser ce deuxième produit formé à partir de la 6-méthoxycoumarine par LC-MS/MS afin d'identifier plus précisément la nature du produit formé. Cette analyse a permis de confirmer que le produit formé a une masse de 192 g/mol (avec des pics de fragmentation majoritaires à 132, 148, 164 et 177), ce qui correspond à une hydroxylation de la 6-méthoxycoumarine (Figure 58-B, p. 179). Nous avons également analysé par LC-MS/MS tous les substrats disponibles au laboratoire ayant une masse de 192 g/mol (5,7-dihydroxy-4-méthylcoumarine, scopolétine et daphnétine-7-méthyléther). Malheureusement, aucune de ces molécules ne possède les mêmes caractéristiques (temps de rétention, spectre d'absorption, spectre de masse) que le produit formé. L'analyse en LC-MS/MS n'a donc pas permis de mettre en évidence la nature exacte du produit formé et la position de l'hydroxylation reste à déterminer. CYP71AZ4 est donc capable, en présence de NADPH, d'hydroxyler la 6-méthoxycoumarine en deux positions différentes conduisant à la formation de scopolétine et d'un produit dont la nature exacte reste à déterminer. Ce produit inconnu pourrait vraisemblablement être la 6-méthoxy-8-hydroxycoumarine, mais cette hypothèse devra être confirmée par des analyses RMN.

Enfin, lorsque l'on incube des microsomes de levure exprimant CYP71AZ4 en présence de NADPH avec de la 7-méthoxy 3-méthylcoumarine, un produit apparait. Ce produit présente un temps de rétention de 24,67 min (Figure 59-A, p. 181) et un spectre d'absorption de 250 nm à 379 nm avec deux maxima d'absorption détectés à 261 nm et 319 nm (Figure 59-B, p. 181). Néanmoins, il n'a pas été possible de corréler ce temps de rétention et ce spectre d'absorption avec ceux des molécules disponibles au laboratoire. Nous avons donc analysé ce composé en LC-MS/MS et les résultats obtenus indiquent que ce produit présente un temps de rétention de 19,18 min. Cette analyse a montré que le produit formé possède une masse de 206 g/mol avec des pics de fragmentation majoritaires à 146, 163, 174 et 191, ce qui correspond à une hydroxylation de la 7-méthoxy 3-méthylcoumarine (Figure 59-B, p. 181). Nous avons également analysé en LC-MS/MS tous les substrats disponibles au laboratoire ayant une masse de 206 g/mol (daphnétine diméthyléther, scoparone et limettine), mais aucune de ces molécules ne possède les mêmes caractéristiques (temps de rétention, spectre d'absorption, spectre de masse) que le produit formé. La position exacte de l'hydroxylation de la 7-méthoxy 3-méthylcoumarine reste donc à déterminer par des analyses RMN. CYP71AZ4 est donc également capable d'hydroxyler la 7-méthoxy 3-méthylcoumarine.

Ces résultats indiquent donc que CYP71AZ4, en plus de sa capacité à synthétiser du xanthotoxol à partir de psoralène est également en mesure de métaboliser différentes coumarines (scopolétine, 7-méthoxycoumarine, 6-méthoxycoumarine, 3-méthyl,7-méthoxycoumarine) avec visiblement une efficacité différente.

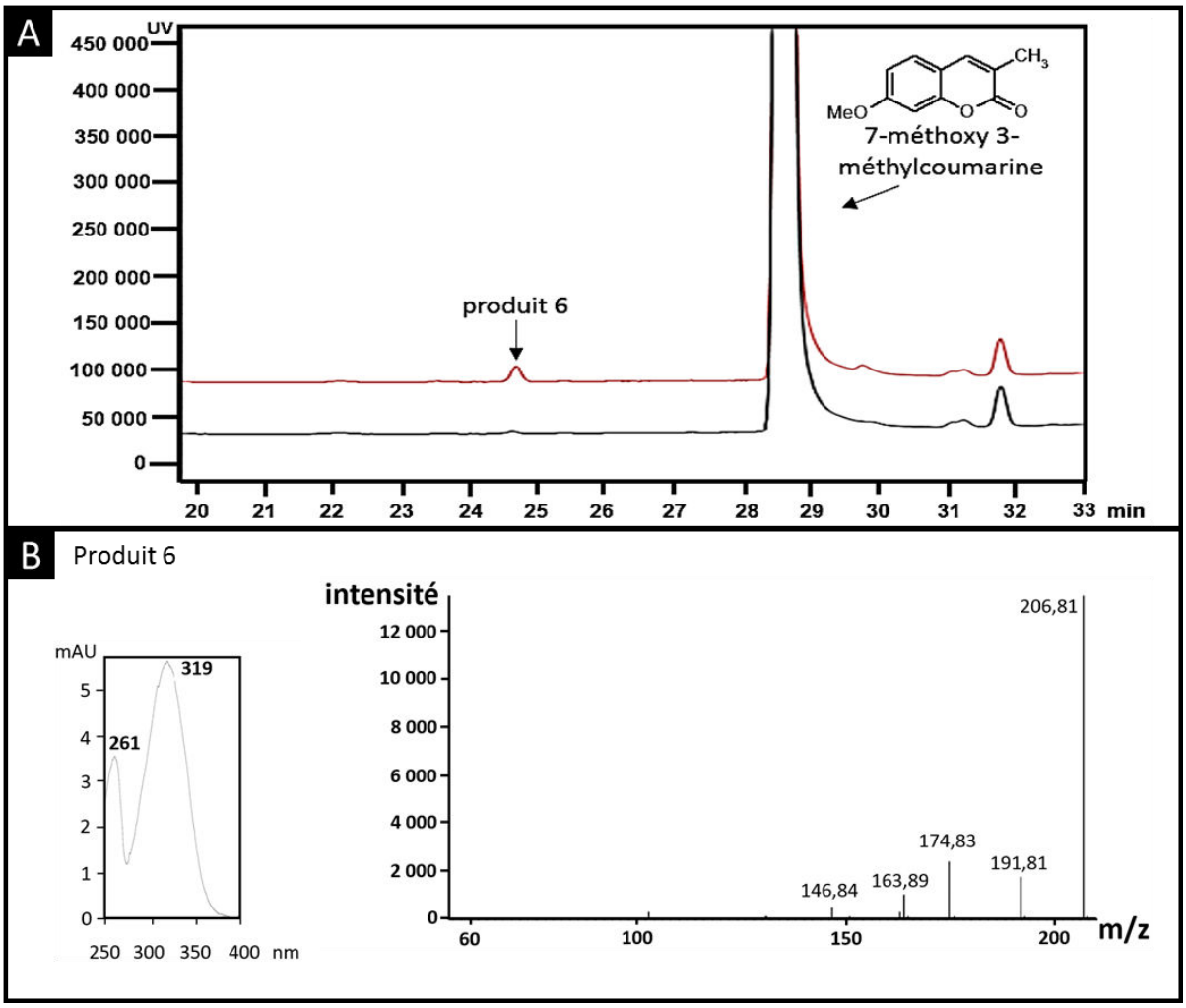


Figure 59 : Détection du produit de la conversion de la 7-méthoxy 3-méthylcoumarine catalysée par CYP71AZ4. A : Détection en HPLC du produit de la métabolisation de la 7-méthoxy 3-méthylcoumarine par CYP71AZ4. En rouge : avec NADPH, en noir : même incubation sans NADPH. B : Spectre d'absorbance et spectre de masse en mode ionisation positive $(M-H)^+$ du produit formé par CYP71AZ4 en présence de 7-méthoxy 3-méthylcoumarine.

CYP71AZ3 et métabolisation de coumarines

Les tests de métabolisation effectués avec des microsomes de levure exprimant CYP71AZ3 ont permis de mettre en évidence la formation d'un produit en présence d'esculétine et de cofacteur (Figure 60-A,

p. 182). Ce produit formé n'apparaît qu'en présence de NADPH (Figure 60-A, p. 182) et n'est pas produit lorsque l'on incube des microsomes de levure transformées avec le vecteur p*Ye*DP60 vide (en absence ou en présence de cofacteur). Le produit ainsi formé possède un temps de rétention de 18,76 min et un spectre d'absorption de 230 à 380 nm avec un maximum d'absorption détecté à 330 nm (Figure 60-B, p. 182). Le produit formé a été analysé par LC-MS/MS afin de déterminer son identité selon la méthode décrite dans le chapitre « Matériel et Méthodes ». Cette analyse a montré que le produit formé (16,25 min) présente une masse totale de 196 g/mol avec des pics majoritaires de fragmentation à 138, 148, 150, 162 et 166 ce qui correspond à une hydroxylation de l'esculétine (Figure 60-B, p. 182). Il n'a pas été possible de comparer ce spectre de masse avec ceux des molécules disponibles au laboratoire, car aucune des molécules disponibles ne possède une masse de 196 g/mol. La position de l'esculétine où s'effectue l'hydroxylation reste donc à déterminer par une analyse RMN.

Néanmoins, ces résultats démontrent que CYP71AZ3 est capable d'hydroxyler l'esculétine.

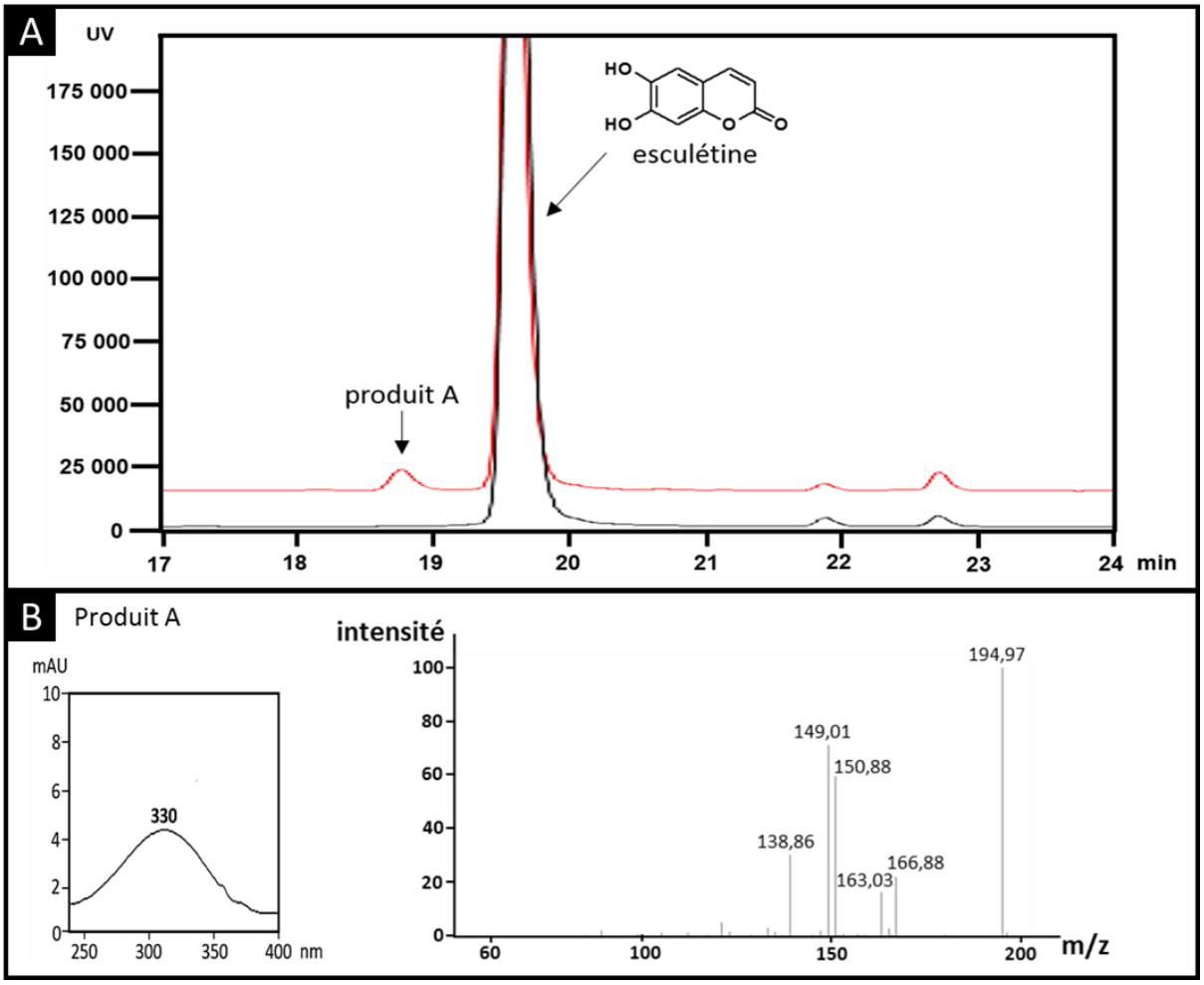


Figure 60 : Détection du produit de la conversion de l'esculétine catalysée par CYP71AZ3. A : Détection en HPLC du produit de la métabolisation de l'esculétine par CYP71AZ3. En rouge : avec NADPH, en noir : même incubation sans NADPH. B : Spectre d'absorbance et spectre de masse en mode ionisation positive $(M-H)^+$ du produit formé par CYP71AZ3 en présence d'esculétine.

2.4.2.2 Détermination des constantes cinétiques de CYP71AZ3 et CYP71AZ4

Le criblage métabolique réalisé avec la totalité des substrats disponibles au laboratoire a donc permis de mettre en évidence plusieurs substrats pour CYP71AZ*4* (psoralène, scopolétine, 7-méthoxycoumarine, 6-méthoxycoumarine et 7-méthoxy 3-méthylcoumarine), alors que l'esculétine est le seul substrat métabolisé par CYP71AZ3. Afin de réaliser la caractérisation enzymatique de ces deux enzymes, les conditions optimales de pH et de température sont déterminées dans un premier temps. Puis, la détermination des paramètres de la linéarité des réactions ainsi que les constantes cinétiques de ces deux enzymes par rapport à leurs substrats seront identifiées. Enfin, une recherche d'inhibition compétitive de CYP71AZ3 et CYP71AZ*4* vis-à-vis des coumarines et des furanocoumarines sera initiée.

2.4.2.2.1 Détermination du pH optimal d'activité

Afin de mesurer le pH optimal d'activité de CYP71AZ3 et CYP71AZ*4*, des milieux réactionnels ont été préparés comme cela a été décrit dans le chapitre « Matériel et Méthodes » excepté pour le pH du tampon NaPi 0,1 M. Une gamme de pH du tampon NaPi s'étendant de 5,8 à 8,0 (5,8 ; 6 ; 6,2 ; 6,4 ; 6,6 ; 6,8 ; 7,0 ; 7,2 ; 7,4 ; 7,6 ; 7,8 et 8) a été utilisée pour réaliser les différentes incubations. Tous les autres paramètres réactionnels restent inchangés et les incubations ont été réalisées comme cela a été décrit précédemment. La détermination du pH optimal d'activité de CYP71AZ3 et de CYP71AZ4 a été réalisée respectivement avec l'esculétine et le psoralène. Toutes les réactions ont été réalisées en triplicats.

CYP71AZ*4* présente une activité sur toute la gamme de pH testée et son optimum d'activité est obtenu à pH 7,3 (Figure 61, p. 184). De la même manière, une activité est détectée pour CYP71AZ3 pour la gamme de pH testée, mais l'optimum de pH n'a pas été atteint avec le pH le plus élevé testé (pH 8,0) (Figure 62, p. 184).

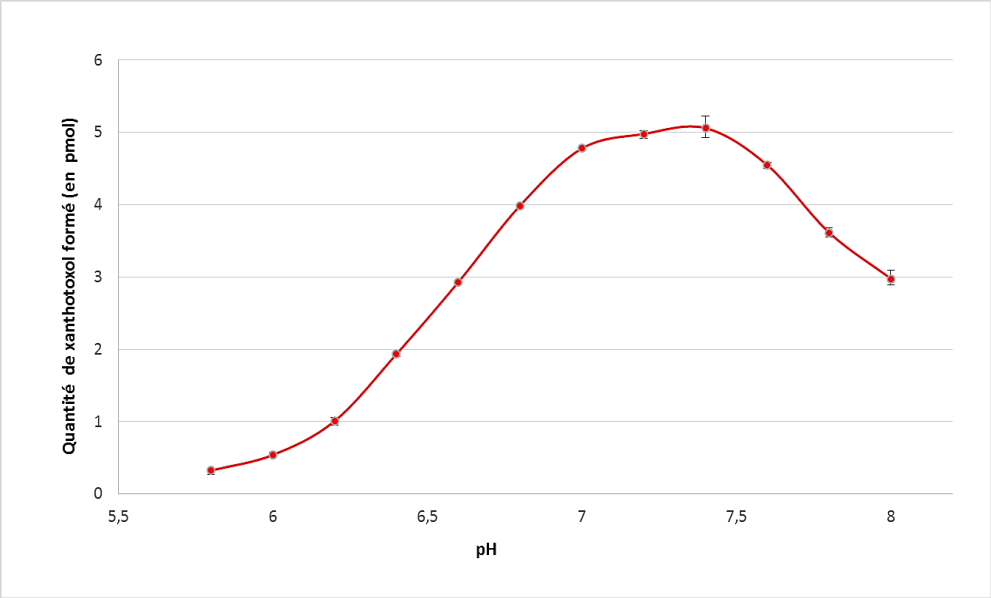


Figure 61 : Détermination du pH optimum pour l'activité de la psoralène 8-monooxygénase.

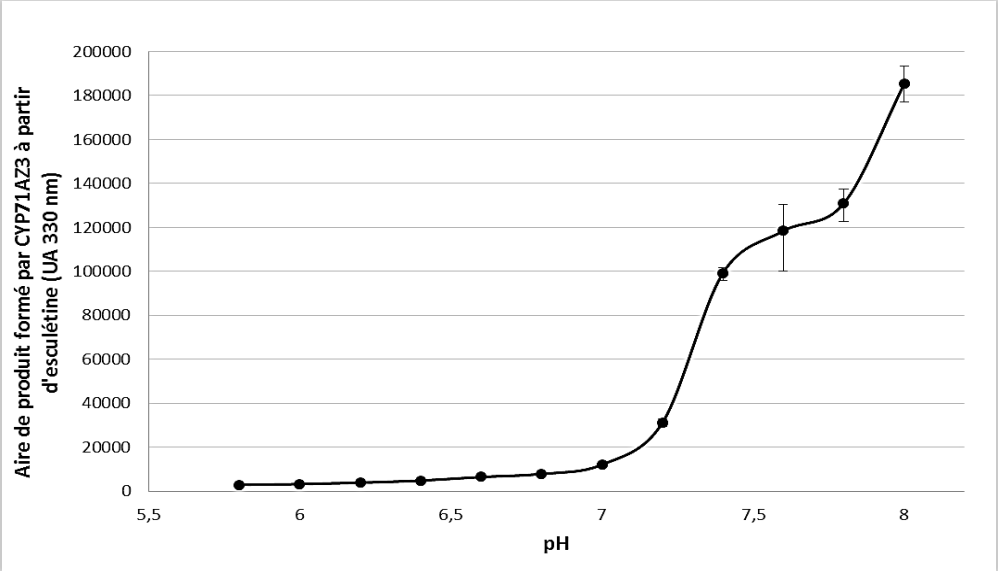


Figure 62 : Détermination du pH optimum d'activité de CYP71AZ3 en présence d'esculétine. Les aires de produit formé ont été mesurées à 330 nm.

2.4.2.2.2 Détermination de la température optimale d'activité

Les températures optimales d'activité pour CYP71AZ3 et CYP71AZ4 ont été déterminées en incubant les différents milieux réactionnels contenant respectivement du psoralène et de l'esculétine à différentes

températures (de 15 à 45 °C). Les incubations ont été réalisées à pH 7,5 comme cela a été décrit dans le chapitre « Matériel et Méthodes ». Toutes les réactions ont été réalisées en triplicats. Les activités maximales de CYP71AZ3 et CYP71AZ4 ont été mesurées pour des températures respectivement comprises entre 25 et 30 °C et entre 25 et 28 °C (Figure 63, p. 185 et Figure 64, p. 185).

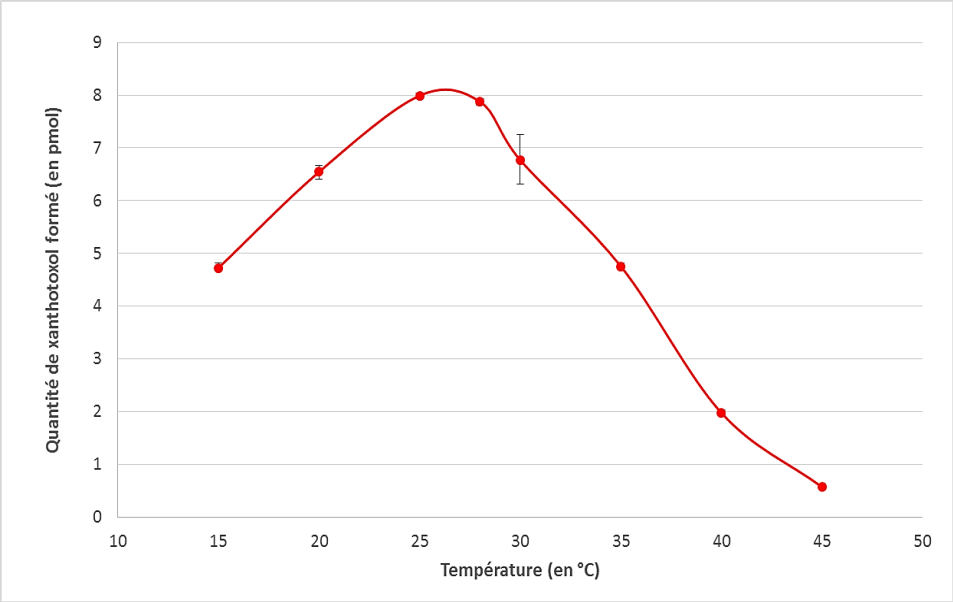


Figure 63 : Détermination de la température optimale d'activité de CYP71AZ4 en présence de psoralène.

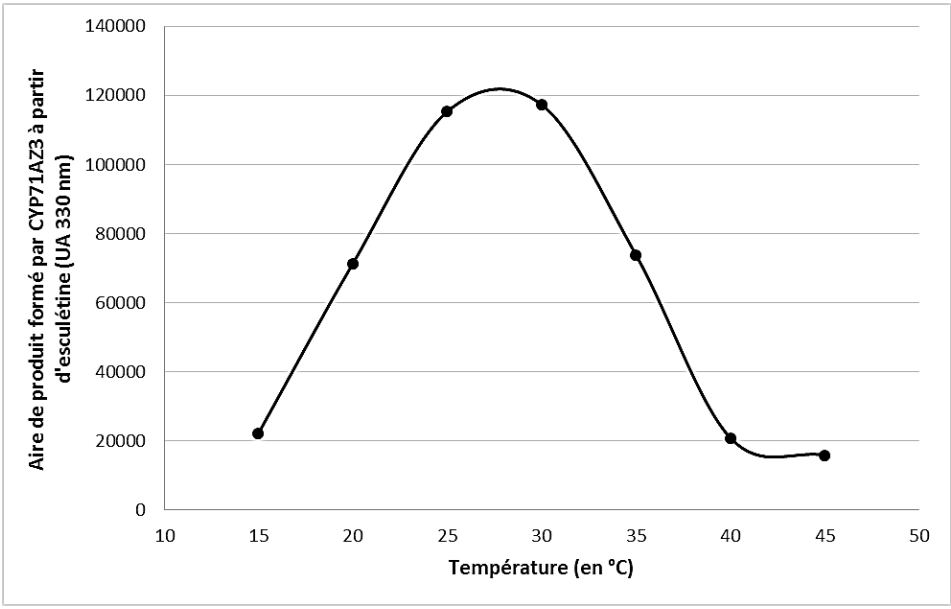


Figure 64 : Détermination de la température optimale d'activité de CYP71AZ3 en présence d'esculétine. Les aires de produit formé ont été mesurées à 330 nm.

2.4.2.2.3 Linéarité de la réaction

Afin de déterminer la linéarité des réactions catalysées par CYP71AZ3 et CYP71AZ4, des milieux réactionnels, contenant les différents substrats de ces enzymes, ont été préparés comme cela a été décrit dans le chapitre « Matériel et Méthodes ». La linéarité des réactions a été mesurée pour les deux enzymes sur une période de temps de 30 min. La formation des différents produits par CYP71AZ4 est linéaire pendant 2 min 30 sec (Figure 65 à Figure 70, p. 186 à p. 188). La formation de xanthotoxol, de daphnétine 7-méthyléther et du produit de métabolisation de la 7-méthoxy 3-méthylcoumarine décroissent par la suite pour atteindre un plateau après 20 min de réaction (Figure 65 p. 186, Figure 67 p. 187 et Figure 70 p. 188). La formation de fraxétine, scopolétine et du deuxième produit formé à partir de la 6-méthoxycoumarine décroissent également après 2 min 30 sec, mais le plateau de la réaction n'est pas encore atteint après 30 min de réaction (Figure 66 p. 187, Figure 68 p. 187 et Figure 70 p. 188). Concernant CYP71AZ3, la formation du produit à partir de l'esculétine est linéaire sur une période de 30 min (Figure 71, p. 188).

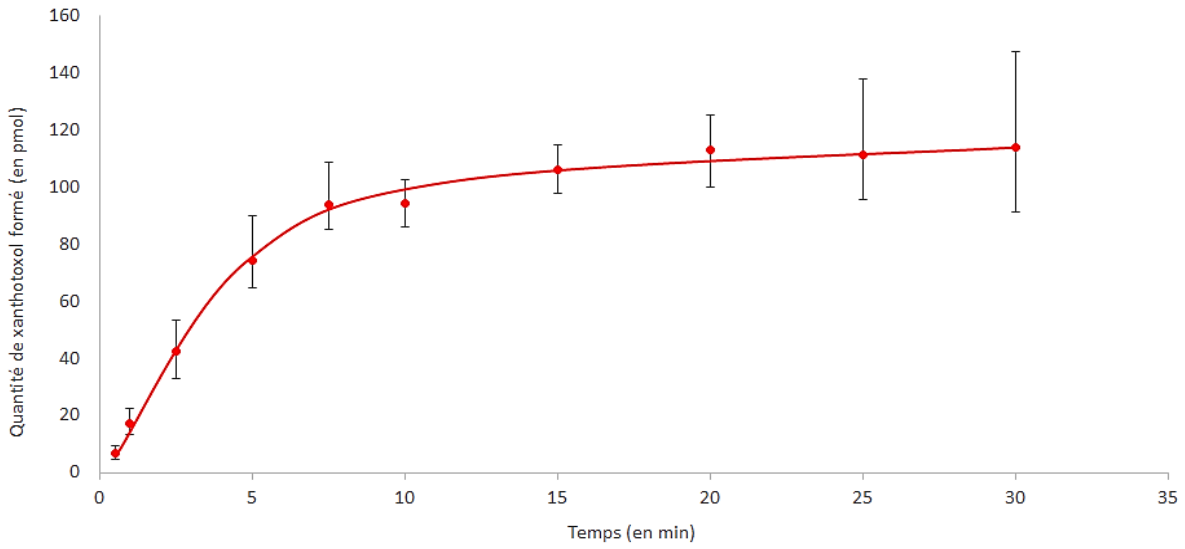


Figure 65 : Accumulation de xanthotoxol à partir de psoralène par CYP71AZ4 en fonction du temps.

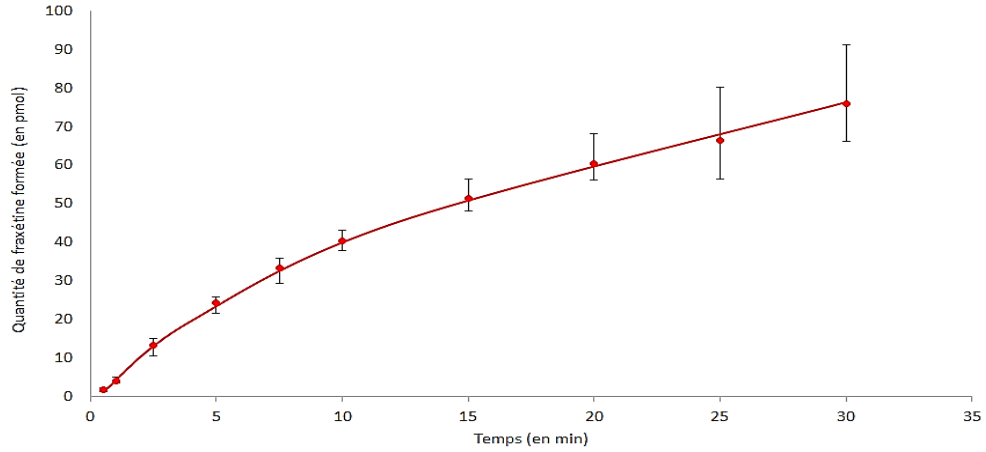


Figure 66 : Accumulation de fraxétine à partir de scopolétine par CYP71AZ4 en fonction du temps.

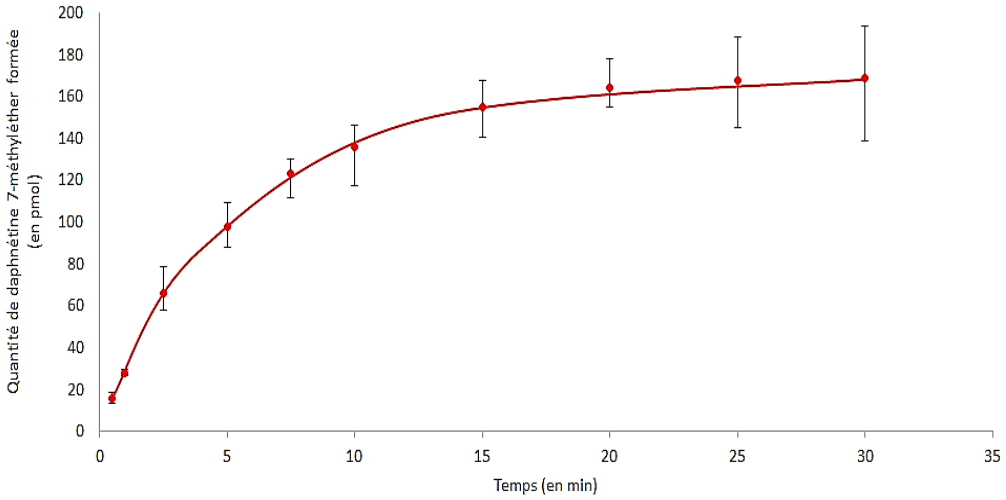


Figure 67 : Accumulation de daphnétine 7-méthyléther à partir de 7-méthoxycoumarine par CYP71AZ4 en fonction du temps.

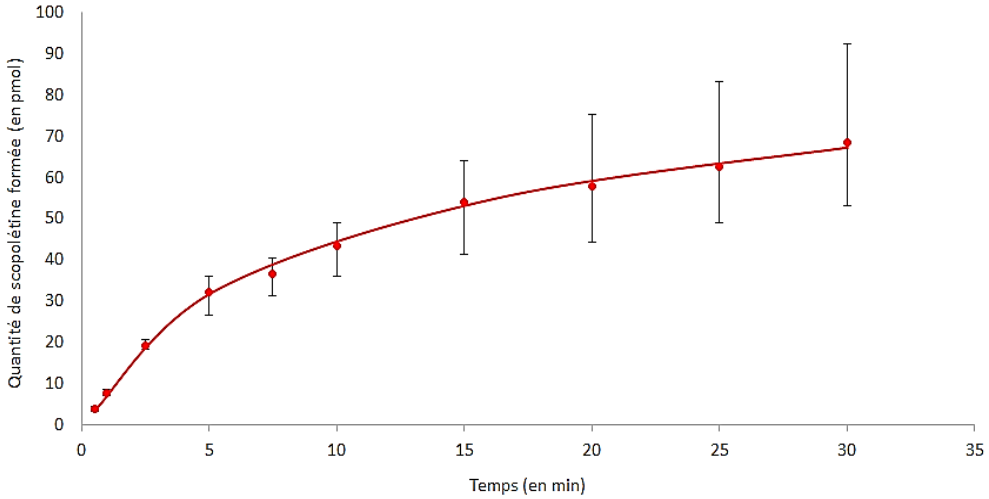


Figure 68 : Accumulation de scopolétine à partir de 6-méthoxycoumarine par CYP71AZ4 en fonction du temps.

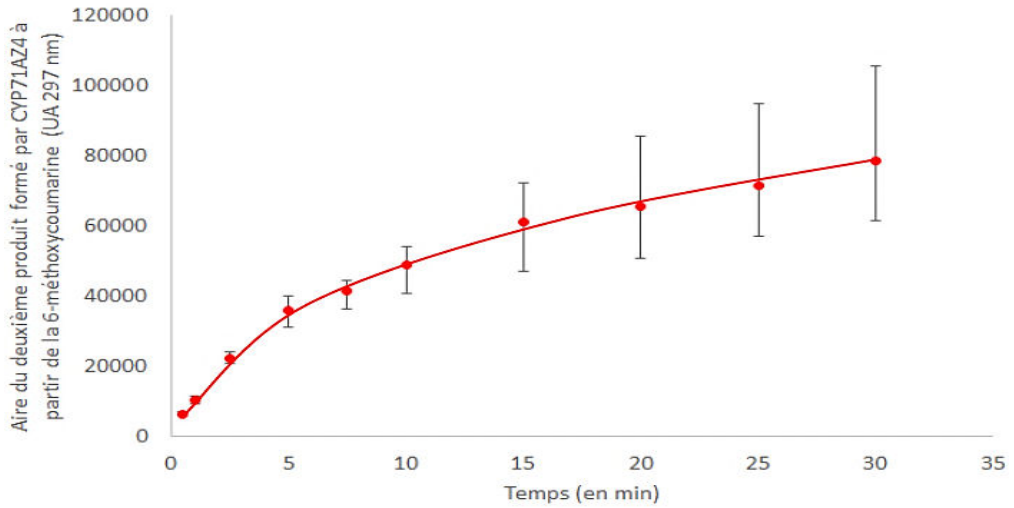


Figure 69 : Accumulation du deuxième produit formé par CYP71AZ4 à partir de la 6-méthoxycoumarine en fonction du temps.

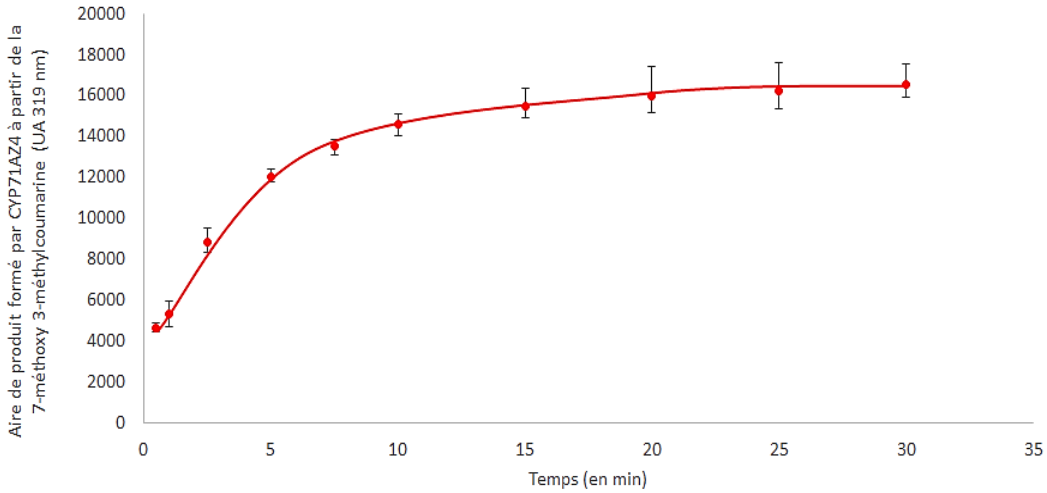


Figure 70 : Accumulation du produit formé par CYP71AZ4 à partir de la 7-méthoxy 3-méthylcoumarine en fonction du temps.

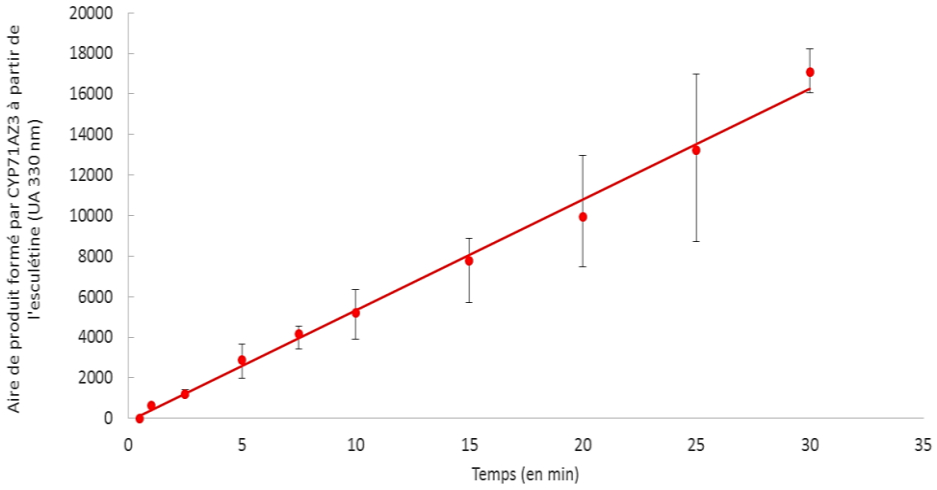


Figure 71 : Accumulation du produit formé par CYP71AZ3 à partir de l'esculétine en fonction du temps.

2.4.2.2.4 Constantes cinétiques

Les conditions d'activité optimales définies précédemment pour CYP71AZ3 et CYP71AZ4 ont été utilisées afin de déterminer les constantes enzymatiques de ces deux enzymes. Les réactions sont conduites dans 76 µl de tampon NaPi pH 7,5 à 0,1 M contenant 200 µM de NADPH et une gamme de concentrations en substrats (de 0,1 à 200 µM). Les incubations ont été réalisées à 28 °C sous agitation à 200 rpm pour CYP71AZ4 et CYP71AZ3 respectivement pendant 1 min 30 sec et 20 min. Les réactions ont ensuite été stoppées par l'ajout de 37,5 µL d'acétonitrile/HCl (99/1). Chaque réaction a été réalisée en triplicats. Les résultats ainsi obtenus représentent la moyenne des triplicats et révèlent des courbes aux allures d'hyperboles qui sont caractéristiques des enzymes de type Michaeliennes (CYP71AZ4 : Figure 72 à Figure 77 p. 190 à p. 191; CYP71AZ3 : Figure 78 p. 192).

Enzyme	Substrat	Produit	K_m (µM)
CYP71AZ3	Esculétine	Inconnu	248,6 ± 51,9
CYP71AZ4	Psoralène	Xanthotoxol	2,1 ± 0,3
	Scopolétine	Fraxétine	762,2 ± 586,4
	7-méthoxycoumarine	Daphnétine 7-méthyléther	72,5 ± 15,2
	6-méthoxycoumarine	Scopolétine	9,5 ± 1,5
	6-méthoxycoumarine	Inconnu	44,8 ± 16,3
	7-méthoxy 3-méthylcoumarine	Inconnu	159,9 ± 41,1

Tableau 17 : Constantes d'affinités de CYP71AZ3 et de CYP71AZ4 calculées à l'aide du logiciel Sigmaplot.

Les constantes d'affinité (K_m) de CYP71AZ3 pour l'esculétine et de CYP71AZ4 pour ses cinq substrats (psoralène, scopolétine, 6-méthoxycoumarine, 7-méthoxycoumarine et 7-méthoxy 3-méthylcoumarine) ont ensuite été déterminées à partir des courbes obtenues à l'aide du logiciel Sigmaplot et sont représentées dans le Tableau 17 (p. 189). D'après les résultats obtenus, **le substrat préférentiel de CYP71AZ4 est le psoralène avec un K_m de 2,1 ± 0,3 µM,** puis la 6-méthoxycoumarine, la 7-méthoxycoumarine, la 7-méthoxy 3-méthylcoumarine et enfin la scopolétine. **Le K_m pour CYP71AZ3 et l'esculétine est de 248,6 ± 51,9 µM.** Etant donné que cette constante d'affinité est assez élevée, il est vraisemblable que l'esculétine ne soit pas le substrat préférentiel de CYP71AZ3.

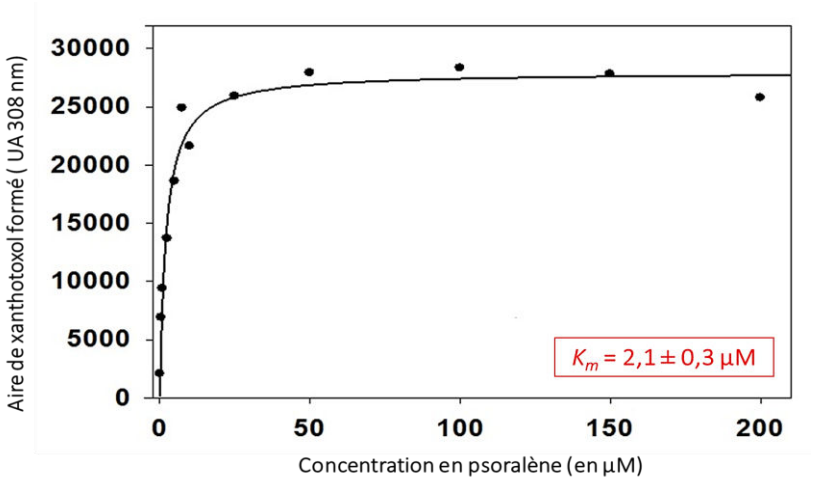


Figure 72 : Quantité de xanthotoxol formé par CYP71AZ4 en fonction de la concentration en psoralène.

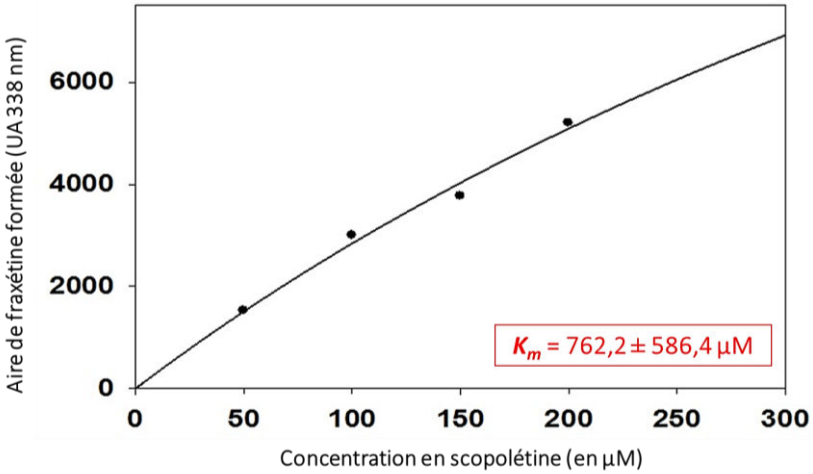


Figure 73 : Quantité de fraxétine formée par CYP71AZ4 en fonction de la concentration en scopolétine.

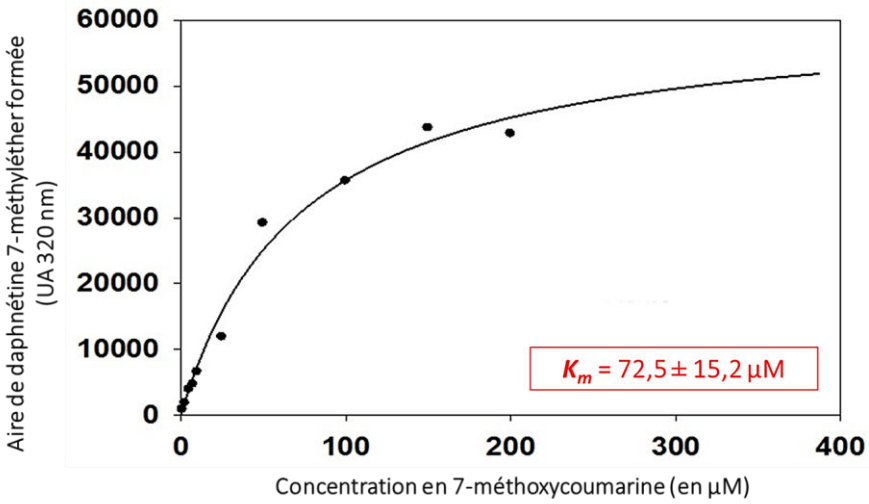


Figure 74 : Quantité de daphnétine 7-méthyléther formée par CYP71AZ4 en fonction de la concentration en 7-méthoxycoumarine.

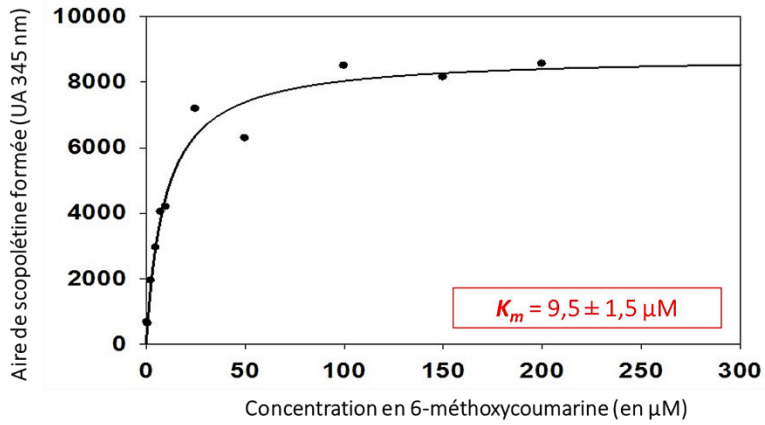


Figure 75 : Quantité de scopolétine formée par CYP71AZ*4* en fonction de la concentration en 6-méthoxycoumarine.

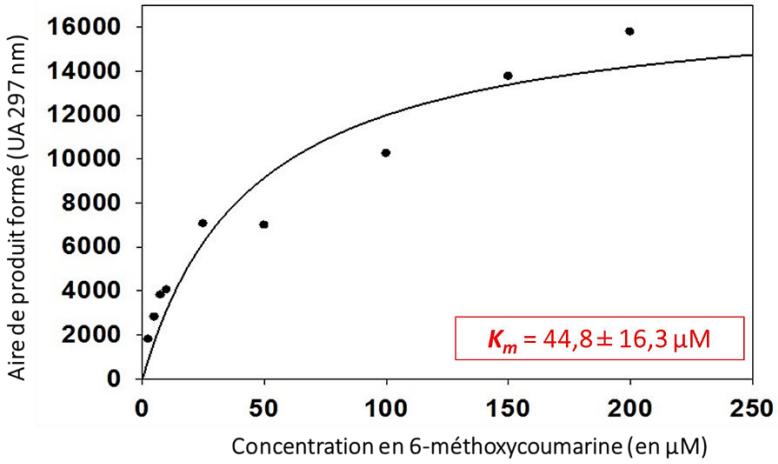


Figure 76 : Quantité de produit inconnu formé par CYP71AZ4 en fonction de la concentration en 6-méthoxycoumarine.

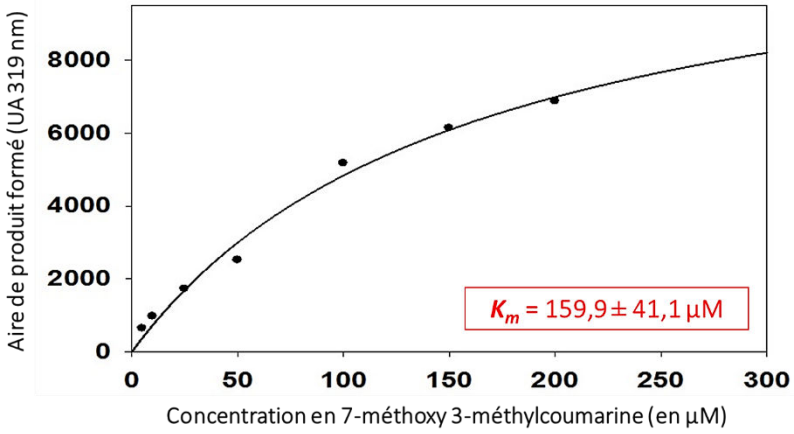


Figure 77 : Quantité de produit inconnu formé par CYP71AZ4 en fonction de la concentration en 7-méthoxy 3-méthylcoumarine.

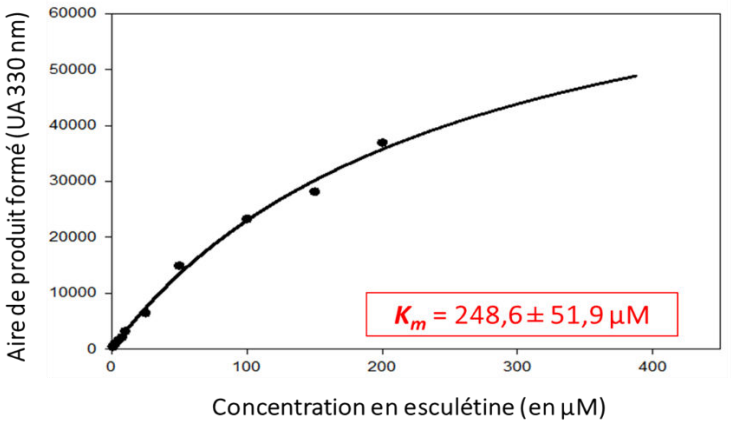


Figure 78 : Quantité de produit inconnu formé par CYP71AZ3 en fonction de la concentration en esculétine.

2.4.2.2.5 Recherche d'une inhibition compétitive de CYP71AZ3 et CYP71AZ4 par des coumarines et furanocoumarines

Une recherche d'inhibition compétitive de CYP71AZ3 et CYP71AZ4 par des coumarines et des furanocoumarines a été initiée en incubant 20 µl de microsomes de levure exprimant respectivement ces deux enzymes pendant 30 minutes à 28 °C en présence de 200 µM de NADPH, 100 µM d'esculétine ou de psoralène et 100 µM d'inhibiteur. Des réactions similaires ont été effectuées en remplaçant la solution contenant l'inhibiteur par une solution de DMSO (qui est le solvant utilisé pour préparer les solutions d'inhibiteurs) afin de constituer un témoin négatif. Un total de 13 coumarines, 15 furanocoumarines et une furoquinoline alcaloïde a été testé pour CYP71AZ3 et 8 coumarines, 12 furanocoumarines et une furoquinoline ont été testées comme inhibiteur pour CYP71AZ4. Les activités obtenues après incubations de CYP71AZ4 avec du psoralène et de CYP71AZ3 avec de l'esculétine en absence d'inhibiteur constituent les témoins de cette expérience d'inhibition et sont considérés comme ayant une activité relative de 100 %. Les résultats de la recherche d'inhibiteurs de CYP71AZ3 et CYP71AZ4 sont représentés dans le Tableau 18 (p. 193).

Inhibiteur	CYP71AZ3	CYP71AZ4
6-méthoxycoumarine	100 %	/
7-méthoxycoumarine	94,09 %	/
Coumarine	76,38 %	/
Daphnétine	30,06 %	69,63 %
Daphnétine 7-méthyléther	95,46 %	88,28 %
Daphnétine diméthyléther	100 %	/
Esculétine	/	57,91 %
Fraxétine	16,45 %	70,81 %
Isoscopolétine	100 %	81,50 %
Limettine	100 %	78,24 %
Osthole	95,62 %	59,21 %
Scoparone	97,65 %	/
Scopolétine	100 %	/
Umbelliférone	89,90 %	87,78 %
Angélicine	100 %	82,32 %
Bergaptol	95,10 %	89,94 %
Bergaptène	100 %	95,18 %
Déméthylsubérosine	100 %	91,76 %
Isobergaptol	100 %	70,26 %
Isobergaptène	100 %	99,11 %
Isopimpinelline	95,69 %	93,46 %
Marmésine	99,77 %	77,43 %
Osthénol	99,68 %	/
Pimpinelline	100 %	94,29 %
Psoralène	100 %	/
Sphondine	100 %	92,74 %
Sphondinol	100 %	/
Xanthotoxine	100 %	100 %
Xanthotoxol	84,90 %	/
Dictamnine	100 %	82,17 %
Sans inhibiteur	100 %	100 %

Tableau 18 : Recherche d'inhibition compétitive des coumarines et furanocoumarines sur CYP71AZ3 et CYP71AZ4. Les résultats sont exprimés en pourcentage d'activité résiduelle par rapport au témoin sans inhibiteur.

Parmi toutes les molécules testées, seules la daphnétine et la fraxétine démontrent une forte activité inhibitrice de CYP71AZ3 avec respectivement 30,06 et 16,45 % d'activité résiduelle par rapport au

témoin sans inhibiteur (Tableau 18, p. 193). De la même manière, les molécules qui démontrent les plus fortes activités inhibitrices de CYP71AZ4 sont deux coumarines : l'osthole (59,21 % d'activité résiduelle) et l'esculétine (57,91 % d'activité résiduelle). Ces résultats restent néanmoins des résultats préliminaires et nécessitent d'être complétés par des mesures de constantes cinétiques de CYP71AZ3 et CYP71AZ4 en présence de leurs inhibiteurs à différentes concentrations afin de vérifier que l'addition de l'inhibiteur modifie bien l'affinité de l'enzyme pour son substrat sans changer la vitesse d'activité maximale qui sont des propriétés caractéristiques d'une inhibition compétitive. Si l'inhibition compétitive est confirmée par des expériences ultérieures, cela démontrera que ces substrats sont capables de rentrer dans le site actif de ces deux enzymes.

2.5 Caractérisation fonctionnelle de CYP71AZ3 et CYP71AZ4 *in vivo*

Les premiers résultats obtenus « *in vitro* » indiquent que CYP71AZ4 est impliqué dans l'hydroxylation de plusieurs substrats (psoralène, scopolétine, 6-méthoxycoumarine, 7-méthoxycoumarine et 7-méthoxy 3-méthylcoumarine), alors que CYP71AZ3 est capable d'hydroxyler uniquement l'esculétine. Afin de démontrer que ces fonctions ne sont pas artéfactuelles, mais bien représentatives de ce qui se passe *in planta*, CYP71AZ3 et CYP71AZ4 a été exprimé chez *N. benthamiana*, une plante qui, à l'état naturel, n'est pas capable de synthétiser du psoralène, de la 6-méthoxycoumarine, de la 7-méthoxycoumarine et de la 7-méthoxy 3-méthylcoumarine.

2.5.1 Expression transitoire de CYP71AZ3 et CYP71AZ4 dans un système d'expression *in planta* : *N. benthamiana*

Le système d'expression transitoire chez *N. benthamiana*, déjà décrit dans la littérature (Sparkes *et al*., 2006), est utilisé et a été optimisé au sein de notre laboratoire pour l'expression de prényltransférases et de dioxygénases. Des tests préliminaires ont été réalisés au laboratoire par le Dr Alexandre Olry pour tenter d'exprimer des enzymes de la famille des P450s.

2.5.1.1 Test préliminaire avec CYP71AJ1 et l'ATR1 d'*A. thaliana*

Le P450 utilisé pour optimiser la méthode d'expression transitoire chez le tabac est CYP71AJ1, isolé chez *A. majus* et qui code la psoralène synthase. Dans un premier temps, la séquence codante de CYP71AJ1 contenue dans le pCR8/TOPO/GW a été transférée par LR recombinaison dans le plasmide binaire pBIN_GW® comme cela est décrit dans le chapitre « Matériel et méthodes ». Des agrobactéries ont ensuite été transformées avec le plasmide recombinant, sélectionnées sur milieu sélectif contenant

de la kanamycine et de la rifampicine, puis mises en culture pendant 24 heures. En parallèle, des agrobactéries exprimant la protéine p19 ont également été mises en culture pendant 24 h. Les faces abaxiales des deuxième et troisième feuilles les plus jeunes de *N. benthamiana* ont été co-infiltrées avec les inocula bactériens exprimant CYP71AJ1 et la protéine p19. Quatre jours après l'inoculation, les feuilles ayant été infiltrées ont subi une deuxième infiltration avec une solution contenant de la marmésine à 100 µM. Trois heures après la deuxième infiltration, les feuilles infiltrées sont broyées à l'azote liquide et les phénylpropanoïdes sont extraits comme décrit dans le chapitre « Matériel et Méthodes » et les extraits sont analysés en LC-MS/MS. Les résultats obtenus en LC-MS/MS permettent de confirmer la présence de marmésine dans les feuilles de tabac, néanmoins la production de psoralène par CYP71AJ1 n'a pas été mise en évidence.

Une optimisation de l'expression transitoire de P450s chez le tabac a été initiée en co-exprimant le P450 avec l'ATR1 d'*A. thaliana* afin de restaurer la stœchiométrie entre l'expression du P450 et de la CPR. Pour cela, le gène codant pour l'ATR1 a été cloné dans le sens direct dans le plasmide pCR8/TOPO/GW, puis transféré dans le pBIN_GW® par LR recombinaison et les plasmides recombinants ainsi obtenus ont été utilisés pour transformer des agrobactéries. Le gène CYP71AJ1 a été co-exprimé chez *N. benthamiana* avec la protéine p19 et l'ATR1 d'*A. thaliana*. Pour cela, trois inocula d'agrobactéries recombinantes ont été préparés : le premier contient le gène codant de la protéine p19, le deuxième avec l'ATR1 et le troisième contenant CYP71AJ1. Un premier lot de plantes de tabac a été co-infiltré avec la protéine p19 et CYP71AJ1 et un deuxième lot de *N. benthamiana* a été co-infiltré avec la protéine p19, l'ATR1 et CYP71AJ1 suivant le protocole décrit dans le chapitre « Matériel et Méthodes ». Quatre jours après la première inoculation, les feuilles des deux lots de plantes ont été infiltrées avec une solution de marmésine à 100 µM. Trois heures après la deuxième inoculation, les feuilles ont été broyées à l'azote liquide et une extraction de polyphénols a été réalisée. La formation de psoralène a uniquement été détectée par LC-MS/MS chez les plantes co-exprimant la protéine p19, l'ATR1 et CYP71AJ1. La co-expression de l'ATR1 avec un P450 permet donc d'améliorer la fonctionnalité du P450 en assurant un meilleur transfert d'électron entre la CPR et le P450.

2.5.1.2 Infiltration de CYP71AZ3 et CYP71AZ4

En se basant sur ces résultats préliminaires, nous avons réalisé l'expression de CYP71AZ3 et CYP71AZ4 chez le tabac. Les gènes codants pour CYP71AZ3 et CYP71AZ4 contenus initialement dans le plasmide pCR8/TOPO/TA/GW, ont été introduits dans le plasmide pBIN_GW® par LR recombinaison. Des agrobactéries *A. tumefaciens* ont été transformées avec les plasmides recombinants

obtenus selon le protocole décrit dans le chapitre « Matériel et Méthodes ». Des agrobactéries ont également été transformées avec le vecteur pBIN_GW® vide afin de constituer un témoin négatif. A l'image des expérimentations préliminaires, j'ai donc choisi de co-exprimer CYP71AZ3 et CYP71AZ4 avec les gènes codants pour la protéine p19 et l'ATR1. Pour cela, cinq inocula bactériens différents ont été préparés : un premier inoculum d'agrobactéries exprimant CYP71AZ3, un deuxième exprimant CYP71AZ4, un troisième inoculum exprimant la protéine p19, un quatrième exprimant l'ATR1 et un dernier correspondant au vecteur vide. Puis, j'ai réalisé la co-infiltration de trois lots de plantes de *N. benthamiana* selon le protocole décrit dans le chapitre « Matériel et Méthodes ». Le premier lot de plantes a été co-infiltré avec des agrobactéries exprimant CYP71AZ3, l'ATR1 et la protéine p19. Le deuxième lot de plantes a été co-infiltré avec des agrobactéries exprimant CYP71AZ4, l'ATR1 et la protéine p19. Enfin un dernier lot a été co-infiltré avec des agrobactéries exprimant l'ATR1, la protéine p19 et le vecteur pBIN_GW® vide afin de constituer un contrôle négatif.

Nous avons décidé d'utiliser deux approches différentes afin de caractériser *in planta* la fonction de ces deux enzymes. Dans un premier temps, nous avons réalisé une préparation de microsomes de tabac à partir des plantes agro-infiltrées et nous avons ensuite effectué un criblage métabolique *in vitro*. Dans un second temps, nous avons réalisé une deuxième infiltration des plants de tabac en utilisant un mélange de différents substrats, puis nous avons extraits les polyphénols.

Préparation de microsomes de tabac et criblage métabolique :

Après quatre jours de co-infiltration, les feuilles des lots de *N. benthamiana* co-infiltrées avec CYP71AZ4, l'ATR1 et la protéine p19, avec CYP71AZ3, l'ATR1 et la protéine p19 ainsi que les feuilles du lot de tabac constituant le contrôle négatif sont prélevées, broyées à ultraturax et une préparation de microsomes est ensuite effectuée selon la méthode décrite dans le chapitre « Matériel et Méthodes ». Des incubations dans du tampon NaPi 0,1 M pH 7,5 ont par la suite été effectuées pendant 30 min avec les trois lots de microsomes avec les différents substrats de ces deux enzymes à 200 µM en présence et en absence de NADPH. Les réactions ont été stoppées par l'addition de 37,5 µl d'acétonitrile/HCl (99/1). Les mélanges réactionnels sont ensuite centrifugés pendant 30 minutes à 10 000 g et les surnageants sont prélevés et placés dans des vials en vue d'être analysés en LC-MS/MS.

Co-infiltration avec un mélange de substrats :

En parallèle, après quatre jours de co-infiltration, les feuilles du lot de *N. benthamiana* co-infiltrées avec CYP71AZ4, l'ATR1 et la protéine p19 ainsi que les feuilles du lot de tabac constituant le contrôle négatif sont à nouveau infiltrées avec un mélange de substrats potentiels (psoralène, scopolétine, 6-

méthoxycoumarine, 7-méthoxycoumarine et 7-méthoxy 3-méthylcoumarine) de CYP71AZ4 à 100 µM. Trois heures après la seconde infiltration, les feuilles ont été prélevées et broyées à l'azote liquide en vue de réaliser une extraction des composés phénylpropanoïdes.

2.5.2 Extraction des polyphénols et identification par spectrométrie de masse

Les polyphénols ont été extraits à partir des feuilles de tabac agro-infiltrées selon le protocole décrit dans le chapitre « Matériel et Méthodes », puis analysés par spectrométrie de masse. Plusieurs molécules ont été recherchées en utilisant un filtre de masse : les différents substrats utilisés ainsi que les masses des molécules hydroxylées.

Concernant les expériences réalisées à partir de la **préparation de microsomes**, les cinq substrats de CYP71AZ4 ainsi que le substrat de CYP71AZ3, l'esculétine, **ont été clairement identifiés dans les échantillons**. Néanmoins, cette approche n'a **pas permis de détecter une métabolisation** que ce soit pour CYP71AZ3 ou pour CYP71AZ4.

La deuxième expérience réalisée en **co-inflitrant les substrats** de CYP71AZ4 dans des tabacs co-infiltrés avec CYP71AZ4, l'ATR1 et la protéine p19 n'a pas non plus abouti à une métabolisation. Cette expérience d'expression transitoire de P450s chez le tabac a été répétée deux fois en utilisant CYP71AJ1 comme contrôle positif, sans plus de succès. Ce protocole reste donc à être optimiser.

Cette approche n'a donc pas permis de déterminer si les substrats de ces deux enzymes identifiés *in vitro* étaient représentatifs de la réalité physiologique chez le panais ou artéfactuels. J'ai donc décidé de réaliser une extraction de polyphénols à partir de feuilles de *P. sativa* âgés de 2 à 3 mois. Les extraits ainsi obtenus ont été analysés en LC-MS/MS afin de rechercher les différents substrats identifiés. **Les résultats obtenus indiquent que seuls l'esculétine et le psoralène sont produits dans des feuilles de *P. sativa*.**

2.6 Discussion

Les expériences de « differential display » réalisées par l'équipe du Pr. Matern de l'université de Marburg sur des cultures cellulaires d'*A. majus* traitées ou non avec des extraits de *Pmg* ont permis de mettre en évidence plusieurs P450s qui pourraient être impliqués dans la synthèse de furanocoumarines (Hamerski *et al*., 1990). Parmi les huit séquences candidates identifiées, je me suis plus intéressée dans ce chapitre à CYP71AZ1 pour qui il existe une corrélation entre l'expression du gène et la synthèse de furanocoumarines (communication personnelle, Kellner, 2008). J'ai donc recherché des membres de cette sous-famille dans les trois banques d'ADNc de plantes produisant des furanocoumarines (*P. sativa*,

R. graveolens et *C. cinereum*) dont nous disposons au laboratoire. J'ai identifié plusieurs membres de CYP71AZ dans la banque d'ADNc de *Pastinaca sativa* et parmi les différentes séquences candidates, deux étaient complètes : CYP71AZ3 et CYP71AZ4.

Pour réaliser la caractérisation fonctionnelle de ces deux candidats potentiels, les protéines ont été exprimées dans des cellules de *S. cerevisiae*. Après vérification de la production de ces deux protéines par western blot (Figure 51, p. 170), j'ai réalisé un criblage métabolique *in vitro* avec des coumarines, des furanocoumarines et diverses autres molécules (Figure 52, p. 172 et Figure 54, p. 174). La caractérisation fonctionnelle de CYP71AZ3 a permis de mettre en évidence *in vitro* l'hydroxylation de l'esculétine en un produit inconnu. CYP71AZ4, quant à lui, présente une spécificité de substrat beaucoup plus faible et est capable de métaboliser *in vitro* la métabolisation du psoralène en xanthotoxol, de la scopolétine en fraxétine, de la 7-méthoxycoumarine en daphnétine 7-méthyléther, de la 6-méthoxycoumarine en scopolétine, mais cette enzyme est également capable d'hydroxyler la 7-méthoxy 3-méthylcoumarine et la 6-méthoxycoumarine en un autre produit que la scopolétine. Néanmoins, la détermination des constantes d'affinité de CYP71AZ4 pour ses différents substrats réalisée *in vitro*, indique que le substrat préférentiel de cette enzyme est le psoralène. Pour CYP71AZ3, un K_m très élevé semble indiquer que l'esculétine n'est pas le substrat préférentiel de l'enzyme.

Afin de déterminer quels sont les substrats physiologiques de ces deux enzymes, j'ai tenté d'exprimer transitoirement ces deux protéines *in planta* dans des feuilles de *N. benthamiana*, une plante qui ne produit pas de furanocoumarines, mais qui synthétise en très faible quantité de la scopolétine et de l'esculétine (Costet *et al*., 2002; Court, 1977). Néanmoins, il n'a pas été possible de détecter les produits formés *in vitro* dans les plantes transformées par agro-infiltration. Etant donné que le contrôle positif de cette expérience (agro-infiltration de bactéries exprimant CYP71AJ1 suivie d'une infiltration de marmésine) n'a pas permis de détecter la formation de psoralène, le protocole d'expression transitoire de P450s dans les feuilles de *N. benthamiana* est encore à optimiser pour obtenir une procédure répétable. J'ai donc réalisé une extraction de polyphénols à partir de feuilles de panais afin de rechercher les substrats de ces deux enzymes identifiés *in vitro*. Les analyses réalisées en LC-MS/MS indiquent que seuls le psoralène et l'esculétine sont produits dans des feuilles de *P. sativa*. Ces résultats sont en adéquation avec les constantes d'affinité mesurées *in vitro*. On peut donc conclure que CYP71AZ3 hydroxyle l'esculétine et que CYP71AZ4 catalyse la formation de xanthotoxol à partir de psoralène chez le panais.

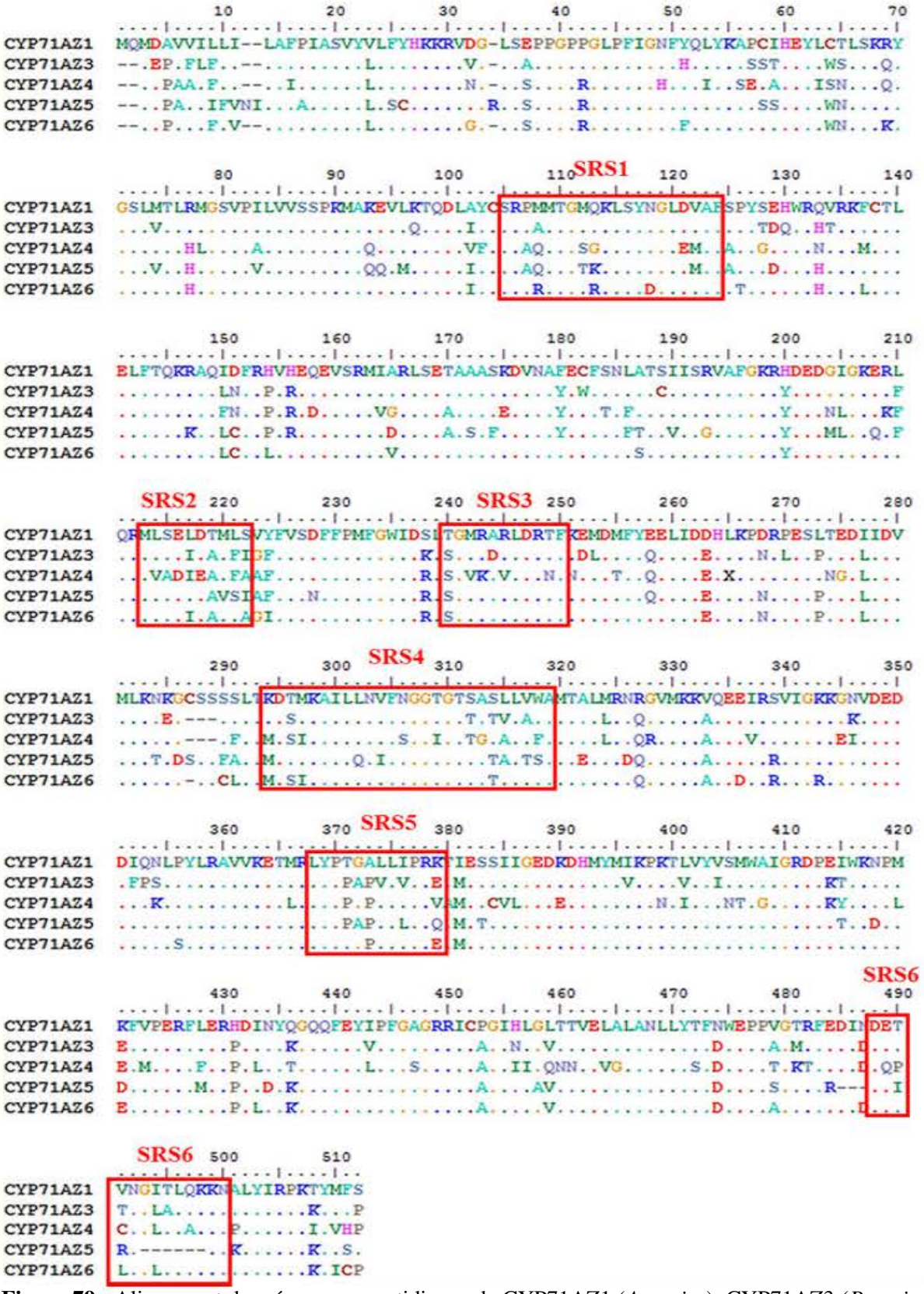


Figure 79 : Alignement des séquences peptidiques de CYP71AZ1 (*A. majus*), CYP71AZ3 (*P. sativa*), CYP71AZ4 (*P. sativa*), CYP71AZ5 (*P. sativa*) et CYP71AZ6 (*P. sativa*). Les six séquences potentielles de reconnaissance du substrat (SRSs) sont encadrées en rouge.

L'analyse du transcriptome du panais a mis en évidence l'existence d'autres cytochromes P450 de cette même sous-famille. Un projet parallèle à cette thèse a été initié au laboratoire et piloté par le Dr. Sandro Roselli, en post-doctorat au LAE. Ce projet vise à mettre en place une banque d'ADN génomique de *P. sativa*. Ce travail a été réalisé dans le cadre d'un contrat INTERREG et d'une collaboration avec le CNRGV de Toulouse. En nous basant sur les résultats obtenus dans le cadre de cette caractérisation fonctionnelle de CYP71AZ3 et CYP71AZ4, nous avons réalisé un criblage de la banque BAC de panais. Ce travail nous a permis d'identifier 2 nouveaux gènes, CYP71AZ5 et CYP71AZ6 pour lequel deux fragments étaient existants dans la banque d'ADNc de *P. sativa* et correspondaient aux contigs P_sat_rep_c5079 et P_sat_rep_c35794. Un alignement des différentes séquences désormais disponibles pour cette famille enzymatique met en évidence un degré d'homologie assez important (Figure 79, p. 199). Ces deux gènes ont été clonés par le Dr. Sandro Roselli et exprimés dans la levure. Le criblage fonctionnel a permis de démontrer que **CYP71AZ6** était également impliqué dans la synthèse de furanocoumarines, mais intervenait dans l'**hydroxylation de la xanthotoxine,** pour former la 5-hydroxyxanthotoxine, le précurseur direct de l'isopimpinelline (5,8-diméthoxypsoralène). Cette activité a aussi été mise en évidence pour **CYP71AZ1** identifiée chez *Ammi majus*. A l'heure actuelle, aucune fonction n'a encore été déterminée pour CYP71AZ5.

L'expression et la caractérisation fonctionnelle de différents membres de CYP71AZ nous a permis de commencer à mieux comprendre l'apparition de la voie de biosynthèse des furanocoumarines chez les Apiacées. L'existence de variants et la mise en évidence de fonctions différentes sur des enzymes structurellement très proches (CYP71AZ3 impliqué dans la synthèse de coumarines, alors que CYP71AZ1, 4 et 6 sont impliqués dans la synthèse de furanocoumarines) nous amènent maintenant à émettre l'hypothèse que des P450s originels ont évolué au cours du temps. Le fait que plusieurs gènes codants pour des P450s appartenant à la même sous-famille aient une fonction différente n'est pas une originalité, mais cela démontre qu'un gène peut acquérir de nouvelles fonctions par des phénomènes de duplication génique et de mutations et ainsi, par des phénomènes de néo-fonctionnalisation, être à l'origine de nouvelles voies de biosynthèse. C'est le cas notamment des P450s impliqués dans la synthèse de benzoxazinoïdes chez le maïs. Les recherches effectuées sur les homologies de séquences, les relations phylogénétiques et l'organisation dans le génome des P450s en cluster de cette voie de biosynthèse suggèrent que ces enzymes ont évolué par des processus de duplications en tandem qui ont ensuite conduit à l'acquisition de nouvelles fonctions de ces P450s (Frey *et al*., 1997, 2009; Gierl and Frey, 2001; Ober, 2010). Dans le cas des P450s impliqués dans la voie de biosynthèse des furanocoumarines, ayant d'abord un spectre de substrat relativement restreint, l'apparition de mutations a probablement élargi la taille de la poche catalytique permettant l'entrée d'une gamme de substrats

beaucoup plus large comme c'est le cas de CYP71AZ4. Dans le cadre d'une collaboration avec le Pr. Simonsen de l'université de Copenhague, un autre membre de CYP71AZ a été identifié. Il s'agit de CYP71AZ2, isolé à partir de *Thapsia laciniata* Rouy, une plante qui à notre connaissance, ne produit pas de furanocoumarines. Un criblage métabolique avec les furanocoumarines et les coumarines disponibles au laboratoire a également été effectué avec cette enzyme, mais aucune fonction n'a été identifiée à ce jour. CYP71AZ2 ne semble donc pas impliqué dans la synthèse de furanocoumarines, mais il est possible que cette enzyme soit impliquée dans la métabolisation de coumarines ou qu'elle ait été recrutée dans une autre voie de biosynthèse. Il en est de même pour CYP71AZ5 isolé à partir du panais pour qui aucune fonction n'a été identifiée à ce jour, alors qu'un criblage métabolique avec des furanocoumarines et des coumarines a été réalisé. Les résultats obtenus dans cette étude sont un exemple parmi d'autres de la plasticité du métabolisme secondaire et surtout des cytochromes P450 à répondre aux modifications des conditions environnementales.

Afin de mieux comprendre les phénomènes évolutifs mis en jeu, il est maintenant envisagé de comparer les structures 3D de ces protéines de sorte à identifier les acides aminés importants dans leur fonction. Pour cela, une collaboration a été initiée avec le Dr. Dave Ritchie du Loria (Laboratoire INRIA-CNRS de l'université de Lorraine, Nancy) pour réaliser une modélisation et une comparaison des structures tridimentionnelles de ces protéines. Cette analyse a permis de pointer un certain nombre d'acides aminés susceptibles de jouer un rôle important dans la spécificité de substrat. Des mutants ont été générés et sont actuellement en cours de caractérisation fonctionnelle.

2.7 Conclusion

Les résultats que j'ai décrits dans cette deuxième partie du chapitre « Résultats » présentent les différents systèmes d'expression hétérologue *in vitro* et *in planta* que j'ai utilisé afin de caractériser les candidats CYP71AZ3 et CYP71AZ4 pour leur potentielle implication dans la voie de biosynthèse des furanocoumarines. Si CYP71AZ3 présente une activité très stricte et est capable de réaliser l'hydroxylation d'une coumarine, l'esculétine, CYP71AZ4 présente un spectre de substrats beaucoup plus large *in vitro* (hydroxylation du psoralène, de la scopolétine, de la 7-méthoxycoumarine, de la 6-méthoxycoumarine en deux positions différentes et de la 7-méthoxy 3-méthylcoumarine). CYP71AZ4 pourrait être un P450 spécialisé dans l'hydroxylation en C8 de la structure de base, la 2H-1-benzopyran-2-one, présente chez les coumarines et les furanocoumarines. L'identification des produits inconnus formés par CYP71AZ3 et 4 a été entreprise avec le Dr. Bernd Schneider de l'Institut Max Planck (Jena – Allemagne) et devrait permettre d'identifier avec précision les positions des hydroxylations. J'ai tenté

de compléter ce travail *in vitro* par des expériences *in planta*. L'expression transitoire de ces deux protéines recombinantes chez le tabac n'a pas permis de mettre en évidence une métabolisation. Ce protocole reste à optimiser, car il n'est pas répétable pour l'instant. J'ai donc réalisé des extraits de polyphénols à partir de feuilles de panais en vue de rechercher les différents substrats identifiés *in vitro*. Les analyses de ces extraits par LC-MS/MS ont montré que seuls l'esculétine et le psoralène sont produits dans les feuilles de *P. sativa*. La caractérisation moléculaire de cette nouvelle sous-famille de P450s impliquée dans la voie de biosynthèse des furanocoumarines et de coumarines est donc une avancée majeure pour la compréhension de l'apparition de cette voie de biosynthèse chez les Apiacées.

CHAPITRE 3 : IDENTIFICATION DE VARIANTS DE CYP71AJ CHEZ *PASTINACA SATIVA*

3.1 Identification de nouveaux P450s appartenant à la sous-famille CYP71AJ

Comme pour la sous-famille CYP71AZ, nous avons réalisé sur la banque d'ADNc de panais une recherche ciblée en utilisant comme séquence sonde CYP71AJ3 et CYP71AJ4. Cette stratégie a conduit à l'identification de 22 contigs identifiés comme appartenant à la sous-famille CYP71AJ en se basant sur la base de données fournie par le Dr. Nelson. Parmi ces 22 contigs, 10 partagent plus de 97 % d'identité de séquence avec CYP71AJ3 et peuvent être considérés comme variants alléliques de ce gène (Tableau 19, p. 203). Deux contigs dont un est prédit complet et l'autre non, partagent respectivement 99 % et 100 % d'identité avec CYP71AJ4 et correspondent à l'angélicine synthase (Tableau 19, p. 203). Le contig P_sat_rap_c41649 présente 100 % d'identité de séquence avec une séquence partielle de CYP71AJ13 isolée à partir de *P. sativa* par Bjørn Dueholm dans le cadre d'une collaboration avec l'équipe du Pr. Simonsen de l'université de Copenhague. Enfin, neuf nouvelles séquences de P450s de la sous-famille CYP71AJ sont identifiées dans cette banque (Tableau 19, p. 203). Toutefois, ces séquences sont incomplètes et ne pourront être classées dans la sous-famille CYP71AJ que lorsque leurs séquences complètes seront obtenues.

Gène	Contigs
CYP71AJ3	P_sat_rep_c2097, c5795, c6145, c9002, c9692, c24823, c39813, c40856, c72040, c77225
CYP71AJ4	P_sat_rep_c2542, c4136
CYP71AJ13	P_sat_rep_c41649
Autre	P_sat_rep_c5642, c10479, c14511, c26434, c29131, c41649, c46810, c61375, c64662

Tableau 19 : Contigs correspondant à des P450s appartenant à la sous-famille CYP71AJ identifiés chez *P. sativa* par tblastn en utilisant comme sondes les séquences peptidiques de CYP71AJ3 et CYP71AJ4 de *P. sativa.*

Ces 9 fragments de P450s orthologues à CYP71AJ3 et CYP71J4 ne sont pas tous localisés sur la même partie de la séquence codante. Néanmoins, si on les aligne avec les séquences disponibles on constate qu'elles en divergent légèrement puisque l'identité des séquences peut aller de 50 % à 78 % et leur degré d'homologie varie de 67 % à 90 %. Ces résultats indiquent une diversité vraisemblablement importante

de cette sous-famille de P450s (Tableau 20, p. 204). Les fonctions de CYP71AJ3 et CYP71J4 étant différentes il est probable que tous les membres de CYP71AJ présentent des fonctions différentes et soient le résultat phénomènes de duplication géniques suivi de néo-fonctionnalisation à l'instar de la sous-famille CYP71AZ décrite dans la partie précédente de ce mémoire. Ils pourraient être issus d'un même gène ancestral, mais cela reste à être démontré.

	5642	10479	14511	26434	29131	41649	46810	61375	64662
CYP71AJ1	**61**	**69**	**62**	**69**	**73**	**50**	**56**	**76**	**69**
	78	**83**	**81**	**84**	**86**	**67**	**71**	**88**	**80**
CYP71AJ2	**61**	**66**	**62**	**65**	**75**	**50**	**57**	**74**	**74**
	80	**83**	**81**	**85**	**87**	**68**	**73**	**89**	**84**
CYP71AJ3	**64**	**68**	**59**	**63**	**72**	**66**	**57**	**74**	**72**
	81	**82**	**81**	**83**	**88**	**78**	**73**	**86**	**83**
CYP71AJ4	**57**	**65**	**58**	**59**	**72**	**55**	**60**	**58**	**78**
	79	**85**	**77**	**84**	**86**	**75**	**75**	**79**	**90**

Tableau 20 : Pourcentages d'identité (en noir) et d'homologie (en rouge) peptidiques entre CYP71AJ1 d'*A. majus*, CYP71AJ2 d'*Apium graveolens*, CYP71AJ3 et 4 isolés chez *P. sativa* par Larbat et ses collaborateurs (Larbat *et al*., 2007, 2009), et les séquences candidates identifiées dans la banque d'ADNc de *P. sativa*.

Afin de réaliser une analyse plus fine de ces 9 séquences correspondant potentiellement à des nouvelles enzymes de la sous-famille des CYP71AJ, il est d'abord nécessaire d'identifier la séquence complète de chacune de ces séquences codantes. Puis, dans un second temps, il faudra exprimer les protéines correspondantes avant d'en rechercher la fonction.

Au vu de leur localisation sur l'alignement présenté dans la Figure 81 (p. 207), il est possible que certains fragments correspondent à 2, voire 3 parties différentes d'une même séquence codante. En effet, aucun chevauchement de séquence n'a été mis en évidence. Si tel est le cas, le clonage des séquences codantes complètes en serait grandement facilité. Il suffirait alors de réaliser une RT-PCR sur des ARNm (ou une PCR sur de l'ADN génomique) extraits de *P. sativa* en utilisant des amorces complémentaires aux extrémités 5' et 3' des différents fragments de gènes et de tester les différentes combinaisons possibles pour vérifier si une amplification est possible. En me basant sur ce postulat, nous avons fait synthétiser des amorces sens spécifiques des contigs P_sat_rep_c41649, c46810, c61375 et c64662 et des amorces antisens spécifiques des contigs P_sat_rep_c5642, c10479, c14511, c29131 et c41649 (Tableau 21, p. 207).

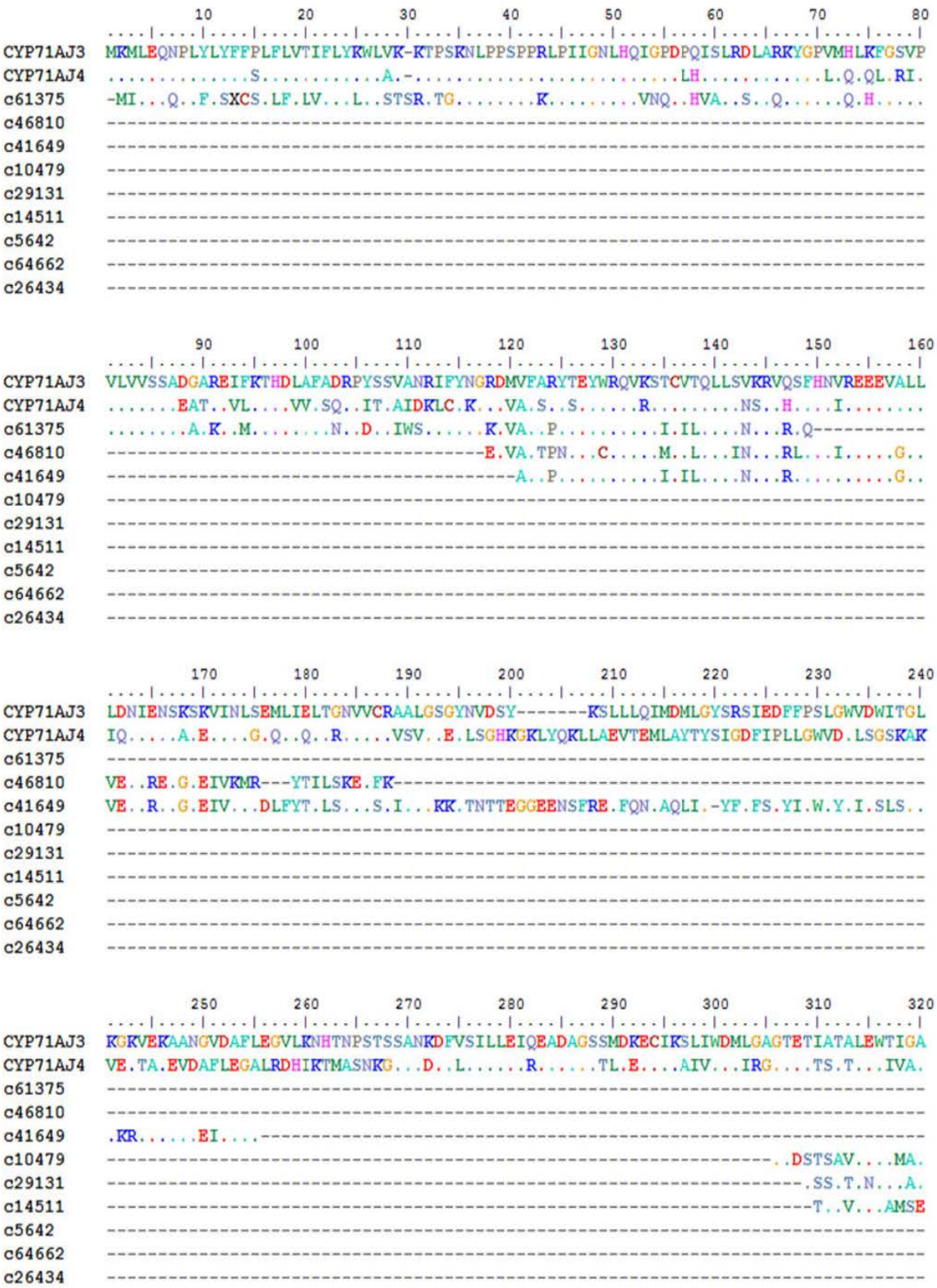


Figure 80 : Alignement des séquences peptidiques partielles des CYP71AJ identifiés à partir de la banque d'ADNc de *P. sativa* avec celles de CYP71AJ3 et CYP71AJ4 de *P. sativa*. X : inconnu.

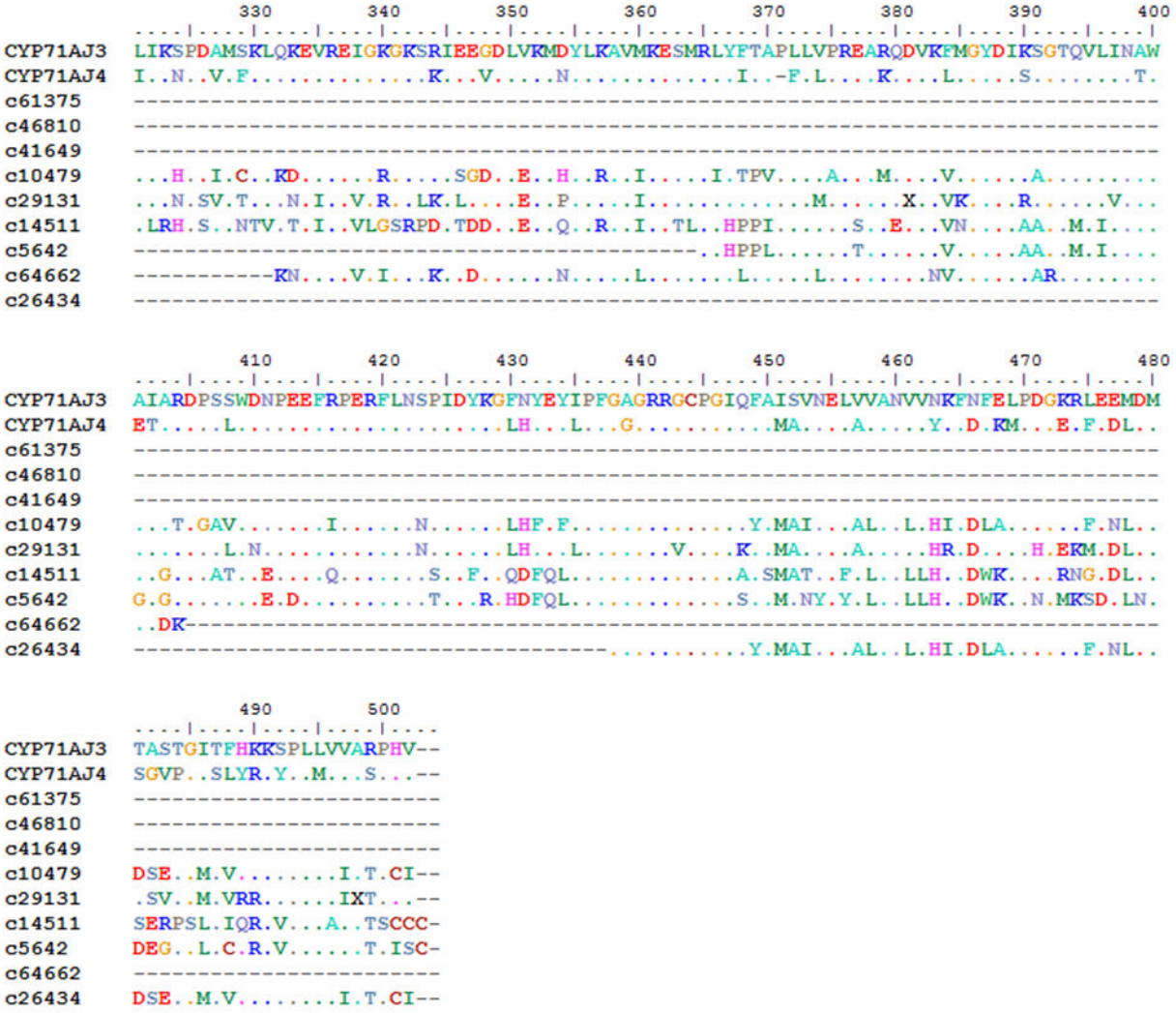


Figure 80 : Alignement des séquences peptidiques partielles des CYP71AJ identifiés à partir de la banque d'ADNc de *P. sativa* avec celles de CYP71AJ3 et CYP71AJ4 de *P. sativa*. X : inconnu.

La première étape de ce travail a donc consisté à identifier les fractions de séquences manquantes pour définir une stratégie de recherche visant à obtenir les séquences complètes. Pour cela, j'ai réalisé un alignement des séquences partielles de CYP71AJ avec les structures primaires de CYP71AJ3 et CYP71AJ4 isolées à partir du panais (Figure 80, p. 205-206). Parmi ces neuf séquences, une seule contient un codon initiateur ATG (P_sat_rep_c61375), deux de ces contigs correspondent au milieu de la séquence codante (P_sat_rep_c41649 et c46810) et six correspondent à la partie 3' terminale de la séquence codante (P_sat_rep_c5642, c10479, c14511, c26434, c29131 et c64662) (Figure 80 p. 205-206). Cet alignement révèle également que le contig P_sat_rep_c26434 présente 100 % d'identité peptidique avec le contig P_sat_rep_c10479 ce qui réduit donc à huit le nombre de séquences partielles de CYP71AJ à étudier dans la banque de panais (Figure 80 p. 205-206).

Contig	Amorce directe (5'→ 3')	Amorce réverse (5'→ 3')
c5642		TCAACAAGAAATTGGAGTTGCAACTACAAG TAG
c10479		CTAGATACATGGCGTTGCAATCACCAATAG
c14511		CTAGCAGCAACATGAAGTTGCGACAG
c29131		GGAGATTTCCTACGGACAGTCATGC
c41649		GCATCAATTTCATTAGCTGCTTTCTCAACCC
c46810	TGAGGACGTGGCTTCA	
c61375	ATGATACTTGAGCAACAACCCTTGT TTCTGTC	
c64662	CCCACAGAGAAGGATCTCTTGTCAA TTGCC	

Tableau 21 : Amorces utilisées dans les réactions de PCR pour déterminer si les différents fragments de CYP71AJ identifiés dans la banque d'ADNc de *P. sativa* appartiennent au même gène ou non.

Pour des raisons de stabilité du matériel, nous avons commencé par tenter de faire des amplifications sur de l'ADN génomique. L'ADN de *P. sativa* a été extrait à partir de feuilles selon le protocole décrit dans le chapitre « Matériel et Méthodes ». Des réactions de PCR ont ensuite été effectuées avec les différents couples d'amorces sens et antisens. Les produits de PCR ont ensuite été analysés sur gel d'agarose 1 % (Figure 82, p. 208).

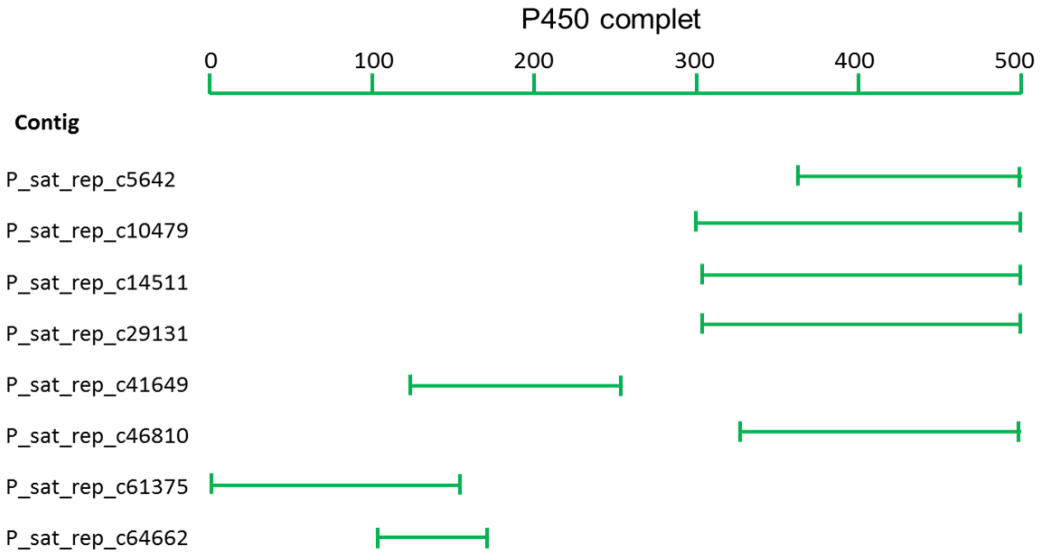


Figure 81 : Représentation schématique des séquences partielles de CYP71AJ identifiées dans la banque d'ADNc de *P. sativa*.

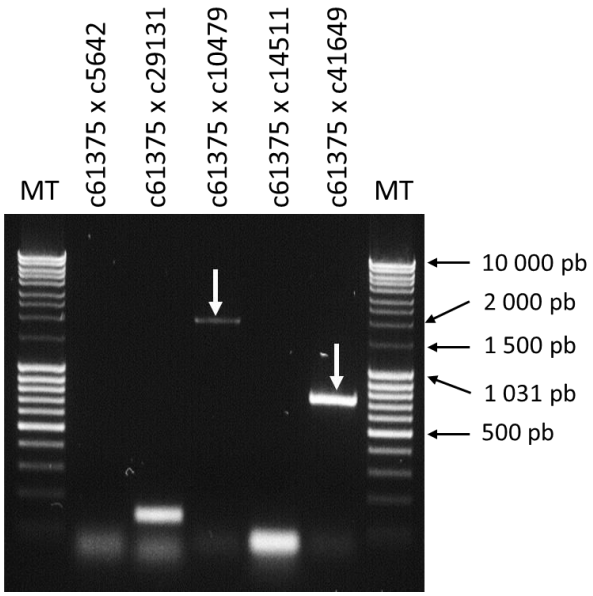


Figure 82 : Analyse sur gel d'agarose à 1 % des produits de PCR réalisés avec les différents couples d'amorces spécifiques des fragments CYP71AJ isolés à partir de la banque d'ADNc de *P. sativa*. MT : marqueur de taille SM00403.

A ce stade, seuls deux couples d'amorces ont permis d'amplifier un fragment d'ADN. Ce premier résultat semble indiquer que les contigs P_sat_rep_c5642, c29131 et c14511 n'appartiennent pas à la même séquence codante que le contig c61375.

Le couple d'amorces c61375 x c10479 a permis d'amplifier un fragment d'environ 2 000 pb et le couple c61375 x c41649 a amplifié un fragment d'ADN d'environ 800 pb (Figure 82, p. 208). Ces deux fragments ont ensuite été extraits du gel d'agarose et purifiés. Pour une raison qui n'a pas été identifiée, le fragment de 2 000 pb n'a pas pu être cloné. Le fragment de 800 pb a été ligué dans le vecteur pCR8/TOPO/GW® et le plasmide contenant l'insert a été séquencé. Les résultats du séquençage indiquent que les contigs P_sat_rep_c41649 et c61375 correspondent à une même séquence ce qui est cohérent avec les observations faites sur l'alignement de séquences. En effet on peut constater un chevauchement de quelques acides aminés. Cette première approche nous a donc permis de rallonger ces deux contigs mais également de confirmer que les constructions faites *in silico* correspondent à la réalité.

Si on se fie aux alignements de séquences de la Figure 80 (p. 205-206), les amorces c61375 et c10479 pourraient correspondre aux extrémités de la séquence codante. D'un côté, on a l'ATG initiateur et dans le second cas, on a le codon stop. La taille d'une séquence codante d'un cytochrome P450 est généralement de 1500 pb environ. Dans le cas, de la PCR faite sur l'ADN génomique, on amplifie un

fragment de 2000 pb environ. Ce résultat est très encourageant et semble indiquer que les 2 amorces s'hybrident sur une même séquence codante. La taille plus importante que celle qui est attendue pourrait refléter la présence d'un intron dans la mesure où la PCR a été réalisée sur de l'ADN génomique. Nous avons dans un second temps fait une amplification sur de l'ARN. Les ARN totaux ont été extraits à partir de feuilles de panais selon le protocole décrit dans le chapitre « Matériel et Méthodes », puis rétrotranscrits en ADNc. Une réaction de PCR a ensuite été effectuée avec le couple d'amorces c61375 x c10479. Le produit de PCR a ensuite été analysé sur gel d'agarose 1 %. Le fragment obtenu d'une taille d'environ 1500 pb a été prélevé, purifié, puis ligué dans le pCR8/TOPO/GW® ce plasmide a été séquencé. La séquence codante de ce gène, d'une longueur de 1 512 pb, a été soumise au Dr. Nelson pour être classé dans la nomenclature internationale, il s'agit du P450 CYP71AJ13. CYP71AJ13 est donc représenté par 4 contigs différents dans la banque d'ADNc de *P. sativa* : P_sat_rep_c10479, c26434, c41649 et c61375. Ce gène complet nous a donc permis de compléter une séquence codante partielle identifiée précédemment par Bjørn Dueholm (voir début de ce chapitre).

3.2 Recherche de gènes complets correspondant aux séquences partielles présentes dans la banque de *P. sativa* : approche TAIL-PCR

Parmi les 5 contigs restants correspondant potentiellement à des fragments de CYP71AJ, aucun ne correspondait à une extrémité 5' terminale de gènes (Figure 81, p. 207) mais à des extrémités 3' terminales ou alors à des parties centrales de la séquence codante. D'un point de vue technique, il est relativement facile de rechercher des extrémités 3' terminale de séquences codantes en utilisant une approche PCR avec un couple d'amorces dirigées contre l'extrémité 5' terminale connue et l'extrémité 3' terminale qui chez les eucaryotes correspond généralement à une séquence poly adénylée. La recherche d'une partie 5' terminale est par contre plus difficile que ce soit à partir d'ADN ou d'ARNm. Différentes méthodes ont été décrites dans la littérature comme les approches de RACE-PCR (Rapid Amplification of cDNA Ends) sur des ARNm ou des approches de TAIL-PCR (Thermal Asymmetric InterLaced) sur de l'ADN génomique. L'ADN génomique étant réputé plus stable et moins sujet à dégradation que l'ARN, nous avons choisi de nous focaliser sur la seconde méthode dont le principe a été décrit dans la partie « Matériel et Méthodes » de ce manuscrit.

Nous nous sommes focalisés sur les contigs P_sat_rep_c5642, c14511, c29131 et c64662. Le contig P_sat_rep_c46810, ayant une taille relativement courte (388 nt) et comportant plusieurs changements de cadre de lecture, n'a pas été retenu pour l'approche de TAIL-PCR. Vue la qualité de la séquence de ce petit contig il est vraisemblable qu'il soit le fruit d'erreurs de séquençage voire d'assemblage.

En me basant sur la littérature, j'ai identifié 8 amorces dégénérées différentes me permettant de tenter d'amplifier les extrémités manquantes des différents P450s. Pour chaque contig, j'ai fait synthétiser trois amorces se succédant sur la séquence nucléique. Les réactions de TAIL-PCR ont ensuite été effectuées indépendamment avec les huit amorces dégénérées en réalisant à l'aveugle 3 PCR successives en utilisant dans les mélanges réactionnels, les amorces spécifiques les plus distantes de l'extrémité pour la première PCR jusqu'à la plus proche pour la troisième réaction. Les produits de l'ensemble de ces TAIL-PCR ont ensuite été analysés sur gel d'agarose 1 % et ce sont les fragments obtenus lors de la troisième réaction de PCR qui ont été prélevés.

Résultats obtenus avec le contig P_sat_rep_c29131

En utilisant l'amorce dégénérée AD5 en association avec le contig P_sat_rep_c29131, j'ai pu générer un fragment d'environ 1 700 pb. Ce fragment a été cloné dans le pCR8/TOPO/GW® et séquencé. Les résultats de séquençage indiquent que la séquence amplifiée permet de remonter jusqu'au codon initiateur et que cette séquence contient un intron ce qui n'est pas incohérent avec ce qui est décrit dans la littérature (Annexe 15-A, p. 315). La séquence codante ainsi obtenue permet la synthèse d'une protéine théorique de 500 acides aminés (Annexe 15-A et B, p. 315-314), compatible avec la taille de séquences correspondant à des cytochromes P450. Cette séquence complète a été soumise à la classification internationale et reste bien affiliée à la sous-famille des CYP71AJ.
Cette approche a donc permis d'identifier un nouveau CYP71AJ différent de ceux déjà identifiés jusque-là. Ce gène est actuellement en cours de clonage à partir d'ARN messager. Des amorces spécifiques de la séquence codante de ce gène ont été générées (DIR : ATGCCGGAGCTATATTCC, REV : TTAAACATGTGGTGTG) et utilisées pour amplifier cette séquence en utilisant comme matrice de l'ADNc. Pour le moment il n'a pas pu être amplifié ce qui peut être expliqué de différentes manières. Il est possible 1) que le gène soit exprimé trop faiblement dans la plante, 2) que les conditions d'induction ne soit pas les bonnes, 3) que les séquences codantes soient légèrement variables d'une plante à une autre (lié à la différence de génotypes) rendant l'amplification impossible avec des conditions de PCR trop stringentes, 4) que le gène ne soit pas exprimé et qu'il corresponde à un pseudogène.

Résultats obtenus avec le contig P_sat_rep_c14511

L'utilisation combinée de l'amorce dégénérée AD4 avec les amorces spécifiques du contig P_sat_rep_c14511 m'a amené à amplifier un fragment d'une taille d'environ 1 200 pb. Comme pour le cas précédent, ce fragment d'ADN a été cloné dans le plasmide pCR8/TOPO/GW® et son séquençage

indique que l'approche de TAIL-PCR a de nouveau permis de remonter jusqu'à différents codons méthionine : 3 en tout (Annexe 16-A, p. 317). La caractérisation fonctionnelle de cette protéine passera donc par un travail préliminaire consistant à définir quel est le vrai codon initiateur de cette séquence codante. Rappelons qu'ici l'amplification a été faite sur de l'ADN génomique et il est possible qu'en partant de l'ARNm on obtienne des informations complémentaires. L'analyse fine de cette séquence indique qu'elle contient également un intron et qu'elle code pour une protéine de 507, 520 ou 521 acides aminés selon le codon initiateur considéré (Annexe 16-A et B, p. 317-316). Une nouvelle recherche de l'affiliation de cette séquence aux cytochromes P450 révèle cependant que le contig que j'avais précédemment associé à la sous-famille CYP71AJ appartient plutôt à la sous-famille des CYP71AQ.
Malgré cela, tout comme pour le contig P_sat_rep_c29131, j'ai cherché à cloner cette séquence codante à partir d'ARNm. Pour cela, plusieurs amorces spécifiques de la séquence codante de ce gène ont été commandées. J'ai fait synthétiser des amorces différentes s'hybridant aux trois codons initiateurs (DIR 1 : ATGATGGTTTTTCTACATGTACTACAGCAA ; DIR 2 : GGATCCATGTTGTTGCAACATCTGTTCCTTGC ; REV : GGTACCCTAGCAGCAACATG). Malgré toutes les tentatives entreprises, aucune amplification à partir d'ARNm n'a pu être obtenue. Les raisons sont vraisemblablement les mêmes que pour le contig P_sat_rep_c29131.

Résultats obtenus avec le contig P_sat_rep_c5642

L'utilisation de l'amorce dégénérée AD3 en association avec les amorces spécifique du contig P_sat_rep_c5642 a permis de générer trois fragments différents lors de la troisième PCR : deux de ces fragments possèdent une taille de 800 pb et la taille du troisième fragment est d'environ 1 000 pb. Ces fragments ont été prélevés du gel, purifiés puis ligués dans le vecteur pCR8/TOPO/GW® avant d'être séquencés.
Plusieurs conclusions peuvent être tirées de ce troisième clonage. Les résultats de séquençage indiquent que dans les trois cas, **nous avons pu rallonger** la séquence du contig P_sat_rep_c5642 pour obtenir un fragment plus grand de la séquence codante qui correspond maintenant à une protéine de 225 acides aminés (alors qu'initialement on ne disposait que d'une séquence codant pour 172 acides aminés). Cette séquence n'est cependant pas encore suffisamment longue pour avoir la totalité de la séquence codante (environ 500 acides aminés).
La seconde information que nous apportent ces clonages est que, dans les trois, cas on met en évidence la présence d'un intron dans la séquence (Annexe 17-A, B et C, p. 319-319) ce qui pourra nous être utile pour la suite du clonage de ce gène. La troisième information indique que les séquences obtenues diffèrent légèrement entre elles par un certain nombre de modifications ponctuelles qui peuvent être

liées à différents facteurs (nombre de cycles de PCR, utilisation d'une polymérase ne disposant d'une activité de « proof-reading », présence de deux allèles différents, aspécificité de l'approche de TAIL-PCR due à une très forte homologie de séquence nucléotidique entre certains P450s…).
Enfin, en comparant les séquences de ces nouvelles protéines avec celles qui sont disponibles dans la base de données fournie par le Dr Nelson, on se rend compte que l'affiliation à la sous-famille CYP71AJ n'est plus exacte. En effet, les résultats indiquent que les deux séquences les plus courtes appartiennent à la sous-famille CYP71BF et la séquence la plus longue appartiendrait à la sous-famille CYP71AQ.

Résultats obtenus avec le contig P_sat_rep_c64662

Enfin, les approches de TAIL-PCR réalisées sur le contig P_sat_rep_c64662 avec l'amorce dégénérées AD5 ont permis d'amplifier un fragment, mais l'analyse des résultats de séquençage ont montré que le gène amplifié n'est pas un P450, mais un gène chloroplastique qui présente une forte homologie de séquence avec le gène *Ycf2* de *N. tabacum*. Ces résultats indiquent que l'approche TAIL-PCR reste une méthode qui n'est pas extrêmement spécifique.

Dans le cadre de cette recherche de séquences complètes par TAIL-PCR à partir de contigs correspondant à l'extrémité 3' terminale de P450s, j'ai donc pu cloner les séquences génomiques de **2 nouveaux cytochromes P450s**. Le premier appartient à la sous-famille **CYP71AJ** comme attendu et le second correspond à un P450 d'une **autre sous-famille, CYP71AQ** qui reste néanmoins un candidat intéressant au vu des résultats obtenus dans la partie précédente de ce manuscrit. Ces 2 P450s devront maintenant être clonés à partir d'ARN messager et leurs fonctions identifiées. Ce travail a également permis de rallonger la séquence connue d'un troisième P450 mais n'a pas été efficace pour le dernier candidat. Dans ces 2 cas, il sera nécessaire maintenant soit de refaire l'expérimentation en utilisant d'autres amorces dégénérées soit en utilisant d'autres approches comme la technique de 5' RACE-PCR par exemple.

3.3 Recherche de la fonction de CYP71AJ13

CYP71AJ13 est le seul P450 de la sous-famille CYP71AJ que nous avons pu cloner dans le cadre de ce travail. Il est nécessaire maintenant d'en identifier la fonction.

3.3.1 Clonage et expression chez la levure

Comme pour CYP71AZ3 et CYP71AZ4, nous avons décidé de réaliser la caractérisation fonctionnelle de CYP71AJ13 dans le système hétérologue *S. cerevisiae*. La première étape a consisté à extraire les ARN totaux de feuilles de panais en utilisant le kit RNeasy® Plant Mini (Qiagen) selon le protocole décrit dans le chapitre « Matériel et Méthodes ». L'intégrité de ces ARN a été vérifiée sur gel d'agarose 1 % (Figure 83, p. 213) et ces ARNm ont été rétro-transcrits en ADNc comme décrit précédemment.

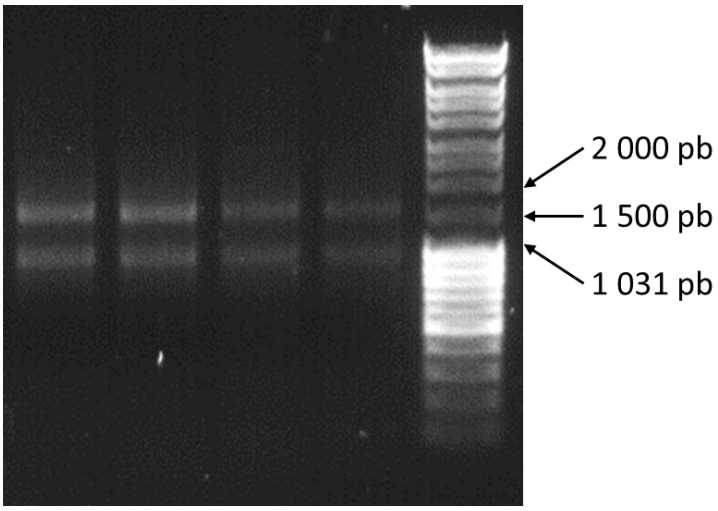


Figure 83 : Vérification de l'intégrité des ARNm extraits à partir des feuilles de *P. sativa* à l'aide du kit RNeasy® Plant Mini (Qiagen). Marqueur de taille : SM0403 (Fermentas).

La séquence codante correspondant à CYP71AJ13 a été amplifiée par PCR avec le couple d'amorces suivant : DIR : GGATCCATGATACTTGAGCAACAACC et REV : GAATTCCTA*GTGGTGATGGTGATGATG*GATACATGGCGTTGC. Ces amorces ont permis d'ajouter un site de restriction de part et d'autre du gène ainsi qu'un motif constitué de 6 histidines afin de permettre ultérieurement la détection de la production de notre protéine dans l'organisme hétérologue. L'amplification a été réalisée en effectuant 35 cycles (94 °C 30 s, 55 °C 30 s, 68 °C 2 minutes) grâce à la Platinium®*Taq* DNA Polymerase High Fidelity(Invitrogen). Le produit de PCR a été déposé sur gel d'agarose 1 % et le fragment ayant une taille d'environ 1 500 pb a été prélevé puis purifié à l'aide du kit « MinElute® Gel Extraction » (Qiagen) selon les recommandations du fournisseur (Figure 84, p. 214).

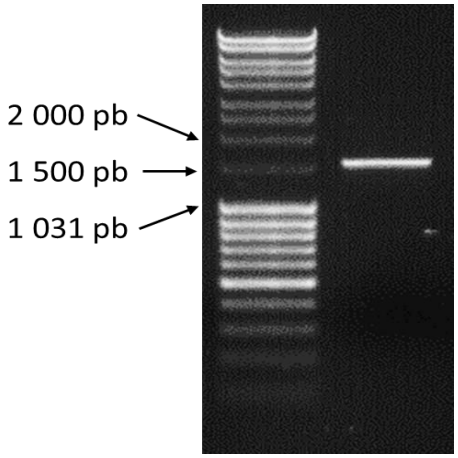


Figure 84 : Amplification de la séquence codante de *CYP71AJ13* isolé à partir de *P. sativa*. Marqueur de taille : SM0403 (Fermentas).

Ce produit de PCR purifié a ensuite été cloné dans le vecteur d'entrée pCR8™/TOPO/GW et séquencé afin de vérifier la séquence de l'insert et son orientation dans le vecteur. La séquence codante de *CYP71AJ13* a ensuite été transférée dans notre plasmide de destination p*Ye*DP60_GW® grâce à une réaction de LR recombinaison selon le protocole décrit dans le chapitre « Matériel et Méthodes » et le plasmide recombinant introduit dans la levure WAT11.

Avant d'initier un criblage fonctionnel, nous avons vérifié que la protéine recombinante était bien exprimée en utilisant une approche de western blot avec des anticorps commerciaux dirigés contre l'étiquette poly histidine présente à l'extrémité C-terminale de la protéine. Pour cela, les protéines totales contenues dans notre préparation de microsomes ont été séparées selon leur poids moléculaire apparent par SDS-PAGE selon le protocole décrit dans le chapitre « Matériel et Méthodes ». Des microsomes de levure transformées avec le plasmide p*Ye*DP60 vide ont servi de contrôle négatif. Après révélation en présence de NBT/BCIP, on constate que les microsomes issus de la culture de deux colonies différentes de levure WAT11 transformées avec le p*Ye*DP60_GW®_CYP71AJ13 semblent renfermer des quantités similaires de protéine (Figure 85, p. 215, puits 1 et 2). Cette protéine CYP71AJ13 semble avoir un poids moléculaire apparent légèrement supérieur à 50 kDa (Figure 85, p. 215, puits 1 et 2). Dans le cas du vecteur vide, aucune bande n'est révélée (Figure 85, p. 215, puits 3 et 4) indiquant que les signaux observés sur les pistes 1 et 2 correspondent bien à une protéine surexprimée.

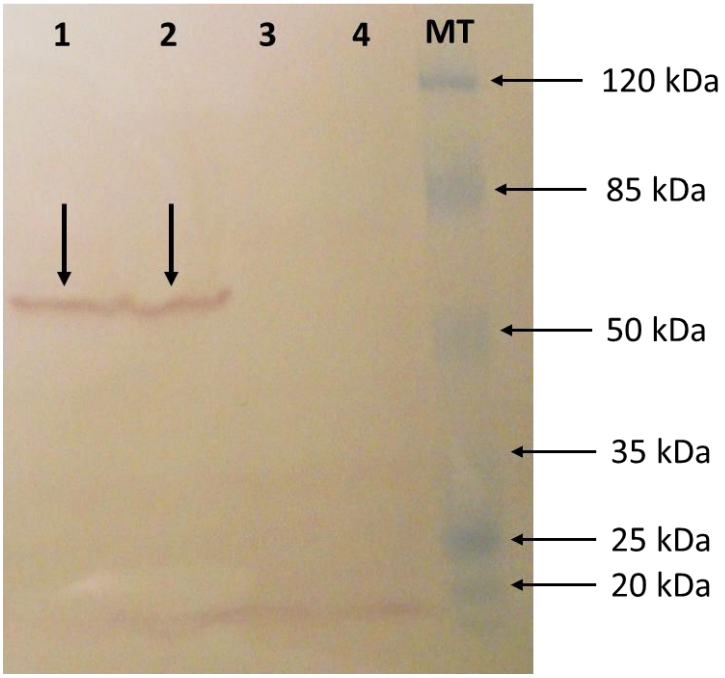


Figure 85 : Détection de la production de la protéine CYP71AJ13 dans le système levure WAT11 par western blot après séparation en SDS-PAGE en comparaison avec le marqueur de taille Prestained Protein Molecular Weight Marker (ready-to-use) (Fermentas). MT : marqueur de taille. Puits 1 et 2 : microsomes de levures trasformées avec le p*Ye*DP60_GW®_CYP71AJ13, puits 3 et 4 : microsomes de levures transformées avec le vecteur vide.

3.3.2 Criblage fonctionnel

Un criblage métabolique a été réalisé pour CYP71AJ13 avec une large collection de 59 molécules correspondant à 18 furanocoumarines linéaires, 12 angulaires, 23 coumarines. En plus de ces molécules déjà testés précédemment, j'ai rajouté des molécules présentant quelques similitudes au niveau structural comme l'acide cinnamique et deux de ses dérivés hydroxylés, deux flavonoïdes, ou encore des molécules dérivées comme une furoquinoline alcaloïde (Annexe 2, p. 284-287) afin de déterminer la fonction de cette protéine.

A l'heure actuelle, aucune métabolisation n'a pas être mise en évidence dans le cas de CYP71AJ13. Le criblage devra donc être élargi à d'autres molécules. Il est vraisemblable qu'une meilleure connaissance du métabolisme des coumarines chez *P. sativa* permettrait de cibler de manière plus pertinente les molécules à utiliser lors du criblage métabolique.

3.3.3 Tentative de compréhension de la fonction par une approche d'analyse bio-informatique

Le criblage métabolique réalisé avec un large panel de molécules disponibles au laboratoire sur des colonies de levure WAT11 n'a pas permis de déterminer la fonction de CYP71AJ13. Nous avons donc décidé d'utiliser une approche d'analyse bio-informatique afin de tenter de mieux comprendre la fonction de ce gène. Dans un premier temps, nous avons comparé les pourcentages d'identité et d'homologie de séquences entre CYP71AJ13 et les membres de cette sous-famille dont la fonction est connue.

A	CYP71AJ1	CYP71AJ2	CYP71AJ3	CYP71AJ4	CYP71AJ13
CYP71AJ1		90 %	87 %	81 %	68 %
CYP71AJ2	90 %		86 %	81 %	67 %
CYP71AJ3	87 %	86 %		83 %	69 %
CYP71AJ4	81 %	81 %	83 %		68 %
CYP71AJ13	68 %	67 %	69 %	68 %	

B	CYP71AJ1	CYP71AJ2	CYP71AJ3	CYP71AJ4	CYP71AJ13
CYP71AJ1		89 %	81 %	70 %	59 %
CYP71AJ2	93 %		81 %	71 %	57 %
CYP71AJ3	89 %	89 %		70 %	60 %
CYP71AJ4	81 %	83 %	83 %		55 %
CYP71AJ13	75 %	76 %	76 %	75 %	

Tableau 22 : Comparaison des pourcentages d'identité (en vert) et d'homologie (en violet) nucléotidiques (A) et peptidiques (B) entre CYP71AJ13 identifié dans la banque d'ADNc de *P. sativa* et CYP71AJ1 (*A. majus*), CYP71AJ2 (*Apium graveolens*), CYP71AJ3 et CYP71AJ4 (*P. sativa*).

La comparaison des pourcentages d'identité et d'homologie de séquences peptidiques entre les différents membres de cette sous-famille indique que CYP71AJ13 est légèrement plus proche de CYP71AJ2 et de CYP71AJ3 (76 % d'identité) que de CYP71AJ1 et CYP71AJ4 (75 % d'identité) (Tableau 22, p. 216). Lorsque l'on compare les pourcentages d'identité nucléotidiques entre les différents membres de CYP71AJ, CYP71AJ13 est plus proche de CYP71AJ3 (69 % d'identité) que des autres membres, néanmoins CYP71AJ13 diverge légèrement des autres CYP71AJ avec seulement 70 % d'identité, alors que CYP71AJ1-4 partagent entre 81 % et 93 % d'homologie de séquences peptidiques (Tableau 22, p. 216). Ces résultats viennent appuyer l'hypothèse que les différents membres de cette sous-famille

dérivent d'un même gène ancestral, toutefois CYP71AJ13 semble être plus éloigné de CYP71AJ1-4 et pourrait donc intervenir dans une autre voie de biosynthèse que celle des furanocoumarines.

Dans un deuxième temps, nous avons généré un alignement de structures primaires de CYP71AJ13 avec les membres de CYP71AJ dont la fonction a été caractérisée (Figure 86, p. 219). On constate également que CYP71AJ13 présente un certain nombre d'acides aminés différents par rapport aux psoralène synthases (CYP71AJ1-3) et à l'angélicine synthase (CYP71AJ4). CYP71AJ13 présente notamment une insertion de 7 acides aminés (KKYTNTT) par rapport aux psoralène synthases en position 192, une délétion d'un acide aminé en position 221 par rapport à CYP71AJ1-4 et une insertion de deux acides aminés en position 291 (AF) par rapport à CYP71AJ1-4 (Figure 86, p. 219).

Afin d'étayer l'hypothèse émise plus tôt, nous nous sommes intéressés de plus près aux différences existant entre CYP71AJ1-4 et CYP71AJ13 au niveau des six sites de reconnaissance du substrat (SRSs) potentiels (Gotoh, 1992; Larbat *et al*., 2007). Nous considérons dans le cas des CYP71AJ, que les structures tertiaires des protéines sont superposables et que les SRSs sont donc localisés dans les mêmes régions de la séquence primaire des P450s. Les six SRSs potentiels sont encadrés en rouge sur la Figure 86 (p. 219). L'analyse comparative des différents SRSs des membres de CYP71AJ, indique qu'*a priori*, CYP71AJ13 est plus proche de CYP71AJ3 (entre 30 et 67 % d'identité) que de CYP71AJ4 (entre 10 et 67 % d'identité) (Tableau 23, p. 217). Ces résultats vont dans le sens des comparaisons de séquence globale des séquences protéiques décrites précédemment. Néanmoins, une analyse plus détaillée des homologies de séquences restreintes à l'un ou l'autre SRS donnent des informations nouvelles. Larbat et de ses collaborateurs avaient démontré que les SRS2 et SRS3 pourraient former un canal permettant l'entrée du substrat dans l'enzyme (Larbat *et al*., 2007). Au niveau du SRS2, CYP71AJ13 présente une grande différence par rapport à CYP71AJ1-4 (30 % d'identité avec CYP71AJ1-3 et seulement 10 % d'identité avec CYP71AJ4) (Tableau 23, p. 217).

	SRS 1	**SRS 2**	**SRS 3**	**SRS 4**	**SRS 5**	**SRS 6**
CYP71AJ1	62	30	50	46	50	50
CYP71AJ2	57	30	58	46	58	50
CYP71AJ3	62	30	67	38	58	64
CYP71AJ4	43	10	67	38	33	29

Tableau 23 : Pourcentages d'identité des séquences peptidiques codant pour les potentiels SRSs des membres de CYP71AJ. Les séquences SRSs ont été identifiées sur la Figure 86. Ce tableau présente les pourcentages d'identité entre CYP71AJ13 et CYP71AJ1 d'*A. majus*, CYP71AJ2 d'*Apium graveolens*, CYP71AJ3 de *P. sativa* et CYP71AJ4 de *P. sativa*.

Par ailleurs, l'analyse détaillée de ce SRS2 montre que la phénylalanine 211 de CYP71AJ13 est remplacée par une isoleucine chez CYP71AJ1 et 2 et par une leucine chez CYP71AJ3 et 4. Or la phénylalanine possède un encombrement stérique plus important que la leucine et l'isoleucine. De plus, la méthionine 215 des psoralène synthases et la thréonine 215 ayant un caractère polaire de part son groupement OH de CYP71AJ4 sont remplacées par une alanine chez CYP71AJ13. Or, l'alanine présente un plus faible encombrement stérique que la méthionine et la thréonine et procure une flexibilité plus importante à la chaîne protéique. De même, la méthionine 217 de CYP71AJ1-4 est remplacée par une leucine chez CYP71AJ13 qui présente également un encombrement stérique plus faible et procure une plus grande flexibilité à la chaîne protéique. Il est donc très probable que cet accès au site actif laisse passer des molécules très différentes. L'analyse détaillée du SRS3 montre que la glycine 242 des psoralène synthases et l'alanine 242 de CYP71AJ4 est remplacée par une lysine chez CYP71AJ13 qui confère un encombrement stérique plus important. De même, au niveau du SRS4 la cystéine 299 de CYP71AJ1-4 est remplacée par une sérine chez CYP71AJ13. Or les cystéines sont des acides aminés souvent impliqués dans la formation de ponts disulfures, son remplacement par un autre acide aminé pourrait induire des changements de conformation de la protéine. Enfin, au niveau du SRS5 la thréonine 374 de CYP71AJ1-4 est remplacée par une proline chez CYP71AJ13 et la proline 380 de CYP71AJ1-4 est remplacée par une adénine chez CYP71AJ13. Or, la proline est un acide aminé qui induit des contraintes conformationnelles chez les protéines ; elle induit un point de rupture dans une hélice α et peut déstabiliser un feuillet β. Ces changements d'acides aminés peuvent induire de grandes modifications de structure secondaire et ces résultats indiquent que CYP71AJ13 aurait une fonction différente de CYP71AJ1-4. Ces analyses *in silico* semblent donc indiquer très clairement que malgré des homologies de séquences assez importantes, il est fort probable que ces enzymes ne soient pas impliquées dans la synthèse des mêmes types de molécules.

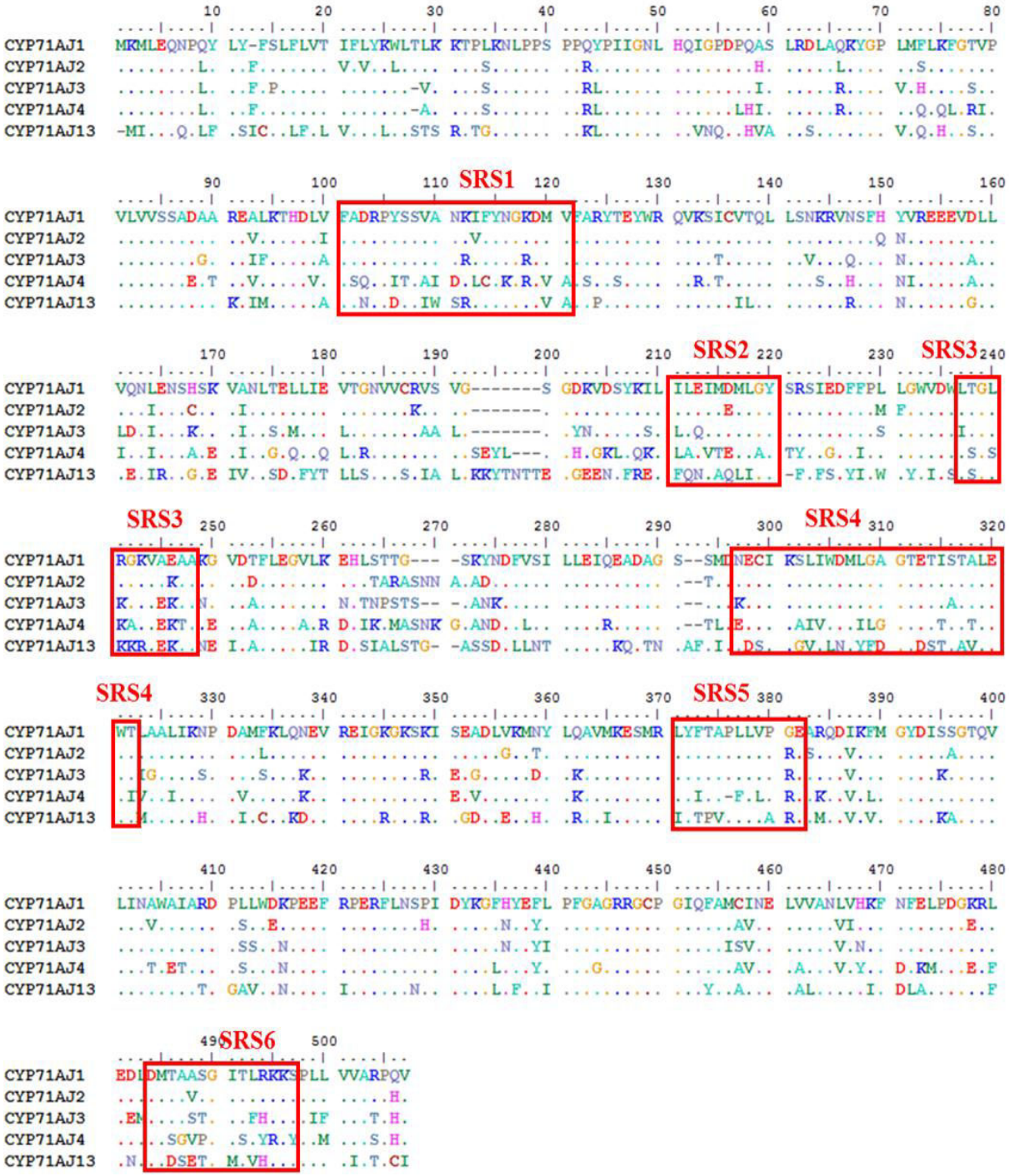


Figure 86 : Alignement des séquences peptidiques de CYP71AJ1 (*Ammi majus*), CYP71AJ2 (*Apium graveolens*), CYP71AJ3 (*P. sativa*), CYP71AJ4 (*P. sativa*) et CYP71AJ13 (*P. sativa*). Les six séquences potentielles de reconnaissance du substrat (SRSs) sont encadrées en rouge.

3.4 Discussion

Dans ce chapitre, nous nous sommes focalisés sur la recherche de membres de la sous-famille CYP71AJ, à laquelle appartiennent la psoralène synthase (CYP71AJ3) et l'angélicine synthase (CYP71AJ4) (Larbat *et al*., 2007, 2009). L'analyse de la banque de données a montré l'existence de nouveaux candidats. Parmi les différents fragments identifiés, un seul était complet : il s'agit de CYP71AJ13. Des approches de TAIL-PCR réalisées sur de l'ADN génomique ont été utilisées en vue d'identifier les séquences complètes correspondant aux autres séquences partielles isolées. Cette approche nous a permis d'identifier 2 nouvelles séquences codantes complètes correspondant à des P450s et de rallonger une séquence partielle. Dans le cas des séquences complètes, il est à présent nécessaire de cloner les séquences codantes à partir d'ARNm pour éliminer les introns mis en évidence sur l'ADN génomique. A ce jour les tentatives d'amplification par une approche de PCR se sont révélées être négatives. Plusieurs arguments peuvent être avancés pour expliquer cet échec. Si on réussit à amplifier le gène à partir de l'ADN génomique, cela signifie qu'il existe dans le génome de la plante. L'absence d'amplification à partir d'ARNm pourrait donc s'expliquer par une absence d'expression ou une trop faible expression du gène au sein des ARN messagers préparés. Il est possible que les plantes utilisées pour la préparation de la banque ait subit des stress particuliers ayant stimulé la synthèse de cet ARNm, stress que nous n'avons pas pu ou su reproduire dans le second lot de plantes. L'absence d'amplification peut également être liée à la nature de la population de plantes sur laquelle nous travaillons. En effet, des analyses faites sur la banque d'ADN génomique de *P. sativa* générée par le Dr. Sandro Roselli indiquent une variabilité génétique assez importante. Il est possible que les amorces commandées ne puissent s'hybrider sur les ARNm extraits de certaines plantes à cause de quelques modifications ponctuelles sur leur site d'hybridation.

Grâce à l'approche de TAIL-PCR, nous avons également rallongé des séquences de contigs disponibles dans la banque d'ADNc. Même si nous n'avons pas pu aller au bout de la démarche, il s'agit tout de même d'un premier pas vers la reconstruction d'un troisième cytochrome P450 de la famille CYP71.

Même si l'approche de TAIL-PCR semble être une solution pour l'identification de séquences complètes, ce n'est pas une solution fiable à 100%. En effet l'amplification aspécifique obtenue à partir du 4^{ième} contig montre que l'utilisation d'amorces dégénérées permet également d'amplifier des gènes qui n'ont rien à voir avec la cible recherchée.

Pour achever le travail de reconstruction des gènes restants, d'autres méthodes peuvent être utilisées comme la 5' RACE-PCR et/ou 3' RACE-PCR ou des approches de PCR inverse.

Ce travail a également mis en évidence les limites de l'affiliation de fragments de gènes à une sous-famille de cytochromes P450 donnée. La proximité des séquences peut générer des erreurs comme cela a été le cas pour deux contigs candidats qui au final appartiennent à la sous-famille CYP71AQ.

Si les approches génomique et transcriptomique permettent de faire des avancées extraordinaires dans l'identification de gènes candidats pour certaines voies de biosynthèse, les travaux décrits dans ce chapitre démontre également la difficulté d'aller ensuite vers la caractérisation fonctionnelle des enzymes correspondantes. De prime abord, CYP71AJ13 semble être un bon candidat pour son implication dans la voie de biosynthèse des furanocoumarines. L'hypothèse que nous avons émise au début de ce travail était que, à l'image de la voie de biosynthèse des benzoxazinoïdes chez le maïs (Frey *et al*., 2009), l'ensemble des étapes de la voie de biosynthèse de furanocoumarines puisse être réalisées par des enzymes de la même sous-famille de cytochrome P450. Les résultats obtenus dans ce chapitre de mon travail de thèse fragilisent cette hypothèse de manière considérable. En effet, le criblage métabolique relativement large n'a pas permis de mettre en évidence de fonction pour CYP71AJ13. Deux hypothèses peuvent être avancées pour expliquer cette absence de métabolisation : 1) le substrat de l'enzyme n'a pas été identifié lors du criblage métabolique, 2) la protéine n'est pas fonctionnelle.

Il est possible que cette absence de métabolisation soit tout simplement liée à une absence de reconnaissance de l'enzyme des différents substrats que nous lui avons présenté. Mais ce n'est pas la seule hypothèse. Il est également envisageable que la qualité de l'enzyme utilisée soit en cause. En effet, il est bien connu par l'homme de l'art que l'expression de ces enzymes n'est pas toujours facile de par la fragilité des protéines, entre autre. Afin de s'affranchir de cette hypothèse, nous avons mis en place divers tests. Nous avons vérifié la production de notre protéine recombinante par western blot, qui est une technique très sensible permettant de révéler une très faible quantité de protéine. Si ce premier test permet de voir que la protéine est bien exprimée, il ne nous indique pas si la protéine est présente en quantité suffisante ou même qu'elle ne soit fonctionnelle. En effet, l'ajout d'une queue poly histidine pourrait altérer la fonction de la protéine. Une solution pour vérifier la fonctionnalité des cytochromes P450 consiste à réaliser un spectre différentiel en présence de CO (spectre de type I). Malheureusement ce test, s'il est réalisable sur des enzymes fortement exprimées dans le système levure (comme c'est le cas par exemple de CYP73A1 du topinambour), n'a pas pu être mis en œuvre dans le cas de ce P450 de panais. Il est également envisageable que la protéine soit fonctionnelle, mais que le transfert d'électrons ne soit pas assuré suite à un mauvais couplage entre le P450 et la CPR. Une alternative possible serait de construire une souche de levure recombinante qui, au lieu de surexprimer les CPRs d'Arabidopsis

(comme c'est le cas des souches WAT11 et WAT21) surexprimerait une CPR de *P. sativa*. Des gènes codants pour ces enzymes ont été mis en évidence dans la banque d'ADNc (voir partie 1 de ce chapitre). Enfin, pour tenter d'améliorer le niveau d'expression de CYP71AJ13 différentes approches déjà décrites pourraient être mise en œuvres. Certaines concernent directement la séquence, d'autres le système d'expression. Dans le premier cas, une première option pourrait être le remplacement de l'extrémité N-terminale de CYP71AJ13 par celle de la C4H de topinambour. Cette approche a déjà été décrite dans la littérature et a notamment permis d'améliorer considérablement le niveau d'expression de ces certains P450s (Batard *et al*., 2000; Larbat *et al*., 2007). Toutefois, cette approche a également tentée dans notre laboratoire par le Dr. Tsunashi Kamo (chercheur invité de l'université de Shinshu (Japon) sur CYP71AJ5 sans succès. La deuxième alternative pourrait consister à entièrement recoder la séquence du gène afin d'optimiser son expression dans le système hétérologue choisi (Gustafsson *et al*., 2004; Perlak *et al*., 1991; Rouwendal *et al*., 1997). Concernant le système d'expression, d'autres approches que la levure peuvent être envisagées. Comme nous l'avons vu dans la partie précédente, le protocole d'expression transitoire de P450s chez *N. benthamiana* reste à optimiser, mais reste néanmoins une piste prometteuse. Dans le cadre d'une collaboration avec le Pr. Mizutani de l'université de Kobé (Japon), un travail d'expression hétérologue a également été initié afin de produire ces P450s dans des cellules d'insectes.

Les résultats obtenus au laboratoire depuis quelques années tendent tout de même à laisser croire que, même si on ne peut pas mettre en évidence de P450s par une approche de spectre différentielle, des enzymes actives sont quand même présentes dans les microsomes préparés. Dans ce cas, les différents tests réalisés dans le cadre de ce chapitre indiquent que le vrai substrat n'est pas présent dans la collection de molécules testées. Ce résultat n'est finalement pas si surprenant au regard des résultats obtenus dans le chapitre 2 où j'ai pu voir la diversité de fonctions des enzymes de la sous-famille CYP71AZ qui jouent un rôle dans la synthèse des furanocoumarines mais sont également impliqués dans la synthèse de coumarines. Il est donc vraisemblable que CYP71AJ13 pourrait également être impliqué dans la synthèse de coumarines. A ce jour, plus d'un millier de coumarines naturelles ont été identifiées dans le règne végétal.

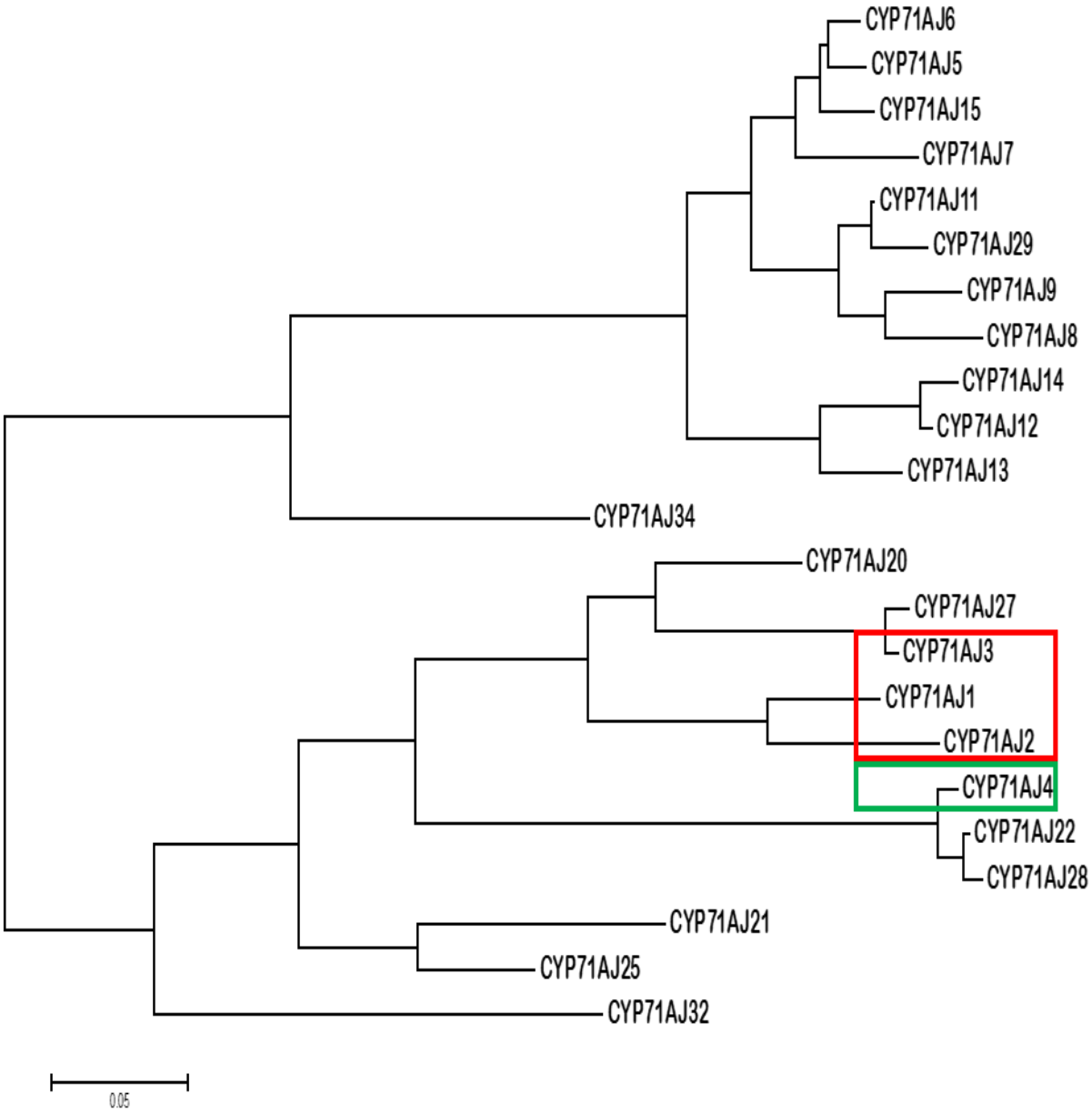


Figure 87 : Arbre phylogénétique des différents membres de CYP71AJ. Alignement de séquence réalisé avec ClustalX et arbre phylogénétique construit avec MEGA6 en utilisant la méthode de maximum likelihood. Les numéros d'accession sur GenBank/EMBL/DDBJ des gènes caractérisés sont les suivants : CYP71AJ1 (psoralène synthase) *A. majus* (Q6QNI4), CYP71AJ2 (psoralène synthase) *A. graveolens* (C0SJS4), CYP71AJ3 (psoralène synthase) *P. sativa* (C0SJS2), CYP71AJ4 (angélicine synthase) *P. sativa* (C0SJS3). En rouge : psoralène synthases ; en vert : angélicine synthase.

Pour tenter d'apporter une réponse à cette question, une collaboration entre le laboratoire Agronomie et Environnement et l'équipe du Pr. Simonsen de l'université de Copenhague a été mise en place en 2009. Dans le cadre de cette collaboration, Bjørn Dueholm a initié la recherche de gènes orthologues à

CYP71AJ chez un grand nombre d'Apiacées de la sous-famille des Apioidées à partir de banques d'ADNc publiques et privées. Il a ainsi pu mettre en évidence une trentaine de membres de cette sous-famille de P450s présents chez 17 différentes Apiacées. Parmis ces plantes, certaines sont connues pour produire des furanocoumarines comme par exemple *A. majus*, *A. graveolens*, *P. sativa*, *Thapsia garganica*, *Daucus carota*, *P. crispum*, *Heracleum mantegazziannum*, *Heracleum lanatum*, *Seseli montanum, Angelica archangelica* ou encore *Ferula communis*. De plus, la sous-famille CYP71AJ est également présente chez des plantes comme *Thapsia laciniata*, *Laserpitium siler*, *Anthriscus sylvestris*, *Peucedanum cervaria* et *Bupleurum chinense*. Or, la litérature n'indique pas clairement la présence de furanocoumarines chez ces plantes. A titre d'exemple, dans le cas de *Bupleurum chinense*, les travaux de Berenbaum et de Liu et de ses collaborateurs indiquent la présence de furanocoumarines linéaires et angulaires et notamment d'angélicine sans citer de sources alors que dans le cadre d'une étude phytochimique récente ces composés sont absents (Ashour and Wink, 2011; Berenbaum, 1983; Liu *et al*., 2013). Des travaux sont actuellement en cours au laboratoire afin de déterminer si cette plante produit des furanocoumarines ou non. En se basant sur les travaux de Downie, Bjørn Dueholm a construit un arbre phylogénétique à l'aide des séquences Internal Transcribed Spacer (ITS) de l'ADN ribosomique d'Apiacées de la sous-famille des Apioidées (Figure 88, p. 225) (Downie *et al*., 2001). L'occurrence de CYP71AJ chez les différentes espèces est indiquée sur la Figure 88 (p. 225). L'Apiacée la plus ancienne serait *B. chinense* chez qui un CYP71AJ inconnu a été identifié. Au cours de l'évolution, certaines espèces auraient acquis la psoralène synthase et c'est chez les espèces les plus évoluées que l'on retrouve les angélicine synthases, ce qui est cohérent avec l'occurrence des formes angulaires qui sont toujours associées aux formes linéaires et qui laissent penser que les formes linéaires sont apparues avant les formes angulaires au cours de l'évolution. Ces résultats appuyent l'hypothèse que cette sous-famille de P450s dérive d'un même gène ancestral et que par des phénomènes de mutations et de duplications ait acquit de nouvelles fonctions comme la psoralène et l'angélicine synthase. L'identification de la présence ou de l'absence de furanocoumarines chez les plantes pour qui les données bibliographiques sont limitées permettra d'apporter des éléments de réponse quant à la fonction des autres CYP71AJ identifiés, car si ces gènes sont présents chez des plantes qui ne produisent pas de furanocumarines, il est probable qu'ils soient impliqués dans une autre voie de biosynthèse que celle des furanocoumarines. Il est envisageable que ces enzymes soient impliquées dans la métabolisation de molécules proches des furanocoumarines, comme par exemples les coumarines, ou encore les pyranocoumarines.

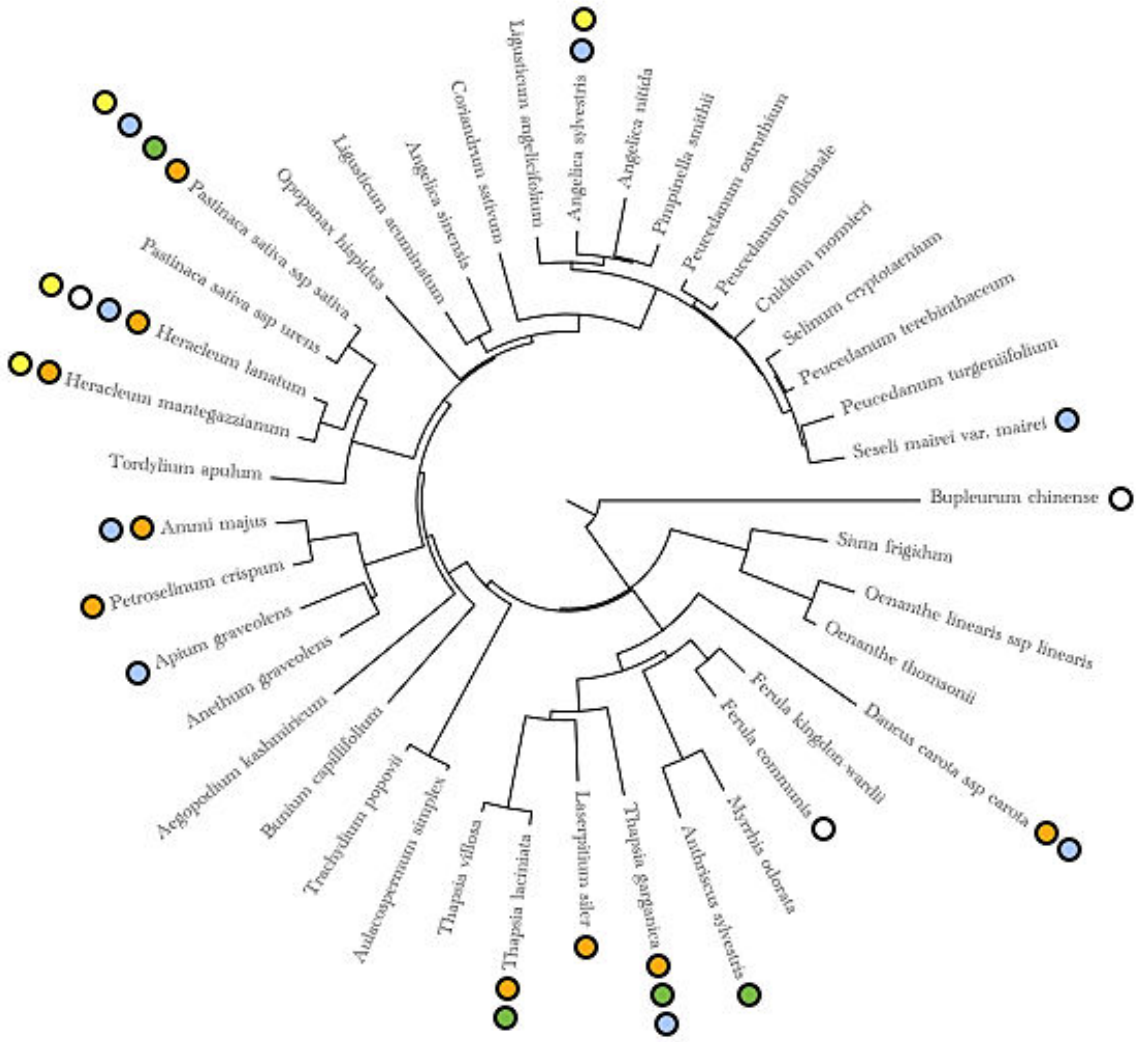


Figure 88 : Arbre phylogénétique de différentes Apioidées et occurrence de la sous-famille CYP71AJ. L'arbre a été construit à l'aide de MEGA5.2.2 à partir des séquences ITS identifiées par Downie et ses collaborateurs (Downie *et al*., 2001). En bleu : espèces chez qui une psoralène synthase ou une psoralène synthase putative a été identifiée ; en jaune : angélicine synthase ou angélicine synthase putative ; en orange : clade 1, inconnu ; en vert : clade 2, inconnu ; en blanc : inconnu.

Parmis les gènes identifiés par l'équipe de Copenhague, des premiers criblages métaboliques avec CYP71AJ5 ont été réalisés par les Dr. Tsunashi Kamo et Guilhem Vialart dans notre laboratoire mais n'ont pas permis de trouver la fonction de ce gène. J'ai également réalisé l'expression hétérologue de certaines de ces enzymes (CYP71AJ5, 6 et 12) au laboratoire mais aucune activité n'a pu être mise en évidence à ce jour, que ce soit sur des molécules de la famille des furanocoumarines ou d'autres. Une analyse phylogénétique réalisée sur ces différents orthologues de CYP71AJ permet de constater que l'arbre généré est subdivisée en deux clades distincts. On constate ainsi clairement que CYP71AJ1-3 et CYP71AJ4 sont regroupés d'un côté de l'arbre tandis qu'une majorité d'entre eux sont localisés sur la

seconde branche dont CYP71AJ5, 6, 12 et 13 (Figure 87, p. 223). Ces résultats appuyent le fait que cette sous-famille de P450s a sûrement évoluée au sein des Apiacées pour être impliquée dans la synthèse de furanocoumarines, de coumarines voire d'autres familles de molécules.

3.5 Conclusion

Dans le cadre de cette troisième partie du chapitre « Résultats et Discussion », je me suis focalisée sur l'identification et la caractérisation fonctionnelle de membres de la sous-famille CYP71AJ.

L'analyse de la banque d'ADNc de *P. sativa* a permis d'identifier un certain nombre de gènes candidats. Parmi les fragments candidats, un seul comportait la totalité de la séquence codante, CYP71AJ13.

Des approches de TAIL-PCR ont été utilisées en vue de compléter les séquences candidates partielles de certains candidats mais leur clonage n'a pas été possible.

J'ai initié la caractérisation fonctionnelle de CYP71AJ13 dans le système d'expression hétérologue *S. cerevisiae*. Le criblage métabolique réalisé sur près de 60 molécules différentes s'est révélé infructueux à ce jour.

Une recherche d'autres gènes codant pour des enzymes de la sous-famille CYP71AJ a été entreprise en collaboration avec Bjørn Dueholm, étudiant du groupe du Pr. Simonsen de l'université de Copenhague. Ces travaux ont permis de mettre en évidence une grande diversité de ces enzymes au sein de nombreuses Apiacées, y compris des plantes pour qui la synthèse de furanocoumarines n'est pas clairement illustrée dans la littérature.
Le criblage fonctionnel et l'analyse des séquences de CYP71AJ orthologues m'amène à conclure qu'il est vraisemblable que CYP71AJ13 ne soit pas impliqué dans le métabolisme des furanocoumarines.

CONCLUSION GENERALE ET PERSPECTIVES

Conclusion générale et perspectives

L'objectif de ce travail de thèse était de mieux comprendre le fonctionnement et d'identifier les chainons moléculaires manquants de la voie de biosynthèse des furanocoumarines. Lorsque ce projet de thèse a débuté, très peu de gènes directement impliqués dans cette voie de biosynthèse avaient été décrits. Un des principaux freins à l'identification de nouveaux gènes jouant un rôle dans la synthèse de ces molécules était que, contrairement à des plantes modèles comme Arabidopsis ou encore la tomate, très peu de données moléculaires étaient disponibles pour les plantes productrices de furanocoumarines. Les avancées technologiques réalisées dans le domaine du séquençage haut débit au cours de ces dernières années ont rendu plus accessible les données de transcriptomique et génomique pour des plantes peu étudiées. Le travail réalisé dans le cadre de cette thèse a été subdivisé en 2 grandes parties. La première a consisté à générer des banques de données moléculaires basées sur les transcriptomes de 3 plantes produisant des furanocoumarines différentes et à en réaliser une analyse globale sans *a priori*. Une analyse plus fine des données ainsi générées a, dans un second temps, permis d'identifier de nombreux gènes candidats dont la fonction dans la synthèse des furanocoumarines a été établie pour un certain nombre d'entre eux.

Construction et analyse de trois banques transcriptomiques de plantes productrices de furanocoumarines en vue d'identifier de nouveaux gènes candidats impliqués dans la synthèse de furanocoumarines.

Les furanocoumarines ont été décrites dans seulement 4 familles botaniques différentes qui, d'un point de vue évolutif, sont très éloignées. L'hypothèse d'une apparition par le biais d'une évolution convergente a souvent été émise mais n'a jamais pu être démontrée jusque-là par manque de données expérimentales. Pour commencer à lever le voile sur cette boite noire, nous avons choisi d'initier l'étude du transcriptome de 3 plantes représentatives de 3 de ces familles : *P. sativa* (Apiacées), *R. graveolens* (Rutacées) et *C. cinereum* (Fabacées). Ces trois plantes présentent des profils « furanocoumariniques » différents : *P. sativa* produit des formes linéaires et angulaires simples et complexes, *R. graveolens* ne produit que des formes linéaires simples et complexes et *C. cinereum* produit uniquement des formes linéaires et angulaires simples (psoralène et angélicine).

La voie de biosynthèse des furanocoumarines est une voie appartenant au métabolisme secondaire qui par définition permet la synthèse de molécules présentes en faible concentration dans les plantes. De fait, il est possible que les gènes qui en permettent la synthèse soient transcrits en faible quantité par

rapport à des gènes impliqués dans des voies métaboliques plus importantes. Afin de maximiser les chances d'avoir des représentants d'un maximum de gènes, nous avons choisi de générer des banques normalisées ce qui limite les possibilités de réaliser des analyses quantitatives de l'expression de gènes mais permet une analyse qualitative plus large.

La comparaison des séquences contenues dans ces trois banques d'ADNc normalisées avec celles de la banque publique Genbank (NCBI) a permis d'annoter la quasi-totalité des contigs. Etant donné que la majorité des étapes de la synthèse de furanocoumarines sont catalysées par des P450s, je me suis focalisée dans le cadre de ce travail sur l'identification de cette famille d'enzymes et de son partenaire redox, les CPRs. Cette approche a permis d'identifier plusieurs centaines de séquences partielles codant pour des P450s (336 chez *P. sativa*, 279 chez *R. graveolens* et 221 chez *C. cinereum*) plus ou moins longues (seules 7, 26 et 50 séquences sont prédites complètes respectivement chez *P. sativa*, *R. graveolens* et *C. cinereum*). L'affiliation de certains de ces P450s a pu être réalisée en réalisant des alignements avec des séquences présentes dans une base de données fournie par le Pr. Nelson. L'analyse de ces trois banques d'ADNc a également permis d'identifier un certain nombre de fragments codant pour des CPR (17 chez *P. sativa*, 13 chez *R. graveolens* et 9 chez *C. cinereum* et une seule est prédite complète dans chacune des banques). Parallèlement à la recherche de séquences codant pour des cytochromes P450s et de CPR, ces banques ont également permis d'identifier un certain nombre d'autres familles de gènes. Ainsi, un travail réalisé dans le cadre de la thèse de Guilhem Vialart a mené à la caractérisation fonctionnelle d'enzymes de la famille des dioxygénases qui catalysent la formation de l'umbelliférone, un précurseur de la voie de biosynthèse des furanocoumarines, et de la scopolétine. Un gène présentant de fortes homologies avec cette dioxygénase identifié dans la banque de *R. graveolens* a également été mis en évidence dans la banque de panais. La fonction de ce gène est actuellement à l'étude. Les résultats préliminaires semblent indiquer que cette enzyme dispose de caractéristiques enzymatiques légèrement différentes de celle codée par le gène orthologue de la rue, car elle est uniquement impliquée dans la synthèse d'umbelliférone et est donc spécifique de la voie de biosynthèse des furanocoumarines. Un autre travail réalisé dans le cadre de la thèse de Fazeelat Karamat au laboratoire a permis d'identifier une prényltransférase catalysant à la fois la synthèse de DMS et d'osthenol à partir de l'umbelliférone chez le persil. L'analyse de la banque d'ADNc de panais indique l'existence de deux gènes codant pour ce même type d'enzyme. Leur caractérisation fonctionnelle actuellement en cours dans le cadre d'une collaboration avec l'équipe du Pr. Yazaki (université de Kyoto, Japon) semble montrer un rôle plus spécifique de chacune dans la synthèse des furanocoumarines linéaires ou angulaires permettant ainsi d'émettre de nouvelles hypothèses sur l'apparition des voies de biosynthèses respectives.

Si les données transcriptomiques sont un prérequis pour identifier de nouveaux gènes candidats, la caractérisation fonctionnelle reste sans doute le verrou le plus compliqué à lever Avec la construction de ces banques, les séquences codantes de nombreux P450s ont été révélées. L'étape suivante a donc été d'identifier celles qui codent pour une enzyme impliquée dans la synthèse des furanocoumarines. Deux approches peuvent être envisagées. La première consiste à cloner la totalité des séquences codantes, à produire les enzymes correspondantes et à tester leur capacité à métaboliser un des intermédiaires de la synthèse des furanocoumarines. Cette première stratégie est longue car, elle nécessite la recherche des séquences complètes de tous les gènes partiels identifiés jusque-là. Cette approche semble irréalisable dans le cadre d'une thèse. La seconde approche consiste à réaliser un premier criblage afin d'éliminer un certain nombre de candidats. Ce premier tri peut se faire en réalisant des analyses différentielles de niveaux d'expression des certains gènes sur des tissus produisant peu ou beaucoup de furanocoumarines. Au cours de ma thèse j'ai tenté de développer un tel système sur l'une des plantes étudiée au laboratoire. Des essais d'élicitations ont été réalisés sur des plantules de panais en utilisant différents types d'inducteurs : les traitements aux UVB ou encore de traitement au MeJa. Ces différentes approches n'ont pas permis d'affecter les teneurs en furanocoumarines présentes dans les plantes. D'autres pistes d'élicitations devront à présent être explorées comme la mise au point, par exemple, de lignées cellulaires plus facilement inductibles qu'une plante entière.

Un travail d'élicitation a été réalisé sur des cultures cellulaires d'*A. majus* dans le cadre de travaux réalisés par Sylvia Specker et Sandra Kellner, des doctorantes du Pr. Matern de l'université de Marburg. Dans cette équipe la plante modèle était *A. majus*, une plante de la famille des Apiacées mais qui ne produit que des furanocoumarines linéaires. Ce travail exploratoire préliminaire a conduit à l'identification de 2 sous-familles de P450s, CYP71AJ et CYP71AZ dont l'expression semblait corrélée avec la synthèse des furanocoumarines. Des travaux réalisés dans le cadre de la thèse de Romain Larbat ont permis de valider cette hypothèse pour la sous-famille CYP71AJ identifiées comme étant des psoralène synthases et l'angélicine synthase. L'analyse des banques d'ADNc montre que ces 2 sous-familles de gènes n'étaient pas présentes chez la rue ou chez *C. cinereum*. Par contre, un certain nombre de gènes présentant de fortes homologies avec les gènes identifiés chez *A. majus* ont été détectés.

Etude et caractérisation fonctionnelle d'enzymes de la sous-famille CYP71AZ chez *P. sativa*.

La recherche de CYP71AZ dans la banque d'ADNc de *P. sativa* a conduit à l'identification de 10 fragments candidats dont deux correspondaient à une séquence codante complète. CYP71AZ3 et CYP71AZ4 ont été exprimées *in vitro* dans le système hétérologue *S. cerevisiae*. Les résultats de ces expérimentations ont montré que CYP71AZ3 était capable d'hydroxyler l'esculétine en un produit inconnu et que CYP71AZ4 présentait un spectre d'activité plus large en hydroxylant le psoralène en position 8 et 5 sur le noyau coumarine. Ces différents substrats de CYP71AZ3 et CYP71AZ4 ont fait l'objet d'une recherche ciblée dans des feuilles de panais en utilisant des approches analytiques fines telles que des analyses LC-MS-LTQ. Ces analyses nous ont montré que seuls l'esculétine et le psoralène sont détectés au sein de la plante. Ce qui semble donc indiquer, qu'*in vivo* ces deux molécules pourraient être les substrats préférentiels de ces enzymes. Ces résultats sont cohérents avec le calcul des paramètres catalytiques de CYP71AZ4 pour les différents substrats qui indiquent que le psoralène semble être le substrat privilégié de cette enzyme avec un K_m conforme à ce qui a été décrit pour d'autres enzymes. Les autres produits formés par CYP71AZ4 seraient donc dus à des métabolisations opportunistes de cette enzyme. Pour CYP71AZ3, les résultats doivent être manipulés avec plus de prudence dans la mesure où le K_m calculé pour l'esculétine est beaucoup plus élevé que ce qui pouvait être attendu, sous entendant que cette enzyme dispose probablement d'un autre substrat préférentiel. Parmi les différents produits formés par ces deux enzymes *in vitro*, un certain nombre de produits n'ont pas encore été identifiés (deux produits formés par CYP71AZ4 et le produit formé à partir de l'esculétine par CYP71AZ3). Leur caractérisation est actuellement en cours par des approches de RMN. L'identification des sites d'hydroxylations devraient permettre de mieux comprendre comment les différents substrats se positionnent au sein du site actif de ces deux enzymes et de confirmer l'hypothèse que CYP71AZ4 est spécialisé dans l'hydroxylation en C8 du noyau 2H-1-benzopyran-2-one commun aux coumarines et furanocoumarines. Si la fonction exacte de CYP71AZ3 n'est pas encore entièrement caractérisée, il est cependant clair que CYP71AZ3 et CYP71AZ4 présentent des fonctions différentes malgré une très forte homologie de séquences. Un travail sur les relations structure-fonction est également entreprit au laboratoire avec la génération de différents mutants. Ces mutants sont actuellement en cours d'analyse et devraient permettre d'identifier les régions impliquées dans les différentes fonctions de ces enzymes.

Identification de variants de CYP71AJ chez *P. sativa*.

L'analyse de la banque d'ADNc de *P. sativa* a permis d'identifier de nouveaux membres de CYP71AJ dont un complet CYP71AJ13. L'existence de plusieurs variants alléliques de CYP71AJ3 et CYP71AJ4 au sein de la même plante me laisse supposer que ces variants pourraient être synthétisés à différents stades de développement de la plante ou en réponse à différents stress environnementaux.

Afin de trouver la fonction de ce nouveau membre de la famille CYP71AJ, j'ai entrepris d'exprimer la protéine dans un système d'expression hétérologue. Il est bien connu de l'homme de l'art que l'expression de P450s dans une optique de caractérisation fonctionnelle est une réelle gageure. Chaque P450 est un nouveau challenge. Une des limites des systèmes d'expression hétérologue existants réside dans le niveau d'expression des protéines mais également dans l'activité des protéines produites. Si le niveau d'expression peut être amélioré en optimisant entre autre les conditions de cultures ou encore le biais de codon, l'activité peut être étroitement liée aux interactions qui s'établissent ou doivent s'établir avec le partenaire privilégié des P450s, à savoir la cytochrome NADPH P450 réductase. Des travaux pionniers réalisés par Pompon, Urban et leurs collaborateurs ont permis de montrer que l'expression de P450s végétaux dans un système levure était sensiblement amélioré lorsque la CPR endogène était remplacée par une CPR d'origine végétale et notamment par une CPR clonée à partir d'*A. thaliana* (Urban *et al*., 1997). Les CPRs sont relativement bien conservées d'une plante à l'autre et par conséquent le fait de remplacer les CPRs d'*Arabidopsis* par une CPR provenant d'une autre plante risque de rester sans effet. Néanmoins, dans la mesure où ces données sont disponibles, il est envisageable dans notre cas, de cloner une ou plusieurs CPR de *P. sativa* afin d'optimiser le système d'expression.

Plusieurs autres systèmes d'expression ont été mis au point et adaptés pour les P450s. Il serait également envisageable d'utiliser d'autres systèmes d'expression hétérologue. Dans le cadre d'une collaboration avec le Pr. Mizutani de l'université de Kobé au Japon, CYP71AJ13 ainsi que d'autres P450s candidats ont été clonés en vue de les exprimer dans des cellules d'insectes par le biais de baculovirus. Une autre alternative serait d'exprimer ces gènes de manière transitoire ou stable dans des plantes productrices de furanocoumarines. De nombreux protocoles ont déjà été décrits dans la littérature pour la transformation de *D. carota* et d'autres Apiacées et pourraient être transposables à *P. sativa* (Baranski, 2008).

Malgré toutes ces pistes, à ce stade les différents systèmes ou stratégies utilisées pour exprimer et caractériser les nouveaux membres de la sous-famille CYP71AJ sont restés infructueux.

Que faire maintenant ?

Le travail réalisé dans le cadre de mon projet de thèse a permis d'identifier 2 nouveaux gènes impliqués dans la synthèse des furanocoumarines chez le panais. De la voie de biosynthèse complète, il ne reste donc plus qu'à identifier un dernier P450 ayant une activité de marmesine synthase. Pour identifier ce gène, il va être nécessaire de mettre au point un système permettant de générer des cellules de panais produisant peu ou beaucoup de furanocoumarines. Une fois ce système établi, un criblage par RT-PCR quantitative pourra être utilisé pour tenter de rechercher les gènes exprimés différentiellement en nous basant sur les séquences de P450s disponibles dans la banque de panais. Une approche similaire a été développée sur d'autres Rutacées, les Citrus, dans le cadre du travail de master de Victoire Coqueret et de la thèse d'Audray Dugrand. Dans ce cas, la recherche différentielle a été effectuée sur des variétés de Citrus disponibles sur le site expérimental de Corse qui produisent des quantités différentes de furanocoumarines. Une analyse différentielle ciblée sur différentes familles de P450s a permis d'identifier des familles de gènes potentiellement intéressantes. Des résultats préliminaires ont permis d'identifier un P450 de la famille CYP82 présentant des similarités fonctionnelles avec CYP71AZ6, un P450 identifié dans une banque génomique de panais construite au laboratoire dans le cadre d'un projet géré par le Dr. Sandro Roselli.

Les travaux réalisés par Hamerski et Matern dans les années 1980 ont mis en évidence que la plupart des étapes de la voie de biosynthèse des furanocoumarines chez les Apiacées étaient catalysées par des P450s. Ces résultats n'ont jamais été confirmés chez les autres plantes produisant ces molécules. A ce jour, les seuls gènes codants pour des cytochromes P450 impliqués dans la synthèse de furanocoumarines sont issus de 2 familles de plantes : les Apiacées et les Brassicacées. *Arabidopsis* dispose de deux enzymes de la sous-famille CYP82C, qui sont capables de transformer la xanthotoxine en 5-hydroxy xanthotoxine (Kruse *et al.*, 2008). Toutefois, *Arabidopsis* n'appartient pas aux familles de plantes décrites comme produisant ce type de molécules. Il est donc très probable que cette activité décrite par Kruse ne soit que le reflet d'une métabolisation opportuniste. Ce type de métabolisation a déjà été décrite pour d'autres P450s et notamment dans le cas de métabolisation d'herbicides comme le chlortoluron qui n'est pas présent chez les plantes en règle générale (Gion *et al*., 2014). Néanmoins, d'autres enzymes sont capables de réaliser le même type de réaction comme par exemple les dioxygénases. En effet, il a été démontré que certaines réactions catalysées par des P450s chez certaines plantes pouvaient l'être par des enzymes de la famille des dioxygénases alpha-cétoglutarate dépendantes chez d'autres plantes. C'est le cas notamment de la F3H, une flavone synthase, qui est un P450 de la famille CYP93 chez certaines plantes comme le soja et une dioxygénase chez le persil (Bredebach *et al*., 2011; Fliegmann *et al*., 2010). L'absence des familles CYP71AJ et CYP71AZ chez les Psoralées et les Rutacées est un argument supplémentaire pouvant renforcer cette hypothèse. De plus, des résultats

récents au laboratoire ont montré la présence d'une activité marmesine synthase chez *N. benthamiana* lorsque les feuilles sont infiltrées avec de la DMS. Ces résultats sont probablement le reflet d'une métabolisation opportuniste de la DMS et laissent penser à une activité de type dioxygénase. Il pourrait donc être intéressant de ne pas se limiter à l'exploration des fonctions de P450s chez ces 2 plantes. Si cette hypothèse venait à être vérifiée, elle confirmerait l'apparition de cette voie de biosynthèse chez différentes familles botaniques par le biais d'une évolution convergente.

Pour finir, les furanocoumarines représentent des molécules très intéressantes pour l'industrie pharmaceutique. En effet, elles sont utilisées pour traiter des affections de la peau comme le psoriasis et le vitiligo (Domínguez *et al*., 2004; Pathak and Fitzpatrick, 1992), mais également dans le traitement de lymphomes cutanés à cellules T (Chiesa-Fuxench and González-Chávez, 2010). Ces molécules semblent également avoir un avenir prometteur dans l'industrie pharmaceutique avec leur potentielle utilisation dans le traitement de la dépression (Chen *et al*., 2007), dans le traitement de certains cancers de par leur activité antiproliférative et apoptotique (Sigurdsson *et al*., 2004; Sumiyoshi *et al*., 2014), et leur propriétés bactériostatiques, fongicides et antimicrobiennes (Adams *et al*., 2006; Girennavar *et al*., 2007; Sardari *et al*., 1999). A ce jour, les furanocoumarines utilisées dans l'industrie pharmaceutique proviennent essentiellement de l'extraction d'huile de bergamote (*Citrus bergamia*), car leur synthèse chimique reste un processus très onéreux. Or, la culture de la bergamote se fait de plus en rare de nos jours et il devient nécessaire de trouver des plantes alternatives pour la production de furanocoumarines. Le panais pourrait être une plante adaptée à la production de ces molécules d'intérêt étant donné qu'elle produit en grande quantité de nombreuses furanocoumarines à la fois linéaires et angulaires. Il serait intéressant de développer un protocole de transformation stable de *P. sativa* en vue de surexprimer les différents gènes déjà caractérisés et impliqués dans la voie de biosynthèse des furanocoumarines afin de déterminer si cette transformation entraîne une augmentation significative de la production de ces molécules *in planta*. De plus, lorsque toutes les enzymes de cette voie de biosynthèse seront identifiées et caractérisées fonctionnellement, il serait intéressant de reconstituer cette voie de biosynthèse chez une plante qui ne produit pas naturellement de furanocoumarines, comme par exemple chez la plante modèle *A. thaliana*. De la même manière, lorsque toutes les enzymes de cette voie seront identifiées, il sera alors possible de produire ces molécules dans des organismes hétérologues (levure, plante, insecte…). La production de nouvelles plantes produisant des furanocoumarines présenterait un avantage agronomique grâce à leur potentiel allélopathique et/ou permettrait de produire en plus grande quantité ces molécules utilisées dans l'industrie thérapeutique. Enfin, les furanocoumarines contenues dans les agrumes et plus particulièrement dans le pomelo et le pamplemousse peuvent engendrer des phénomènes de toxicité lorsqu'elles sont ingérées avec certains médicaments. D'autres plantes comme le céleri contiennent de

fortes concentrations en furanocoumarines dans les feuilles pouvant engendrer de sérieuses photodermatites lorsqu'elles sont manipulées. Il serait donc intéressant de créer des plantes dépourvues de furanocoumarines par ingénierie métabolique afin d'assurer la sécurité des consommateurs et de celles des producteurs de plantes contenant ces molécules.

REFERENCES BIBLIOGRAPHIQUES

Références bibliographiques

Abdalla, D.S., Monteiro, H.P., Oliveira, J.A., and Bechara, E.J. (1986). Activities of superoxide dismutase and glutathione peroxidase in schizophrenic and manic-depressive patients. Clin. Chem. *32*, 805–807.

Abreu, M.T., von Tirpitz, C., Hardi, R., Kaatz, M., Van Assche, G., Rutgeerts, P., Bisaccia, E., Goerdt, S., Hanauer, S., and Knobler, R. (2009). Extracorporeal photopheresis for the treatment of refractory Crohn's disease: Results of an open-label pilot study. Inflamm. Bowel Dis. *15*, 829–836.

Abyshev, A.Z., Gindin, V.A., Kerimov, Y.B., Ismailov, É.S., Agaev, É.M., and Isaev, N.Y. (1992). Furocoumarins of *Ruta graveolens*. Chem. Nat. Compd. *28*, 382–383.

Adams, M., Ettl, S., Kunert, O., Wube, A.A., Haslinger, E., Bucar, F., and Bauer, R. (2006). Antimycobacterial activity of geranylated furocoumarins from *Tetradium daniellii*. Planta Med. *72*, 1132.

Afek, U., Carmeli, S., and Aharoni, N. (1995). Columbianetin, a phytoalexin associated with celery resistance to pathogens during storage. Phytochemistry *39*, 1347–1350.

Agati, G., and Tattini, M. (2010). Multiple functional roles of flavonoids in photoprotection. New Phytol. *186*, 786–793.

Akashi, T., Aoki, T., Takahashi, T., Kameya, N., Nakamura, I., and Ayabe, S. (1997). Cloning of cytochrome P450 cDNAs from cultured *Glycyrrhiza echinata* L. cells and their transcriptional activation by elicitor-treatment. Plant Sci. *126*, 39–47.

Ali, M.B., Yu, K.-W., Hahn, E.-J., and Paek, K.-Y. (2006). Methyl jasmonate and salicylic acid elicitation induces ginsenosides accumulation, enzymatic and non-enzymatic antioxidant in suspension culture Panax ginseng roots in bioreactors. Plant Cell Rep. *25*, 613–620.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.

Anaya, A.L., Macías-Rubalcava, M., Cruz-Ortega, R., García-Santana, C., Sánchez-Monterrubio, P.N., Hernández-Bautista, B.E., and Mata, R. (2005). Allelochemicals from *Stauranthus perforatus*, a Rutaceous tree of the Yucatan Peninsula, Mexico. Phytochemistry *66*, 487–494.

Anthony, F.A., Laboda, H.M., and Costlow, M.E. (1997). Psoralen-fatty acid adducts activate melanocyte protein kinase C: a proposed mechanism for melanogenesis induced by 8-methoxypsoralen and ultraviolet A light. Photodermatol. Photoimmunol. Photomed. *13*, 9–16.

Appendino, G., Bianchi, F., Bader, A., Campagnuolo, C., Fattorusso, E., Taglialatela-Scafati, O., Blanco-Molina, M., Macho, A., Fiebich, B.L., and Bremner, P. (2004). Coumarins from *Opopanax c hironium*. New Dihydrofuranocoumarins and Differential Induction of Apoptosis by Imperatorin and Heraclenin. J. Nat. Prod. *67*, 532–536.

Arakaki, A.K., Huang, Y., and Skolnick, J. (2009). EFICAz2: enzyme function inference by a combined approach enhanced by machine learning. BMC Bioinformatics *10*, 107.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., and Eppig, J.T. (2000). Gene Ontology: tool for the unification of biology. Nat. Genet. *25*, 25–29.

Ashour, M.L., and Wink, M. (2011). Genus Bupleurum: a review of its phytochemistry, pharmacology and modes of action. J. Pharm. Pharmacol. *63*, 305–321.

Bairoch, A. (2000). The ENZYME database in 2000. Nucleic Acids Res. *28*, 304–305.

Bak, S., Kahn, R.A., Olsen, C.E., and Halkier, B.A. (1997). Cloning and expression in *Escherichia coli* of the obtusifoliol 14α-demethylase of *Sorghum bicolor* (L.) Moench, a cytochrome P450 orthologous to the sterol 14α-demethylases (CYP51) from fungi and mammals. Plant J. *11*, 191–201.

Bak, S., Nielsen, H.L., and Halkier, B.A. (1998). The presence of CYP79 homologues in glucosinolate-producing plants shows evolutionary conservation of the enzymes in the conversion of amino acid to aldoxime in the biosynthesis of cyanogenic glucosides and glucosinolates. Plant Mol. Biol. *38*, 725–734.

Bak, S., Olsen, C.E., Halkier, B.A., and Møller, B.L. (2000). Transgenic Tobacco and Arabidopsis Plants Expressing the Two Multifunctional Sorghum Cytochrome P450 Enzymes, CYP79A1 and CYP71E1, Are Cyanogenic and Accumulate Metabolites Derived from Intermediates in Dhurrin Biosynthesis. Plant Physiol. *123*, 1437–1448.

Baranski, R. (2008). Genetic transformation of carrot (*Daucus carota*) and other apiaceae species. Trans Plant J *2*, 18–38.

Barouki, R., and Morel, Y. (2001). Repression of cytochrome P450 1A1 gene expression by oxidative stress: mechanisms and biological implications. Biochem. Pharmacol. *61*, 511–516.

Barrell, D., Dimmer, E., Huntley, R.P., Binns, D., O'Donovan, C., and Apweiler, R. (2009). The GOA database in 2009—an integrated Gene Ontology Annotation resource. Nucleic Acids Res. *37*, D396–D403.

Baskin, J.M., Ludlow, C.J., Harris, Tm., and Wolf, F.T. (1967). Psoralen, an inhibitor in the seeds of *Psoralea subacaulis* (Leguminosae). Phytochemistry *6*, 1209–1213.

Bastide, M., and McCombie, W.R. (2007). Assembling genomic DNA sequences with PHRAP. Curr. Protoc. Bioinforma. 11–14.

Batard, Y., Hehn, A., Nedelkina, S., Schalk, M., Pallett, K., Schaller, H., and Werck-Reichhart, D. (2000). Increasing expression of P450 and P450-reductase proteins from monocots in heterologous systems. Arch. Biochem. Biophys. *379*, 161–169.

Baudry, J., Li, W., Pan, L., Berenbaum, M.R., and Schuler, M.A. (2003). Molecular docking of substrates and inhibitors in the catalytic site of CYP6B1, an insect cytochrome P450 monooxygenase. Protein Eng. *16*, 577–587.

Bauer, N., Fulgosi, H., and Jelaska, S. (2011). Overexpression of Phenylalanine Ammonia-Lyase in Transgenic Roots of *Coleus blumei* Alters Growth and Rosmarinic Acid Synthesis. Food Technol. Biotechnol. *49*, 24–31.

Becerra, J.X. (2007). The impact of herbivore–plant coevolution on plant community structure. Proc. Natl. Acad. Sci. *104*, 7483–7488.

Becerra, J.X., Venable, D.L., Evans, P.H., and Bowers, W.S. (2001). Interactions between chemical and mechanical defenses in the plant genus Bursera and their implications for herbivores. Am. Zool. *41*, 865–876.

Becerra, J.X., Noge, K., and Venable, D.L. (2009). Macroevolutionary chemical escalation in an ancient plant–herbivore arms race. Proc. Natl. Acad. Sci. *106*, 18062–18066.

Beier, R.C., and Oertli, E.H. (1983). Psoralen and other linear furocoumarins as phytoalexins in celery. Phytochemistry *22*, 2595–2597.

Bendtsen, J., Nielsen, H., von Heijne, G., and Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. *340*, 783–795.

Bennett, R.G., Ryan, M.H., Colmer, T.D., and Real, D. (2011). Prioritisation of novel pasture species for use in water-limited agriculture: a case study of Cullen in the Western Australian wheatbelt. Genet. Resour. Crop Evol. *58*, 83–100.

Benveniste, I., Lesot, A., Hasenfratz, M.-P., Kochs, G., and Durst, F. (1991). Multiple forms of NADPH-cytochrome P450 reductase in higher plants. Biochem. Biophys. Res. Commun. *177*, 105–112.

Berenbaum, M. (1983). Coumarins and caterpillars: a case for coevolution. Evolution 163–179.

Berenbaum, M.R. (1990). Evolution of specialization in insect-umbellifer associations. Annu. Rev. Entomol. *35*, 319–343.

Berenbaum, M.R. (2002). Postgenomic chemical ecology: from genetic code to ecological interactions. J. Chem. Ecol. *28*, 873–896.

Berenbaum, M., and Feeny, P. (1981). Toxicity of angular furanocoumarins to swallowtail butterflies: escalation in a coevolutionary arms race? Science *212*, 927–929.

Berenbaum, M.R., Zangerl, A.R., and Nitao, J.K. (1984). Furanocoumarins in seeds of wild and cultivated parsnip. Phytochemistry *23*, 1809–1810.

Berenbaum, M.R., Nitao, J.K., and Zangerl, A.R. (1991). Adaptive significance of furanocoumarin diversity in *Pastinaca sativa* (Apiaceae). J. Chem. Ecol. *17*, 207–215.

Bernhardt, R. (2006). Cytochromes P450 as versatile biocatalysts. J. Biotechnol. *124*, 128–145.

Bertea, C.M., Schalk, M., Karp, F., Maffei, M., and Croteau, R. (2001). Demonstration That Menthofuran Synthase of Mint (*Mentha*) Is a Cytochrome P450 Monooxygenase: Cloning, Functional Expression, and Characterization of the Responsible Gene. Arch. Biochem. Biophys. *390*, 279–286.

Besemer, J., Lomsadze, A., and Borodovsky, M. (2001). GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res. *29*, 2607–2618.

Bilgili, S.G., Akdeniz, N., Akbayram, S., Ceylan, A., Çalka, Ö., and Karaman, K. (2011). Phototoxic dermatitis due to *Chenopodium album* in a child. Pediatr. Dermatol. *28*, 674–676.

Bilici, M., Efe, H., Köroğlu, M.A., Uydu, H.A., Bekaroğlu, M., and Değer, O. (2001). Antioxidative enzyme activities and lipid peroxidation in major depression: alterations by antidepressant treatments. J. Affect. Disord. *64*, 43–51.

Biteau, F. (2009). Production de protéines recombinantes par des plantes carnivores génétiquement transformées: application à *Drosera rotundifolia* et transfert de la technologie à *Nepenthes alata*. Vandoeuvre-les-Nancy, INPL.

Blanco, E., Pignatelli, M., Beltran, S., Punset, A., Pérez-Lluch, S., Serras, F., Guigó, R., and Corominas, M. (2008). Conserved chromosomal clustering of genes governed by chromatin regulators in Drosophila. Genome Biol *9*, R134.

Boardley, M., Stirton, C.H., and Harborne, J.B. (1986). A chemotaxonomic survey of the tribe Psoraleeae in Africa. Biochem. Syst. Ecol. *14*, 603–613.

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.-C., Estreicher, A., Gasteiger, E., Martin, M.J., Michoud, K., O'Donovan, C., and Phan, I. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res. *31*, 365–370.

Bohlmann, J., Gibraltarskaya, E., and Eilert, U. (1995). Elicitor induction of furanocoumarin biosynthetic pathway in cell cultures of *Ruta graveolens*. Plant Cell Tissue Organ Cult. *43*, 155–161.

Bono, J.M., Matzkin, L.M., Castrezana, S., and Markow, T.A. (2008). Molecular evolution and population genetics of two *Drosophila mettleri* cytochrome P450 genes involved in host plant utilization. Mol. Ecol. *17*, 3211–3221.

Boudet, A.-M. (2007). Evolution and current status of research in phenolic compounds. Phytochemistry *68*, 2722–2735.

Bourgaud, F., Brunel, M.C., Guckert, A., and Forlot, P. (1992). Effect of nitrogen nutrition and environmental conditions on the production of pharmaceutically useful metabolites by *Psoralea cinerea* [fucocoumarin, secondary metabolites, chemical stress, psoralen, angelicin]. Eur. J. Agron. Fr.

Bourgaud, F., Bouque, V., and Guckert, A. (1999). Production of flavonoids by Psoralea hairy root cultures. Plant Cell Tissue Organ Cult. *56*, 96–103.

Bourgaud, F., Hehn, A., Larbat, R., Doerper, S., Gontier, E., Kellner, S., and Matern, U. (2006). Biosynthesis of coumarins in plants: a major pathway still to be unravelled for cytochrome P450 enzymes. Phytochem. Rev. *5*, 293–308.

Bredebach, M., Matern, U., and Martens, S. (2011). Three 2-oxoglutarate-dependent dioxygenase activities of *Equisetum arvense* L. forming flavone and flavonol from (*2S*)-naringenin. Phytochemistry *72*, 557–563.

Brown, S.A. (1963). Biosynthesis of the coumarins IV. The formation of coumarin and herniarin in lavender. Phytochemistry *2*, 137–144.

Brown, S.A., and Sampathkumar, S. (1977). The biosynthesis of isopimpinellin. Can. J. Biochem. *55*, 686–692.

Burge, C., and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. J. Mol. Biol. *268*, 78–94.

Cabello-Hurtado, F., Durst, F., Jorrín, J.V., and Werck-Reichhart, D. (1998a). Coumarins in *Helianthus tuberosus* : characterization, induced accumulation and biosynthesis. Phytochemistry *49*, 1029–1036.

Cabello-Hurtado, F., Batard, Y., Salaün, J.-P., Durst, F., Pinot, F., and Werck-Reichhart, D. (1998b). Cloning, expression in yeast, and functional characterization of CYP81B1, a plant cytochrome P450 that catalyzes in-chain hydroxylation of fatty acids. J. Biol. Chem. *273*, 7260–7267.

Cadet, J., Anselmino, C., Douki, T., and Voituriez, L. (1992). New trends in photobiology: Photochemistry of nucleic acids in cells. J. Photochem. Photobiol. B *15*, 277–298.

Caffieri, S. (2002). Furocoumarin photolysis: chemical and biological aspects. Photochem. Photobiol. Sci. *1*, 149–157.

Cannone, J.J., Subramanian, S., Schnare, M.N., Collett, J.R., D'Souza, L.M., Du, Y., Feng, B., Lin, N., Madabusi, L.V., and Müller, K.M. (2002). The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. BMC Bioinformatics *3*, 2.

Cappelletti, E.M., Innocenti, G., and Caporale, G. (1984). Furocoumarins localization in the fruit and seed of *Psoralea corylifolia* L. Pl Med Phytother *18*, 181–190.

Carpinella, M.C., Ferrayoli, C.G., and Palacios, S.M. (2005). Antifungal Synergistic Effect of Scopoletin, a Hydroxycoumarin Isolated from *Melia azedarach* L. Fruits. J. Agric. Food Chem. *53*, 2922–2927.

Casadaban, M.J., and Cohen, S.N. (1980). Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. J. Mol. Biol. *138*, 179–207.

Caspi, R., Foerster, H., Fulcher, C.A., Kaipa, P., Krummenacker, M., Latendresse, M., Paley, S., Rhee, S.Y., Shearer, A.G., and Tissier, C. (2008). The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Res. *36*, D623–D631.

Chalmel, F., Lardenois, A., Thompson, J.D., Muller, J., Sahel, J.-A., Léveillard, T., and Poch, O. (2005). GOAnno: GO annotation based on multiple alignment. Bioinformatics *21*, 2095–2096.

Chang, C.-L., Zhang, L.-J., Chen, R.Y., Kuo, L.-M.Y., Huang, J.-P., Huang, H.-C., Lee, K.-H., Wu, Y.-C., and Kuo, Y.-H. (2010). Antioxidant and Anti-inflammatory Phenylpropanoid Derivatives from *Calamus quiquesetinervius*. J. Nat. Prod. *73*, 1482–1488.

Chapple, C. (1998). Molecular-genetic analysis of plant cytochrome P450-dependent monooxygenases. Annu. Rev. Plant Biol. *49*, 311–343.

Chen, C.-D., and Kemper, B. (1996). Different structural requirements at specific proline residue positions in the conserved proline-rich region of cytochrome P450 2C2. J. Biol. Chem. *271*, 28607–28611.

Chen, J.-J., Wang, T.-Y., and Hwang, T.-L. (2008). Neolignans, a Coumarinolignan, Lignan Derivatives, and a Chromene: Anti-inflammatory Constituents from *Zanthoxylum avicennae*. J. Nat. Prod. *71*, 212–217.

Chen, S., Glawischnig, E., Jørgensen, K., Naur, P., Jørgensen, B., Olsen, C.-E., Hansen, C.H., Rasmussen, H., Pickett, J.A., and Halkier, B.A. (2003). CYP79F1 and CYP79F2 have distinct functions in the biosynthesis of aliphatic glucosinolates in Arabidopsis. Plant J. *33*, 923–937.

Chen, Y., Wang, H.-D., Xia, X., Kung, H.-F., Pan, Y., and Kong, L.-D. (2007). Behavioral and biochemical studies of total furocoumarins from seeds of *Psoralea corylifolia* in the chronic mild stress model of depression in mice. Phytomedicine *14*, 523–529.

Chiesa-Fuxench, Z.C., and González-Chávez, J. (2010). Extracorporeal photopheresis: a review on the immunological aspects and clinical applications. PR Health Sci. J. *29*.

Chong, J., Baltz, R., Schmitt, C., Beffa, R., Fritig, B., and Saindrenan, P. (2002). Downregulation of a Pathogen-Responsive Tobacco UDP-Glc:Phenylpropanoid Glucosyltransferase Reduces Scopoletin Glucoside Accumulation, Enhances Oxidative Stress, and Weakens Virus Resistance. Plant Cell Online *14*, 1093–1107.

Churngchow, N., and Rattarasarn, M. (2001). Biosynthesis of scopoletin in *Hevea brasiliensis* leaves inoculated with *Phytophthora palmivora*. J. Plant Physiol. *158*, 875–882.

Cianfrogna, J.A., Zangerl, A.R., and Berenbaum, M.R. (2002a). Effects of furanocoumarins on feeding behavior of parsnip webworms *Depressaria pastinacella*. J. Chem. Ecol. *28*, 1365–1375.

Cianfrogna, J.A., Zangerl, A.R., and Berenbaum, M.R. (2002b). Dietary and developmental influences on induced detoxification in an oligophage. J. Chem. Ecol. *28*, 1349–1364.

Clark, G.S. (1995). Coumarin. An aroma chemical profile. Perfum Flavor *20*, 23–34.

Claudel-Renard, C., Chevalet, C., Faraut, T., and Kahn, D. (2003). Enzyme-specific profiles for genome annotation: PRIAM. Nucleic Acids Res. *31*, 6633–6639.

Cokol, M., Nair, R., and Rost, B. (2000). Finding nuclear localization signals. EMBO Rep. *1*, 411–415.

Collu, G., Unver, N., Peltenburg-Looman, A.M.G., van der Heijden, R., Verpoorte, R., and Memelink, J. (2001). Geraniol 10-hydroxylase, a cytochrome P450 enzyme involved in terpenoid indole alkaloid biosynthesis. FEBS Lett. *508*, 215–220.

Comino, C., Lanteri, S., Portis, E., Acquadro, A., Romani, A., Hehn, A., Larbat, R., and Bourgaud, F. (2007). Isolation and functional characterization of a cDNA coding a hydroxycinnamoyltransferase involved in phenylpropanoid biosynthesis in *Cynara cardunculus* L. BMC Plant Biol. *7*, 14.

Comino, C., Hehn, A., Moglia, A., Menin, B., Bourgaud, F., Lanteri, S., and Portis, E. (2009). The isolation and mapping of a novel hydroxycinnamoyltransferase in the globe artichoke chlorogenic acid pathway. BMC Plant Biol. *9*, 30.

Conesa, A., and Götz, S. (2008). Blast2GO: A comprehensive suite for functional analysis in plant genomics. Int. J. Plant Genomics *2008*.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics *21*, 3674–3676.

Costa, M.A., Bedgar, D.L., Moinuddin, S.G.A., Kim, K.-W., Cardenas, C.L., Cochrane, F.C., Shockey, J.M., Helms, G.L., Amakura, Y., Takahashi, H., *et al*. (2005). Characterization *in vitro* and *in vivo* of the putative multigene 4-coumarate:CoA ligase network in Arabidopsis: syringyl lignin and sinapate/sinapyl alcohol derivative formation. Phytochemistry *66*, 2072–2091.

Costet, L., Fritig, B., and Kauffmann, S. (2002). Scopoletin expression in elicitor-treated and tobacco mosaic virus-infected tobacco plants. Physiol. Plant. *115*, 228–235.

Court, W.A. (1977). High-performance reversed-phase liquid chromatography of naturally occurring phenolic compounds. J. Chromatogr. A *130*, 287–291.

Cowan, M.M. (1999). Plant Products as Antimicrobial Agents. Clin. Microbiol. Rev. *12*, 564–582.

Cravotto, G., Nano, G.M., Palmisano, G., and Tagliapietra, S. (2001). An asymmetric approach to coumarin anticoagulants via hetero-Diels–Alder cycloaddition. Tetrahedron Asymmetry *12*, 707–709.

Curini, M., Cravotto, G., Epifano, F., and Giannone, G. (2006). Chemistry and biological activity of natural and synthetic prenyloxycoumarins. Curr. Med. Chem. *13*, 199–222.

Daglia, M. (2012). Polyphenols as antimicrobial agents. Curr. Opin. Biotechnol. *23*, 174–181.

Dahan, A., and Altman, H. (2004). Food–drug interaction: grapefruit juice augments drug bioavailability—mechanism, extent and relevance. Eur. J. Clin. Nutr. *58*, 1–9.

Dall'Acqua, F., and Martelli, P. (1991). Photosensitizing action of furocoumarins on membrane components and consequent intracellular events. J. Photochem. Photobiol. B *8*, 235–254.

Danno, K. (1999). PUVA therapy: current concerns in Japan. J. Dermatol. Sci. *19*, 89–105.

Dardalhon, M., de Massy, B., Nicolas, A., and Averbeck, D. (1998). Mitotic recombination and localized DNA double-strand breaks are induced after 8-methoxypsoralen and UVA irradiation in *Saccharomyces cerevisiae*. Curr. Genet. *34*, 30–42.

David, J.P., Boyer, S., Mesneau, A., Ball, A., Ranson, H., and Dauphin-Villemant, C. (2006). Involvement of cytochrome P450 monooxygenases in the response of mosquito larvae to dietary plant xenobiotics. Insect Biochem. Mol. Biol. *36*, 410–420.

De Castro, E.M., Pinto, J., Bertolucci, S.K., Malta, M.R., Cardoso, M. das G., and de MSilva, F.A. (2007). Coumarin contents in young *Mikania glomerata* plants (guaco) under different radiation levels and photoperiod. Acta Farm. Bonaer. *25*, 387.

De Feo, V., and Senatore, F. (1993). Medicinal plants and phytotherapy in the Amalfitan coast, Salerno province, Campania, Southern Italy. J. Ethnopharmacol. *39*, 39–51.

De Graaf, C., Vermeulen, N.P., and Feenstra, K.A. (2005). Cytochrome P450 *in silico*: an integrative modeling approach. J. Med. Chem. *48*, 2725–2755.

Delcher, A.L., Bratke, K.A., Powers, E.C., and Salzberg, S.L. (2007). Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics *23*, 673–679.

De Montellano, P.R.O. (2005). Cytochrome P450: structure, mechanism, and biochemistry (Springer).

Després, L., David, J.-P., and Gallet, C. (2007). The evolutionary ecology of insect resistance to plant chemicals. Trends Ecol. Evol. *22*, 298–307.

Dhillon, D.S., and Brown, S.A. (1976). Localization, purification, and characterization of dimethylallylpyrophosphate: Umbelliferone dimethylallyltransferase from *Ruta graveolens*. Arch. Biochem. Biophys. *177*, 74–83.

Didierjean, L., Gondet, L., Perkins, R., Lau, S.-M.C., Schaller, H., O'Keefe, D.P., and Werck-Reichhart, D. (2002). Engineering Herbicide Metabolism in Tobacco and Arabidopsis with CYP76B1, a Cytochrome P450 Enzyme from Jerusalem Artichoke. Plant Physiol. *130*, 179–189.

Diesperger, H., Müller, C.R., and Sandermann Jr, H. (1974). Rapid isolation of a plant microsomal fraction by Mg^{2+} precipitation. FEBS Lett. *43*, 155–158.

Diwan, R., and Malpathak, N. (2009). Furanocoumarins: Novel topoisomerase I inhibitors from *Ruta graveolens* L. Bioorg. Med. Chem. *17*, 7052–7055.

Diwan, R., and Malpathak, N. (2010). Histochemical localization in *Ruta graveolens* cell cultures: elucidating the relationship between cellular differentiation and furanocoumarin production. Vitro Cell. Dev. Biol.-Plant *46*, 108–116.

Dixon, R., and Paiva, N. (1995). Stress-Induced Phenylpropanoid Metabolism. Plant Cell *7*, 1085–1097.

Domínguez, M.A., Membrillo, V.G., and Ramos, G.A. (2004). Parapsoriasis en grandes placas tratada con PUVA. Reporte de un caso. Rev Cent Dermatol Pascua *13*, 143–146.

Downie, S.R., Plunkett, G.M., Watson, M.F., Spalik, K., Terentieva, E., Troitsky, A., Lee, B., Lahham, J., and others (2001). Tribes and clades within Apiaceae subfamily Apioideae: the contribution of molecular data. Edinb. J. Bot. *58*, 301–330.

Dugrand, A., Olry, A., Duval, T., Hehn, A., Froelicher, Y., and Bourgaud, F. (2013). Coumarin and Furanocoumarin Quantitation in Citrus Peel via Ultraperformance Liquid Chromatography Coupled with Mass Spectrometry (UPLC-MS). J. Agric. Food Chem. *61*, 10677–10684.

Eckey-Kaltenbach, H., Ernst, D., Heller, W., and Sandermann Jr, H. (1994). Biochemical Plant Responses to Ozone (IV. Cross-Induction of Defensive Pathways in Parsley (*Petroselinum crispum* L.) Plants). Plant Physiol. *104*, 67–74.

Edelson, R., Berger, C., Gasparro, F., Jegasothy, B., Heald, P., Wintroub, B., Vonderheid, E., Knobler, R., Wolff, K., and Plewig, G. (1987). Treatment of cutaneous T-cell lymphoma by extracorporeal photochemotherapy. N. Engl. J. Med. *316*, 297–303.

Egan, D., O'kennedy, R., Moran, E., Cox, D., Prosser, E., and Thornes, R.D. (1990). The Pharmacology, Metabolism, Analysis, and Applications of Coumarin and Coumarin-Related Compounds. Drug Metab. Rev. *22*, 503–529.

Ehlers, D., Pfister, M., Bork, W.-R., and Toffel-Nadolny, P. (1995). HPLC analysis of tonka bean extracts. Z. Für Lebensm.-Unters. Forsch. *201*, 278–282.

Ehrlich, P.R., and Raven, P.H. (1964). Butterflies and plants: a study in coevolution. Evolution 586–608.

Ekiert, H., and Gomólka, E. (1999). Effect of light on contents of coumarin compounds in shoots of *Ruta graveolens* L. cultivated *in vitro*. Acta Soc. Bot. Pol. *68*, 197–200.

Ekiert, H., and Kisiel, W. (2000). Isolation of furanocoumarins from *Pastinaca sativa* L. callus culture. Acta Soc. Bot. Pol. *69*, 193–195.

Eljounaidi, K., Cankar, K., Comino, C., Moglia, A., Hehn, A., Bourgaud, F., Bouwmeester, H., Menin, B., Lanteri, S., and Beekwilder, J. (2014). Cytochrome P450s from *Cynara cardunculus* L. CYP71AV9 and CYP71BL5, catalyze distinct hydroxylations in the sesquiterpene lactone biosynthetic pathway. Plant Sci. *223*, 59–68.

Ellis, B.E., and Brown, S.A. (1974). Isolation of dimethylallylpyrophosphate: umbelliferone dimethylallyltransferase from *Ruta graveolens.* Can. J. Biochem. *52*, 734–738.

Endler, A., Martens, S., Wellmann, F., and Matern, U. (2008). Unusually divergent 4-coumarate:CoA-ligases from *Ruta graveolens* L. Plant Mol. Biol. *67*, 335–346.

Epifano, F., Genovese, S., Menghini, L., and Curini, M. (2007). Chemistry and pharmacology of oxyprenylated secondary plant metabolites. Phytochemistry *68*, 939–953.

Evans, P.H., and Becerra, J.X. (2006). Non-terpenoid essential oils from *Bursera chemapodicta*. Flavour Fragr. J. *21*, 616–618.

Evans, P.H., Becerra, J.X., Venable, D.L., and Bowers, W.S. (2000). Chemical analysis of squirt-gun defense in Bursera and counterdefense by chrysomelid beetles. J. Chem. Ecol. *26*, 745–754.

Facchini, P.J. (2001). Alkaloid biosynthesis in plants: biochemistry, cell biology, molecular regulation, and metabolic engineering applications. Annu. Rev. Plant Biol. *52*, 29–66.

Fahmy, I.R., and Abu-Shady, H. (1946). *Ammi majus* Linn.; pharmacognostical study and isolation of a crystalline constituent, ammoidin. Q. J. Pharm. Pharmacol. *20*, 281–291.

Fahrendorf, T., and Dixon, R.A. (1993). Stress Responses in Alfalfa (*Medicago sativa* L.) XVIII: Molecular Cloning and Expression of the Elicitor-Inducible Cinnamic Acid 4-Hydroxylase Cytochrome P450. Arch. Biochem. Biophys. *305*, 509–515.

Ferrazzano, G.F., Amato, I., Ingenito, A., Zarrelli, A., Pinto, G., and Pollio, A. (2011). Plant Polyphenols and Their Anti-Cariogenic Properties: A Review. Molecules *16*, 1486–1507.

Feskanich, D., Willett, W.C., Hunter, D.J., and Colditz, G.A. (2003). Dietary intakes of vitamins A, C, and E and risk of melanoma in two cohorts of women. Br. J. Cancer *88*, 1381–1387.

Finn, R.D., Tate, J., Mistry, J., Coggill, P.C., Sammut, S.J., Hotz, H.-R., Ceric, G., Forslund, K., Eddy, S.R., and Sonnhammer, E.L. (2008). The Pfam protein families database. Nucleic Acids Res. *36*, D281–D288.

Fliegmann, J., Furtwängler, K., Malterer, G., Cantarello, C., Schüler, G., Ebel, J., and Mithöfer, A. (2010). Flavone synthase II (CYP93B16) from soybean (*Glycine max* L.). Phytochemistry *71*, 508–514.

Forslund, K., Morant, M., Jørgensen, B., Olsen, C.E., Asamizu, E., Sato, S., Tabata, S., and Bak, S. (2004). Biosynthesis of the nitrile glucosides rhodiocyanoside A and D and the cyanogenic glucosides lotaustralin and linamarin in *Lotus japonicus*. Plant Physiol. *135*, 71–84.

Fouin-Fortunet, H., Tinel, M., Descatoire, V., Letteron, P., Larrey, D., Geneve, J., and Pessayre, D. (1986). Inactivation of cytochrome P-450 by the drug methoxsalen. J. Pharmacol. Exp. Ther. *236*, 237–247.

Fraenkel, G.S. (1959). The Raison d'Être of Secondary Plant Substances These odd chemicals arose as a means of protecting plants from insects and now guide insects to food. Science *129*, 1466–1470.

Frank, S., Caffieri, S., Raffaelli, A., Vedaldi, D., and Dall'Acqua, F. (1998). Characterization of psoralen-oleic acid cycloadducts and their possible involvement in membrane photodamage. J. Photochem. Photobiol. B *44*, 39–44.

Frey, M., Chomet, P., Glawischnig, E., Stettner, C., Grün, S., Winklmair, A., Eisenreich, W., Bacher, A., Meeley, R.B., and Briggs, S.P. (1997). Analysis of a chemical plant defense mechanism in grasses. Science *277*, 696–699.

Frey, M., Huber, K., Park, W.J., Sicker, D., Lindberg, P., Meeley, R.B., Simmons, C.R., Yalpani, N., and Gierl, A. (2003). A 2-oxoglutarate-dependent dioxygenase is integrated in DIMBOA-biosynthesis. Phytochemistry *62*, 371–376.

Frey, M., Schullehner, K., Dick, R., Fiesselmann, A., and Gierl, A. (2009). Benzoxazinoid biosynthesis, a model for evolution of secondary metabolic pathways in plants. Phytochemistry *70*, 1645–1651.

Friedman, J., Rushkin, E., and Waller, G.R. (1982). Highly potent germination inhibitors in aqueous eluate of fruits of bishop's weed (*Ammi majus* L.) and avoidance of autoinhibition. J. Chem. Ecol. *8*, 55–65.

Fritig, B., Hirth, L., and Ourisson, G. (1970). Biosynthesis of the coumarins: Scopoletin formation in tobacco tissue cultures. Phytochemistry *9*, 1963–1975.

Fukunishi, Y., and Hayashizaki, Y. (2001). Amino acid translation program for full-length cDNA sequences with frameshift errors. Physiol. Genomics *5*, 81–87.

Gabriac, B., Werck-Reichhart, D., Teutsch, H., and Durst, F. (1991). Purification and immunocharacterization of a plant cytochrome P450: The cinnamic acid 4-hydroxylase. Arch. Biochem. Biophys. *288*, 302–309.

Garcia-Molina, M. of the S., Munoz-Munoz, J.L., Garcia-Molina, F., Rodriguez-Lopez, J.N., and Garcia-Canovas, F. (2013). Study of Umbelliferone Hydroxylation to Esculetin Catalyzed by Polyphenol Oxidase. Biol. Pharm. Bull. *36*, 1140–1145.

Gardner, P.P., Daub, J., Tate, J.G., Nawrocki, E.P., Kolbe, D.L., Lindgreen, S., Wilkinson, A.C., Finn, R.D., Griffiths-Jones, S., and Eddy, S.R. (2009). Rfam: updates to the RNA families database. Nucleic Acids Res. *37*, D136–D140.

Gardy, J.L., Laird, M.R., Chen, F., Rey, S., Walsh, C.J., Ester, M., and Brinkman, F.S. (2005). PSORTb v. 2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. Bioinformatics *21*, 617–623.

Gattuso, G., Barreca, D., Caristi, C., Gargiulli, C., and Leuzzi, U. (2007). Distribution of Flavonoids and Furocoumarins in Juices from Cultivars of *Citrus bergamia* Risso. J. Agric. Food Chem. *55*, 9921–9927.

Giacomo Calzavara-Pinton, P., Zane, C., Carlino, A., and De Panfilis, G. (1997). Bath-5-methoxypsoralen-UVA therapy for psoriasis. J. Am. Acad. Dermatol. *36*, 945–949.

Giddings, L.-A., Liscombe, D.K., Hamilton, J.P., Childs, K.L., DellaPenna, D., Buell, C.R., and O'Connor, S.E. (2011). A stereoselective hydroxylation step of alkaloid biosynthesis by a unique cytochrome P450 in *Catharanthus roseus.* J. Biol. Chem. *286*, 16751–16757.

Gierl, A., and Frey, M. (2001). Evolution of benzoxazinone biosynthesis and indole production in maize. Planta *213*, 493–498.

Gion, K., Inui, H., Takakuma, K., Yamada, T., Kambara, Y., Nakai, S., Fujiwara, H., Miyamura, T., Imaishi, H., and Ohkawa, H. (2014). Molecular mechanisms of herbicide-inducible gene expression of tobacco CYP71AH11 metabolizing the herbicide chlorotoluron. Pestic. Biochem. Physiol.

Girennavar, B., Jayaprakasha, G.K., and Patil, B.S. (2007). Potent inhibition of human cytochrome P450 3A4, 2D6, and 2C9 isoenzymes by grapefruit juice and its furocoumarins. J. Food Sci. *72*, C417–C421.

Girennavar, B., Cepeda, M.L., Soni, K.A., Vikram, A., Jesudhasan, P., Jayaprakasha, G.K., Pillai, S.D., and Patil, B.S. (2008). Grapefruit juice and its furocoumarins inhibits autoinducer signaling and biofilm formation in bacteria. Int. J. Food Microbiol. *125*, 204–208.

Glawischnig, E., Grün, S., Frey, M., and Gierl, A. (1999). Cytochrome P450 monooxygenases of DIBOA biosynthesis: specificity and conservation among grasses. Phytochemistry *50*, 925–930.

Gliszczyńska, A., and Brodelius, P.E. (2012). Sesquiterpene coumarins. Phytochem. Rev. *11*, 77–96.

Godar, D.E. (1999). UVA1 radiation triggers two different final apoptotic pathways. J. Invest. Dermatol. *112*, 3–12.

Gontier, E., Fliniaux, M. A., Barbotin, J. N., and Sangwan-Norreel, B. S. (1993). Tropane alkaloid levels in the leaves of micropropagated *Datura innoxia* Mill. plants. Planta Medica *59*, 432-435.

Gordon, P.R., and Gilchrest, B.A. (1989). Human melanogenesis is stimulated by diacylglycerol. J. Invest. Dermatol. *93*.

Gorz, H.J., and Haskins, F.A. (1962). Translocation of Coumarin Across a Graft Union m Sweetclover. Agron. Publ. 243.

Gotoh, O. (1992). Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. J. Biol. Chem. *267*, 83–90.

Götz, S., García-Gómez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talón, M., Dopazo, J., and Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. *36*, 3420–3435.

Gouzy, J., Carrere, S., and Schiex, T. (2009). FrameDP: sensitive peptide detection on noisy matured sequences. Bioinformatics *25*, 670–671.

Gravot, A. (2002). Etude de P450s impliques dans la biosynthese des furocoumarines chez *Ruta graveolens.*

Gravot, A., Larbat, R., Hehn, A., Lièvre, K., Gontier, E., Goergen, J.-L., and Bourgaud, F. (2004). Cinnamic acid 4-hydroxylase mechanism-based inactivation by psoralen derivatives: cloning and characterization of a C4H from a psoralen producing plant—*Ruta graveolens*—exhibiting low sensitivity to psoralen inactivation. Arch. Biochem. Biophys. *422*, 71–80.

Gray, A.I., and Waterman, P.G. (1978). Coumarins in the Rutaceae. Phytochemistry *17*, 845–864.

Grimes, J.W. (1997). A revision of Cullen (Leguminosae: Papilionoideae). Aust. Syst. Bot. *10*, 565–648.

Gross, S.S., Do, C.B., Sirota, M., and Batzoglou, S. (2007). CONTRAST: a discriminative, phylogeny-free approach to multiple informant *de novo* gene prediction. Genome Biol. *8*, R269.

Grotewold, E., Drummond, B.J., Bowen, B., and Peterson, T. (1994). The myb-homologous P gene controls phlobaphene pigmentation in maize floral organs by directly activating a flavonoid biosynthetic gene subset. Cell *76*, 543–553.

Guarrera, P.M. (1999). Traditional antihelmintic, antiparasitic and repellent uses of plants in Central Italy. J. Ethnopharmacol. *68*, 183–192.

Guarrera, P.M. (2005). Traditional phytotherapy in Central Italy (Marche, Abruzzo, and Latium). Fitoterapia *76*, 1–25.

Guengerich, F.P. (2007). Mechanisms of cytochrome P450 substrate oxidation: MiniReview. J. Biochem. Mol. Toxicol. *21*, 163–168.

Guirimand, G., Guihur, A., Ginis, O., Poutrain, P., Héricourt, F., Oudin, A., Lanoue, A., St-Pierre, B., Burlat, V., and Courdavault, V. (2011). The subcellular organization of strictosidine biosynthesis in *Catharanthus roseus* epidermis highlights several trans-tonoplast translocations of intermediate metabolites. FEBS J. *278*, 749–763.

Guo, F., Lei, J., Sun, Y., Chi, Y.H., Ge, F., Patil, B.S., Koiwa, H., Zeng, R., and Zhu-Salzman, K. (2012). Antagonistic regulation, yet synergistic defense: effect of bergapten and protease inhibitor on development of cowpea bruchid *Callosobruchus maculatus*. PloS One *7*, e41877.

Guo, L.-Q., Fukuda, K., Ohta, T., and Yamazoe, Y. (2000). Role of furanocoumarin derivatives on grapefruit juice-mediated inhibition of human CYP3A activity. Drug Metab. Dispos. *28*, 766–771.

Gustafsson, C., Govindarajan, S., and Minshull, J. (2004). Codon bias and heterologous protein expression. Trends Biotechnol. *22*, 346–353.

Gutiérrez-Mellado, M.-C., Edwards, R., Tena, M., Cabello, F., Serghini, K., and Jorrín, J. (1996). The production of coumarin phytoalexins in different plant organs of sunflower (*Helianthus annuus* L.). J. Plant Physiol. *149*, 261–266.

Guyot, A.D., Farhi, D., Ingen-Housz-Oro, S., Bussel, A., Parquet, N., Rabian, C., Bachelez, H., and Francès, C. (2007). Treatment of refractory erosive oral lichen planus with extracorporeal photochemotherapy: 12 cases. Br. J. Dermatol. *156*, 553–556.

Hadaček, F., Müller, C., Werner, A., Greger, H., and Proksch, P. (1994). Analysis, isolation and insecticidal activity of linear furanocoumarins and other coumarin derivatives from Peucedanum (Apiaceae: Apioideae). J. Chem. Ecol. *20*, 2035–2054.

Hale, A.L., Meepagala, K.M., Oliva, A., Aliotta, G., and Duke, S.O. (2004). Phytotoxins from the leaves of *Ruta graveolens*. J. Agric. Food Chem. *52*, 3345–3349.

Halkier, B.A., and Gershenzon, J. (2006). Biology and biochemistry of glucosinolates. Annu Rev Plant Biol *57*, 303–333.

Hamberger, B., and Bak, S. (2013). Plant P450s as versatile drivers for evolution of specific-specific chemical diversity.

Hamberger, B., Ellis, M., Friedmann, M., de Azevedo Souza, C., Barbazuk, B., and Douglas, C.J. (2007). Genome-wide analyses of phenylpropanoid-related genes in *Populus trichocarpa*, *Arabidopsis thaliana*, and *Oryza sativa*: the Populus lignin toolbox and conservation and diversification of angiosperm gene familiesThis article is one of a selection of papers published in the Special Issue on Poplar Research in Canada. Can. J. Bot. *85*, 1182–1201.

Hamerski, D., and Matern, U. (1988a). Biosynthesis of psoralens Psoralen 5-monooxygenase activity from elicitor-treated *Ammi majus* cells. FEBS Lett. *239*, 263–265.

Hamerski, D., and Matern, U. (1988b). Elicitor-induced biosynthesis of psoralens in *Ammi majus* L. suspension cultures. Eur. J. Biochem. *171*, 369–375.

Hamerski, D., Beier, R.C., Kneusel, R.E., Matern, U., and Himmelspacht, K. (1990). Accumulation of coumarins in elicitor-treated cell suspension cultures of *Ammi majus*. Phytochemistry *29*, 1137–1142.

Hatzigeorgiou, A.G., Fiziev, P., and Reczko, M. (2001). DIANA-EST: a statistical analysis. Bioinformatics *17*, 913–919.

Hehmann, M., Lukačin, R., Ekiert, H., and Matern, U. (2004). Furanocoumarin biosynthesis in *Ammi majus* L. Eur. J. Biochem. *271*, 932–940.

Heinke, R., Franke, K., Michels, K., Wessjohann, L., Ali, N.A.A., and Schmidt, J. (2012). Analysis of furanocoumarins from Yemenite Dorstenia species by liquid chromatography/electrospray tandem mass spectrometry. J. Mass Spectrom. *47*, 7–22.

Heinrich, U., Moore, C.E., Spirt, S.D., Tronnier, H., and Stahl, W. (2011). Green Tea Polyphenols Provide Photoprotection, Increase Microcirculation, and Modulate Skin Properties of Women. J. Nutr. *141*, 1202–1208.

Helliwell, C.A., Chandler, P.M., Poole, A., Dennis, E.S., and Peacock, W.J. (2001). The CYP88A cytochrome P450, ent-kaurenoic acid oxidase, catalyzes three steps of the gibberellin biosynthesis pathway. Proc. Natl. Acad. Sci. *98*, 2065–2070.

Helmkamp, G.M., Rando, R.R., Brock, D.J.H., and Bloch, K. (1968). β-Hydroxydecanoyl thioester dehydrase specificity of substrates and acetylenic inhibitors. J. Biol. Chem. *243*, 3229–3231.

Henegouwen, V., Beijersbergen, G.M., Wijn, E.T., Schoonderwoerd, S.A., and Dall'Acqua, F. (1989). A method for the determination of PUVA-induced *in vivo* irreversible binding of 8-methoxypsoralen (8-MOP) to epidermal lipids, proteins and DNA/RNA. J. Photochem. Photobiol. B *3*, 631–635.

Hivelin, M., Siemionow, M., Grimbert, P., and Lantieri, L. (2009). Extracorporeal photopheresis: from solid organs to face transplantation. Transpl. Immunol. *21*, 117–128.

Höfer, R., Briesen, I., Beck, M., Pinot, F., Schreiber, L., and Franke, R. (2008). The Arabidopsis cytochrome P450 CYP86A1 encodes a fatty acid ω-hydroxylase involved in suberin monomer biosynthesis. J. Exp. Bot. *59*, 2347–2360.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2008). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

Hübner, S., Hehmann, M., Schreiner, S., Martens, S., Lukačin, R., and Matern, U. (2003). Functional expression of cinnamate 4-hydroxylase from *Ammi majus* L. Phytochemistry *64*, 445–452.

Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., Cuche, B.A., De Castro, E., Lachaize, C., Langendijk-Genevaux, P.S., and Sigrist, C.J. (2008). The 20 years of PROSITE. Nucleic Acids Res. *36*, D245–D249.

Humphreys, J.M., Hemm, M.R., and Chapple, C. (1999). New routes for lignin biosynthesis defined by biochemical characterization of recombinant ferulate 5-hydroxylase, a multifunctional cytochrome P450-dependent monooxygenase. Proc. Natl. Acad. Sci. *96*, 10045–10050.

Hung, C.-F., Prapaipong, H., Berenbaum, M.R., and Schuler, M.A. (1995). Differential induction of cytochrome P450 transcripts in *Papilio polyxenes* by linear and angular furanocoumarins. Insect Biochem. Mol. Biol. *25*, 89–99.

Hunter, S., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U., Daugherty, L., and Duquenne, L. (2009). InterPro: the integrative protein signature database. Nucleic Acids Res. *37*, D211–D215.

Irmler, S., Schröder, G., St-Pierre, B., Crouch, N.P., Hotze, M., Schmidt, J., Strack, D., Matern, U., and Schröder, J. (2000). Indole alkaloid biosynthesis in *Catharanthus roseus*: new enzyme activities and identification of cytochrome P450 CYP72A1 as secologanin synthase. Plant J. *24*, 797–804.

Jagasia, M.H., Savani, B.N., Stricklin, G., Engelhardt, B., Kassim, A., Dixon, S., Chen, H., Chinratanalab, W., Goodman, S., and Greer, J.P. (2009). Classic and overlap chronic graft-versus-host disease (cGVHD) is associated with superior outcome after extracorporeal photopheresis (ECP). Biol. Blood Marrow Transplant. *15*, 1288–1295.

Janzen, D.H., and Muller, C.H. (1969). Coevolution. Science *165*, 415–416.

Joos, H.-J., and Hahlbrock, K. (1992). Phenylalanine ammonia-lyase in potato (*Solanum tuberosum* L.). Eur. J. Biochem. *204*, 621–629.

Jørgensen, K., Rasmussen, A.V., Morant, M., Nielsen, A.H., Bjarnholt, N., Zagrobelny, M., Bak, S., and Møller, B.L. (2005). Metabolon formation and metabolic channeling in the biosynthesis of plant natural products. Curr. Opin. Plant Biol. *8*, 280–291.

Kai, K., Shimizu, B., Mizutani, M., Watanabe, K., and Sakata, K. (2006). Accumulation of coumarins in *Arabidopsis thaliana*. Phytochemistry *67*, 379–386.

Kai, K., Mizutani, M., Kawamura, N., Yamamoto, R., Tamai, M., Yamaguchi, H., Sakata, K., and Shimizu, B. (2008). Scopoletin is biosynthesized via ortho-hydroxylation of feruloyl CoA by a 2-oxoglutarate-dependent dioxygenase in *Arabidopsis thaliana*. Plant J. *55*, 989–999.

Kandel, S., Sauveplane, V., Compagnon, V., Franke, R., Millet, Y., Schreiber, L., Werck-Reichhart, D., and Pinot, F. (2007). Characterization of a methyl jasmonate and wounding-responsive cytochrome P450 of *Arabidopsis thaliana* catalyzing dicarboxylic fatty acid formation in vitro. FEBS J. *274*, 5116–5127.

Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., Katayama, T., Kawashima, S., Okuda, S., and Tokimatsu, T. (2008). KEGG for linking genomes to life and the environment. Nucleic Acids Res. *36*, D480–D484.

Kang, H.K., Shin, E.J., and Shim, S.C. (1992). Transient absorption spectra and quenching of coumarin excited states by nucleic acid bases. J. Photochem. Photobiol. B *13*, 19–28.

Kao, Y.-Y., Harding, S.A., and Tsai, C.-J. (2002). Differential expression of two distinct phenylalanine ammonia-lyase genes in condensed tannin-accumulating and lignifying cells of quaking aspen. Plant Physiol. *130*, 796–807.

Karamat, F., Olry, A., Munakata, R., Koeduka, T., Sugiyama, A., Paris, C., Hehn, A., Bourgaud, F., and Yazaki, K. (2014). A coumarin-specific prenyltransferase catalyzes the crucial biosynthetic reaction for furanocoumarin formation in parsley. Plant J.

Kellner, S. (2008). Molekulare und biochemische Aspekte der Furanocumarinbiosynthese in *Ammi majus* L. Philipps-Universität Marburg.

Kemper, B. (2004). Structural basis for the role in protein folding of conserved proline-rich regions in cytochromes P450. Toxicol. Appl. Pharmacol. *199*, 305–315.

Kent, U.M., Lin, H., Noon, K.R., Harris, D.L., and Hollenberg, P.F. (2006a). Metabolism of bergamottin by cytochromes P450 2B6 and 3A5. J. Pharmacol. Exp. Ther. *318*, 992–1005.

Kent, U.M., Lin, H., Mills, D.E., Regal, K.A., and Hollenberg, P.F. (2006b). Identification of 17-α-ethynylestradiol-modified active site peptides and glutathione conjugates formed during metabolism and inactivation of P450s 2B1 and 2B6. Chem. Res. Toxicol. *19*, 279–287.

Kevekordes, S., Mersch-Sundermann, V., Burghaus, C.M., Spielberger, J., Schmeiser, H.H., Arlt, V.M., and Dunkelberg, H. (1999). SOS induction of selected naturally occurring substances in *Escherichia coli* (SOS chromotest). Mutat. Res. Toxicol. Environ. Mutagen. *445*, 81–91.

Kharasch, E.D., Hankins, D.C., and Taraday, J.K. (2000). Single-dose methoxsalen effects on human cytochrome P-450 2A6 activity. Drug Metab. Dispos. *28*, 28–33.

Kim, T.-W., Hwang, J.-Y., Kim, Y.-S., Joo, S.-H., Chang, S.C., Lee, J.S., Takatsuto, S., and Kim, S.-K. (2005). Arabidopsis CYP85A2, a Cytochrome P450, Mediates the Baeyer-Villiger Oxidation of Castasterone to Brassinolide in Brassinosteroid Biosynthesis. Plant Cell Online *17*, 2397–2412.

Kim, Y.-K., Kim, Y.S., and Ryu, S.Y. (2007). Antiproliferative effect of furanocoumarins from the root of *Angelica dahurica* on cultured human tumor cell lines. Phytother. Res. *21*, 288–290.

Kitamura, N., Kohtani, S., and Nakagaki, R. (2005). Molecular aspects of furocoumarin reactions: Photophysics, photochemistry, photobiology, and structural analysis. J. Photochem. Photobiol. C Photochem. Rev. *6*, 168–185.

Kitamura, Y., Ikenaga, T., Ooe, Y., Hiraoka, N., and Mizukami, H. (1998). Induction of furanocoumarin biosynthesis in *Glehnia littoralis* cell suspension cultures by elicitor treatment. Phytochemistry *48*, 113–117.

Klingenberg, M. (1958). Pigments of rat liver microsomes. Arch. Biochem. Biophys. *75*, 376–386.

Koenigs, L.L., Peter, R.M., Thompson, S.J., Rettie, A.E., and Trager, W.F. (1997). Mechanism-based inactivation of human liver cytochrome P450 2A6 by 8-methoxypsoralen. Drug Metab. Dispos. *25*, 1407–1415.

Koes, R., Verweij, W., and Quattrocchio, F. (2005). Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. Trends Plant Sci. *10*, 236–242.

Korkina, L.G. (2007). Phenylpropanoids as naturally occurring antioxidants: from plant defense to human health. Cell Mol Biol *53*, 15–25.

Krogh, A., Larsson, B., Von Heijne, G., and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. *305*, 567–580.

Krueger, J.G. (2002). The immunologic basis for the treatment of psoriasis with new biologic agents. J. Am. Acad. Dermatol. *46*, 1–26.

Kruse, T., Ho, K., Yoo, H.-D., Johnson, T., Hippely, M., Park, J.-H., Flavell, R., and Bobzin, S. (2008). *In planta* biocatalysis screen of P450s identifies 8-methoxypsoralen as a substrate for the CYP82C subfamily, yielding original chemical structures. Chem. Biol. *15*, 149–156.

Kumar, A., and Ellis, B.E. (2001). The Phenylalanine Ammonia-Lyase Gene Family in Raspberry. Structure, Expression, and Evolution. Plant Physiol. *127*, 230–239.

Kung, A.C., Stephens, M.B., and Darling, T. (2009). Phytophotodermatitis: bulla formation and hyperpigmentation during spring break. Mil. Med. *174*, 657–661.

La Camera, S., Gouzerh, G., Dhondt, S., Hoffmann, L., Fritig, B., Legrand, M., and Heitz, T. (2004). Metabolic reprogramming in plant innate immunity: the contributions of phenylpropanoid and oxylipin pathways. Immunol. Rev. *198*, 267–284.

Lagesen, K., Hallin, P., Rødland, E.A., St\a erfeldt, H.-H., Rognes, T., and Ussery, D.W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. *35*, 3100–3108.

Lampe, J.W., King, I.B., Li, S., Grate, M.T., Barale, K.V., Chen, C., Feng, Z., and Potter, J.D. (2000). Brassica vegetables increase and apiaceous vegetables decrease cytochrome P450 1A2 activity in humans: changes in caffeine metabolite ratios in response to controlled vegetable diets. Carcinogenesis *21*, 1157–1162.

Larbat, R., Kellner, S., Specker, S., Hehn, A., Gontier, E., Hans, J., Bourgaud, F., and Matern, U. (2007). Molecular Cloning and Functional Characterization of Psoralen Synthase, the First Committed Monooxygenase of Furanocoumarin Biosynthesis. J. Biol. Chem. *282*, 542–554.

Larbat, R., Hehn, A., Hans, J., Schneider, S., Jugdé, H., Schneider, B., Matern, U., and Bourgaud, F. (2009). Isolation and Functional Characterization of CYP71AJ4 Encoding for the First P450 Monooxygenase of Angular Furanocoumarin Biosynthesis. J. Biol. Chem. *284*, 4776–4785.

Larson, E.D., Groskreutz, S.R., Harmes, D.C., Gibbs-Hall, I.C., Trudo, S.P., Allen, R.C., Rutan, S.C., and Stoll, D.R. (2013). Development of selective comprehensive two-dimensional liquid chromatography with parallel first-dimension sampling and second-dimension separation—application to the quantitative analysis of furanocoumarins in apiaceous vegetables. Anal. Bioanal. Chem. *405*, 4639–4653.

Latunde-Dada, A.O., Cabello-Hurtado, F., Czittrich, N., Didierjean, L., Schopfer, C., Hertkorn, N., Werck-Reichhart, D., and Ebel, J. (2001). Flavonoid 6-Hydroxylase from Soybean (*Glycine max* L.), a Novel Plant P-450 Monooxygenase. J. Biol. Chem. *276*, 1688–1695.

Le Bouquin, R., Skrabs, M., Kahn, R., Benveniste, I., Salaün, J.-P., Schreiber, L., Durst, F., and Pinot, F. (2001). CYP94A5, a new cytochrome P450 from *Nicotiana tabacum* is able to catalyze the oxidation of fatty acids to the ω-alcohol and to the corresponding diacid. Eur. J. Biochem. *268*, 3083–3090.

Lee-Parsons, C.W., Ertürk, S., and Tengtrakool, J. (2004). Enhancement of ajmalicine production in *Catharanthus roseus* cell cultures with methyl jasmonate is dependent on timing and dosage of elicitation. Biotechnol. Lett. *26*, 1595–1599.

Levy, S., Sutton, G., Ng, P.C., Feuk, L., Halpern, A.L., Walenz, B.P., Axelrod, N., Huang, J., Kirkness, E.F., and Denisov, G. (2007). The diploid genome sequence of an individual human. PLoS Biol. *5*, e254.

Li, H., Pinot, F., Sauveplane, V., Werck-Reichhart, D., Diehl, P., Schreiber, L., Franke, R., Zhang, P., Chen, L., Gao, Y., *et al*. (2010). Cytochrome P450 family member CYP704B2 catalyzes the omega-hydroxylation of fatty acids and is required for anther cutin biosynthesis and pollen exine formation in rice. Plant Cell Online *22*, 173–190.

Li, W., Schuler, M.A., and Berenbaum, M.R. (2003). Diversification of furanocoumarin-metabolizing cytochrome P450 monooxygenases in two papilionids: Specificity and substrate encounter rate. Proc. Natl. Acad. Sci. *100*, 14593–14598.

Lievre, K., Hehn, A., Tran, T.L.M., Gravot, A., Thomasset, B., Bourgaud, F., and Gontier, E. (2005). Genetic transformation of the medicinal plant *Ruta graveolens* L. by an *Agrobacterium tumefaciens* mediated method. Plant Sci. *168*, 883–888.

Lin, H., Kent, U.M., and Hollenberg, P.F. (2002). Mechanism-based inactivation of cytochrome P450 3A4 by 17α-ethynylestradiol: evidence for heme destruction and covalent binding to protein. J. Pharmacol. Exp. Ther. *301*, 160–167.

Lin, H.-M., Wang, W.-K., Hsiung, P.-A., and Shyu, S.-G. (2010). Light-sensitive intelligent drug delivery systems of coumarin-modified mesoporous bioactive glass. Acta Biomater. *6*, 3256–3263.

Liu, Y.-G., and Whittier, R.F. (1995). Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. Genomics *25*, 674–681.

Liu, F., Sun, G., Gao, H., Li, R., Soromou, L.-W., Chen, N., Deng, Y.-H., and Feng, H. (2013). Angelicin regulates LPS-induced inflammation *via* inhibiting MAPK/NF-κB pathways. J. Surg. Res. *185*, 300–309.

Liu, Y.-G., Chen, Y., and Zhang, Q. (2004). Amplification of genomic sequences flanking T-DNA insertions by thermal asymmetric interlaced polymerase chain reaction. In Transgenic Plants: Methods and Protocols, (Springer), pp. 341–348.

Logemann, E., Parniske, M., and Hahlbrock, K. (1995). Modes of expression and common structural features of the complete phenylalanine ammonia-lyase gene family in parsley. Proc. Natl. Acad. Sci. *92*, 5905–5909.

Lottaz, C., Iseli, C., Jongeneel, C.V., and Bucher, P. (2003). Modeling sequencing errors by combining Hidden Markov models. Bioinformatics *19*, ii103–ii112.

Lovell, C.R. (1997). Phytodermatitis. Clin. Dermatol. *15*, 607–613.

Lowe, T.M., and Eddy, S.R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. *25*, 0955–0964.

Lukashin, A.V., and Borodovsky, M. (1998). GeneMark. hmm: new solutions for gene finding. Nucleic Acids Res. *26*, 1107–1115.

Majoros, W.H., Pertea, M., and Salzberg, S.L. (2004). TigrScan and GlimmerHMM: two open source *ab initio* eukaryotic gene-finders. Bioinformatics *20*, 2878–2879.

Mandel, M.J., Stabb, E.V., and Ruby, E.G. (2008). Comparative genomics-based investigation of resequencing targets in *Vibrio fischeri*: focus on point miscalls and artefactual expansions. BMC Genomics *9*, 138.

Mansuy, D. (1994). Cytochromes P-450 and model systems: great diversity of catalyzed reactions. Pure Appl. Chem. *66*, 737–744.

Mansuy, D. (1998). The great diversity of reactions catalyzed by cytochromes P450. Comp. Biochem. Physiol. C Pharmacol. Toxicol. Endocrinol. *121*, 5–14.

Mao, W., Schuler, M.A., and Berenbaum, M.R. (2007). Cytochrome P450s in *Papilio multicaudatus* and the transition from oligophagy to polyphagy in the Papilionidae. Insect Mol. Biol. *16*, 481–490.

Marcotte, C.J.V., and Marcotte, E.M. (2002). Predicting functional linkages from gene fusions with confidence. Appl. Bioinformatics *1*, 93–100.

Marí, M., and Cederbaum, A.I. (2001). Induction of catalase, alpha, and microsomal glutathione S-transferase in CYP2E1 overexpressing HepG2 cells and protection against short-term oxidative stress. Hepatology *33*, 652–661.

Marrelli, M., Menichini, F., Statti, G.A., Bonesi, M., Duez, P., Menichini, F., and Conforti, F. (2012). Changes in the phenolic and lipophilic composition, in the enzyme inhibition and antiproliferative activity of *Ficus carica* L. cultivar Dottato fruits during maturation. Food Chem. Toxicol. *50*, 726–733.

Marzano, C., Chilin, A., Bordin, F., Baccichetti, F., and Guiotto, A. (2002). DNA damage and biological effects induced by photosensitization with new N_1-unsubstituted furo [2, 3-*h*] quinolin-2 (1*H*)-ones. Bioorg. Med. Chem. *10*, 2835–2844.

Massot, B., Milesi, S., Gontier, E., Bourgaud, F., and Guckert, A. (2000). Optimized culture conditions for the production of furanocoumarins by micropropagated shoots of *Ruta graveolens*. Plant Cell Tissue Organ Cult. *62*, 11–19.

Matern, O., Strasser, H., Wendorff, H., and Hamerski, D. (1988). Coumarins and furanocoumarins.

Matsubara, T., Baron, J., Peterson, L.L., and Peterson, J.A. (1976). NADPH-cytochrome P450 reductase. Arch. Biochem. Biophys. *172*, 463–469.

Matsumoto, S., Mizutani, M., Sakata, K., and Shimizu, B.-I. (2012). Molecular cloning and functional analysis of the ortho-hydroxylases of p-coumaroyl coenzyme A/feruloyl coenzyme A involved in formation of umbelliferone and scopoletin in sweet potato, *Ipomoea batatas* (L.) Lam. Phytochemistry *74*, 49–57.

Matthews, L., Gopinath, G., Gillespie, M., Caudy, M., Croft, D., de Bono, B., Garapati, P., Hemish, J., Hermjakob, H., and Jassal, B. (2009). Reactome knowledgebase of human biological pathways and processes. Nucleic Acids Res. *37*, D619–D622.

McCloud, E.S., Berenbaum, M.R., and Tuveson, R.W. (1992). Furanocoumarin content and phototoxicity of rough lemon (*Citrus jambhiri*) foliage exposed to enhanced ultraviolet-B (UVB) irradiation. J. Chem. Ecol. *18*, 1125–1137.

McKey, D. (1974). Adaptive patterns in alkaloid physiology. Am. Nat. 305–320.

Meng, H., Xu, J., Guo, W., Xu, Q., Li, D., and Deng, X. (2005). A protocol for rapid isolation of flanking regions from short known sequences. Plant Mol. Biol. Report. *23*, 75–75.

Meyer, K., Cusumano, J.C., Somerville, C., and Chapple, C.C. (1996). Ferulate-5-hydroxylase from *Arabidopsis thaliana* defines a new family of cytochrome P450-dependent monooxygenases. Proc. Natl. Acad. Sci. *93*, 6869–6874.

Michalak, P. (2008). Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes. Genomics *91*, 243–248.

Mignotte, B., and Vayssiere, J.-L. (1998). Mitochondria and apoptosis. Eur. J. Biochem. *252*, 1–15.

Milesi, S., Massot, B., Gontier, E., Bourgaud, F., and Guckert, A. (2001). *Ruta graveolens* L.: a promising species for the production of furanocoumarins. Plant Sci. *161*, 189–199.

Mizutani, M., and Ohta, D. (1998). Two Isoforms of NADPH: Cytochrome P450 Reductase in *Arabidopsis thaliana* Gene Structure, Heterologous Expression in Insect Cells, and Differential Regulation. Plant Physiol. *116*, 357–367.

Mizutani, M., and Ohta, D. (2010). Diversification of P450 genes during land plant evolution. Annu. Rev. Plant Biol. *61*, 291–315.

Mizutani, M., Ward, E., Dimaio, J., Ohta, D., Ryals, J., and Sato, R. (1993). Molecular Cloning and Sequencing of a cDNA Encoding Mung Bean Cytochrome P450 (P450C4H) Possessing Cinnamate 4-Hydroxylase Activity. Biochem. Biophys. Res. Commun. *190*, 875–880.

Moor, A.C., and Gasparro, F.P. (1996). Biochemical aspects of psoralen photochemotherapy. Clin. Dermatol. *14*, 353–365.

Morant, M., Hehn, A., and Werck-Reichhart, D. (2002). Conservation and diversity of gene families explored using the CODEHOP strategy in higher plants. BMC Plant Biol. *2*, 7.

Moreira, A.I.D., Leite, I., Guedes, R., Baptista, A., and Ferreira, E.O. (2010). Phytophotodermatitis: an occupational and recreational skin disease. Rev. Assoc. Médica Bras. *56*, 269–270.

Morikawa, T., Mizutani, M., Aoki, N., Watanabe, B., Saga, H., Saito, S., Oikawa, A., Suzuki, H., Sakurai, N., and Shibata, D. (2006). Cytochrome P450 CYP710A encodes the sterol C-22 desaturase in Arabidopsis and tomato. Plant Cell Online *18*, 1008–1022.

Mulder, N., and Apweiler, R. (2007). InterPro and InterProScan. In Comparative Genomics, (Springer), pp. 59–70.

Murray, M. (2000). Mechanisms of inhibitory and regulatory effects of methylenedioxyphenyl compounds on cytochrome P450-dependent drug oxidation. Curr. Drug Metab. *1*, 67–84.

Murray, R.D.H., Méndez, J., and Brown, S.A. (1982). The natura, coumarins: Occurrence, Chemistry and Biochemistry.

Musajo, L., Rodighiero, G., Colombo, G., Torlone, V., and Dall'Acqua, F. (1965). Photosensitizing furocoumarins: interaction with DNA and photo-inactivation of DNA containing viruses. Experientia *21*, 22–24.

Nafisi, M., Goregaoker, S., Botanga, C.J., Glawischnig, E., Olsen, C.E., Halkier, B.A., and Glazebrook, J. (2007). Arabidopsis cytochrome P450 monooxygenase 71A13 catalyzes the conversion of indole-3-acetaldoxime in camalexin synthesis. Plant Cell Online *19*, 2039–2052.

Nair, R.B., Xia, Q., Kartha, C.J., Kurylo, E., Hirji, R.N., Datla, R., and Selvaraj, G. (2002). Arabidopsis CYP98A3 Mediating Aromatic 3-Hydroxylation. Developmental Regulation of the Gene, and Expression in Yeast. Plant Physiol. *130*, 210–220.

Naur, P., Petersen, B.L., Mikkelsen, M.D., Bak, S., Rasmussen, H., Olsen, C.E., and Halkier, B.A. (2003). CYP83A1 and CYP83B1, Two Nonredundant Cytochrome P450 Enzymes Metabolizing Oximes in the Biosynthesis of Glucosinolates in Arabidopsis. Plant Physiol. *133*, 63–72.

Neal, J.J., and Wu, D. (1994). Inhibition of insect cytochromes P450 by furanocoumarins. Pestic. Biochem. Physiol. *50*, 43–50.

Nelson, D.R. (1999). Cytochrome P450 and the individuality of species. Arch. Biochem. Biophys. *369*, 1–10.

Nelson, D.R., Kamataki, T., Waxman, D.J., Guengerich, F.P., Estabrook, R.W., Feyereisen, R., Gonzalez, F.J., Coon, M.J., Gunsalus, I.C., and Gotoh, O. (1993). The P450 superfamily: update on new sequences, gene mapping, accession numbers, early trivial names of enzymes, and nomenclature. DNA Cell Biol. *12*, 1–51.

Nelson, D.R., Schuler, M.A., Paquette, S.M., Werck-Reichhart, D., and Bak, S. (2004). Comparative genomics of rice and Arabidopsis. Analysis of 727 cytochrome P450 genes and pseudogenes from a monocot and a dicot. Plant Physiol. *135*, 756–772.

Nicol, D.L., Finlayson, J., Colmer, T.D., and Ryan, M.H. (2013). Opportunistic Mediterranean agriculture–Using ephemeral pasture legumes to utilize summer rainfall. Agric. Syst. *120*, 76–84.

Niggeweg, R., Michael, A.J., and Martin, C. (2004). Engineering plants with increased levels of the antioxidant chlorogenic acid. Nat. Biotechnol. *22*, 746–754.

Nishiyama, Y., Yun, C.-S., Matsuda, F., Sasaki, T., Saito, K., and Tozawa, Y. (2010). Expression of bacterial tyrosine ammonia-lyase creates a novel p-coumaric acid pathway in the biosynthesis of phenylpropanoids in Arabidopsis. Planta *232*, 209–218.

Nitao, J.K., and Zangerl, A.R. (1987). Floral development and chemical defense allocation in wild parsnip (*Pastinaca sativa*). Ecology *68*, 521–529.

Nograles, K.E., Davidovici, B., and Krueger, J.G. (2010). New insights in the immunologic basis of psoriasis. In Seminars in Cutaneous Medicine and Surgery, (NIH Public Access), p. 3.

Nuringtyas, T.R., Choi, Y.H., Verpoorte, R., Klinkhamer, P.G.L., and Leiss, K.A. (2012). Differential tissue distribution of metabolites in *Jacobaea vulgaris*, *Jacobaea aquatica* and their crosses. Phytochemistry *78*, 89–97.

Obach, R.S., Walsky, R.L., and Venkatakrishnan, K. (2007). Mechanism-based inactivation of human cytochrome P450 enzymes and the prediction of drug-drug interactions. Drug Metab. Dispos. *35*, 246–255.

Ober, D. (2010). Gene duplications and the time thereafter–examples from plant secondary metabolism. Plant Biol. *12*, 570–577.

Ohnishi, A., Matsuo, H., Yamada, S., Takanaga, H., Morimoto, S., Shoyama, Y., Ohtani, H., and Sawada, Y. (2000). Effect of furanocoumarin derivatives in grapefruit juice on the uptake of vinblastine by Caco-2 cells and on the activity of cytochrome P450 3A4. Br. J. Pharmacol. *130*, 1369–1377.

Ohta, D., and Mizutani, M. (2004). Redundancy or flexibility: molecular diversity of the electron transfer components for P450 monooxygenases in higher plants. Front. Biosci. J. Virtual Libr. *9*, 1587–1597.

Ojala, T. (2001). Biological screening of plant coumarins (University of Helsinki Helsinki, Finland).

Ojala, T., Remes, S., Haansuu, P., Vuorela, H., Hiltunen, R., Haahtela, K., and Vuorela, P. (2000). Antimicrobial activity of some coumarin containing herbal plants growing in Finland. J. Ethnopharmacol. *73*, 299–305.

Okada, A., Okada, K., Miyamoto, K., Koga, J., Shibuya, N., Nojiri, H., and Yamane, H. (2009). OsTGAP1, a bZIP transcription factor, coordinately regulates the inductive production of diterpenoid phytoalexins in rice. J. Biol. Chem. *284*, 26510–26518.

Oliva, A., Meepagala, K.M., Wedge, D.E., Harries, D., Hale, A.L., Aliotta, G., and Duke, S.O. (2003). Natural fungicides from *Ruta graveolens* L. leaves, including a new quinolone alkaloid. J. Agric. Food Chem. *51*, 890–896.

Olsen, K.M., Lea, U.S., Slimestad, R., Verheul, M., and Lillo, C. (2008). Differential expression of four *Arabidopsis PAL* genes; *PAL1* and *PAL2* have functional specialization in abiotic environmental-triggered flavonoid synthesis. J. Plant Physiol. *165*, 1491–1499.

Olsen, K.M., Hehn, A., Jugdé, H., Slimestad, R., Larbat, R., Bourgaud, F., and Lillo, C. (2010). Identification and characterisation of CYP75A31, a new flavonoid 3'5'-hydroxylase, isolated from *Solanum lycopersicum*. BMC Plant Biol. *10*, 21.

Omura, T., and Sato, R. (1964). The carbon monoxide-binding pigment of liver microsomes. I. Evidence for its hemoprotein nature. J Biol Chem *239*, 2370–2378.

Orlita, A., Sidwa-Gorycka, M., Malinski, E., Czerwicka, M., Kumirska, J., Golebiowski, M., Lojkowska, E., and Stepnowski, P. (2008a). Effective biotic elicitation of *Ruta graveolens* L. shoot cultures by lysates from *Pectobacterium atrosepticum* and *Bacillus sp.* Biotechnol. Lett. *30*, 541–545.

Orlita, A., Sidwa-Gorycka, M., Kumirska, J., Maliński, E., Siedlecka, E.M., Gajdus, J., \Lojkowska, E., and Stepnowski, P. (2008b). Identification of *Ruta graveolens* L. metabolites accumulated in the presence of abiotic elicitors. Biotechnol. Prog. *24*, 128–133.

Osbourn, A.E., and Field, B. (2009). Operons. Cell. Mol. Life Sci. *66*, 3755–3775.

Overbeek, R., Fonstein, M., D'souza, M., Pusch, G.D., and Maltsev, N. (1999). The use of gene clusters to infer functional coupling. Proc. Natl. Acad. Sci. *96*, 2896–2901.

Pae, H.-O., Oh, H., Yun, Y.-G., Oh, G.-S., Jang, S.I., Hwang, K.-M., Kwon, T.-O., Lee, H.-S., and Chung, H.-T. (2002). Imperatorin, a Furanocoumarin from *Angelica dahurica* (Umbelliferae), Induces Cytochrome c-Dependent Apoptosis in Human Promyelocytic Leukaemia, HL-60 Cells. Pharmacol. Toxicol. *91*, 40–48.

Page, N.A., Wall, R.E., Darbyshire, S.J., and Mulligan, G.A. (2006). The biology of invasive alien plants in Canada. 4. *Heracleum mantegazzianum* Sommier & Levier. Can. J. Plant Sci. *86*, 569–589.

Paquette, S.M., Bak, S., and Feyereisen, R. (2000). Intron-exon organization and phylogeny in a large superfamily, the paralogous cytochrome P450 genes of *Arabidopsis thaliana*. DNA Cell Biol. *19*, 307–317.

Park, J.H., Park, N.I., Xu, H., and Park, S.U. (2010). Cloning and Characterization of Phenylalanine Ammonia-Lyase and Cinnamate 4-Hydroxylase and Pyranocoumarin Biosynthesis in *Angelica gigas*. J. Nat. Prod. *73*, 1394–1397.

Parrish, J.A., Fitzpatrick, T.B., Tanenbaum, L., and Pathak, M.A. (1974). Photochemotherapy of psoriasis with oral methoxsalen and longwave ultraviolet light. N. Engl. J. Med. *291*, 1207–1211.

Pathak, M.A., and Fitzpatrick, T.B. (1992). The evolution of photochemotherapy with psoralens and UVA (PUVA): 2000 BC to 1992 AD. J. Photochem. Photobiol. B *14*, 3–22.

Pathak, M.A., Daniels, F., and Fitzpatrick, T.B. (1962). The Presently Known Distribution of Furocoumarins (Psoralens) in Plants1. J. Invest. Dermatol. *39*, 225–239.

Pavela, R., and Vrchotová, N. (2013). Insecticidal effect of furanocoumarins from fruits of *Angelica archangelica* L. against larvae *Spodoptera littoralis* Boisd. Ind. Crops Prod. *43*, 33–39.

Perlak, F.J., Fuchs, R.L., Dean, D.A., McPherson, S.L., and Fischhoff, D.A. (1991). Modification of the coding sequence enhances plant expression of insect control protein genes. Proc. Natl. Acad. Sci. *88*, 3324–3328.

Peroutka, R., Schulzová, V., Botek, P., and Hajšlová, J. (2007). Analysis of furanocoumarins in vegetables (Apiaceae) and citrus fruits (Rutaceae). J. Sci. Food Agric. *87*, 2152–2163.

Petersen, R.A., Zangerl, A.R., Berenbaum, M.R., and Schuler, M.A. (2001). Expression of *CYP6B1* and *CYP6B3* cytochrome P450 monooxygenases and furanocoumarin metabolism in different tissues of *Papilio polyxenes* (Lepidoptera: Papilionidae). Insect Biochem. Mol. Biol. *31*, 679–690.

Pierrel, M.A., Batard, Y., Kazmaier, M., Mignotte-Vieux, C., Durst, F., and Werck-Reichhart, D. (1994). Catalytic Properties of the Plant Cytochrome P450 CYP73 Expressed in Yeast. Eur. J. Biochem. *224*, 835–844.

Poiesz, B.J., Ruscetti, F.W., Gazdar, A.F., Bunn, P.A., Minna, J.D., and Gallo, R.C. (1980). Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. Proc. Natl. Acad. Sci. *77*, 7415–7419.

Pokrovskii, O.I., Markoliya, A.A., Lepeshkin, F.D., Kuvykin, I.V., Parenago, O.O., and Gonchukov, S.A. (2009). Extraction of linear furocoumarins from *Ammi majus* seeds by means of supercritical fluid extraction and supercritical fluid chromatography. Russ. J. Phys. Chem. B *3*, 1165–1171.

Pompon, D., Louerat, B., Bronine, A., and Urban, P. (1996). Yeast expression of animal and plant P450s in optimized redox environments. Methods Enzymol. *272*, 51–64.

Poupardin, R., Reynaud, S., Strode, C., Ranson, H., Vontas, J., and David, J.-P. (2008). Cross-induction of detoxification genes by environmental xenobiotics and insecticides in the mosquito *Aedes aegypti*: Impact on larval tolerance to chemical insecticides. Insect Biochem. Mol. Biol. *38*, 540–551.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., and Glöckner, F.O. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. *35*, 7188–7196.

Qi, X., Bakht, S., Leggett, M., Maxwell, C., Melton, R., and Osbourn, A. (2004). A gene cluster for secondary metabolism in oat: implications for the evolution of metabolic diversity in plants. Proc. Natl. Acad. Sci. U. S. A. *101*, 8233–8238.

Qi, X., Bakht, S., Qin, B., Leggett, M., Hemmings, A., Mellon, F., Eagles, J., Werck-Reichhart, D., Schaller, H., and Lesot, A. (2006). A different function for a member of an ancient and highly conserved cytochrome P450 family: from essential sterols to plant defense. Proc. Natl. Acad. Sci. *103*, 18848–18853.

Quinlan, R.F., Shumskaya, M., Bradbury, L.M., Beltrán, J., Ma, C., Kennelly, E.J., and Wurtzel, E.T. (2012). Synergistic interactions between carotene ring hydroxylases drive lutein formation in plant carotenoid biosynthesis. Plant Physiol. *160*, 204–214.

Ranea, J.A., Yeats, C., Grant, A., and Orengo, C.A. (2007). Predicting protein function with hierarchical phylogenetic profiles: the Gene3D Phylo-Tuner method applied to eukaryotic genomes. PLoS Comput. Biol. *3*, e237.

Rani, A., Singh, K., Sood, P., Kumar, S., and Ahuja, P.S. (2009). p-Coumarate:CoA ligase as a key gene in the yield of catechins in tea [*Camellia sinensis* (L.) O. Kuntze]. Funct. Integr. Genomics *9*, 271–275.

Rice-Evans, C.A., Miller, N.J., and Paganga, G. (1996). Structure-antioxidant activity relationships of flavonoids and phenolic acids. Free Radic. Biol. Med. *20*, 933–956.

Richards, L.A., Lampert, E.C., Bowers, M.D., Dodson, C.D., Smilanich, A.M., and Dyer, L.A. (2012). Synergistic effects of iridoid glycosides on the survival, development and immune response of a specialist caterpillar, *Junonia coenia* (Nymphalidae). J. Chem. Ecol. *38*, 1276–1284.

Ro, D.K., Mah, N., Ellis, B.E., and Douglas, C.J. (2001). Functional characterization and subcellular localization of poplar (*Populus trichocarpa* x *Populus deltoides*) cinnamate 4-hydroxylase. Plant Physiol. *126*, 317–329.

Roberts, E.S., Hopkins, N.E., Alworth, W.L., and Hollenberg, P.F. (1993). Mechanism-based inactivation of cytochrome P450 2B1 by 2-ethynylnaphthalene: identification of an active-site peptide. Chem. Res. Toxicol. *6*, 470–479.

Roberts, E.S., Hopkins, N.E., Foroozesh, M., Alworth, W.L., Halpert, J.R., and Hollenberg, P.F. (1997). Inactivation of cytochrome P450s 2B1, 2B4, 2B6, and 2B11 by arylalkynes. Drug Metab. Dispos. *25*, 1242–1248.

Robineau, T., Batard, Y., Nedelkina, S., Cabello-Hurtado, F., LeRet, M., Sorokine, O., Didierjean, L., and Werck-Reichhart, D. (1998). The chemically inducible plant cytochrome P450 CYP76B1 actively metabolizes phenylureas and other xenobiotics. Plant Physiol. *118*, 1049–1056.

Rojas-Lima, S., Santillan, R.L., Domínguez, M.-A., and Gutiérrez, A. (1999). Furocoumarins of three species of the genus *Dorstenia*. Phytochemistry *50*, 863–868.

Ronaghi, M., Uhlén, M., and Nyren, P. (1998). A sequencing method based on real-time pyrophosphate. Science *281*, 363–365.

Rosler, J., Krekel, F., Amrhein, N., and Schmid, J. (1997). Maize Phenylalanine Ammonia-Lyase Has Tyrosine Ammonia-Lyase Activity. Plant Physiol. *113*, 175–179.

Rouwendal, G.J., Mendes, O., Wolbert, E.J., and de Boer, A.D. (1997). Enhanced expression in tobacco of the gene encoding green fluorescent protein by modification of its codon usage. Plant Mol. Biol. *33*, 989–999.

Rusk, N. (2009). Cheap third-generation sequencing. Nat. Methods *6*, 244–244.

Sahali-Sahly, Y., Balani, S.K., Lin, J.H., and Baillie, T.A. (1996). *In vitro* studies on the metabolic activation of the furanopyridine L-754,394, a highly potent and selective mechanism-based inhibitor of cytochrome P450 3A4. Chem. Res. Toxicol. *9*, 1007–1012.

Saito, S., Hirai, N., Matsumoto, C., Ohigashi, H., Ohta, D., Sakata, K., and Mizutani, M. (2004). Arabidopsis CYP707As Encode (+)-Abscisic Acid 8'-Hydroxylase, a Key Enzyme in the Oxidative Catabolism of Abscisic Acid. Plant Physiol. *134*, 1439–1449.

Sandermann Jr, H. (1992). Plant metabolism of xenobiotics. Trends Biochem. Sci. *17*, 82–84.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. *74*, 5463–5467.

Santosh, K. (2012). Cytochrome P450-mediated phytoremediation using transgenic plants: a need for engineered cytochrome P450 enzymes. J. Pet. Environ. Biotechnol.

Sardari, S., Mori, Y., Horita, K., Micetich, R.G., Nishibe, S., and Daneshtalab, M. (1999). Synthesis and antifungal activity of coumarins and angular furanocoumarins. Bioorg. Med. Chem. *7*, 1933–1940.

Sasaki, M., Meguro, F., Kumazawa, E., Fujita, H., Kakishima, H., and Sakata, T. (1988). Evidence for uptake of 8-methoxypsoralen and 5-methoxypsoralen by cellular nuclei. Mutat. Res. Mol. Mech. Mutagen. *197*, 51–58.

Sashidhara, K.V., Kumar, A., Chatterjee, M., Rao, K.B., Singh, S., Verma, A.K., and Palit, G. (2011). Discovery and synthesis of novel 3-phenylcoumarin derivatives as antidepressant agents. Bioorg. Med. Chem. Lett. *21*, 1937–1941.

Sastry, S.S. (1997). Isolation and Partial Characterization of a Novel Psoralen-Tyrosine Photoconjugate from a Photoreaction of Psoralen with a Natural Protein. Photochem. Photobiol. *65*, 937–944.

Sayre, R.M., and Dowdy, J.C. (2008). The increase in melanoma: Are dietary furocoumarins responsible? Med. Hypotheses *70*, 855–859.

Schalk, M., Cabello-Hurtado, F., Pierrel, M.-A., Atanassova, R., Saindrenan, P., and Werck-Reichhart, D. (1998). Piperonylic Acid, a Selective, Mechanism-Based Inactivator of the trans-Cinnamate 4-Hydroxylase: A New Tool to Control the Flux of Metabolites in the Phenylpropanoid Pathway. Plant Physiol. *118*, 209–218.

Schimmer, O., and Kühne, I. (1990). Mutagenic compounds in an extract from Rutae Herba (*Ruta graveolens* L.) II. UV-A mediated mutagenicity in the green alga *Chlamydomonas reinhardtii* by furoquinoline alkaloids and furocoumarins present in a commercial tincture from Rutae Herba. Mutat. Res. Lett. *243*, 57–62.

Schmitt, I.M., Maxwell, V.M., Olack, G.A., Edelson, R.L., Chimenti, S., and Gasparro, F.P. (1994). The specific effects of 8-methoxypsoralen photoadducts on cell growth: HPLC analysis of monoadduct and crosslink formation in cells exposed to split-dose treatment. J. Photochem. Photobiol. B *22*, 17–21.

Schmitt, I.M., Chimenti, S., and Gasparro, F.P. (1995). Psoralen-protein photochemistry—a forgotten field. J. Photochem. Photobiol. B *27*, 101–107.

Schoch, G., Goepfert, S., Morant, M., Hehn, A., Meyer, D., Ullmann, P., and Werck-Reichhart, D. (2001). CYP98A3 from *Arabidopsis thaliana* Is a 3′-Hydroxylase of Phenolic Esters, a Missing Link in the Phenylpropanoid Pathway. J. Biol. Chem. *276*, 36566–36574.

Schoch, G.A., Yano, J.K., Wester, M.R., Griffin, K.J., Stout, C.D., and Johnson, E.F. (2004). Structure of human microsomal cytochrome P450 2C8 Evidence for a peripheral fatty acid binding site. J. Biol. Chem. *279*, 9497–9503.

Schonberg, A., Schenck, G.O., Neumuller, O.-A., and Schönberg, A. (1968). Preparative organic photochemistry.

Schoonderwoerd, S.A., van Henegouwen, G.M.J., Persons, C.C.M., Caffieri, S., and Dall'Acqua, F. (1991). Photobinding of 8-methoxypsoralen, 4, 6, 4′-trimethylangelicin and chlorpromazine to Wistar rat epidermal biomacromolecules *in vivo*. J. Photochem. Photobiol. B *10*, 257–268.

Schopfer, C.R., and Ebel, J. (1998). Identification of elicitor-induced cytochrome P450s of soybean (*Glycine max* L.) using differential display of mRNA. Mol. Gen. Genet. MGG *258*, 315–322.

Schröder, G., Unterbusch, E., Kaltenbach, M., Schmidt, J., Strack, D., De Luca, V., and Schröder, J. (1999). Light-induced cytochrome P450-dependent enzyme in indole alkaloid biosynthesis: tabersonine 16-hydroxylase. FEBS Lett. *458*, 97–102.

Schröder, J., Schröder, H., Puglisi, S.J., Sinha, R., and Schmidt, B. (2009). SHREC: a short-read error correction method. Bioinformatics *25*, 2157–2163.

Schuler, M.A. (1996). The role of cytochrome P450 monooxygenases in plant-insect interactions. Plant Physiol. *112*, 1411.

Schuler, M.A. (2011). P450s in plant–insect interactions. Biochim. Biophys. Acta BBA-Proteins Proteomics *1814*, 36–45.

Schuler, M.A., and Werck-Reichhart, D. (2003). Functional genomics of P450s. Annu. Rev. Plant Biol. *54*, 629–667.

Scott, E.E., He, Y.A., Wester, M.R., White, M.A., Chin, C.C., Halpert, J.R., Johnson, E.F., and Stout, C.D. (2003). An open conformation of mammalian cytochrome P450 2B4 at 1.6-\AA resolution. Proc. Natl. Acad. Sci. *100*, 13196–13201.

Seitz, C., Ameres, S., and Forkmann, G. (2007). Identification of the molecular basis for the functional difference between flavonoid 3′-hydroxylase and flavonoid 3′,5′-hydroxylase. FEBS Lett. *581*, 3429–3434.

Selengut, J.D., Haft, D.H., Davidsen, T., Ganapathy, A., Gwinn-Giglio, M., Nelson, W.C., Richter, A.R., and White, O. (2007). TIGRFAMs and Genome Properties: tools for the assignment of molecular function and biological process in prokaryotic genomes. Nucleic Acids Res. *35*, D260–D264.

Serrano-Pérez, J.J., Merchan, M., and Serrano-Andres, L. (2008). Photoreactivity of Furocoumarins and DNA in PUVA Therapy: Formation of Psoralen- Thymine Adducts. J. Phys. Chem. B *112*, 14002–14010.

Sharan, M., Taguchi, G., Gonda, K., Jouke, T., Shimosaka, M., Hayashida, N., and Okazaki, M. (1998). Effects of methyl jasmonate and elicitor on the activation of phenylalanine ammonia-lyase and the accumulation of scopoletin and scopolin in tobacco cell cultures. Plant Sci. *132*, 13–19.

Shilling, D.G., Dusky, J.A., Mossier, M.A., and Bewick, T.A. (1992). Allelopathic potential of celery residues on lettuce. J. Am. Soc. Hortic. Sci. *117*, 308–312.

Shimada, Y., Fujioka, S., Miyauchi, N., Kushiro, M., Takatsuto, S., Nomura, T., Yokota, T., Kamiya, Y., Bishop, G.J., and Yoshida, S. (2001). Brassinosteroid-6-Oxidases from Arabidopsis and Tomato Catalyze Multiple C-6 Oxidations in Brassinosteroid Biosynthesis. Plant Physiol. *126*, 770–779.

Shimura, K., Okada, A., Okada, K., Jikumaru, Y., Ko, K.-W., Toyomasu, T., Sassa, T., Hasegawa, M., Kodama, O., and Shibuya, N. (2007). Identification of a biosynthetic gene cluster in rice for momilactones. J. Biol. Chem. *282*, 34013–34018.

Sibbesen, O., Koch, B., Halkier, B.A., and Møller, B.L. (1995). Cytochrome P-450 is a multifunctional heme-thiolate enzyme catalyzing the conversion of L-tyrosine to p-hydroxyphenylacetaldehyde oxime in the biosynthesis of the cyanogenic glucoside dhurrin in *Sorghum bicolor* (L.) Moench. J. Biol. Chem. *270*, 3506–3511.

Sidwa-Gorycka, M., Krolicka, A., Orlita, A., Malinski, E., Golebiowski, M., Kumirska, J., Chromik, A., Biskup, E., Stepnowski, P., and Lojkowska, E. (2009). Genetic transformation of *Ruta graveolens* L. by *Agrobacterium rhizogenes*: hairy root cultures a promising approach for production of coumarins and furanocoumarins. Plant Cell Tissue Organ Cult. PCTOC *97*, 59–69.

Sigurdsson, S., Ogmundsdottir, H.M., and Gudbjarnason, S. (2004). Antiproliferative effect of *Angelica archangelica* fruits. Z. NATURFORSCHUNG C *59*, 523–527.

Sinkel, C., Schwarzer, M.C., Frenking, G., Greiner, A., and Agarwal, S. (2011). Polymer-bound 4-methylcoumarin/1-heptanoyl-5-fluorouracil photodimers: NMR elucidation of dimer structure. Magn. Reson. Chem. *49*, 70–75.

Sjölander, K. (2004). Phylogenomic inference of protein molecular function: advances and challenges. Bioinformatics *20*, 170–179.

Sparkes, I.A., Runions, J., Kearns, A., and Hawes, C. (2006). Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. Nat. Protoc. *1*, 2019–2025.

Staden, R., Beal, K.F., and Bonfield, J.K. (1999). The staden package, 1998. In Bioinformatics Methods and Protocols, (Springer), pp. 115–130.

Stanfill, S.B., Brown, C.R., Yan, X. (Jane), Watson, C.H., and Ashley, D.L. (2006). Quantification of Flavor-Related Compounds in the Unburned Contents of Bidi and Clove Cigarettes. J. Agric. Food Chem. *54*, 8580–8588.

Stange Jr., R.R., Sims, J.J., Midland, S.L., and McDonald, R.E. (1999). Isolation of a phytoalexin, trans-p-coumaryl aldehyde, from *Cucurbita maxima*, Cucurbitaceae. Phytochemistry *52*, 41–43.

Stashenko, E.E., Acosta, R., and Martínez, J.R. (2000). High-resolution gas-chromatographic analysis of the secondary metabolites obtained by subcritical-fluid extraction from Colombian rue (*Ruta graveolens* L.). J. Biochem. Biophys. Methods *43*, 379–390.

Steck, W., and Brown, S.A. (1970). Biosynthesis of angular furanocoumarins. Can. J. Biochem. *48*, 872–880.

Stern, R.S., Nichols, K.T., and Väkevä, L.H. (1997). Malignant melanoma in patients treated for psoriasis with methoxsalen (psoralen) and ultraviolet A radiation (PUVA). N. Engl. J. Med. *336*, 1041–1045.

Stevenson, P.C., Simmonds, M.S., Yule, M.A., Veitch, N.C., Kite, G.C., Irwin, D., and Legg, M. (2003). Insect antifeedant furanocoumarins from *Tetradium daniellii*. Phytochemistry *63*, 41–46.

Sturm, J., and Sturm, J.G. (1796). Deutschlands Flora in Abbildungen. Gedruckt Auf Kosten Verfass. Nurnberg Ger.

Sumiyoshi, M., Sakanaka, M., Taniguchi, M., Baba, K., and Kimura, Y. (2014). Anti-tumor effects of various furocoumarins isolated from the roots, seeds and fruits of Angelica and Cnidium species under ultraviolet A irradiation. J. Nat. Med. *68*, 83–94.

Szopa, A., Ekiert, H., Szewczyk, A., and Fugas, E. (2012). Production of bioactive phenolic acids and furanocoumarins in in vitro cultures of *Ruta graveolens* L. and *Ruta graveolens* ssp. *divaricata* (Tenore) Gams. under different light conditions. Plant Cell Tissue Organ Cult. PCTOC *110*, 329–336.

Taguchi, G., Fujikawa, S., Yazawa, T., Kodaira, R., Hayashida, N., Shimosaka, M., and Okazaki, M. (2000). Scopoletin uptake from culture medium and accumulation in the vacuoles after conversion to scopolin in 2, 4-D-treated tobacco cells. Plant Sci. *151*, 153–161.

Tal, B., and Robeson, D.J. (1985). The induction, by fungal inoculation, of ayapin and scopoletin biosynthesis in *Helianthus annuus*. Phytochemistry *25*, 77–79.

Tal, B., and Robeson, D.J. (1986). The Metabolism of Sunflower Phytoalexins Ayapin and Scopoletin Plant-Fungus Interactions. Plant Physiol. *82*, 167–172.

Tassaneeyakul, W., Guo, L.-Q., Fukuda, K., Ohta, T., and Yamazoe, Y. (2000). Inhibition selectivity of grapefruit juice components on human cytochromes P450. Arch. Biochem. Biophys. *378*, 356–363.

Taylor, L.P., and Grotewold, E. (2005). Flavonoids as developmental regulators. Curr. Opin. Plant Biol. *8*, 317–323.

Thomé, O.W. (1885). Flora von Deutschland. Österr. Schweiz Wort Bild Für Sch. Haus Gera Dtschl.

Thompson, H.J., Sharma, S.K., and Brown, S.A. (1978). *O*-methyltransferases of furanocoumarin biosynthesis. Arch. Biochem. Biophys. *188*, 272–281.

Tian, W., and Skolnick, J. (2003). How well is enzyme function conserved as a function of pairwise sequence identity? J. Mol. Biol. *333*, 863–882.

Tietjen, K.G., Hunkler, D., and Matern, U. (1983). Differential Response of Cultured Parsley Cells to Elicitors from Two Non-pathogenic Strains of Fungi. Eur. J. Biochem. *131*, 401–407.

Tijet, N., Helvig, C., and Feyereisen, R. (2001). The cytochrome P450 gene superfamily in *Drosophila melanogaster*: Annotation, intron-exon organization and phylogeny. Gene *262*, 189–198.

Tohge, T., Watanabe, M., Hoefgen, R., and Fernie, A.R. (2013). The evolution of phenylpropanoid metabolism in the green lineage. Crit. Rev. Biochem. Mol. Biol. *48*, 123–152.

Urban, P., Mignotte, C., Kazmaier, M., Delorme, F., and Pompon, D. (1997). Cloning, yeast expression, and characterization of the coupling of two distantly related *Arabidopsis thaliana* NADPH-cytochrome P450 reductases with P450 CYP73A5. J. Biol. Chem. *272*, 19176–19186.

Valle, T., López, J.L., Hernández, J.M., and Corchete, P. (1997). Antifungal activity of scopoletin and its differential accumulation in *Ulmus pumila* and *Ulmus campestris* cell suspension cultures infected with *Ophiostoma ulmi* spores. Plant Sci. *125*, 97–101.

Ververidis, F., Trantas, E., Douglas, C., Vollmer, G., Kretzschmar, G., and Panopoulos, N. (2007). Biotechnology of flavonoids and other phenylpropanoid-derived natural products. Part I: Chemical diversity, impacts on plant biology and human health. Biotechnol. J. *2*, 1214–1234.

Vialart, G., Hehn, A., Olry, A., Ito, K., Krieger, C., Larbat, R., Paris, C., Shimizu, B., Sugimoto, Y., Mizutani, M., *et al*. (2012). A 2-oxoglutarate-dependent dioxygenase from *Ruta graveolens* L. exhibits p-coumaroyl CoA 2'-hydroxylase activity (C2'H): a missing step in the synthesis of umbelliferone in plants. Plant J. *70*, 460–470.

Vogt, T. (2010). Phenylpropanoid Biosynthesis. Mol. Plant *3*, 2–20.

Waksmundzka-Hajnos, M., Petruczynik, A., Dragan, A., Wianowska, D., Dawidowicz, A.L., and Sowa, I. (2004). Influence of the extraction mode on the yield of some furanocoumarins from *Pastinaca sativa* fruits. J. Chromatogr. B *800*, 181–187.

Wang, Q., Hillwig, M.L., Wu, Y., and Peters, R.J. (2012). CYP701A8: A Rice ent-Kaurene Oxidase Paralog Diverted to More Specialized Diterpenoid Metabolism. Plant Physiol. *158*, 1418–1425.

Wasmuth, J.D., and Blaxter, M.L. (2004). prot4EST: translating expressed sequence tags from neglected genomes. BMC Bioinformatics *5*, 187.

Wen, Z., Pan, L., Berenbaum, M.R., and Schuler, M.A. (2003). Metabolism of linear and angular furanocoumarins by *Papilio polyxenes* CYP6B1 co-expressed with NADPH cytochrome P450 reductase. Insect Biochem. Mol. Biol. *33*, 937–947.

Wen, Z., Rupasinghe, S., Niu, G., Berenbaum, M.R., and Schuler, M.A. (2006). CYP6B1 and CYP6B3 of the black swallowtail (*Papilio polyxenes*): adaptive evolution through subfunctionalization. Mol. Biol. Evol. *23*, 2434–2443.

Wen, Z., Zeng, R.S., Niu, G., Berenbaum, M.R., and Schuler, M.A. (2009). Ecological significance of induction of broad-substrate cytochrome P450s by natural and synthetic inducers in *Helicoverpa zea*. J. Chem. Ecol. *35*, 183–189.

Wen-Sheng, X., Xiang-Jing, W., Tian-Rui, R., and Su-Qin, C. (2006). Purification of recombinant wheat cytochrome P450 monooxygenase expressed in yeast and its properties. Protein Expr. Purif. *45*, 54–59.

Werck-Reichhart, D., and Feyereisen, R. (2000). Cytochromes P450: a success story. Genome Biol *1*, 3003–1.

Werck-Reichhart, D., Hehn, A., and Didierjean, L. (2000). Cytochromes P450 for engineering herbicide tolerance. Trends Plant Sci. *5*, 116–123.

Wester, M.R., Yano, J.K., Schoch, G.A., Yang, C., Griffin, K.J., Stout, C.D., and Johnson, E.F. (2004). The structure of human cytochrome P450 2C9 complexed with flurbiprofen at 2.0-\AA resolution. J. Biol. Chem. *279*, 35630–35637.

Wheat, C.W., Vogel, H., Wittstock, U., Braby, M.F., Underwood, D., and Mitchell-Olds, T. (2007). The genetic basis of a plant–insect coevolutionary key innovation. Proc. Natl. Acad. Sci. *104*, 20427–20431.

Wheeler, D.A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., He, W., Chen, Y.-J., Makhijani, V., and Roth, G.T. (2008). The complete genome of an individual by massively parallel DNA sequencing. Nature *452*, 872–876.

Whitehead, S.R., and Bowers, M.D. (2013). Iridoid and secoiridoid glycosides in a hybrid complex of bush honeysuckles (*Lonicera spp.*, Caprifolicaceae): Implications for evolutionary ecology and invasion biology. Phytochemistry *86*, 57–63.

Wickramasinghe, R.H., and Villee, C.A. (1975). Early role during chemical evolution for cytochrome P450 in oxygen detoxification. Nature *256*, 509–511.

Williams, P.A., Cosme, J., Sridhar, V., Johnson, E.F., and McRee, D.E. (2000a). Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity. Mol. Cell *5*, 121–131.

Williams, P.A., Cosme, J., Sridhar, V., Johnson, E.F., and McRee, D.E. (2000b). Microsomal cytochrome P450 2C5: comparison to microbial P450s and unique features. J. Inorg. Biochem. *81*, 183–190.

Williams, P.A., Cosme, J., Vinković, D.M., Ward, A., Angove, H.C., Day, P.J., Vonrhein, C., Tickle, I.J., and Jhoti, H. (2004). Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone. Science *305*, 683–686.

Wilmouth, R.C., Turnbull, J.J., Welford, R.W., Clifton, I.J., Prescott, A.G., and Schofield, C.J. (2002). Structure and Mechanism of Anthocyanidin Synthase from *Arabidopsis thaliana*. Structure *10*, 93–103.

Wittstock, U., and Gershenzon, J. (2002). Constitutive plant toxins and their role in defense against herbivores and pathogens. Curr. Opin. Plant Biol. *5*, 300–307.

Wu, S., and Chappell, J. (2008). Metabolic engineering of natural products in plants; tools of the trade and challenges for the future. Curr. Opin. Biotechnol. *19*, 145–152.

Xu, H., Park, N.I., Li, X., Kim, Y.K., Lee, S.Y., and Park, S.U. (2010). Molecular cloning and characterization of phenylalanine ammonia-lyase, cinnamate 4-hydroxylase and genes involved in flavone biosynthesis in *Scutellaria baicalensis*. Bioresour. Technol. *101*, 9715–9722.

Yano, J.K., Wester, M.R., Schoch, G.A., Griffin, K.J., Stout, C.D., and Johnson, E.F. (2004). The structure of human microsomal cytochrome P450 3A4 determined by X-ray crystallography to 2.05-\AA resolution. J. Biol. Chem. *279*, 38091–38094.

Yano, J.K., Hsu, M.-H., Griffin, K.J., Stout, C.D., and Johnson, E.F. (2005). Structures of human microsomal cytochrome P450 2A6 complexed with coumarin and methoxsalen. Nat. Struct. Mol. Biol. *12*, 822–823.

Zangar, R.C., Davydov, D.R., and Verma, S. (2004). Mechanisms that regulate production of reactive oxygen species by cytochrome *P*450. Toxicol. Appl. Pharmacol. *199*, 316–331.

Zangerl, A.R., and Berenbaum, M.R. (1987). Furanocoumarins in wild parsnip: effects of photosynthetically active radiation, ultraviolet light, and nutrients. Ecology *68*, 516–520.

Zangerl, A.R., and Berenbaum, M.R. (1998). Damage-inducibility of primary and secondary metabolites in the wild parsnip (*Pastinaca sativa*). Chemoecology *8*, 187–193.

Zangerl, A.R., and Rutledge, C.E. (1996). The probability of attack and patterns of constitutive and induced defense: a test of optimal defense theory. Am. Nat. 599–608.

Zangerl, A.R., Green, E.S., Lampman, R.L., and Berenbaum, M.R. (1997). Phenological changes in primary and secondary chemistry of reproductive parts in wild parsnip. Phytochemistry *44*, 825–831.

Zangerl, A.R., Hamilton, J.G., Miller, T.J., Crofts, A.R., Oxborough, K., Berenbaum, M.R., and De Lucia, E.H. (2002). Impact of folivory on photosynthesis is greater than the sum of its holes. Proc. Natl. Acad. Sci. *99*, 1088–1091.

Zarebska, Z., Waszkowska, E., Caffieri, S., and Dall'Acqua, F. (2000). PUVA (psoralen+ UVA) photochemotherapy: processes triggered in the cells. Il Farm. *55*, 515–520.

Zdero, C., Bohlmann, F., and Niemeyer, H.M. (1990). *ent*-Labdane glycosides from *Baccharis pingraea*. Phytochemistry *29*, 2611–2616.

Zeng, R.S., Wen, Z., Niu, G., Schuler, M.A., and Berenbaum, M.R. (2009). Enhanced toxicity and induction of cytochrome P450s suggest a cost of "eavesdropping" in a multitrophic interaction. J. Chem. Ecol. *35*, 526–532.

Zhou, N., Tootle, T.L., and Glazebrook, J. (1999). Arabidopsis PAD3, a gene required for camalexin biosynthesis, encodes a putative cytochrome P450 monooxygenase. Plant Cell Online *11*, 2419–2428.

Ziegler, J., and Facchini, P.J. (2008). Alkaloid biosynthesis: metabolism and trafficking. Annu Rev Plant Biol *59*, 735–769.

Zobel, A.M., and Brown, S.A. (1990a). Dermatitis-inducing furanocoumarins on leaf surfaces of eight species of rutaceous and umbelliferous plants. J. Chem. Ecol. *16*, 693–700.

Zobel, A.M., and Brown, S.A. (1990b). Seasonal changes of furanocoumarin concentrations in leaves of *Heracleum lanatum*. J. Chem. Ecol. *16*, 1623–1634.

Zobel, A.M., and Brown, S.A. (1993). Influence of low-intensity ultraviolet radiation on extrusion of furanocoumarins to the leaf surface. J. Chem. Ecol. *19*, 939–952.

Zobel, A.M., and March, R.E. (1993). Autofluorescence reveals different histological localizations of furanocoumarins in fruit of some Umbelliferae and Leguminosae. Ann. Bot. *71*, 251–255.

Zobel, A.M., Brown, S.A., and Glowniak, K. (1990). Localization of furanocoumarins in leaves, fruits, and seeds of plants causing contact photodermatitis. Planta Med. *56*, 571–572.

Zobel, A.M., Wang, J., March, R.E., and Brown, S.A. (1991). Identification of eight coumarins occurring with psoralen, xanthotoxin, and bergapten on leaf surfaces. J. Chem. Ecol. *17*, 1859–1870.

Zulak, K.G., and Bohlmann, J. (2010). Terpenoid biosynthesis and specialized vascular cells of conifer defense. J. Integr. Plant Biol. *52*, 86–97.

Zumwalt, J.G., and Neal, J.J. (1993). Cytochromes P450 from *Papilio polyxenes*: Adaptations to host plant allelochemicals. Comp. Biochem. Physiol. C Pharmacol. Toxicol. Endocrinol. *106*, 111–118.

ANNEXES

Annexe 1 : Vialart, G., Hehn, A., Olry, A., Ito, K., Krieger, C., Larbat, R., Paris, C., Shimizu, B., Sugimoto, Y., Mizutani, M., *et al.* (2012). A 2-oxoglutarate-dependent dioxygenase from *Ruta graveolens* L. exhibits p-coumaroyl CoA 2′-hydroxylase activity (C2′H): a missing step in the synthesis of umbelliferone in plants. Plant J. 70, 460–470.

the plant journal

SEB Society for Experimental Biology

The Plant Journal (2012) **70**, 460–470 doi: 10.1111/j.1365-313X.2011.04879.x

A 2-oxoglutarate-dependent dioxygenase from *Ruta graveolens* L. exhibits *p*-coumaroyl CoA 2′-hydroxylase activity (C2′H): a missing step in the synthesis of umbelliferone in plants

Guilhem Vialart[1,†], Alain Hehn[1,†,*], Alexandre Olry[1], Kyoko Ito[2], Celia Krieger[1], Romain Larbat[1], Cedric Paris[3], Bun-ichi Shimizu[4], Yukihiro Sugimoto[2], Masaharu Mizutani[1] and Frederic Bourgaud[1]

[1]*UMR 1121 Lorraine University, INRA Agronomie et Environnement Nancy –Colmar, ENSAIA, 2 avenue de la forêt de Haye, 54505 Vandœuvre-lès-Nancy, France,*

[2]*Graduate School of Agricultural Science, Kobe University, Rokkodai, Nada, Kobe 657-8501, Japan,*

[3]*Laboratoire Commun d'Analyse, Lorraine University, ENSAIA, 2 avenue de la forêt de Haye, 54505 Vandœuvre-lès-Nancy, France, and*

[4]*Graduate School of Life Science, Toyo University, Izumino 1-1-1, Itakura, Gunma 374-0193, Japan*

Received 26 July 2011; revised 6 December 2011; accepted 8 December 2011; published online 16 January 2012.
*For correspondence (fax +33 3 83 59 57 99; e-mail alain.hehn@ensaia.inpl-nancy.fr).
†These authors contributed equally to this work.
The GenBank accession number for the *RgJF799117* sequence is JF799117.

SUMMARY

Coumarins are important compounds that contribute to the adaptation of plants to biotic or abiotic stresses. Among coumarins, umbelliferone occupies a pivotal position in the plant phenylpropanoid network. Previous studies indicated that umbelliferone is derived from the *ortho*-hydroxylation of *p*-coumaric acid by an unknown biochemical step to yield 2,4-dihydroxycinnamic acid, which then undergoes spontaneous lactonization. Based on a recent report of a gene encoding a 2-oxoglutarate-dependent dioxygenase from *Arabidopsis thaliana* that exhibited feruloyl CoA 6′-hydroxylase activity (Bourgaud *et al.*, 2006), we combined a bioinformatic approach and a cDNA library screen to identify an orthologous ORF (Genbank accession number JF799117) from *Ruta graveolens* L. This ORF shares 59% amino acid identity with feruloyl CoA 6′-hydroxylase, was functionally expressed in *Escherichia coli*, and converted feruloyl CoA into scopoletin and *p*-coumaroyl CoA into umbelliferone with equal activity. Its bi-functionality was further confirmed *in planta*: transient expression of *JF799117* in *Nicotiana benthamiana* yielded plants with leaves containing high levels of umbelliferone and scopoletin when compared to control plants, which contained barely detectable traces of these compounds. The expression of *JF799117* was also tightly correlated to the amount of umbelliferone that was found in UV-elicited *R. graveolens* leaves. Therefore, *JF799117* encodes a *p*-coumaroyl CoA 2′-hydroxylase in *R. graveolens*, which represents a previously uncharacterized step in the synthesis of umbelliferone in plants. Psoralen, which is an important furanocoumarin in *R. graveolens*, was found to be a competitive inhibitor of the enzyme, and it may exert this effect through negative feedback on the enzyme at an upstream position in the pathway.

Keywords: umbelliferone, 2-oxoglutarate-dependent dioxygenase, coumarin, furanocoumarin, *Ruta graveolens*, scopoletin.

INTRODUCTION

Coumarins (1,2-benzopyrones) are ubiquitous in higher plants, where they display important biological and ecological functions. They contribute to the environmental adaptation of plants, and are primarily involved in defense against phytopathogens (Kùc, 1982), limitation of oxidative bursts (Chong *et al.*, 2002) and tolerance to abiotic stresses (Gutierrez *et al.*, 1995). Furanocoumarins (also known as furobenzo-α-pyrones or furocoumarins) constitute a subclass of coumarins that are generally found in four higher plant families: Apiaceae, Rutaceae, Moraceae and Fabaceae.

© 2011 The Authors
The Plant Journal © 2011 Blackwell Publishing Ltd

Annexe 1 : Vialart, G., Hehn, A., Olry, A., Ito, K., Krieger, C., Larbat, R., Paris, C., Shimizu, B., Sugimoto, Y., Mizutani, M., *et al.* (2012). A 2-oxoglutarate-dependent dioxygenase from *Ruta graveolens* L. exhibits p-coumaroyl CoA 2′-hydroxylase activity (C2′H): a missing step in the synthesis of umbelliferone in plants. Plant J. 70, 460–470.

These compounds demonstrate stronger allelochemical properties than the simple coumarins, and have also been characterized for their contribution to resistance against plant pathogens (Beier and Oertli, 1983), insects (Schuler and Berenbaum, 2003) and environmental stresses (Zobel and Brown, 1993).

Umbelliferone (Figure 1) is a coumarin that is found in all higher plants; it is a pivotal intermediate in the synthesis of more complex derivatives, such as poly-hydroxylated/methoxylated/prenylated coumarins, furanocoumarins and pyranocoumarins (Bourgaud *et al.*, 2006). Early investigations were based on plant feeding and use of radiolabeled precursors; these studies provided data concerning the synthesis of umbelliferone in various model plants. Indeed, the first such experiment, performed on *Hierochloe odorata* and *Melilotus officinalis* (Brown *et al.*, 1960), as well as a complementary study also performed on *M. officinalis* (Ellis and Amrhein, 1971), demonstrated incorporation of cinnamic acid into coumarin and 7-hydroxycoumarins (i.e. umbelliferone derivatives). In addition, other tracer experiments were performed on *Lavandula officinalis*, and clearly showed that *para*-hydroxylation was a prerequisite for *ortho*-hydroxylation (Brown, 1962). This result unequivocally demonstrated that umbelliferone is formed after the *ortho*-hydroxylation of *p*-coumaric that leads to synthesis of 2,4-dihydroxycinnamic acid, which is subsequently transformed into umbelliferone following the spontaneous closure of the lactone ring by a non-enzymatic reaction that occurs under acidic or neutral conditions (Figure 1) (reviewed by Bourgaud *et al.*, 2006). In another set of pioneering experiments, double-labeled (*ortho*-^{3}H, *ring*-1-^{14}C) cinnamic acid was fed to *Melilotus alba* or *Gaultheria*

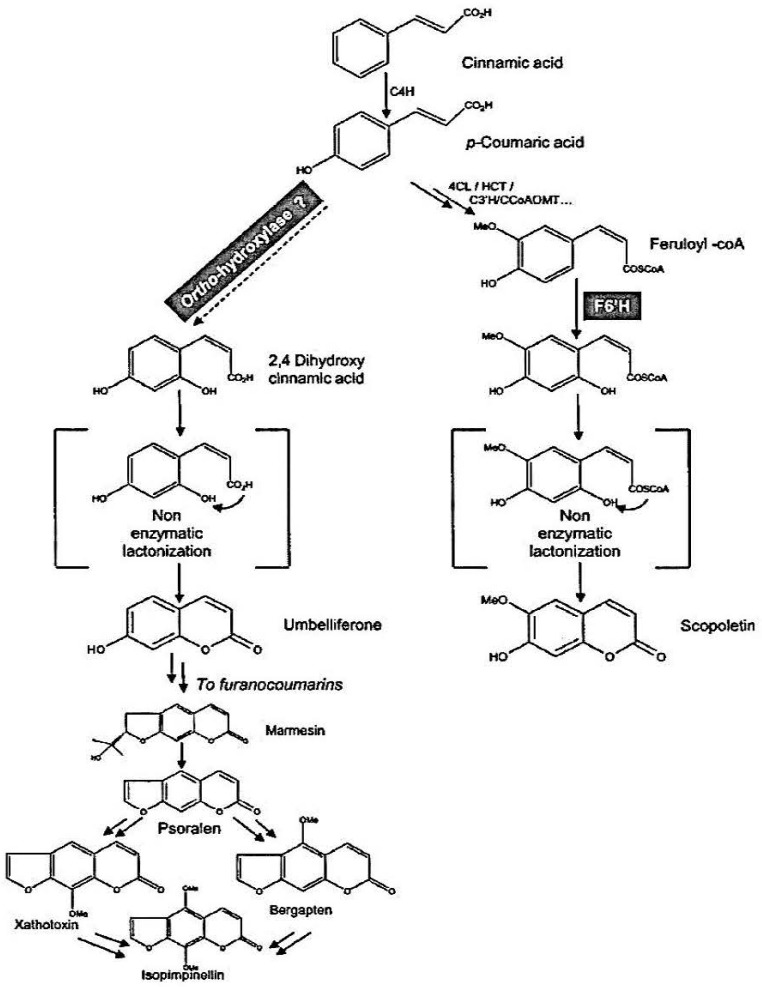


Figure 1. Schematic representation of the umbelliferone/furanocoumarins and scopoletin biosynthetic pathways.
Multiple arrows represent multi-enzymatic steps.
Dashed arrow: *ortho*-hydroxylation step to umbelliferone.

© 2011 The Authors
The Plant Journal © 2011 Blackwell Publishing Ltd,

Annexe 1 : Vialart, G., Hehn, A., Olry, A., Ito, K., Krieger, C., Larbat, R., Paris, C., Shimizu, B., Sugimoto, Y., Mizutani, M., *et al.* (2012). A 2-oxoglutarate-dependent dioxygenase from *Ruta graveolens* L. exhibits p-coumaroyl CoA 2′-hydroxylase activity (C2′H): a missing step in the synthesis of umbelliferone in plants. Plant J. 70, 460–470.

procumbens, and retention of the label was monitored upon its conversion to *o*-coumaric acid, which is the immediate precursor of coumarin (Ellis and Amrhein, 1971). The significant decrease in the ^{3}H:^{14}C ratio was interpreted as an indication of a cytochrome P450 mono-oxygenase reaction mechanism. Another study of *M. alba* assigned the *ortho*-hydroxylation of cinnamic acid to the chloroplast; again, the results suggested the involvement of a P450-dependent mechanism (Gestetner and Conn, 1974). Unfortunately, these *in vitro* results could not be reproduced, and the enzyme family involved in the *ortho*-hydroxylation of *p*-coumaric acid remained a missing link in the process of phenylpropanoid biosynthesis.

The first evidence demonstrating the pivotal role of a 2-oxoglutarate-dependent dioxygenase (2-OGD) in the *ortho*-hydroxylation of cinnamic acid derivatives was only obtained recently, when the biosynthetic pathway of scopoletin (6-methoxy,7-hydroxy-coumarin) was established in *Arabidopsis thaliana*. The first set of experiments utilized T-DNA insertion mutants within the gene that encodes CYP98A3, which is a P450-dependent enzyme that catalyzes the 3′-hydroxylation of *p*-coumaroyl CoA shikimate or quinate (Kai *et al.*, 2006). Transgenic knockout plants exhibited dramatic decreases in scopoletin, confirming its origin from ferulic acid in *A. thaliana* (Figure 1). The feruloyl CoA 6′-hydroxylase (F6′H) that is responsible for conversion of feruloyl CoA into scopoletin was then isolated and characterized as a 2-OGD enzyme (Kai *et al.*, 2008). Because scopoletin is structurally similar to umbelliferone, it is tempting to assume that the *ortho*-hydroxylation of *p*-coumaric acid that facilitates the formation of umbelliferone may be catalyzed by an enzyme that is similar to F6′H. However, F6′H displayed a narrow specificity; feruloyl CoA was the only substrate that was accepted among cinnamic acid derivatives. This finding left the method of *ortho*-hydroxylation of *p*-coumaric acid unresolved, and the process of synthesis of umbelliferone continued to be debated.

Here we describe the identification of a gene that encodes a 2-OGD in *R. graveolens*, which is a furanocoumarin-producing plant; it catalyzes the conversion of feruloyl CoA into scopoletin and *p*-coumaroyl CoA into 2,4-dihydroxycinnamic acid. This latter activity corresponds to a *p*-coumaroyl CoA 2′-hydroxylase (C2′H), which is the missing step of umbelliferone synthesis in plants. This bi-functional activity was confirmed *in vitro* using a purified enzyme that was heterologously produced in *Escherichia coli* and *in planta* following transient expression of the corresponding gene in *Nicotiana benthamiana*. Further *in vitro* studies revealed that this enzyme is negatively regulated by psoralen, which is a furanocoumarin with high toxicity potential in *R. graveolens*; this result demonstrates possible feedback control by this compound upstream of the pathway.

RESULTS

Involvement of a 2-OGD in the furanocoumarin pathway of *Ruta graveolens* and identification of a candidate gene from this plant

The occurrence of a 2-OGD in a *R. graveolens* furanocoumarin pathway was investigated in plantlets that had been sprayed with prohexadione-calcium, which is an inhibitor of this enzyme family (Puhl *et al.*, 2008). The decrease in furanocoumarin content was statistically significant in the leaves of treated plants when compared to control plants: the levels in the control and treated plants were 22.6 ± 2.6 and 15.1 ± 7.1 mg g^{-1} FW (fresh weight) for bergapten ($P < 0.05$), 1.0 ± 0.2 and 0.6 ± 0.3 mg g^{-1} FW for psoralen ($P < 0.05$) and 6.7 ± 0.5 and 5.9 ± 0.6 mg g^{-1} FW for xanthotoxin ($P < 0.05$), respectively. The umbelliferone content also decreased significantly ($P < 0.05$) between control and treated plants (0.05 ± 0.02 and 0.03 ± 0.01 mg g^{-1} FW, respectively), which are unexpectedly low concentrations.

As the first attempt to identify a putative 2-OGD from *R. graveolens*, the *Arabidopsis thaliana* F6′H nucleotide sequence (*At3g13610*) (Kai *et al.*, 2008) was used as a template to identify *in silico* candidates in the Rutaceae family that shared a high level of homology. The first tBLASTn search performed in the GenBank EST database revealed no *R. graveolens* accessions because only a small number of genes are available for this plant species. However, the same *in silico* analysis identified several candidates belonging to the Rutaceae family, and two gene fragments (GenBank accession numbers CX667249 and CX667579) that corresponded to opposite ends of the same sequence revealed the presence of a putative 2-OGD in a *Citrus paradisi* × *Poncirus trifoliata* hybrid. The ORF was amplified from cDNAs prepared from RNA isolated from *Citrus limetta*, using primers based on the two gene fragments. The resulting sequence (GenBank accession number JN245615) was not completely identical to the gene described in the EST database (93% nucleotide identity and 94% amino acid identity), but it showed 63% nucleotide identity and 61% amino acid identity with *A. thaliana* At3g13610. To determine its activity, the gene was introduced into a bacterial expression system. Although the corresponding protein was expressed efficiently in this heterologous system (data not shown), no reaction products of cinnamic acid derivatives were detected.

Based on its high level of identity with the citrus and *A. thaliana* genes previously described, the *JN245615* sequence was used to screen a bacteriophage library to identify an orthologous 2-OGD-encoding gene. The bacteriophage library contained cDNAs prepared from RNA extracted from *R. graveolens* leaves that had been exposed to UV light. This experiment enabled isolation of a new gene (GenBank accession number JF799117, which is hereafter referred to as *RgJF799117*) (Figure 2).

© 2011 The Authors

Annexe 1 : Vialart, G., Hehn, A., Olry, A., Ito, K., Krieger, C., Larbat, R., Paris, C., Shimizu, B., Sugimoto, Y., Mizutani, M., *et al.* (2012). A 2-oxoglutarate-dependent dioxygenase from *Ruta graveolens* L. exhibits p-coumaroyl CoA 2′-hydroxylase activity (C2′H): a missing step in the synthesis of umbelliferone in plants. Plant J. 70, 460–470.


```
At3g13610    MAPT---LLTTQFSNPAEVTDFVVYKGNGVKGLSETGIKALPEQYIQPLE 47
Ruta 2ODG    MAPTKDSVIHMGAESWDEISEFVTKKGHGVKGLSELGIKTLPKQFHQPLE 50
             ****   ::    ..  *:;.**. **;******* ***;**;*; ****

At3g13610    ERLINKFVNETDEAIPVIDMSNPDEDRVAEAVCDAAEKWGFFQVINHGVP 97
Ruta 2OGD    ERFSEKKILER-ASIPLIDMSKWDSPEVVKSICDAAEHWGFFQIVNHGVP 99
             **: :* : *   :**:****: *. .*.:::*****:*****::*****

At3g13610    LEVLDDVKAATHKFFNLPVEEKRKFTKENSLSTTVRFGTSFSPLAEQALE 147
Ruta 2OGD    LETLQRVKEATHRFFALPAEEKNKYSKENSPINNVRFGSSFVPHVEKALE 149
             **.*. ** ***.** **.***.*::****  ..****:** * .*:***

At3g13610    WKDYLSLFFVSEAEAEQFWPDICRNETLEYINKSKKMVRRLLEYLG-KNL 196
Ruta 2OGD    WKDFLSMFYVSEEETNTYWPPICRDEMLEYMRSSEVLIKRLMEVLVVKGL 199
             ***:**:*:*** *:: :** ***.* ***:..*: ::.**:* *  *.*

At3g13610    NVKELDETKESLFMGSIRVNLNYYPICPNPDLTVGVGRHSDVSSLTILLQ 246
Ruta 2OGD    KVKQIDEIREPMLVGSRRINLNYYPKCPNPELTLGVGPHSDISTFTILLQ 249
             :**::** .*.:::** *:****** ****.**:*******:*::*****

At3g13610    DQIGGLHVRSLASGG--NWVHVPPVAGSFVINIGDAMQIMSNGLYKSVEHR 294
Ruta 2OGD    DEIGGLHVRKLDDTGNTWVHVTPISGSLIINIGDALQIMSNGRYKSIEHM 299
             *:*******.* .   .****.*::**::******:****** ***:**

At3g13610    VLANGYNDRISVPIFVNPKPESVIGPLPEVIANGEEPIYRDVLYSDYVKY 344
Ruta 2OGD    VVANGTQPRISVPLFVNPKPQAILCPFPEVLANGEKPVYKPVLCSDYSRH 349
             *:*** :.*****:******:::: *:***:****:*:*. ** *** .:

At3g13610    FFRKAHDGKKTVDYAKI- 361
Ruta 2OGD    FYTKPHDGKKTVDFALMN 367
             *: *.********:* :
```

Figure 2. Sequence alignment of the *Ruta graveolens* putative 2-OGD (Rg 2-OGD GenBank accession number *JF799117*) and the *Arabidopsis thaliana* 2-OGD (*At3g13610*). The sequences were aligned using Clustal W. Asterisks indicate identity between the two sequences; colons indicate conservative substitutions; dots indicate semi-conservative substitutions. The dots and triangles above the alignment indicate amino acids (highlighted) that are involved in recognition sites for oxoglutarate and ferrous iron, respectively.

A phylogenic analysis of 2-OGD sequences (Figure 3) revealed that this enzyme family is organized in clusters, and that each cluster is associated with a specific enzymatic function. In this phylogenic tree, the cloned *RgJF799117* sequence belongs to the same cluster as the two *A. thaliana* sequences that were previously described as F6′Hs (Kai *et al.*, 2008). RgJF799117 shares 59% amino acid identity with At3g13610 (Figure 2) and 81% amino acid identity with citrus JN245615.

In vitro functional characterization

The coding region of the candidate gene *RgJF799117* was cloned and fused with a histidine tag in *E. coli*. Protein extracts were purified by affinity chromatography and incubated in the presence of cinnamic acid, caffeic acid, *p*-coumaric acid, ferulic acid, cinnamoyl CoA, caffeoyl CoA, *p*-coumaroyl CoA or feruloyl CoA. HPLC analysis of the incubations revealed that no conversion occurred with either the four acids or the cinnamoyl and caffeoyl esters. However, the incubations with *p*-coumaroyl CoA and feruloyl CoA led to the accumulation of two products, P1 and P2 (Figure 4a,d), that were absent from the control samples. The UV and MS/MS spectra of P1, which formed from *p*-coumaroyl CoA, perfectly matched the spectra from a commercial umbelliferone standard (Figure 4b,c), whereas P2, which formed from feruloyl CoA, was identified as scopoletin (Figure 4e,f).

Kinetic assays revealed a similar affinity and efficiency for *p*-coumaroyl CoA (K_m = 50 ± 10 μM, $k_{cat\ coum}/K_m$ = 0.85 μmol min^{-1}) and feruloyl CoA (K_m = 37 ± 13 μM, $k_{cat\ fer}/K_m$ = 0.75 μmol min^{-1}) (Table 1). To complete the kinetic study, inhibition tests were performed using umbelliferone, psoralen, xanthotoxin, bergapten and isopimpinellin. Only psoralen had a competitive inhibitory effect on the reaction, characterized by an increase in the K_m in a concentration-dependent manner, and no modification of the V_{max} (Table 1). The estimated K_i values were 111 μM for *p*-coumaroyl CoA and 122 μM for feruloyl CoA.

In planta functional characterization

Further functional characterization of *RgJF799117* was achieved by its transient expression in *Nicotiana benthamiana*. The *RgJF799117* ORF was introduced into a pBIN binary plasmid under the control of the CaMV 35S promoter. The recombinant plasmid was then introduced into *Agrobacterium tumefaciens* strain LBA4404, and was co-inoculated into *N. benthamiana* leaves with an *Agrobacterium* strain expressing the P19 silencing suppressor viral protein of tomato bushy stunt virus (Voinnet *et al.*, 2003). Phenolic extracts from 3-day-old infiltrated leaves contained significant

© 2011 The Authors
The Plant Journal © 2011 Blackwell Publishing Ltd,

Annexe 1 : Vialart, G., Hehn, A., Olry, A., Ito, K., Krieger, C., Larbat, R., Paris, C., Shimizu, B., Sugimoto, Y., Mizutani, M., *et al*. (2012). A 2-oxoglutarate-dependent dioxygenase from *Ruta graveolens* L. exhibits p-coumaroyl CoA 2′-hydroxylase activity (C2′H): a missing step in the synthesis of umbelliferone in plants. Plant J. 70, 460–470.

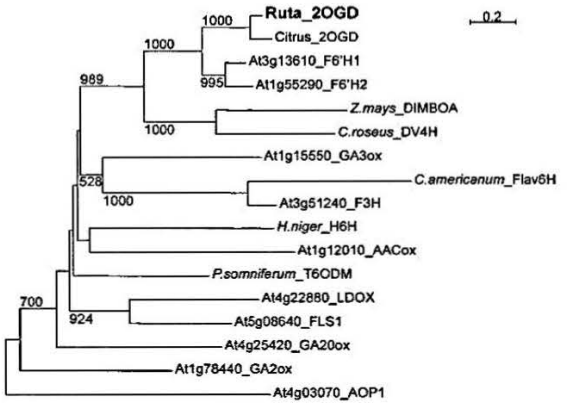


Figure 3. Phylogenetic tree of the 2-OGD family.
The accession numbers in GenBank/EMBL/DDBJ database are given in parentheses: Ruta_2OGD (JF799117); Citrus_2OGD (JN245615); At1g78440_GA2ox (NM_106491); At1g15550_GA3ox (NM_101424); At4g25420_GA20ox (NM_118674); At1g12010_ACCox (NM_101073); At3g13610_F6′H1 (NM_112207); At1g55290_F6′H2 (NM_104404); At3g51240_F3H (NM_114983); At4g22880_LDOX (NM_118417); At5g08640_FLS1 (NM_120951); At4g03070_AOP1 (AF417855); *Zea mays* DIBOA 7-hydroxylase (NM_001111630) (Frey *et al.*, 2003); *Catharanthus roseus* desacetoxyvindoline-4-hydroxylase (AF008597) (Vazquez-Flota *et al.*, 1997); *Chrysosplenium americanum* flavonol 6-hydroxylase (AY605048) (Anzelotti and Ibrahim, 2004); *Papaver somniferum* thebaine 6-*O*-demethylase (GQ500139) (Hagel and Facchini, 2010) and *Hyoscyamus niger* hyoscyamine 6β-hydroxylase (HYSH6H) (Matsuda *et al.*, 1991).

levels of umbelliferone and scopoletin, as revealed by UV and mass spectral analyses; only traces of these products were detected in the control plants (Figure 5). These results from *N. benthamiana* are consistent with the *in vitro* enzymatic investigations, and support the bi-functionality of the enzyme encoded by *RgJF799117* in synthesis of umbelliferone and scopoletin.

Given the potential bi-functionality of *RgJF799117*, the umbelliferone/scopoletin balance was investigated in *R. graveolens* tissues. Scopoletin was not detected by HPLC-DAD in any tissue, but umbelliferone was present in the leaves (0.4 ± 0.1 mg g^{-1} FW) and petioles (0.5 ± 0.1 mg g^{-1} FW), and in trace amounts (<0.05 mg g^{-1} FW) in the stems and roots. Complementary experiments using HPLC-MS on the same plant tissues indicated that scopoletin was present at low concentrations (<10 ng g^{-1} FW) in addition to scopolin and skimmin, which are the glycosylated derivatives of scopoletin and umbelliferone.

Previous experiments indicated that UV light can induce the production of furanocoumarin in *R. graveolens* seedlings (Zobel and Brown, 1993; Zobel *et al.*, 1998). To determine whether there was a correlation between the accumulation of umbelliferone/furanocoumarins and the expression of *RgJF99117*, 3-month-old plantlets were irradiated with UV-B light (312 nm) for 24 h, and compared to non-irradiated control plants. Quantification of the transcripts by real-time PCR revealed a significant induction of *RgJF799117* in the treated leaves and petioles (Figure 6), but no change was detected in the other vegetative tissues (roots and stem). Similarly, the umbelliferone concentration was significantly enhanced ($P < 0.01$) in the UV-treated leaves (1.4 ± 0.1 mg g^{-1} FW) compared to the control plants (0.4 ± 0.1 mg g^{-1} FW). However, the levels were only slightly increased ($0.1 > P > 0.05$) in the UV-treated petioles (0.8 ± 0.2 mg g^{-1} FW) compared to the control plants (0.5 ± 0.1 mg g^{-1} FW), and did not change in the roots and stems (<0.05 mg g^{-1} FW). No significant difference was observed for the levels of furanocoumarins between the UV-treated and control plants. Together, these results show a correlation between the expression of *RgJF799117* and the increase in umbelliferone content in *R. graveolens* leaves and petioles. In a second set of experiments, we analyzed the effect of prohexadion-calcium in UV-treated plants. The results confirmed a decrease in umbelliferone and furanocoumarins for plants treated with both UV and prohexadion-calcium compared to plants with a single UV treatment (umbelliferone, 0.04 ± 0.02 versus 0.08 ± 0.01 mg g^{-1} FW, $P < 0.01$; xanthotoxin, 6.1 ± 0.7 versus 6.8 ± 0.4 mg g^{-1} FW, $P < 0.05$; bergapten, 16.8 ± 5.9 versus 24.4 ± 1.5 mg g^{-1} FW, $P < 0.01$; marmesin, 1.2 ± 0.3 versus 1.8 ± 0.4 mg g^{-1} FW, $P < 0.01$; psoralen, 0.7 ± 0.3 versus 1.1 ± 0.2 mg g^{-1} FW, $P < 0.01$).

DISCUSSION

Umbelliferone is an important coumarin that is ubiquitously produced in plants. Previous studies have demonstrated that umbelliferone is formed following the *ortho*-hydroxylation of *p*-coumaric acid, but molecular evidence regarding its synthesis was unavailable until recently. Elucidation of this *ortho*-hydroxylation step began in 2008 with studies demonstrating the role of a 2-OGD in hydroxylation of feruloyl CoA (F6′H), leading to the formation of scopoletin in *A. thaliana* (Kai *et al.*, 2008). However, it was shown experimentally that *A. thaliana* F6′H is not committed to the formation of umbelliferone (Kai *et al.*, 2008), and the method of biosynthesis of this compound remains unresolved.

Due to obvious structural similarities between scopoletin and umbelliferone, we decided to examine the possible involvement of a 2-OGD in the synthesis of umbelliferone. Investigations were performed in *R. graveolens*, a plant that synthesizes large amounts of furanocoumarins (Milesi *et al.*, 2001), and for which umbelliferone is an important intermediate (Figure 1). The occurrence of a 2-OGD in the furanocoumarin pathway of *R. graveolens* plants was investigated in experiments that utilized prohexadione-calcium, a 2-OGD inhibitor (Puhl *et al.*, 2008), with or without UV treatment. The prohexadione-calcium-treated plants showed a significant decrease in leaf umbelliferone and furanocoumarin

© 2011 The Authors

Annexe 1 : Vialart, G., Hehn, A., Olry, A., Ito, K., Krieger, C., Larbat, R., Paris, C., Shimizu, B., Sugimoto, Y., Mizutani, M., *et al.* (2012). A 2-oxoglutarate-dependent dioxygenase from *Ruta graveolens* L. exhibits p-coumaroyl CoA 2′-hydroxylase activity (C2′H): a missing step in the synthesis of umbelliferone in plants. Plant J. 70, 460–470.

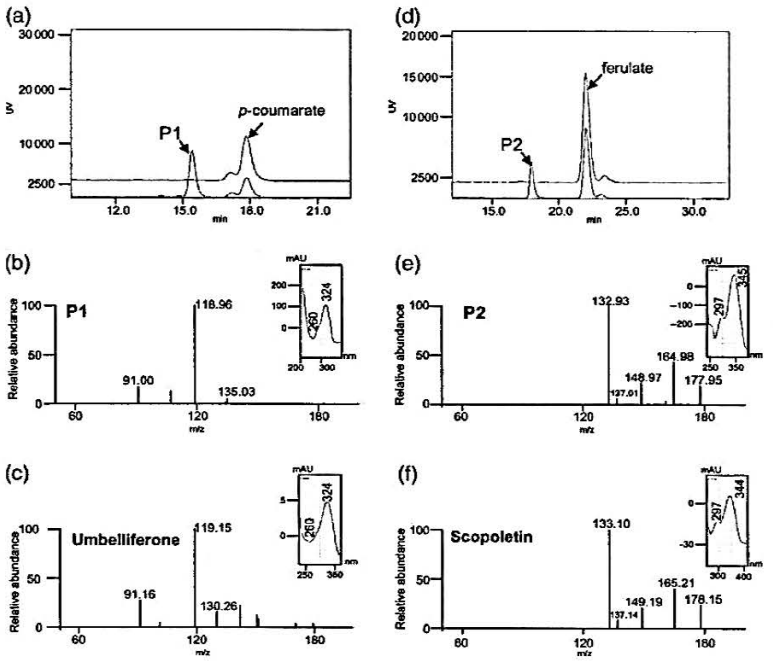


Figure 4. *In vitro* functional characterization of the enzyme encoded by *RgJF799117*.
(a,d) HPLC analysis of *p*-coumaroyl CoA (a) and feruloyl CoA (d) metabolites P1 and P2, respectively.
(b,e) UV and MS spectra of the products P1 (b) and P2 (e).
(c,f) UV and MS spectra analysis of authentic umbelliferone (c) and scopoletin (f) standards.

Table 1 Kinetic constants for Rg 2-OGD heterologously expressed in *Escherichia coli*

	Feruloyl CoA	*p*-coumaroyl CoA
Kinetic constants		
V_{max} (sec^{-1})	0.46 ± 0.05	0.71 ± 0.06
K_m (μM)	37 ± 13	50 ± 10
K_{cat}/K_m (μmol min^{-1})	0.75	0.85
Inhibition with 50 μM psoralen		
$V_{max\ app}$ (sec^{-1})	0.47 ± 0.05	0.69 ± 0.08
$K_{m\ app}$ (μM)	64 ± 14	77 ± 9
Inhibition with 100 μM psoralen		
$V_{max\ app}$ (sec^{-1})	0.45 ± 0.05	0.69 ± 0.08
$K_{m\ app}$ (μM)	73 ± 14	96 ± 9
Inhibition with 200 μM psoralen		
$V_{max\ app}$ (sec^{-1})	0.46 ± 0.06	0.65 ± 0.08
$K_{m\ app}$ (μM)	107 ± 17	108 ± 9

contents compared to untreated plants, supporting the hypothesis that a 2-OGD is involved in this pathway. *In silico* studies and a screen of a *R. graveolens* cDNA library based on the F6′H sequence revealed a putative 2-OGD gene (GenBank accession number JF799117) that was 59% identical to the *A. thaliana* F6′H sequence. Heterologous expression of *RgJF799117* in *E. coli* enabled an *in vitro* screen for enzymatic activity, and clearly showed that the corresponding enzyme accepts two substrates, feruloyl CoA and *p*-coumaroyl CoA, leading to the formation of scopoletin and umbelliferone, respectively, with almost the same efficiency for each substrate (Table 1). The bi-functionality of this enzyme was also tested in *N. benthamiana* transiently expressing the gene under the control of the CaMV 35S promoter. It is well known that tobacco accumulates scopoletin and scopolin, which is the glycoside derivative of scopoletin, but not umbelliferone (Chong *et al.*, 2002; Costet *et al.*, 2002). Treated plants harvested 3 days post-infection had similarly high levels of umbelliferone and scopoletin, and these compounds were both absent in the inoculated control plants. The accumulation of coumarin products observed in *N. benthamiana* clearly indicates that the enzyme encoded by *RgJF799117* is fully functional in plants, and that it can synthesize scopoletin and umbelliferone. The high level of umbelliferone found in tobacco plants expressing *RgJF799117* is particularly interesting with respect to the role of this enzyme in the synthesis of umbelliferone, because this plant does not normally accumulate umbelliferone. In addition, quantification of *RgJF799117* transcripts in *R. graveolens* demonstrated that expression of this gene

© 2011 The Authors

Annexe 1 : Vialart, G., Hehn, A., Olry, A., Ito, K., Krieger, C., Larbat, R., Paris, C., Shimizu, B., Sugimoto, Y., Mizutani, M., *et al.* (2012). A 2-oxoglutarate-dependent dioxygenase from *Ruta graveolens* L. exhibits p-coumaroyl CoA 2′-hydroxylase activity (C2′H): a missing step in the synthesis of umbelliferone in plants. Plant J. 70, 460–470.

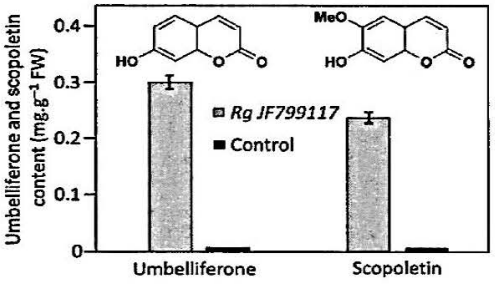


Figure 5. Quantification of umbelliferone and scopoletin in the leaves of *Nicotiana benthamiana* plants transiently expressing *RgJF799117*

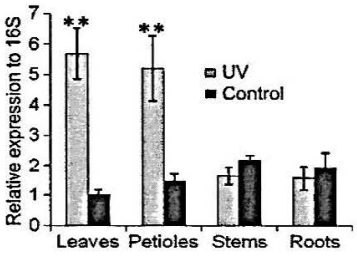


Figure 6. UV-B effects on the *RgJF799117* expression level in *Ruta graveolens* vegetative tissues.
Error bars represent the SE for three replicates. The statistical significance was assessed by Student's *t* test. Asterisks indicate a statistically significant difference ($P < 0.01$) compared to the control.

is positively correlated with the accumulation of umbelliferone in UV-exposed leaves and petioles of this plant. Taken together, the *in planta* results confirm the bi-functional activity of the enzyme. Therefore, *RgJF799117* encodes a 2-OGD that displays both F6′H and C2′H activities in *R. graveolens*. Because the umbelliferone produced by this 2-OGD is derived from *p*-coumaroyl CoA rather than free *p*-coumaric acid, this enzyme probably operates closely with a 4-coumarate CoA-ligase (4CL) that is responsible for synthesis of the *p*-coumaroyl CoA ester. This theory is consistent with other studies in *R. graveolens*, in which two different 4CLs were described (Endler *et al.*, 2008). One of the two *R. graveolens* 4CLs, which is specifically induced by elicitation of the cells, has been suggested to be specifically engaged in coumarin biosynthesis (Endler *et al.*, 2008), and may operate in tandem with the C2′H described above.

Previously, Lukacin *et al.* (2003) also identified a bi-functional 2-OGD that is involved in citrus flavonoid metabolism (Lukacin *et al.*, 2003). Multi-functional enzymes are essential elements that explain the complexity of the metabolic grids found in plants (Schwab, 2003). In addition to F6′H, the presence of C2′H activity in *R. graveolens* allows the generation of umbelliferone and more complex 7-hydroxycoumarin derivatives, such as furanocoumarins, which have important defense functions in plants. These furanocoumarins are absent from *A. thaliana*, which contains only a mono-functional F6′H (Kai *et al.*, 2008). The presence of scopoletin, which is the immediate product of F6′H activity, was investigated, but it was not detected at significant levels in any *R. graveolens* tissues.

The lack of scopoletin in our experiments is puzzling in light of the F6′H activity observed for *RgJF799117* because F6′H was found to be equivalent to C2′H in terms of its activity *in vitro* (heterologous expression in *E. coli*, Table 1) and *in planta* (transient expression in *N. benthamiana*, Figure 5); it would be reasonable to expect co-occurrence of scopoletin and umbelliferone in *R. graveolens*. The formation of associated products downstream of scopoletin may explain the lack of scopoletin itself. However, scopolin (the glycosidic end-product that is synthesized from scopoletin) was also present in only trace amounts in *R. graveolens*, contradicting the above hypothesis. Other phytochemical studies have confirmed the low level of scopoletin in *R. graveolens*: scopoletin has frequently been described to be a minor coumarin in *R. graveolens* (Steck *et al.*, 1971; Paszkiewicz *et al.*, 2008), and a recent report revealed that its concentration is below the HPLC-UV quantification limit (<2.5 μg g^{-1} FW) (Adam *et al.*, 2009). Although the F6′H activity of the protein that is encoded by *RgJF799117* seems to be fully functional, the reason for the absence of significant levels of scopoletin in *R. graveolens* is currently unresolved, and will require further investigation.

Phylogenetic analyses facilitated the construction of a tree in which 2-OGDs are organized according to their functional diversity (Figure 3). To product of *RgJF99117* shares the highest identity with a peptide sequence isolated from *Citrus limetta* (NCBI accession number JN245615). For an unknown reason, functional characterization of this gene could not be achieved using the *E. coli* heterologous expression system. However, the fact that this citrus gene is the closest neighbor of *RgJF99117* is consistent with the observation that citrus plants synthesize large amounts of furanocoumarins and other umbelliferone derivatives; therefore, it is probably necessary for these plants to possess a C2′H function similar to that in *R. graveolens*. Our sequence corresponding to a C2′H and the two *A. thaliana* F6′H sequences are grouped together and constitute a separate clade.

Kinetic experiments with various coumarins and furanocoumarins revealed that psoralen is a competitive inhibitor of the C2′H/F6′H encoded by *RgJF99117* (Table 1). Furanocoumarins are natural compounds with potent allochemical functions because of their severe cellular toxicity. The severe genotoxicity of linear furanocoumarins (i.e. psoralen derivatives) is a consequence of their ability to intercalate with double-stranded DNA and create covalent

© 2011 The Authors

Annexe 1 : Vialart, G., Hehn, A., Olry, A., Ito, K., Krieger, C., Larbat, R., Paris, C., Shimizu, B., Sugimoto, Y., Mizutani, M., *et al.* (2012). A 2-oxoglutarate-dependent dioxygenase from *Ruta graveolens* L. exhibits p-coumaroyl CoA 2′-hydroxylase activity (C2′H): a missing step in the synthesis of umbelliferone in plants. Plant J. 70, 460–470.

cross-links between thymine residues (Dall'Acqua *et al.*, 1978). Linear furanocoumarins are therefore potentially toxic to any living organism, including plants that synthesize large quantities of these compounds, such as *R. graveolens*. The K_i that was experimentally determined for psoralen using the *R. graveolens* C2′H/F6′H enzyme was approximately twice the K_m value (Table 1). Although the enzyme affinity is less pronounced for psoralen than for its two substrates (feruloyl and *p*-coumaroyl CoA), psoralen may exert a feedback inhibition on C2′H at higher concentrations, and this may down-regulate the general umbelliferone/furanocoumarin pathway. In *R. graveolens*, this down-regulation pattern is consistent with reports showing constitutive synthesis of furanocoumarins in all the tissues from this plant (Eilert, 1989; Milesi *et al.*, 2001) and the UV elicitation experiments reported here, in which furanocoumarin concentrations in vegetative organs did not significantly increase. In contrast to most plants that accumulate furanocoumarins and exhibit a highly inducible biosynthetic pathway (Hamerski and Matern, 1988) based on differentially expressed genes (Larbat *et al.*, 2007), *R. graveolens* may have developed an alternative regulation system, such as enzyme retro-inhibition.

Finally, two independent studies from the 1960s and 1970s hypothesized that the *ortho*-hydroxylation of *p*-coumaric acid leading to formation of umbelliferone was controlled by a P450-dependent mono-oxygenase. Here, we provide evidence that a 2-OGD is committed to this synthesis in *R. graveolens*, and describe the first *p*-coumaroyl CoA 2′-hydroxylase to have been characterized in plants. In addition to the formation of umbelliferone, this enzyme converts feruloyl CoA into scopoletin (F6′H activity) with equivalent efficiency, and the enzyme is therefore fully bi-functional. The C2′H and F6′H activities are equally repressed by psoralen, which is a competitive inhibitor located downstream of the umbelliferone pathway. Psoralen down-regulates the C2′H function, subsequently reducing the metabolic flux through the umbelliferone/furanocoumarin pathway and preserving the plant tissues from excessive psoralen concentrations. Some higher plants display enzymatic function that may be controlled by a P450 or a 2-OGD enzyme. In Apiaceae (Martens *et al.*, 2001) and *Equisetum arvense* (Bredebach *et al.*, 2011), [2*S*]-naringenin is converted into apigenin by a flavone synthase I that belongs to the oxoglutarate-dependent dioxygenase family. However, in Fabaceae species such as *Glycyrrhiza echinita* (Akashi *et al.*, 1998) and soybean (*Glycine max*) (Fliegmann *et al.*, 2010), this flavone formation is performed by enzymes of the CYP93B P450 family. Further investigations are required to determine whether C2′H is encoded by the 2-oxaloglutarate-dependent dioxygenase superfamily alone, or whether other mono-oxygenase enzymes (such as P450s) are also involved, as previously assumed.

EXPERIMENTAL PROCEDURES

Plant material

Ruta graveolens and *Nicotiana benthamiana* seeds were sown and germinated in a greenhouse on potting soil. The seedlings were grown for 1–2 months before being used in the experiments.

UV-B light elicitation of the plants was achieved as follows: 15 cm tall *R. graveolens* plants were cultured under greenhouse conditions, and were irradiated for 8 h in the dark using a monochromatic lamp (312 nm) that was placed 25 cm from the plants. The plants were rotated every 2 h within the irradiation area to minimize any positional effects prior to RNA extraction.

In vivo inhibition of 2-oxoglutarate-dependent dioxygenases

Prohexadione-calcium (Regalis; BASF, http://www.basf.fr) is a non-specific inhibitor of 2-oxoglutarate-dependent dioxygenases (2-OGD) (Puhl *et al.*, 2008). Prohexadione-calcium was diluted with ultrapure water to a final concentration of 500 ppm, and was administered to the aerial parts of the plant by spraying. Two treatments were administered with a 6 h interval. The control plants were sprayed with ultrapure water. Furanocoumarins were extracted from leaves 4 days post-inoculation, and analyzed by HPLC-DAD and LC-MS. Each treatment was performed on six plants.

Substrate synthesis

Enzymatic synthesis of the CoA substrates cinnamoyl CoA, caffeoyl CoA, feruloyl CoA and *p*-coumaroyl CoA was performed using *A. thaliana* 4-coumarate CoA ligase (4CL) produced in *E. coli* M15 transformed with the pQE30/4CL plasmid (Lee and Douglas, 1996). LB medium (Bacto-Tryptone, 10 g/l; Bacto-yeast extract, 5 g/l; NaCl, 10 g/l) (100 ml) was supplemented with ampicillin (100 mg L^{-1}), and then inoculated at 37°C with transformed bacteria until the OD_{600} reached 0.7. Induction of 4CL synthesis was achieved by adding isopropyl-β-D-thiogalactopyranoside to a final concentration of 1 mM for 20 h at 30°C. 4CL was purified using the Ni-NTA purification system (Qiagen, http://www.qiagen.com/) according to the manufacturer's instructions.

The CoA esters were synthesized in a final volume of 1 ml at room temperature from the corresponding acids (0.2 mM): ferulic acid, *p*-coumaric acid, caffeic acid and cinnamic acid (Sigma, http://www.sigmaaldrich.com/). The reaction mixture contained 0.2 mM coenzyme A, 2.5 mM ATP, 1 mM DTT, 2.5 mM $MgCl_2$ and purified enzyme (20 mg protein ml^{-1}). The kinetics of the chemical reaction were followed spectrophotometrically at 346 nm for feruloyl CoA and caffeoyl CoA, at 333 nm for *p*-coumaroyl CoA, and at 311 nm for cinnamoyl CoA.

cDNA library screening

Total RNA was extracted from *R. graveolens* leaves that had been exposed to UV light for 24 h using an RNeasy plant mini kit (Qiagen). The total RNA (600 ng) was reverse-transcribed and used to construct a Lambda TriplEx2 cDNA library using a SMART cDNA library construction kit (Clontech, http://www.clontech.com/) according to the manufacturer's instructions.

The protein sequence of *A. thaliana* F6′H identified by Kai *et al.* (2008) was used to perform a tBLASTn search on the Expressed Sequence Tags Databases available in GenBank. This *in silico* screen permitted the reconstruction of a citrus gene that had a high degree of homology with *A. thaliana* F6′H. Primers were designed to amplify this gene from RNA extracted from *Citrus limetta*

© 2011 The Authors
The Plant Journal © 2011 Blackwell Publishing Ltd,

Annexe 1 : Vialart, G., Hehn, A., Olry, A., Ito, K., Krieger, C., Larbat, R., Paris, C., Shimizu, B., Sugimoto, Y., Mizutani, M., *et al*. (2012). A 2-oxoglutarate-dependent dioxygenase from *Ruta graveolens* L. exhibits p-coumaroyl CoA 2′-hydroxylase activity (C2′H): a missing step in the synthesis of umbelliferone in plants. Plant J. 70, 460–470.

(CitrusFwd, 5′-CCATATGGCACCAACAAATGGCT-3′; CitrusRev, 5′-CGGATCCTCACCTATCTTTGCAAAGCT-3′). Additional *Nde*I and *Bam*HI restriction sites (underlined) were added at the 5′ and 3′ ends to facilitate subsequent cloning into an expression plasmid. Total RNA was extracted from *Citrus limetta* leaves using an RNeasy plant extraction kit (Qiagen) according to the manufacturer's instructions. The gene was amplified using SuperScript™ One-Step RT-PCR with Platinum® *Taq* polymerase (Invitrogen, http://www.invitrogen.com/). The reverse transcription step was performed at 45°C for 30 min. The PCR amplification was performed with an initial step of 94°C for 5 min, five cycles of 94°C for 15 sec, 40°C for 30 sec and 72°C for 2 min, then 25 cycles of 94°C for 15 sec, 52°C for 30 sec and 72°C for 2 min, and a final step at 72°C for 10 min. The resulting amplified product was cloned into the pCR8® plasmid (Invitrogen) and sequenced (accession number JN245615). This gene was then used to synthesize a probe using an AlkPhos Direct™ labeling and detection systems labeling kit (GE Healthcare, http://www.gelifesciences.com). Library screening was performed using the cDNA library from *R. graveolens*. The bacteriophages were inoculated into *E. coli* XL1 Blue, and spread on an LB plate that contained 10 mM $MgSO_4$ to generate 200 000 lysis plaques (150 mm Petri dishes). The bacteriophages were blotted onto Amersham Hybond™ membrane, denatured for 3 min in 0.5 M NaOH and 1.5 M NaCl, neutralized for 3 min in 10 mM Tris/HCl and 1.5 M NaCl, and washed in 2× SSC for 3 min. Finally, the phage DNA was linked to the membrane using a BIO-LINK BLX254 crosslinker (Fisher Scientific - http://www.fr.fishersci.com) at 254 nm and 0.12 J for 30 sec. The probe was hybridized at 52°C overnight, and two washes with 2x SSC for 10 min were performed at 55°C. Detection was achieved by chemiluminescence using CDP Star (Amersham). The plasmids were extracted from lysis plates that were identified as positive as described in the protocol for the SMART cDNA library construction kit using a Cre/Lox excision system in *E. coli* BM25.8. The isolated gene was sequenced (GenBank accession number *JF799117*).

Bacterial heterologous expression of Rg 2-OGD

The coding sequence for the *R. graveolens RgJF799117* gene was amplified from the pTriplEx plasmid using PCR with primers Rg2OGDFor (5′-CGGATCCATGGCACCGACAAAAGA-3′) and Rg2OGDRev (5′-CCTCGAGTTAATTCATCAAAGCGAAGTCGACAGTTTTC-3′). An additional *Eco*RI site at the 5′ end and an additional *Xho*I site at the 3′ end (underlined) were added to facilitate sub-cloning into the bacterial expression vector. PCR amplification was performed with an initial step of 94°C for 5 min, five cycles of 94°C for 15 sec, 40°C for 30 sec and 72°C for 2 min, then 25 cycles of 94°C for 15 sec, 52°C for 30 sec and 72°C for 2 min, and a final step at 72°C for 10 min. The gene was cloned into pET28a plasmid (Clontech) to produce a gene that allowed expression of a fusion protein with a C-terminal histidine tag. *JN245615* was sub-cloned from the pCR8 plasmid and introduced into the pET3a plasmid using the *Nde*I and *Bam*HI restriction sites that had been added to the 5′ and 3′ ends during the RT-PCR step. The recombinant plasmids were finally introduced into the *E. coli* BL21 (DE3) strain.

To produce the recombinant proteins, the bacteria were cultured in 100 ml LB medium supplemented with kanamycin (50 μg ml^{-1}) until the OD_{600} reached 0.7. Protein synthesis was then induced by adding isopropyl-thio-β-D-galactopyranoside to a concentration of 1 mM for 20 h at 26°C. The culture was then centrifuged at 5000 *g* for 10 min at 4°C. The cell pellets were resuspended in 5 ml cold PBS, sonicated using a Bandelin Sonopuls HD 2070 ultrasonic homogenizer (Sigma) type MS 73 at a sound intensity of 200 W cm^{-2}, three times for 30 sec each on ice, and centrifuged at 16 100 *g* for 10 min at 4°C. The His-tagged RgJF799117 proteins present in the supernatant were purified using an Ni-NTA purification system kit (Invitrogen) according to the manufacturer's instructions. After three column washes, the adsorbed proteins were eluted in 300 μl of a solution containing 500 mM imidazole, 300 mM NaCl and 50 mM NaH_2PO_4 at pH 8. The concentration of the purified proteins was determined by fluorometry using the Invitrogen Qubit system. The purified recombinant proteins were visualized by SDS–PAGE. The proteins were revealed by staining the gel using Coomassie brilliant blue R-250. The proteins were stored at −80°C for further analyses.

Enzymatic assays and HPLC analysis

The enzyme assay reaction contained 1 μg of protein, 0.1 mM Tris/HCl pH 7.2, 0.2 mM $FeSO_4$, 5 mM α-ketoglutarate, 10 mM sodium ascorbate and substrate at concentrations from 2 to 200 μM to determinate the kinetic parameters. For the inhibition assays, three concentrations of psoralen (50, 100 and 200 μM) were added to this enzymatic assay reaction. The kinetic and inhibition assays were performed in triplicate. The reaction was performed for 3 min at 30°C, stopped by addition of 20 μl of 3 mM NaOH, before neutralization with 10 μl glacial acetic acid. This treatment resulted in lysis of the CoA esters and therefore did not permit visualization of these esters; only the corresponding acids could be visualized. The enzymatic mixture was centrifuged for 10 min at 16 000 *g*, and the supernatant was analyzed by HPLC-DAD on a Cosmosil 5C18-AR-II column (Nacalai Tesque Inc., http://www.nacalai.co.jp) on a linear gradient of 10–70% methanol for 35 min at 1 ml min^{-1}. The UV spectra and retention times of the products formed were compared to commercial standards. The products formed were also analyzed by mass spectrometry. The kinetic parameters were calculated using the SIGMAPLOT software program (Systat Software Inc., http://www.sigmaplot.com/)

Heterologous expression in *N. benthamiana*

The pBIN-GW plasmid was obtained using pBIN-mGFP5ER (Haseloff *et al.*, 1997) in which the *mGFP5ER* gene was replaced by a Multi Cloning Site linker that contained *Bam*HI, *Sma*I, *Eco*RI, *Kpn*I and *Eco*RI sites, producing the plasmid pBIN-Linker. To obtain pBIN-GW, the pBIN-Linker was digested with *Bam*HI, and an RfB cassette was introduced to generate a Gateway destination vector as described by the manufacturer (Invitrogen). The gene encoding Rg 2-OGD was first cloned into the pCR8®/GW/TOPO® TA plasmid (Invitrogen, http://www.invitrogen.com). It was then transferred into the destination plasmid pBIN-GW by LR recombination using Gateway® technology according to the manufacturer's instructions, to produce pBIN-Rg 2-OGD.

The three recombinant plasmids pBIN-Rg 2-OGD, pBIN-GW and pBIN-mGFP5ER (Haseloff *et al.*, 1997) were introduced into *Agrobacterium* strain LBA4404 to produce three *Agrobacterium* strains. In parallel, *Agrobacterium* strain C5851 containing pBIN61-P19 was generated (Voinnet *et al.*, 2003). All of these strains were individually grown in YEB culture medium (5.0 g sucrose. 5.0 g peptone. 5.0 g beef extract. 1.0 g yeast extract. 0.049 g MgSO4 . 7H20. pH to 7.2) for 24 h at 30°C. The bacterial cultures were centrifuged for 5 min at 5000 g, and the pellets were washed three times with sterile distilled water. Finally, the bacterial pellet was resuspended in water to obtain an OD_{600} between 0.3 and 0.4. To avoid gene silencing during the transient expression experiment, each LBA4404 strain (containing the various plasmids) was infiltrated together with the C5851 pBIN61-P19 strain in *N. benthamiana* leaves. Leaves transformed with the three plasmid combinations were harvested 3 days post-inoculation. The leaves of the GFP-transformed plant were used as a positive control to ascertain the agroinfection efficiency, and were observed under UV light. The leaves from

© 2011 The Authors

Annexe 1 : Vialart, G., Hehn, A., Olry, A., Ito, K., Krieger, C., Larbat, R., Paris, C., Shimizu, B., Sugimoto, Y., Mizutani, M., *et al.* (2012). A 2-oxoglutarate-dependent dioxygenase from *Ruta graveolens* L. exhibits p-coumaroyl CoA 2′-hydroxylase activity (C2′H): a missing step in the synthesis of umbelliferone in plants. Plant J. 70, 460–470.

plants transformed with the empty pBIN-GW plasmid (used as a negative control) or pBIN-Rg 2-OGD were analyzed for their coumarin/furanocoumarin content. All the infiltrations were reproduced in triplicate.

Coumarin extraction and analysis

The leaves, petioles, stems and roots of *R. graveolens* seedlings and the leaves of *N. benthamiana* were harvested, frozen in liquid nitrogen and crushed immediately. Fresh material (200 mg) was collected for each organ. Extraction of the compounds was performed in 1.5 ml of 80% ethanol, followed by an initial grinding step in liquid nitrogen and a subsequent grinding step using a benchtop homogenizer Polytron PT2100 (Kinematica, http://www.kinematica.ch) for 1 min. After centrifugation at 16 000 ***g*** for 10 min, the supernatant was collected and concentrated by overnight ethanol evaporation in an RC10.10 speed vacuum (Jouan, http://www.thermoscientific.com). The pellet was resuspended in 200 µl methanol and analyzed by HPLC-DAD (Shimadzu, http://www.shimadzu.fr) on a Cosmosil 5C18-AR-II column (Nacalai Tesque Inc., http://www.nacalai.co.jp). The mobile phase consisted of a binary solvent system of water (solvent A) and methanol (solvent B), each containing 0.1% v/v formic acid. The gradient (0–5 min, 10% solvent B; 5–45 min, 20–60% solvent B; 45–75 min, 60–75% solvent B; 75–76 min, 75–100% solvent B) was eluted at a flow rate of 0.7 ml min^{-1}. Scopoletin was detected at 346 nm and umbelliferone was detected at 330 nm. To determine statistical significance, Student's *t* test was performed on the data using the MS Office Excel software program (Microsoft, http://www.microsoft.com/).

Real-time PCR

Total RNA was extracted from the leaves, roots, stems and petioles of *R. graveolens*. Extractions were performed on three independent plants. The extraction was performed using the RNeasy plant mini kit (Qiagen) according to the manufacturer's instructions. High-capacity RNA-to-cDNA master mix (Applied Biosystems, http://www.appliedbiosystems.com/) was used for first-strand cDNA synthesis with oligo(dT) as the primer. The resulting cDNAs were diluted 100-fold before use for real-time PCR under the conditions described below. Ribosomal 16S RNA from *R. graveolens* was used as an internal control for data normalization (forward primer, 5′-CATTCGGCCCGTCTTGAA-3′; reverse primer, 5′-CCGTTGACTCGCACACATGT-3′). Real-time PCR was performed using the primer pair qPCRD4Dir (5′-CCACCTCTCAGAGCGACTATCA-3′) and qPCRD4Rev (5′-CCCATGTGAATGACGGAGTCT-3′) designed using the Primer Express Software program (Applied Biosystems, http://www.appliedbiosystems.com). The efficiency of the probes was checked using a standard linear curve method. The relative transcript level of the same gene in the various organs was expressed as the $\log_2$ ratio of the 16S normalized transcript levels in a given organ relative to the mean expression of the gene in organs tested. Student's *t* test was performed on the data using MS Office Excel.

MS analysis

The HPLC-MS system comprised a binary solvent delivery pump and a linear ion trap mass spectrometer (LTQ-MS, Thermo Scientific, http://www.thermoscientific.com). The LTQ linear ion trap was equipped with an atmospheric pressure ionization interface and operated in ESI positive ion mode. The data were processed using the XCALIBUR software program (version 2.1, http://www.thermoscientific.com). For the mass spectrometer, the spray voltage was 5 kV, and the temperature of the heated capillary was set at 300°C. The flow rates for sheath gas, auxiliary gas and sweep gas were set to 40, 10 and 10 arbitrary units min^{-1}, respectively. The capillary voltage was 36 V, the split lens was −44 V, and the front lens was −3.25 V. All of the parameters were optimized by infusing a standard solution of isopimpinellin (0.1 g l^{-1}) in the mobile phase [water + 0.1% acetic acid/methanol + 0.1% acetic acid (ratio 90/10)] at a flow rate of 5 ml min^{-1}. The coumarins and furanocoumarins of interest were monitored by specific MS/MS scans in addition to a full scan (50–2000 *m/z*): MS^2(163) for umbelliferone and MS^2(193) for scopoletin.

Phylogenetic analysis

Multiple sequence alignment of the amino acid sequences of the 2-OGD family members was performed at the Bioinformatics Center, Institute for Chemical Research, Kyoto University, Japan (http://www.genome.jp/) using the ClustalW program (http://www.genome.jp/tools/clustalw/). *A. thaliana* AOP1 was included in the alignment as an outgroup. Based on the alignment, a molecular phylogenetic tree was constructed using the neighbor-joining method. The statistical significance of the neighbor-joining tree topology was evaluated using bootstrap analysis with 1000 iterative tree constructions. The tree was visualized using the TREE VIEW software program (http://taxonomy.zoology.gla.ac.uk/rod/treeview.html).

ACKNOWLEDGEMENTS

The experiments performed in Japan by G.V. were funded by INRA/JSPS 2009–2011. The authors are grateful to Professor Jim Haselhoff (Cambridge, UK) who provided the pBIN-mGFP5ER vector, Professor David Baulcombe (Cambridge, UK) who provided pBIN61-P19, and to Professor Carl Douglas (Vancouver, Canada) who provided the pQE30/4CL plasmid. The authors would like to acknowledge the technical help of L. Calvier, A. Courteille, N. Delacroix, L. Eliasson, A. Kutrowska, E. Martins and M. Lourdin for their assistance in the Furoc2008 project (ENSAIA).

REFERENCES

Adam, M., Dobias, P., Eisner, A. and Ventura, K. (2009) Extraction of antioxidants from plants using ultrasonic methods and their antioxidant capacity. *J. Sep. Sci.* **32**(2), 288–294.

Akashi, T., Aoki, T. and Ayabe, S.I. (1998) Identification of a cytochrome P450 cDNA encoding (2*S*)-flavanone 2-hydroxylase of licorice (*Glycyrrhiza echinata* L.; Fabaceae) which represents licodione synthase and flavone synthase II. *FEBS Lett.* **431**(2), 287–290.

Anzelotti, D. and Ibrahim, R.K. (2004) Molecular characterization and functional expression of flavonol 6-hydroxylase. *BMC Plant Biol.* **4**, 20.

Beier, R.C. and Oertli, E.H. (1983) Psoralen and other linear furocoumarins as phytoalexins in celery. *Phytochemistry*, **22**(11), 2595–2597.

Bourgaud, F., Hehn, A., Larbat, R., Doerper, S., Gontier, E., Kellner, S. and Matern, U. (2006) Biosynthesis of coumarins in plants: a major pathway still to be unravelled for cytochrome P450 enzymes. *Phytochem. Rev.* **5**(2–3), 293–308.

Bredebach, M., Matern, U. and Martens, S. (2011) Three 2-oxoglutarate-dependent dioxygenase activities of *Equisetum arvense* L. forming flavone and flavonol from (2*S*)-naringenin. *Phytochemistry*, **72**(7), 557–563.

Brown, S.A. (1962) Biosynthesis of coumarin and herniarin in lavender. *Science*, **137**, 977–978.

Brown, S.A., Towers, G.H. and Wright, D. (1960) Biosynthesis of the coumarins. Tracer studies on coumarin formation in *Hierochloe odorata* and *Melilotus officinalis*. *Can. J. Biochem. Physiol.* **38**, 143–156.

Chong, J., Baltz, R., Schmitt, C., Beffa, R., Fritig, B. and Saindrenan, P. (2002) Downregulation of a pathogen-responsive tobacco UDP-Glc: phenylpropanoid glucosyltransferase reduces scopoletin glucoside accumulation, enhances oxidative stress, and weakens virus resistance. *Plant Cell*, **14**(5), 1093–1107.

Costet, L., Fritig, B. and Kauffmann, S. (2002) Scopoletin expression in elicitor-treated and tobacco mosaic virus-infected tobacco plants. *Physiol. Plant.* **115**(2), 228–235.

© 2011 The Authors

Annexe 1 : Vialart, G., Hehn, A., Olry, A., Ito, K., Krieger, C., Larbat, R., Paris, C., Shimizu, B., Sugimoto, Y., Mizutani, M., *et al*. (2012). A 2-oxoglutarate-dependent dioxygenase from *Ruta graveolens* L. exhibits p-coumaroyl CoA 2′-hydroxylase activity (C2′H): a missing step in the synthesis of umbelliferone in plants. Plant J. 70, 460–470.

Dall'Acqua, F., Vedaldi, D. and Recher, M. (1978) The photoreaction between furocoumarins and various DNA with different base compositions. *Photochem. Photobiol.* **27**, 33–36.

Eilert, U. (1989) Elicitor induction of secondary metabolism in dedifferentiated *in vitro* system of *Ruta graveolens*. In *Primary and Secondary Metabolism of Plant Cell Cultures* (Kurz, W.G.W., ed). Berlin: Springer, pp. 219–228.

Ellis, B.E. and Amrhein, N. (1971) The 'NIH-shift' during aromatic *ortho*-hydroxylation in higher plants. *Phytochemistry*, **10**, 3069–3072.

Endler, A., Martens, S., Wellmann, F. and Matern, U. (2008) Unusually divergent 4-coumarate: CoA-ligases from *Ruta graveolens* L. *Plant Mol. Biol.* **67**(4), 335–346.

Fliegmann, J., Furtwangler, K., Malterer, G., Cantarello, C., Schuler, G., Ebel, J. and Mithofer, A. (2010) Flavone synthase II (CYP93B16) from soybean (*Glycine max* L.). *Phytochemistry*, **71**(5–6), 508–514.

Frey, M., Huber, K., Park, W.J., Sicker, D., Lindberg, P., Meeley, R.B., Simmons, C.R., Yalpani, N. and Gierl, A. (2003) A 2-oxoglutarate-dependent dioxygenase is integrated in DIMBOA-biosynthesis. *Phytochemistry*, **62**(3), 371–376.

Gestetner, B. and Conn, E.E. (1974) 2-hydroxylation of *trans*-cinnamic acid by chloroplasts from *Melilotus alba*. *Arch. Biochem. Biophys.* **163**, 617–624.

Gutierrez, M.-C., Parry, A., Tena, M., Jorrin, J. and Edwards, R. (1995) Abiotic elicitation of coumarin phytoalexins in sunflower. *Phytochemistry*, **38**(5), 1185–1191.

Hagel, J.M. and Facchini, P.J. (2010) Dioxygenases catalyse the *O*-demethylation steps of morphine biosynthesis in opium poppy. *Nat. Chem. Biol.* **6**(4), 273–275.

Hamerski, D. and Matern, U. (1988) Elicitor-induced biosynthesis of psoralens in *Ammi majus* L. suspension cultures. Microsomal conversion of demethylsuberosin into (+)marmesin and psoralen. *Eur. J. Biochem.* **171**(1–2), 369–375.

Haseloff, J., Siemering, K.R., Prasher, D.C. and Hodge, S. (1997) Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic Arabidopsis plants brightly. *Proc. Natl Acad. Sci. USA*, **94**(6), 2122–2127.

Kai, K., Shimizu, B.-I., Mizutani, M., Watanabe, K. and Sakata, K. (2006) Accumulation of coumarins in *Arabidopsis thaliana*. *Phytochemistry*, **67**(4), 379–386.

Kai, K., Mizutani, M., Kawamura, N., Yamamoto, R., Tamai, M., Yamaguchi, H., Sakata, K. and Shimizu, B.I. (2008) Scopoletin is biosynthesized via *ortho*-hydroxylation of feruloyl CoA by a 2-oxoglutarate-dependent dioxygenase in *Arabidopsis thaliana*. *Plant J.* **55**(6), 989–999.

Kùc, J. (1982) Phytoalexins from the Solanaceae. In *Phytoalexins* (Bailey, J.A. and Mansfield, J.W., eds). Glasgow: Blackie & Sons, pp. 81–105.

Larbat, R., Kellner, S., Specker, S., Hehn, A., Gontier, E., Hans, J., Bourgaud, F. and Matern, U. (2007) Molecular cloning and functional characterization of psoralen synthase, the first committed monooxygenase of furanocoumarin biosynthesis. *J. Biol. Chem.* **282**(1), 542–554.

Lee and Douglas (1996) Two divergent members of a tobacco 4-coumarate: Coenzyme A ligase (4CL) gene family. *Plant Physiol.* **112**, 193–205.

Lukacin, R., Wellmanna, F., Britsch, L., Martens, S. and Matern, U. (2003) Flavonol synthase from *Citrus unshiu* is a bifunctional dioxygenase. *Phytochemistry*, **62**(3), 287–292.

Martens, S., Forkmann, G., Matern, U. and Lukacin, R. (2001) Cloning of parsley flavone synthase I. *Phytochemistry*, **58**(1), 43–46.

Matsuda, J., Okabe, S., Hashimoto, T. and Yamada, Y. (1991) Molecular cloning of hyoscyamine 6β-hydroxylase, a 2-oxoglutarate-dependent dioxygenase, from cultured roots of *Hyoscyamus niger*. *J. Biol. Chem.* **266**(15), 9460–9464.

Milesi, S., Massot, B., Gontier, E., Bourgaud, F. and Guckert, A. (2001) *Ruta graveolens* L.: a promising species for the production of furocoumarins. *Plant Sci.* **161**, 189–199.

Paszkiewicz, M., Orlita, A., Dziabas, A., Golebiowski, M., Lojkowska, E., Szafranek, J., Malinski, E. and Stepnowski, P. (2008) Simplex optimized LC analysis of plant coumarins and furanocoumarins. *Chromatographia*, **67** (7–8), 653–657.

Puhl, I., Stadler, F. and Treutter, D. (2008) Alterations of flavonoid biosynthesis in young grapevine (*Vitis vinifera* L.) leaves, flowers, and berries induced by the dioxygenase inhibitor prohexadione-Ca. *J. Agric. Food. Chem.* **56**(7), 2498–2504.

Schuler, M.A. and Berenbaum, M.R. (2003) Diversification of furanocoumarin-metabolizing cytochrome P450 monooxygenases in two papilionids: specificity and substrate encounter rate. *Proc. Natl Acad. Sci. USA*, **100**(Suppl. 2), 14593–14598.

Schwab, W. (2003) Metabolome diversity: too few genes, too many metabolites? *Phytochemistry*, **62**(6), 837–849.

Steck, W., Bailey, B.K., Shyluk, J.P. and Gamborg, O.L. (1971) Coumarins and alkaloids from cell cultures of *Ruta graveolens*. *Phytochemistry*, **10**(1), 191–194.

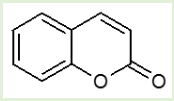
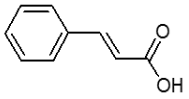
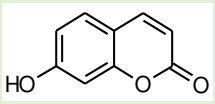
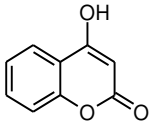
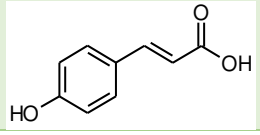
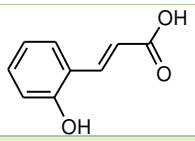
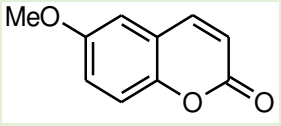
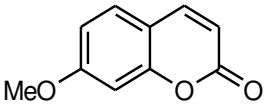
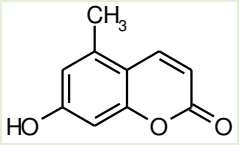
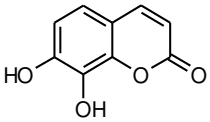
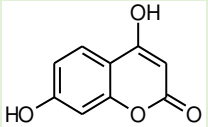
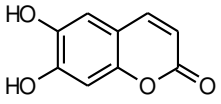
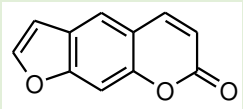
Vazquez-Flota, F., De Carolis, E., Alarco, A.M. and De Luca, V. (1997) Molecular cloning and characterization of desacetoxyvindoline-4-hydroxylase, a 2-oxoglutarate dependent-dioxygenase involved in the biosynthesis of vindoline in *Catharanthus roseus* (L.) G. Don. *Plant Mol. Biol.* **34**(6), 935–948.

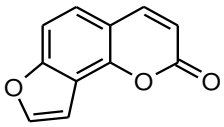
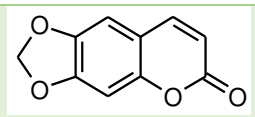
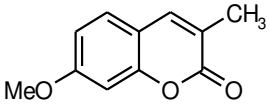
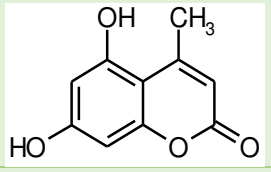
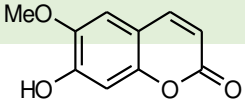
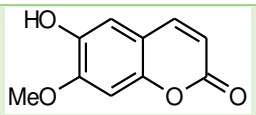
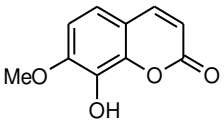
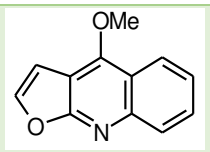
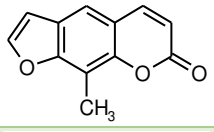
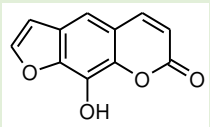
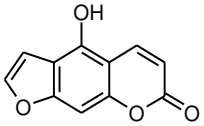
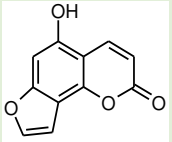
Voinnet, O., Rivas, S., Mestre, P. and Baulcombe, D. (2003) An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *Plant J.* **33**(5), 949–956.

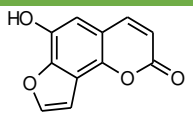
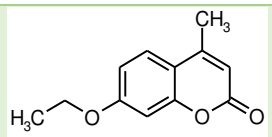
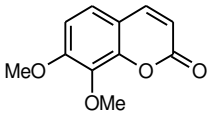
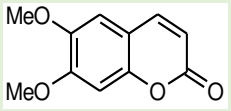
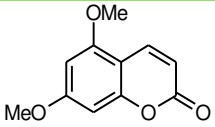
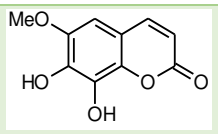
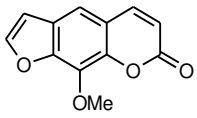
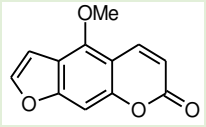
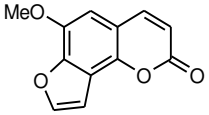
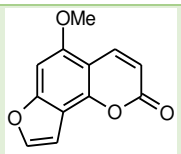
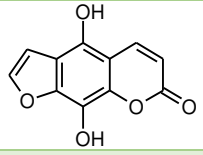
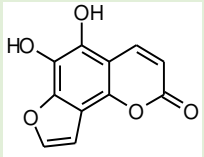
Zobel, A.M. and Brown, S.A. (1993) Influence of low-intensity ultraviolet radiation on extrusion of furanocoumarins to the leaf surface. *J. Chem. Ecol.* **19**(5), 939–952.

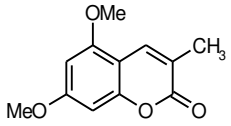
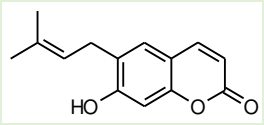
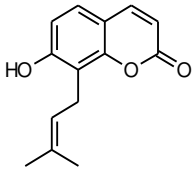
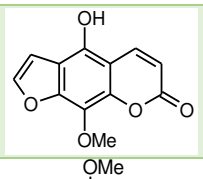
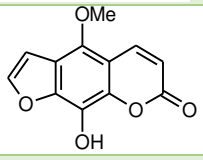
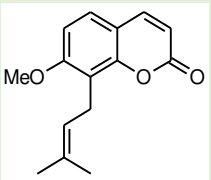
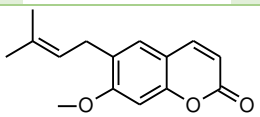
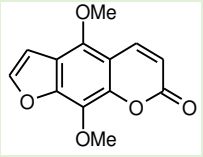
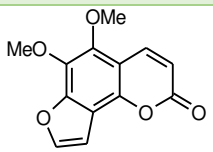
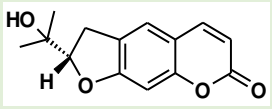
Zobel, A.M., Zalewski, A., Renke, K.W. and Growniak, K. (1998) Influence of UV-radiation on xanthotoxin and bergapten contents in *Ruta graveolens* L. and *Angelica arachangelica* L. seedlings. *Allelopathy J.* **5**(2), 183–189.

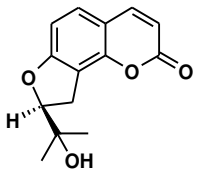
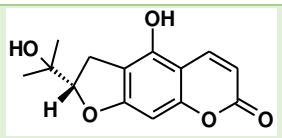
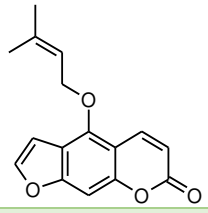
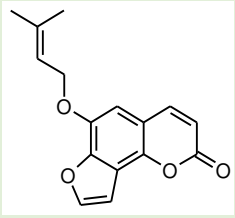
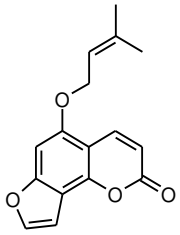
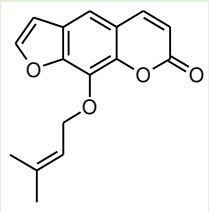
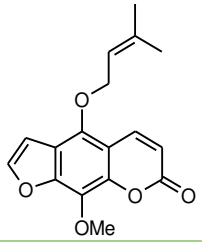
Annexe 2 : Liste des substrats disponibles au laboratoire.

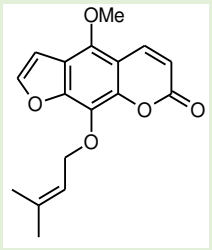
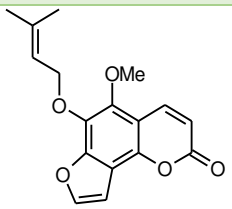
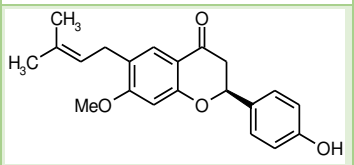
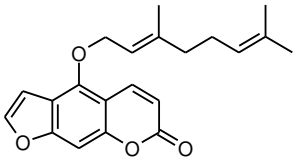
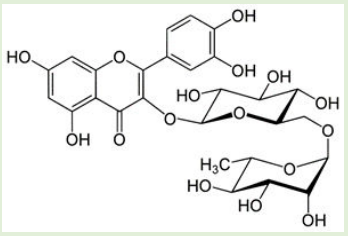
Nom	Structure	Poids moléculaire (g/mol)	Numéro CAS ou InChIKey
Coumarine		146,15	91-64-5
Cinnamate		148,16	140-10-3
Umbelliférone		162,14	93-35-6
4-hydroxycoumarine		162,14	1076-38-6
***p*-coumarate**		164,16	501-98-4
***o*-coumarate**		164,16	614-60-8
6-méthoxycoumarine		176,17	17372-53-1
7-méthoxycoumarine		176,17	531-59-9
5-méthylumbelliférone		176,17	
Daphnétine		178,14	486-35-1
4,7-dihydroxycoumarine		178,14	1983-81-9
Esculétine		178,14	305-01-1
Psoralène		186,16	523-50-2

Nom	Structure	Poids moléculaire (g/mol)	Numéro CAS ou InChIKey
Angélicine		186,16	66-97-7
Ayapine		190,15	494-56-4
7-méthoxy-3-méthylcoumarine		190,20	
5,7-dihydroxy-4-méthylcoumarine		192.17	2107-76-8
Scopolétine		192,17	05-10 3067
Isoscopolétine		192,17	776-86-3
Daphnétine 7-méthyléther		192,17	19492-03-6
Dictamnine		199,20	484-29-7
8-méthylpsoralène		200,19	15912-88-6
Xanthotoxol		202,17	486-60-2
Bergaptol		202,17	2009-24-7
Isobergaptol		202,17	21339-45-7

Nom	Structure	Poids moléculaire (g/mol)	Numéro CAS ou InChIKey
Sphondinol		202,17	61265-07-4
4-méthyl-7-éthoxycoumarine		204,22	87-05-8
Daphnétine diméthyléther		206,20	2445-80-9
Scoparone		206,20	120-08-1
Limettine		206,20	487-06-9
Fraxétine		208,17	574-84-5
Xanthotoxine		216,19	484-20-8
Bergaptène		216,19	298-81-7
Sphondine		216,19	483-66-9
Isobergaptène		216,19	482-48-4
5,8-dihydroxypsoralène		218,17	14348-23-3
5,6-dihydroxyangélicine		218,17	857019-99-9

Nom	Structure	Poids moléculaire (g/mol)	Numéro CAS ou InChIKey
5,7-diméthoxy-3-méthylcoumarine		220,22	
Déméthylsubérosine		230,26	21422-04-8
Osthénol		230,26	484-14-0
5-hydroxanthotoxine		232,19	7471-73-0
8-hydroxybergaptène		232,19	1603-47-0
Osthole		244,29	484-12-8
Subérosine		244,29	581-31-7
Isopimpinelline		246,22	482-27-9
Pimpinelline		246,22	131-12-4
***(+)*-Marmésine**		246,28	13849-08-6

Nom	Structure	Poids moléculaire (g/mol)	Numéro CAS ou InChIKey
***(+)*-Columbianétine**		246,28	1147-29-1
5-hydroxymarmésine		262,21	YVFCHGMDKP GYIK-UHFFFAOYSA-N
Isoimpératorine		270,28	482-45-1
Hératomine		270,28	61265-06-3
Lanatine		270,28	76026-24-9
Impératorine		270,29	482-44-0
Cnidiline		300,31	14348-22-2

Nom	Structure	Poids moléculaire (g/mol)	Numéro CAS ou InChIKey
Phelloptérine		300,31	2543-94-4
6-isopentényloxyisobergaptène		300,31	24099-29-4
Bavachinine		338,40	19879-30-2
Bergamottine		338,41	7380-40-7
Rutine		610,50	153-18-4

Annexe 3 : Annotations des champs de Gene Ontology obtenus pour les cytochromes P450 identifiés à partir des trois banques d'ADNc.

Annotation GO	*P. sativa*	*R. graveolens*	*C. cinereum*
2-methylbutanal oxime monooxygenase-like	4		
3-epi-6-deoxocathasterone 23-monooxygenase-like			3
7-ethoxycoumarin o-deethylase-like			3
13-hydroperoxide lyase			1
aba 8-hydroxylase		3	
aba 8-hydroxylase cyp707a1	1		
aba 8-hydroxylase cyp707a3	1		
aba 8-oxidase	1		
abscisic acid 8 -hydroxylase			1
abscisic acid 8 -hydroxylase 2		1	
abscisic acid 8 -hydroxylase 4-like			2
af092915_1cytochrome p450-dependent fatty acid hydroxylase	1		
af258262_1cytochrome p450-like protein		1	
af264037_1cytochrome p4501a			1
af332974_1p450 monooxygenase		1	
af426451_1cytochrome p450		2	
allene oxide synthase	1	1	
c71a8_menpi ame: full=cytochrome p450 71a8	2		
c76a2_solme ame: full=cytochrome p450 76a2 ame: full=cyplxxvia2 ame: full=cytochrome p-450eg7	1		
c77a1_solme ame: full=cytochrome p450 77a1 ame: full=cyplxxviia1 ame: full=cytochrome p-450eg6	1		
c85a_phavu ame: full=cytochrome p450 85a ame: full=c6-oxidase			1
c93e1_soybn ame: full=beta-amyrin 24-hydroxylase ame: full=cytochrome p450 93e1 ame: full=sophoradiol 24-hydroxylase			1
carotene epsilon- chloroplastic-like	1		4
cinnamate 4-hydroxylase	7	3	2
coniferaldehyde 5-hydroxylase		1	
cytochrome	1	1	3
cytochrome family subfamily polypeptide 3	2	1	
cytochrome family subfamily polypeptide 13			1

Annotation GO	*P. sativa*	*R. graveolens*	*C. cinereum*
cytochrome partial		1	
cytochrome p450	137	155	35
cytochrome p450 1a1			1
cytochrome p450 2c23-like			1
cytochrome p450 2j2-like			2
cytochrome p450 2j6-like			1
cytochrome p450 2u1- partial			1
cytochrome p450 4g15-like			1
cytochrome p450 4v2-like			1
cytochrome p450 71 family protein		2	
cytochrome p450 71a1-like	9		17
cytochrome p450 71a2-like isoform 1	1		
cytochrome p450 71a4	2		
cytochrome p450 71a9 precursor			1
cytochrome p450 71a22	1		
cytochrome p450 71a24	1		
cytochrome p450 71a24-like			1
cytochrome p450 71a26-like			5
cytochrome p450 71a28	1		
cytochrome p450 71b2	1		
cytochrome p450 71b9	1	1	
cytochrome p450 71b10	2		
cytochrome p450 71b34-like	1		
cytochrome p450 71b35-like	1		
cytochrome p450 71b35-like isoform 2	2		
cytochrome p450 71b37			2
cytochrome p450 71d8-like		1	3
cytochrome p450 71d10		1	
cytochrome p450 71d10-like		2	1
cytochrome p450 71d11-like		11	10
cytochrome p450 cyp72a129	4		

Annotation GO	*P. sativa*	*R. graveolens*	*C. cinereum*
cytochrome p450 76a2	1		
cytochrome p450 76a2-like			1
cytochrome p450 76c1-like		1	
cytochrome p450 76c4	1		
cytochrome p450 76c4-like		1	1
cytochrome p450 77a2	4		
cytochrome p450 77a3	1		
cytochrome p450 77a3-like	1		2
cytochrome p450 78a3-like	1		4
cytochrome p450 78a4 isoform 1	2		
cytochrome p450 78a4-like		1	1
cytochrome p450 81d1			1
cytochrome p450 81f1-like		1	
cytochrome p450 82a1-like			1
cytochrome p450 82a3	1		
cytochrome p450 82a3-like	4		1
cytochrome p450 82a4-like	1		7
cytochrome p450 82c4	3	1	1
cytochrome p450 82c4-like			2
cytochrome p450 83b1	8	4	
cytochrome p450 83b1-like	5		17
cytochrome p450 85a	1	1	
cytochrome p450 85a-like			2
cytochrome p450 86a2	3		
cytochrome p450 86a2-like	2		5
cytochrome p450 86b1	1	2	
cytochrome p450 86b1-like	1		3
cytochrome p450 87a3-like		1	
cytochrome p450 89a2	1		
cytochrome p450 89a2-like	7	2	1
cytochrome p450 90a1		2	

Annotation GO	*P. sativa*	*R. graveolens*	*C. cinereum*
cytochrome p450 90a1-like			3
cytochrome p450 90b1-like		2	
cytochrome p450 93a1-like	1		2
cytochrome p450 94a1	3		
cytochrome p450 94a1-like	4	1	3
cytochrome p450 94a2-like	1	1	2
cytochrome p450 704c1-like		3	
cytochrome p450 710a1-like			3
cytochrome p450 716b1-like			1
cytochrome p450 716b2-like			4
cytochrome p450 734a6-like	3		
cytochrome p450 716b2		1	
cytochrome p450 734a1-like	1	1	
cytochrome p450 734a2	1	2	
cytochrome p450 734a4		3	
cytochrome p450 734a6	1		
cytochrome p450-1	1		
cytochrome p450 like_tbp	7	2	2
cytochrome p450 like protein	3	6	
cytochrome p450-type monooxygenase 97a29	3	3	
cytochrome p450-type monooxygenase 97c11		1	
cytochrome p450 5-hydroxylase for coniferyl alcohol and ferulic acid		1	
cytochrome p450 allene oxide synthase		1	
cytochrome p450 monooxygenase	4	1	2
cytochrome p450 monooxygenase cyp51g1			1
cytochrome p450 monooxygenase cyp72a61			1
cytochrome p450 monooxygenase cyp72a65			1
cytochrome p450 monooxygenase cyp72a67			1
cytochrome p450 monooxygenase cyp97a10	1		
cytochrome p450 monooxygenase cyp704g7		2	
cytochrome p450 monooxygenase cyp704g9		1	1

Annotation GO	*P. sativa*	*R. graveolens*	*C. cinereum*
cytochrome p450 monooxygenase cyp706a12	1		
cytochrome p450 monooxygenase cyp710a15	2		
cytochrome p450 a	1		
cytochrome p450 chloroplastic	1		1
cytochrome p450 cyp736a12	1		
cytochrome p450 cypm3r9			1
cytochrome p450 obtusifoliol 14-alpha-demethylase	1	2	
cytochrome p450 probable 6-deoxocathasterone to 6-deoxoteasterone or cathasterone to teasterone	1	4	
cytochrome p450 probable 6-deoxoteasterone to 3-dehydro 6-deoxoteasterone or teasterone to 3-dehydro teasterone		1	
cytochrome p450 probable campestanol to 6-deoxocathasterone or 6-oxocampestanol to cathasterone		2	
cytochrome p450 probable ent-kaurenoic acid oxidase	1	1	
cytochrome p450 reductase			1
cytochrome p450-dependent monooxygenase-like protein	6		
ent-kaurene oxidase	1	1	
ent-kaurenoic acid 13-hydroxylase	1		
ent-kaurenoic acid oxidase	1		1
ent-kaurenoic acid hydroxylase 1		1	
fatty acid hydroperoxide lyase	1		
ferulate 5-hydroxylase	2		
ferulate 5- partial			1
flavonoid 3-	2	3	
flavonoid 3-hydroxylase	3	1	1
flavonoid 3 -monooxygenase		1	
flavonoid 3 -monooxygenase-like	7	5	4
geraniol 8-hydroxylase-like			1
geraniol 10-hydroxylase	12		
histone deacetylase	1		
hydroperoxide lyase		1	1
isoflavone 2 -hydroxylase		2	
isoflavone 2 -hydroxylase-like		1	5

Annotation GO	*P. sativa*	*R. graveolens*	*C. cinereum*
isoflavone synthase 1			2
licodione synthase			1
low quality protein: cytochrome p450 71a4-like	1		
low quality protein: cytochrome p450 71d10-like		1	
low quality protein: cytochrome p450 71d11-like			1
low quality protein: cytochrome p450 82c4-like	1	1	
low quality protein: cytochrome p450 83b1-like	1		
low quality protein: secologanin synthase-like	1		
obtusifoliol 14-alpha demethylase	2	1	
p-coumaroyl quinate shikimate 3 –hydroxylase	2		
p-coumaryl- 3 -hydroxylase	1		
premnaspirodiene oxygenase-like		1	
probable cytochrome p450 49a1-like			1
protein lutein deficient chloroplastic-like			1
secologanin synthase		1	
secologanin synthase isoform 1		1	
secologanin synthase partial		1	
secologanin synthase-like	2	7	12
steroid c-6 oxidase			1
thromboxane-a synthase-like	1		
trans-cinnamate 4-monooxygenase	1		
unnamed protein product [Vitis vinifera]	1		

Annexe 4 : Séquences complètes correspondant à des P450s isolées à partir de la banque d'ADNc de *P. sativa*.

Contig	Annotation	Longueur contig (nt)	Changement cadre de lecture
P_sat_rep_c169	trans-cinnamate 4-monooxygenase	1793	Non
P_sat_rep_c588	cytochrome p450 83b1	1774	Non
P_sat_rep_c729	cytochrome p450 83b1	1681	Non
P_sat_rep_c1089	cytochrome p450	1717	Non
P_sat_rep_c2306	cytochrome p450 77a3-like	2069	Oui
P_sat_rep_c2505	cytochrome p450 86a2	1899	Non
P_sat_rep_c2542	cytochrome p450	1554	Non

Annexe 5 : Séquences complètes correspondant à des P450s isolées à partir de la banque d'ADNc de *R. graveolens*.

Contig	Annotation	Longueur contig (nt)	Changement de cadre de lecture
R_gra_c89	cytochrome p450 71d11-like	1676	Oui
R_gra_c149	low quality protein: cytochrome p450 71d10-like	1679	Non
R_gra_c578	flavonoid 3 -monooxygenase-like	1685	Non
R_gra_c643	cytochrome p450 71d11-like	1967	Non
R_gra_c696	cytochrome p450	1881	Oui
R_gra_c1526	isoflavone 2 -hydroxylase	1925	Non
R_gra_c1542	secologanin synthase-like	1682	Non
R_gra_c1672	cytochrome p450	1728	Non
R_gra_c1737	cytochrome p450 83b1	1804	Non
R_gra_c1884	cytochrome p450 86b1	1652	Non
R_gra_c2181	cytochrome p450 83b1	1873	Non
R_gra_c2737	cytochrome p450	1930	Oui
R_gra_c3128	cytochrome p450	1999	Oui
R_gra_c3135	flavonoid 3 -monooxygenase-like	1778	Non
R_gra_c3823	secologanin synthase	1875	Non
R_gra_c4163	cytochrome p450	1898	Non
R_gra_c4541	cytochrome p450	1933	Oui
R_gra_c5032	cytochrome p450	1559	Oui
R_gra_c5041	cytochrome p450	1831	Non
R_gra_c6403	secologanin synthase-like	1827	Non
R_gra_c7149	secologanin synthase-like	1701	Non
R_gra_c7779	ent-kaurene oxidase	1595	Non
R_gra_c10009	cytochrome p450	1695	Non
R_gra_c12789	cytochrome p450 monooxygenase	1648	Non
R_gra_c31430	flavonoid 3 -monooxygenase-like	1668	Oui
R_gra_c33406	cytochrome p450-type monooxygenase 97c11	1672	Oui

Annexe 6 : Séquences complètes correspondant à des P450s isolées à partir de la banque d'ADNc de *C. cinereum*.

Contig	Annotation	Longueur contig (nt)	Changement de cadre de lecture
P_cin_c9	7-ethoxycoumarin o-deethylase-like	1837	Non
P_cin_c17	cytochrome p450 71a26-like	1715	Non
P_cin_c96	protein lutein deficient chloroplastic-like	2161	Non
P_cin_c475	ent-kaurenoic acid oxidase	1628	Non
P_cin_c843	cinnamate 4-hydroxylase	1817	Non
P_cin_c877	cytochrome p450 monooxygenase cyp51g1	1934	Non
P_cin_c2997	cytochrome p450 71a26-like	1929	Non
P_cin_c3613	cytochrome p450 2j6-like	1663	Oui
P_cin_c4080	cytochrome p450 chloroplastic	1684	Oui
P_cin_c4264	cytochrome p450 85a-like	1886	Non
P_cin_c5419	cytochrome p450	1997	Oui
P_cin_c5649	cytochrome p450 86a2-like	2155	Non
P_cin_c6189	cytochrome p450	1793	Non
P_cin_c6257	cytochrome p450 77a3-like	2146	Oui
P_cin_c7100	flavonoid 3 -hydroxylase	1621	Non
P_cin_c8594	cytochrome p450 78a3-like	1624	Oui
P_cin_c8834	cytochrome p450 78a4-like	1636	Oui
P_cin_rep_c9422	secologanin synthase-like	1870	Non
P_cin_rep_c9447	cytochrome p450 71d11-like	2083	Non
P_cin_rep_c9476	cytochrome p450 83b1-like	1919	Non
P_cin_rep_c9592	cytochrome p450 83b1-like	1824	Oui
P_cin_rep_c9657	cytochrome p450 liketbp	1713	Oui
P_cin_rep_c9698	cytochrome p450	2222	Oui
P_cin_rep_c9737	flavonoid 3 -monooxygenase-like	1872	Non
P_cin_rep_c9822	7-ethoxycoumarin o-deethylase-like	1713	Non
P_cin_rep_c9915	licodione synthase	1953	Non
P_cin_rep_c9972	cytochrome p450 82a4-like	1815	Non
P_cin_rep_c10023	isoflavone synthase 1	1763	Non

Contig	Annotation	Longueur contig (nt)	Changement de cadre de lecture
P_cin_rep_c10482	cytochrome p450 89a2-like	2073	Non
P_cin_rep_c10498	cytochrome p450 716b2-like	1727	Non
P_cin_rep_c10705	flavonoid 3 -monooxygenase-like	1699	Non
P_cin_rep_c10717	cytochrome p450 94a2-like	1957	Non
P_cin_rep_c11026	cytochrome p450 94a1-like	1764	Oui
P_cin_rep_c11200	cytochrome p450 71a1-like	1720	Non
P_cin_rep_c11261	cytochrome p450 71a1-like	1674	Non
P_cin_rep_c11530	cytochrome p450	1615	Non
P_cin_rep_c11636	13-hydroperoxide lyase	1618	Non
P_cin_rep_c11713	carotene epsilon- chloroplastic-like	1914	Oui
P_cin_rep_c11978	cytochrome p450 83b1-like	1714	Oui
P_cin_rep_c12074	cytochrome p450 82a4-like	1831	Oui
P_cin_rep_c12157	cytochrome p450 83b1-like	1888	Non
P_cin_rep_c12308	cytochrome p450 82a4-like	1944	Oui
P_cin_rep_c13071	cytochrome p450	1669	Oui
P_cin_rep_c13417	cytochrome p450	1971	Oui
P_cin_rep_c13565	cytochrome p450 76c4-like	1697	Non
P_cin_rep_c16429	secologanin synthase-like	1675	Non
P_cin_rep_c21518	cytochrome p450 71d11-like	1878	Non
P_cin_c24241	cytochrome p450 716b1-like	1654	Oui
P_cin_c25341	cytochrome p450 4v2-like	1911	Oui
P_cin_c35496	cytochrome p450 86b1-like	1607	Oui

Annexe 7 : Séquences peptidiques complètes de cytochromes P450 isolées à partir de la banque d'ADNc de *Pastinaca sativa*.

>P_sat_rep_c169_pep 506 aa
MDFLLLEKALLGLFVATIVAITISKLRGKKLKLPPGPIPVPVFGNWLQVGDDLNQRNLVDYAKKFGDL
FLLRMGQRNLVVVSSPDLARDVLHTQGVEFGSRTRNVVFDIFTGKGQDMVFTIYSEHWRKMRRIMT
VPFFTNKVVQQYRFGWEDEAARVVADVKANPEAATNGIVLRNRLQLLMYNNMYRIMFDRRFESVD
DPLFLKLKALNGERSRLAQSFEYNFGDFIPILRPFLRGYLKLCQEIKDKRLKLFKDYFVDERKKLESSL
KRVDNNSLKCAIDHILEAQEKGEINEDNVLYIVENINVAAIETTLWSIEWGIAELVNNPEIQKKLRHEM
DTVLGVGVQICEPDVQKLPYLQAVIKETLRYRMAIPLLVPHMNLHDAKLAGYDIPAESKILVNAWWL
ANNPAHWNNPDEFRPERFLEEESKVEANGNDFKYIPFGVGRRSCPGIILALPILGITIGRMVQNFELFPP
PGQSKIDTAEKGGQFSLQILKHSTIVCKPRSF*
>P_sat_rep_c588_pep 506 aa
MDPAAIFLILAIPIASVYLLFYHKKRVNGLSSPPGPRGLPFIGHFYQIYKSECAHEYISNLSKQYGSLMT
LHLGSVPALVVSSPKMAQEVLKTQDLVFCSRAQMTGSGKLSYNGLEMAFAPYGEHWRNVRKMCTL
ELFTQKRAQFNFRPVREDEVSRMVGRLSEAAAASEDVNAYECFTNFATSIISRVAFGKRYDEDNLGK
EKFQRMVADIEAMFAAFFVSDFFPMFGWIDRLSGVKAVLDRNFNEMDTFYQELIDEHXKPDRPESLN
GDLIDVMLKNKGSFLTMDSIKAILLNVFSGGIGTTGSALVFAMTALLRNQRVMKKAQEEVRSVIGKK
EIVDEDDIQKLPYLRAVVKETLRLYPPGPLLIPRVAMESCVLGEDEDHMYMIKPNTIVYVNTWGIGRD
PKYWKNPLEFMPERFFERPDLNYTGQQFEYLPFGSGRRICAGIIIGQNNVEVGLANLLYSFDWEPPTG
KTFEDIDDQPCNGLTLAKKNPLYIRPKIYVHP*
>P_sat_rep_c729_pep 506 aa
MEPVFLFLILAFPIASVYLLFYHKKRVVGLSAPPGPPGLPFIGNFHQLYKASSTHEYLWSLSKQYGSLV
TLRMGSVPILVVSSPKMAKEVLQTQDLIYCSRPAMTGMQKLSYNGLDVAFSPYTDQWRHTRKFCTL
ELFTQKRAQLNFRPVREQEVSRMIARLSETAAASKDVNAYEWFSNLATCIISRVAFGKRYDEDGIGKE
RFQRMLSEIDAMFIGFFVSDFFPMFGWIDKLSGMRDRLDRTFKDLDMFYQELIDEHLKPNRLESPTED
LIDVMLKNEGSSLTKDSMKAILLNVFNGGTGTTATVLAWAMTALLRNQGVMKKAQEEIRSVIGKKG
KVDEDDFPSLPYLRAVVKETMRLYPPAPVLVPRETMESSIIGEDKDHMYMVKPKTVVYISMWAIGRD
PKTWKNPMEFVPERFLERPDINYKGQQFEYVPFGAGRRICAGINLGVTTVELALANLLYTFDWEPPA
GMRFEDIDDETTNGLALQKKNALYIRPKKYMFP*
>P_sat_rep_c1089_pep 506 aa
MEFPSPFSIISSLIFILVIFQLVTKSKRKAKSNLPPGPWKLPILGNLLQVAGKIPHLGLRELAKIYGPLMH
LQLGEISAIVISNPRVAKEVLRTNDLCFADRPELILGTIILANCRDIVLAIYGDYWRQMRKICTLELLSS
HKVRSFRFIREDEAWQLIQSIKSSSGLPVNISDKVSKVSNAITCRSTIGEKCKYQDELIELVEEIAHWGS
GFFMADLFRSVKFLPQLTGMKPALKKIRSKLDVIFDSIIKEHKERLNKRKEGTEYANEEEDLVDVLLRI
NESQRLEFPISSGDIQGLVLDMFTAGTDTSSAVLEWAMSELMKHPRVMKKVQSEVREVVKGKERIEE
TDIQKMSYLKLVVKETLRLHAPVPLLLPRECRKECEVDGYTIPVGTKVIVNAWAIGRDPEYWADSESF
IPERFENSSIDYIGANFEFIPFGAGRRMCAGISFGVATVELPLAQMLYSFDWKLPNEMKPEDLDMDET
NAATSKRKNNLYLLATEYPIEVVE*
>P_sat_rep_c2505_pep 545 aa
MDASTVMMVLAMMAAYLMWFRTMVRPLKGPRVWPLVGSLPGLIENSNKMHDWIAENLRACGGT
YQTCIAAVPFLAQKQGLVTVTCDPKNLEHILKLRFENYPKGPTWQGVFHDLLGEGIFNSDGDTWKFQ
RKTAALEFTTRTLRQAMARWVSRAIKLRFCPILKTAQLEGKPVDLQDLLLRLTFDNICGLAFGKDPET
LSPGLPENHFASSFDRATEATLQRFILPEMIWKLRRWLRLGMEVSLSESIGHVDKYLTNVINTRKLELI
TQQNGGGGIPHDDLLSRFMKKKESYTDQFLQHVALNFILAGRDTSSVALSWFFWLVTKNPRVEEKIL
MEICAVLMETRGKDISKWLDDPLVFEEVDRLVYLKAALSETLRLYPSVPEDSKHVTADDILPDGTXV
LAGSSITYSIYSSGRMKFIWGEDCLEFRPERWLTGDGTKFETKDQFKFVSFNAGPRICLGKDLAYLQM
KSIAAALLLRHRLTVAAGHRVEQKMSLTLFMKYGLKVDLHPRDLTPIVAKIGENYACMGNTCDEELE
YVKLVAEVA*
>P_sat_rep_c2542_pep 500 aa
MKMLEQNPLYLYFFSLFLVTIFLYKWLAKKTPSKNLPPSPPRLPIIGNLHQIGPDLHISLRDLARKYGPL
MQLQLGRIPVLVVSSAEATREVLKTHDVVFSQRPITSAIDKLCYKGRDVAFSRYSEYWRQVRSTCVT
QLLSNSRVHSFHNIREEEVALLIQNIENSASEVINLGEQLIQLTRNVVCRVSVGSEYLSGHKGKLYQKL
LAEVTEMLAYTYSIGDFIPLLGWVDWLSGSKAKVEKTAKEVDAFLEGALRDHIKTMASNKGSANDD
FLSILLEIREADAGSTLDEECIKAIVWDMILGGTETTSTTLEWIVAAIIKNPDVMFKLQKEVREIGKGKS
KIEEVDLVKMNYLKAVMKESMRLYITAFLLPREAKQDVKLMGYDISSGTQVLINTWETARDPSLWD
NPEEFRPERFLNSPIDYKGLHYEYLPFGGGRRGCPGIQFAMAVNELAVANVVYKFDFKMPDGERFED

Annexe 8 : Séquences peptidiques complètes de cytochromes P450 isolées à partir de la banque d'ADNc de *Ruta graveolens*.

>R_gra_c149_pep 533 aa
MELHFPSIVFIFSFLVFLFTIVKILLQSKTLQGTSEERLPPGPWRLPLIGNLHQLFGSLPHHRLRDLAKQH
GPLMHLQLGEVSTVVVSSPEVAKEVLKTHDLIFAQRPLLVSAKFTSYDYTNIGLAPYGNYWRQLRRIC
TVELLSANRVQSFRSIREEEVSNLIKAVGSNSEGSAINLSEGIFTMTFGITARAAFGKKCKDQESFINVV
KEFIQIMSGFSIADLYPSVKLLQEISGVRHKLEKLHLESDRIFEEIIDEHKRKRKKAKTEESAEAGEDLV
DVFLNLQEDGDLEFPLTDNNIKAVIWDIFVAGSDTSSTTVEWAMAEMLKNPRVMKEAQAEVRRVFN
GKENVDEIGIHELKFLKLVVKETLRLHPPAPLLLPRESRESCRINGYTIPEKTRVIVNAWAIGRDPEYW
TEAETFCPERFMDSSVDYRGTDFEFIPFGAGRRVCPGITFALPNIELPLAQLLYHFDWELPSAGMKHQD
LDMTEDFGLTIRRKNDLLLIPSSRGXSSDFSNVENNHHCXIVSRCSVFCF*
>R_gra_c578_pep 502 aa
MWYYYSVLLIIPVFLFLLRSKKKLNLPPGPKPWPLIGNLNQIGPLPHMSTHALSRKYGPLMHLKFGLD
PVVVVSSVEMAELFLKTHDLSFSSRPQLLAGKHLTYDYTGMGTAXYSPYWRQARKICVMELLSTKR
LDQFEYMRVEERRIFLNGMFKAANTPVDLKDRLYMLNLSAMSRMVLGKRYTEEDENNIVTPKEFTE
MLEELFLVHGLLDIGDSIPWLAPLDLQGHVKRMKDVNRRLDKFYEHIXDERYARKDSIKKLGAQDM
VDVLIQFADDPTLEVKLQRHQLKAIIQDLLTAGTDTSALNVEWSMSLLLKNPEAFRKANEELDRVIGK
DRWVEEKDIVNLPYIEAIVKETMRLHPVVPLLIPRVAREDCKVAGYDITKGTRVLVNIWTIGRDPALW
ENPTEFQPERFIGKDIDIKGHHYQLMPFGSGRRMCLGYGLGLKLVQTSLANLLHGFEWKLPEGMKKE
DLDMEERFGLTTCKKNPLVVVPKPRLSPHLYTL*
>R_gra_c643_pep 507 aa
MEFLPISIAFLLFAITLLKLVKKSKNRNSCAKPPPPGPWKLPVIGNLHQLAFSDLPHHRLRALSKKYGP
LMLLKLGEVQAVIITSPEVAKEVLKTNDVVFASRPHSDPAEIISWKNSGVVFAPYGESWKQLRKMCA
SLLLTPKRVQSFRAIREEEVSELYDWICSNGGSEINLSDKIYTLMCGITSRAAFGNKTGDSEVFVRIIQE
ATKLISGFNIEELFPSCGILGWFTGMRPKLERLHEKSDIIVENIINEHKKRKAASEDRKVDDSEEDLVDV
LLRIQEQGDLDSILTDDQIRAIIFDMLSAGSETAATTIDWAMCELMKNPRVMKKAQAEVREVFTRRG
KVDEEGTNEMTYLKLICKETLRLHPAAPLLIPRECGERCLINGFEIPEKAKVMVNAWAIGRDPELWTE
PESFMPERFIDDSVDYKAVNFKYIPFGAGRRICPGISFALALMELPLAMFLYHFDWQLPNGLKPGDMD
MTESFGMTVKRKQDLSIVATPTYPLPVA*
>R_gra_c1526_pep 510 aa
MEETTILYSALSLIFLVIALNKLSLIFRRGKNLPPSPPAFPIVGHLPFIKQPMHRILHRFSQKYGPVISLRF
GSRLVIVVSSYQAAEECFTKNDVVFANRPRFLTGKYLGYNYTVLTQASYGDHWRNLRRITAIEVLST
HRLNKFLPIRKQEIQRLLKKLSSSGSGHEDQFSKVEMKSMLSELTFNIMMRMIAGKRYYGDDVENEE
EARRFRKIMKEVIAYGGATNPGDFLPILKWIDGGDFEKKIMRFSEAIDGFLQGLIDEHRTNKAGPESNN
TMIDHMLTLQESEPEYYTDQIIKGLILVLLLAGTDTSSLTVEWAMSNLVNNPQVLEKARAELDNKVG
LECLIDEPDLSKLHYLNNIISETLRLYPVGPLLVPHQSSDDCTVGGYHVPAGAILLVSAWSIHRDPKLW
NDPNNFRPERFEEDAGEGETRSKLMPFGLGRRACPGSGLAQRVVGLTLGSLIQCFEWQRIDEDKVDL
TEGRGLTLPKANPLEVMCRARPIVKNVLEHLD*
>R_gra_c1542_pep 513 aa
MGSLSFAFAVLLIAAASAYAAWKLLNWVWVKPRKLEKLLRQQGFSGNPYRFFHGDMKEFAHLTRQ
AKSKPIGLDDDITHRILPFHNHIIAKFGKSSFIWMGPRPNINITDPKMIREVLLKKDIFTKLSSPIDKLLVS
GMAVYEGEQWFKVRKIATPAFHHDKLKNMLPEIHKSCCDMLRKWEISVSKEGSCELDVWPDIQTLT
ADVISRTAFGSSFDEGRKIFQVLREQINLLIRASLLSYIPGWRFLPNRLNRQMKKNYNEMSKLVKAIID
KRQEALKDGKVSTHDDLLGIMMESNKKLIKEKGTGMSIDEVIEECKLFYLAGQETTATLLVWTMVLL
CWYPEWQERAREEVFQVFGVKEPKFDDLNQLKVVSKILNEVLRLYTPAPMLTRVNKEEIKLGEIVLP
PGVALSLPLILVHHDHEYWGDDAKEFNPDRFSEGVSKASKNQISFFSFGGGPRICIGQNFALLEAKLA
MAMILQKFTFQLSPTYAHAPTRGVTLFPQYGANIILHKI*
>R_gra_c1672_pep 526 aa
MALILSPESAVAAAIAALLIFLSVSASISRYGVRRNSNNTRKLARDQAPKARGAWPLIGHLHLLAGSEP
AHRVLGNMADKYGPIFTIKLGINPALVVSNWELAKECFTTNDRAFANRPKTMAMELLGYNFLVFAFS
QYGHYWRQSRXIATAELLSNHRLEKLKHVFEHEVKGSIKKLCENCRDSSSSKVFVEMKHWLEGLTL
DIILSIIVGKSHDSLEGKDWRQKMANFFALSGKFVVSDALPFLGWLDIGGDERLMRKTAKEFDIFLQQ
WLDEHKMKRASGQVKGDEDLMYVLLSLLDGNAEQLTDHDADTVIKALCLNLMLAASETTTTTMT
WAISLLLNNRDVLKKAQAEIDTQVGTERQVNELDMKNLIYLHAIIKETMRLYPAAPLFIPHECREECT
VGGYHVPIGTQLYVNAWKIHRDPRVWEEASEFKPERFLTTHKDIDYRGQNPELIPFGSGRRMCPGISF
GIHVMQLVLASLLHGFDLATPSGEEVDMGERAAITIVKATPVEVALTPRLAGSLYG*

Annexe 8 : Séquences peptidiques complètes de cytochromes P450 isolées à partir de la banque d'ADNc de *Ruta graveolens*.

>R_gra_c1737_pep 507 aa
MEVVSLLVVTALFFLFITHMVMRYRSTNRRLKSTKALPLPPGPRGLPFIGNLHQFSNSNPHLYLLKLSK
QFGPLLSLRLGSVPVIVVSSAKMAEQVLKVHDLQFCSRPAQLGLQKLSYNGSDLGFAPYNSYWREIR
KICVVHLFNSSRTRSFRPVREDEVSRMIEKVSRSVVASKPVNLSEATVSLACSITSRIGFGKKYEEEGSG
RSRFHTLLNEAQALLASFFFTDYFPSIGFWVDKLTGMISRLERIFKELDAFYQALIDDHLDPKRIRTDQ
QETIVDVLLEAQKEHGKVDLTYDNIKGVLMNVFLAGTDTIAATMVWAMSYLMKFPRAMKKVQEEI
RSSMGTKGFVDEDDLKNLSYLRAVVKETLRLQPALIARETMETCIIDGYEIPRKTQVLVNSYAVGRDP
DAWECPDEFYPERFIDSSVDFKGQHFQFIPFGAGRRICPGMTMGVATVDLALANLLYKFDWEMPPGM
DKEEIDFDVLPGVTMHKKNLLSLMAKHHNSSC*
>R_gra_c1884_pep 495 aa
MIESSAFLQYLIFPVIFFAFLYGFPKFNSTSWNMPLIGMIVSLILNINQLHDRVAEMLERKKGSFMIKCS
WIAQKDMFLTSDPANVQHIMSTNFSAYHKGSEVKKRFDIFGDTGLFNSDFDEWKNQRKLARAFINHH
TFQQLMAKIVPEMIENGLIPVLDHVCEEKMVVDLQDLLKRSTFDFACMIATGCNPNSLSVGFPKMEFP
DAVDDACEAILIRHLVPESLWKLQSWLGFGKEQKLANGWRTIDQFWEKQKSVKLLDPDHNYSLLDC
YLSGHEVTGPTPSEKVMRDNITGLIFATEGTTSIALTWFFWLISKNPFAECKIREELQRNNFPQEGSLKL
VYLHAALCETLRLYPPVPFEIRIPLQDDVLPSGHQVNRNTNVLISAYSMARMSSVWGEDCREFKPER
WITEDGLRIKHEPAHKFFSFNAGPRICIGKDVAFTLMKAMAANIIHNYDVKVEKNHPVTYNLSVSFRM
KYGLLARIYKRSDVKKTES*
>R_gra_c2181_pep 505 aa
MAVVSLLVVPALFFLFVTHMIMRYRSTNRRVKSTKALPLPPGPRGLPFIGNLHQFHNSNPHQYLLKLS
KQFGPLLSLRLGSVPVLVVSSAKMAEQVLKVHDLQFCSRPALLGLQKLSYNGSDLAFAPYNSYWREI
RKICVVHIFNSSRTRSFRPVREVEVSRMIKKVSRSVAASKPVNLSEATLSLACSITYRIGFGKKYEDQGS
ERSRFHTLLNDLLASFFFTDYFPSIGSWVDKLTGTISRLDRIFKELDAFYQAVIDEHLDPKRIRTDQQQT
IVDVLLEAQKEHGLKVDLTYDNIKAVLMNVFIAGTDTSAATLVWAMSYLMKFPRAMKKVQEEIRSS
MGTKGFVDEDDLKNLSYLRAVVKETLRLQPALIVRETMETCIIDRYEIPRKTQVLVNAYAVGRDPDA
WESPDEFYPERFIDSSVDFKGHHFQFIPFGAGRRICPGMSMGMAAVDLALANLLYKFDWEMPPGMD
KEEIDFDVLPGVTMHKKNLLSLMAKHRNSSC*
>R_gra_c3135_pep 508 aa
MEKLAVCCYSILFLIPVLLLLLRRRKNKPNLPPGPKPWPFIGNLNLIGPLPHVSIHSLSQKYGPLMHLKF
GLSSVVVGSSVDVAQLLLKTHDISFASRPPLLAGKYTTYDFSGMGTAPYGPYWRQARKICVMELLSP
KRLDQFEYIRVEERKSFLFELXKSASNPVHLKDRLYTLNLTSICRMTLGKRYTDESEGNAITPKEFTEM
LEELFLLNGVLDIGDSIPWLAPLDLQGQVKRMKAVAKKLDGFYESILDEHFARRKSVRDHGIRDMVD
VLVQLSDDPSLEVKLERHHIKALIQDLLTAGTDTSAINVEWAMSELLKNPEAIKKATEELDRVIGKDK
WVEEADIVKLPFIEAIVKETMRLHPVAPLLVPRIAREDCKVAGYDIPKDTRVLINVWAIGRDPTVWEN
PNEFKPERFIGKAIDVKGHDYELLPFGSGRRMCLGYGLGLKVVQSSLANLLHGFEWELAGDMKGXD
LNMDDKFXLTMSRKIPLVVVPXPRLPLHLYXCQ*
>R_gra_3823_pep 513 aa
MGSLSFAFVVLLIAAASAYAAWKLLNWVWLKPRKLEKLLRQQGFSGNPYRFLHGDMKEFAHLTKQ
AKSKPIGLDDDIIHRILPFRSHIIAKFGKNSFIWMGTRPNINITDPKMIREVLLKQDIFTKLRSPIGELLLT
GMLVYEGEQWFKVRKVVTPAFHHDKLKNMLPEIHKSCCDMLRKWEILVLKEGSFELDVWPDIQTLT
ADVISRTAFGSSFEDGRKIFQVLREQINLVFRASLVSYIPGWRFLPNRINRQMKQNYNDMSKLVKAIID
KRQEALQDGKASTHDDLLGIMMESNKKLIEEKGAGMSIDEVIEECKLFYLAGQETTATLLVWTMVLL
CKHPEWQERAREEVFQVFGVCEPKFDELNQLKVVSRILYEVLRLYSPAPVVNRVTKREVKLGEIVLPP
GLSLTLPLIVVHHDHEYWGDDAKEFNPDRFSEGVSMASKNQISFFSFGGGPRICIGQNFALLEAKLAM
AMILQKFSFQLSPTYAHAPTRGVTLYPQYGVNIILHKI*
>R_gra_c4163_pep 530 aa
MSSPLSWLCNANGSTDKCQNLMAIIASLVLVSAAVVWFTKKSKKESSFRLPPGPRGLPLVGFLPFLGT
DLHKSFTDLAGAYGHIYKFWLANQLWVVVSSPELAREVVRDKDTIFGNREATIAARLSTFGGNDIVW
CPYGAHWRKLRKVMVGEMMSNSILDSCSALRKQELKNAIRDLHANKIGKPVDIGEFATSASINVLQN
MLWGGVHDLGDDRQGTNVNAEVKQMLEDIMDLIGTPNISDIFPVLSWLDLQGIESKTKKIALWFENF
FNTVVEKRKNLIAKGKLDFHGGKDFLQTFLELLDGEDSSTSLTKIQLMALLMDVITGGTDTTTTILEW
TMTMLLQYPEEMKKVQAELAEVVGKTNIVEESQLPRLKYLDAVVKETFRLHPPSKLLPPRTPSRETTI
GGYSIPKGASVIINLGAIQLDPQLWDNPLEFRPERFLMGGPDKFDYTGNNFQYMPFGSGRRICAGLPL
AERLVMYLLASLLHSFDWELPHGENLDLSDKYGLVVKKKKPLVAVPTPRLANAELYN*

Annexe 8 : Séquences peptidiques complètes de cytochromes P450 isolées à partir de la banque d'ADNc de *Ruta graveolens*.

>R_gra_c5041_pep 526 aa
MEFSIELQEINAXWGILFAVTCLYVXTKMFIRKRTRGRRPPEAEGAWPFIGHLHLLGANQLLHQTLGE
MADKYGKIFMIRLGIRQALVISSWEVAKECFTTNDKLFSTRPRSLAMKLLGYDHSVLGFAPYGPYWR
DIRKLVSVELLSSHRLELLKHVHDKEVKIFIKELHGLWTQNGGTVSVEMKQRFGNLAMNIIVRVLAG
KRYYGSGTNADDHESRRCQKALGDFFYLIGLFLASDSVPFLGWLDILQGYTSKMKRTAQELDIVLGS
WVREHRCRRINNSITESKQDFIDVMLSVMDDGNVSADVVDNIIKSTCLTLIAGGNDTTVITITWALSLL
LNNRHELKKAQDELDTHVGKHRLVEESDIKNLVHLQAIAKETLRLYPAAPCSVPHEAMEDCTVAGF
HIPAGTRLIANLWKIHRDXRIWAKPSVFIPERFLTENANLDVRGQDFEYLPFGSGRRKCPGISFALQVV
HLTLARLLHAFELGTVSDTLVDMSESPGMTTPKATPLDIVLTPRLPSKLYDEL*
>R_gra_c6403_pep 513 aa
MGSFSFAFAVLLIAAASAYAAWKLLNWVWLKPRNLEKLLRQQGFSGNPYRFLHGDMKEFAHLTKQ
AKSKPIGLDNDITHRILPFHNHIISKFGKSSFIWMGPRPNINITDPKMIREVLLKQDIFTKLRSPIGKLLVS
GMVVYEGEQWFKVRKIATPAFHHDKLKNMLPEIHKSCCDMLRKWEISVLKEGSCELDVWPDIQTFT
ADVISRTAFGSSFEEGRKIFQVLREQISLLIRASLLSYIPGWRFLPNRINRQMKKNYNEMSQLVKAVVD
KRQEALKDGKVSTHDDLLGIMMESNKKLIEEKGTGMSIDEVIEECKLFYLAGQETTATLLVWTMVLL
CRYPEWQERAREEVFQVFGVKEPKFDELNQLKVVSKILNEVLRLYTPAPMLTRVNKEEIKLGEIVLPP
GVVLSLPLILVHHDHEYWGDDAKEFNPDRFSEGVSKASKNQISFFSFGGGPRICIGQNFALLEAKLAM
AMILQKFTFQLSPTYAHAPTRGVTLFPQYGANIILHKI*
>R_gra_c7149_pep 513 aa
MGSFSFGFAVLLIAAASAYAAWKLLDWVWMKPRKLEKLLRQQGFSGNPYRFFHGDMKEFAHLTKQ
AKSKPIGLDDDIIHRILPFRSRIIAKFGKKSFIWMGPRPNINITDPRMIREVFMKQDIFTKLRSPIGELLLT
GMLVYEGEQWFKVRKVATPAFHHDKLKNMLPEIHKSCCDMLRKWEILVLKEGSCELDVWPDIQTLT
ADVISRTAFGSSFEDGRKIFQVLREQINLVVRASLVSYIPGWRFLPNRINRQMKKNYNEMSKLVKAIID
KRQEALQDGKACTHDDLLGIMMESNKKLIEEKGAGMSIDEVIEECKLFYLAGQETTATLLVWTMVL
LCKHPEWQERAREEVFQVFGVNEPKFDELNQLKVVSKILYEVLRLYSPAPLVNRVTKKETKLGEIVLP
PGVVLSLPLILVHHDHEYWGDDAEEFNPDRFSEGVSKASKNQISFFSFGGGPRICIGQNFALLEAKLA
MAMILQKFSFQLSPTYAHAPTRGVTLYPQYGVNIILHKI*
>R_gra_c7779_pep 506 aa
MQAVPFATSIAIGXLSVFFFLFLXRRLISXPKRMEAVNLPXVPEVPGRLPLIGNLLQLKEKKPYMTFKK
WAEIYGPIYSIKTGASTMIVLNSTEAAKEAMVTRYQSISXRKLSNALKTLTSDKCMVAISDYDEFHKM
AKRCILTNVLGANAQKKHRIHRDTMMENLSSQFHAHIKDNPLQLVNFRKTFESELFGLSMKQAVGK
VVESIYVKELDATLSKNELLKILVLDIMEGAIEVDWRDFFPYLRWVPNNAVERKIQKLCSRRQAVMN
ALINDAKQRISSGQELNCYVDYLLSEAKSLTEQQLMMLLWETIIETSDTTLVTTEWAMYELAKDQSR
QERLYQEIQAVCGSNKFTEENISQIPYLGAVFHETLRKHSPAPIVPLRLAHEDTEIGGYSIPAGSEIAINI
YGCNMDKKLWEDPEEWKPERFLNGKYDPADLYKTMAFGAGKRVCAGSLQASLIAGTAIGRLVQEF
QWSLRDGVEESVDTVGLTTHKLHPLHAMIRPRN*
>R_gra_c10009_pep 522 aa
MEFSIDLQEINALWGILFAVTCLYVATKMFIRKRTRGRRPPEAEGAWPFIGHLHLLGANQLLHQTLGE
MADKXGXIFMIRLGIRQALVISSWEVAKECFTTNDKLFSTRPRSLAVKLMGYDHSMLGFAPYGPYWR
XIRKLVSVELLSSHRLELLKHVRDKEIKIFIKELYGLCTQNGGTVSVEMKQRFGNLAMNIIVRVLAGK
RYYGSGTNADDHESRRCQKALGDFFYLIGLFLASDCVPFLGWLDVLQGYSSKMKRTAQELDIVLGS
WVREHRCRRISKSITESKQDFIDVMLSVMDDGNVSAEVDNIIKSTCLSLIAGGNDTTVITITWALSLLL
NNRHVLKKAQDELDTHVGKHRLVEESDIKNLVYLQAIAKETLRLYPALPCSVPHEAMEDCTVAGFHI
PVGTRLIANLWKIHRDPRIWAKPSEFIPERFLTENANLDVRGHDFEYLPFGSGRRKCPGISFALQVVHL
TLARLLHAFELGTVSDTLVDMSESPGMTTPKATPLDVVLTPRLPSELY*
>R_gra_c12789_pep 475 aa
MAGTLTILYHLLLCISISLIAFFLLKRSRTRRLRLPPGSLGLPFLGETLQLISAYKTENPEPFIDERVRRFG
SIFTTHVFGEPTVFSTDPETNRFILQNEGKLFDCSYPGSISNLLGKHSLLLMKGSLHKRMHSLTMSFAN
SSIIRDHLLVDIDRLVRLHMDSWTDRVLLMEEAKKITFELTVKQLMSFDPNEWTESLRKEYVLVIEGF
FTVPFPLFSTTYRRAIQARTKVAEALSLIVRQRRKERDESGEKKKDMLAALLAGDDGFSDEEIVDFMV
ALLVAGYETTPTIMTLAVKFLTETPLALAQLKEEHDEIRAKKSQQEPLERSDYKSMPFTQCVVNETLR
VANIISGIFRRAMTDINIKGYTIPKGWKVFASFRAVHLDHDHFKDARSFNPWRWQNNSGAASAANVF
TPFGGGPRLCPGYELARVELSVFLHHLVTRFSWVAAEQDKLVFFPTTRTQKRYPIIVQRRNQS*

Annexe 9 : Séquences peptidiques complètes de cytochromes P450 isolées à partir de la banque d'ADNc de *Cullen cinereum.*

>P_cin_c9_pep 507 aa
MMETETFTCILLFLFSIIFLKCLTSLLTRNHEKPKLPPGPTGLPIVGNLLQLGSKPHQSMATLSKIHGPIMTLKLGQIT
TIVISSAQTAKEVLQIHDQFLSNRKTPDAVTGANHHLYSLPFLPLSHRWRHLRKLCNTLLFSNKTLDANQGPRLNK
LQDLFNHIHYSSIKGEPVDIGRLAFKTTINLLSNTIYSEDLIESTDKAGEFKELAANIMKEIGSPNLADCFPVLKVFDP
HGIRRRTGNYFWKLLTILKGLVDKRLKLRKEVGYCSKNDMLDNILNDAEHNNSQDMYKDKIERLSLDLFVAGTD
TVTSTVEWAMTELLRNPKIMAKAKQELEKTIGKGKIVEESDIGRLPYLEAIVKETFRLHPAAPLLLPRQAEVELEM
HGYTIPKGAQVLVNVWAIGRDPNLWDNPNLFSPERFLGSEIDFKGRSFELIPFGAGRRICPGHPLAIRLLFLMIGLLI
NSFDWELVGGVKPEDMNMDEKFGLTLEKAQPVLVVPIKATNQNIS*
>P_cin_c17_pep 492 aa
MKHSKKGKNNSPPSPRGLPILGNLHQLGSYPHRTLQSLAQKHGPLMLLHLGNSPLLVVSSSAAAREVMKTHDRVF
SDRPHSRIFDVLLYGSKDVAAAPYGEYWRQIRSISVFYLLSSKRVHSFRSVREEEIQTTIQDITTSCTLRQKVSSSSCS
SEDLPVNVSDILSTITSNIICRVALGRKYGGERGREFKKLLLEFTELLGKFTVGDYVPWLGWCDIVSGLYGRANRV
AKEFDEVLNMVIEEHLNREKGDNNNEEHDHFVDVLLWIQRTNAIGFPIDRIVLKALILDIFVAGTDTTSTSLEWAM
TELLRHPIVMKKLQDEITSVVGVDRKHITENDLGHMHYLKAVIKETFRLHPPIPLLVPRQSIQDINLKNNNYHIKAG
TRVIVNAWAIARDPMNWDDPEEFKPERFLNSSIDVKGNDFELIPFGAGRRGCPGIMFAMVVNELVLANLVHQFD
WALLSGDKNLDMSETTGLTMHRKSPLMAVAVPTKK*
>P_cin_c96_pep 651 aa
MASHVALFHAPPLSISSQRFHAKHISINGLKFTTTNSCSSSCSSSSSSCSSSSSFFPCSMTTQRGSCFSVVACSSSNGR
DPKSVKDEDLQQVERILEEKRRAALAAKIASGEFTVKQKSVLRSVLEGLAKVGVPNQVLEFLFSWAEGGGAHPKI
PEAKGSIKAVRSVAFFIPLYELYLTYGGIFRLTFGPKSFLIVSDPSIAKHILRDNAKAYSKGILAEILDFVMGKGLIPA
DGEIWRVRRRAIVPALHQKYVAAMIGLFGKAADRLCQKLDAAASDGEDVEMESLFSRLTLDIIGKAVFNYDFDSL
SNDTGIVEAVYTVLREAEDRSVAPIPVWEIPIWKDISPRLRKVNAALKFINDTLDDLIAICKRMVDEEELQFHEEYM
NEKDPSILHFLLASGDDVSSKQLRDDLMTMLIAGHETSAAVLTWTFYLLSKEPRVVSKLQEEVDSVLGDRFPTIED
MKKLKYTTRVINESLRLYPQPPVLIRRSIEEDVLGDYPIKRGEDIFISIWNLHRSPKLWDDADKFEPERWPLDGPNP
NETNQNFKYLPFGGGQRKCIGDLFASYETVVALAMLVRRFNFQMAVGAPPVDMTTGATIHTTQGLKMTVTRRTK
PPILPSLQMSTLEVDPSISLSDQEEITQKGEVYQAQPQS*
>P_cin_c475_pep 493 aa
MDLDSNSMKVLVVVATAGALLVLRSILKNVNWWLYESKLGEKQHYLPPGDMGWPVIGVMWSFLSAFKTKNPD
SFMASFYSRFGKTGMYKTLMFGNPSVLVTTPETCKRVLTDDEKFEPGWPQSTVDLVGKRSFISMSFEEHRRLRRL
TSASINGFDALSLYLTYIEKNVKISLDKWSKMGEIEFLTEMRRLTFKIIMHIFLSSESEPVMEALEREYTTLNIGVRS
MRINIPGFAYHNALKARKNLVGVFQSIVDERRSLRKGYSKPGKAKDMMDALVDVEDTDGRKLDDETIIDIMLMY
LNAGHESSGHVTMWATYFLQKYPEYMQKAREEQEEILRKRPPTQEGMTLKEFRQMEFLSKVIDETMRLITFSLVV
FRQAKSDVNINGYLIPKGWKALVWFRSVHLDPEIYPNPKEFNPYRWNKEHKVGEFLPFGAGTRLCPGNDLAKLEI
AVFLHHYVMNYRFVPTNPDGPVRFLPHTRPLDNCLGRFIKCPST*
>P_cin_c843_pep 506 aa
MDLLFLEKTLTGLFLAAVVAIIVSKLRGKKFKLPPGPLPVPIFGNWLQVGDDLNHRNLTDLAKKFGDIFLLRMGQR
NLVVVSSPELAKEVLHTQGVEFGSRTRNVVFDIFTGKGQDMVFTVYGEHWRKMRRIMTVPFFTNKVVQQYRQG
WEAEAAAVVEDLKKDPTASTSGSIIRQRLQLMMYNNMYRIMFDRRFESQDDPIFQRLRALNGERSRLAQSFEYNY
GDFIPILRPFLKGYLKICKEVKETRLKLFKDYFVDERKKLGSTKSTNNNNELKCAIDHILDAQKKGEINEDNVLYIV
ENINVAAIETTLWSIEWGIAELVNHPEIQQKVRDEIDRVLGAGKQVTEPDLQNLPYLQAVIKETLRLRMAIPLLVPH
MNLHDAKLGGYDIPAESKILVNAWWLANNPEHWKKPEEFRPERFFEEEKHVEANGNDFRYLPFGVGRRSCPGIIL
ALPILGITLGRLVQNFELLPPPGQSQIDTSEKGGQFSLHILKHSTIVAKPRSF*
>P_cin_c877_pep 487 aa
MEIDGRFLNTGLLIVATILVAKLISAFIVPKSRKRVPXFVKGWPLIGGLIRFLKGPIFMIREEYPKLGSVFSLKLFHKN
ITFLIGPEVSAHFFKAPETDLSQQEVYQFNVPTFGPGVVFDVDYSVRQEQFRFFTEALRVNKLKSYVNQMVAEAEE
YFSKWGPSGEVDLKYELEHLIILTASRCLLGREVRDKLFDDVSALFHDLDNGMLPISVLFPYLPIPAHKRRDQARK
KLAEIFASIIASRKSANKSEEDMLQCFIDSRYKDGRPTTEAEVTGLLIAALFAGQHTSSITSTWTGAYLLSNNQYLSA
VQEEQKMLMEKHGDKVDHDVLAEMDVLYRCIKEALRLHPPLIMLMRSSHTDFSVTTREGKEYDIPKGHIIATSPA
FANRLGHIFKDPDRYDPDRFAVGREEDKVAGAFSYISFGGGRHGCLGEPFAYLQIKAIWSHLLRNFELELVSPFPEI
DWNAMVVGVKGKVMVRYKRRELLVNQ*
>P_cin_c2997_pep 523 aa
MRISNLHDYSSNWFFLSVFTFIILFILRTLLNLFTKWNNNKNSSNSATNSPPSPPKLPIIGNLHQLGTNIHRSLHSLAQ
SYGPSLMLLHFGKVPVLVVSTAESARQVLKTHDAVFANKPHRKMFDILLYGSRDVASAPYGNYWRQTRSICVLH
LLSPKKVQSFCAVRREEIAMMMESIGKCCSSMTPVNLTDLFATVTNDIVCRTALGRRYSGEGGSKLRGPISVMAEL
MGTPVVGDYIPWLDWLGRVNGMYGRAERAAKKVDEFFDEVVDEHVRKGGHDVDDDAKDDDQSDLVDVLLRI
QKTNAMGFEVDRTTIKALIMDMFGAGSETTYTVLGWTMTELLRHPIMMQKLQDEVRNMVSGKSEISEEDLSSMH
YLKAVVKESLRFHPPVSILAPRECMENTKVMGYDIAVGTQVMVNAWAIARDPSYWDNPQEFQPERFLNSSIDVK
GHDFQLIPFGAGRRACPGYTFSIVVIELVIANLVHKFNWKVPTGVVGDQTLDITEAPGLSIQRKFPLIAIATPHT*

Annexe 9 : Séquences peptidiques complètes de cytochromes P450 isolées à partir de la banque d'ADNc de *Cullen cinereum.*

>P_cin_c4264_pep 464 aa
MALYMAVLGVVVLVLCFCSALLKWNEVRFRRKGLPPGTMGWPVFGETTEFLKQGPNFMKNKRARY
GSFFKSHILGCPTVVSMDPELNRYILMNEAKGLVPGYPQSMLDILGKRNIAAVHGSTHKYMRGALLSI
ISPTLIRDQLLPKIDEFMRTHLSHWDNKLINIQEKTKEMAFLSSLKQIAGMESSSISQPFMTEFFKLVLG
TLSLPINLPGTNYRRGLQARKSIISILSQLLEERKASQEAHVDMLGCLMNTDGNRYKLTDDEIIDLIITI
MYSGYETVSTTSMMALKYLHDHPKVLEEIRKEHFSIREKKRPEDPIDCNDLKSMRFTRAVIFETSRLA
TIVNGVLRKTTHDMELNGYLIPKGWRIYVYTREINYDPFLYHDPLTFNPWRWLGNSLESQSHFLIFGG
GTRQCPGKELGIAEISTFLHYFVTRYRWEEVGGDKLMKFPRVVAPNGLHIRVSSY*
>P_cin_5649_pep 526 aa
MDASTALMILSAIAAYLIWFTFVTRSLKGPRVWPLLGSLPGLIQHANRMHDWIADNLRACGGTYQTC
ICALPFLARKQCLVTVTCDPQNLEHILKLRFDNYPKGPTWQSAFHDLLGEGIFNSDGDIWLFQRKTAA
LEFTTRTLRQAMARWVNRAIKHRFCPILATAQNQNNPVDFQDLLLRLTFDNICGLAFGQDPQTLSEGL
PDNDFALSFDRATEATLQRFILPEILWKLKRSLRLGMEGSLSRSLKHIDQYLSHIIKNRKLELLNGTGSL
HDDLLSRFMKKKESYSEEFLQHVALNFILAGRDTSSVALSWFFWLCIKNPHVEENILHELCTVLLSTR
GDNVSTWLEEPLGFEEVDRLVYLKAALSETLRLYPSVPEDSKHVVKDDVLPNGTFVPAGSSITYSIYS
VGRMKFIWGEDCLEFKPERWLSPEGDKIQLQDSYKFVSFNAGPRLCLGKDLAYLQMKSIAAAVLLRH
RLSVAPGHRVEQKMSLTLFMKYGLKVNVYHRDLKPLLEKITSMPTKTVNY*
>P_cin_c6189_pep 498 aa
MVGLFSYLFLSLVFIFTVKSLLQTQLRRLKNLPPGPPTLPIIGNLHHSKQRPHRVYSALSRTYGDIISLW
FGTRLVVVVSSLPLVQECFTKNDTVLANRPRFLAGKYILFNYTTVSAISYGDLWRNLRRVITLDVLSN
HRINSFGGVRRDEALRLLRRLRGDGNGGYKKVELRPGLKEMTFNAMMRSISEMKYTGEGEVTDPED
AKEFRKILSEMLLLYGANNKGDFLPILRLFDFEDLEKRLKRIGKRADAFLQNLIEEHRSGKHGSGDSMI
NHLLKLGESQPQYYSVQIIKGLIQAMFLAGTDTTSEAIEWGMGELLNNPEVLKKAKEEIETQIGIDRLV
DEQDLSKLAYLQNIVNETLRLHPPAPLLLPHESAKDCTVGGYNIPQDTIVLINVWHVHRDPKLWSDPT
SFKPERFEKEGEVNKLITFGIGRRACPGSGLAQRIVGLTMALLIQCFEWKRVSEEKIDMMEDHGGATM
PKLIPVEAMCKPLPIVNGIM*
>P_cin_c7100_pep 513 aa
MSLFTVILATLVTAIIVYRILKFITRPISLPLPPGPKPWPILGNXPHMGPVPHHSLAALARVHGPLMHLR
LGFVDVVVAASASVAEQFLKIHDSNFSSRPPNAGAKYIAYNYQDLVFAPYGPRWRLLRKLTSVHLFS
SRAMDEFTHLRQEEVARLTRNLESSHTKAVNLGQLLNVCTTNALARVVIGRRVFDDGNGGCDSRAD
EFKAMVVELMVLAGVFNIGDFIPFLEWLDLQGVQAKMKKLHKKFDAFLSSIIEEHNSSSKNENNKNL
LSILLSLKDVRDDDGNHLTDTEIKALLLNMFTAGTDTSSSTTEWAMAELIRSPRILAQVREELDNVVG
RDRRVKEEDLTHLPYLQAVVKETFRLHPSXPLSVPRVASDSCEIFGYHIPKGSTLLVNIWAIARDPKE
WADPLEFRPERFLPGGEKADVDIRGNDFEVIPFGAGRRICVGLSLGLRMVQLLTATLAHSFDWELKD
GMNPEKLNMDEAYGLTLQRAVPLSVHPRPRLAPHVYSMSS*
>P_cin_rep_c9422_pep 518 aa
MEAAWATILTLIVSLICAWKVLNWLWLKPKRLEKLLREQGLHGNPYTLFVGDSKEFLKMRKDALSK
PINLSDDIIPRVVSYDQHSVNKHGKNHFIWLGPIPRVTILDPEQIKDVFNKIYDFPKPKMNPLVKLLATG
LAVYGGEKWSKHRRIINPAFNLEKLKIMLPLFFKSCNDLVCKWEAMLSSEGSCELDAWPFLQNLASD
VIARSAFGSSYEEGRRIFQLQREQAELLIRVVLKIQIPGWRFLPTKTHRRMKEIDRDIKSSLNDMINKRE
KALKAGEATKNDLLGILLESNHKEIQEHGNRNNKNAGMSIDEVIEECKLFYFAGQETTSVLLNWTMV
LLSRHLHWQERARDEVLQVFGNQKPDFDGLSRLKIVTMILYEVLRLYPPAPSMTRSVEKDVKLGSITL
PAGVHVFLPTILVHHDSELWGEDAKQFNPERFSEGVLKATNGKVSYFPFGWGPRICIGQNFSLLEAKM
ALSMILQNFSFELSPAYAHAPTALITLQPQYGAHIILRKVTI*
>P_cin_rep_c9447_pep 514 aa
MDSEVLSSLALVVPFFLFMVLALKIGNNLKKAESSPIPNIPPGPWKLPIIGSLHHLITSTPHTKLRDLAKI
YGALMHLQLGEIFTIIVSSPEYAKEIMKTHDLTFASRPHILLAEIACFDSTNIAFSPYGNYWRQVRKICA
AELFSQKRVKSFQPIREDELSNLVKMVSSHNGSPINLTEAVPSSIFNIISRAAFGSKCKHQEEFLSVVKE
AILVASGFEIGDLFPSAKWFQVVTGLRPKIERLRRKVDRILEDIINQHKDKNGQGEADHEDLVDVLLK
FQDVHTTNHISLTTTNIKAIIQDVFTAGGESTAATIDWAMAEMVKNPQVMKKAQVEVREVFDMKGM
VDETCIYQLKYLKSVVKETLRLHPPGPLLLPRECAQTCQIDGYQIPVKSKVIVNAWAIGRDPNYWSEA
ERFHPERFIDSSIDYNGTNFQYIPFGAGRRICPGISLALVNVELTLAFLLYHFDWKLPNGMKSEDLNMT
EDFGLTVKRKDDLYLIPVISRPLQVTKAIYNY*

Annexe 9 : Séquences peptidiques complètes de cytochromes P450 isolées à partir de la banque d'ADNc de *Cullen cinereum.*

>P_cin_rep_c9476_pep 527 aa
MTSCISFSSTDIELYDSQLLLSSLVTIMLVTLVLCLTLPLLFLFFIQHLRAFKTPYLPPGPKGLPIIGNLHQ
LDNSILSQQLWQISKKYGPIFSLQLGLRQAIVVSSSKLAKEVLKNHDLEFSRRPKLLPLQKLSYNGSDI
AFSPYNDEYWREMRKICVANIFSSKCVSRFSSIRESEVKLMIKTILGHASSSGVTNLSELLISLSSTIVCR
VAFGRRYEDEGSERSRFHGLLNELQQLMGTFFVSDYIPLTGWIDKLKGFQARLETNFEEFDKFYQEIID
EHMDPNRQHMEEHDMVDLLLRLKNDRSLSIHITFDHIKGVLMNLLAAGTDTTAAASVWAMTALVK
NPRVMKKVQKEIRNVGATKDFLDEDDVQKLPYLKAMIKETLRLHLPSQLLVPRESSQECIIGGYRIPA
KTIVYMNAWAIQRDHNVWKNPLEFFPERFLDSTIDFRGQDFELIPFGAGRRICPGIPMATVTLELVLAN
LLHSFDWELPQGMTKEDIDLEVLPGITQHKKNHLCLCAKTRSHI*
>P_cin_rep_c9737_pep 527 aa
MAIEFPAPNWVPTAATCLAAFAVLLLFIHRLRRHKYNLPPGPKPWPIIGNLNLIGSLPHQSIHALSKKY
GPIMYMWFGSNPVMVGSSVEMAKAILKTHDATLAGRPKYAAGKHTTYNYSDITWSQYGPYWRQAR
RMCLLELFSAKRLEEYEYIRKEELRGVLNELFNSANKTILLKDHLSNLSLNVISRMVLGKKYLEESEN
AIVSPDEFKKMLDELFLLNGVYNIGDFIPWIDFLDLQGYIKRMKALSKKFDRFMEHVLDEHIARKKGV
KDYVAKDMVDVLLQLAEDPTLEVKLERHGVKAFTQDLIAGGTESSAVTVEWAICELLRKPEIFKKAT
EELDRVIGRERWVEEKDIVNLPYVNAIAKETMRLHPVAPMLVPRLAREDCKIGEYDIPKGTQVLVNV
WTIGRDPSVWDNPNEFQPERFLDKEIDVRGHDYELLPFGAGRRMCPGYPLGLKVIQSSLANLLHGFN
WRLPHNVQKEDLNMDEIFGLSTPKKIPLETVAEPRLPPHLYSL*
>P_cin_rep_c9822_pep 491 aa
MDTVLSTTLLVLLACSIMHVLRSLHSRSKKSNYNLPPGPSLLTIIRNSSQLYKKPQQTMAKLAKTYGPI
MRFTIGQSTTVVISSIEATQEILQTQESSFSDRTNPDITTSYNHNRYSLVFLPVSALWQELRKICHGNLFS
VKTLDASTDLRRKKMQELLTDIRERSLKGEVVDIGRAAFMACINFLSYTFLSMDFVPSVGDGEYKHIV
GTLLKATGTPNMVDYFPMLRVFDPQGIRRHTTNYIDKLFDVLDPLIDERMRKRKEKGYVTNHDMLD
SLLDISDQSSEKIHRKQIKHLFLDLFVAGTDTTAYGLERALTELMHNPEVMVKAKKEIAETIGVGKPV
EESDVARLPYLKAVIKETLRMHPPAPMLLPRKAKSDVEVCGYRIPKGTQILINEWAIGRNPDIWEDAH
VFSPERFLNCDIDVKGRHFKLTPFGSGRRICPGSPLAIRMLHNMLGSLINNFDWKLENNMDPNHMNL
DLSLRAIPLIVNAN*
>P_cin_rep_c9915_pep 528 aa
MENEVVLVLVLISGPLLYLLYLNYTKQKGREALPPSPAFQLPVIGHMHILGPLLHHSFQRLSFKHGPLF
SLRFGSVPCVVASSPHFAKQLLQTNELAFNCRIETAAVKRLTYESSLAFAPYGEYWRFIKKLSMNQLL
GSRSISGFQHLRMQETRHFLGLLARKARAHEPVNVTEELLKLTNNVISRMMLGEAEEARDVVRGVTR
IFGEFNVSDFIWVFKKLDLQGFGKRIEDLFQNFDALVERVISKREEMRREKRKVNGDTLNPQGEDEVR
DFLDILLDCVEDAKSEVKINRVHIKALIMDFFTAGTDTTAISTEWALVELIKNPKLLEEARKEIEKVVG
KERLVEESDCANMPYLQAIVKETFRLHPPVPLVSRRCVTDCIIENYMIPENTLLFVNVWAIGRNPKYW
DNPLEFRPERFLRSGEEEEEEVSSMATGVIDVRGQHFQLLPFGSGRRMCPGVSLAMQEVPALLGAIIQ
CFDFHVIDSKGQVLKGDDIDVSERPGLTAPRAQDLICIPLQTSTHDKILPS*
>P_cin_rep_c9972_pep 523 aa
MDSVLNFLIATTIGLLFIILFFLYLYRPLKSSDDKIKNPPTVAGAWPILGHLPLLSRSKLIHRTLVALAYE
YGPLFTIKIGSKEALVINNWQLAKECFTTNDIVVSSRPKLIASEHLGYNQALFAMAPYGPYWRELRKV
TTSAILSTHRVEQLRHVFVSEVQISIKELYDVWCGKKNESGIPLELKQWFSHLSLNIILRMVAGKRYSS
DTTVDDEKAQRCIGVVKEFLRLFGVFTVGDAIPFLRWFDFGGHEKAMKETAKEFDEIVGEWLEEHRQ
HKALGENNERDFMDVMFSIFDGKAIEGIDADTLIKSTVVTLIVAATDTTSSTLTWAICLILRNPLVLEK
VKKELDTQVGKERSISESDISKLIYLQAIVKETLRLYPTAPLSTPREFTDDCIISGYNVKKGTRLITNLCK
IHTDPNVWSDPLEFKPERFLTTSKDIDIKGHHFELLPFGSGRRICPGISFGLQVVHLTLASILHSFEILNPS
TEPIDMSEAFGVTNSKATPLEVLVKPHLPPSCFQSV*
>P_cin_rep_c10023_pep 520 aa
MLLEVALGLLVLAVFVYLRPTPTAKSKALRHLPNPPSPTRLPLIGHLHLLKDKLLHYALIDLSKKHGP
LFSLYFGSMPTVVASTPELFKLFLQTHEATSFNTRFQTSAIRRLTYDASVAMVPFGPYWKFIRKLIMND
LLNATTVAKLRPLRTQQIRKFLGVMAQSAEAQKPLDVTEELLKWTNSTISMMMLGEAEEIRDIAREV
LKIFGEYSLTDFIWPLKYLKVGKYEKRIDDILNKFDPVVERVIKKRREIVRRRKNGEVIENEASGVFLD
TLLEFAEDETMEIKITKEQIKGLVVDFFSAGTDSTAVATEWALAELINNPRVLQKARDEIYSVVGKDR
LVDEVDTQNLPYIRAIVKETFRMHPPLPVVKRKCTEECEIDGFVIPEGALILFNVWQVGRDPKYWVKP
LEFRPERFLESGGEGEAGPLDLRGQHFQLLPFGSGRRMCPGVNLATSGMATLLASIIQNFDLQVPGPH
GQLLKGDDAKVSMEERGGLTVPRAHSLILVPLARTSVTSKLLS*

Annexe 9 : Séquences peptidiques complètes de cytochromes P450 isolées à partir de la banque d'ADNc de *Cullen cinereum.*

>P_cin_rep_c10482_pep 502 aa
MEAWFLILVSLCLCFLVRAIFLSPIKEKTITTPPGPPNIPIITSFLWIRKSFSELEPILRTLYAKYGPIVTLRI
GHHRSIFICDRALAHKALVHNGSVFSDRPKSLATSRILSSNQHNISSCPYGATWRTLRRNLVSEMLHPS
RVKSFSGIRSSVLHTLIARLKSASESNQSVKVMDHFQHAMFCLLVFMCFGEPLDDGKVRDIERVLRQL
LIGFSRFSVLNFWPSVTRVLFRKRWKLMLRFRREQENVFIPLIRGRKDKRAKEGGEVVAYVDTLLDLE
LPEEDNRKLSDDEMVTLCSELMNAGTDTTSTALQWIMANLVKYPEVQERVVNEIKDNLGERQQNRE
IKEEELQKLPYLKAVILEGLRRHPPGHFVLPHAVTEDVIFNDYLIPKNGTVNFAVAEMGWDPKVWED
PMAFKPDRFLNNEGFDITGSKEIKMMPFGAGRRICPGYNLALLHLEYFVANLVSNFDWKVPQGGDVD
FSEKQEFTVVMKNPLQVHLTPRL*
>P_cin_rep_c10498_pep 482 aa
MEVTNLVVLPAVLAFFFLSLHFITKVFRLRKDPNLNLPPGRLGWPIVGETLEFMRTMVEGNMLKFIQE
RKEKYDSRVFKTSLFGDPIAVFCGPAGNKFLFSNENKNVQVWWPSSVKKLLRLSLANRVGNEAKMM
RRLLMSFLNAETLRNYLPKMDTIAQRHIDKYWEGKEQVFVYSVVQLYTFELACCLFLSIEDIDYISKLS
LNFHELLKGLIGFSLNIPGTRFYRAMKAADLLRKEIKTIMNKRKVELEKKRASPTQDILSHMLVTSDPN
GRFMTEMEIIDNILLLLFAGHDTSRSVLSSVMKYLGELPQVYENALKEQVEISEGKEAGELLQWEDVQ
KMKYSWNVVSEVMRLSPPVSGAYREAIKDFTYADYNIPKGWKLHWNTSSSHMDPELFPNPETFDAS
RFEGSGPAPFSYVPFGGGPRMCLGQEFARLEMLVFLHNIVKRFKWDLVIPDEKFKYDPMLEPEKGLAI
RLHPNSHF*
>P_cin_rep_c10705_pep 516 aa
MVAILSVTTWFQNQNLNLVFALCIIIFTVAWYAWLYFLKPNNIQKLPPGPLGFPIFGNLLSLDPELHSY
FAGLAQIHGPIFKVRLGSKLCMVITSPGLAREVLKDQDTVFANHDVPAAGRAATYGGSDIAWTPYGP
EWRMLRKVCVQKMLSNATLDSVYDLRRNEMRKTVAYLHGRVGNPVKVGEQVFLTMMNVITNMM
WGAAVEGAERESLGAEFRQLVAEITRLLGKPNVSDFFPGLARLDLQGVVREIKRLVWQLDEIFEGMIS
ERRKVEKESIERKDFLQFLLKLVLEGRDSKTPLTITHVKAILMDMVTGGTDTSSNTVEFAMAEMMHK
PDIMKRVQEELEAVVGKDNMVEESHIHKLSYLQAVMKETLRLHPVLPLLIPHYPSENTIVGGYTIPKG
SRIFVNVWAVHRDPTIWENPLEFDPTRFLDAKWDFSGNDFNYFPFGSGRRICAGVAMAERTVLYSLA
TLVHLFDWTIPQGEKLDVIEKFGMVLKKKTPLVAIPTPRLSNPDLYI*
>P_cin_rep_c10717_pep 506 aa
MEAETPALLFLVLIATLVWFLVTTLSKRKSAPASSPPPTTTTIPKAYPIIGSVFSIVANKHRRIHWISDIL
HASPSSTFVLHRTFGSRQVFTANPAVVQHILKTNFPVYHKGPTLNRALGDFLGHGIFNSDGSNWKHQ
RQISSHEFNTRALRKFVETVVDAELSDRLLPLLASAAENKTLIPDFQDILQRFTFDNICKIAFGFDPEYL
LPSLPATSFATAFDDATRISSERFNAAFPLFWKIKSLLNLGSEKRLKDAISEVRSLARRIILEKKKEFQEN
KTLDSVDLLSRFLCSGHSDEDFVMDIVISFILAGRDTTSTALTWFFWLLSQHPHVQDEVLREVAEKDA
HTYDQVKEMVYTQAALCESMRLYPPVPVDTKEAMNDDVLPDGTAVKKGWRVAYHIYAMGRSENI
WGSDWKEFRPERWLSWDKTEGRWKFEGVDGFTPVFQAGPRVCLGKEMAFLQMKRMVAGIIKRFKV
VSQVTEPEYSGYLTSFMIGGFPVRIENRI*
>P_cin_rep_c11200_pep 512 aa
MKGIFSSTFYLSLSFLIALFFLFKLARRTKPKSKSNLNLPPSPPKLPIIGNIHQFGTLPHRSLRDLSLKYGD
MMMLQLGKMPTPTLVVSSVDVAMEIIKTHDLAFSDRPQNTAAKILLYGCTDVGFASYGENWRLKRK
ICVLELLSTKRVQSFRAIREEEVAQLVNKLREAISSGAGYVNLSEMIMSTSNNIVCKCALGRIFSENGY
NSVKELARDAMIHLTAFTVRDYFPRLGWIDVLTGKIQKYKATARAMDALFDLAIEEHLTKKREIGEY
SKRKDLVDILLQLQEDNMVGFHLTKNDIKALLTDMFLGGTDTTSTALEWAMSELARNPMIMKKVQE
EVRRVVGHKSNVEENDITQMHYLKCVIKEILRTHPPTPLLAPRLTMSDVKLKGYDIPAKTTVYINSWA
MLRDSRFWESPEEFLPERFENSEVDFRSQEYFQFIPFGFGRRRCPGMNFAIASIEYVLANLLYWFDWK
LPESNTLKQDIDMSEVFGLVVSKKVPLFLKPKAFTF*
>P_cin_rep_c11261_pep 510 aa
MAIPYEQMKLGSSSIFYLYLSIFITILLLFKLATKTKTKTKTDLNLPPYPPKLPIIGNLHQLGTLPHRSLR
DLSLKCGDTMMLQLGQRQTPTLVVSSAEVAMEIMKTHDLAFSNRTQNIAAKIFLYGCIDVAFGPYGE
SWRQKRKICVLQLLRPKMVQSFGSIRVEEVGEMVDKLRGAGSYVNLSEMLTSTSMNIIFKCIFGRKYS
RDDLKRVKELRGVFVELAGFTLSDYFPLLGWVDVLIGNIRKFEASSEALDAVFDKLIEEHLRAKTHSQ
NKDFVDILLQLQEDDAMLNFELTKNDVKALILDMFAAATDTTSVIMEWAISELMRNPIIMKKVQEEV
RNIVGHKSNIEENDISKMHYLKCIIKETLRLHSPGPLLVPRETTSSVKLNGYDIPAKTMVYINAWTIQR
DPKTWDNPEEFVPERFESIQVDFKGKDFQFIPFGFGRRGCPGMNFGLASVEYMLANLLYWFDWELPK
TDTIKQDIDMDEKFGMVVSKKVSLNVKPVVFSF*

Annexe 9 : Séquences peptidiques complètes de cytochromes P450 isolées à partir de la banque d'ADNc de *Cullen cinereum.*

>P_cin_rep_c11530_pep 518 aa
MDLVINLLNSTTIGVASLILLLCVFLYRPFKSKKEAPTVAGAWPILGHLPLLNGSKTPHRTLGDLADKY
GPIFNIKLGVKKALVISNWEMAKECYTTNDIAVSTLPKLLSAELLCYQGTMILVAPYGSYWRELRRIII
SEFLSPTRVEQRQHVRVSEVRDSIKELVNVGKKNESGYALVELKEWYSLLVFNMILRMVIGKRYFGA
AVALDDKAKRCVKAVDEFVRLAGVFMVADAIPSLRWFDFGGYEKAMRETGKELDDIIGEWLEEHR
QNKNSGENVQDFMNVVLSLLEGKTIEGMNVDTVVKSFVLTIIQAGTEASITTLIWTTCLILKNPHVLE
KIKAELDVQIGKERYISESDLSKLTYLQAVVKETLRLYPPGPLSRPREFAEDCTVGGYTVKKGTRLITN
LSKIHIDPNVWSDPLEFKPERFLTTYKDIDIKGQHYQLLPFGSGRMVCPGINLGLQTVQLTLATFFHSF
DILNPSTEPLDMTESFGVTNTKATPLEILIKPRFPPSFYENM*
>P_cin_rep_c11636_pep 473 aa
MSLPPPSLVKAATTTELPIRQIPGSYGFPLLGPMADRLDYFWFQKPESFFRKRMEKYKSSVFRTNVPPS
FPFFLNVNPNVIAVLDVKSFSHLFDMDLVDKKDVLVGDFVPSLAFTGNLRVGVYQDTTEPQHSKVKS
YSMDILKRSSGIWVEELVKNLDIMCNKIEKSLSNSSSVSYLFPLQQFLFSFLCKTIAGADPSLDPKIAES
GYSMLNSWLALQLLPTVSIGILQPLEEIFLHSFAYPFFLVSSNYNNLYNFIKQQGKDTVKRGQVGFGLT
EEEAIHNLLFVLGFNAYGGFSVFLPSLIDAIATNSALQQKLIKEAREKGGETLTFDSVKEMDLINSIVYE
TLRMNPPVPLQYARARKDFRLSSHDSVFDVKKGELMCGFQKLVMRDPVIFDEPERFKPERFTGEKGA
QLLNYLYWSNGPQTGSPSVSNKQCSAKSVVTLTAALIVAYLFRRYDSIKADATSITALQKAN*
>P_cin_rep_c12157_pep 509 aa
MNCKKVGMEAVAYGVIGLAGLLMVCMVLRKKKAAEGGRPPGPRGLPIIGNLHQLDSSKLNYQLHN
LSKTYGPLVSLRLGFKEAIVVSSRKLAREVLKDHDLDVCTRPPSRGPLRLTYNGLELIFSPYSDHWRET
RKICVVHFFSSKRLSSFSHVRLSEVKRMLEIVRGHVLSSKTTNLSELLMSVSGGIICRVAFGRKYDDEA
AEKSSFHHMLNDAQAMFLSFFVSDYVPFLGWIDNLTGMVSRLENTFKALDLFFQRVIDDHLHPNSLT
HIHDDQDIVHLLLQLKNQGRLSIDLTNDQIKAIILDILVAGTDTTAATSVWVMTALIKNPRVMAKVQE
EITSVCGNKEFLEEKDVQKLVYLKAVIKETLRFYPPTPLVPREAIRSFRLEGYEIKAKTIVYVNGWSIQR
DPGAWKDPEEFYPERFLENEIDFKGQDFELIPFGAGRRICPGISLGIASVELITANLLNSFHWELPQGMK
PEHIDTEGLPGLARHKKNHLCLVAKKINKGL*
>P_cin_rep_c13565_pep 499 aa
MEYLLLVLVITIMSISIHVFISSFKPVKSSKKYPPGPHPFPIIGNILELGKQPHQALAKLSQIYGPLMSLKL
GNTNTIVISSPQVAKEVLQKNDQIFANRMVPDSLRALDHHSFSVVWMAPLAQWRNLRRVCATKVFS
SQKLDSTQVLRQRKVQGLMDYVKERCKKGEALDIGEASFTTVLNSISNTFFSMDLAHYSSDKSQEFK
DIIWGVMEEAGRPNVVDFFPILRLLDPQGARRRMNGYFAKLTAFFDGLIDERLRLRALQNESREFNDV
LDSVLEIMLEDDSQVTRPHVLHLFLDLFVAGIDTTSSTIEWAMAELLNNPEKLEKVRRELKEVIAKGE
QLEESHLSKLPYLQAVVKETFRMHPPIPMLVPHKSEVDFELCGFMVPKSSQILVNVWATGRDSSVWS
NPNEFTPERFLECDTDFKGQDFELIPFGAGRRICPGLPLASRTMHIVLASLLYNYDWKLMDGQKPEDM
DMSEKSGITLHKAQPLLVIPIQA*
>P_cin_rep_c16429_pep 514 aa
METCIVVTLTVIVALTLAWRVLNWLWLRPKRLERLLRDQGLQGNPYRLFVGDLKEIMKMQKEVTSK
PMNLSHDIVPRVFSYLQHTLNKHGKNSFIWFGLKPRVIITEPELIKDVLNKIHDFPKPDTNPLVKLLAT
GLINHEGEKWNKHRRIISPAFNLEKLKNMLPIFFKSCNDLIIKWEEMLLSDGSCEMDVWPFLQNLASD
VISRTAFGSSYEEGRRIFQLLEEQAQLATKVAIKVYVPGWRFLPTATHRRMKEIDREIKASLTDMINHR
EKALKEGGATEDDLLGILLESNHKEIEEHGKNKNIGMSLEDVIEECKLFYFAGQETTSALLVWTMVLL
SRYPDWQARARKEVLQVFGKQKPNFEGISHLNIVTMILNEVLRLYPPAVGLSRNVHRDMKIGNLSLP
AGVQVSLPTTVVHHDCELWGDEANEFNPERFSEGVLKATNGRVSFFPFGWGPRICIGQNFSLLEAKM
ALSTILQRFSFQLSPSYAHAPVTVTTLRPQYGAHVILRKV*
>P_cin_rep_c21518_pep 510 aa
MDSEVLKTLTLTMSFFILIIVSLKLGNNLKKAESSPNIPPGPWKLPIIGNIPHLVTSTPHRKLRDLAKVH
GALMHLQLGEIFTIIVSSAEYAKEIMKTHDLIFASRPKILASDIVSYDSTNIIFSPYGNYWRQLRKICTVE
LFTQKRVNSFKPIREEELNNLVKMIDSHKGSPINLTQAVLLSVFNIISRSAFGMKCKDQEDFISLVKEAI
RLGAGFNIGDLFPSSKWLQHVTGLRPKLERMHRQLDRILEDIINEHKEAKEGHGEADEDLVDVLLKF
QDGNDTDQDISLTINNVKAIILDVFGAGGETSATTINWAMAEMVKNPRVMEKAQVEVREVFELKGR
VDEIGISELKYLKSVVKETLRLHPPAPLLLPRECGQSCEINGYHIPAKSKVIVNAWAIGRDPNYWSEAE
RFYPERFIDSTIDYKGTNFEYIPFGAGRRICPGSTFGLINVELALAFLLYHFDWKLPNGTKSGDLDMTE
QFGATVTRKADLYLIPITSSPFLASVAD*

Annexe 10 : Classification dans le clan 71 des P450s identifiés à partir des trois banques d'ADNc.

P. sativa			*R. graveolens*			*C. cinereum*		
Famille	**Sous-famille (nombre de séquences)**	**Total**	**Famille**	**Sous-famille (nombre de séquences)**	**Total**	**Famille**	**Sous-famille (nombre de séquences)**	**Total**
71	AH (1), AJ(18), AP (1), AQ (6), AS (2), AT (8), AU (4), AZ (10), B (2), BC (7), D (4), J (1), M (1), N (1)	66	**71**	AH (4), AL (1), AN (1), AP (1), AS (5), AT (1), AV (1), B (1), BE (33), BG (2), D (3), W (2)	55	**71**	AH (1), AN (14), AP (1), AQ (2), AR (1), AS (1), BE (13), BF (3), BG (1), D (4)	41
76	A (4), B (5), C (2), E (3), F (6), S (1),	23	**82**	A (1), D (3), H (9), K (1), Q (4), R (1), S	21	**82**	A (10), D (3), J (2), K (2), Q (3)	20
82	H (10), J (1), K (1), L (5), Q (4)	21	**89**	A (19)	19	**83**	E (16), G (3)	19
736	A (16)	16	**706**	D (1), G (9), J (1), K (1)	12	**76**	B (4), E (1), F (5), G (1), X (4)	15
92	A (11), B (3)	14	**81**	C (1), F (1), Q (6), W (2)	10	**81**	E (8), F (2), Q (1)	11
77	A (7), B (4)	11	**83**	A/B (2), E (1), F (7)	10	**93**	A (5), B (1), C (1), E (2), F (1)	10
89	A (7), B (1), C (3)	11	**92**	A (6), B (4)	10	**706**	A (1), B (1), C (1), E (1), J (1)	5
706	A (1), B (1), G (1), J (7)	10	**76**	B (1), C (1), F (1), T (1), Z (1)	5	**78**	A (3), D (2)	5
98	A (9)	9	**73**	A (3)	3	**736**	A (2)	2
73	A (8)	8	**98**	A (3)	3	**73**	A (2)	2
83	E (2), F (4)	6	**712**	A (2)	2	**77**	A (2)	2
78	A (3), D (2)	5	**79**	A (1), D (1)	2	**92**	A (2)	2
81	B (2), H (1), Q (2)	5	**84**	A (2)	2	**75**	B (1)	1
84	A (4)	4	**701**	A (1)	1	**79**	B (1)	1
75	B (3)	3	**703**	A (1)	1	**84**	A (1)	1
80	D (2)	2	**75**	B (1)	1	**89**	A (1)	1
701	C (1)	1						
712	E (1)	1						
79	E (1)	1						

Annexe 11 : Classification dans les autres clans des P450s identifiés à partir des trois banques d'ADNc. SF : Sous-famille.

P. sativa				*R. graveolens*				*C. cinereum*			
Clan	**Famille**	**SF (nombre de séquences)**	**Total**	**Clan**	**Famille**	**SF (nombre de séquences)**	**Total**	**Clan**	**Famille**	**SF (nombre de séquences)**	**Total**
86	94	A (5), B (10), C (1), D (5), F (1)	22	**85**	90	A (10), B (4), C (5), D (1)	20	**85**	90	A (5), C (3)	8
	704	A (5), G (6)	11		707	A (11)	11		716	D (5), G (1)	6
	96	A (4), F (4), K (1)	9		724	B (3)	3		707	A (5)	5
	86	A (4), C (2), F (1)	7		85	A (3)	3		85	A (4)	4
72	72	A (11)	11		88	A (3)	3		722	C (1)	1
	714	A (1), E (4), F (4)	9		718	A (2)	2		88	A (1)	1
	709	E (1)	1		87	C (2)	2	**72**	72	A (16)	16
	715	B (1)	1		708	A (1)	1		721	A (3)	3
	734	A (1)	1		716	E (1)	1		714	A (2)	2
	735	A (1)	1		720	A (1)	1	**86**	86	A (5), B (1)	6
85	707	A (5), C (1)	6		728	E (1)	1		94	A (5)	5
	716	A (1), B (1), D (1), P (1)	4	**72**	72	A (15), D (1)	16		704	A (1), J (1)	2
	87	D (2)	2		714	A (3), E (3), J (5)	11		96	A (1), F (1)	2
	88	A (2)	2		715	A (2)	2	**51**	51	G (1)	1
	90	A (1), B (1)	2		721	B (2)	2	**74**	74	A (1), B (1)	2
	728	E (1)	1	**86**	94	A (1), B (6), D (3), F (2)	12	**97**	97	A (2), B (1), C (4)	7
	85	A (1)	1		704	A (2), B (1), C (2), G (5)	10	**710**	710	A (3)	3
97	97	A (5), B (3), C (1)	9		96	K (2)	2	**85**	90	A (5), C (3)	8
51	51	G (3)	3	**51**	51	G (3)	3		716	D (5), G (1)	6
74	74	A (1), B (1)	2	**74**	74	A (2)	2		707	A (5)	5
710	710	A (2)	2	**97**	97	A (4), B (3), C (1)	8		85	A (4)	4
711	711	A (1)	1								

Annexe 12 : Classification des séquences prédites complètes de P450s isolées à partir des trois banques d'ADNc. P_sat : *P.sativa*, R_gra : *R. graveolens*, P_cin : *C. cinereum*.

Contig	Annotation B2GO	Meilleur hit avec les séquences du Dr. Nelson	Pourcentage d'identité
P_sat_rep_c169	Trans-cinnamate 4-monooxygenase	CYP73A5 *A. thaliana*	84 %
P_sat_rep_c588	cytochrome p450 83b1	CYP71AZ1 *A. majus*	71 %
P_sat_rep_c729	cytochrome p450 83b1	CYP71AZ1 *A. majus*	81 %
P_sat_rep_c1089	cytochrome p450	CYP71D97 *A. majus*	70 %
P_sat_rep_c2505	cytochrome p450 86a2	CYP86A1 *A. thaliana*	63 %
P_sat_rep_c2542	cytochrome p450	CYP71AJ1 *A. majus*	70 %
R_gra_c149	low quality protein: cytochrome p450 71d10-like	CYP71BE1 *Vitis vinifera*	62 %
R_gra_c578	flavonoid 3 -monooxygenase-like	CYP92A1 *Zea mays*	52 %
R_gra_c643	cytochrome p450 71d11-like	CYP71BE1 *Vitis vinifera*	51 %
R_gra_c1526	isoflavone 2 -hydroxylase	CYP81Q7 *Vitis vinifera*	58 %
R_gra_c1542	secologanin synthase-like	CYP72A61 *Glycine max*	50 %
R_gra_c1672	cytochrome p450	CYP82H1 *A. majus*	51 %
R_gra_c1737	cytochrome p450 83b1	CYP83F1 *P. trichocarpus*	56 %
R_gra_c1884	cytochrome p450 86b1	CYP96K1 *G. hirsutum*	46 %
R_gra_c2181	cytochrome p450 83b1	CYP83F1 *P. trichocarpus*	56 %
R_gra_c3135	flavonoid 3 -monooxygenase-like	CYP92A1 *Zea mays*	53 %
R_gra_c3823	secologanin synthase	CYP72A1 *C. roseus*	47 %
R_gra_c4163	cytochrome p450	CYP706G1 *V. vinifera*	50 %
R_gra_c5041	cytochrome p450	CYP82H1 *A. majus*	55 %
R_gra_c6403	secologanin synthase-like	CYP72A61 *Glycine max*	50 %
R_gra_c7149	secologanin synthase-like	CYP72A7 *A. thaliana*	50 %
R_gra_c7779	ent-kaurene oxidase	CYP701A3 *A. thaliana*	66 %
R_gra_c10009	cytochrome p450	CYP82H1 *A. majus*	55 %
R_gra_c12789	cytochrome p450 monooxygenase	CYP90A1 *A. thaliana*	80 %
P_cin_c9	7-ethoxycoumarin o-deethylase-like	CYP76B4 *M. sativa*	57 %
P_cin_c17	cytochrome p450 71a26-like	CYP71BF1 *C. papaya*	52 %

Contig	Annotation B2GO	Meilleur hit avec les séquences du Dr. Nelson	Pourcentage d'identité
P_cin_c96	protein lutein deficient chloroplastic-like	CYP97A3 *A. thaliana*	80 %
P_cin_c475	ent-kaurenoic acid oxidase	CYP88A3 *A. thaliana*	64 %
P_cin_c843	cinnamate 4-hydroxylase	CYP73A5 *A. thaliana*	83 %
P_cin_c877	cytochrome p450 monooxygenase cyp51g1	CYP51G1 *A. thaliana*	81 %
P_cin_c2997	cytochrome p450 71a26-like	CYP71BF1 *C. papaya*	54 %
P_cin_c4264	cytochrome p450 85a-like	CYP85A1 *A. thaliana*	68 %
P_cin_c5649	cytochrome p450 86a2-like	CYP86A1 *A. thaliana*	63 %
P_cin_c6189	cytochrome p450	CYP81E1 *G. echinata*	58 %
P_cin_c7100	flavonoid 3 -hydroxylase	CYP75B1 *A. thaliana*	66 %
P_cin_rep_c9422	secologanin synthase-like	CYP72A61 *G. max*	58 %
P_cin_rep_c9447	cytochrome p450 71d11-like	CYP71BE1 *V. vinifera*	54 %
P_cin_rep_c9476	cytochrome p450 83b1-like	CYP83E1 *M. sativa*	67 %
P_cin_rep_c9737	flavonoid 3 -monooxygenase-like	CYP92A1 *Z. mays*	57 %
P_cin_rep_c9822	7-ethoxycoumarin o-deethylase-like	CYP76X1 *T. pratense*	66 %
P_cin_rep_c9915	licodione synthase	CYP93B1 *G. echinata*	72 %
P_cin_rep_c9972	cytochrome p450 82a4-like	CYP82A6 *P. sativum*	56 %
P_cin_rep_c10023	isoflavone synthase 1	CYP93C4 *G. max*	88 %
P_cin_rep_c10482	cytochrome p450 89a2-like	CYP89A2 *A. thaliana*	52 %
P_cin_rep_c10498	cytochrome p450 716b2-like	CYP716D1 *P. trichocarpus*	56 %
P_cin_rep_c10705	flavonoid 3 -monooxygenase-like	CYP706J1 *V. vinifera*	53 %
P_cin_rep_c10717	cytochrome p450 94a2-like	CYP94A1 *V. sativa*	58 %
P_cin_rep_c11200	cytochrome p450 71a1-like	CYP71AN1 *P. trichocarpus*	54 %
P_cin_rep_c11261	cytochrome p450 71a1-like	CYP71AN1 *P. trichocarpus*	53 %
P_cin_rep_c11530	cytochrome p450	CYP82A6 *P. sativum*	53 %
P_cin_rep_c11636	13-hydroperoxide lyase	CYP74B2 *A. thaliana*	60 %
P_cin_rep_c12157	cytochrome p450 83b1-like	CYP83E1 *M. sativa*	56 %
P_cin_rep_c13565	cytochrome p450 76c4-like	CYP76E1 *M. sativa*	60 %
P_cin_rep_c16429	secologanin synthase-like	CYP72A61 *G. max*	57 %
P_cin_rep_c21518	cytochrome p450 71d11-like	CYP71BE1 *V. vinifera*	55 %

Annexe 13 : Séquences peptidiques complètes de CPRs isolées à partir des trois banques d'ADNc. P_sat : *P. sativa*, R_gra : *R. graveolens*, P_cin : *C. cinereum*. **M** : possibles codons initiateurs. **X** : phénylalanine ou leucine.

>P_sat_rep_c249_pep 709 aa
MQSEEIKLSPFDFMAGFLKRGSGDLTNSSDSVVSLSPSLALLLENRELLMLLTTSFAVLIGCVFVFVWR
RSNDKSSKSFEPVKLVTPKIEPEEEVDDGKKKVTIFFGTQTGTAEGFAKALAEEAKARYEKATFKVID
LDDYAAEDDEYETKFKKETIAFFFLATYGDGEPTDNAARFCKWFSEGKEKGEWLNNLQYGVFGLGN
RQYEHFNKIAKDVDDGLAEQGAKRLVPVGMGDDDQCIEDDFTAWRELVWPELDQLLRDEEDGAVA
TPYTAAVLEYRVVFHDQTDSSLLDRTFSMSNGHAIYDAQHPCRANVYXKRELHTSESDRSCTHLEFDI
SHTGLTYETGDHVGVYTENLIEIVEEAERLLAISPDTYFSIHAEAEDGTPLSGGSLLPPFPPCTLRQALT
KYADLLSSPKKSALLALAAHASDPTEAERLRFLASPAGKDEYAQWIVASQRSLLEVLTEFPSAKPPLG
VFFASVVPRLQPRYYSISSSPRIAPSRIHVTCALVYERTPAGRIHKGVCSTWMKNAVPLEEGNDCSWA
PIFVRQSNFKLPSDSKVPIIMIGPGTGLAPFRGFLQERLALKESGVELGPAVLYFGCRNRKLDFIYNDEL
KNFVETGAMSEMVVAFSREGPTKEYVQHKMSQKAAEIWDMVSQGAYIYVCGDAKGMAKDVHRML
HTIAQEQGGLDSSKAESLVKNLQISGRYLRDVW*
>R_gra_c3549_pep 713 aa
MQSSASEESGMVKVSPLDLMSAIIRGGKVDPSNISSSGSGLEVASIMLENKEFVMILTTSIAVLIGCVIV
FFMRRSSSSKPKKIEPLKPLVVKEREVEVDDGKQKITVFFGTQTGTAEGFAKALADEAKARYDKAIFK
VVDIDDYAAEEDEYEEKLKKENIVFFFLATYGDGEPTDNAARFYKWFTEQKERDGWLQNLKYGVFG
LGNKQYEHFNKIAKVVDEILADQGAKRLVPVGLGDDDDCIEDDFSAWRELVWPELDSLLRDDDDTT
VSTPYTAAISEYRVVFYDSADASIEKKSSGNANGHAVCDAQHPCRSNVAVRKELHTQLSDRSCTHLE
FDIAGTGLTYETGDHVGVYCENLCETVEEALSLLGLSPDTYFSIHTETEDGTPLGGSSLPPPFPPCSLRT
ALSKYADLLGSPKKSALLALAAHASDPFEADRLRHLASPAGKDEYAQWIVASQRSLMEVMSQFPSA
KPPLGVFFAAVVPRLQPRYYSISSSPRVAPSRIHVTCALVYEKTPAGRVHRGVCSTWMKNAVPMEKS
DDCSWAPIFVRQSNFKLPADAKVPIIMIGPGTGLAPFRGFLQERFAMKEAGTELGHSILFFGCRNRQM
DYIYENELNNFVQSGALSELFVAFSREGPTKEYVQHKMMEKASEIWNMLSQGAYLYVCGDAKGMA
KDVHRTLHTIVQEQGSLDGSKAESMV*
>P_cin_c6685_pep 704 aa
MQDSNSIKISPLDLMSAMIKGKFDPSNISSTATATVLLQNRELVMVLTTSIAVLIGCVVVFIWRRSSAP
KPKPLETPKRVVEKLPEIDLDDGTKKVTILFGTQTGTAEGFAKAIAEEAKVRYEKAKFRVVDMDDYA
ADDDEYEEKFKKETHVLFFLATYGDGEPTDNAATFYKWFTEGGEKGEGWLQNLHYGVFGLGNRQY
EHFNKVAKVVDEMLAEQGGKRLVPVGLGDDDQCIEDDFTAWKEELWPALDQLLRDDDGTTVSTPY
TAAVLEYRVVVHDPLEASVDEKKWHNVNGHAIVDAQHPVRANVAVIKELHTPASDRSCTHLEFDIS
GTGVRYETGDHVGVYCENLSEIVEEAIKLIGLPPDTYFSIHSDDEDGKPLSGSSLPPTFPPCTLRTALAR
YADVLSSPKKSALLALAAHASDPSEADRLRHLASPAGKDEYAEWVIASQRSLLEVMAEFPSAKPPIGV
FFAAVAPRLQPRFYSISSSPRMVPNRIHVTCALVHDKMPTGRIHKGVCSTWMKNSVPLEKSQDCSWA
PIFVRTSNFRLPADNKVPIIMIGPGTGLAPFRGFLQERLALKEGGAELGPSVLFFGCRNRQLDYIYEDEL
NHFSNTGALSELIVAFSREGPTKEYVQHKMMEKASEIWSMISQGAYIYVCGDAKGMARDVHRALHTI
LQEQGSLDSSKAESMVKNLQTTGRYLRDVW*

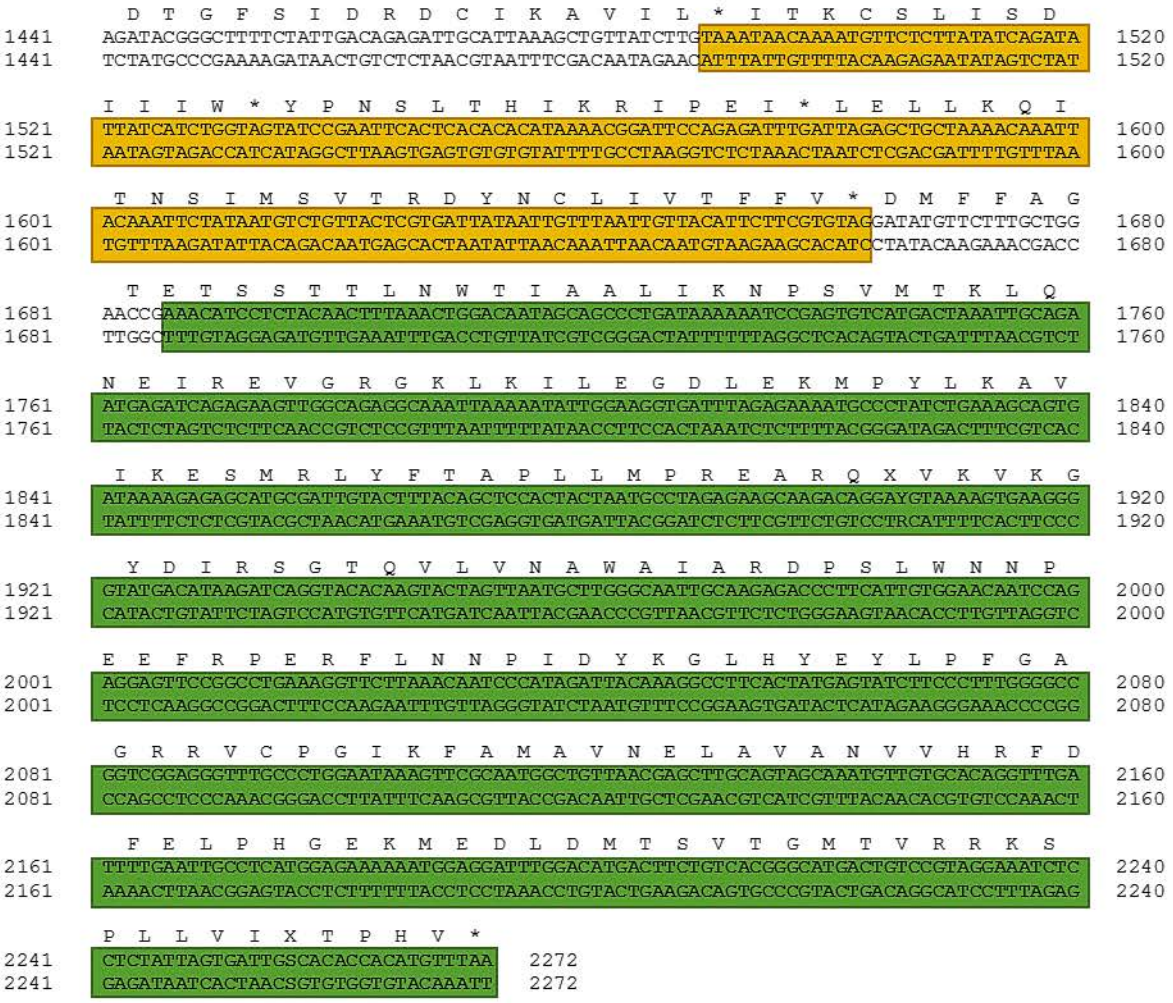
Annexe 14 : Séquence nucléotidique de la région des constructions 1 et 2 du plasmide p*Ye*DP60_GW® contenant la cassette de clonage RfA. Les nucléotides surlignés en jaune correspondent au plasmide p*Ye*DP60, les nucléotides non surlignés correspondent à la cassette de clonage RfA.

```
1       CTATACTTCTATAGACACGCAAACACAAATACACACACTAAATTACCGGATCATCACAAGTTTGTACAAAAAAGCTGAAC   80
1       GATATGAAGATATCTGTGCGTTTGTGTTTATGTGTGTGATTTAATGGCCTAGTAGTGTTCAAACATGTTTTTTCGACTTG   80

81      GAGAAACGTAAAATGATATAAATATCAATATATTAAATTAGATTTTGCATAAAAAACAGACTACATAATACTGTAAAACA   160
81      CTCTTTGCATTTTACTATATTTATAGTTATATAATTTAATCTAAAACGTATTTTTTGTCTGATGTATTATGACATTTTGT   160

161     CAACATATCCAGTCACTATGGCGGCCGCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATAATGTGTGG   240
161     GTTGTATAGGTCAGTGATACCGCCGGCGTAATCCGTGGGGTCCGAAATGTGAAATACGAAGGCCGAGCATATTACACACC   240

241     ATTTTGAGTTAGGATCCGTCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGAAAAAAATCACTGGATATACCACCGT   320
241     TAAAACTCAATCCTAGGCAGCTCTAAAAGTCCTCGATTCCTTCGATTTTACCTCTTTTTTTAGTGACCTATATGGTGGCA   320

321     TGATATATCCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTC   400
321     ACTATATAGGGTTACCGTAGCATTTCTTGTAAAACTCCGTAAAGTCAGTCAACGAGTTACATGGATATTGGTCTGGCAAG   400

401     AGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTATCCGGCCTTTATTCACATTCTTGCC   480
401     TCGACCTATAATGCCGGAAAAATTTCTGGCATTTCTTTTTATTCGTGTTCAAAATAGGCCGGAAATAAGTGTAAGAACGG   480

481     CGCCTGATGAATGCTCATCCGGAATTCCGTATGGCAATGAAAGACGGTGAGCTGGTGATATGGGATAGTGTTCACCCTTG   560
481     GCGGACTACTTACGAGTAGGCCTTAAGGCATACCGTTACTTTCTGCCACTCGACCACTATACCCTATCACAAGTGGGAAC   560

561     TTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCTAC   640
561     AATGTGGCAAAAGGTACTCGTTTGACTTTGCAAAAGTAGCGAGACCTCACTTATGGTGCTGCTAAAGGCCGTCAAAGATG   640

641     ACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTC   720
641     TGTATATAAGCGTTCTACACCGCACAATGCCACTTTTGGACCGGATAAAGGGATTTCCCAAATAACTCTTATACAAAAAG   720

721     GTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTTT   800
721     CAGAGTCGGTTAGGGACCCACTCAAAGTGGTCAAAACTAAATTTGCACCGGTTATACCTGTTGAAGAAGCGGGGGCAAAA   800

801     CACCATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTCAGGTTCATCATGCCGTTTGTGATG   880
801     GTGGTACCCGTTTATAATATGCGTTCCGCTGTTCCACGACTACGGCGACCGCTAAGTCCAAGTAGTACGGCAAACACTAC   880

881     GCTTCCATGTCGGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAAACGCGTGGATCC   960
881     CGAAGGTACAGCCGTCTTACGAATTACTTAATGTTGTCATGACGCTACTCACCGTCCCGCCCCGCATTTGCGCACCTAGG   960

961     GGCTTACTAAAAGCCAGATAACAGTATGCGTATTTGCGCGCTGATTTTTGCGGTATAAGAATATATACTGATATGTATAC   1040
961     CCGAATGATTTTCGGTCTATTGTCATACGCATAAACGCGCGACTAAAAACGCCATATTCTTATATATGACTATACATATG   1040

1041    CCGAAGTATGTCAAAAAGAGGTATGCTATGAAGCAGCGTATTACAGTGACAGTTGACAGCGACAGCTATCAGTTGCTCAA   1120
1041    GGCTTCATACAGTTTTTCTCCATACGATACTTCGTCGCATAATGTCACTGTCAACTGTCGCTGTCGATAGTCAACGAGTT   1120

1121    GGCATATATGATGTCAATATCTCCGGTCTGGTAAGCACAACCATGCAGAATGAAGCCCGTCGTCTGCGTGCCGAACGCTG   1200
1121    CCGTATATACTACAGTTATAGAGGCCAGACCATTCGTGTTGGTACGTCTTACTTCGGGCAGCAGACGCACGGCTTGCGAC   1200

1201    GAAAGCGGAAAATCAGGAAGGGATGGCTGAGGTCGCCCGGTTTATTGAAATGAACGGCTCTTTTGCTGACGAGAACAGGG   1280
1201    CTTTCGCCTTTTAGTCCTTCCCTACCGACTCCAGCGGGCCAAATAACTTTACTTGCCGAGAAAACGACTGCTCTTGTCCC   1280

1281    GCTGGTGAAATGCAGTTTAAGGTTTACACCTATAAAAGAGAGAGCCGTTATCGTCTGTTTGTGGATGTACAGAGTGATAT   1360
1281    CGACCACTTTACGTCAAATTCCAAATGTGGATATTTTCTCTCTCGGCAATAGCAGACAAACACCTACATGTCTCACTATA   1360

1361    TATTGACACGCCCGGGCGACGGATGGTGATCCCCCTGGCCAGTGCACGTCTGCTGTCAGATAAAGTCTCCCGTGAACTTT   1440
1361    ATAACTGTGCGGGCCCGCTGCCTACCACTAGGGGGACCGGTCACGTGCAGACGACAGTCTATTTCAGAGGGCACTTGAAA   1440

1441    ACCCGGTGGTGCATATCGGGGATGAAAGCTGGCGCATGATGACCACCGATATGGCCAGTGTGCCGGTCTCCGTTATCGGG   1520
1441    TGGGCCACCACGTATAGCCCCTACTTTCGACCGCGTACTACTGGTGGCTATACCGGTCACACGGCCAGAGGCAATAGCCC   1520

1521    GAAGAAGTGGCTGATCTCAGCCACCGCGAAAATGACATCAAAAACGCCATTAACCTGATGTTCTGGGGAATATAAATGTC   1600
1521    CTTCTTCACCGACTAGAGTCGGTGGCGCTTTTACTGTAGTTTTTGCGGTAATTGGACTACAAGACCCCTTATATTTACAG   1600

1601    AGGCTCCCTTATACACAGCCAGTCTGCAGGTCGACCATAGTGACTGGATATGTTGTGTTTTACAGTATTATGTAGTCTGT   1680
1601    TCCGAGGGAATATGTGTCGGTCAGACGTCCAGCTGGTATCACTGACCTATACAACACAAAATGTCATAATACATCAGACA   1680

1681    TTTTTATGCAAAATCTAATTTAATATATTGATATTTATATCATTTTACGTTTCTCGTTCAGCTTTCTTGTACAAAGTGGT   1760
1681    AAAAATACGTTTTAGATTAAATTATATAACTATAAATATAGTAAAATGCAAAGAGCAAGTCGAAAGAACATGTTTCACCA   1760

1761    GATGATCCATCGATGCATCGATGCCCCGGGTACCGAGCTCGAATTCGCGGGGGATCTCCCATGTCTCTACTGGTGGTGGT   1840
1761    CTACTAGGTAGCTACGTAGCTACGGGGCCCATGGCTCGAGCTTAAGCGCCCCCTAGAGGGTACAGAGATGACCACCACCA   1840

1841    GCT   1843
1841    CGA   1843
```

Annexe 15 - A : Séquences nucléotidiques et peptidiques obtenues par une réaction de TAIL-PCR grâce à l'amorce dégénérée AD5 sur le contig P_sat_rep_c29131. En jaune : intron, en vert : séquence correspondant au contig P_sat_rep_c29131, en bleu : position du codon intiateur.

```
          G  C  Y  C  P  V  *  S  C  E  D  V  S  V  P  A  K  N  I  S  Y  T  K  N  V  T
1     GGGCTGCTATTGTCCAGTTTAAAGTTGTGAGGATGTTTCGGTTCCAGCAAAGAACATATCCTACACGAAGAATGTAACAG   80
1     CCCGACGATAACAGGTCAAATTTCAACACTCCTACAAAGCCAAGGTCGTTTCTTGTATAGGATGTGCTTCTTACATTGTC   80

      V  K  Q  L  *  S  R  V  T  D  I  I  E  F  I  N  C  F  R  Y  V  I  L  S  *  P  L
81    TTAAACAATTATAATCACGAGTAACAGACATTATAGAATTTATAAATTGCTTTAGATATGTTATATTGTCGTAACCACTT  160
81    AATTTGTTAATATTAGTGCTCATTGTCTGTAATATCTTAAATATTTAACGAAATCTATACAATATAACAGCATTGGTGAA  160

       E  I  *  L  F  *  T  I  S  *  K  P  G  A  G  H  G  P  A  P  S  T  Q  Q  V  V  V
161   GAAATTTGACTTTTTTAAACGATATCATGAAAGCCCGGGGCGGGACACGGCCCGGCTCCGAGTACTCAACAGGTTGTGGT  240
161   CTTTAAACTGAAAAAATTTGCTATAGTACTTTCGGGCCCCGCCCTGTGCCGGGCCGAGGCTCATGAGTTGTCCAACACCA  240

        E  K  S  *  A  T  I  Q  L  V  Y  N  Y  Y  V  *  T  P  L  K  V  H  S  R  E  S
241   AGAAAAATCTTGAGCTACAATTCAACTGGTATATAATTATTATGTGTAGACGCCATTAAAAGTTCATTCAAGAGAATCAA  320
241   TCTTTTTAGAACTCGATGTTAAGTTGACCATATATTAATAATACACATCTGCGGTAATTTTCAAGTAAGTTCTCTTAGTT  320

      K  N  Y  Q  E  G  R  Q  G  L  S  H  E  *  I  Y  L  S  C  R  V  L  E  L  Y  V  *
321   AAAATTATCAAGAAGGCCGGCAAGGTTTGAGCCATGAATGAATATACCTTTCTTGTAGGGTTTTGGAATTATATGTATAA  400
321   TTTTAATAGTTCTTCCGGCCGTTCCAAACTCGGTACTTACTTATATGGAAAGAACATCCCAAAACCTTAATATACATATT  400

       F  D  K  K  K  P  C  *  V  R  A  K  P  K  P  L  L  F  F  C  Y  F  I  *  L  L  D
401   TTTGACAAAAAAAAAACCATGCTAAGTACGGGCGAAGCCTAAACCTCTTCTTTTTTTTTGTTATTTTATCTAGTTATTAGA  480
401   AAACTGTTTTTTTTTGGTACGATTCATGCCCGCTTCGGATTTGGAGAAGAAAAAAAAACAATAAAATAGATCAATAATCT  480

        F  F  Y  P  I  C  G  C  K  L  F  Q  T  L  D  V  E  L  L  E  S  F  G  V  Y  L
481   TTTTTTTTACCCCATTTGTGGTTGCAAACTGTTTCAGACTTTGGATGTAGAGTTGCTAGAAAGTTTTGGGGTGTATCTTT  560
481   AAAAAAAATGGGGTAAACACCAACGTTTGACAAAGTCTGAAACCTACATCTCAACGATCTTTCAAAACCCCACATAGAAA  560

      Y  L  G  R  V  N  L  L  Y  S  K  T  M  P  E  L  Y  S  L  Y  L  C  F  F  S  L  F
561   ATTTAGGCAGAGTTAACCTTCTGTATAGCAAGACCATGCCGGAGCTATATTCCCTTTATCTATGTTTCTTTTCGTTGTTC  640
561   TAAATCCGTCTCAATTGGAAGACATATCGTTCTGGTACGGCCTCGATATAAGGGAAATAGATACAAAGAAAAGCAACAAG  640

       F  V  T  V  F  L  Y  K  W  L  T  T  G  K  T  A  L  K  K  L  P  P  S  P  R  K  F
641   TTTGTCACAGTTTTTCTCTATAAATGGTTAACAACTGGAAAAACAGCCTTGAAAAAATTGCCACCTTCTCCTCGTAAATT  720
641   AAACAGTGTCAAAAAGAGATATTTACCAATTGTTGACCTTTTTGTCGGAACTTTTTTAACGGTGGAAGAGGAGCATTTAA  720

        P  I  I  G  N  L  H  Q  I  G  P  D  P  H  I  S  L  G  A  L  A  Q  K  Y  G  P
721   TCCTATAATCGGAAACCTTCATCAAATTGGTCCTGACCCTCACATTTCACTCGGGGCCTTGGCTCAAAAGTATGGCCCGT  800
721   AGGATATTAGCCTTTGGAAGTAGTTTAACCAGGACTGGGAGTGTAAAGTGAGCCCCGGAACCGAGTTTTCATACCGGGCA  800

      L  M  L  L  K  F  G  S  V  P  V  L  V  V  S  S  A  D  A  A  R  E  I  M  K  T  H
801   TGATGCTACTCAAGTTTGGAAGTGTACCTGTACTTGTTGTTTCTTCAGCTGATGCAGCTCGAGAGATTATGAAAACGCAT  880
801   ACTACGATGAGTTCAAACCTTCACATGGACATGAACAACAAAGAAGTCGACTACGTCGAGCTCTCTAATACTTTTGCGTA  880

       D  L  V  F  A  D  R  P  I  T  S  I  T  D  R  I  F  Y  N  S  R  D  V  G  F  A  R
881   GATTTGGTTTTTGCAGATAGACCTATTACAAGTATTACTGATAGAATTTTCTATAACAGCAGGGACGTGGGTTTTGCTCG  960
881   CTAAACCAAAAACGTCTATCTGGATAATGTTCATAATGACTATCTTAAAAGATATTGTCGTCCCTGCACCCAAAACGAGC  960

        Y  S  E  Y  W  R  Q  V  K  S  T  C  V  I  Q  L  L  S  S  K  R  V  H  S  F  R
961   TTATAGTGAGTATTGGAGACAAGTGAAGAGTACATGTGTTATTCAGCTTCTAAGTAGTAAAAGGGTTCATTCATTTCGGA 1040
961   AATATCACTCATAACCTCTGTTCACTTCTCATGTACACAATAAGTCGAAGATTCATCATTTTCCCAAGTAAGTAAAGCCT 1040

      N  V  R  E  E  E  V  D  L  L  I  Q  S  I  E  Y  S  H  S  K  I  V  N  L  S  D  L
1041  ATGTCAGGGAAGAAGAAGTTGACCTTTTGATCCAGAGTATTGAATATTCTCACTCGAAAATAGTTAATTTGAGTGATTTG 1120
1041  TACAGTCCCTTCTTCTTCAACTGGAAAACTAGGTCTCATAACTTATAAGAGTGAGCTTTTATCAATTAAACTCACTAAAC 1120

       L  S  E  L  T  N  N  V  V  C  R  V  A  L  G  K  K  Y  G  N  D  K  G  N  S  Y  K
1121  TTATCGGAGTTAACAAATAATGTAGTCTGCAGGGTTGCACTTGGGAAGAAATATGGAAATGATAAAGGGAATTCGTATAA 1200
1121  AATAGCCTCAATTGTTTATTACATCAGACGTCCCAACGTGAACCCTTCTTTATACCTTTACTATTTCCCTTAAGCATATT 1200

        Y  L  I  G  E  L  N  K  L  L  G  Y  F  S  S  I  G  E  Y  I  P  L  L  Y  W  V
1201  GTACTTGATTGGGGAATTGAATAAGCTGTTAGGTTACTTCAGCAGCATAGGGGAATACATTCCGTTGCTTTATTGGGTTG 1280
1201  CATGAACTAACCCCTTAACTTATTCGACAATCCAATGAAGTCGTCGTATCCCCTTATGTAAGGCAACGAAATAACCCAAC 1280

      D  C  L  R  G  L  R  G  R  V  E  K  V  A  D  E  V  D  A  F  L  E  S  V  L  G  D
1281  ATTGCCTTAGAGGTTTAAGGGGAAGGGTGGAAAAAGTAGCTGATGAGGTTGATGCTTTTCTGGAGAGTGTTCTTGGAGAT 1360
1281  TAACGGAATCTCCAAATTCCCCTTCCCACCTTTTTCATCGACTACTCCAACTACGAAAAGACCTCTCACAAGAACCTCTA 1360

       H  R  L  G  E  S  D  N  D  D  A  N  K  D  F  V  S  I  L  L  E  I  Q  K  Q  N  T
1361  CATCGACTGGGTGAGTCGGATAATGATGATGCAAACAAAGACTTTGTATCCATTTTGCTTGAAATTCAAAAACAGAACAC 1440
1361  GTAGCTGACCCACTCAGCCTATTACTACTACGTTTGTTTCTGAAACATAGGTAAAACGAACTTTAAGTTTTTGTCTTGTG 1440
```

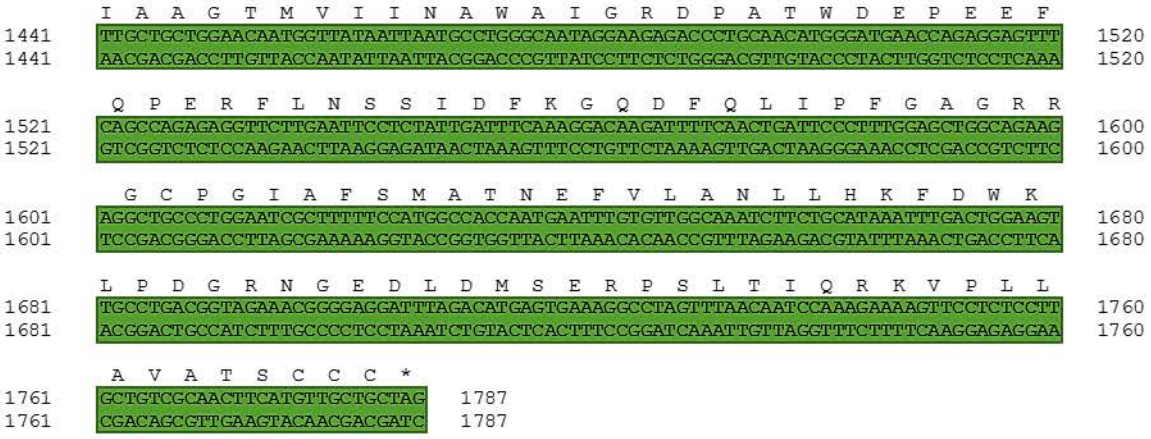
Annexe 15 – B : Séquences nucléotidiques et peptidiques obtenues par une réaction de TAIL-PCR grâce à l'amorce dégénérée AD5 sur le contig P_sat_rep_c29131. En jaune : intron, en vert : séquence correspondant au contig P_sat_rep_c29131, en bleu : position du codon intiateur.



Annexe 16 – A : Séquences nucléotidiques et peptidiques obtenues par une réaction de TAIL-PCR grâce à l'amorce dégénérée AD4 sur le contig P_sat_rep_c14511. En jaune : intron, en vert : séquence correspondant au contig P_sat_rep_c29131, en bleu : positions potentielles du codon initiateur.

```
          W  R  T  R  D  K  *  P  K  G  K  K  Y  E  M  M  V  F  L  H  V  L  Q  Q  E  D
1         TGTGGAGGACCAGAGACAAGTGACCCAAAGGAAAGAAGTATGAAATGATGGTTTTTCTACATGTACTACAGCAAGAAGAT   80
1         ACACCTCCTGGTCTCTGTTCACTGGGTTTCCTTTCTTCATACTTTACTACCAAAAAGATGTACATGATGTCGTTCTTCTA   80

          A  R  M  L  L  Q  H  L  F  L  A  I  L  P  L  L  L  L  F  L  V  K  W  I  F  F
81        GCAAGAATGTTGTTGCAACATCTGTTCCTTGCTATACTACCATTACTCCTCTTACTCTTCCTTGTCAAATGGATATTCTT  160
81        CGTTCTTACAACAACGTTGTAGACAAGGAACGATATGATGGTAATGAGGAGAATGAGAAGGAACAGTTTACCTATAAGAA  160

            N  A  P  V  S  N  K  N  L  P  P  S  P  A  R  F  P  V  L  G  N  F  H  Q  L  G
161       CAATGCCCCAGTAAGCAACAAAAACTTGCCACCTTCACCAGCAAGATTCCCAGTACTAGGAAACTTCCATCAACTAGGTT  240
161       GTTACGGGGTCATTCGTTGTTTTTGAACGGTGGAAGTGGTCGTTCTAAGGGTCATGATCCTTTGAAGGTAGTTGATCCAA  240

          L  L  P  R  R  N  L  Q  S  L  A  R  K  H  G  S  V  M  L  L  H  L  G  S  V  P  T
241       TGCTTCCGCGCCGAAATCTCCAGTCTTTGGCTCGAAAACATGGCTCCGTCATGCTTCTTCACCTTGGTAGTGTGCCTACA  320
241       ACGAAGGCGCGGCTTTAGAGGTCAGAAACCGAGCTTTTGTACCGAGGCAGTACGAAGAAGTGGAACCATCACACGGATGT  320

            L  V  L  S  S  A  D  A  A  R  E  V  M  K  I  Q  D  I  F  F  C  D  R  P  E  S  S
321       CTTGTGCTCTCCTCTGCTGATGCAGCTCGCGAAGTCATGAAAATTCAGGACATTTTCTTTTGTGACAGGCCTGAATCGAG  400
321       GAACACGAGAGGAGACGACTACGTCGAGCGCTTCAGTACTTTTAAGTCCTGTAAAAGAAAACACTGTCCGGACTTAGCTC  400

            V  G  R  R  L  L  Y  D  M  K  D  I  S  A  A  L  Y  G  E  Y  W  R  Q  L  K  S
401       TGTTGGCAGACGATTGCTCTATGACATGAAGGATATATCTGCAGCGCTTTATGGAGAGTACTGGAGGCAGCTGAAAAGCA  480
401       ACAACCGTCTGCTAACGAGATACTGTACTTCCTATATAGACGTCGCGAAATACCTCTCATGACCTCCGTCGACTTTTCGT  480

          I  F  I  L  Q  I  A  S  N  K  R  V  Q  S  F  R  K  V  R  E  E  E  T  A  L  M  V
481       TCTTTATTCTCCAGATAGCAAGCAACAAAAGGGTACAGTCATTTCGTAAAGTAAGAGAGGAAGAAACTGCTCTTATGGTG  560
481       AGAAATAAGAGGTCTATCGTTCGTTGTTTTCCCATGTCAGTAAAGCATTTCATTCTCTCCTTCTTTGACGAGAATACCAC  560

            N  N  I  K  D  L  S  S  S  S  L  P  V  D  L  S  E  L  F  M  T  L  T  N  D  V  S
561       AACAATATTAAAGACTTGTCTTCTTCCTCATTACCCGTGGATCTCAGCGAGTTGTTCATGACACTAACTAATGACGTATC  640
561       TTGTTATAATTTCTGAACAGAAGAAGGAGTAATGGGCACCTAGAGTCGCTCAACAAGTACTGTGATTGATTACTGCATAG  640

            C  R  S  A  F  G  R  K  Y  S  E  G  G  S  G  R  E  F  K  K  L  L  R  E  F  L
641       CTGTAGATCTGCGTTCGGAAGGAAGTACAGTGAAGGAGGAAGTGGGAGAGAATTCAAAAAACTATTGAGGGAGTTTTTGG  720
641       GACATCTAGACGCAAGCCTTCCTTCATGTCACTTCCTCCTTCACCCTCTCTTAAGTTTTTTGATAACTCCCTCAAAAACC  720

          E  L  L  G  G  L  S  F  G  D  F  I  P  W  L  R  W  V  D  R  I  S  G  L  D  A  G
721       AGTTACTAGGTGGTCTAAGTTTTGGGGATTTTATTCCATGGCTTAGATGGGTCGATCGAATTAGCGGTTTGGATGCTGGA  800
721       TCAATGATCCACCAGATTCAAAACCCCTAAAATAAGGTACCGAATCTACCCAGCTAGCTTAATCGCCAAACCTACGACCT  800

            L  D  R  V  S  K  Q  L  D  D  F  L  Q  G  V  V  Q  E  H  I  Q  L  G  Q  S  N  I
801       CTGGACAGAGTTTCTAAGCAACTTGATGACTTTCTCCAAGGTGTTGTTCAAGAACACATACAACTTGGTCAAAGCAACAT  880
801       GACCTGTCTCAAAGATTCGTTGAACTACTGAAAGAGGTTCCACAACAAGTTCTTGTGTATGTTGAACCAGTTTCGTTGTA  880

            H  T  E  H  K  E  D  F  V  D  I  L  L  R  I  Q  K  E  T  T  H  G  I  S  I  D
881       TCATACTGAACATAAAGAAGACTTTGTTGATATCTTACTTAGGATTCAGAAGGAAACTACTCATGGGATATCAATCGACA  960
881       AGTATGACTTGTATTTCTTCTGAAACAACTATAGAATGAATCCTAAGTCTTCCTTTGATGAGTACCCTATAGTTAGCTGT  960

          N  D  S  V  K  A  I  L  L  V  R  T  N  *  E  L  F  H  S  S  L  K  N  K  C  M  L
961       ATGACAGTGTCAAGGCTATTCTCTTGGTAAGAACTAATTGAGAATTATTTCATTCTTCTTTAAAGAATAAATGCATGTTA  1040
961       TACTGTCACAGTTCCGATAAGAGAACCATTCTTGATTAACTCTTAATAAAGTAAGAAGAAATTTCTTATTTACGTACAAT  1040

            S  L  V  S  W  *  L  K  D  D  G  S  N  M  G  M  *  V  F  M  I  P  Q  L  I  I  S
1041      TCCCTAGTCTCTTGGTAATTAAAGGATGATGGCTCAAACATGGGCATGTGAGTGTTTATGATTCCGCAACTAATCATTTC  1120
1041      AGGGATCAGAGAACCATTAATTTCCTACTACCGAGTTTGTACCCGTACACTCACAAATACTAAGGCGTTGATTAGTAAAG  1120

            R  Q  K  Y  V  *  V  H  L  N  I  N  Y  *  F  L  F  *  D  I  Y  S  A  G  T  D
1121      CAGACAAAAGTATGTGTGAGTGCATCTTAATATCAATTACTGATTTTTGTTCTAGGATATATTCTGCTGGAACTGATA  1200
1121      GTCTGTTTTCATACACACTCACGTAGAATTATAGTTAATGACTAAAAACAAGATCCTATATATAAGACGACCTTGACTAT  1200

          T  T  A  T  A  L  E  W  A  M  S  E  L  L  R  H  P  S  A  M  N  T  V  Q  T  E  I
1201      CCACAGCTACAGCCCTAGAATGGCAATGTCCGAACTCTTGAGACATCCATCTGCTATGAATACTGTGCAAACGGAGATC  1280
1201      GGTGTCGATGTCGGGATCTTACCCGTTACAGCCTTGAGAACTCTGTAGGTAGACGATACTTATGACACGTTTGCCTCTAG  1280

            R  E  V  L  G  S  R  P  D  I  T  D  D  D  L  E  K  M  Q  Y  L  R  A  V  I  K  E
1281      AGAGAAGTTCTTGGTAGCAGACCTGATATAACAGACGATGATTTGGAGAAAATGCAGTATCTGAGAGCAGTGATCAAAGA  1360
1281      TCTCTTCAAGAACCATCGTCTGGACTATATTGTCTGCTACTAAACCTCTTTTACGTCATAGACTCTCGTCACTAGTTTCT  1360

            T  L  R  L  H  P  P  I  P  L  L  V  P  R  S  A  R  E  D  V  K  V  N  G  Y  D
1361      AACTCTCCGTCTGCATCCACCAATTCCGTTGTTGGTTCCAAGATCAGCACGCGAAGATGTAAAAGTAAATGGCTACGACA  1440
1361      TTGAGAGGCAGACGTAGGTGGTTAAGGCAACAACCAAGGTTCTAGTCGTGCGCTTCTACATTTTCATTTACCGATGCTGT  1440
```

Annexe 16 – B : Séquences nucléotidiques et peptidiques obtenues par une réaction de TAIL-PCR grâce à l'amorce dégénérée AD4 sur le contig P_sat_rep_c14511. En jaune : intron, en vert : séquence correspondant au contig P_sat_rep_c29131, en bleu : positions potentielles du codon initiateur.



Annexe 17 - A : Séquences nucléotidiques et peptidiques obtenues par des réactions de TAIL-PCR grâce à l'amorce dégénérée AD3 sur le contig P_sat_rep_c5642. En jaune : intron, en vert : séquence correspondant au contig P_sat_rep_c5642.

```
          V  E  F  G  V  G  T  F  T  P  W  L  A  W  I  N  T  V  S  G  F  N  G  K  V  K  N
1         GTCGAGTTTGGTGTTGGAACCTTTACTCCATGGCTTGCATGGATCAATACTGTTTCAGGTTTCAATGGCAAAGTAAAAAA   80
1         CAGCTCAAACCACAACCTTGGAAATGAGGTACCGAACGTACCTAGTTATGACAAAGTCCAAAGTTACCGTTTCATTTTTT   80

            I  A  K  E  N  D  E  F  F  E  R  I  V  E  E  H  I  N  A  S  T  R  K  K  K  G
81        CATTGCAAAAGAGAACGATGAATTTTTCGAACGGATTGTGGAAGAGCACATAAATGCTTCAACAAGAAAGAAGAAAGGGG  160
81        GTAACGTTTTCTCTTGCTACTTAAAAAGCTTGCCTAACACCTTCTCGTGTATTTACGAAGTTGTTCTTTCTTCTTTCCCC  160

          G  E  K  I  K  E  T  E  V  F  L  D  I  L  L  G  A  Q  E  D  A  I  G  G  F  H  I
161       GTGAGAAAATTAAAGAAACAGAGGTTTTTCTCGATATCTTGCTTGGAGCACAAGAGGATGCCATCGGTGGCTTTCATATT  240
161       CACTCTTTTAATTTCTTTGTCTCCAAAAAGAGCTATAGAACGAACCTCGTGTTCTCCTACGGTAGCCACCGAAAGTATAA  240

          D  R  D  G  I  K  A  L  I  L  V  N  S  K  T  L  H  F  K  *  H  D  L  K  F  R  F
241       GATAGAGATGGCATCAAAGCTCTGATCCTGGTAAATTCTAAAACACTTCATTTCAAATGACATGATTTGAAATTCAGGTT  320
241       CTATCTCTACCGTAGTTTCGAGACTAGGACCATTTAAGATTTTGTGAAGTAAAGTTTACTGTACTAAACTTTAAGTCCAA  320

            P  T  D  H  N  I  F  L  S  V  K  Y  I  H  I  Q  N  C  *  Y  F  M  K  L  G  V
321       CCCAACGGACCATAACATCTTTCTGTCTGTCAAATATATTCATATTCAGAATTGTTGATACTTCATGAAACTCGGAGTTA  400
321       GGGTTGCCTGGTATTGTAGAAAGACAGACAGTTTATATAAGTATAAGTCTTAACAACTATGAAGTACTTTGAGCCTCAAT  400

          N  L  *  T  N  N  L  H  S  S  S  A  T  N  F  F  K  N  K  V  H  H  C  N  L  S  L
401       ATTTATAGACCAATAACTTACATTCTTCTTCTGCCACAAATTTCTTTAAGAATAAGGTGCATCATTGCAATCTTTCACTC  480
401       TAAATATCTGGTTATTGAATGTAAGAAGAAGACGGTGTTTAAAGAAATTCTTATTCCACGTAGTAACGTTAGAAAGTGAG  480

          *  G  I  D  M  H  L  F  K  K  I  S  T  H  R  P  N  F  *  C  I  T  I  I  V  T  Q
481       TGAGGAATTGACATGCATCTTTTCAAAAAAATTAGTACCCACAGACCTAACTTTTGATGTATCACCATCATTGTAACACA  560
481       ACTCCTTAACTGTACGTAGAAAAGTTTTTTTTAATCATGGGTGTCTGGATTGAAAACTACATAGTGGTAGTAACATTGTGT  560

            D  M  I  A  G  G  T  D  T  T  Y  T  A  L  E  W  T  M  T  E  L  L  R  H  P  K
561       GGATATGATTGCTGGTGGAACTGACACTACCTACACAGCTCTAGAATGGACAATGACGGAACTTTTAAGGCACCCGAAAG  640
561       CCTATACTAACGACCACCTTGACTGTGATGGATGTGTCGAGATCTTACCTGTTACTGCCTTGAAAATTCCGTGGGCTTTC  640

          A  M  K  E  V  Q  N  E  I  R  N  I  C  R  G  K  S  N  I  S  Y  P  D  D  L  E  N
641       CCATGAAGGAAGTGCAAAATGAGATAAGGAATATTTGCAGAGGAAAATCTAACATATCTTATCCAGATGACTTGGAAAAC  720
641       GGTACTTCCTTCACGTTTTACTCTATTCCTTATAAACGTCTCCTTTTAGATTGTATAGAATAGGTCTACTGAACCTTTTG  720

          M  H  Y  L  K  A  V  I  K  E  T  L  R  L  H  P  P  L  P  L  L  V  P  R  T  A  R
721       ATGCACTACTTGAAAGCAGTGATTAAAGAGACACTACGGTTACATCCACCACTTCCGTTATTAGTCCCACGAACTGCAAG  800
721       TACGTGATGAACTTTCGTCACTAATTTCTCTGTGATGCCAATGTAGGTGGTGAAGGCAATAATCAGGGTGCTTGACGTTC  800

            Q  D  V  K  V  M  G  Y  D  I  A  A  G  T  M  V  I  I  N  A  W  G  I  G  R  D
801       ACAGGATGTTAAAGTTATGGGCTACGATATTGCAGCTGGAACTATGGTCATCATCAACGCTTGGGGAATCGGAAGGGACC  880
801       TGTCCTACAATTTCAATACCCGATGCTATAACGTCGACCTTGATACCAGTAGTAGTTGCGAACCCCTTAGCCTTCCCTGG  880

          P  S  S  W  D  E  P  D  E  F  R  P  E  R  F  L  N  S  T  I  D  Y  R  G  H  D  F
881       CTAGTTCATGGGATGAACCGGATGAGTTTAGGCCAGAAAGGTTCTTGAATTCGACAATAGATTACAGAGGTCATGACTTC  960
881       GATCAAGTACCCTACTTGGCCTACTCAAATCCGGTCTTTCCAAGAACTTAAGCTGTTATCTAATGTCTCCAGTACTGAAG  960

          Q  L  I  P  F  G  A  G  R  R  G  C  P  G  I  S  F  A  M  S  N  Y  E  Y  V  L  A
961       CAGTTGATCCCATTTGGTGCTGGGAGGAGGGGTTGTCCCGGAATAAGTTTTGCCATGTCAAATTACGAGTATGTGTTGGC 1040
961       GTCAACTAGGGTAAACCACGACCCTCCTCCCCAACAGGGCCTTATTCAAAACGGTACAGTTTAATGCTCATACACAACCG 1040

            N  L  L  H  K  F  D  W  K  L  P  N  G  M  K  S  D  E  L  N  M  D  E  G  T  G
1041      AAATCTTCTGCATAAGTTTGATTGGAAACTTCCTAATGGAATGAAGAGTGATGAGCTTAATATGGATGAGGGCACTGGTC 1120
1041      TTTAGAAGACGTATTCAAACTAACCTTTGAAGGATTACCTTACTTCTCACTACTCGAATTATACCTACTCCCGTGACCAG 1120

          L  T  C  H  R  K  V  P  L  L  V  V  A  T  P  I  S  C  *
1121      TTACTTGTCATAGAAAAGTCCCTCTACTTGTAGTTGCAACTCCAATTTCTTGTTGA  1176
1121      AATGAACAGTATCTTTTCAGGGAGATGAACATCAACGTTGAGGTTAAAGAACAACT  1176
```

Annexe 17 – B : Séquences nucléotidiques et peptidiques obtenues par des réactions de TAIL-PCR grâce à l'amorce dégénérée AD3 sur le contig P_sat_rep_c5642. En jaune : intron, en vert : séquence correspondant au contig P_sat_rep_c5642.

```
          I  E  F  G  V  G  T  F  T  P  W  L  A  W  I  N  T  V  S  G  F  N  A  K  V  K  N
1         ATCGAGTTTGGTGTTGGAACCTTTACTCCATGGCTTGCATGGATCAATACTGTTTCAGGTTTCAATGCCAAAGTAAAAAA   80
1         TAGCTCAAACCACAACCTTGGAAATGAGGTACCGAACGTACCTAGTTATGACAAAGTCCAAAGTTACGGTTTCATTTTTT   80

          I  A  K  E  N  D  E  F  F  E  R  I  V  E  E  H  I  N  A  S  T  R  K  K  K  G
81        CATTGCAAAAGAGAACGATGAATTTTTCGAACGGATTGTGGAAGAGCACATAAATGCTTCAACAAGAAAGAAGAAAGGGG   160
81        GTAACGTTTTCTCTTGCTACTTAAAAAGCTTGCCTAACACCTTCTCGTGTATTTACGAAGTTGTTCTTTCTTCTTTCCCC   160

          G  E  K  I  K  E  T  E  V  F  L  D  I  L  L  G  A  Q  E  D  A  I  G  G  F  H  I
161       GTGAGAAAATTAAAGAAACAGAGGTTTTTCTCGATATCTTGCTTGGAGCACAAGAGGATGCCATCGGTGGCTTTCATATT   240
161       CACTCTTTTAATTTCTTTGTCTCCAAAAAGAGCTATAGAACGAACCTCGTGTTCTCCTACGGTAGCCACCGAAAGTATAA   240

          D  R  D  G  I  K  A  L  I  L  V  N  S  K  T  L  H  F  K  *  H  D  L  K  F  R  F
241       GATAGAGATGGCATCAAAGCTCTGATCCTGGTGAATTCTAAAACACTTCATTTCAAATGACATGATTTGAAATTCAGGTT   320
241       CTATCTCTACCGTAGTTTCGAGACTAGGACCACTTAAGATTTTGTGAAGTAAAGTTTACTGTACTAAACTTTAAGTCCAA   320

            P  T  D  H  N  I  F  L  S  V  K  Y  I  H  I  Q  N  C  *  Y  F  M  Q  L  G  V
321       CCCAACGGACCATAACATCTTTCTGTCTGTCAAATATATTCATATTCAGAATTGTTGATACTTCATGCAACTCGGAGTTA   400
321       GGGTTGCCTGGTATTGTAGAAAGACAGACAGTTTATATAAGTATAAGTCTTAACAACTATGAAGTACGTTGAGCCTCAAT   400

          N  L  *  T  N  N  L  H  S  S  S  A  T  N  F  F  K  N  K  V  H  H  C  N  L  S  L
401       ATTTATAGACCAATAACTTGCATTCTTCTTCAGCCACAAATTTCTTTAAGAATAAGGTGCATCATTGCAATCTTTCACTC   480
401       TAAATATCTGGTTATTGAACGTAAGAAGAAGTCGGTGTTTAAAGAAATTCTTATTCCACGTAGTAACGTTAGAAAGTGAG   480

            *  G  I  D  M  H  L  F  K  K  I  S  T  H  R  P  N  F  *  C  I  T  I  I  V  T  Q
481       TGAGGAATTGACATGCATCTTTTCAAAAAAATTAGTACCCACAGACCTAACTTTTGATGTATCACCATCATTGTAACACA   560
481       ACTCCTTAACTGTACGTAGAAAAGTTTTTTTAATCATGGGTGTCTGGATTGAAAACTACATAGTGGTAGTAACATTGTGT   560

            D  M  I  A  G  G  T  D  T  T  Y  T  A  L  E  W  T  M  T  E  L  L  R  R  P  K
561       GGATATGATTGCTGGTGGAACTGACACTACCTACACAGCTCTAGAATGGACAATGACGGAACTTTTAAGGCGCCCGAAAG   640
561       CCTATACTAACGACCACCTTGACTGTGATGGATGTGTCGAGATCTTACCTGTTACTGCCTTGAAAATTCCGCGGGCTTTC   640

          A  M  K  E  V  Q  N  E  I  R  N  I  C  R  G  K  S  N  I  S  Y  P  D  D  L  E  N
641       CCATGAAGGAAGTGCAAAATGAGATAAGGAATATTTGCAGAGGAAAATCTAACATATCCTATCCAGATGACTTGGAAAAC   720
641       GGTACTTCCTTCACGTTTTACTCTATTCCTTATAAACGTCTCCTTTTAGATTGTATAGGATAGGTCTACTGAACCTTTTG   720

            M  H  Y  L  K  A  V  I  K  E  T  L  R  L  H  P  P  L  P  L  L  V  P  R  T  A  R
721       ATGCACTACTTGAAAGCAGTGATTAAAGAGACACTACGGTTACATCCACCACTTCCGTTATTAGTCCCACGAACTGCAAG   800
721       TACGTGATGAACTTTCGTCACTAATTTCTCTGTGATGCCAATGTAGGTGGTGAAGGCAATAATCAGGGTGCTTGACGTTC   800

            Q  D  V  K  V  M  G  Y  D  I  A  A  G  T  M  V  I  I  N  A  W  G  I  G  R  D
801       ACAGGATGTTAAAGTTATGGGCTACGATATTGCAGCTGGAACTATGGTCATCATCAACGCTTGGGGAATCGGAAGGGACC   880
801       TGTCCTACAATTTCAATACCCGATGCTATAACGTCGACCTTGATACCAGTAGTAGTTGCGAACCCCTTAGCCTTCCCTGG   880

          P  S  S  W  D  E  P  D  E  F  R  P  E  R  F  L  N  S  T  I  D  Y  R  G  H  D  F
881       CTAGTTCATGGGATGAACCGGATGAGTTTAGGCCAGAAAGGTTCTTGAATTCGACAATAGATTACAGAGGTCATGACTTC   960
881       GATCAAGTACCCTACTTGGCCTACTCAAATCCGGTCTTTCCAAGAACTTAAGCTGTTATCTAATGTCTCCAGTACTGAAG   960

            Q  L  I  P  F  G  A  G  R  R  G  C  P  G  I  S  F  A  M  S  N  Y  E  Y  V  L  A
961       CAGTTGATCCCATTTGGTGCTGGGAGGAGGGGTTGTCCCGGAATAAGTTTTGCCATGTCAAATTACGAGTATGTGTTGGC   1040
961       GTCAACTAGGGTAAACCACGACCCTCCTCCCCAACAGGGCCTTATTCAAAACGGTACAGTTTAATGCTCATACACAACCG   1040

            N  L  L  H  K  F  D  W  K  L  P  N  G  M  K  S  D  E  L  N  M  D  E  G  T  G
1041      AAATCTTCTGCATAAGTTTGATTGGAAACTTCCTAATGGAATGAAGAGTGATGAGCTTAATATGGATGAGGGCACTGGTC   1120
1041      TTTAGAAGACGTATTCAAACTAACCTTTGAAGGATTACCTTACTTCTCACTACTCGAATTATACCTACTCCCGTGACCAG   1120

          L  T  C  H  R  K  V  P  L  L  V  V  A  T  P  I  S  C  *
1121      TTACTTGTCATAGAAAAGTCCCTCTACTTGTAGTTGCAACTCCAATTTCTTGTTGA   1176
1121      AATGAACAGTATCTTTTCAGGGAGATGAACATCAACGTTGAGGTTAAAGAACAACT   1176
```

Annexe 17 - C : Séquences nucléotidiques et peptidiques obtenues par des réactions de TAIL-PCR grâce à l'amorce dégénérée AD3 sur le contig P_sat_rep_c5642. En jaune : intron, en vert : séquence correspondant au contig P_sat_rep_c5642.

```
          V  E  F  G  V  V  Q  L  L  N  N  K  G  V  E  D  M  G  T  I  R  E  E  E  T  A  F
1         GTCGAGTTTGGTGTTGTCCAGCTTTTGAATAACAAAGGGGTTGAAGATATGGGAACGATAAGAGAGGAAGAAACTGCTTT   80
1         CAGCTCAAACCACAACAGGTCGAAAACTTATTGTTTCCCCAACTTCTATACCCTTGCTATTCTCTCCTTCTTTGACGAAA   80

            L  L  K  N  I  L  D  S  N  S  S  V  V  N  L  S  E  L  L  I  T  L  T  F  N  L
81        TCTACTGAAAAATATTCTAGACTCTAATTCTTCGGTTGTGAACTTGAGTGAACTGTTAATCACACTCACATTCAATCTTG  160
81        AGATGACTTTTTTATAAGATCTGAGATTAAGAAGCCAACACTTGAACTCACTTGACAATTAGTGTGAGTGTAAGTTAGAAC  160

          V  S  T  A  A  F  G  K  K  V  D  E  W  G  K  G  R  K  L  W  M  L  I  K  E  T  E
161       TTTCTACAGCAGCTTTTGGGAAAAAAGTTGATGAATGGGGGAAAGGCAGAAAGTTATGGATGCTAATAAAGGAGACAGAG  240
161       AAAGATGTCGTCGAAAACCCTTTTTTCAACTACTTACCCCCTTTCCGTCTTTCAATACCTACGATTATTTCCTCTGTCTC  240

            E  L  L  G  R  F  D  V  G  T  F  T  P  W  L  A  W  I  N  T  V  S  G  F  N  A  K
241       GAATTACTTGGCCGATTTGATGTTGGAACCTTTACTCCATGGCTTGCATGGATCAATACTGTTTCAGGTTTCAATGCCAA  320
241       CTTAATGAACCGGCTAAACTACAACCTTGGAAATGAGGTACCGAACGTACCTAGTTATGACAAAGTCCAAAGTTACGGTT  320

            V  K  N  I  A  K  E  N  D  E  F  F  E  R  I  V  E  E  H  I  N  A  S  T  R  K
321       AGTAAAAAACATTGCAAAAGAGAACGATGAATTTTTCGAACGGATTGTGGAAGAGCACATAAATGCTTCAACAAGAAAGA  400
321       TCATTTTTTGTAACGTTTTCTCTTGCTACTTAAAAAGCTTGCCTAACACCTTCTCGTGTATTTACGAAGTTGTTCTTTCT  400

          K  K  G  G  E  K  I  K  E  T  E  V  F  L  D  I  L  L  G  A  Q  E  D  A  I  G  G
401       AGAAAGGCGGTGAGAAAATTAAAGAAACAGAGGTTTTTCTCGATATCTTGCTTGGAGCACAAGAGGATGCCATCGGTGGC  480
401       TCTTTCCGCCACTCTTTTAATTTCTTTGTCTCCAAAAAGAGCTATAGAACGAACCTCGTGTTCTCCTACGGTAGCCACCG  480

            F  H  I  D  R  D  G  I  K  A  L  I  L  V  N  S  I  T  L  H  F  K  *  H  D  L  K
481       TTTCATATTGATAGAGATGGCATCAAAGCTCTGATCCTGGTAAATTCTATAACACTTCATTTCAAATGACATGATTTGAA  560
481       AAAGTATAACTATCTCTACCGTAGTTTCGAGACTAGGACCATTTAAGATATTGTGAAGTAAAGTTTACTGTACTAAACTT  560

              F  R  F  P  T  D  H  N  I  F  L  S  V  K  Y  I  H  I  Q  N  C  *  Y  F  M  K
561       ATTCAGGTTCCCAACGGACCATAACATCTTTCTGTCTGTCAAATATATTCATATTCAGAATTGTTGATACTTCATGAAAC  640
561       TAAGTCCAAGGGTTGCCTGGTATTGTAGAAAGACAGACAGTTTATATAAGTATAAGTCTTAACAACTATGAAGTACTTTG  640

          L  G  V  N  L  *  T  N  N  L  H  S  S  S  A  T  N  F  F  K  N  K  V  H  H  C  N
641       TCGGAGTTAATTTATAGACCAATAACTTGCATTCTTCTTCTGCCACAAATTTCTTTAAGAATAAGGTGCATCATTGCAAT  720
641       AGCCTCAATTAAATATCTGGTTATTGAACGTAAGAAGAAGACGGTGTTTAAAGAAATTCTTATTCCACGTAGTAACGTTA  720

            L  S  L  *  G  I  D  M  H  L  F  K  K  I  S  T  H  K  P  N  F  *  C  I  T  I  I
721       CTTTCACTCTGAGGAATTGACATGCATCTTTTCAAAAAAATTAGTACCCACAAACCTAACTTTTGATGTATCACCATCAT  800
721       GAAAGTGAGACTCCTTAACTGTACGTAGAAAAGTTTTTTTAATCATGGGTGTTTGGATTGAAAACTACATAGTGGTAGTA  800

              V  T  Q  D  M  I  A  G  G  T  D  T  T  Y  T  A  L  E  W  T  M  T  E  L  L  R
801       TGTAACACAGGATATGATTGCTGGTGGAACTGACACTACCTACACAGCTCTAGAATGGACAATGACCGAACTTTTAAGGC  880
801       ACATTGTGTCCTATACTAACGACCACCTTGACTGTGATGGATGTGTCGAGATCTTACCTGTTACTGGCTTGAAAATTCCG  880

          H  P  K  A  M  K  E  V  Q  I  E  I  R  N  I  C  R  G  K  S  N  I  S  Y  P  D  D
881       ACCCGAAAGCCATGAAGGAAGTGCAAATTGAGATAAGGAATATTTGCAGAGGAAAATCTAACATATCTTATCCAGATGAC  960
881       TGGGCTTTCGGTACTTCCTTCACGTTTAACTCTATTCCTTATAAACGTCTCCTTTTAGATTGTATAGAATAGGTCTACTG  960

            L  E  N  M  H  Y  L  K  A  V  I  K  E  A  L  R  L  H  P  P  L  P  L  L  V  P  R
961       TTGGAAAACATGCACTACTTGAAAGCAGTGATTAAAGAGGCACTACGGTTACATCCACCACTTCCGTTATTAGTCCCACG 1040
961       AACCTTTTGTACGTGATGAACTTTCGTCACTAATTTCTCCGTGATGCCAATGTAGGTGGTGAAGGCAATAATCAGGGTGC 1040

              T  A  R  Q  D  V  K  V  M  G  Y  D  I  A  A  G  T  M  V  I  I  N  A  W  G  I
1041      AACTGCAAGACAGGATGTTAAAGTTATGGGCTACGATATTGCAGCTGGAACTATGGTCATCATCAACGCTTGGGGAATCG 1120
1041      TTGACGTTCTGTCCTACAATTTCAATACCCGATGCTATAACGTCGACCTTGATACCAGTAGTAGTTGCGAACCCCTTAGC 1120

          G  R  D  P  S  S  W  D  E  P  D  E  F  R  P  E  R  F  L  N  S  T  I  D  Y  R  G
1121      GAAGGGACCCTAGTTCATGGGATGAACCGGATGAGTTTAGGCCAGAAAGGTTCTTGAATTCGACAATAGATTACAGAGGT 1200
1121      CTTCCCTGGGATCAAGTACCCTACTTGGCCTACTCAAATCCGGTCTTTCCAAGAACTTAAGCTGTTATCTAATGTCTCCA 1200

            H  D  F  Q  L  I  P  F  G  A  G  R  R  G  C  P  G  I  S  F  A  M  S  N  Y  E  Y
1201      CATGACTTCCAGTTGATCCCATTTGGTGCTGGGAGGAGGGGTTGTCCCGGAATAAGTTTTGCCATGTCAAATTACGAGTA 1280
1201      GTACTGAAGGTCAACTAGGGTAAACCACGACCCTCCTCCCCAACAGGGCCTTATTCAAAACGGTACAGTTTAATGCTCAT 1280

              V  L  A  N  L  L  H  K  F  D  W  K  L  P  N  G  M  K  S  D  E  L  N  M  D  E
1281      TGTGTTGGCAAATCTTCTGCATAAGTTTGATTGGAAACTTCCTAATGGAATGAAGAGTGATGAGCTTAATATGGATGAGG 1360
1281      ACACAACCGTTTAGAAGACGTATTCAAACTAACCTTTGAAGGATTACCTTACTTCTCACTACTCGAATTATACCTACTCC 1360

          G  T  G  L  T  C  H  R  K  V  P  L  L  V  V  A  T  P  I  S  C  *
1361      GCACTGGTCTTACTTGTCATAGAAAAGTCCCTCTACTTGTAGTTGCAACTCCAATTTCTTGTTGA  1425
1361      CGTGACCAGAATGAACAGTATCTTTTCAGGGAGATGAACATCAACGTTGAGGTTAAAGAACAACT  1425
```

Liste de publications

Articles dans des revues internationales à comité de lecture :

Bushakra, J. M., Krieger, C., Deng, D., Stephens, M. J., Allan, A. C., Storey, R., Symonds, V. V., Stevenson, D., McGuie, T. K., Buck, E. J., and Gardiner, S. E. (2013). *QTL involved in the modification of cyanidin compounds in black and red raspberry fruit*. Theoretical and Applied Genetics, 126(3), 847-865.

Chagné, D., Krieger, C., Rassam, M., Sullivan, M., Fraser, J., André, C., Pindo, M., Troggio, M., Gardiner, S. E., Henry, R. A., Allan, A. C., McGuie, T. K., and Laing, W. A. (2012). *QTL and candidate gene mapping for polyphenolic composition in apple fruit*. BMC plant biology, 12(1), 12.

Vialart G., Hehn A., Olry A., Ito K., Krieger C., Larbat R., Paris C., Shimizu B., Sugimoto Y., Mizutani M and Bourgaud F. (2012). *A 2-oxoglutarate-dependent dioxygenase from Ruta graveolens L. exhibits p-coumaroyle CoA 2'-hydroxylase activity (C2'H): a missing step in the synthesis of umbelliferone in plants*. The plant Journal, 70(3), 460,470.

Communications nationales :

Présentation orale :

Krieger C., Hehn A., Olry A., Dueholm B., Weitzel C., Simonsen H. T., Kellner S., Matern U., Bourgaud F. *Identification de gènes codants pour des cytochromes P450 à partir de trois banques d'ADNc de plantes produisant des furanocoumarines et caractérisation fonctionnelle de la sous-famille CYP71AZ chez Pastinaca sativa.* Séminaire de recherche inter-unités labellisé par le labex Arbre de l'Université de Lorraine, le 18 octobre 2013.

Par voie d'affiche :

Krieger C., Hehn A., Olry A., Roselli S., Weitzel C., Simonsen H. T., Kellner S., Matern U., Bourgaud F. *Les cytochromes P450s de la famille CYP71AZ impliqués dans la synthèse de furanocoumarines et de coumarines chez les Apiacées*. Journée interpôle Scientifique CPM-BMS-A2F de l'Université de Lorraine, le 08 novembre 2013.

Krieger C., Hehn A., Olry A., Menin B., Simonsen H. T., Bourgaud F. *Identification des chaînons moléculaires manquants dans la voie de biosynthèse des furocoumarines chez le panais (Pastinaca sativa)*. Ecole doctorale RP2E, 19 janvier 2012, Nancy.

Identification moléculaire et caractérisation fonctionnelle d'une nouvelle sous-famille de cytochromes P450, CYP71AZ, impliquée dans la synthèse de furanocoumarines et de coumarines chez *Pastinaca sativa*.

Les furanocoumarines (FCs) sont des métabolites secondaires principalement synthétisés chez quatre familles botaniques et dérivent de la voie de biosynthèse des phénylpropanoïdes. Ces phytoalexines interviennent dans les processus de défense de la plante et présentent un fort potentiel thérapeutique. Des travaux réalisés dans les années 1960 sur des cultures cellulaires en parallèle de l'utilisation de précurseurs radiomarqués ont permis de démontrer que de nombreuses enzymes impliquées dans cette voie appartenaient à la famille des cytochromes P450 (P450s). Seules deux d'entre elles avaient pu être identifiées d'un point de vue moléculaire au début de ce travail de thèse.
Afin de générer des informations concernant le génome de plantes productrices de FCs, nous avons fait séquencer les ARNm extraits de feuilles de *Pastinaca sativa*, de *Ruta graveolens* et de *Cullen cinereum*. L'analyse *in silico* de ces trois banques de données a permis d'identifier près de 800 fragments d'ADNc codants pour des P450s. Des travaux antérieurs réalisés au laboratoire et l'analyse comparative des transcriptomes de ces 3 plantes nous ont amenés à nous focaliser sur la sous-famille CYP71AZ au travers d'une étude fine de CYP71AZ3 et CYP71AZ4.
La caractérisation fonctionnelle de ces enzymes a été réalisée dans un système d'expression hétérologue eucaryote : *Saccharomyces cerevisiae*. Les résultats obtenus ont permis de montrer que CYP71AZ4 avait une spécificité de substrat assez large puisqu'elle pouvait métaboliser au moins une FC et 4 coumarines. L'analyse et la comparaison des constantes cinétiques pour chacun de ces substrats indiquent néanmoins que le psoralène est le substrat préférentiel. La caractérisation fonctionnelle de CYP71AZ3 a mis en évidence que cette enzyme pouvait hydroxyler l'esculétine, une coumarine, mais ne jouait aucun rôle dans la synthèse de FCs.
Ces travaux mettent en évidence la diversité fonctionnelle au sein d'une même sous-famille enzymatique et permettent d'émettre des hypothèses nouvelles quant à l'apparition de cette voie de biosynthèse chez les Apiacées d'une part, et chez les autres familles botaniques d'autre part.

Mots clefs : cytochrome P450, psoralène 8-monooxygénase, métabolite secondaire, *Pastinaca sativa*, *Ruta graveolens*, *Cullen cinereum*, banque d'ADNc, furanocoumarines, psoralène, xanthotoxol, esculétine.

Molecular isolation and functional characterization of a novel cytochrome P450 subfamily, CYP71AZ, involved in the biosynthesis of furanocoumarins and coumarins in *Pastinaca sativa*.

Furanocoumarins (FCs) are secondary metabolites mainly synthetized in four botanical families deriving from the phenylpropanoid biosynthetic pathway. These phytoalexins are involved in plant defense mechanisms and present strong therapeutic potential. Early studies in the 1960s based on cell cultures and the use of radiolabeled precursors have shown that many enzymes involved in this pathway belong to the cytochrome P450 family (P450s). Only two of them had been identified from a molecular point of view at the beginning of this thesis.
In order to generate information regarding the genome of plants producing FCs, we sequenced the mRNA extracted from leaves of *Pastinaca sativa*, *Ruta graveolens*, and *Cullen cinereum*. *In silico* analysis of these three libraries identified nearly 800 cDNA fragments encoding for P450s. Previous studies in the laboratory and comparative transcriptome analysis of these three plants have led us to focus on the subfamily CYP71AZ through a detailed study of CYP71AZ3 and CYP71AZ4.
Functional characterization of these enzymes was performed in an eukaryote heterologous expression system: *Saccharomyces cerevisiae*. The results showed that CYP71AZ4 had a broad substrate specificity enough as it could metabolize one FC and 4 coumarins. The analysis and comparison of the kinetic constants for each of these substrates indicate, however, that the preferred substrate is psoralen. The functional characterization of CYP71AZ3 showed that this enzyme could hydroxylate esculetin, a coumarin, but played no role in the synthesis of FCs.
This study highlights the functional diversity within a single enzyme subfamily and allows to issue new hypotheses about the emergence of this biosynthetic pathway in Apiaceae on one hand, and among other botanical families on the other hand.

Key words : cytochrome P450, psoralen 8-monooxygenase, secondary metabolite, *Pastinaca sativa*, *Ruta graveolens*, *Cullen cinereum*, cDNA databank, furanocoumarins, psoralen, xanthotoxol, esculetin.