

UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

IDENTIFICATION D'ENZYMES IMPLIQUÉES DANS LA BIOSYNTHÈSE
DE MOLÉCULES ANALOGUES AUX CANNABINOÏDES ET
LEUR CARACTÉRISATION DANS LES MICROORGANISMES

MÉMOIRE PRÉSENTÉ
COMME EXIGENCE PARTIELLE DE LA
MAÎTRISE EN BIOLOGIE CELLULAIRE ET MOLÉCULAIRE

PAR
ARIANE GARAND

AOÛT 2022

Université du Québec à Trois-Rivières

Service de la bibliothèque

Avertissement

L'auteur de ce mémoire, de cette thèse ou de cet essai a autorisé l'Université du Québec à Trois-Rivières à diffuser, à des fins non lucratives, une copie de son mémoire, de sa thèse ou de son essai.

Cette diffusion n'entraîne pas une renonciation de la part de l'auteur à ses droits de propriété intellectuelle, incluant le droit d'auteur, sur ce mémoire, cette thèse ou cet essai. Notamment, la reproduction ou la publication de la totalité ou d'une partie importante de ce mémoire, de cette thèse et de son essai requiert son autorisation.

It's not going to be easy, it's going to be worth it

Zig Ziglar

REMERCIEMENTS

Je souhaite remercier toutes les personnes qui ont contribué de près ou de loin à la concrétisation de ma maîtrise en recherche et de ce mémoire. En premier lieu, je tiens à remercier la professeure Isabel Desgagné-Penix, ma directrice de recherche. J'ai pour Isabel un profond respect envers son parcours et ses accomplissements ainsi qu'une grande admiration pour les valeurs qu'elle s'efforce de transmettre chaque jour en tant que femme chercheuse issue d'une minorité. Un grand merci pour son soutien, ses conseils, ainsi que pour les beaux moments que nous avons partagés tout au long de mon parcours universitaire.

Je tiens spécialement à remercier Dre Fatma Meddeb, professionnelle de recherche, qui m'a accompagnée tout au long de ce projet. Je la remercie pour sa gentillesse, sa disponibilité, son aide et le dévouement qu'elle porte envers les étudiants du laboratoire.

Je remercie mes nombreux collègues et membres de l'équipe, actuels ou passés, pour leur aide et leurs conseils. Je tiens à remercier plus particulièrement mon stagiaire Mikaël Desrosiers qui m'a énormément aidé lors de mon dernier été au laboratoire et dans la finalisation de ce projet.

Je remercie les professeurs Hugo Germain, Lionel Berthoux, Tagnon Missihoun et Marc Beauregard, de nous avoir permis d'utiliser leur matériel et leurs appareils.

Je tiens à remercier les organismes subventionnaires ainsi que les entreprises et groupes de recherche qui ont collaboré à ce projet : Algae-C, le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG), le Fonds de recherche du Québec Nature et technologies (FRQnet), Mitacs, le groupe de recherche en biologie végétale (GRBV) ainsi que la Société québécoise de bryologie (SQB). Un merci spécial à

Marc Favreau, président de la Société québécoise de bryologie, de m'avoir partagé ses connaissances et son intérêt pour les bryophytes du Québec.

Enfin, je remercie mes parents, ma famille et mes amis pour leurs encouragements et leur support au cours de mon cheminement académique. Je suis extrêmement choyée de vous avoir à mes côtés.

AVANT-PROPOS

Ce mémoire fait l'objet du travail réalisé dans le cadre de la maîtrise en biologie cellulaire et moléculaire, de septembre 2019 à décembre 2021. La recherche a été réalisée sous la direction de la professeure Isabel Desgagné-Penix.

Les travaux présentés dans ce mémoire correspondent à une petite fraction d'un bien plus grand projet de collaboration entre l'UQTR et l'entreprise Algae-C. L'objectif primaire de cette collaboration visait la production de métabolites spécialisés, et plus spécialement de cannabinoïdes, grâce à la création d'une plateforme utilisant des microalgues modifiées par ingénierie métabolique. Le projet a été mis sur pied en raison d'un besoin émergeant en molécules actives d'origine végétale pour les secteurs médical, pharmaceutique et récréatif. La mission première de cette entreprise en démarrage est de produire ces différents principes actifs de manière plus verte, durable et à faible coût. Puisque le développement d'une telle plateforme d'ingénierie métabolique nécessite une certaine expertise, la professeure-chercheuse de l'UQTR, Isabel Desgagné-Penix, a accepté de mettre à profit les diverses connaissances de sa nombreuse équipe pour résoudre cette problématique. Une fois que la base de cette plateforme a été mise en place, il a été possible d'étudier d'autres avenues. C'est ainsi que mon projet a été mis en place visant l'étude d'enzymes analogues à celles du cannabis impliquées dans la biosynthèse de cannabinoïdes.

RÉSUMÉ

Depuis octobre 2018, la légalisation du cannabis au Canada a amené plusieurs entreprises à s'intéresser aux différentes méthodes possibles pour la production de cannabinoïdes. Les cannabinoïdes étant des molécules actives capables de se lier aux récepteurs cannabinoïdes humains, ils sont produits de façon sélective *in planta*. En effet, comme les molécules actives se retrouvent en quantité variable dans les plantes, les industries se sont tournées vers la production synthétique, par chimie organique, de ces molécules. Ce projet de recherche propose de développer une plateforme utilisant des microorganismes modifiés par ingénierie métabolique pour produire des cannabinoïdes. Toutefois, un brevet accordé à Prof. Jonathan E. Page sur la séquence de nucléotides de la prényltransférase aromatique 1 de *Cannabis sativa* (*CsPT1*), une enzyme impliquée de la voie de biosynthèse des cannabinoïdes, empêche pour l'instant leur production par biologie synthétique.

Le projet proposé a donc pour but de trouver, dans une autre espèce végétale, une enzyme pouvant imiter la fonction de la *CsPT1* et de l'insérer dans un microorganisme afin de la caractériser. Des molécules analogues aux cannabinoïdes identifiées récemment dans l'espèce *Radula marginata*, une plante hépatique indigène de Nouvelle-Zélande et de Tasmanie, ont concentré nos recherches d'enzymes sur cette espèce.

Ainsi, dans cette étude, différents candidats enzymatiques ont été identifiés dans les données transcriptomiques de *Radula marginata*. Ces derniers ont été soumis à différentes analyses *in silico* afin d'obtenir des prédictions de structure, de fonction, de localisation et de phylogénie. Notre approche a révélé deux protéines particulièrement intéressantes, soit RmPT280 et RmPT130. Les séquences protéiques de ces candidats ont été synthétisées et assemblées dans des vecteurs d'expression. Les vecteurs assemblés ont ensuite été transformés dans la levure *Saccharomyces cerevisiae*. Les protéines ont ensuite été induites et détectées par immunobuvardage de type Western blot. Cependant, l'absence d'une activité enzymatique reproductible n'a pas permis de véritablement caractériser les enzymes et ainsi infirmer ou confirmer la possibilité qu'elles puissent catalyser la même réaction que la *CsPT1*.

Mots-clés : Cannabinoïdes, expression hétérologue, *Radula marginata*, biologie synthétique, *Saccharomyces cerevisiae*, transcriptomique.

TABLE DES MATIÈRES

REMERCIEMENTS	iii
AVANT-PROPOS	v
RÉSUMÉ	vi
LISTE DES TABLEAUX	x
LISTE DES FIGURES	xi
LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES	xiii
LISTE DES SYMBOLES	xvi
CHAPITRE I	
INTRODUCTION	1
1.1 Le cannabis (<i>Cannabis sativa</i>)	1
1.1.1 Généralités	1
1.1.2 Voie de biosynthèse de cannabinoïdes	3
1.1.3 Autres prényltransférases de <i>C. sativa</i>	5
1.1.4 Les trichomes	6
1.2 Les métabolites spécialisés	8
1.2.1 Généralités	8
1.2.2 Les métabolites spécialisés produits par <i>C. sativa</i>	10
1.3 L'industrie du cannabis au Québec	11
1.3.1 Contexte actuel	11
1.3.2 Revenus générés au Québec en 2021	12
1.3.3 Enjeux liés à la production de cannabis	13
1.4 <i>Radula</i> sp.	15
1.4.1 Généralités	15
1.4.2 <i>Radula perrottetii</i>	18
1.4.3 <i>Radula marginata</i>	19
1.5 Hypothèse et objectifs du projet de recherche.....	21
CHAPITRE II	
MÉTHODES EXPÉRIMENTALES	23
2.1 Démarche expérimentale	23

2.2	Obtention et assemblage des données de séquençage Illumina de <i>R. marginata</i>	24
2.3	Analyses <i>in silico</i>	25
2.3.1	Détermination des domaines conservés chez les prényltransférases	25
2.3.2	Identification des séquences pertinentes chez <i>R. marginata</i>	26
2.3.3	Prédictions des structures des protéines d'intérêts	26
2.4	Conception des cassettes d'expression	27
2.5	Sélection du vecteur d'expression	28
2.6	Assemblage du vecteur d'expression par Gibson Assembly	32
2.7	Transformation dans la bactérie <i>E. coli</i>	34
2.8	Choix de la levure <i>S. cerevisiae</i> comme système d'expression hétérologue	35
2.9	Transformation dans la levure	37
2.10	Expression des protéines recombinantes	37
2.11	Extraction et enrichissement des protéines recombinantes	38
2.12	Détection par immunobuvardage de type Western blot	39
2.13	Essais enzymatiques	41
CHAPITRE III		
RÉSULTATS		43
3.1	Obtention et assemblage des données de séquençage Illumina de <i>R. marginata</i>	43
3.2	Analyses <i>in silico</i>	44
3.1.1	Détermination des domaines conservés chez les prényltransférases	44
3.2.1	Identification des séquences pertinentes chez <i>R. marginata</i>	46
3.2.2	Prédictions des structures des protéines d'intérêts	48
3.3	Assemblage du vecteur d'expression par Gibson Assembly	50
3.4	Transformation dans la bactérie <i>E. coli</i>	53
3.5	Transformation dans la levure	53
3.6	Expression des protéines recombinantes	55
3.7	Détection par immunobuvardage de type Western blot	57
3.8	Essais enzymatiques	60

CHAPITRE IV	
DISCUSSION	63
4.1 Obtention et assemblage des données de séquençage Illumina de <i>R. marginata</i>	63
4.2 Analyses <i>in silico</i>	64
4.2.1 Détermination des domaines conservés chez les prényltransférases	65
4.2.2 Identification des séquences pertinentes chez <i>R. marginata</i>	65
4.2.3 Prédiction des structures des protéines d'intérêts	65
4.3 Transformation dans la levure	67
4.4 Expression et extraction des protéines recombinantes	67
4.5 Immunobuvardage de type Western blot.....	70
4.6 Essais enzymatiques	72
4.7 Hypothèses.....	75
4.8 Améliorations	76
CHAPITRE V	
CONCLUSION.....	78
ANNEXE A	
SÉQUENCES DES CASSETTES D'EXPRESSION UTILISÉES DANS LE CADRE DE CE PROJET.....	81
ANNEXE B	
LISTE DES ABRÉVIATIONS ET DES NUMÉROS D'ACCESSION GENBANK POUR LES 11 PRÉNYLTRANSFÉRASES UTILISÉES LORS DE L'ALIGNEMENT MULTIPLE	85
ANNEXE C	
LISTE DES ABRÉVIATIONS ET DES NUMÉROS D'ACCESSION GENBANK DES PROTÉINES UTILISÉES POUR L'ANALYSE PHYLOGÉNÉTIQUE	86
RÉFÉRENCES BIBLIOGRAPHIQUES.....	88

LISTE DES TABLEAUX

Tableau		Page
1.1	Résultats de l'exercice financier de la SQDC pour l'année financière 2020-2021 en milliers de dollars canadiens	12
1.2	Empreinte carbone de la production intérieure de cannabis	14
2.1	Éléments du vecteur commercial pYES2/CT pour expression dans la levure	29
2.2	Résumé des différentes conditions testées pour les essais enzymatiques	42
3.1	Résumé de l'assemblage des données brutes du séquençage de <i>R. marginata</i> par les logiciels Trimmomatic et Trinity	43
4.1	Résumé des différentes conditions utilisées dans la littérature pour des essais enzymatiques sur des prényltransférases.....	73
4.2	Liste des substrats intéressants à tester lors des essais enzymatiques.....	75

LISTE DES FIGURES

Figure		Page
1.1	Voie de biosynthèse menant à la formation des deux principaux cannabinoïdes	4
1.2	Trois types de trichomes du cannabis représentés par microscopie électronique	7
1.3	Informations génomiques disponibles sur les Embryophytes depuis les années 2000	17
1.4	Effets induits <i>in vivo</i> par le PET, identifié chez <i>R. perrottetii</i> , sur le récepteur de cannabinoïdes CB1R de la souris	18
1.5	Proposition d'une voie menant à la biosynthèse du PET chez <i>R. marginata</i> selon les métabolites identifiés.....	20
2.1	Représentation visuelle du plan expérimental de la programmation de recherche	23
2.2	Schématisation du processus d'assemblage par les différents modules du logiciel Trinity.....	25
2.3	Représentation visuelle d'une cassette d'expression synthétisée dans le cadre de cette programmation de recherche.....	28
2.4	Représentation schématique du vecteur d'expression pYES2/CT et du site de clonage multiple	29
2.5	Représentation de la méthode d'assemblage Gibson	33
2.6	Structure de la paroi cellulaire de la levure <i>S. cerevisiae</i>	36
2.7	Représentation du processus de culture, d'induction et d'enrichissement des protéines	39
3.1	Alignement multiple partiel de la séquence en acides aminés de 11 prényltransférases.....	45
3.2	Alignement multiple de la CsPT1 et des deux protéines de <i>R. marginata</i>	47
3.3	Prédiction des structures de la CsPT1 et des deux protéines de <i>R. marginata</i>	49

3.4	Analyse phylogénétique des diverses prényltransférases aromatiques	50
3.5	Test de digestion enzymatique du vecteur pYES2/CT.....	51
3.6	Représentation visuelle du vecteur assemblé par la méthode Gibson.....	52
3.7	Amplification des gènes dans des colonies positives résultant de la transformation dans <i>E. coli</i> DH5 α	53
3.8	Colonies positives de <i>S. cerevisiae</i> après transformation	54
3.9	Amplification des gènes dans des colonies positives résultant de la transformation dans la souche <i>S. cerevisiae</i> INVSc.....	55
3.10	Détermination de la durée d'induction optimale pour l'obtention des protéines d'intérêt.....	56
3.11	Optimisation de la détection des protéines d'intérêt	57
3.12	Amplification des gènes dans des colonies positives.....	59
3.13	Résultats de l'identification de protéines dans les bandes envoyées à la Plateforme de Protéomique	60
3.14	Détection de 5 ppm de CBGA dans un essai enzymatique effectué sur les protéines extraites d'un clone contenant le vecteur pYES2/CT + CsPT1.....	61

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

ADN	Acide Désoxyribonucléique
ARN	Acide Ribonucléique
av. J.-C.	Avant Jésus-Christ
BLAST	Basic Local Alignment Search Tool
BLASTP	Basic Local Alignment Search Tool Protein Database
BSA	Albumine De Sérum Bovin
CB1R et CB2R	Récepteurs Humains De Cannabinoïdes 1 Et 2
CBD	Cannabidiol
CBDA	Acide Cannabidiolique
CBDAS	Acide Cannabidiolique Synthase
CBGA	Acide Cannabigérolique
CBNA	Acide Cannabinolique
CHU	Centre Hospitalier Universitaire
CO ₂	Dioxyde De Carbone
Covid-19	Maladie À Coronavirus 2019
CsPT1-8	Prényltransférase Aromatique De <i>Cannabis Sativa</i> 1 À 8
D164	Acide Aspartique En Position 164
D168	Acide Aspartique En Position 168
DMAPP	Diméthylallyl-Pyrophosphate
DTT	Dithiothréitol
DVA	Acide Divarinolique
EDTA	Acide Éthylènediaminetétraacétique
Endo H	Endoglycosidase H

G222	Glycine En Position 222
GAM-HRP	Anticorps De Chèvre Anti-Souris Conjugué À La Peroxydase De Raifort
GPP	Géranyl-Pyrophosphate
HEPES	Acide 4-(2-Hydroxyéthyl)-1-Pipérazine Éthane Sulfonique
HPLC	Chromatographie En Phase Liquide Haute Performance
I159	Isoleucine En Position 159
IDT	Integrated DNA Technologies
IPP	Isopentényl-Pyrophosphate
L159	Leucine En Position 159
LB	Luria-Bertani
LC-MS-MS	Chromatographie En Phase Liquide Avec Spectrométrie De Masse En Tandem
MEP	Méthylérythritol Phosphate
Mg ²⁺	Cation Magnésium
MgCl ₂	Chlorure De Magnésium
Mn ²⁺	Cation Manganèse
N160	Asparagine En Position 160
NCBI	National Center For Biotechnology Information
NEB	New England Biolabs
O/N	Pendant Une Nuit
OA	Acide Olivetolique
OAC	Acide Olivetolique Cyclase
pb	Paires De Bases
PCR	Réaction En Chaîne Par Polymérase
PDB	Protein Data Bank

PEG	Polyéthylène Glycol
PET	<i>Cis</i> -Perrottétinène
PMSF	Fluorure De Phénylméthylsulfonyle
PT	Prényltransférase Aromatique
PVDF	Polyfluorure De Vinylidène
R108	Arginine En Position 108
RmPT130	Prényltransférase Aromatique De <i>Radula Marginata</i> Contig R4826130
RmPT280	Prényltransférase Aromatique De <i>Radula Marginata</i> Contig R4869280
rpm	Révolutions Par Minute
RT	Température Pièce
SC	Synthétique Complet
SDS-PAGE	Électrophorèse En Gel De Polyacrylamide Avec Laurylsulfate De Sodium
SQDC	Société Québécoise Du Cannabis
TBLASTN	Translated Basic Local Alignment Search Tool Nucleotide Database
TBST	Tampon Tris Salin Avec Tween
TCA	Acide Trichloroacétique
THC	Δ -9-Tétrahydrocannabinol
THCA	Acide Δ -9-Tétrahydrocannabinolique
THCAS	Acide Δ -9-Tétrahydrocannabinolique Synthase
Tris	Tris(Hydroxyméthyl)Aminométhane
Tris-HCl	Chlorhydrate De Tris(Hydroxyméthyl)Aminométhane
UTP	Uridine Triphosphate

LISTE DES SYMBOLES

#	Numéro
%	Pourcentage
% p/p	Pourcentage Massique
°C	Degré Celsius
μL	Microlitre
g	Accélération Gravitationnelle
gr	Gramme
h	Heure
kDa	Kilodalton
mL	Millilitre
mM	Millimolaire
ng	Nanogramme
pH	Potentiel Hydrogène
ppm	Partie Par Million
V	Volt
X	Fois
α	Alpha
β	Bêta

CHAPITRE I

INTRODUCTION

1.1 Le cannabis (*Cannabis sativa*)

1.1.1 Généralités

Cannabis sativa est une plante annuelle à fleurs apparue il y a 38 millions d'années en Asie. Il s'agit de l'une des plus vieilles plantes cultivées et sa présence est dorénavant très répandue sur le globe [1]. Elle fait partie de la famille des Cannabacées et les plants peuvent atteindre une hauteur de 5 mètres dans les meilleures conditions de croissance [2]. Les feuilles emblématiques du cannabis sont généralement composées de 5 à 7 folioles.

Les premières traces de l'utilisation de plantes de type cannabis datent de 7000 ans avant notre ère, en Chine. À cette époque, le cannabis, sous forme de chanvre, était utilisé pour ses fibres dans différents textiles. Ses fibres ont également été identifiées en Égypte sur la momie de Ramsès II, 1200 ans av. J.-C. [1] Le chanvre est de nos jours encore utilisé dans certains textiles et papiers, mais il est également employé dans la composition de plastiques, de produits de beauté et de produits alimentaires.

En comparaison, les premières utilisations du cannabis pour usage curatif datent de 2700 ans av. J.-C. Un ancien cimetière, contenant la tombe d'un shaman datant de 2700 ans av. J.-C., a été découvert en Chine il y a quelques années. L'homme avait été mis sous terre avec de la matière végétale. Des analyses botaniques, phytochimiques et génétiques ont révélé qu'il s'agissait de *C. sativa* avec un pourcentage assez élevé en molécules psychoactives pour présumer un usage médical ou à motif divinatoire [3].

Il est connu que le cannabis est utilisé depuis des milliers d'années dans la médecine traditionnelle orientale [4]. Cependant, son apparition est plus tardive en Amérique. Son utilisation à des fins médicales dans les teintures et différentes concoctions est rapportée dans les années 1800 jusqu'au début des années 1900 [1]. Toutefois, en 1923, par l'ajout du cannabis à la loi fédérale sur l'Opium de 1908, le cannabis devient illégal au Canada [5]. Peu de temps après, dans les années 1930, un Américain nommé Harry Anslinger donne un nouveau nom au cannabis, *marijuana*, afin de faciliter son passage entre les États-Unis et le Mexique. En 1937, après le passage de la Loi sur la taxe sur la marijuana, le cannabis est également considéré comme une drogue illicite aux États-Unis [1].

Il est important de faire la distinction entre le chanvre et le cannabis, qui appartiennent tous deux à la même espèce scientifique *C. sativa*. Tout comme le cannabis, le chanvre a longtemps été considéré comme illégal au Canada jusqu'à sa légalisation en 1998 après la Convention contre le trafic illicite de stupéfiants et de substances psychotropes. La différenciation entre le chanvre et le cannabis est exclusivement basée sur sa teneur en Δ -9-tétrahydrocannabinol (THC), une molécule psychoactive de la plante. Au Canada, la loi indique qu'un plant de chanvre est « une plante de cannabis — ou de toute partie d'une telle plante — dont la concentration de THC dans les têtes florales et les feuilles est d'au plus 0,3 % p/p » [6]. La composition en THC et autres cannabinoïdes, c'est-à-dire les molécules chimiques produites par les plants de *C. sativa* ayant la capacité d'activer les récepteurs cannabinoïdes chez les mammifères, peut être grandement influencée par l'environnement [7]. L'exposition au soleil, l'humidité de l'air et l'irrigation sont des facteurs qui peuvent influencer la teneur en cannabinoïdes tout comme la sous-espèce du plant. Il existe trois grandes sous-espèces informelles connues du genre cannabis, dont la classification n'est toujours pas approuvée par la communauté scientifique, soit *C. sativa*, *C. sativa indica*, *C. sativa ruderalis*.

C. sativa désigne les plants originaires d'Europe et d'Eurasie occidentale, normalement grands et avec une sommité florale longue. À la consommation, il provoque des effets énergisants, euphorisants et stimulants. *C. sativa indica* désigne les

plants originaires d'Inde et du Moyen-Orient. Ils sont normalement plus petits et touffus avec une sommité florale compacte. À la consommation, il provoque des effets calmants et relaxants dus à sa teneur élevée en cannabidiol (CBD). *C. sativa ruderalis* est la sous-espèce communément associée au chanvre sauvage et désigne donc les plants de *C. sativa* ayant un très faible taux en THC [8].

Depuis plusieurs années, différents croisements spécifiques sont effectués entre certains plants pour favoriser une production enrichie en certains types de cannabinoïdes, donnant naissance à ce qu'on appelle des souches ou chimiotypes. Les souches sont de nos jours classées dans une des trois sous-espèces selon leur prédominance en cannabinoïdes. Il existe donc une panoplie de souches disponibles sur le marché. Au Québec, La Société québécoise du cannabis (SQDC) propose plus de 140 souches différentes dans ses succursales [8].

Plusieurs noms seront donnés au cannabis psychoactif sous ses différentes formes. On parle de cannabis, marijuana, pot, weed ou herbe lorsqu'on parle de la forme séchée destinée à être fumée. On peut également retrouver la résine ou l'huile, communément appelé haschish ou l'huile de haschish. Il existe également sous forme de concentré, comme le shatter, ou sous forme vaporisée [9].

1.1.2 Voie de biosynthèse de cannabinoïdes

Comme mentionné dans la section précédente, les cannabinoïdes sont les molécules produites originellement par les plants de *C. sativa* ayant la capacité de se lier aux récepteurs cannabinoïdes chez les mammifères. À ce jour, plus d'une centaine de cannabinoïdes ont été identifiés dans *C. sativa*, et dans d'autres organismes, et leur abondance varie selon la souche utilisée [10]. Les cannabinoïdes sont classés comme polycétides ou composés terpénophénoliques puisqu'ils résultent de la condensation entre un terpène et un acide phénolique [11].

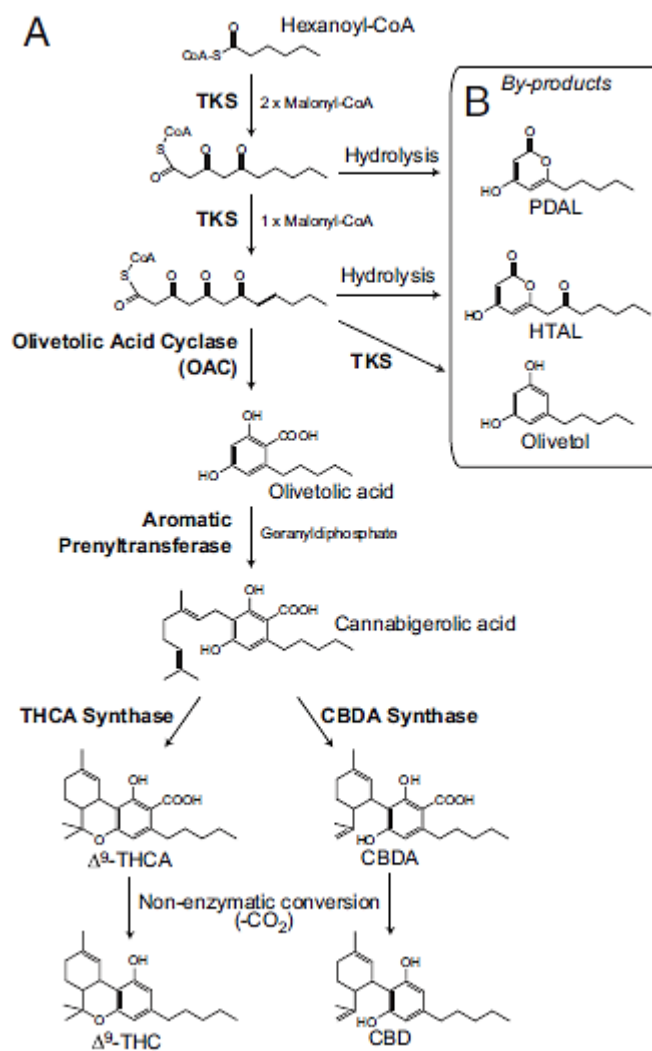


Figure 1.1 Voie de biosynthèse menant à la formation des deux principaux cannabinoïdes.

Les abréviations sont détaillées dans le texte ci-bas. D'après [12].

La biosynthèse de l'acide cannabigérolique, le précurseur des cannabinoïdes, est réalisée par la prénylation du géranyl-pyrophosphate (GPP) sur l'acide olivetolique (OA) par une enzyme appelée prényltransférase aromatique (PT) ou acide cannabigérolique synthase (Figure 1.1). Le GPP provient principalement de la voie du méthylérythritol phosphate (MEP), une voie métabolique dédiée à la synthèse de l'isopentényl-pyrophosphate (IPP) et du diméthylallyl-pyrophosphate (DMAPP). D'autre part, l'OA provient de la biosynthèse des acides gras et résulte de la condensation

d'une unité d'hexanoyl-CoA avec trois molécules de malonyl-CoA par l'acide olivetolique cyclase (OAC) [13].

L'acide cannabigérolique (CBGA) peut ensuite être utilisé comme substrat par différentes enzymes pour mener à la formation de différents cannabinoïdes. C'est entre autres le cas de l'acide Δ -9-tétrahydrocannabinolique synthase (THCAS) et de l'acide cannabidiolique synthase (CBDAS) qui produisent respectivement l'acide Δ -9-tétrahydrocannabinolique (THCA) et l'acide cannabidiolique (CBDA), qui seront ensuite convertis en THC et CBD par une réaction de décarboxylation non enzymatique, comme une simple exposition à la chaleur [12].

La voie de biosynthèse peut également menée à la production de sous-produit comme le lactone de l'acide hexanoyl triacétique (HTAL), le lactone de l'acide pentyldiacétique (PDAL) et l'olivétol [12].

1.1.3 Autres prényltransférases de *C. sativa*

Les récentes études sur le cannabis ont démontré que *C. sativa* produit huit ARN messagers menant à la synthèse de huit protéines complètes différentes de type prényltransférases aromatiques. En d'autres mots, il existe chez *C. sativa* huit homologues appartenant à la famille des prényltransférases, toutes localisées dans les plastides et dont les fonctions sont différentes. En effet, les prényltransférases du cannabis (CsPT) se trouvent dans 3 groupes fonctionnels différents selon leur rôle métabolique. Les CsPT1, 3, 4, 7 et 8 sont dans le groupe menant à la formation de composés terpénophénoliques. La CsPT1 est plus précisément impliquée dans la prénylation de l'acide olivetolique en acide cannabigérolique dans la voie de biosynthèse des cannabinoïdes. La CsPT3 est impliquée dans la production des cannflavines A et B. Les cannflavines sont des flavonoïdes connus pour réduire l'inflammation dans les cellules animales. La CsPT2 ainsi que la CsPT6 sont impliquées dans la biosynthèse des tocophérols, des molécules dérivées de la vitamine E. Ce sont des antioxydants puissants qui jouent un rôle dans la protection des cellules contre le stress oxydatif. La CsPT5, quant à elle, sert à la production

de plastoquinones, des coenzymes qui sont impliquées dans la chaîne de transport des électrons au cours de la photosynthèse [14].

1.1.4 Les trichomes

Les trichomes sont des petits poils translucides présents sur les feuilles et les fleurs de certaines plantes comme le cannabis, mais également la tomate, la menthe et la lavande. Chez *C. sativa*, les trichomes glandulaires sont plus précisément le lieu de production et de stockage des cannabinoïdes. Il existe trois types de trichomes glandulaires sur les plants de cannabis producteurs de cannabinoïdes, c'est-à-dire les plants femelles (Figure 1.2). Les trichomes pédonculés, sessiles et bulbeux sont différenciables par leur morphologie, leur développement et leur physiologie. Les trichomes pédonculés sont les plus hauts, s'élevant à quelques microns au-dessus de l'épiderme grâce à une tige constituée de plusieurs cellules. C'est le type de trichome qui possède la plus grosse tête, ce qui pourrait expliquer que le contenu en cannabinoïdes y soit plus élevé. Le trichome sessile est similaire au trichome pédonculé. En revanche, il possède une très petite tige, ce qui donne l'impression qu'il repose sur l'épiderme. Il a été amené que le trichome pédonculé serait en fait la forme mature du trichome sessile sur les tissus floraux. Le trichome bulbeux est le plus petit en taille et produit une petite quantité de métabolites spécialisés [15,16].

Le trichome glandulaire pédonculé est composé de plusieurs cellules. Les cellules de l'épiderme, les cellules de l'hypoderme et la cellule basale forment ensemble la tige. Les cellules du pédicule et les cellules sécrétrices se retrouvent dans la tête du trichome, aussi appelée cavité de stockage. Les enzymes nécessaires à la biosynthèse des cannabinoïdes se trouveraient principalement dans les cellules sécrétrices. Les cannabinoïdes seraient ensuite entreposés dans des vésicules sécrétrices, dans la cavité de stockage. Les cannabinoïdes sont entreposés à cet endroit, car comme une majorité des métabolites spécialisés, ils présentent une cytotoxicité pour les cellules végétales. Il a été rapporté que le THCA et le CBGA induisent la mort cellulaire par apoptose des cellules végétales. L'hypothèse proposée par la communauté scientifique est donc que les

cannabinoïdes sont produits et entreposés dans les trichomes glandulaires pédonculés afin de prévenir le dommage cellulaire à la plante de cannabis [17].

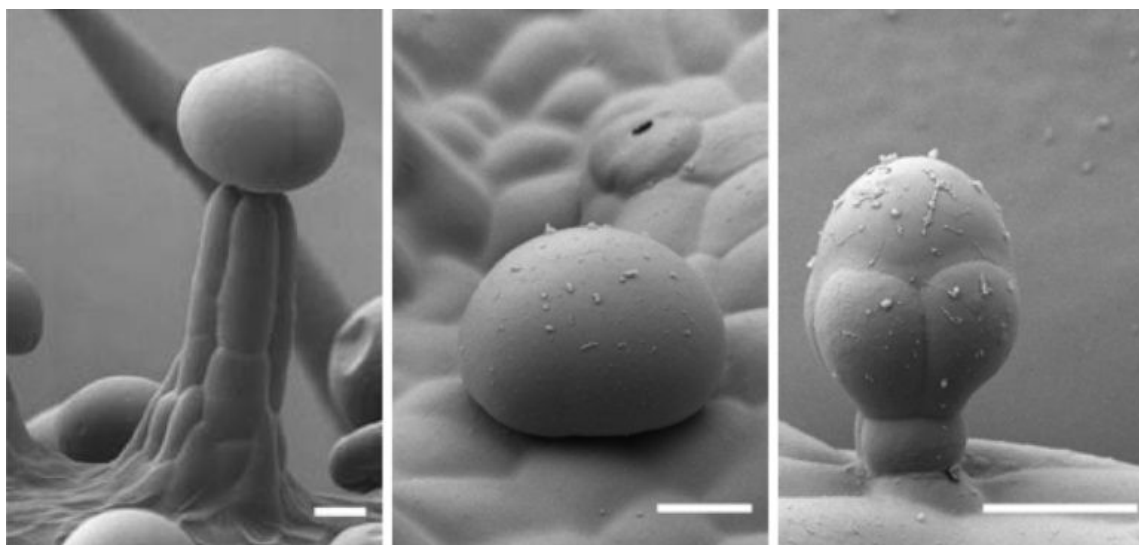


Figure 1.2 Trois types de trichomes du cannabis représentés par microscopie électronique.

De gauche à droite : trichome pédonculé, trichome sessile et trichome bulbeux. Tirée de [16].

Par ailleurs, la couleur des trichomes peut également indiquer la prévalence des cannabinoïdes ainsi que le meilleur moment pour récolter les plants. En effet, la couleur des trichomes varie de transparent à blanc à ambre pendant la maturation du plant. Le stage transparent indique que la plante est en pleine croissance et produit des métabolites spécialisés. Le stage où les trichomes sont blancs serait le moment idéal pour la récolte avec une prévalence de THCA et de CBDA tandis que la couleur ambre indiquerait la dégradation du THCA et du CBDA en acide cannabinoïque (CBNA) [18].

Certains trichomes glandulaires vont plutôt présenter une couleur de rouge à violet et cette coloration ne serait pas causée par le contenu en cannabinoïdes. En effet, c'est l'accumulation d'anthocyanines, un flavonoïde, qui cause cette pigmentation chez différents fruits et fleurs [19].

1.2 Les métabolites spécialisés

1.2.1 Généralités

Tous les organismes vivants possèdent un métabolisme primaire assurant leur croissance, leur développement et leur reproduction. Le métabolisme primaire est essentiel pour la survie de l'organisme et il est constitué de métabolites primaires impliqués dans différents processus métaboliques tels que les voies de signalisation ou de régulation. Les métabolites primaires sont essentiellement des composés organiques comme des protéines, des lipides, des sucres simples, des acides aminés et des acides nucléiques.

En comparaison, le métabolisme spécialisé est présent chez un nombre restreint d'organismes. On le retrouve plus particulièrement chez les plantes, les champignons et certaines bactéries. Le métabolisme spécialisé n'est pas essentiel à la survie de l'organisme, mais lui procure plutôt un avantage. Il existe trois types distincts d'avantages. Les avantages concernant l'attraction, par exemple la plante produit des pigments particuliers dans ses fleurs permettant d'attirer certains pollinisateurs. Il peut également s'agir d'avantages en ce qui concerne la protection, plus particulièrement des métabolites permettant de repousser les herbivores et les agents pathogènes. Finalement, il peut être question d'avantages concernant des interactions, par exemple via une interaction allélopathique. Un exemple d'interaction allélopathique implique la libération de composés biochimiques au niveau du sol, visant à inhiber la croissance des plantes voisines.

Même si au départ les métabolites spécialisés visaient à procurer un avantage à l'organisme qui les produit, il s'avère que les êtres humains ont rapidement compris qu'ils pouvaient les utiliser pour bénéficier de leurs propriétés. En effet, les métabolites spécialisés sont utilisés dans plusieurs types de médecines traditionnelles, mais on les utilise aussi dans les cosmétiques, comme colorants ou encore comme aromates [20]. Il existe trois grandes classes de métabolites spécialisés, soit les terpénoïdes, les composés phénoliques et les alcaloïdes.

Les terpénoïdes sont le groupe le plus abondant des métabolites spécialisés, dénombrant plus de 40 000 molécules. Ils peuvent avoir des structures et des fonctions très variées. L'origine du nom terpène provient de l'allemand *terpentin* désignant la térébenthine, une essence extraite de la résine initialement présente sur le pistachier térébinthe *Pistacia terebinthus* L. La résine de térébinthe était traditionnellement utilisée dans la médecine orientale pour soigner la toux et d'autres problèmes respiratoires. La térébenthine a ensuite été retrouvée dans la résine de plusieurs autres conifères comme le sapin, le pin et le mélèze. L'essence de térébenthine se révèle être un mélange d'hydrocarbures, qui seront nommés terpènes. Les terpènes sont donc des molécules linéaires ou cycliques constituées d'atomes de carbone et d'hydrogène. Ils sont tous dérivés du même précurseur, l'IPP [21].

Si les terpènes sont des hydrocarbures avec une structure moléculaire assez simple, les terpénoïdes sont des molécules souvent plus complexes. Les terpénoïdes sont un regroupement de plusieurs unités d'isoprène qui ont ensuite subi diverses modifications, réarrangements ou cyclisations [22]. C'est pourquoi les terpénoïdes sont également appelés isoprénoïdes. On retrouve entre autres dans la famille des terpènes le menthol, donnant le parfum si caractéristique de la menthe ou encore le carotène, donnant la coloration jaune-orange de certains aliments comme la carotte [23].

Les composés phénoliques sont des molécules aromatiques comportant une fonction alcool. Ils sont divisés en quatre grandes catégories, soit les phénols acides, les flavones, les anthocyanes et les tanins. Ils sont particulièrement efficaces en tant qu'antiseptique, mais ils sont également régulièrement employés dans la composition de parfums et de produits pharmaceutiques.

Les alcaloïdes sont des composés majoritairement cycliques, basiques et contenant une molécule d'azote. Ils sont presque toujours d'origine végétale et font partie intégrante de notre quotidien. On retrouve parmi les alcaloïdes la caféine, la morphine et la nicotine.

Cependant, les cannabinoïdes étant le résultat de la fusion entre deux de ces familles, ils ne sont pas considérés comme une classe à part entière des métabolites spécialisés. À ce jour, plus de 500 composés ont été identifiés dans *Cannabis sativa*, soit plus de 100 cannabinoïdes et 120 terpénoïdes [10].

1.2.2 Les métabolites spécialisés produits par *C. sativa*

Outre les cannabinoïdes, *C. sativa* produit d'autres métabolites spécialisés qui revêtent d'une importance particulière pour les consommateurs, comme les terpènes. Les terpènes volatils, comme les monoterpènes et sesquiterpènes, sont partiellement responsables de l'arôme du cannabis et ont des propriétés thérapeutiques qui pourraient agir en synergie avec celles des cannabinoïdes [10,24]. Étonnamment, le plant de cannabis produit à lui seul plus de 200 terpènes, dont 120 terpénoïdes, et les terpènes peuvent représenter jusqu'à 10 % du contenu des trichomes [25].

Les monoterpènes sont normalement le type de terpènes le plus abondant dans les plants de cannabis. On retrouve principalement comme monoterpènes le limonène, le myrcène et le pinène. Cependant, puisqu'il s'agit de composés volatils, une grande proportion est perdue lors du processus de séchage et d'entreposage du cannabis. Ce sont donc souvent les sesquiterpènes qu'on retrouve en plus grande prévalence dans les produits consommables [25].

Les profils terpéniques peuvent changer d'un cultivar à l'autre, ce qui crée les différences aromatiques entre les variétés [10,26]. Cette variabilité dans les profils chimiques des plants de cannabis est causée par les différences environnementales lors de la croissance.

La famille des composés phénoliques est également présente dans les plants de cannabis. Parmi les quatre grandes catégories de composés phénoliques, ce sont les flavonoïdes qui sont dominants. Les flavonoïdes ont normalement comme fonction de donner la couleur vive aux fleurs, aux fruits et aux légumes. Il existe une vingtaine de

flavonoïdes dans le cannabis et ils représentent environ 10 % du total des molécules présents dans la plante. Cette concentration peut varier d'un cultivar à l'autre [27].

Cependant, parmi tous les composés contenus dans le plant de cannabis, ce ne serait pas un métabolite spécialisé qui serait responsable de son odeur de moufette si caractéristique chez certaines variétés. En effet, la principale molécule responsable de l'odeur serait le 3-méthyl-2-butène-1-thiol, un composé soufré volatil ne faisant pas partie d'une des trois grandes familles de métabolites spécialisés [28].

1.3 L'industrie du cannabis au Québec

1.3.1 Contexte actuel

En octobre 2018, le Canada devient le premier pays d'Occident à légaliser le cannabis. Depuis, l'industrie du cannabis et de ses dérivés occupe une place grandissante dans l'économie canadienne. Avec une contribution au produit intérieur brut évaluée à 12,5 milliards de dollars en 2020, on peut dire qu'il s'agit d'un marché en pleine expansion [29]. Depuis la légalisation, le Canada a édicté des lois concernant la production, la transformation et la vente de cannabis et de ses dérivés. Chaque province peut cependant appliquer la loi sur le cannabis selon ses propres modalités. Au Québec, le cannabis est régi par les lois fédérales ainsi que des réglementations supplémentaires selon des lois nationales. Ainsi, au Québec, seule la Société québécoise du cannabis (SQDC) est autorisée à vendre du cannabis, alors que pour le reste du Canada, les personnes détenant une licence délivrée par Santé Canada peuvent cultiver, transformer et vendre au Canada [30]. De plus, au Québec, tous les producteurs et transformateurs doivent vendre leurs produits à la SQDC, qui elle, les vend par la suite aux consommateurs.

La SQDC a été mise en place au Québec avec comme objectif « [d'] assurer la vente du cannabis dans une perspective de protection de la santé, afin d'intégrer les consommateurs au marché licite du cannabis et de les y maintenir, sans favoriser la

consommation de cannabis ». Depuis la mise en place de la SQDC, celle-ci affirme qu'elle a répondu à sa mission en convertissant 53 % des ventes du marché illicite [31].

1.3.2 Revenus générés au Québec en 2021

Tableau 1.1

Résultats de l'exercice financier de la SQDC pour
l'année financière 2020-2021 en milliers de dollars canadiens
Tiré de [31].

Résultats financiers	
Ventes	537 236
Bénéfice brut	133 770
Charges nettes ⁽¹⁾	67 266
Résultat net	66 504
Situation financière	
Actif total	105 673
Immobilisations corporelles, incorporelles et actifs au titre de droits d'utilisation des baux	61 131
Capitaux propres	2
Ventes par réseaux	
Succursales	486 754
Site transactionnel SQDC.ca	50 482
Total	537 236
Ventes par catégories de produits	
Fleurs séchées	484 247
	82 837 kg
Autres produits (huiles, atomiseurs, etc.)	52 989
	8 692 kg
Total	537 236
	91 529 kg

(1) Les charges nettes représentent les frais de vente et les frais d'administration. Elles incluent également les produits nets de financement.

Au total, l'industrie légale du cannabis au Québec a généré des ventes de plus de 530 millions de dollars pour l'année financière 2020-2021, ce qui représente une augmentation de 72 % par rapport à l'année financière précédente (Tableau 1.1).

La SQDC s'engage chaque année à remettre l'entièreté des ventes globales pour financer la recherche, l'éducation et la prévention en matière de cannabis. Donc, en cumulant les ventes et les revenus fiscaux provenant des taxes, c'est environ 188 millions de dollars qui ont été versés au gouvernement québécois en 2021. Le bénéfice net global s'élève donc à 66,5 millions de dollars, soit un bénéfice nettement supérieur à l'objectif fixé de 50,1 millions. Pour répondre à la demande grandissante envers les produits du cannabis, la SQDC possède maintenant un total de 66 succursales, donc 25 ont ouvert dans la dernière année [31].

1.3.3 Enjeux liés à la production de cannabis

Outre les enjeux associés à la légalisation de cette drogue, la production de la plante en tant que telle engendre à elle seule certains enjeux. D'abord, l'espace utilisé pour la culture du cannabis. Dans un monde où règnent pauvreté et famine, il y a certainement un enjeu lié à l'utilisation d'espace extérieur pour la culture d'une plante à usage médicinal et récréatif. Ce même espace pourrait également être utilisé pour nourrir la population. Au Canada, on estimait en mars 2021 que les zones extérieures détenues par des détenteurs de licence fédérale s'étendaient à une superficie de plus de 7 millions de mètres carrés [32]. Toutefois, une partie de la culture est intérieure pour des raisons de sécurité, pour assurer plusieurs récoltes par année ainsi que pour garantir des conditions stables qui génèrent un produit de qualité constante. Selon le rapport produit par l'entreprise Cannabis Business Times en 2020, 42 % des producteurs de cannabis cultiveraient uniquement à l'intérieur [33]. Or, la culture intérieure engendre de grandes dépenses énergétiques. En Californie, le plus grand état producteur de cannabis des États-Unis, la production intérieure est responsable de 3 % de la consommation en électricité, soit l'équivalent de l'électricité requise pour alimenter 1 million de maisons [34].

Donc, bien sûr, la culture de cannabis est également un enjeu environnemental. À la consommation électrique s'ajoute la grande émission de dioxyde de carbone (CO₂). On estime que pour l'état du Colorado, convertir entièrement la culture intérieure pour la culture extérieure réduirait de 1,3 % l'émission annuelle de gaz à effet de serre.

De plus, comme le cannabis est une plante qui doit être transformée avant d'être consommé sous n'importe quelle forme, la méthode de culture additionnée au processus de transformation engendre un grand impact environnemental. Une étude sur les émissions de dioxyde de carbone associées à la culture du cannabis aux États-Unis a révélé que la production de chaque kilogramme de fleurs séchées peut libérer jusqu'à 5 tonnes de CO₂ (Tableau 1.2). Si on applique ces statistiques au Québec pour l'année financière 2020-2021, les 82 000 kg de fleurs séchées vendues auraient engendré jusqu'à 410 000 tonnes de CO₂ [35].

Tableau 1.2

Empreinte carbone de la production intérieure de cannabis
Tiré de [34].

	Energy intensity (kW/h/kg yield)	Emissions factor (kgCO ₂ emissions/kg yield)	
Lighting	2283	1520	33%
Ventilation & dehumid.	1848	1231	27%
Air conditioning	1284	855	19%
Space heat	304	202	4%
CO ₂ injected to increase foliage	93	82	2%
Water handling	173	115	2%
Drying	90	60	1%
Vehicles		546	12%
Total	6074	4612	100%

Note: The calculations are based on U.S.-average carbon burdens of 0.666 kg/kW/h. "CO₂ injected to increase foliage" represents combustion fuel to make on-site CO₂. Assumes 15% of electricity is produced in off-grid generators.

Puis, il y a la quantité de métabolites spécialisés produits par la plante. Bien souvent, l'accumulation de molécules actives est minime dans les plantes [36], ce qui est une problématique puisqu'il faut une grande biomasse de plantes pour produire une petite quantité de molécules d'intérêt purifiées comme le THC ou le CBD. Grâce aux différents croisements effectués avec les diverses souches de cannabis, le pourcentage de THC a considérablement augmenté dans les dernières années passant de 10 % p/p en 2009 à

14 % p/p en 2019. L'accumulation *in planta* des autres cannabinoïdes demeure cependant faible à moins de 1 % p/p pour le CBD par exemple [37,38].

Finalement, un dernier enjeu qui n'est pas relié à la production, mais qui a un impact sur ce projet de recherche, est que très peu d'études ont été effectuées par le passé sur le cannabis vu son statut pénal. Ainsi, beaucoup de connaissances sont toujours manquantes.

1.4 Radula sp.

1.4.1 Généralités

Les recherches des dernières années ont permis de confirmer qu'il existe des cannabinoïdes ou des analogues chez d'autres espèces végétales comme *Linum usitatissimum*, *Helichrysum umbraculigerum*, *Piper nigrum*, *Radula marginata* et *Radula perrottetii* [39,40].

Les plantes du genre *Radula* font partie de la famille des bryophytes, une famille grandement utilisée dans la médecine traditionnelle chinoise. Les bryophytes sont des plantes ancestrales, c'est-à-dire qu'elles sont dans les premières plantes à avoir évolué du milieu aquatique pour coloniser la terre [41]. Elles ont évolué des algues vertes il y a 472 millions d'années [42]. La famille des bryophytes, qui compte plus de 23 000 espèces, est composée de trois embranchements, soit les hépatiques, les mousses et les anthocérotes. L'embranchement des mousses étant le plus nombreux avec près de 14 000 espèces, suivi par les hépatiques avec 6000 espèces et les anthocérotes à 300 espèces [42].

Les plantes hépatiques peuvent se présenter sous deux formes, les hépatiques feuillues et les hépatiques thalloïdes. C'est d'ailleurs à cause de l'apparence similaire au foie des hépatiques thalloïdes que l'embranchement des hépatiques porte ce nom. Ce sont également des plantes qui ne sont pas vascularisées, c'est-à-dire qu'elles ne possèdent pas de vaisseaux conducteurs de sève comme le xylème et le phloème.

Sans ces vaisseaux, les plantes demeurent très petites et très proches du sol puisqu'elles n'ont pas la possibilité de transporter l'eau sur de longues distances comme les plantes vascularisées. On dit alors que ce sont des plantes qui sont dépendantes de l'eau et on les retrouve surtout en milieu humide. Puisque les hépatiques ne sont pas des plantes évoluées, elles doivent normalement vivre en symbiose avec des champignons capables de leur fournir des nutriments qu'elles ne peuvent pas capter en échange de sucres produits pendant la photosynthèse [39,41].

Les hépatiques sont particulièrement intéressantes puisqu'elles peuvent posséder des corps oléiques. Les corps oléiques sont des organelles à part entière formées de plusieurs gouttelettes lipidiques entourées par une membrane. Ils sont facilement distinguables au microscope par leur taille qui est normalement assez importante et ils sont présents uniquement chez les plantes hépatiques [43]. Chez certaines espèces de *Radula*, comme *Radula complanata*, une espèce qu'on retrouve au Québec, les corps oléiques peuvent être tellement imposants qu'ils obscurcissent l'apparence des feuilles [44]. D'une manière similaire aux trichomes de *C. sativa*, les hépatiques peuvent entreposer les métabolites spécialisés dans ce compartiment distinct. Il est certain que les corps oléiques peuvent prévenir le dommage cellulaire que pourraient induire certaines molécules produites par l'organisme, mais il est également suggéré que ces organelles joueraient un rôle dans la protection contre les herbivores et les agents pathogènes, dans la protection au niveau des changements de température et dans la protection au niveau de la lumière ultraviolette et des radiations [43]. Des rôles similaires donc, aux rôles potentiels joués par les trichomes.

Le genre *Radula* est un genre de plantes hépatiques qui a piqué la curiosité de la communauté scientifique par le contenu retrouvé dans ses corps oléiques dans les dernières années. Il s'agit d'un genre assez répandu dans le monde entier et il regroupe environ 285 espèces différentes [41].

Bien que le genre *Radula* ait piqué la curiosité de plusieurs scientifiques, à ce jour encore très peu d'informations génomiques sont disponibles sur les bryophytes.

Entre 2000 et 2018, seulement 2 génomes ont été séquencés pour cette famille (Figure 1.3). Parmi ces deux génomes, un appartient à l'embranchement des hépatiques et l'autre à l'embranchement des mousses. Le génome séquencé en 2015 de l'embranchement des hépatiques est celui de *Marchantia polymorpha*, et donc n'appartient pas au genre *Radula*. Il est intéressant de comparer la pauvreté des informations disponibles sur les hépatiques avec l'abondance d'informations disponibles sur les trachéophytes et plus précisément sur l'embranchement des eudicotylédones. Dans les eudicotylédones, on retrouve notamment *Cannabis sativa*, dont le génome a été séquencé à une dizaine de reprises depuis 2013.

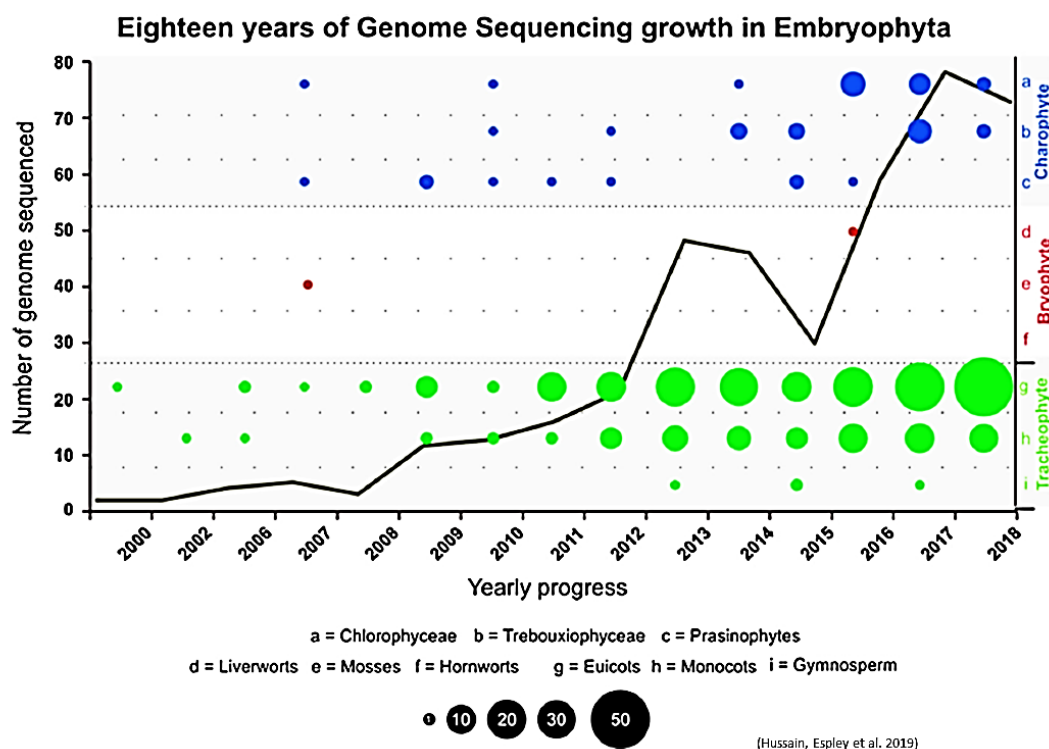


Figure 1.3 Informations génomiques disponibles sur les Embryophytes depuis les années 2000. Tirée de [41].

Alors que très peu d'informations génomiques sont disponibles pour le genre *Radula*, des analyses de métabolites ont toutefois été réalisées. Ainsi, nous avons une certaine connaissance sur les molécules qui les composent.

1.4.2 *Radula perrottetii*

En 1994, une extraction de métabolites a été réalisée sur *R. perrottetii*, une espèce endémique au Japon, par une équipe de chercheurs de Tokushima. Parmi les molécules extraites de l'organisme, l'équipe a rapporté l'identification d'une nouvelle molécule, un bibenzyle analogue aux cannabinoïdes qu'ils ont nommé *cis*-perrottétinène (PET) [45]. Le PET est un analogue structurel du THC produit par *C. sativa*. En effet, les deux molécules sont très similaires, à l'exception qu'ils ont une configuration stéréochimique opposée, *cis* pour le PET et *trans* pour le THC, et que le PET possède un cycle aromatique supplémentaire qu'on ne retrouve pas sur le THC (Figure 1.4).

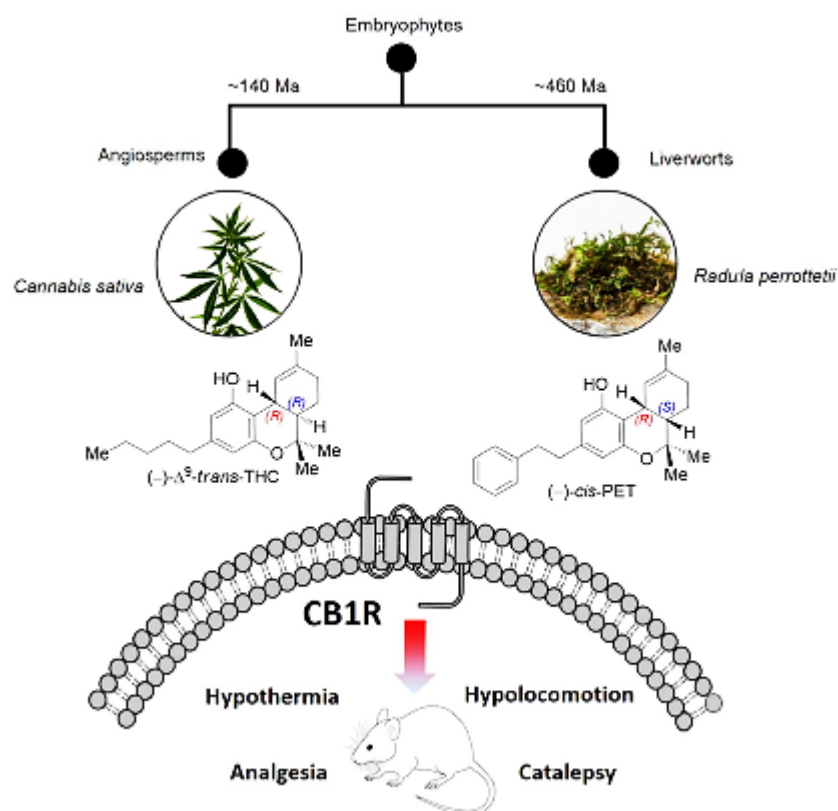


Figure 1.4 Effets induits *in vivo* par le PET, identifié chez *R. perrottetii*, sur le récepteur de cannabinoïdes CB1R de la souris. Tirée de [46].

En 2018, dû à la similitude entre le THC et le PET, une équipe de chercheurs en Suisse s'est intéressée aux propriétés pharmacologiques du PET, et plus particulièrement

à sa psychoactivité. Un composé est classé psychoactif lorsqu'il a la capacité d'influer sur l'activité mentale. Les cannabinoïdes psychoactifs, comme le THC, provoquent des effets chez l'homme grâce à leur capacité à interagir avec les récepteurs humains de cannabinoïdes CB1R et CB2R. Le récepteur CB1R a été localisé dans le cerveau et le système nerveux tandis que le récepteur CB2R a été localisé dans les cellules du système immunitaire. Alors que la liaison entre les cannabinoïdes psychoactifs et le récepteur CB1R semble être responsable des effets mentaux et perceptuels, la liaison au récepteur CB2R induirait la réduction de l'inflammation et de la douleur [1].

Les recherches sur le PET ont rapporté que la molécule avait, tout comme le THC, la capacité de pénétrer dans le cerveau de la souris pour se lier au récepteur cannabinoïde CB1R et induire *in vivo* de l'hypothermie, de la catalepsie, de l'hypolocomotion et de l'analgésie. Il a finalement été possible d'établir que le PET est la première molécule psychoactive découverte parmi toute la division des bryophytes [46].

1.4.3 *Radula marginata*

Après avoir identifié le PET dans une autre espèce de *Radula*, *R. laxiramea*, en 1999, c'est en 2002 que le même composé a été identifié dans *R. marginata*, une espèce endémique de Nouvelle-Zélande et de Tasmanie [46-48]. *R. marginata* possède des feuilles d'un vert foncé dont la bordure épaisse est bien distinguable. On la retrouve principalement sur les roches et les troncs d'arbre [49].

Outre le PET qui a été identifié dans *R. marginata*, plusieurs enzymes et précurseurs appartenant à la voie métabolique ont également été identifiés par différentes équipes de recherche. Ainsi, il est possible, en se basant sur *C. sativa*, de suggérer une voie métabolique menant à la synthèse du PET (Figure 1.5).

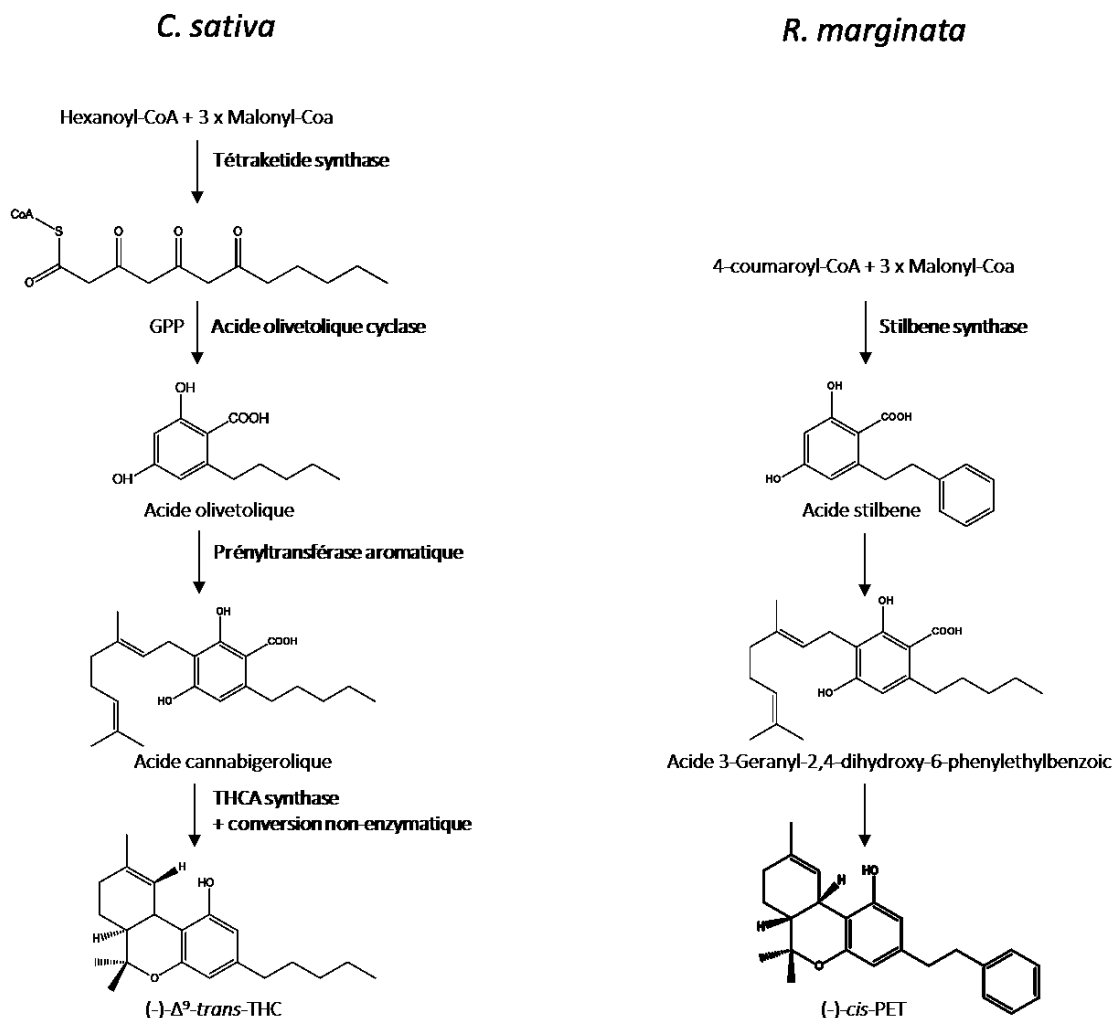


Figure 1.5 Proposition d'une voie menant à la biosynthèse du PET chez *R. marginata* selon les métabolites identifiés.

C. sativa et *R. marginata* sont séparés de 320 millions d'années sur l'échelle de l'évolution [46]. Toutefois, les deux plantes parviennent à produire des molécules actives très similaires. Les enzymes utilisées pour parvenir à cette biosynthèse sont cependant différentes, mais elles dérivent parfois d'un ancêtre commun comme c'est le cas pour l'OAC et la stilbène synthase [41]. De plus, les deux produits résultant de ces voies métaboliques, le PET et le THC ont assez de similarité pour activer tous deux les récepteurs cannabinoïdes CB1R et CB2R.

Les enzymes de *R. marginata* s'avèrent donc très intéressantes compte tenu du brevet américain accordé à Prof. Jonathan E. Page en 2014 sur la séquence de nucléotides de la prényltransférase aromatique 1 de *C. sativa* (*CsPTI*). Le brevet indique que toute protéine ayant une séquence de nucléotides partageant plus de 85 % de similitude avec la *CsPTI* ne peut être exploitée [50]. Ainsi, il n'est pas envisageable d'utiliser les autres prényltransférases aromatiques du cannabis pour remplacer celle concernée par le brevet dû à une trop grande ressemblance pour ce qui est de la séquence en nucléotides.

1.5 Hypothèse et objectifs du projet de recherche

Vu la demande grandissante pour le cannabis et ses produits dérivés, il devient nécessaire d'étudier les différentes options de production afin d'éventuellement pouvoir réduire l'empreinte environnementale et les enjeux éthiques liés à la culture de cannabis. Subséquemment, la biologie synthétique offre des outils permettant le développement d'une plateforme biotechnologique verte basée sur l'utilisation de microorganismes comme milieux de production de cannabinoïdes.

La biologie synthétique est un domaine émergent de la biotechnologie qui combine les principes d'ingénierie à la biologie et vise notamment le transfert d'informations génétiques d'un organisme vers un autre. Cette nouvelle méthode a permis au cours des dernières années de reconstruire et d'étudier des voies métaboliques dans des systèmes d'expression hétérologues. La biologie synthétique bénéficie grandement au secteur industriel par sa facilité à reproduire en quantité des protéines et des biocomposés d'origines diverses sans devoir passer par toutes les étapes de fractionnement, de purification ou de concentration nécessaires pour obtenir ces molécules d'intérêt *in planta*. Les organismes les plus souvent utilisés pour la production de protéines recombinantes et de molécules à haute valeur sont les microorganismes comme la bactérie *Escherichia coli* et la levure *Saccharomyces cerevisiae*.

Compte tenu de la demande de brevet déposée sur la *CsPTI* impliquée dans la voie de biosynthèse des cannabinoïdes chez *C. sativa*, les entreprises se sont mises à rechercher

d'autres voies possibles pour parvenir à la production de cannabinoïdes à grande échelle. Les enzymes impliquées dans la biosynthèse des molécules analogues aux cannabinoïdes s'avèrent intéressantes pour contrer cette problématique.

L'hypothèse de travail associé à la programmation de recherche est qu'il existe une enzyme dans *R. marginata* pouvant effectuer un rôle similaire à la CsPT1 de *C. sativa*. Cette hypothèse est appuyée par la découverte, dans *R. marginata*, d'enzymes analogues à celle impliquée dans la voie de biosynthèse du THC et du CBD de *C. sativa* [39,51].

L'objectif général de la recherche est d'identifier les enzymes impliquées dans la biosynthèse de molécules analogues aux cannabinoïdes chez *R. marginata*, et plus précisément celles pouvant effectuer une réaction de prénylation, dans le but de les caractériser dans la levure *S. cerevisiae*.

Pour ce faire, l'objectif principal de la programmation se décompose en quatre étapes. La première étape étant d'identifier, dans le transcriptome assemblé de *R. marginata*, des séquences d'ARN messagers codant pour des prényltransférases. Ensuite, ces séquences seront soumises à des analyses *in silico* afin de modéliser les structures, de comparer les séquences et d'identifier les domaines et sites de liaison. Puis, les séquences intéressantes seront clonées dans un plasmide et transformées dans la levure afin d'exprimer les protéines. Dans le même ordre d'idée, les protéines seront ensuite purifiées avant d'être caractérisées par différents essais enzymatiques.

CHAPITRE II

MÉTHODES EXPÉRIMENTALES

2.1 Démarche expérimentale

Comme mentionné dans le chapitre précédent, l'objectif général de ce projet de recherche est d'identifier les enzymes, chez *R. marginata*, impliquées dans la voie de biosynthèse menant à la formation de molécules similaires aux cannabinoïdes. La programmation de recherche est principalement axée sur la découverte d'enzymes pouvant effectuer une réaction de prénylation ainsi que leur caractérisation dans la levure *S. cerevisiae*. Ci-dessous est présenté le plan expérimental établi pour atteindre l'objectif fixé (Figure 2.1).

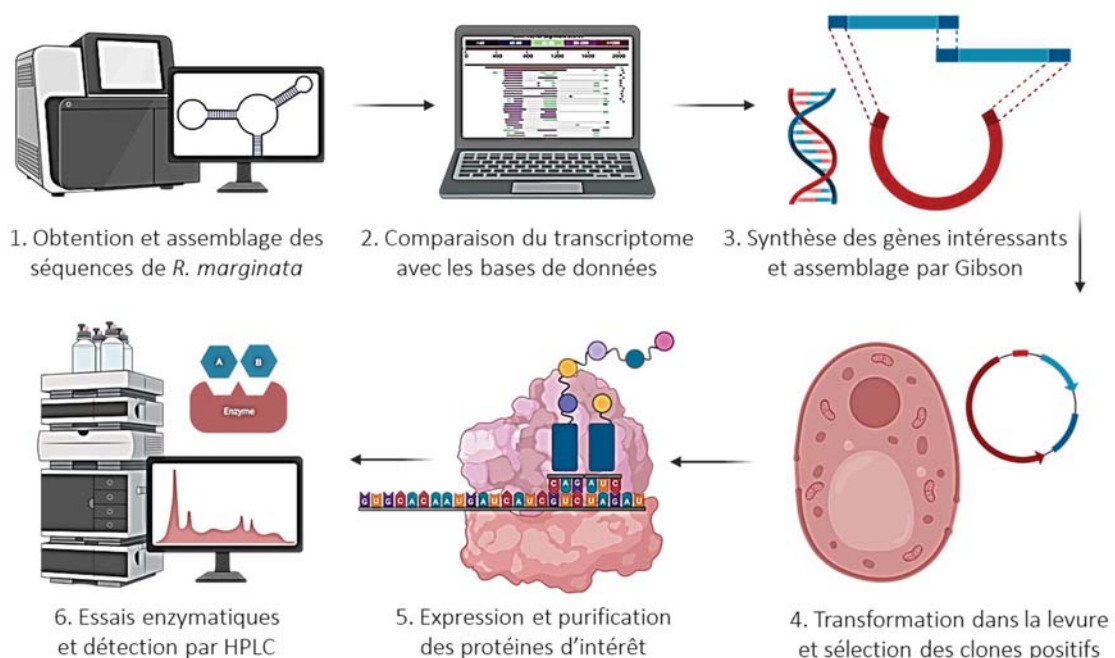


Figure 2.1 Représentation visuelle du plan expérimental de la programmation de recherche.

2.2 Obtention et assemblage des données de séquençage Illumina de *R. marginata*

La première étape du projet visait à obtenir les séquences brutes résultantes du séquençage de l'ARN de *R. marginata*. La plante a été collectée en Nouvelle-Zélande par une équipe de chercheurs d'Allemagne. Ils ont ensuite procédé à l'extraction de l'ARN total et au séquençage avec le système Illumina HiSeq-2000. Les données brutes du séquençage ont par la suite été déposées sur le site du National Center for Biotechnology Information (NCBI) avec les numéros d'identification suivants : BioProject N° PRJNA430820, BioSample N° 08378951 et SRA N° SRR6489347 [39].

Une fois que l'entièreté des données a été téléversée sur la plateforme de calcul Cedar, les données ont été traitées avec le logiciel Trimmomatic v0.38. Ce logiciel permet de nettoyer et filtrer les séquences. Premièrement, il permet de retirer les adaptateurs utilisés lors du séquençage Illumina. Puis, à l'aide d'une fenêtre glissante, les séquences sont scannées et celles présentant des extrémités de mauvaise qualité ou encore les séquences complètes de mauvaise qualité sont retirées. Les séquences trop petites pour procéder à l'assemblage sont également retirées. Les séquences restantes sont ensuite utilisées pour générer un transcriptome *de novo* puisqu'aucun génome de référence n'est disponible à ce jour.

Les séquences ont été assemblées en paires à l'aide du logiciel Trinity v2.9.1. Trinity est divisé en trois modules majeurs soit Inchworm, Chrysalis et Butterfly. La première étape, nommée Jellyfish, consiste à couper les séquences en courts fragments selon la taille désirée (Figure 2.2). Ce paramètre de taille, qu'on désigne *k-mer*, a été ajusté à 32 pour cet assemblage. Ainsi, toutes les 32 bases, les séquences sont coupées. Les séquences sont ensuite traitées par le premier module, Inchworm. Inchworm permet d'assembler les petits fragments en fragments plus longs nommés *contigs*. Ensuite, les *contigs* sont assemblés en *scaffolds* par Chrysalis. Ce module permet aussi de générer des graphiques de Bruijn. Le dernier module, Butterfly, utilise les *scaffolds* et les graphiques pour générer des séquences linéaires en tenant compte de l'épissage et des possibles paralogues [52].

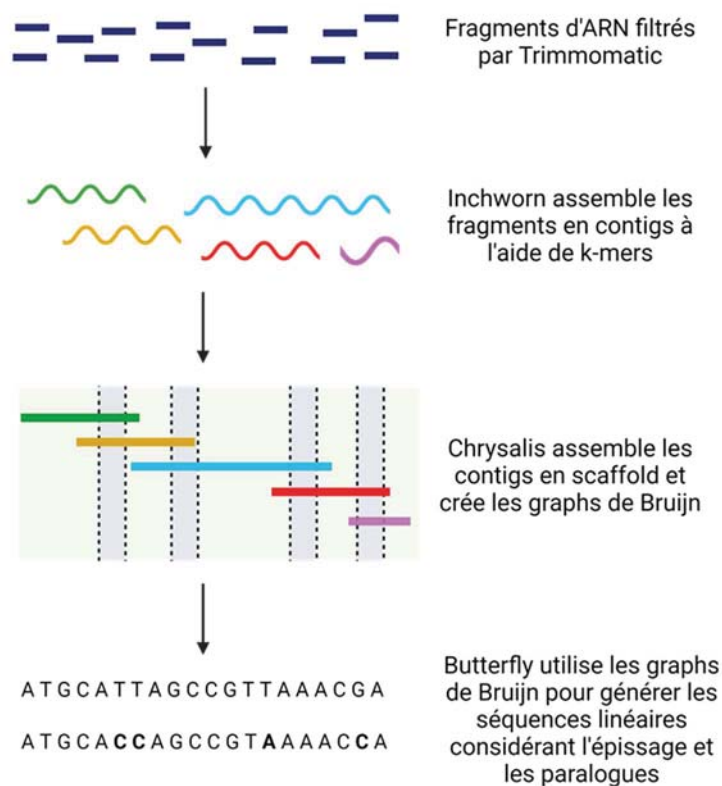


Figure 2.2 Schématisation du processus d'assemblage par les différents modules du logiciel Trinity.

2.3 Analyses *in silico*

2.3.1 Détermination des domaines conservés chez les prényltransférases

Dans le but de résoudre le premier sous-objectif qui est d'identifier des séquences d'ARN messagers codant pour des prényltransférases, la première étape est de déterminer une portion de séquence qui est conservée dans les protéines de la famille des prényltransférases.

Premièrement, il faut utiliser la séquence en acides aminés de la CsPT1 pour effectuer une recherche dans la base de données protéique à l'aide de BLASTP. BLAST est une méthode de recherche basée sur l'alignement local entre une séquence requête et une base de données. Dans ce cas-ci, l'acronyme BLASTP indique qu'il s'agit d'une recherche depuis une séquence protéique dans une base de données protéique. Parmi les

résultats de la recherche, 10 autres protéines ont été sélectionnées afin de réaliser un alignement multiple des séquences d'acides aminés. L'alignement multiple par ClustalW a révélé une séquence conservée dont le domaine a ensuite été identifié par Interpro. Interpro est une ressource en ligne qui offre une classification des protéines et une identification des domaines en utilisant la collaboration de diverses bases de données.

2.3.2 Identification des séquences pertinentes chez *R. marginata*

La séquence en acide aminé du domaine conservé a ensuite été utilisée pour faire une recherche dans la base de nucléotides du transcriptome assemblé de *R. marginata* via une opération qu'on désigne TBLASTN [53,54]. TBLASTN est une opération disponible dans le programme de recherche heuristique BLAST où une séquence protéique est alignée avec une base de données de nucléotides traduite dans les six cadres de lecture possibles. Parmi les résultats de la recherche, seules les séquences ayant un cadre de lecture complet ont été conservées. Un simple alignement multiple par ClustalW a finalement été réalisé afin de donner une représentation visuelle de l'homologie entre les séquences des protéines identifiées et la CsPT1.

2.3.3 Prédiction des structures des protéines d'intérêts

Par la suite, plusieurs analyses de prédiction ont été réalisées avec le logiciel PredictProtein, dont des prédictions de structure secondaire, d'accessibilité aux solvants, les sites de liaisons et la localisation cellulaire. Les prédictions générées par PredictProtein ont été validées par le logiciel SWISS-MODEL. SWISS-MODEL est un serveur de modélisation d'homologie de structure. Il permet de produire une prédiction de structure tridimensionnelle et d'identifier les sites de liaisons des ligands. Puis, le logiciel de visualisation tridimensionnel PyMOL a été utilisé pour identifier les acides aminés formant la pochette catalytique autour de l'un des substrats. Finalement, un arbre phylogénétique a été réalisé afin de créer une représentation visuelle permettant de visualiser les relations entre différentes protéines. Afin de réaliser cette analyse, la séquence en acides aminés de 41 prényltransférases, dont certaines ont préalablement été utilisées dans une publication pour faire une analyse phylogénétique, ont été alignés

avec le programme d'alignement multiple ClustalW en utilisant les paramètres par défaut [14,55]. L'alignement multiple a ensuite été optimisé avec le programme Gblocks, ce qui a permis de retirer les positions faiblement alignées et les régions divergentes [56]. Le fichier résultant a été utilisé pour créer l'arbre phylogénétique dans le logiciel MEGA7 selon la méthode Neighbor-Joining [57]. Les pourcentages indiqués sur les branches de l'arbre ont été calculés avec le test bootstrap réglé à 1000 réplicats [58].

2.4 Conception des cassettes d'expression

Lors de la construction d'une cassette d'expression, il faut songer à la manière dont la cassette sera utilisée. En plus d'ajouter la séquence codante pour la protéine, il est parfois nécessaire d'ajouter certains éléments favorisant l'expression, la détection ou encore la purification de la protéine. Dans ce sens, l'ensemble des éléments sélectionnés pour la création des cassettes d'expression dans cette programmation de recherche a été minutieusement pensé (Figure 2.3).

Pour la création des cassettes d'expression, les séquences utilisées sont les séquences d'acides nucléiques codant pour les protéines complètes identifiées dans le transcriptome de *R. marginata* ainsi que la séquence de la protéine CsPT1 de *C. sativa*. Selon les recommandations du fournisseur du vecteur d'expression, une séquence consensus pour la levure a été ajoutée devant le codon d'initiation de la transcription ATG. L'ajout de la séquence consensus chez la levure augmente la stabilisation de la protéine à l'extrémité N-terminale et aurait donc un impact dans la prévention de dégradation de la protéine. Aussi, l'ajout de la séquence consensus permettrait d'augmenter l'efficacité de l'initiation de la transcription par 2 ou 3 fois [59].

Puis, on retrouve le codon d'initiation suivi de la séquence codante. À la fin de la séquence codante, un site de clivage pour la thrombine a été ajouté. La thrombine est une protéase, c'est-à-dire une enzyme capable de cliver d'autres séquences protéiques. Ce type de clivage est souvent utilisé en biologie synthétique afin de retirer une protéine porteuse ou un épitope. Il a été démontré par le passé que ces éléments, soit une protéine porteuse

ou un épitope, peuvent affecter le repliement de la protéine et ainsi influencer sur sa fonction biologique. La thrombine reconnaît spécifiquement la séquence LVPR[^]GS et clive au niveau de la liaison entre l'arginine et la glycine. Il est ensuite facile de la retirer par une simple chromatographie d'affinité sur billes de sépharose activée à la benzamidine [60]. De plus, l'utilisation de la thrombine s'avère intéressante lors de la manipulation de protéines membranaires puisqu'elle résiste à plusieurs détergents. Les détergents sont essentiels lors de l'extraction de protéines membranaires. Ce sont des molécules amphiphiles, c'est-à-dire possédant une tête hydrophile et un corps hydrophobe, qui vont former des micelles en milieu aqueux. Les micelles vont recréer un environnement similaire à la membrane cellulaire, ce qui va permettre de solubiliser les protéines membranaires extraites [61].

Enfin, on retrouve l'épitope 1D4. Cette courte séquence hydrophile provient de l'extrémité C-terminale de la rhodopsine bovine. Elle est couramment utilisée, conjointement avec l'anticorps monoclonal 1D4, pour la localisation, la purification et la détection de protéines membranaires [60]. Le positionnement de l'épitope 1D4 en fin de cassette est stratégique afin de ne pas interférer avec le possible signal peptide des protéines recombinantes. Les différentes cassettes d'expression nécessaires pour ce projet de recherche ont toutes été synthétisées par la compagnie Bio Basic, un fabricant de gènes (Annexe A).

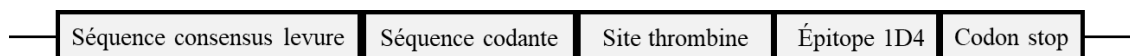


Figure 2.3 Représentation visuelle d'une cassette d'expression synthétisée dans le cadre de cette programmation de recherche.

2.5 Sélection du vecteur d'expression

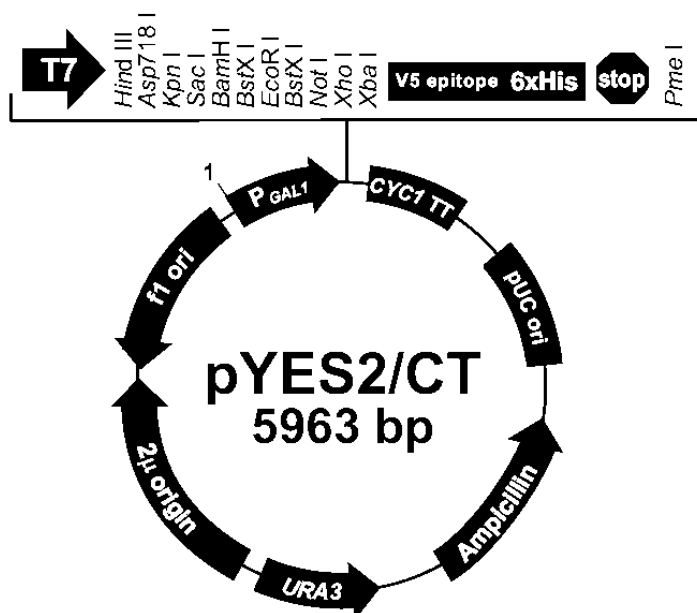
Le vecteur utilisé pour l'expression du transgène dans la levure est le vecteur pYES2/CT (Figure 2.4). Ce vecteur d'expression a précisément été choisi pour sa disponibilité dans notre laboratoire. Il fait 5963 paires de bases (pb) et contient les différents éléments listés dans le tableau ci-dessous (Tableau 2.1).

Tableau 2.1

Éléments du vecteur commercial pYES2/CT pour expression dans la levure

Éléments du vecteur	Position
Promoteur GAL1	1-451
Promoteur T7	475-494
Site de clonage multiple	501-594
Épitope V5	607-648
Région polyhistidine	658-675
Termineur CYC1	708-961
Origine pUC	1145-1818
Gène de résistance à l'ampicilline*	1963-2823*
Gène URA3*	2841-3948*
Origine 2 μ	3952-5423
Origine f1*	5491-5946*

* Brin complémentaire

**Figure 2.4** Représentation schématique du vecteur d'expression pYES2/CT et du site de clonage multiple.

Le promoteur est une séquence qui indique le site d'initiation de la transcription d'un transgène. Dans le vecteur pYES2/CT, le promoteur GAL1 est un promoteur inductible endogène de la levure *S. cerevisiae* utilisé pour exprimer à haut niveau des protéines recombinantes. Le promoteur est activé en présence de galactose, mais réprimé en présence de glucose. Il est donc très important de retirer complètement le milieu de culture initial de la levure contenant du glucose lors de l'induction de la transcription. À ce moment, il faut remplacer le milieu initial par un milieu de culture dont la source de sucre est le galactose. Le vecteur d'expression comprend également une séquence pour le promoteur T7. Ce promoteur est utilisé pour exprimer des protéines recombinantes dans la bactérie, ce qui ne sera pas nécessaire dans le cadre de ce projet de recherche.

Puis, on retrouve le site de clonage multiple. Il s'agit d'une courte séquence, qui fait ici 93 paires de bases, dans laquelle on retrouve une grande quantité de sites de restriction. Les sites de restriction sont des séquences de quelques acides nucléiques reconnus par des enzymes de restriction. Ce système est communément utilisé en biologie synthétique pour couper des fragments d'ADN à doubles brins. Pour ce projet, le site de clonage multiple est complètement retiré ainsi que les épitopes V5 et 6x His (région polyhistidine) afin de réduire le risque d'erreur possible dans le cadre de lecture lors de la traduction de la protéine. Aussi, le retrait des épitopes permet d'utiliser celui préalablement choisi pour les cassettes d'expression, soit l'épitope 1D4. Les enzymes de restriction utilisées pour effectuer le retrait du site de clonage multiple sont Hind III et Pme I. Hind III reconnaît la séquence A[^]AGCTT au début du site de clonage multiple et l'enzyme Pme I reconnaît la séquence GTTT[^]AAAC située après la région polyhistidine. Le site de clonage multiple est suivi du terminateur CYC1 qui est endogène chez la levure. Un terminateur est une courte séquence d'ADN reconnue par la machinerie de transcription comme signal de terminaison de la transcription d'un gène.

On retrouve également dans le vecteur pYES2/CT différentes origines de réplication. Une origine de réplication est une courte séquence unique qui permet d'indiquer, à la machinerie du système d'expression hétérologue, le site d'initiation de la réplication de l'ADN. Par la suite, le vecteur sera donc répliqué dans l'hôte et disponible

en plusieurs copies. La séquence de l'origine de réplication exerce une grande influence sur le nombre de copies qui sera produit dans l'hôte. En effet, les séquences sont soumises au phénomène de régulation par le système d'expression hétérologue et certaines séquences seront régulées à la hausse alors que d'autres seront régulées à la baisse. Il y a trois origines de réplication présentes dans le vecteur dont l'origine pUC, une origine permettant le maintien et la réplication du vecteur chez la bactérie *E. coli*. L'origine pUC est une origine permettant de générer un grand nombre de copies du vecteur dans la bactérie, soit environ 500 copies. Cette origine de réplication est essentielle dans le cadre de cette programmation de recherche. Afin de pouvoir exprimer des protéines recombinantes dans la levure, les vecteurs d'expression assemblés sont d'abord répliqués dans la bactérie *E. coli* afin de récolter le plus grand nombre possible de copies du vecteur avant la transformation dans la levure. La seconde origine est l'origine 2μ , une séquence initialement identifiée dans un plasmide naturellement présent chez la levure. Lorsque la séquence correspondant à l'origine 2μ est insérée dans un vecteur d'expression hétérologue, le vecteur est maintenu de façon épisomale et se réplique avec un haut nombre de copies, soit entre 10 et 40 copies en moyenne par cellule. Finalement, on retrouve l'origine de réplication fl dont le rôle dans le vecteur pYES2/CT est de permettre le sauvetage de l'ADN simple brin.

Ensuite, on retrouve dans le vecteur le gène de résistance à l'ampicilline permettant une sélection des colonies positives de la bactérie *E. coli* après l'étape de transformation. Les colonies ayant correctement intégré au moins une copie du vecteur pourront synthétiser le gène de résistance à l'ampicilline et survivre sur un milieu de culture contenant cet antibiotique.

Finalement, on retrouve le gène URA3 qui est utilisé ici comme marqueur auxotrophe. Ce gène est utilisé pour permettre une sélection des colonies positives de la levure après l'étape de transformation. Il code pour l'orotidylate décarboxylase, une enzyme permettant la synthèse de l'uridine triphosphate (UTP). L'UTP est un nucléotide ayant comme base azotée l'uracile. Les levures ayant intégré le vecteur et exprimant le gène URA3 ont donc la capacité de survivre dans un milieu de culture déficient en uracile.

2.6 Assemblage du vecteur d'expression par Gibson Assembly

Tel que mentionné précédemment, le vecteur d'expression pYES2/CT a été digéré avec les enzymes de restrictions Hind III et Pme I afin de retirer le site de clonage multiple, laissant place à un vecteur linéarisé. Cependant, les deux enzymes de restriction ne sont pas efficaces dans le même tampon commercial. L'enzyme Hind III est optimale dans le tampon NEB 2.1 alors que l'enzyme Pme I est optimale dans le tampon Cutsmart. Les deux enzymes de restriction et les tampons de réaction proviennent de la compagnie New England BioLabs (NEB). Cette étape de digestion a dû être optimisée afin d'obtenir une quantité suffisante de vecteurs digérés pour pouvoir procéder à l'assemblage du vecteur et des cassettes d'expression. Afin de déterminer les meilleures conditions pour la digestion, différentes quantités de vecteurs ont été digérées, dans différents tampons et avec une durée d'incubation différente. Différentes quantités, soit 200 ng, 1000 ng et 2500 ng ont été testées pour la digestion dans les tampons NEB 2.1 et Cutsmart pour une période d'incubation variant de 3 h à toute une nuit. Le premier test consistait à digérer 200 ng de vecteur avec l'enzyme Hind III à 37 °C pour 3 heures. Pour le deuxième test, 1000 ng de vecteur ont été digérés en premier lieu par Hind III à 37 °C pour 3 heures dans le tampon NEB 2.1, suivi de l'ajout de Pme I et du tampon Cutsmart dans le mélange réactionnel pour une incubation d'une nuit à 37 °C. Le dernier test consistait à digérer simultanément 2500 ng de vecteur par Hind III et Pme I pendant toute la nuit à 37 °C dans le tampon NEB 2.1. Les différentes réactions de digestion ont ensuite été migrées sur un gel d'agarose 1 % afin de déterminer la méthode optimale pour la linéarisation du vecteur pYES2/CT. Après la migration, la bande correspondant à la taille du plasmide digérée a été extraite du gel et purifiée avec la trousse Gel + PCR extraction kit de Bio Basic. La concentration du vecteur après purification a été mesurée à l'aide du NanoDrop, un spectrophotomètre à microvolumes.

Les différentes cassettes d'expression contenant les gènes d'intérêt ont été amplifiées par PCR afin d'ajouter les extrémités qui serviront à faire la recombinaison homologue lors de l'assemblage de l'insert dans le vecteur. Un prolongement d'environ une vingtaine de nucléotides est nécessaire à chaque extrémité des cassettes. Toutes les

amorces utilisées lors de cette programmation de recherche ont été commandées chez Integrated DNA Technologies (IDT).

L'assemblage par la méthode Gibson a été réalisé à l'aide de la trousse HiFi DNA Assembly Cloning Kit de NEBuilder (Figure 2.5). L'assemblage a été réalisé selon les instructions du fabricant avec un ratio vecteur : insert de 1 :2. La méthode Gibson a été choisie dans le cadre de ce projet pour sa simplicité et la disponibilité des différentes composantes dans le laboratoire. La méthode consiste simplement à mélanger le plasmide linéaire, l'insert amplifié et un mélange commercial contenant une exonucléase, une ADN polymérase et une ADN ligase. Dans un premier temps, l'exonucléase vient dégrader l'extrémité 5' du vecteur et de l'insert. Puis, à la suite de l'hybridation entre l'insert et le vecteur, l'ADN polymérase vient synthétiser les bases manquantes et l'ADN ligase vient lier le tout pour compléter l'assemblage. Il est possible d'insérer plusieurs inserts dans un vecteur, donc le ratio vecteur : insert dépend du nombre et de la taille des inserts à assembler [62].

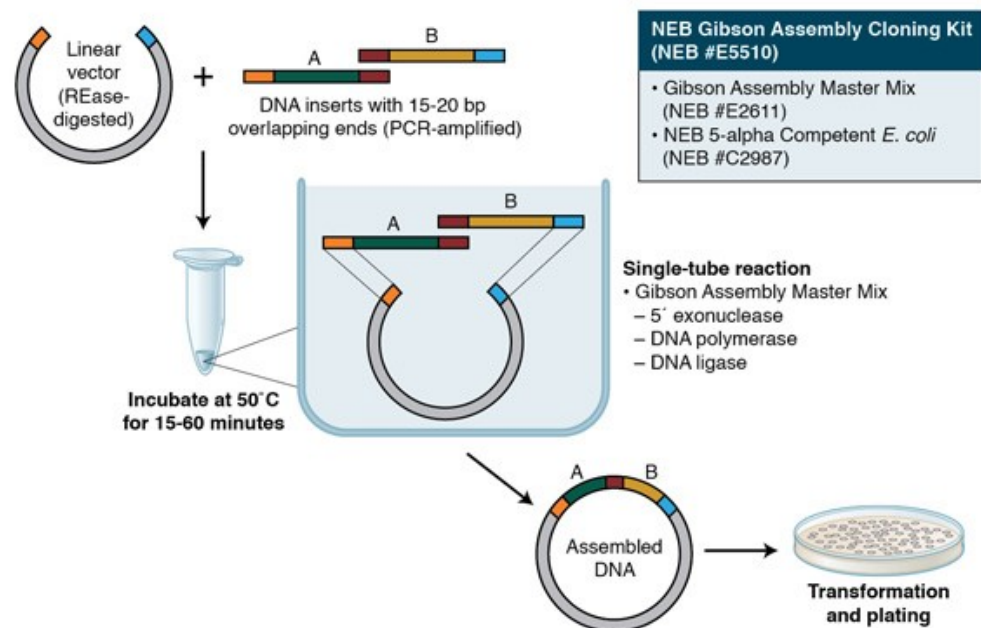


Figure 2.5 Représentation de la méthode d'assemblage Gibson.
Tirée de [63].

2.7 Transformation dans la bactérie *E. coli*

Une fois la réaction d'assemblage terminée, le vecteur assemblé a directement été transformé dans la bactérie *E. coli* DH5 α chimiocompétente de la compagnie NEB. Une transformation est une technique communément employée en biologie synthétique qui consiste à insérer un fragment d'ADN exogène dans une cellule. Plusieurs méthodes peuvent être employées pour effectuer une transformation dans la bactérie. Ici, c'est une transformation par choc thermique qui a été effectuée selon le protocole One Shot TOP10 Competent Cells de Invitrogen. La méthode consiste à faire subir aux cellules un changement rapide de température, ce qui engendre une dilatation de la paroi cellulaire et de la membrane plasmique pour permettre l'entrée de l'ADN exogène. Pour chaque cassette d'expression, 50 μ L de bactéries *E. coli* DH5 α chimiocompétentes ont été mis en présence de 2 μ L du mélange réactionnel de l'assemblage par la méthode Gibson.

Des colonies positives ont été obtenues après une incubation d'une nuit à 37 °C sur une gélose Luria-Bertani (LB) contenant de l'ampicilline 1X. Parmi toutes les colonies ayant poussé sur le milieu de sélection, quelques-unes ont été choisies pour valider l'insertion du vecteur complet, avec l'insert, par la méthode de PCR sur colonie. Après la réaction d'amplification en chaîne par polymérase, des colonies se sont révélées être positives pour chaque vecteur transformé contenant les différentes cassettes.

Les colonies positives validées ont été mises en culture dans 10 mL de milieu liquide LB avec ampicilline. Après avoir cultivé les bactéries pendant une nuit pour permettre la réplication des vecteurs, ils ont finalement été extraits de la bactérie à l'aide de l'ensemble Presto mini kit de Geneaid. Le protocole d'extraction a été réalisé selon les recommandations du fabricant en utilisant 7 mL de culture bactérienne et en suivant les directives spécifiques pour avoir de meilleures réactions de séquençage en aval.

Les différents vecteurs purifiés ont été envoyés à la plateforme d'analyse génomique de l'Université Laval pour un séquençage de type Sanger afin de vérifier l'intégrité des séquences. Des amorces spécifiques ont été conçues pour chaque vecteur afin de pouvoir

visualiser l'entièreté des séquences. Les vecteurs dont l'intégrité de l'insert a été confirmée par séquençage ont été utilisés pour être transformés dans la levure.

2.8 Choix de la levure *S. cerevisiae* comme système d'expression hétérologue

La levure est un organisme eucaryote unicellulaire appartenant au règne des champignons. Elle se reproduit normalement de manière asexuée par mitose. La reproduction est dite asymétrique puisqu'elle consiste en la formation d'un bourgeon qui se détachera ensuite de la cellule mère [64].

La paroi cellulaire de levure est particulièrement rigide puisqu'elle est faite de trois couches (Figure 2.6). La couche externe est composée de mannoprotéines alors que les deux couches externes sont constituées de polymères de glucose, soit de β -glucanes et de chitines [64]. À elle seule, la paroi cellulaire peut représenter jusqu'à 30 % de la masse sèche de la cellule et jusqu'à 50 % du volume [65].

Les mannoprotéines peuvent représenter jusqu'à 40 % de la paroi cellulaire. À ce jour, 40 mannoprotéines différentes ont été identifiées dans la paroi cellulaire de la levure [66]. Puisqu'elles forment la couche externe, les mannoprotéines sont impliquées en ce qui touche à la reconnaissance entre les cellules, mais elles ont également un rôle de protection en restreignant l'accès à la paroi interne et à la membrane cellulaire [67].

Les β -glucanes peuvent représenter jusqu'à 60 % de la paroi cellulaire et sont présentes sous deux formes : 1,6 β -glucane et 1,3 β -glucane. Les 1,3 β -glucanes forment la couche fibreuse de la paroi cellulaire tandis que les 1,6 β -glucanes joueraient un rôle de liaison avec les autres composants de la paroi cellulaire et auraient également la capacité d'augmenter la réponse immunitaire en se liant avec des macrophages ou des leucocytes granulaires [68-70].

Les chitines représentent environ 2 % de la paroi cellulaire et jouent un rôle dans l'imperméabilité de la paroi cellulaire puisqu'il s'agit de polymères insolubles.

Les chitines joueraient également possiblement un rôle dans la conservation de l'intégrité structurale de la cellule et dans le processus de bourgeonnement [65,68,69].

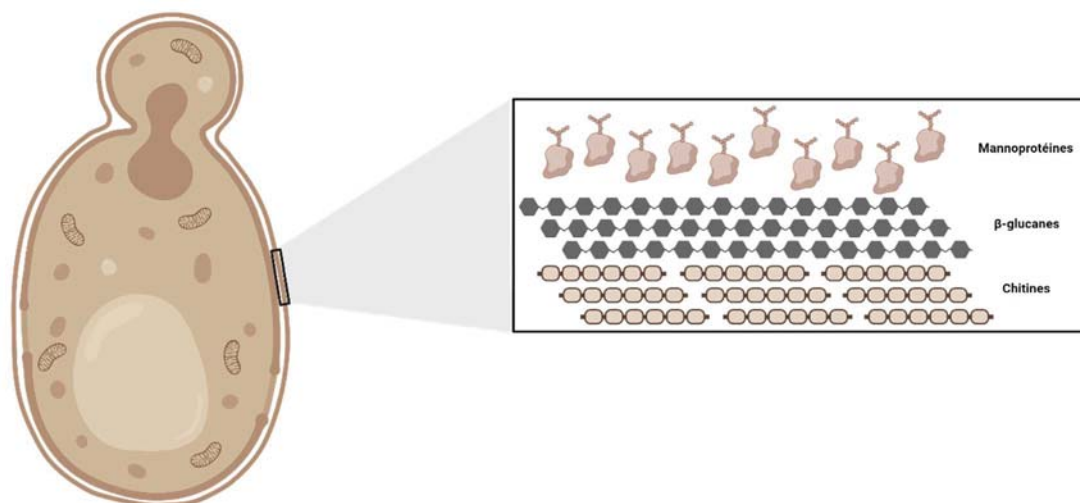


Figure 2.6 Structure de la paroi cellulaire de la levure *S. cerevisiae*.

La levure *S. cerevisiae* est particulièrement connue depuis des milliers d'années pour son utilisation dans la fermentation de l'alcool et du pain. Elle a la capacité de générer de l'alcool et du gaz pendant la fermentation en métabolisant le sucre disponible dans le milieu de culture. Cette espèce de levure est d'ailleurs devenue un organisme modèle pour la science.

L'utilisation de la levure *S. cerevisiae* comme système d'expression hétérologue dans le cadre de la réalisation de ce projet s'est avérée être une décision réfléchie. Comme l'objectif général est d'identifier et de caractériser des enzymes dans *R. marginata*, l'utilisation d'un organisme également eucaryote comme *S. cerevisiae* pour exprimer les protéines d'intérêts s'avère intéressant dû à la similarité des mécanismes de translation et de traduction. De plus, *S. cerevisiae* est un organisme modèle ce qui implique qu'il s'agit d'un organisme bien étudié par la communauté scientifique et qu'il existe plusieurs outils disponibles sur le marché pour parvenir à y exprimer les protéines d'intérêt. Il est donc plus facile d'utiliser un organisme bien documenté et référencé comme *S. cerevisiae* pour caractériser de nouvelles protéines.

2.9 Transformation dans la levure

Les vecteurs purifiés extraits de la bactérie ont été transformés dans la levure. 100 ng de chaque vecteur a été incubé en présence de 50 μ L de cellules compétentes de *S. cerevisiae* INVSc. La transformation par choc thermique a été effectuée selon le protocole Yeastmaker Yeast Transformation System 2 de Clontech. Après quelques jours à 30 °C, des colonies ont été obtenues sur une gélose Synthetic Complete (SC) minimale sans uracile pour chaque vecteur transformé. De la même façon qu'après la transformation dans la bactérie, certaines colonies de levure ayant poussé sur le milieu de sélection ont été choisies pour valider l'insertion du vecteur complet avec l'insert par la méthode de PCR sur colonie. L'amplification en chaîne par polymérase a révélé que chaque vecteur a réussi à intégrer des cellules de levure lors du choc thermique.

Après avoir confirmé les colonies positives par la méthode PCR, la réaction d'amplification a été répétée et les fragments amplifiés ont été purifiés à l'aide de la trousse Gel/PCR DNA fragments extraction kit de Geneaid. Les fragments purifiés ont été envoyés à la plateforme d'analyse génomique de l'Université Laval pour un séquençage de type Sanger afin de vérifier l'intégrité des séquences. Les amorces utilisées pour le séquençage sont les mêmes que celles qui avaient été conçues et utilisées après la transformation bactérienne. Les clones positifs confirmés par séquençage ont été sélectionnés pour être cultivés et induits.

2.10 Expression des protéines recombinantes

L'induction des protéines a été réalisée selon les instructions d'Invitrogen, le fabricant du vecteur d'expression pYES2/CT. Les clones utilisés pour exprimer les protéines recombinantes sont le clone 4.1 pour RmPT280, le clone 4 pour RmPT130, le clone 4 pour CsPT1 et le clone 2 pour pYES2/CT. Les cellules de *S. cerevisiae* ont d'abord été cultivées dans un milieu de culture SC minimal sans uracile, dont la source de sucre est du glucose 2 %. Puis, les cellules ont été centrifugées et le milieu de culture a été remplacé par le milieu d'induction, dont la source de sucre est du galactose 2 %,

afin d'activer le promoteur inductible. Selon les recommandations du fabricant, les cellules ont été récoltées après 0 h, 4 h, 8 h, 12 h, 16 h et 24 h d'induction. Avec l'arrivée de la Covid-19 et les différentes mesures imposées par le gouvernement du Québec, dont les couvre-feux, les heures de prélèvement ont été modifiées pour 0 h, 4 h, 8 h, 12 h, 24 h et 48 h. Les cellules récupérées ont été centrifugées et les culots ont été conservés à -80 °C.

2.11 Extraction et enrichissement des protéines recombinantes

Selon les analyses *in silico* effectuées précédemment, les protéines d'intérêt seraient localisées dans les plastides. Après l'expression de ces protéines dans un système d'expression hétérologue qui ne contient pas de plastides, comme la levure *S. cerevisiae*, les protéines recombinantes sont plus susceptibles de se retrouver dans le réticulum endoplasmique. Le réticulum endoplasmique est un organite cellulaire eucaryote annexé à la membrane nucléaire et il s'agit du lieu de la synthèse protéique. Des fragments de cet organite peuvent être enrichis et récoltés après la lyse cellulaire par ultracentrifugation différentielle dans une fraction nommée la fraction microsomale.

Le protocole utilisé pour l'extraction et l'enrichissement des protéines recombinantes est un protocole créé en se basant sur les recommandations d'Invitrogen pour le vecteur pYES2/CT et sur la demande de brevet déposée sur la *CsPTI*. Le protocole est basé sur les culots d'une culture de 200 mL de levure dans le milieu d'induction (Figure 2.7). Les culots sont resuspendus dans 500 µL de tampon de lyse et transférés dans des microtubes à fond rond contenant 0,4 gr de billes de verres. Les cellules sont ensuite lysées 30 secondes par le vortex en alternance avec une incubation de 30 secondes sur glace pour un total de 4 minutes. Ensuite, les cellules sont centrifugées à 14 000 rpm pour 30 minutes à 4 °C afin de précipiter les débris cellulaires. Le surnageant est ensuite récupéré (fraction *c*) et ultracentrifugé à une vitesse de 100 000 g pour 90 minutes à 4 °C. Après l'ultracentrifugation, le surnageant est récupéré et concentré à l'aide de tubes Amicon (fraction *sa*). Lors des premières tentatives d'extraction de protéines, le surnageant après ultracentrifugation était concentré à l'aide

du système de concentration sous vide SpeedVac (fraction *s*). Le culot contient la fraction microsomale, mais également la membrane plasmique et des polyribosomes. Le culot est resuspendu dans un tampon de resuspension et conservé à -80 °C. Au cours de cette programmation de recherche, deux tampons de resuspension différents ont été testés, soit le tampon *p* ou *p1* composé de Tris base, EDTA, sorbitol, PMSF et glycérol et le tampon *p2* constitué de Tris base, DTT et PMSF.

Les différentes fractions sont quantifiées par méthode colorimétrique dans une plaque de 96 puits à l'aide de l'ensemble DC Protein Assay de Bio-Rad selon les instructions du fabricant.

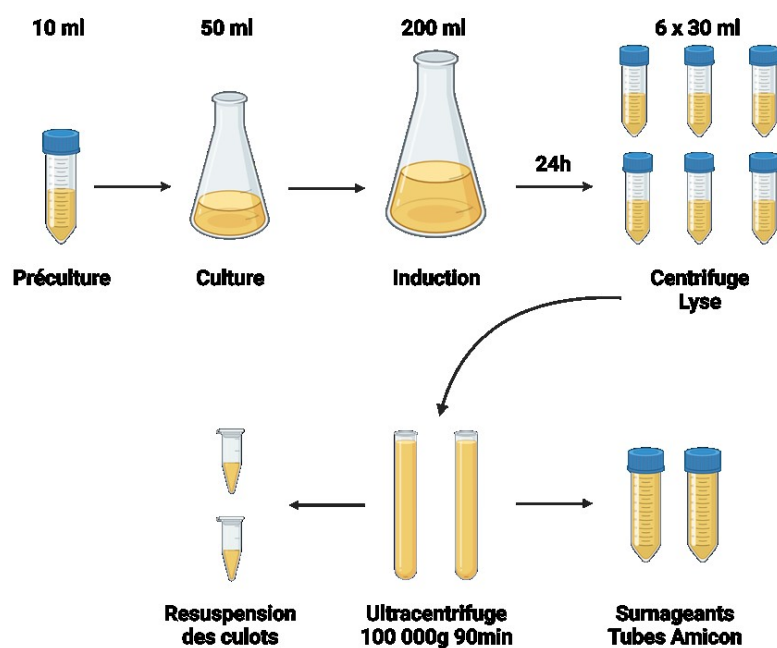


Figure 2.7 Représentation du processus de culture, d'induction et d'enrichissement des protéines.

2.12 Détection par immunobuvardage de type Western blot

Dans un premier temps, un immunobuvardage de type Western blot a été réalisé afin de tenter d'optimiser la détection des protéines d'intérêt en réduisant la quantité de bandes

non spécifiques. Cette expérimentation a été couplée avec un test de vérification afin de s'assurer que les protéines d'intérêt ne subissent pas de glycosylation post-traductionnelle. Donc, la fraction cytosolique (fraction *c*) a été traitée avec l'Endoglycosidase H (Endo H), une enzyme clivant le lien entre les oligosaccharides et le résidu asparagine d'une protéine ayant subi une N-glycosylation. Pour se faire, 10 µg de protéines de la fraction *c* ont été dénaturés à 100 °C pour 10 minutes dans un tampon de dénaturation de glycoprotéines de la compagnie NEB. Puis, l'Endo H a été ajoutée avec un tampon réactionnel et le mélange a été incubé 1 h à 37 °C. Le résultat a été déposé en duplicata sur un gel de polyacrylamide en plus de 4 puits contenant la fraction *c* non traitée à l'Endo H. Le gel de polyacrylamide de 12 % a été migré puis transféré sur une membrane PVDF selon le protocole de transfert standard du Trans-blot Turbo Transfer System. La membrane a ensuite été séparée en quatre sections afin de tester différentes concentrations d'anticorps. Les trois premières ont été bloquées pendant 1 heure avec une dilution 1:1000, 1:2000 ou 1:5000 de l'anticorps primaire 1D4. Après 6 lavages de 10 minutes dans du TBST, l'ensemble des sections ont été traités pendant 1 heure avec l'anticorps secondaire GAM-HRP dilué 1:20000 dans du lait 5 %. Puis, les bandes des quatre sections ont été révélées avec les réactifs de détection Clarity Western ECL de Bio-Rad. Le but de cette première expérimentation étant d'optimiser la détection afin d'obtenir la meilleure révélation possible, les conditions suivantes sont celles qui ont été appliquées par la suite pour les analyses par Western blot.

Les fractions protéiques sont d'abord dénaturées par un chauffage à 95 °C pour 10 minutes en présence du tampon de dénaturation Laemmli buffer 4X. Chaque fraction est déposée dans le gel SDS-PAGE dont la concentration du gel de séparation est de 12 %. Le gel migre à 100 V pour 20 minutes suivis de 150 V jusqu'à la disparition du front de migration. Ensuite, les protéines sont transférées sur une membrane PVDF préalablement activée dans le méthanol. Le transfert est effectué par transfert standard de 30 minutes à 1 ampère constant à l'aide du Trans-Blot® Turbo™ de Bio-Rad. Après avoir bloqué la membrane pendant 1 heure à température pièce avec du TBST 1X + 7 % lait, elle est ensuite incubée pour la nuit en présence de l'anticorps primaire monoclonal 1D4 de Novus Biologicals (Cat. # NBP1-47602) dilué à 1:2000 dans du TBST 1X + 3 % BSA. Le lendemain, la membrane est lavée à 6 reprises pour 10 minutes avec du TBST 1X avant

d'être incubée pour 1 heure à température pièce avec l'anticorps secondaire Immun-Star GAM-HRP de Bio-Rad (Cat. # 170-5047) dilué à 1 : 20000 dans du TBST 1X + 5 % lait. La révélation des bandes sur la membrane par chimioluminescence a été effectuée avec le substrat Clarity™ Western ECL de Bio-Rad et l'image a pu être visualisée grâce au système d'imagerie ChemiDoc™ Touch Imaging System de Bio-Rad.

Toutefois, même avec l'optimisation de la détection des protéines d'intérêt par Western blot, il y a toujours présence de bandes non spécifiques. Afin de valider si les bandes révélées de tailles diverses comprenaient les protéines recombinantes, diverses bandes ont été extraites et envoyées pour identification à la Plateforme de Protéomique du Centre de Recherche du CHU de Québec. Les protéines contenues dans les bandes extraites du gel de polyacrylamide et envoyées à la Plateforme de Protéomique ont été digérées par la trypsine avant d'être injectées dans un HPLC couplé à un spectromètre de masse. Puis, les données sont interprétées par des logiciels spécialisés. Cette analyse à haute résolution et haute précision permet de fournir la liste des peptides présents dans les différentes bandes extraites.

2.13 Essais enzymatiques

Les essais enzymatiques ont été réalisés sur les différentes fractions protéiques après la détection des protéines d'intérêt par Western blot. Le choix des substrats s'est arrêté sur l'acide olivetolique (OA) et le GPP en présence du cofacteur $MgCl_2$ basé sur les substrats utilisés par la CsPT1. Par contre, les tampons de réaction, les concentrations et les temps d'incubation ont été modifiés pour tester un total de 4 conditions différentes. Les différentes conditions ont été choisies en fonction de ce qui a été fait récemment dans la littérature concernant l'activité enzymatique de prényltransférases (Tableau 2.2).

Tableau 2.2

Résumé des différentes conditions testées pour les essais enzymatiques

Conditions	1	2	3	4
[OA] (mM)	1	2	2	0,2
[GPP] (mM)	1	4	2	1
[MgCl ₂] (mM)	10	25	25	5
Tampon	75 mM Tris-HCl pH 7,5	100 mM HEPES pH 7,5	100 mM HEPES pH 7,5	100 mM Tris-HCl pH 7,5
Quantité protéines (µg)	13-400	20-200	20	63-65
Incubation	1 h at 30 °C	16 h at 30 °C	2-8-16-24 h at 30 °C	1 h at 37 °C
Arrêt de la réaction	TCA 2 %	TCA 2 %	TCA 2 %	TCA 2 %
Volume final (µL)	100	50	50	100

Lors des essais enzymatiques, une solution mère de réaction a été créée avec les substrats de réaction, le cofacteur et le tampon. Une quantité de 10 ppm de CBGA a été ajoutée dans les échantillons servant de contrôles positifs. Les protéines ont été ajoutées à la toute fin, au moment de commencer l'incubation. L'incubation a été réalisée à température constante dans un thermocycleur réglé en mode incubation. Les réactions enzymatiques ont été arrêtées par l'ajout de TCA à une concentration finale de 2 %. Les échantillons ont ensuite été filtrés sur colonne avant d'être injectés au HPLC selon la méthode de détection des cannabinoïdes du laboratoire.

CHAPITRE III

RÉSULTATS

3.1 Obtention et assemblage des données de séquençage Illumina de *R. marginata*

Puisqu'aucun génome de référence n'est disponible, le transcriptome de *R. marginata* a dû être assemblé *de novo* depuis les données brutes du séquençage Illumina de l'ARN disponible sur la plateforme NCBI (BioProject N° PRJNA430820, BioSample N° 08378951 et SRA N° SRR6489347) [39]. Les données ont d'abord été filtrées par Trimmomatic avant d'être assemblées par le logiciel Trinity. Sur 30 981 418 séquences brutes, il a été possible d'assembler 1 641 458 transcrits. La taille médiane des transcrits est de 397 paires de bases alors que la taille moyenne est de 1515 paires de bases. Ces statistiques d'assemblage indiquent que l'assemblage est suffisamment de bonne qualité pour pouvoir y rechercher des protéines candidates.

Tableau 3.1

Résumé de l'assemblage des données brutes du séquençage de *R. marginata* par les logiciels Trimmomatic et Trinity

Données de l'assemblage

Nombre de séquences brutes	30 981 418
Nombre de bases assemblées	485 136 993
Nombre de transcrits assemblé	1 641 458
Taille médiane des transcrits (pb)	397
Taille moyenne des transcrits (pb)	1515

3.2 Analyses *in silico*

3.1.1 Détermination des domaines conservés chez les prényltransférases

Avant de rechercher des protéines intéressantes dans le transcriptome assemblé de *R. marginata*, la séquence en acide aminé de la CsPT1 a été utilisée pour effectuer une recherche d'homologie par la méthode BLASTP. Les séquences d'acides aminés de dix prényltransférases ont été sélectionnées, en plus de la CsPT1, pour effectuer l'alignement multiple par ClustalW (Annexe B). Cette analyse a comme objectif de révéler s'il existe un domaine conservé entre les prényltransférases partageant une certaine homologie. Ce domaine pourra alors être utilisé lors de la recherche de protéines d'intérêts dans le transcriptome assemblé.

L'alignement multiple a révélé un domaine conservé entre les positions 108 et 383 de la séquence d'acides aminés de la CsPT1 (Figure 3.1). L'analyse de cette dernière séquence par le logiciel Interpro a identifié le domaine conservé comme étant un domaine UbiA. Un domaine UbiA est un domaine composé de 8 ou 9 hélices transmembranaires qui permet à la protéine de réaliser une prénylation en présence du cofacteur magnésium [71].

```

CoPT      87  E M N K I S T F K D A C W R F L R P H T I R G T A L G S C A L V A R A L T E N S D L I K W S L I K A F S G L A L I
CsPT1    91  S V A T K L N E G R A C W L Q R E W T I E F T S C A C E L F E S E L I H N N L I S W S I M E K A F F L A V I
CsPT4    94  S V A T K L N E G H T Q W L Q R E Y V K G M I S I A C L F E R E L F N N R H I E S G I M M K A F F L I F E L
CsPT5    80  E V N K V S D E F D A C W R F L R P H T I R G T A L G S P A L V A R A L T E N P N L I K W S L I K A F S G L A L I
HbPTlike 87  E V N K V S D E F D A C W R F L R P H T I R G T A L G S P A L V A R A L T E N P N L I K W S L I K A L S G L F A L I
H1PT1    107 S A L T N K D F V S V C M E Y R E Y T A G V I C S S C L F E E L E N P I F S R E L I F A L L G L A L I
H1PT2    101 T V T R G S N E G H S M F R P T P F V V S I C T S L F S A E L R N P N L F S W L M E D A F G L V I L
MmPTlike 85  E V N K V S D E F D A C W R F L R P H T I R G T A L G S P A L V A R A L T E N P N L I K W S L I K A L S G L F A L I
MvPTlike 85  E V N K V S D E F D A C W R F L R P H T I R G T A L G S P A L V S R A L E N S N L I K W S L I K A F A G L F L I
PaPT     132  E V N K V S D E F D A C W R F L R P H T I R G T A L G S T A L V A R A L T E N P N L I K W S L I K A F S G L F A L I
PePTlike 85  E V N K V S D E F D A C W R F L R P H T I R G T A L G S C A L V A R A L T E N T H L I K W S L I K A F G L F A L I

CoPT     147  C G N G Y I V G I N Q I Y D I G I D K V N K P Y L P I A A G D L S V C S A W L V I F F A V A G L I - I V G F N F G P F
CsPT1    151  G I A S F T T T I N Q I Y D H I D R I N K F D L P A S G L S V N A W M S T I V A E G L I T T K K G G F L
CsPT4    154  S F N F A A I N Q I Y D I D R I N K E D L P V S G S S E A W L S T I V A T G L V T K K S L E L
CsPT5    140  C G N G Y I V G I N Q I Y D I G I D K V N K P Y L P I A A G D L S V C S A W L V I L F A V A G L I - T V G F N F G P F
HbPTlike 147  C G N G Y I V G I N Q I Y D I G I D K V N K P Y L P I A A G D L S V C S A W L V I F F A V A G L S - I V G L N F G P F
H1PT1    167  G S C F Y T A G I N Q I D D I D R I N K F D L P V S G R S V E S A W L T L S P A I C F L I T K L N S G F L
H1PT2    161  L Y H I Y T N G I N Q I Y D E S F R I N K E D L P A E E L S V K S A W L I T F S A V A S L L M I K K G C L F
MmPTlike 145  C G N G Y I V G I N Q I Y D I G I D K V N K P Y L P I A A G D L S V C S A W L V I F F A V A G L S - I V L N F G P F
MvPTlike 145  C G N G Y I V G I N Q I Y D I G I D K V N K P Y L P I A A G D L S V K S A W L V I F F A V A G V I - I V G L N F G P F
PaPT     192  C G N G Y I V G I N Q I Y D I G I D K V N K P Y L P I A A G D L S V C S A W L V I F F A V A G L I - I V G L N F G P F
PePTlike 145  C G N G Y I V G I N Q I Y D I G I D K V N K P Y L P I A A G D L S V C S A W L V I F F A V A G L I - I V G L N F G P F

CoPT     206  I T S L Y S L G L F L G T I Y S V P P F R M K R F P V A A F L I I A T V R G F L L N F G V Y Y A T R A A L G L F P F O W S
CsPT1    211  Y I F G Y C F G L F G C I Y S V P P F R M K N E S T A P L I N F L A - H I T N F T F Y Y A R A A L G L F P F E L R
CsPT4    214  F V E Y L F C F A G F A Y S V P P F R M K N E F T N E L L I T I S - H V G L A T S V S A T S A L G L F P F M R
CsPT5    199  I T S L Y C L G L V L G T I Y S V P P F R M K R F P V A A F L I I A T V R G F L L N F G V Y Y A T R A A L G L F P F O W S
HbPTlike 206  I T S L Y C L G L F L G T I Y S V P P F R M K R F P V A A F L I I A T V R G F L L N F G V Y Y A T R A A L G L F P F O W S
H1PT1    227  I T S L Y C L A L S G T I Y S V P P F R M K N E T A F I C L M H A - G L N F S V Y Y A R A A L G L F A F A W S
H1PT2    221  I T C Y C Y L V G A Y S V P P F R M M N T F T S I W N F E I G L N F L N Y A S R A I L G L F P F O M R
MmPTlike 204  I T S L Y C L G L F L G T I Y S V P P F R M K R F P V A A F L I I A T V R G F L L N F G V Y Y A T R A A L G L F P F O W S
MvPTlike 204  I T S L Y C L G L F L G T I Y S V P P F R M K R F P V A A F L I I A T V R G F L L N F G V Y Y A T R A A L G L F A F A W S
PaPT     251  I T S L Y S L G L F L G T I Y S V P P F R M K R F P V A A F L I I A T V R G F L L N F G V Y Y A T R A A L G L F P F O W S
PePTlike 204  I T S L Y S L G L F L G T I Y S V P P F R M K R F P V A A F L I I A T V R G F L L N F G V Y Y A T R A A L G L F A F O W S

CoPT     266  S P V F I T T F V T F A L V I A I T K D L P D V E G D R K F Q I S T L A T K L G V R N I A L G S G L L L N Y V A
CsPT1    270  P S F T F L A F M K S M S A L A T K D A S D V E G L R F E H S T L A K V G S R N T L F G S G L L L S Y V A
CsPT4    273  P A F S F T I A F T M G M T A F A K D I S D E G D A R K G S T A T K L G A R N T E V S G L L L L N Y I V
CsPT5    259  S A V A F I T T F V T F A L V I A I T K D L P D V E G D R K F Q I S T A T K L G V R N I A L G S G L L L N Y V G
HbPTlike 266  S P V F I T T F V T F A L V I A I T K D L P D V E G D R K F Q I S T A T K L G V R N I A L G S G L L L N Y V A
H1PT1    286  P S F S F I T A F T F M T I T A S K D L S D N G D R K F G E T F A T K L G A N I T L L G L L L L N Y V A
H1PT2    281  E F E T F I G E V T L S I I S L K D P D V E G D K V G S T L P V I F G A T I V L G S G F L L N Y V A
MmPTlike 264  S P V F I T T F V T F A L V I A I T K D L P D V E G D R K F Q I S T L A T L G V R N I A L G S G L L N Y V A
MvPTlike 264  S P V F I T T F V T F A L V I A I T K D L P D V E G D R K F Q I S T L A T K L G V R N I A L G S G L L L N Y V A
PaPT     311  S P V A F I T T F V T F A L V I A I T K D L P D V E G D R K F E I S T A T K L G V R N I A L G S G L L L N Y V A
PePTlike 264  S P V F I T T F V T F A L V I A I T K D L P D V E G D R K F Q I S T L A T K L G V R N I A L G S G L L L N Y V A

CoPT     326  A L A A I Y M P Q A F R N L M I F A H V E L A L G L F O T W L V L E Q A N Y T K E A I S G F Y R F I W N L F Y A E E
CsPT1    330  A I L A I I W P Q A E N S N M I L S H A L A F W L I L Q T R D F A L T N Y D F E A G R R E V E F M K L Y A E Y
CsPT4    333  S I S I I I W P Q V E R S N M I L S H A L A F C L I F O T R E L A L A N Y A S A P S R Q F E F I W L I Y A E Y
CsPT5    319  A I A A I Y M P Q A F R N L M I F E H T I L A L S L F C A W V L E Q A N Y T K E A I A G F Y R F I W N L F Y V E Y
HbPTlike 326  S I L A A I Y M P Q A F R S L M I F A H T I L A L S L I F O T R V L E Q A N Y T K E A I S G F Y R F I W N L F Y A E Y
H1PT1    346  A I S T A I I W P K A F R S N M I L S H A L A F S L I P Q A R E L R T N Y T E A C K S F Y E F I W L F S A E Y
H1PT2    341  A I G V A I M P Q A F R G Y I M I P A H A F E S A L I R K T W L K A N Y A K E A S D S Y H F I W L M I A E Y
MmPTlike 324  S I L A A I Y M P Q A F R N L M I F V E S I L A L S L I F O T R V L E Q A N Y T K E A I S G F Y R F I W N L F Y A E Y
MvPTlike 324  A L A A I Y M P Q A F R S L M I F A H T I L A L S L I F O T R V L E Q A N Y T K E A I S G F Y R F I W N L F Y E Y
PaPT     371  A I A A I Y M P Q A F R N L M I F T H A V L A L C L I F O T R V L E Q A N Y K E A T A G F Y R F I W N L F Y E Y
PePTlike 324  S I L A A I Y M P Q A F R N L M I F V E T M A L S L I F O T R A L E Q A N Y T K A I S G F Y R F I W N L F Y A E Y

```

Figure 3.1 Alignement multiple partiel de la séquence en acides aminés de 11 prényltransférases.

Les séquences ont été alignées avec ClustalW. Une boîte noire indique que l'acide aminé est conservé dans au moins la moitié des protéines alignées. Une boîte grise indique une similitude avec d'autres acides aminés à la même position. La liste des espèces est répertoriée en annexe B avec le numéro d'accèsion GenBank correspondant.

3.2.1 Identification des séquences pertinentes chez *R. marginata*

Le domaine conservé de la CsPT1, soit les acides aminés 108 à 383, a été utilisé pour réaliser une recherche par homologie dans le transcriptome de *R. marginata* par la méthode TBASTN. Cette analyse a permis d'identifier deux protéines, ayant une certaine homologie avec le domaine conservé, et un cadre de lecture complet. Le meilleur résultat a été obtenu avec la protéine dénommée RmPT280 qui a montré 40 % d'identité et 58 % de similarités avec le domaine conservé. L'autre protéine identifiée, nommée RmPT130, a montré 30 % d'identité et 45 % de similarités avec le domaine conservé. Pour une comparaison visuelle, un alignement multiple a été réalisé avec la séquence d'acides aminés de la CsPT1 et les deux protéines identifiées dans le transcriptome (Figure 3.2). Cet alignement a permis de déterminer que les acides aminés qui lient le magnésium sont conservés. Les cinq acides aminés qui entourent le GPP dans la poche catalytique sont également majoritairement conservés, à l'exception de I159 qui est changée pour L159 chez la protéine RmPT130.

```

CsPT1      1  MGLS SVCTFS FQNTY-----HTLLNPHNN-NPKTSLLCYRH-----PKTPI
RmPT280    1  MATA LTRIASGCCSHGGASVVAIPTGLHGHAFLFSR-TPSIAPLCRHSDDRRLRHLVLSG
RmPT130    1  MGS LQRSVPTQLCSK-----RLLLIIPSCGAPPFRARLLTY-----AAVPR

CsPT1      41  KYS LNNEPSK-HCS-TKSFHLQNKCS-----SLSTAKNS-----IRAAT
RmPT280    60  AEAFESVDVHGCGQWYRRSSTTGLRSRRSGSRTPRFVGS-----GGEQS
RmPT130    40  PVSIEGYPHIE-GCG-IRRGSLDGLCPGPSGLKLSRDGEISGQSVVNKRSLPIRRGGVAT

CsPT1      79  TNQTEPPES-----DNLSVATKIINFGKACWKLQRFYTTIAFTSCACGLFGKE
RmPT280    105  SDDAEPPSE-----FTRSSELSVSERFFSFKLAFWRFLRPHTIRGILLGSVALVCRA
RmPT130    98  TFAFAEPIDDDDDDAASVPEGA-KTLVVRDGVDWLDALHRSRPHTVIGSAMSTVSVS-LL

CsPT1      127  LLHNTNLSWSLMFKAFEFLLVAVLCTASFTTTINQIYDLHIDRINKPDLPLASGEISVNT
RmPT280    157  LLENPRVDAVLLPRALWGIIALCGNGYIVGINQIYDVEIDRINKPFLPIAAGDLSIPQ
RmPT130    156  ALQSFAELSSAFFGGLVQALTPSLFMNIYIVGINQIYDVIDDKVNKPLPLASGEFSMAT

CsPT1      187  AWIMSTIVALFGIITTKMKGGPIYIEGYCFGIEGLVYSV--PPFRWKQNPSTAFLL-N
RmPT280    217  AWSIVLGLAALGTTIV-GTNFGPLITVLYFGLFLGFIYSV--PPLHLKRFVAVPAFLIITA
RmPT130    216  GVAIVALSAAISVAIGIMVGSRPLL-WTLLVSLFLGFAYSVDVPEFLRWKRSAVAATCIL

CsPT1      244  FLAHLITNFTFYYSR-AALGLPFELRPSFTELLAFMKSMGSALALIKDASDVEGDTKFG
RmPT280    274  TVRFLLNFGVYYAIR-AALGLPFEWNPPIVETISFVTVFATVIAITKDLDPTEGDRKFK
RmPT130    275  ATRAVVQLGFYCHMQGFVFGRAATITROLLIATFMSFFSVVIALAKDLDPVEGDKVYG

CsPT1      303  ISTLASKYGSRNLTLCSGIVLLSYVAAILAGIIWPOAFNSNVMILSHAILAFWLLQTR
RmPT280    333  ISTFAIKLGVRNISFLGSGLLILANYAAATVAATYLPFCFRQVMIPAHAILATAVYQTW
RmPT130    335  IIRSESVRLGQEKVFWSCVWLLIQAAAYAVAMSVGITSPTLFGKVMILGHASLSSLLWMAK

CsPT1      363  DFA LTNYPAGRRFYEFMWKLYAEYLYVVF-
RmPT280    393  LLDSAKYTKDSTIGFYQFIWKLFYSEYTLIPFI-
RmPT130    395  ATNL--NIKTEIVSFYMFVWKLFYAEYFLIPFVR

```

Figure 3.2 Alignement multiple de la CsPT1 et des deux protéines de *R. marginata*.

La séquence en acides aminés de la CsPT1 a été utilisée pour générer un nouvel alignement multiple avec les deux protéines trouvées. CsPT1 (*C. sativa*, numéro d'accèsion GenBank DAC76711); RmPT280 (protéine de *R. marginata* du contig R4869280) et RmPT130 (protéine de *R. marginata* du contig R4826130) ont été alignés pour visualiser la conservation des acides aminés essentiels pour la réaction catalytique. Les quatre acides aminés qui lient le ligand Mg^{2+} sont représentés par les rectangles bleus et les cinq acides aminés qui entourent le GPP dans la poche catalytique sont représentés par des rectangles rouges. La méthode d'alignement et la signification des boîtes sont identiques à la figure 3.1. Les traits d'union indiquent l'insertion d'un espace pour un meilleur alignement.

3.2.2 Prédiction des structures des protéines d'intérêts

Une analyse *in silico* plus approfondie des deux protéines et de CsPT1, à l'aide de PredicProtein, a prédit une location cellulaire plastidique et une structure secondaire principalement hydrophobe composée de nombreuses hélices. SWISS-MODEL a été utilisé pour prédire la structure tertiaire des deux protéines intéressantes et de la CsPT1 (Figure 3.3). L'analyse a révélé, pour chaque protéine, un ligand de magnésium essentiel pour la catalyse. Les protéines sont donc des prényltransférases composées d'hélices transmembranaires et Mg^{2+} -dépendante, ce qui vient valider les prédictions précédentes du logiciel Interpro sur le domaine conservé de la famille UbiA. Le ligand Mg^{2+} serait lié à quatre acides aminés dans la séquence CsPT1, soit R108, N160, D164 et D168.

SWISS-MODEL a utilisé la structure cristallisée de la prényltransférase d'*Archaeoglobus fulgidus* (numéro d'accès PDB 4tq3.1) et d'un fragment de la structure du domaine B-box de la protéine humaine Midline-2 (numéro d'accès PDB 2dja.1) pour générer la prédiction de structure de CsPT1. Le modèle de prédiction de RmPT280 était également basé sur la structure cristallisée de la prényltransférase d'*A. fulgidus* (numéro d'accès PDB 4tq3.1) tandis que le modèle de prédiction de RmPT130 est basé sur la conformation de la Protéine b de résistance à l'acriflavine d'*E. coli* (numéro d'accession PDB 2hqf.1) et l'octaprényltransférase d'*Aeropyrum pernix* (accession PDB numéro 4od5.1).

Comme la prényltransférase d'*A. fulgidus* (numéro d'accès PDB 4tq3.1) lie le GPP tout comme la CsPT1, sa structure cristallisée a été étudiée avec le logiciel PyMOL pour déterminer quels acides aminés entourent le GPP dans la poche catalytique. Ces acides aminés seraient I159, N160, D164, D168 et G222.

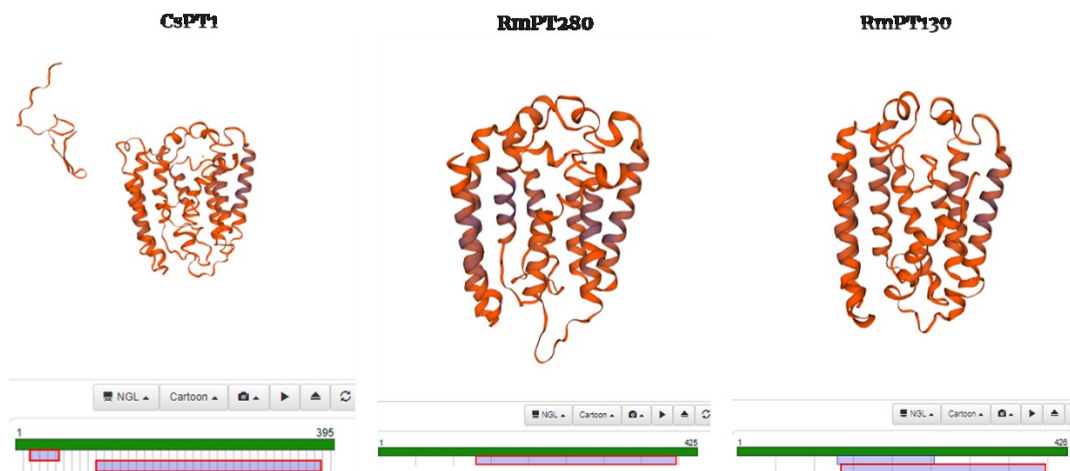


Figure 3.3 Prédiction des structures de la CsPT1 et des deux protéines de *R. marginata*.

Les modèles de structure de la CsPT1 (*C. sativa*, numéro d'accession GenBank DAC76711); RmPT280 (protéine de *R. marginata* du contig R4869280) et RmPT130 (protéine de *R. marginata* du contig R4826130) ont été réalisés avec SWISS-MODEL.

Pour représenter les relations évolutives, des protéines identifiées dans le transcriptome de *R. marginata*, un arbre phylogénétique a été réalisé avec les deux protéines d'intérêt et les protéines utilisées pour générer l'arbre phylogénétique de l'article sur l'étude de la biosynthèse des cannflavines chez *C. sativa* (Figure 3.4 et Annexe C) [14]. Une octaprényltransférase de *R. marginata* a été utilisée comme exogroupe. Après la génération de l'arbre phylogénétique, la CsPT1 se retrouve dans le groupe des terpénophénols tandis que la protéine RmPT280 se retrouve dans le groupe de protéines normalement impliquées dans la biosynthèse des plastoquinones. Ces protéines font partie de la chaîne de transport d'électrons pendant la photosynthèse. RmPT130 se retrouve, quant à elle, dans le groupe des protéines impliquées dans la biosynthèse des tocophérols, comme la vitamine E. Les séquences d'acides nucléiques RmPT280 et RmPT130 partagent 53 % d'identité.

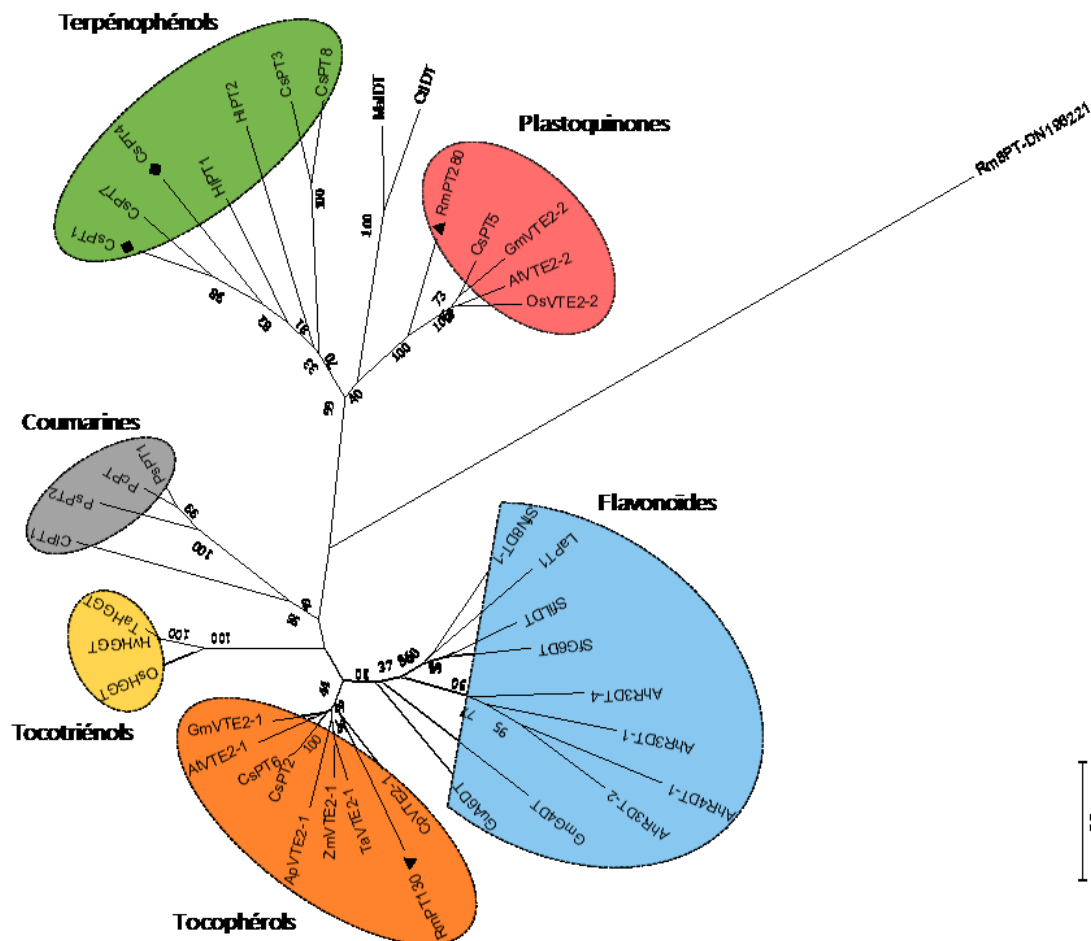


Figure 3.4 Analyse phylogénétique des diverses prényltransférases aromatiques. Les séquences d'acides aminés des différentes prényltransférases ont été alignées avec ClustalW, puis l'alignement a été optimisé avec GBlocks en supprimant les positions mal alignées et les régions divergentes. L'arbre phylogénétique a été réalisé par MEGA7 en utilisant la méthode Neighbor-Joining (1000 bootstrap) pour illustrer leurs relations évolutives. La liste des espèces est répertoriée en annexe C avec le numéro d'accèsion GenBank correspondant.

3.3 Assemblage du vecteur d'expression par Gibson Assembly

Afin de pouvoir créer les différents vecteurs conçus *in silico*, la première étape est de linéariser le vecteur pYES2/CT en retirant le site de clonage multiple par digestion enzymatique avec les enzymes de restriction Hind III et Pme I. Cependant, selon les informations fournies par le fabricant, les deux enzymes de restriction choisies ne sont pas

optimales dans le même tampon de réaction. Ainsi, des tests de digestion ont été effectués afin de déterminer les meilleures conditions pour linéariser le vecteur (Figure 3.5).

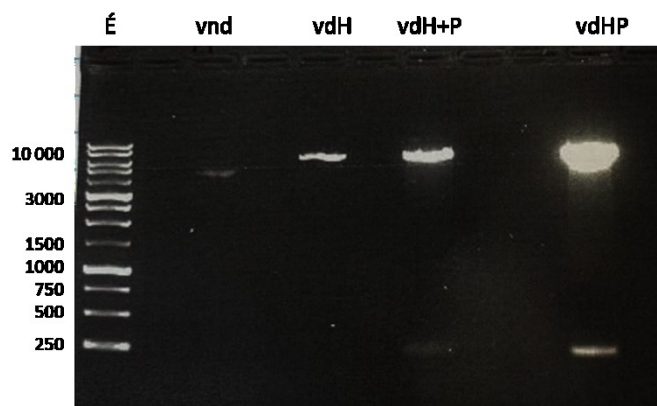


Figure 3.5 Test de digestion enzymatique du vecteur pYES2/CT.

vnd : 38 ng de vecteur pYES2/CT non digéré; vdH : 200 ng de vecteur digéré avec l'enzyme de restriction Hind III pendant 3H à 37 °C dans le tampon NEB 2.1; vdH+P : 1000 ng de vecteur digéré préalablement par Hind III suivi d'une digestion par Pme I toute la nuit à 37 °C dans un mélange de tampons NEB 2.1 et Cutsmart; vdHP : 2500 ng de vecteur pYES2/CT digéré simultanément par Hind III et Pme I pendant toute la nuit à 37 °C dans le tampon NEB 2.1. Les différentes réactions ont été migrées sur un gel d'agarose 1 %.

Le premier puits vnd, qui contient le vecteur non digéré, présente une bande de taille plus petite que la bande correspondant au vecteur linéarisé, ce qui est tout à fait normal puisque le vecteur digéré, bien qu'il soit plus petit, a plus de difficulté à migrer au travers des polymères du gel d'agarose, ce qui donne l'impression que sa taille est plus élevée qu'en réalité. Le deuxième puits vdH contient seulement une seule bande puisque le vecteur a seulement été linéarisé par l'enzyme Hind III donc le site de clonage multiple n'a pas été retiré. Dans le troisième puits vdH+P, il est possible d'apercevoir deux bandes, soit une correspondante au vecteur linéarisé et la seconde, sous 250 paires de bases, correspondant au site de clonage multiple. Dans le dernier puits, la quantité initiale de vecteurs dans le mélange réactionnel était initialement plus élevée, ce qui transparait par l'intensité des bandes obtenues. C'est donc la bande supérieure du dernier puits qui a été extraite du gel d'agarose et purifiée pour isoler le vecteur digéré nécessaire pour l'assemblage par la méthode Gibson.

Un exemple du vecteur obtenu après l'assemblage d'une cassette d'expression dans le vecteur commercial pYES2/CT à l'aide de la méthode Gibson est schématisé ci-dessous (Figure 3.6). Comme les trois cassettes d'expression utilisées dans cette programmation de recherche sont insérées à la même position du même plasmide pYES2/CT, seul le vecteur contenant la cassette RmPT280 est représenté ici. Il a été renommé pYES2/CT + RmPT280 et fait 7117 paires de bases. Les deux autres vecteurs ont été conçus de la même manière pour les deux autres cassettes d'expression, soit le vecteur pYES2/CT + RmPT130 et pYES2/CT + CsPT1, qui font respectivement 7120 pb et 7027 pb. Ces derniers seront utilisés pour la transformation dans la bactérie et ultimement dans la levure.

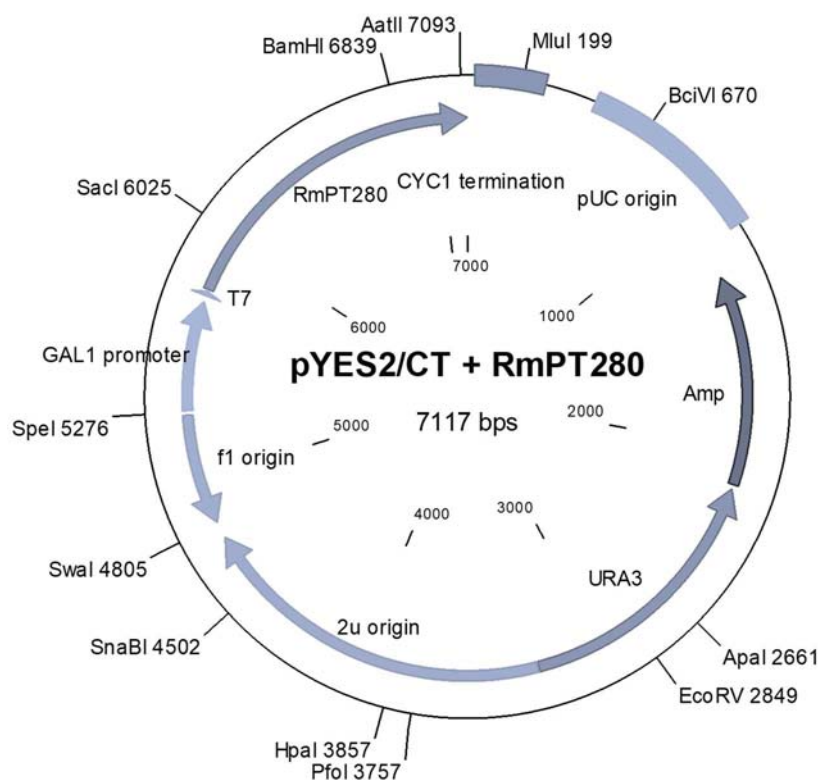


Figure 3.6 Représentation visuelle du vecteur assemblé par la méthode Gibson.
 pUC origin : origine de répllication bactérienne; amp : gène de résistance à l'ampicilline; URA3 : marqueur de sélection auxotrophe pour la levure; 2 μ origin : origine de répllication pour la levure; f1 origin : origine de répllication f1; Gal1 promoter : promoteur inductible de la levure pour l'expression de la cassette; T7 : promoteur bactérien; RmPT280 : cassette d'expression pour le gène d'intérêt RmPT280; CYC1 termination : termineur de la cassette d'expression chez la levure.

3.4 Transformation dans la bactérie *E. coli*

Pour les trois différentes transformations par choc thermique, le mélange de bactéries et de milieu de réaction d'assemblage Gibson ayant subi un changement rapide de température a été incubé sur un milieu de sélection pendant une nuit. Une réaction de PCR par colonie a été effectuée sur certaines colonies isolées afin de vérifier que le vecteur exogène a réussi à pénétrer dans les cellules bactériennes (Figure 3.7).

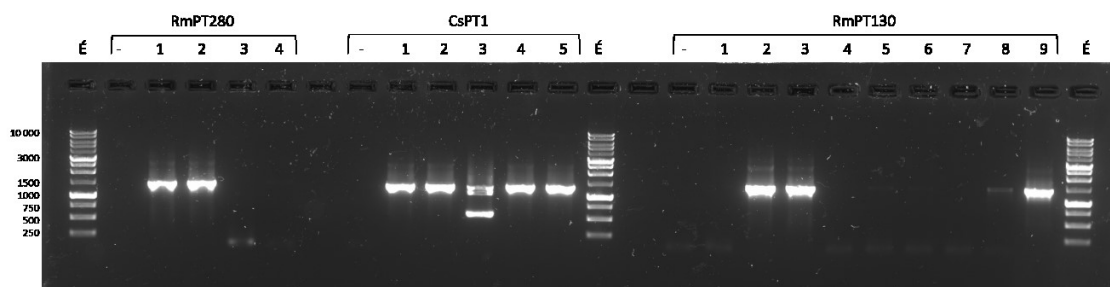


Figure 3.7 Amplification des gènes dans des colonies positives résultant de la transformation dans *E. coli* DH5 α .

Le résultat des réactions PCR a été chargé sur un gel d'agarose à 1 %. Les témoins négatifs ont été réalisés sans ADN. L'échelle moléculaire utilisée est le 1Kb Plus DNA Ladder.

Les tailles attendues sont de 1385 pb pour RmPT280, 1245 pb pour CsPT1 et de 1388 pb pour RmPT130. Les bandes intenses d'amplification indiquent que les clones 1&2 contiennent le vecteur pour RmPT280, les clones 1,2,4&5 contiennent le vecteur CsPT1 et les clones 2,3&9 contiennent le vecteur RmPT130. Le clone 2 de RmPT280, le clone 4 de CsPT1 et le clone 2 de RmPT130 ont été sélectionnés pour procéder à l'extraction et la purification du vecteur d'expression afin d'être ensuite insérés dans la levure pour l'expression hétérologue. L'intégrité des cassettes d'expression a été validée par séquençage Sanger. Les trois cassettes d'expression se sont révélées parfaitement intactes, sans aucune modification.

3.5 Transformation dans la levure

Après la transformation par choc thermique de chaque vecteur purifié dans les cellules compétentes de *S. cerevisiae* INVSc et une incubation sur un milieu de culture

sélectif, certaines colonies isolées ont été sélectionnées pour vérifier l'insertion du vecteur par amplification PCR sur colonie (Figure 3.8 et Figure 3.9).

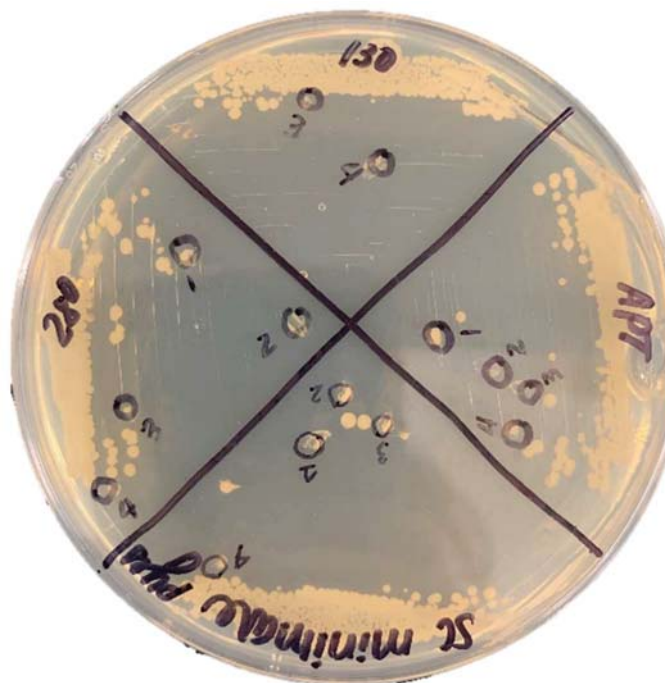


Figure 3.8 Colonies positives de *S. cerevisiae* après transformation.

La transformation a été effectuée avec le protocole de Clontech pour les 4 constructions différentes (pYES2/CT + RmPT280, pYES2/CT + RmPT130, pYES2/CT + CsPT1 et pYES2/CT). Les cellules de levure transformées ont été centrifugées et le surnageant a été retiré. Les cellules ont été étalées sur $\frac{1}{4}$ de la gélose SC minimale sans uracile. La plaque a été incubée à 37 °C jusqu'à l'apparition des colonies.

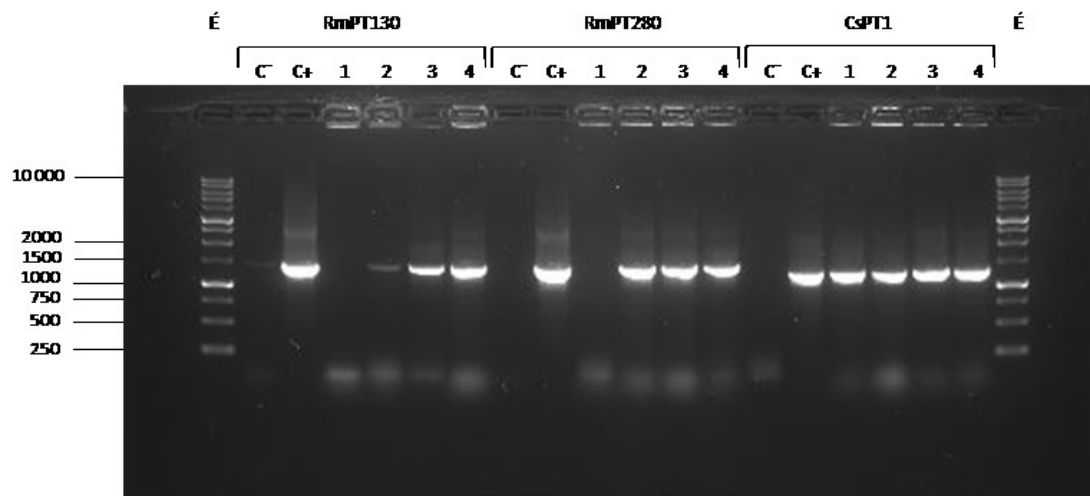


Figure 3.9 Amplification des gènes dans des colonies positives résultant de la transformation dans la souche *S. cerevisiae* INVSc.

Le résultat des réactions PCR a été chargé sur un gel d'agarose à 1 %. Les témoins négatifs ont été réalisés sans ADN. Les témoins positifs ont été réalisés avec les différents plasmides purifiés.

Les tailles attendues sont les mêmes que lors de l'amplification PCR sur les colonies bactériennes, soit de 1388 pb pour RmPT130, 1385 pb pour RmPT280 et de 1245 pb pour CsPT1. Les bandes intenses d'amplification indiquent que les clones 3&4 contiennent le vecteur pour RmPT130, les clones 2,3&4 contiennent le vecteur RmPT280 et les clones 1 à 4 contiennent le vecteur CsPT1. Les clones sélectionnés, après validation par séquençage Sanger, pour poursuivre avec l'expression des protéines recombinantes sont le clone 4 pour RmPT280, le clone 4 pour RmPT130 et le clone 4 pour CsPT1.

3.6 Expression des protéines recombinantes

Cette première analyse, effectuée après la lyse cellulaire, a permis d'étudier le temps d'induction optimal pour la production des protéines d'intérêt dans le système d'expression hétérologue. Parmi les trois fractions protéiques, soit les fractions *c*, *s* et *p*, la fraction protéique *s* était impossible à quantifier dû à sa viscosité après son passage dans le système de concentration sous vide SpeedVac. La fraction *p* était également impossible à quantifier dû à la présence de glycérol dans le tampon de resuspension des microsomes 1. Le glycérol interférant avec l'ensemble des méthodes de quantification disponibles dans

le laboratoire. C'est pourquoi le même volume de protéines a été utilisé, plutôt que la même quantité, pour réaliser la membrane suivante (Figure 3.10).

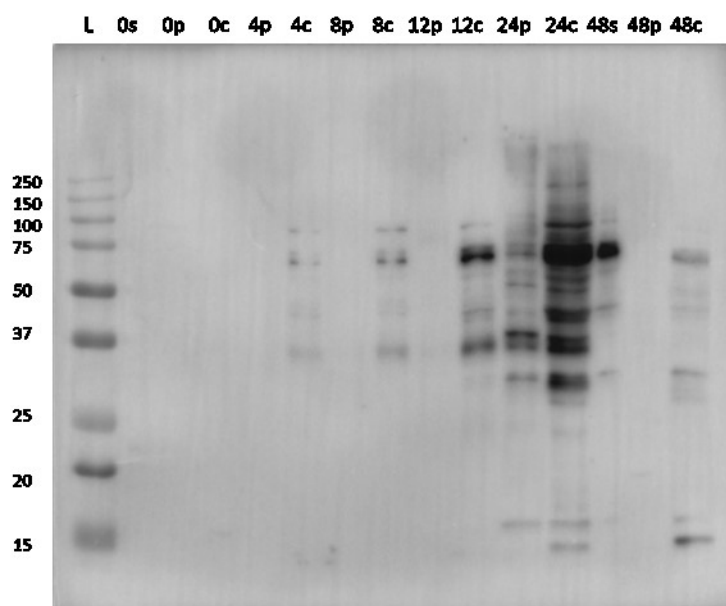


Figure 3.10 Détermination de la durée d'induction optimale pour l'obtention des protéines d'intérêt.

Immunobuvardage de type Western blot avec des protéines extraites de *S. cerevisiae* INVSc du clone 4.1 possédant le vecteur pYES2/CT+RmPT280. Les protéines ont été extraites 0, 4, 8, 12, 24 et 48 heures après induction au galactose 2 %. 12 μ l de protéines ont été chargés pour chaque puits. Le transfert a été effectué avec le système de transfert turbo Trans-blot à 1,3 ampère constant pendant 15 minutes. La membrane a été bloquée 1 heure avec du lait 5 %, puis une nuit avec l'anticorps primaire 1D4 dilué 1:1000. La membrane a été traitée pendant 1 heure avec l'anticorps secondaire GAM-HRP dilué 1:2000, puis révélé avec les réactifs de détection Clarity Western ECL de Bio-Rad. *c* : surnageant après lyse cellulaire; *s* : surnageant concentré par le Système de concentration sous vide SpeedVac après l'ultracentrifugation; *p* : tampon de resuspension des microsomes 1. La taille attendue de la protéine RmPT280 est de 49 kDa.

Le résultat de cet immunobuvardage de type Western blot, bien qu'il n'ait pas été réalisé dans les conditions optimales de détection, indique que 24 heures est le temps d'induction auquel le plus de protéines sont détectées. Ce temps d'induction a donc été sélectionné pour la suite des expérimentations.

3.7 Détection par immunobuvardage de type Western blot

Après avoir remarqué qu'un grand nombre de bandes non spécifiques sont détectées par l'anticorps primaire lors de l'analyse précédente, cette expérimentation visait à déterminer les conditions optimales de détection des protéines d'intérêt contenant l'épitope 1D4. Les protéines extraites après la lyse cellulaire (fraction *c*) ont été utilisées pour réaliser cette analyse. Dans un premier temps, une partie de cette fraction a été traitée par l'Endo H, une enzyme retirant les oligosaccharides d'une protéine glycosylée. La glycosylation est une modification post-traductionnelle courante chez les protéines membranaires et elle peut modifier la taille attendue des protéines. Après la migration et le transfert des protéines sur une membrane PVDF, celle-ci a été découpée en 4 sections afin de tester différentes concentrations de l'anticorps primaire (Figure 3.11).

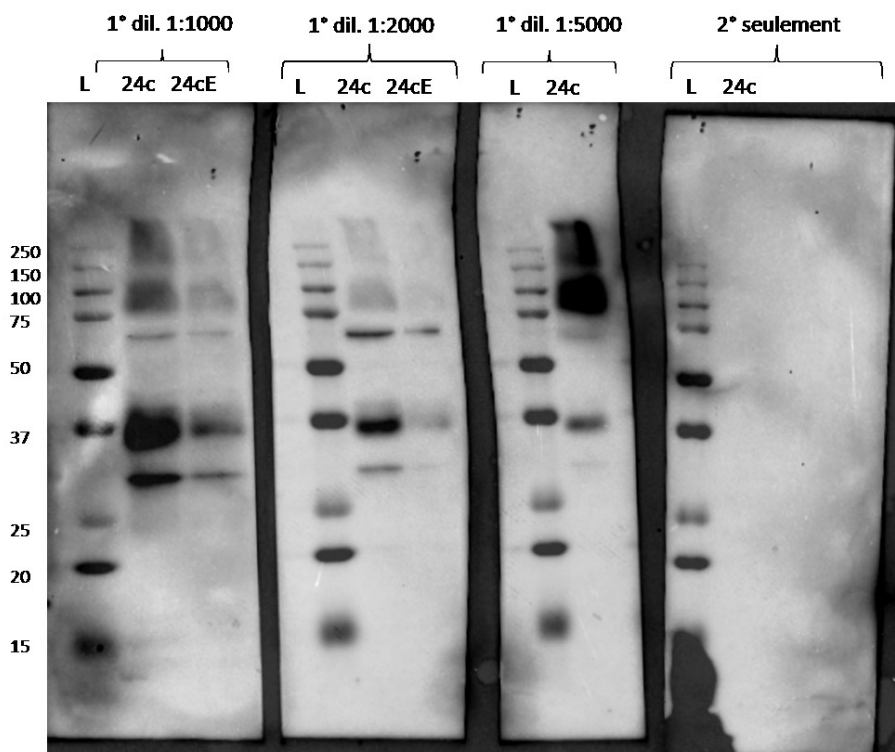


Figure 3.11 Optimisation de la détection des protéines d'intérêt.

Immunobuvardage de type Western blot avec les protéines extraites de *S. cerevisiae* INVSc du clone 4.1 possédant le vecteur pYES2/CT+RmPT280. Les protéines ont été extraites 24 heures après induction au galactose 2 % (24c). Certains échantillons ont été traités avec EndoH pour vérifier s'il y a présence de glycosylation sur les protéines

d'intérêt (24cE). 10 µl de protéines ont été chargés pour chaque puits. Le transfert a été effectué avec le Trans-blot Turbo Transfer System avec le protocole de transfert standard. La membrane a été bloquée pendant une nuit avec du lait à 7 %, puis découpée. Les trois premières bandes coupées ont été bloquées pendant 1 heure avec différentes concentrations de l'anticorps primaire 1D4, soit une dilution 1:1000, 1:2000 et 1:5000. Après 6 lavages de 10 minutes, les quatre fragments de membrane ont été traités pendant 1 heure avec l'anticorps secondaire GAM-HRP dilué 1:20000 dans du lait 5 %, puis révélé avec les réactifs de détection Clarity Western ECL de Bio-Rad. La taille attendue de la protéine RmPT280 est de 49 kDa.

Ici, il n'y a pas de différence de taille entre les protéines révélées dans les puits sans traitement à l'Endo H et les puits traités à l'Endo H. Ce résultat indique que les protéines d'intérêt n'ont probablement pas subi de N-glycosylation post-traductionnelle. Cependant, il existe d'autres modifications post-traductionnelles qui peuvent affecter la taille attendue d'une protéine comme l'acétylation ou encore la méthylation. Dans le cadre de cette programmation de recherche, il n'était pas pertinent de chercher à identifier quels types de modifications post-traductionnelles les protéines exprimées auraient pu possiblement subir, c'est pourquoi nous nous sommes arrêtés après un test.

Pour ce qui est de la meilleure concentration d'anticorps primaire assurant une détection optimale, la dilution 1 : 2000 est celle qui permet l'obtention des bandes les plus nettes. Cette concentration sera donc utilisée pour la suite des expérimentations. La quatrième section, non traitée avec l'anticorps primaire, mais seulement avec l'anticorps secondaire, indique que la détection de bandes non spécifique est uniquement causée par l'anticorps primaire.

Même avec l'optimisation des conditions de détection pour l'immunobuvardage de type Western blot, il est toujours possible de détecter des bandes qui ne correspondent pas à la taille attendue, mais qui ne sont pas présentes dans le contrôle négatif (non représenté sur la Figure 3.12). Ainsi, pour en avoir le cœur net quant à la présence ou non des protéines d'intérêt dans ces bandes, les fractions protéiques ont été migrées et les bandes ont été extraites du gel de polyacrylamide et envoyé pour identification des protéines à la Plateforme de Protéomique du Centre de Recherche du CHU de Québec.

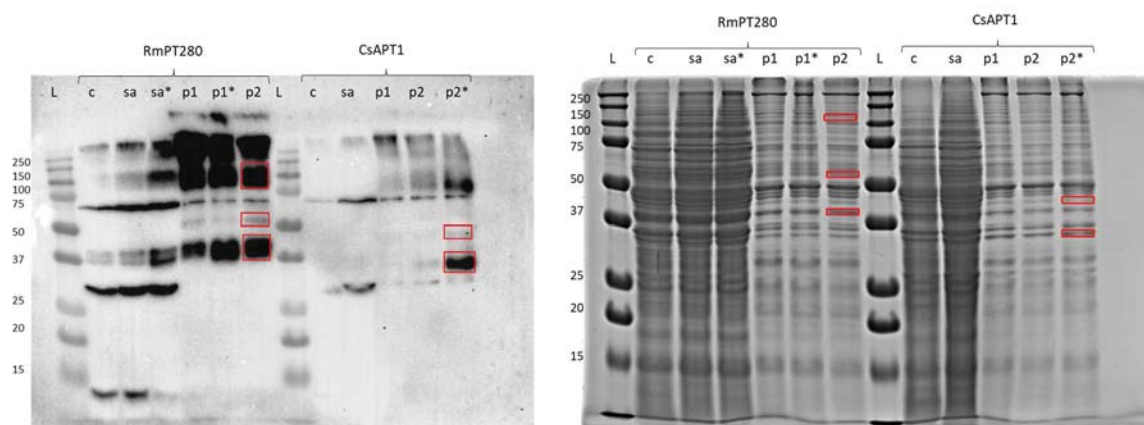


Figure 3.12 Amplification des gènes dans des colonies positives.

Gauche : Immunobuvardage de type Western blot avec des protéines extraites de *S. cerevisiae* INVSc avec le clone 4.1 du vecteur pYES2/CT + RmPT280 et le clone 4 du vecteur pYES2/CT + CsPT1. Les tailles attendues pour les protéines recombinantes sont respectivement de 49 kDa et de 46 kDa. Les protéines ont été extraites 24 heures après induction au galactose 2 %. 12 μ l de protéines ont été chargés pour chaque puits. Le transfert a été effectué avec le Trans-blot Turbo Transfer System en utilisant le protocole de transfert standard. La membrane a été bloquée pendant une nuit avec du lait 7 %, puis traité pendant 1 heure avec l'anticorps primaire 1D4 dilué 1:2000. La membrane a été traitée pendant 1 heure avec l'anticorps secondaire GAM-HRP dilué 1:20000, puis révélé avec les réactifs de détection Clarity Western ECL de Bio-Rad.

Droite : Coloration de 5 minutes d'un gel SDS-PAGE 12 % avec du bleu de Coomassie filtré. Le gel a été décoloré avec la solution de décoloration pendant une heure, puis avec de l'eau MilliQ pendant une nuit. Ce gel est la copie exacte du gel utilisé pour l'immunobuvardage de type Western blot. La même quantité de protéines a été chargée et les gels ont migré simultanément dans la même cuve. Les rectangles rouges correspondent aux bandes extraites et envoyées pour le séquençage des protéines. *c* : surnageant après lyse cellulaire; *sa* : surnageant concentré par tube Amicon après l'ultracentrifugation; *p1* : tampon de resuspension des microsomes 1; *p2* : tampon de resuspension des microsomes 2. * Tampon Laemmli sans β -mercaptoéthanol.

Le résultat de l'identification par le centre de protéomique a révélé que la protéine RmPT280 est présente dans les trois bandes envoyées pour analyse (Figure 3.13). On a retrouvé pour chacune des trois bandes entre 4 et 7 peptides différents correspondant à la séquence de RmPT280 avec plus de 95 % de probabilité. La protéine CsPT1 est également présente dans les deux bandes envoyées. On y retrouve entre 2 et 4 peptides différents avec une probabilité à plus de 95 %. La présence des protéines d'intérêt dans

l'entièreté des bandes envoyées pour identification confirme que ce ne sont pas des bandes non spécifiques.



Figure 3.13 Résultats de l'identification de protéines dans les bandes envoyées à la Plateforme de Protéomique.

Les bandes envoyées pour analyser sont représentées par des rectangles rouges sur la figure précédente. Trois bandes ont été envoyées pour la protéine recombinante RmPT280, soit une bande dont la taille est environ de 50 kDa, une bande à environ 37 kDa et une bande à environ 100 kDa. Dans la bande à 50 kDa, 4 peptides ont été identifiés comme appartenant à la protéine RmPT280 avec plus de 95 % de probabilité. Dans la bande à 37 kDa et à 100 kDa, 7 peptides et 4 peptides ont respectivement été identifiés avec plus de 95 % de probabilité. Deux bandes ont été envoyées pour la protéine recombinante CsPT1, soit une bande à environ 37 kDa et une bande à environ 45 kDa. 4 peptides et 2 peptides ont respectivement été identifiés comme appartenant à la protéine CsPT1 avec plus de 95 % de probabilité.

Parmi les peptides identifiés pour les protéines RmPT280 et CsPT1, aucun des sites de liaison au GPP ou au magnésium n'a été identifié. Pour la protéine RmPT280, le début du domaine prényltransférase UbiA a été identifié. Les peptides identifiés ne correspondent majoritairement pas à une structure secondaire, c'est-à-dire à une hélice ou à un feuillet, mais correspondent plutôt à des connecteurs également nommés coudes. Les coudes sont des sections variables qui permettent un changement drastique d'orientation du squelette protéique.

3.8 Essais enzymatiques

L'objectif avec les essais enzymatiques est de parvenir à identifier quels sont les meilleurs substrats pour nos protéines d'intérêt, mais également quelles sont les meilleures

conditions de réaction. Dans un premier temps, il est important de parvenir à détecter une activité enzymatique avec la CsPT1 qui pourra servir de contrôle positif tout au long des essais enzymatiques. Donc, les premiers substrats à être testé avec toutes les protéines recombinantes ont certainement été l'OA et le GPP, soit les deux substrats de la CsPT1.

Différentes conditions enzymatiques ont été testées au cours de cette programmation de recherche. Le choix des conditions a été déterminé à la suite d'une recherche dans la littérature sur les conditions préalablement utilisées par d'autres équipes de recherche pour des enzymes similaires ou de la même famille. Au total, quatre conditions différentes ont été testées (Tableau 2.2).

Parmi les conditions essayées, la condition 1 n'a pas donné de produits intéressants. La condition 2 n'a rien donné de concluant également. Cependant, avec la condition #3, 5 ppm de CBGA ont été détectées dans la fraction microsomale *p1* de la protéine recombinante CsPT1 (Figure 3.14). Malheureusement, il n'a jamais été possible de reproduire ce résultat en faisant d'autres essais enzymatiques dans ces mêmes conditions.

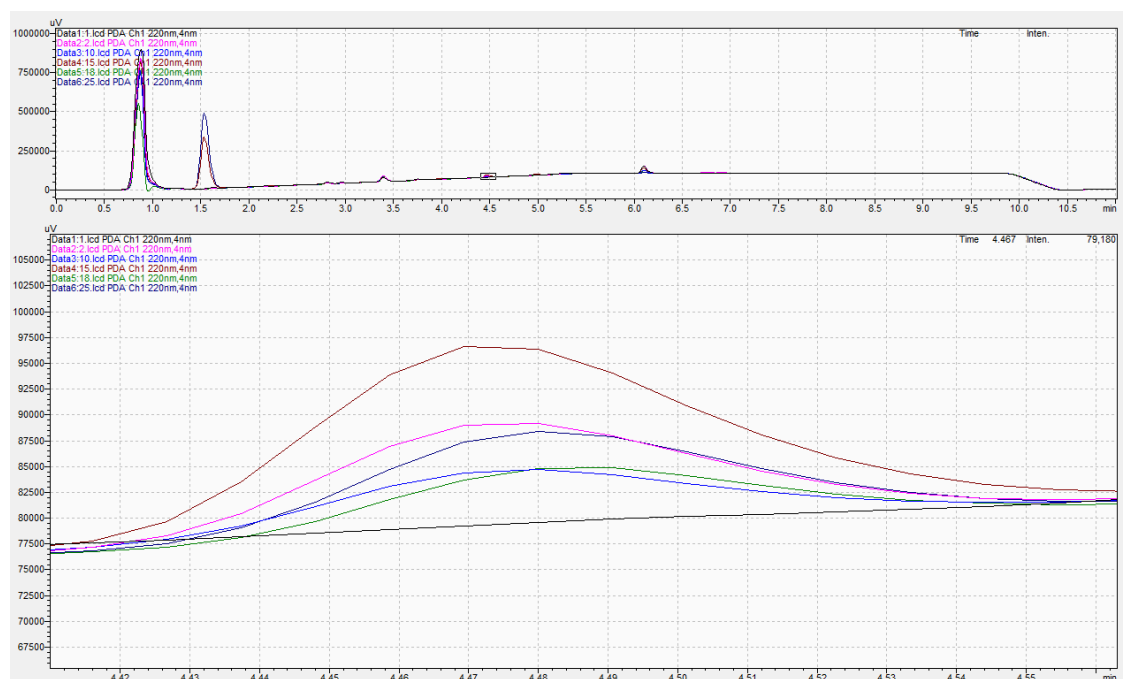


Figure 3.14 Détection de 5 ppm de CBGA dans un essai enzymatique effectué sur les protéines extraites d'un clone contenant le vecteur pYES2/CT + CsPT1.

La réaction enzymatique a été effectuée à 30 °C pendant 16 heures dans un volume final de 50 µL. Les réactifs utilisés sont 2 mM OA, 2 mM GPP, 25 mM MgCl₂, 100 mM HEPES et 14,3 µL de protéines du tampon de resuspension de la fraction microsomale *p1*. La réaction a été arrêtée avec du TCA à une concentration finale de 2 % avant d'être filtrée sur colonne et analysée par HPLC selon la méthode optimisée dans le laboratoire pour la détection de cannabinoïdes.

Avec la condition 4, les essais enzymatiques ont révélé, au HPLC, la présence d'un composé ayant le même temps de rétention que le CBGA, mais avec une absorbance différente. Ce résultat a été observé pour des fractions protéiques extraites de la levure non transformée et de la levure transformée avec le vecteur pYES2/CT + RmPT280. Toutefois, les échantillons n'ont pas été analysés par le LC-MS-MS du laboratoire afin d'identifier le composé synthétisé.

CHAPITRE IV

DISCUSSION

Les métabolites spécialisés, comme les cannabinoïdes, sont souvent des molécules d'intérêt pour le domaine médical. Les cannabinoïdes sont utilisés comme traitement depuis des milliers d'années alors que les analogues aux cannabinoïdes ne sont pas encore très bien connus. C'est pourquoi dans cette programmation de recherche, nous nous sommes intéressées à la voie de biosynthèse des molécules analogues aux cannabinoïdes et plus précisément aux enzymes catalysant une réaction de prénylation. Les molécules trouvées précédemment chez *R. marginata* suggèrent que cette plante hépatique possède des enzymes qui peuvent jouer un rôle similaire à celle présente chez le cannabis. Pour les identifier, le transcriptome de *R. marginata* a été assemblé et étudié afin de déterminer les protéines d'intérêt. Nous avons pu identifier, dans le transcriptome de *R. marginata*, deux protéines possiblement responsables de la prénylation du GPP sur l'acide stilbène. Ces protéines ont été analysées et ensuite transformées dans un système d'expression hétérologue afin de pouvoir les caractériser.

4.1 Obtention et assemblage des données de séquençage Illumina de *R. marginata*

L'obtention et l'assemblage des données de séquençage de *R. marginata* se sont révélés concluants. Les données déposées par l'équipe de chercheurs en Allemagne avaient également été assemblées par leur équipe et le résultat de leur assemblage est disponible dans leur publication. En utilisant sensiblement la même méthode d'assemblage, ils sont parvenus à assembler 1 580 612 transcrits, dont la taille médiane est de 311 pb et la taille moyenne est de 557 pb [39]. En comparaison, il a été possible pour nous d'assembler 1 641 458 transcrits, dont la taille médiane est de 397 pb et la taille moyenne est 1515 pb. Bien que le nombre de transcrits assemblé et la taille médiane soient similaires, nous avons obtenu une taille moyenne significativement supérieure, ce qui

s'avère intéressant lorsque l'objectif est d'identifier des séquences protéiques avec un cadre de lecture complet.

Cependant, ce type d'assemblage *de novo* comprend certaines limitations. La première provient du fait que les plantes hépatiques, dont *R. marginata*, vivent communément en symbiose avec certaines espèces de champignons [39]. Donc, lors de l'extraction de l'ARN d'une plante de si petite taille, il est pratiquement impossible de s'assurer que la biomasse récoltée provient uniquement de l'espèce visée. Il est donc très probable que certains transcrits assemblés proviennent d'espèces fongiques ou même bactériennes puisque les échantillons n'ont pas été stérilisés avant l'extraction. Cependant, les protéines intéressantes identifiées dans le transcriptome se sont révélées majoritairement homologues à des protéines d'origine végétale lors de la recherche dans la base de données protéique de BLASTP. Le résultat serait certainement différent si les protéines identifiées provenaient d'un champignon ou d'une bactérie.

La deuxième limitation majeure dans ce type d'assemblage est qu'en l'absence d'un génome de référence, il est possible que les fragments d'ARN de certaines protéines qui se seraient avérées intéressantes n'aient pas été correctement assemblés et donc les protéines ont été rejetées lorsque nous avons sélectionné celles avec un cadre de lecture complet seulement.

Toutefois, dans un contexte de pandémie mondiale, avec une espèce endémique de la Nouvelle-Zélande et de Tasmanie et en absence d'un génome de référence, c'est en toute connaissance des limitations de la méthode d'assemblage que nous avons décidé de poursuivre la programmation de recherche.

4.2 Analyses *in silico*

Compte tenu de toutes les analyses *in silico* effectuées, on peut dire que les modèles prédits sont satisfaisants. En effet, les prédictions sont similaires parmi les différents

logiciels d'analyse. Bien évidemment, tout ce qui est présenté dans cette section de la programmation de recherche demeure simplement une prédiction.

4.2.1 Détermination des domaines conservés chez les prényltransférases

L'analyse des séquences des prényltransférases a révélé la présence d'un domaine conservé. La portion conservée, allant des acides aminés 108 à 383 de la CsPT1, a été identifiée comme un domaine UbiA. Le domaine UbiA est un domaine catalysant une prénylation Mg^{2+} -dépendante, dans ce cas-ci, la prénylation du GPP sur l'OA en présence de $MgCl_2$.

4.2.2 Identification des séquences pertinentes chez *R. marginata*

Ce domaine conservé a été utilisé pour trouver des protéines pouvant possiblement catalyser une réaction similaire dans le transcriptome de *R. marginata*. À la suite de l'analyse transcriptomique, deux protéines intéressantes ont été identifiées, soit RmPT280 et RmPT130.

L'utilisation du domaine conservé était une tactique afin d'obtenir possiblement plus de protéines candidates en se basant seulement sur l'homologie avec ce domaine précis. En réalité, il nous importe peu que les protéines identifiées dans le transcriptome aient un haut niveau d'homologie avec la CsPT1. Ce qui nous importe réellement c'est que ces protéines possèdent une bonne homologie avec le domaine conservé, qui s'avère être le domaine nécessaire pour réaliser la réaction de prénylation qui nous intéresse.

4.2.3 Prédictions des structures des protéines d'intérêts

Les deux protéines identifiées chez *R. marginata* possèdent des similarités avec la CsPT1 au niveau de leur séquence et de leur conformation structurale. Ces similitudes suggèrent que le type de réaction de prénylation qui a lieu dans les plants de cannabis existe depuis très longtemps puisqu'on retrouve des protéines très similaires dans une

plante ancestrale comme *R. marginata*. Cependant, il se peut que la réaction de prénylation serve un objectif bien différent *in vivo* chez les deux organismes puisqu'il a bien été démontré que les prényltransférases font preuve d'une grande promiscuité enzymatique. Ainsi, même si elles ont la capacité d'utiliser les mêmes substrats, elles ne sont pas nécessairement utilisées pour catalyser la même réaction. En effet, comme les analyses phylogénétiques l'ont démontré, aucune des deux protéines identifiées ne fait partie du groupe menant à la biosynthèse des terpénophénols comme la CsPT1. L'arbre phylogénétique démontre que la protéine RmPT280 serait d'avantage lié évolutivement aux protéines impliquées dans la biosynthèse des plastoquinones tandis que la protéine RmPT130 serait davantage liée aux protéines impliquées en ce qui concerne la biosynthèse des tocophérols. Les trois protéines ne sont pas situées dans les mêmes groupes phylogénétiques, ce qui pourrait suggérer qu'elles n'ont pas la même fonction *in vivo*. La protéine RmPT280 se révèle quand même plus proche, au point de vue évolutif, de la CsPT1 que la protéine RmPT130. Il est donc toujours possible que les protéines identifiées dans le transcriptome soient capables de catalyser la réaction de prénylation à l'étude chez *R. marginata*, ou encore la réaction de prénylation menant à la synthèse du CBGA.

Les alignements multiples ont indiqué que les sites de liaison du GPP (I159, N160, D164, D168 et G222) sont préservés dans la pochette catalytique ainsi que les sites de liaison du magnésium (R108, N160, D164 et D168), ce qui laisse croire que la fonction pourrait être conservée chez ces deux protéines de *R. marginata*. Cependant, on ignore toujours quels acides aminés entourent le substrat aromatique et s'ils sont préservés chez les différentes prényltransférases. Puisqu'aucune prényltransférase du cannabis n'a été cristallisée, il est certain que les modèles de prédiction établis par les différents logiciels ne donnent qu'une idée globale de ce que pourrait être la structure des protéines identifiées dans *R. marginata*. Ainsi, même si toutes analyses se révèlent encourageantes, rien ne peut être considéré comme absolument véridique.

4.3 Transformation dans la levure

À la suite des analyses *in silico*, de l'assemblage des vecteurs d'expression et d'une première réplication dans la bactérie, les vecteurs ont finalement été transformés dans la levure. La transformation dans la levure est une méthode couramment utilisée et bien documentée. Les méthodes les plus populaires sont la transformation par choc thermique ou par électroporation, mais il existe aussi d'autres méthodes comme la méthode des sphéroplast, par biolistique et avec billes de verre. Peu importe la méthode, les études ont démontré que les cations monovalents, et plus précisément le lithium, améliorent grandement l'efficacité de la transformation chez la levure. Le PEG est également un élément important dans la transformation puisqu'il permet de lier les molécules d'ADN à la paroi des cellules. Nous nous sommes donc assurés que ces composants font bien partie du protocole utilisé pour la transformation par choc thermique [72].

Les méthodes de transformation par choc thermique et par électroporation sont toutes les deux très répandues. Évidemment, elles contiennent chacun leurs avantages et inconvénients. La méthode de transformation par choc thermique est légèrement plus complexe et nécessite plus de matériel, mais pour nous, elle s'est avérée efficace dès le premier essai. La méthode de transformation par électroporation a également été testée préalablement à la méthode par choc thermique, mais elle ne s'est pas avérée efficace. La salinité du milieu de culture a conduit à la production d'étincelles lors du choc électrique, ce qui a mené à l'échec de la transformation. L'optimisation de cette méthode de transformation aurait pu nous coûter un temps précieux, c'est pourquoi nous avons opté pour le protocole de transformation par choc thermique qui s'était déjà avéré efficace pour la levure dans notre laboratoire.

4.4 Expression et extraction des protéines recombinantes

L'expression des protéines débute lorsque le glucose du milieu de culture est remplacé par du galactose afin d'activer le promoteur GAL1. Le promoteur GAL1 est un des promoteurs les plus forts pour la levure [73]. Il n'existe cependant aucun contrôle dans

cette programmation de recherche pour s'assurer que le galactose active correctement le promoteur et que les protéines sont fortement exprimées.

Selon les résultats obtenus par les Western blots, une induction de 24 heures semble être la condition idéale pour extraire les protéines d'intérêt. Bien évidemment, la détection n'était pas optimisée à ce moment-là pour la détection des protéines par l'anticorps, c'est la raison pour laquelle on retrouve beaucoup de bandes non spécifiques.

Avant de procéder à l'extraction des protéines, il faut se rappeler que les cellules de levures sont particulièrement difficiles à briser dû à la présence d'une paroi cellulaire rigide en plus d'une membrane plasmique. Ainsi, les simples méthodes de lyse, comme la lyse chimique ou la lyse physique, à l'aide d'un Potter par exemple, ne sont pas efficaces ici. Dans la littérature, les méthodes communément utilisées pour lyser les cellules de *S. cerevisiae* sont soit par homogénéisation haute pression, par l'utilisation de billes de verre, par sonication, par l'utilisation d'azote liquide ou encore par lyse chimique. Certaines de ces méthodes peuvent être préalablement jumelées à une digestion enzymatique par la zymolyase, celle-ci étant efficace pour détruire la paroi cellulaire et mener à la formation des sphéroplastes [50,71,74].

Dans cette programmation de recherche, il n'était pas possible d'utiliser la méthode de lyse par homogénéisation haute pression due à l'indisponibilité de ce type d'équipement. La sonication est une méthode plus ou moins appréciée dans la communauté scientifique étant donné son efficacité moyenne, elle n'a donc pas été prise en compte. La lyse des cellules par l'azote liquide n'a pas été considérée afin d'éviter le bris ou la dénaturation des protéines. La lyse chimique n'a également pas été considérée par crainte que les réactifs interfèrent ensuite lors des essais enzymatiques. Nous nous sommes donc tournés vers la lyse par l'utilisation de billes de verres, une méthode efficace et très accessible.

Des analyses au microscope ont démontré que le protocole de lyse utilisé initialement n'était pas efficace. En effet, très peu de cellules étaient lysées avec des billes

de verre en alternant le vortex et une incubation sur glace pour 4 minutes. Cette découverte remet en doute le moment d'extraction des protéines après 24 heures d'induction. Et si une induction de 48 heures permettait d'obtenir plus de protéines recombinantes, mais qu'il était seulement impossible de bien lyser les cellules?

Un autre test a alors été réalisé afin d'optimiser la méthode de lyse. Les cellules de levure ont été lysées par TissueLyser. La quantité de protéines obtenue pendant la lyse a grandement augmenté, mais il était ensuite impossible de détecter les protéines d'intérêt par Western blot. Nous avons donc pensé que la lyse était trop agressive et endommageait les protéines extraites. La méthode de lyse par billes de verre a donc été reprise puisque cette méthode permettait au moins de détecter les bandes d'intérêt par Western blot. Le nombre d'alternances entre le vortex et l'incubation sur glace a été augmenté à 10 minutes pour tenter de récupérer plus de protéines lors de la lyse. Cependant, cette augmentation n'a pas permis d'obtenir de meilleurs résultats en ce qui concerne la détection par Western blot. Ainsi, tous les résultats présentés dans ce document concernant l'extraction et l'enrichissement de protéines sont basés sur une lyse de 4 minutes.

L'extraction des protéines est une étape critique dans cette programmation de recherche. Elle est complexe, mais sa réussite demeure essentielle pour la caractérisation des protéines recombinantes. La complexité de cette étape provient d'une part de la difficulté à lyser efficacement les cellules de levure, mais également des importantes limitations qu'il faut considérer. En effet, lors de l'extraction des protéines, il faut s'assurer que la température à laquelle les échantillons sont manipulés demeure à environ 4 °C afin de protéger les protéines de la dénaturation ou de la dégradation par des protéases. De plus, afin de prévenir la dégradation par les protéases, du PMSF est ajouté au tampon de lyse, mais également dans les tampons de resuspension. Le PMSF est un inhibiteur de protéases à sérine. Il se lie donc spécifiquement au résidu sérine du site actif d'une protéase à sérine. En revanche, il s'hydrolyse et perd son efficacité au contact de l'eau. En effet, sa durée de vie dans l'eau est d'environ 1 h à 25 °C et d'environ 22 heures à 4 °C pour une solution à pH 8 [75]. Il doit donc être ajouté au dernier moment et renouvelé au besoin afin de maximiser son efficacité. Ainsi, dans le protocole d'extraction

des protéines, l'ultracentrifugation de 1 h 30 est une étape cruciale et elle doit être faite rapidement après l'ajout du PMSF.

4.5 Immunobuvardage de type Western blot

Le Western blot dans lequel nous avons testé l'Endo H a démontré que les protéines d'intérêt n'ont pas subi de N-glycosylation puisque les tailles des bandes révélées par Western blot sont restées les mêmes. Toutefois, il est impossible de s'assurer que les protéines intéressantes n'ont pas subi d'autres modifications post-traductionnelles pouvant affecter leur taille. L'ensemble des analyses par Western blot a également permis de démontrer que la méthode d'extraction permet réellement l'enrichissement des protéines d'intérêt dans les fractions microsomales *p1* et *p2*, ce qui vient appuyer l'hypothèse que les protéines recombinantes initialement exprimées dans les plastides se retrouvent dans le réticulum endoplasmique lors de l'expression hétérologue dans la levure.

Cependant, ce type d'analyse comporte également certaines limitations. Une des limitations majeures est l'absence d'un contrôle positif pour la détection des protéines recombinantes par l'anticorps 1D4. Ce contrôle aurait permis de valider l'efficacité de l'anticorps primaire monoclonal 1D4 ainsi que de permettre une comparaison entre le contrôle positif et les bandes révélées lors de nos analyses. Nous avons fait le choix de ne pas utiliser de contrôle positif en raison des options disponibles sur le marché. Il aurait été possible de commander un petit peptide dont la séquence est celle des neuf acides aminés de l'épitope détecté par l'anticorps primaire. Cependant, ce peptide de très petite taille aurait certainement migré hors du gel d'acrylamide lors de l'électrophorèse. Ainsi, il n'aurait pas été possible de le détecter simultanément aux protéines d'intérêt, donc il perd son utilité. Puisque, selon la littérature, l'anticorps 1D4 est hautement spécifique et se lie très rarement à des bandes non spécifiques, nous avons pris la décision de poursuivre la programmation de recherche sans ce contrôle positif [76]. Il aurait toutefois été possible de contacter une autre institution afin de leur demander s'ils

possèdent des protéines recombinantes purifiées avec le même épitope afin de les utiliser comme contrôle positif.

Comme mentionné dans le paragraphe précédent, l'anticorps 1D4 se lie très rarement à des bandes non spécifiques, donc nous avons voulu investiguer davantage la raison pour laquelle plusieurs bandes de tailles différentes sont révélées par l'immunobuvardage de type Western blot. C'est pourquoi cinq bandes ont été extraites d'un gel de polyacrylamide après sa migration et envoyée à la Plateforme de Protéomique du Centre de Recherche du CHU de Québec. Ces analyses ne permettent pas d'obtenir un séquençage complet des protéines, mais servent plutôt à identifier toutes celles présentes dans les bandes extraites par un séquençage de peptides. Pour chacune des cinq bandes envoyées à La Plateforme de Protéomique, entre 2 à 7 peptides différents ont été identifiées avec plus de 95 % d'identité comme appartenant aux protéines d'intérêt. Il est donc possible de considérer que les protéines d'intérêt se trouvent bien dans chacune des bandes extraites. En effet, selon les indications des membres de La Plateforme de Protéomique, il est nécessaire d'identifier au moins 2 peptides différents appartenant, avec plus de 95 % de probabilité, à une même protéine afin de considérer cette protéine comme étant présente dans l'échantillon analysé. En revanche, comme seulement quelques peptides ont été identifiés pour chaque protéine d'intérêt, il est impossible de connaître la séquence complète de la protéine et ainsi expliquer pourquoi on les retrouve également dans des bandes de tailles différentes de celle attendue. Il est difficile de dire si la séquence est tronquée et à quelle position. Le résultat de ces analyses fait naître une multitude de questions.

En premier lieu, il faut se rappeler que ces bandes sont révélées par Western blot et donc contiennent toujours l'épitope ciblé par l'anticorps primaire 1D4. Donc, la fin de la séquence ne semble pas être la section affectée. La protéine recombinante RmPT280 a été identifiée dans des bandes d'environ 37 kDa, 50 kDa et 100 kDa alors que sa taille attendue est de 49 kDa. Si nous prenons le premier peptide identifié dans la bande à 37 kDa, soit le peptide TPSIAPLCR qui apparaît à la position 39, et nous supposons que la partie N-terminale de la protéine recombinante soit tronquée jusqu'à cette position,

la taille attendue de la protéine RmPT280 serait tout de même de 45 kDa. Alors pourquoi retrouvons-nous la protéine dans une bande à 37 kDa? L'hypothèse actuelle est qu'il pourrait s'agir d'un phénomène d'épissage qui retirerait une section intrinsèque de la protéine recombinante. L'épissage est une modification post-transcriptionnelle qui permet de retirer un intron, une portion non codante, d'un ARN messenger précurseur. Nous présumons que nos séquences ne possèdent pas d'introns puisqu'elles possèdent un cadre de lecture ouvert et continu. Nous pensons donc que ce sont des séquences qui ont déjà subi l'épissage. Cependant, une sorte de quelconque épissage pourrait expliquer que les protéines soient tronquées en milieu de séquence. Il n'est pas commun de voir un second épissage sur un ARN mature, ce phénomène a seulement été recensé sur des cellules cancéreuses [77]. Comme c'est exactement le même phénomène qui se produit pour la protéine recombinante CsPT1, cela laisse croire que l'origine de la protéine n'en est pas la cause, mais qu'il s'agirait probablement du système d'expression hétérologue. Afin de lever le voile sur ce mystère, il aurait pu être intéressant de réaliser un séquençage de l'ARN des cellules de levure transformées et ainsi savoir si le phénomène survient réellement au moment de l'expression.

Finalement, les peptides principalement identifiés lors du séquençage correspondent à des coudes, ce qui mène à d'autres réflexions concernant la récurrence du genre de peptide identifié. Puisque les coudes sont les séquences qui permettent un changement drastique de direction et que pour une protéine membranaire, cela signifie que ce sont des séquences qui se retrouvent principalement à l'extérieur des membranes et donc des séquences hydrophiles, il faut se demander si ce type d'analyse est vraiment optimal pour des protéines membranaires. Peut-être qu'avec un tampon différent de celui utilisé par La Plateforme, un tampon contenant un détergent par exemple, il aurait été possible d'identifier plus de peptides pour chaque protéine d'intérêt.

4.6 Essais enzymatiques

Afin de choisir les conditions à utiliser pour les essais enzymatiques, une revue de la littérature a été effectuée afin de rassembler les conditions fréquemment utilisées pour

mesurer l'activité des prényltransférases. Le résultat de la recherche est exposé dans le tableau suivant (Tableau 4.1).

Tableau 4.1

Résumé des différentes conditions utilisées dans la littérature pour des essais enzymatiques sur des prényltransférases
Tiré de [50,73,78-85].

	Page et al., 2014	Manukata et al., 2014	Zirpel et al., 2017	Zhong et al., 2018	Luo X et al., 2019
Substrat 1	0,2 mM OA	1 mM umbelliférone	1 mM OA	0,2 mM oxyresveratrol	0,5 mM OA
Substrat 2	1 mM GPP	1 mM GPP	1 mM GPP	0,2 mM GPP	0,5 mM GPP
Cofacteur	5 mM MgCl ₂	10 mM Mgcl ₂	5 mM MgCl ₂	10 mM Mgcl ₂	10 mM Mgcl ₂
Volume (µL)	100	200	-	100	50
Qt protéines (µg)	64	-	-	50	-
Tampon	100 mM Tris-HCl	100 mM Tris-HCl	50 mM Tris-HCl	100 mM Tris-HCl	50 mM Tris-HCl
pH	7,5	7,5	7,5	8	8,5
Température (°C)	37	30	37	30	30
Durée (h)	1	1	4	2	1
Provenance	<i>C. sativa</i>	<i>C. limon</i>	<i>Streptomyces sp.</i>	<i>M. alba</i>	<i>C. sativa</i>
Expression	<i>S. cerevisiae</i>	<i>S. cerevisiae</i>	<i>S. cerevisiae</i>	<i>S. cerevisiae</i>	<i>S. cerevisiae</i>

	Qian et al. 2019	Bai et al., 2020	He et al., 2020	Yang et al., 2020	Ni et al., 2020
Substrat 1	2 mM OA et DVA	1 mM p-formylphénol	1 mM indoles	0,5 mM flavones	0,5 mM menadione hydroquinol
Substrat 2	2 mM GPP	1 mm DMAPP	0,1 mM DMAPP	1 mM DMAPP	0,25 mM DMAPP
Cofacteur	5 mM MgCl ₂	-	4 mM MgCl ₂	10 mM Mg ²⁺ et Mn ²⁺	2,5 mM MgCl ₂
Volume (µL)	200	100	100	200	200
Qt protéines (µg)	200	30 à 50	368	30	200
Tampon	50 mM HEPES	50 mM Tris-HCl	50 mM Tris-HCl	50 mM Tris-HCl	70 mM Tris-HCl
pH	7,5	7,5	8,5	9	8
Température (°C)	RT	37	30	37	32
Durée (h)	18	4	O/N	4	1
Provenance	<i>Streptomyces sp.</i>	<i>S. vibrans</i>	<i>A. alternata</i>	<i>A. heterophyllus</i>	<i>S. niveus</i>
Expression	<i>E. coli</i>	<i>E. coli</i>	<i>E. coli</i>	<i>S. cerevisiae</i>	<i>E. coli</i>

Selon les informations rassemblées, le substrat 1 varie d'une publication à l'autre, mais en général, il s'agit très souvent d'un composé aromatique. Le substrat 2 est majoritairement du GPP ou du DMAPP. Le cofacteur MgCl₂ est le cofacteur utilisé dans la majorité des essais enzymatiques pour les prényltransférases. Le volume de réaction

varie entre 50 et 200 μL et la quantité de protéines varie entre 30 et 368 μg . Le tampon de réaction est majoritairement le Tris-HCl, à une concentration de 100 mM ou moins. Le pH tourne généralement dans les environs de 7,5 ou 8, la température d'incubation est majoritairement comprise entre 30 et 37 °C et le temps d'incubation est souvent entre 1 et 4 heures.

Basés sur cette analyse, nous avons choisi les différentes conditions à tester lors de nos essais enzymatiques. Cependant, aucune des conditions testées n'a pu fournir de résultats reproductibles. En effet, sur les cinq conditions testées, la condition #3 avait mené à la détection de CBGA par HPLC pour la protéine CsPT1 et la condition #4 avait mené à la détection d'un produit ayant le même temps de rétention que le CBGA, mais avec une absorbance différente. Puisque ce dernier composé a été détecté dans la levure transformée et la levure non transformée, il pourrait s'agir d'un produit synthétisé par les protéines endogènes de la levure en présence des substrats fournis. Cependant, comme ce composé se trouve également chez la levure non transformée, les analyses n'ont pas été poussées plus loin.

Ces deux conditions, les conditions #3 et #4, ont été répétées à plusieurs reprises, sans succès. Le manque de reproductibilité pour ce qui est des essais enzymatiques pourrait indiquer un problème en ce qui concerne l'activité des enzymes. En effet, ce genre de problème a déjà été reporté par le passé. Une autre équipe de recherche a tenté l'expression hétérologue de la CsPT1 dans la levure *S. cerevisiae* et ils ne sont jamais parvenus à obtenir l'enzyme fonctionnellement active [73].

Toutefois, nous avons prévu tester différents substrats afin de pouvoir caractériser les protéines recombinantes. Ci-dessous sont dressées la liste des composés et la raison pour laquelle nous les avons sélectionnés (Tableau 4.2).

Tableau 4.2

Liste des substrats intéressants à tester lors des essais enzymatiques

Groupe aromatique accepteur	Justification
Acide olivetolique	Substrat de la CsPT1
Acide stilbène	Substrat supposé de la PT de <i>R. marginata</i>
Resvératrol	Similaire à l'acide stilbène, mais plus courant
Acide homogentisique	Substrat dans la biosynthèse des tocophérols
Acide caféique	Acide aromatique simple et très courant
Groupe prényl donneur	Justification
GPP	Substrat de la CsPT1
DMAPP	Fréquemment utilisé dans la littérature

4.7 Hypothèses

Les résultats non concluants obtenus lors de l'activité des protéines recombinantes laissent place à plusieurs pistes de discussion et d'hypothèses.

Selon les résultats du séquençage sur les différentes colonies sélectionnées après la transformation dans la levure, les différentes cassettes d'expression insérées dans le vecteur pYES2/CT ne possédaient aucune mutation, insertion ou délétion. Donc, les levures transformées possèdent bien la bonne cassette pour exprimer la protéine d'intérêt correctement. La section suivante permet de développer les différentes hypothèses pouvant expliquer l'absence d'activité enzymatique.

Puisqu'aucune activité enzymatique n'a été détectée pour les protéines recombinantes provenant de *R. marginata* et de *C. sativa*, il est probable que le problème ne provienne pas des séquences originelles, mais bien de l'expression, de l'extraction ou des conditions de caractérisation de ces protéines. Comme mentionner précédemment, il n'existe aucun contrôle pour ce qui est de l'expression des protéines d'intérêt. Nous ignorons si la transcription des gènes est bien induite par l'ajout de galactose dans

le milieu de culture. Nous ignorons également si les protéines subissent des modifications post-traductionnelles ou si elles parviennent à se replier correctement. Le mauvais repliement de la protéine est également un phénomène qui est possible et qui peut affecter l'activité. En effet, il n'est pas rare de voir dans la littérature que l'épitope affecte le repliement de la protéine. C'est pourquoi dans cette programmation de recherche nous avons inséré un site de clivage par la thrombine devant l'épitope 1D4. Toutefois, cette option n'a pas été testée. Nos efforts ont plutôt été déployés afin d'élucider le mystère concernant la présence de nos protéines d'intérêt dans des bandes de tailles différentes lors des analyses de type Western blot.

Il est également possible que l'absence d'activité soit causée par la lyse. En effet, peut-être est-ce la lyse ou les tampons utilisés lors de la lyse qui entraîne une dénaturation des protéines ou nuit à leur efficacité catalytique.

La dernière hypothèse expliquant l'absence d'activité consiste en une mauvaise sélection des conditions de réaction. Avec les prédictions obtenues par les analyses *in silico* et les informations rassemblées dans la littérature concernant les prényltransférases, nous avons choisi des conditions, mais peut-être celles-ci n'étaient vraiment pas optimales pour que les enzymes de *R. marginata* soient actives. En revanche, avec les conditions choisies, la protéine recombinante CsPT1 devrait démontrer une certaine activité, ce qui laisse croire que les conditions enzymatiques ne sont pas la véritable cause de l'absence d'activité.

4.8 Améliorations

À l'aube des discussions entamées dans ce présent chapitre, voici l'ensemble des améliorations suggérées dans le cadre de cette programmation de recherche.

En premier lieu, il aurait été intéressant de procéder au séquençage complet de nos vecteurs d'expression après la transformation dans la levure. En effet, il a été porté à notre attention récemment que la compagnie Plasmidsaurus offrait un séquençage rapide et peu

coûteux de vecteurs complets. Dans ce projet de recherche, nous avons procédé au séquençage Sanger des cassettes d'expression après la transformation dans la levure. Or, vu les problèmes liés à la détection et à l'activité des protéines d'intérêt, il aurait été pertinent de procéder au séquençage complet des vecteurs afin de s'assurer qu'il n'y a pas de mutations ailleurs dans la séquence et qui pourrait affecter, par exemple, le promoteur ou le terminateur.

Comme mentionner dans la section précédente, il aurait également pu s'avérer intéressant de procéder à l'extraction et au séquençage de l'ARN de la levure transformée afin de comprendre si les protéines subissent un clivage intrinsèque affectant leur taille et à quelle position dans leur séquence protéique.

Finalement, le manque de contrôles positifs en ce qui touche la détection par Western blot et pour l'activité enzymatique a été un facteur limitant durant ces travaux de recherche. Pour ce qui concerne la détection, la petite taille de l'épitope 1D4 fait en sorte qu'il aurait été difficile de le révéler par Western blot sans modifier le pourcentage du gel de séparation et ainsi affecter la révélation de nos protéines d'intérêt. L'amélioration suggérée est donc de trouver et de contacter une autre équipe de recherche qui possède une protéine purifiée avec l'épitope 1D4 afin de comparer et valider l'efficacité de notre anticorps primaire. Pour ce qui est de l'activité enzymatique, la protéine recombinante CsPT1 devait nous servir de contrôle positif. Cependant, l'absence d'activité enzymatique pour cette protéine recombinante vient troubler nos plans. Comme la séquence de la prényltransférase du cannabis est brevetée, il est difficile de demander à une autre équipe de recherche de nous fournir la protéine purifiée active pouvant être utilisée comme contrôle positif. L'amélioration suggérée dans ce cas est de créer une cassette d'expression avec la séquence de la prényltransférase 4 du cannabis (CsPT4). Cette protéine recombinante a démontré la capacité de catalyser la même réaction catalytique que la CsPT1 à quelques reprises lorsqu'elle est exprimée dans un système d'expression hétérologue [80,86,87]. Peut-être cette protéine aurait-elle été plus adéquate pour servir de contrôle positif.

CHAPITRE V

CONCLUSION

Connaissant aujourd'hui le potentiel curatif des cannabinoïdes, il est maintenant temps de découvrir s'il en va de même pour leurs analogues. Afin de révéler leur potentiel, il est dans l'intérêt de la communauté scientifique d'élucider leurs voies de biosynthèse et de comprendre davantage leur fonctionnement sur notre métabolisme. Leur découverte dans différentes espèces végétales au cours des dernières années génère un défi de taille pour les chercheurs. Par exemple, dans les dernières années, des analogues aux cannabinoïdes ont été identifiés dans *R. perrottetii*, *R. marginata*, *R. laxiramea*, *L. usitatissimum*, *H. umbraculigerum* et *P. nigrum* [39,45,47,48]. Afin d'élucider les voies métaboliques, les enzymes susceptibles d'être impliquées dans la biosynthèse des analogues sont majoritairement exprimées dans un système d'expression hétérologue afin d'être caractérisées et ainsi valider ou infirmer l'hypothèse. La découverte de ces nouvelles protéines permet une meilleure compréhension des voies de biosynthèse, mais permet également de fournir des outils pour la production de cannabinoïdes dans des systèmes hétérologues par ingénierie métabolique.

Dans le but de participer à l'élucidation des voies métaboliques menant à la biosynthèse des molécules analogues aux cannabinoïdes, ce projet de maîtrise a présenté deux protéines intéressantes potentiellement capables de catalyser la réaction de prénylation menant à la formation du PET. Nous avons utilisé les données transcriptomiques de la plante hépatique *R. marginata* afin d'assembler un transcriptome dans lesquelles il serait possible d'identifier ces protéines. Les protéines d'intérêt ont été analysées *in silico* et les cassettes d'expression ont été générées par bio-informatique et fournies par la compagnie BioBasic. Les vecteurs d'expression ont été assemblés par la méthode Gibson et finalement transformés dans le système d'expression hétérologue, la levure *Saccharomyces cerevisiae*. Les protéines recombinantes ont été extraites et

révélées par immunobuvardage de type Western blot avant de débiter la caractérisation par essais enzymatiques.

En regard avec l'objectif principal de cette programmation de recherche qui était de caractériser les enzymes impliquées au niveau de la réaction de prénylation menant à la formation du PET chez *R. marginata*, nous pouvons dire globalement qu'il n'a pas été atteint. Cependant, sur les quatre étapes que nous nous étions fixées pour parvenir à atteindre cet objectif, les trois premières ont été complétées avec succès. En effet, la première étape qui consistait à récupérer et assembler les données transcriptomiques de *R. marginata* a permis d'assembler 1 641 458 transcrits parmi lesquels il a été possible, par la suite, d'effectuer une recherche et d'identifier des protéines intéressantes. Par la suite, la deuxième étape qui consistait à effectuer les différentes analyses *in silico* afin de prédire les structures, les domaines catalytiques et les sites de liaisons a permis de prendre des décisions éclairées pour la conception des cassettes d'expression et des stratégies à employer lors de la production des protéines recombinantes. La troisième étape consistait justement au clonage, à la transformation et à l'expression des protéines recombinantes. Cette étape a permis d'obtenir des colonies positives, de lancer des cultures de levures transformées et d'induire la production de protéines recombinantes. Nous avons même pu définir la meilleure durée d'induction et détecter les protéines d'intérêt par immunobuvardage de type Western blot. Nous ne sommes pas parvenus à finaliser la dernière étape qui consistait à caractériser les protéines recombinantes. D'autres tentatives devront être réalisées pour arriver à trouver les bonnes conditions pour les essais enzymatiques, à détecter un produit de réaction catalytique et à caractériser les protéines d'intérêt en testant différentes conditions de réactions et différents substrats.

D'un point de vue général, ce projet de recherche a toutefois permis d'enrichir nos compétences en lien avec le traitement de données brutes provenant du séquençage de l'ARN et l'assemblage de ces données en un transcriptome. Ce projet a également permis d'apporter une nouvelle expertise dans notre laboratoire en ce qui touche à la manipulation de levures comme système d'expression hétérologue. Nous avons testé, adopté et optimisé

des protocoles en lien avec la croissance, l'induction et l'extraction de protéines provenant de levures transformées.

Bref, nous considérons que cette programmation de recherche a tout de même permis de contribuer au début de l'élucidation des voies métaboliques menant à la biosynthèse des molécules analogues aux cannabinoïdes dans des espèces différentes de *C. sativa*. Nous avons proposé un début de piste quant aux enzymes possiblement impliquées au niveau de la prénylation des analogues chez *R. marginata* et proposé des outils afin de développer une méthode permettant l'expression de protéines recombinantes en se servant de levures comme milieux de production.

ANNEXE A

SÉQUENCES DES CASSETTES D'EXPRESSION UTILISÉES DANS LE CADRE DE CE PROJET

Cassette RmPT130

Séquence consensus levure

Codon d'initiation

TACAAAATGTCTATGGGAAGCCTTCAAAGGTCGGTCCCAACCCAGCTTTGCTCAAAGCGGCTTC
TGCTCATTCCTTCTGCGGAGCTCCACCCCGCGCTCGGCTTCTCACTTATGCGGCAGTTCCACG
ACCGGTCTCAATAGAAGGGTATCACATCGAAGGGTGTGGCATAAGACGCGGAAGTCTCGATGGT
TTGTGCCCGATTCCGAGTGGTCTGAAGTTGGGAAGCCGAGATGGGGAGATCTCAGGACAGTCGG
TGGTCAACAAGCGCAGCTTACCAATTCGTCGAGGGGGAGTCGCGACTACTTTTCGCGACTGCTAG
GCCGATTGATGACGATGACGATGACGCTGCCTCTGTTCCGGAAGGAGCAAAAACCTTGGTGCGC
GACGGGGTCGATTGGTTGGACGCTCTTCATAGATTTTCTCGGCCGCACACCGTGATAGGATCTG
CTATGAGTATCGTTTCAGTTTCGCTCCTTGCTCTTCAAAGTCCGGCCGAGCTTTTCGTCGGCGTT
TTTCGGAGGGCTTGTCCAGGCTCTCATACCATCCTTATTCATGAACATCTACATCGTGGGATTG
AACCAACTCTACGACGTGGACATTGACAAGGTCAATAAGCCGTACCTTCCACTTGCCTCCGGGG
AATTTTCTATGGCAACTGGTGTGCGCAATTGTGGCAATTTCCGCCGCTTTGAGTGTTGCAATCGG
ACTAATGGTGGGGTCAAGGCCACTGCTCTGGACACTTTTGGTCAGCCTTTTCTTGGGGACTGCG
TATTCTGTTGATGTTTCCTTTCCTGAGGTGGAAAAGGTCAGCCGTGGCTGCAGCCACCTGTATAC
TTGCTATCCGGGCTATAGTGGTCCAGCTTGGCTTCTACTGTACATGCAGGGTTTTGTATTTCGG
GAGAGCGGCAACTATTACCAGACAGCTTCTATTTGCAACAGGATTCATGTCCTTTTTTCTCCGTA

GTCATTGCTCTGGCAAAGGATCTGCCGGATGTGGAGGGTGACAAAGTGTACGGCATTTCGTTCTT
 TCAGTGTTCGACTAGGCCAAGAGAAGGTGTTTTGGTCATGTGTTTGGCTCATCCAGGCAGCATA
 TGCTGTAGCAATGTCAGTGGGGATAACGTCTCCCACATTGTTTCGGAAAGGTTGCGATGATTTTA
 GGCCATGCAAGCCTTTCGAGCTTACTGTGGATGAAGGCAAAGGCAACAACTTGAATATCAAGA
 CCGAGATCGTGTCCCTTCTACATGTTTATATGGAAGTTATTCTATGCAGAATATTTTCATCATAACC
 GTTTGTGAGGCTAGTACCCAGAGGATCTACAGAGACGTCCCAAGTGGCTCCTGCATAG

Cassette RmPT280

Séquence consensus levure

TACAAAATGTCATTAGGCTACGGCAACTCTGCGCATCGCCTCTGGCTGCTGCAGCCATGGCGGCG
 CCTCCGTGGTCGCGATTCCCACCGGTCTCCACGGGCACGCCTTCTCTTCTCACGCACGCCGTC
 AATCGCTCCGCTATGCCGCCATAGCTGCGACCGTGCCTCAGACACCTGGTCATCTCTGGGGCC
 GAGGCCTTTGAATCAGTGGATGTGCACGGGGGGTGCCAGTGGTATCGTCGGAGCTCCACGACTG
 GTCTTCGGTCGCGACGAAGCGGCTCTCGCACTCCCCGCTTTGTTGGCAGCGCCGGCGGCGAGCA
 ATCGAGCGACGATGCCGAGCCGCCGTCGGAATTTACGAGATCCTCCGAGCTGTCCGTCTCTGAA
 AGGTTTTTCTCGTTCAAGGAAGCGTTTTGGAGGTTTCTCAGGCCGCACACCATTTCGCGGAACGC
 TGCTCGGGTCCGTTGCACTCGTCTGCAGGGCGCTATTGGAGAATCCTCAGCGCGTTGATGCGGT
 TTTACTTCCGAGGGCCTTGTGGGGCATTATTGCTCTTCTCTGCGGAAATGGTTACATTGTGGGC
 ATCAACCAAATCTATGACGTTGAAATTGACAGAATCAACAAGCCGTTCTTGCCGATTGCGGCTG
 GAGACCTCTCCATTCCGCAGGCCTGGAGTTTGGTTTTAGGGCTAGCAGCACTCGGGATCACAAAT
 TGTCGGCACCAATTTGCGGCCTCTCATCACTGTGCTATACTTCTTGGGTCTCTTTCTGGGGACG
 ATTTACTCTGTTCCCTCCGTTGCACCTGAAGAGATTGCGCGTGCCGGCATTCTCATCATTGCCA

CTGTCAGGGGATTTCTCCTCAACTTCGGCGTGTATTACGCCACCAGGGCGGCACTCGGACTTCC
 GTTTGAATGGAACCCCCAATAGTCTTCATCACTTCTTTTGTACGGTTTTTGGCACTGTGATC
 GCCATCACCAAAGATCTCCCCGACACAGAAGGGGACCGTAAATTCAAATCTCCACCTTCGCCA
 CCAAGCTCGGAGTGCGCAACATTTCATTTTTAGGATCCGGCTTAATTCTTGCAAATTACGCCGC
 AGCCATCGTGGCGGCCATTTACTTGCCTCACTGCTTTAGGCAGAGAGTCATGATCCCTGCGCAC
 GCGATTCTGGCCACGGCTGTGGTGTACCAGACATGGCTGTTGGATTTCGGCCAAGTACACCAAGG
 ATAGCATAACCGGATTCTACCAGTTCATATGGAAACTCTTCTACTCAGAGTACACGTTGCTACC
 TTTCAATTCTAGTACCCAGAGGATCTACAGAGACGTCCCAAGTGGCTCCTGCATGA

Site clivage thrombine Épitope 1D4 Codon stop

Cassette CsPT1

Séquence consensus levure

Codon d'initiation

TACAAAATGTCATTATGGGACTCTCATCAGTTTGTACCTTTTCATTTCAAATAATTACCATACTT
 TATTAAATCCTCACAAATAATAATCCCAAACCTCATTATTATGTTATCGACACCCCCAAAACACC
 AATTAAATACTCTTACAATAATTTCCCTCTAAACATTGCTCCACCAAGAGTTTTTCATCTACAA
 AACAAATGCTCAGAATCATTATCAATCGCAAAAAATTCATTAGGGCAGCTACTACAAATCAA
 CTGAGCCTCCAGAATCTGATAATCATTAGTAGCAACTAAAATTTTAACTTTGGGAAGGCATG
 TTGGAAACTTCAAAGACCATATACAATCATAGCATTTACTTCATGCGCTTGTGGATTGTTTTGGG
 AAAGAGTTGTTGCATAACACAAATTTAATAAGTTGGTCTCTGATGTTCAAGGCATTCTTTTTTT
 TGGTGGCTGTATTATGCATTGCTTCTTTTACAACCTACCATCAATCAGATTTACGATCTTCACAT
 TGACAGAATAAACAAGCCTGATCTACCACTAGCTTCAGGGGAAATATCAGTAAACACAGCTTGG
 ATTATGAGCATAATTGTGGCACTGTTTGGATTGATAATAACTATAAAAATGAAGGGTGGACCAC
 TCTATATATTTGGCTACTGTTTTGGTATTTTTGGTGGGATTGTCTATTCTGTTCCACCATTTAG

ATGGAAGCAAATCCTTCCACTGCATTTCTTCTCAATTCCTGGCCATATTATTACAAATTC
ACATTTTATTATGCCAGCAGAGCAGCTCTTGGCTACCATTTGAGTTGAGGCCTTCTTTACTT
TCCTGCTAGCATTTATGAAATCAATGGGTTGAGCTTTGGCTTTAATCAAAGATGCTTCAGACGT
TGAAGGCGACACTAAAATTTGGCATATCAACCTTGGCAAGTAAATATGGTTCAGAAACTTGACA
TTATTTTGTCTGGAATTGTTCTCCTATCCTATGTGGCTGCTATACTTGCTGGGATTATCTGGC
CCCAGGCTTTCAACAGTAACGTAATGTTACTTTCTCATGCAATCTTAGCATTTTGGTTAATCCT
CCAGACTCGAGATTTTGCGTTAACAAATTACGACCCGGAAGCAGGCAGAAGATTTTACGAGTTC
ATGTGGAAGCTTTATTATGCTGAATATTTAGTATATGTTTTTCATACTAGTACCCAGAGGATCT
ACAGAGACGTCCCAAGTGGCTCCTGCATGA

Site clivage thrombine

Épitope 1D4

Codon stop

ANNEXE B

LISTE DES ABRÉVIATIONS ET DES NUMÉROS D'ACCESSION GENBANK POUR LES 11 PRÉNYLTRANSFÉRASES UTILISÉES LORS DE L'ALIGNEMENT MULTIPLE

Abréviation	Organisme	Numéro d'accession GenBank
CoPT	<i>Corchorus olitorius</i>	OMO68733
CsPT1	<i>Cannabis sativa</i>	DAC76711
CsPT4	<i>Cannabis sativa</i>	DAC76710
CsPT5	<i>Cannabis sativa</i>	DAC76714
HbPTlike	<i>Hypseocharis bilobate</i>	AKM76568
HIPT1	<i>Humulus lupulus</i>	E5RP65
HIPT2	<i>Humulus lupulus</i>	A0A0B4ZTQ2
MmPTlike	<i>Monsonia marlothii</i>	AKM76570
MvPTlike	<i>Melianthus villosus</i>	AKM76571
PaPT	<i>Parasponia andersonii</i>	PON44906
PEPTlike	<i>Pelargonium echinatum</i>	AKM76576

ANNEXE C

LISTE DES ABRÉVIATIONS ET DES NUMÉROS D'ACCESSION GENBANK DES PROTÉINES UTILISÉES POUR L'ANALYSE PHYLOGÉNÉTIQUE

Abréviation	Organisme	Numéro d'accension GenBank
CsPT1	<i>Cannabis sativa</i>	PK28436
CsPT2	<i>Cannabis sativa</i>	PK02092
CsPT3	<i>Cannabis sativa</i>	PK17697
CsPT4	<i>Cannabis sativa</i>	PK15523
CsPT5	<i>Cannabis sativa</i>	PK11068
CsPT6	<i>Cannabis sativa</i>	PK13891
CsPT7	<i>Cannabis sativa</i>	PK29226
CsPT8	<i>Cannabis sativa</i>	PK07278
HIPT1	<i>Humulus lupulus</i>	AB543053
HIPT2	<i>Humulus lupulus</i>	KM222442
AtVTE2-1	<i>Arabidopsis thaliana</i>	AAM10489
AtVTE2-2	<i>Arabidopsis thaliana</i>	ABB70127
GmVTE2-1	<i>Glycine max</i>	ABB70126
GmVTE2-2	<i>Glycine max</i>	ABB70128
GmG4DT	<i>Glycine max</i>	BAH22520
TaVTE2-1	<i>Triticum aestivum</i>	ABB70123
TaHGGT	<i>Triticum aestivum</i>	AAP43912
ApVTE2-1	<i>Allium porrum</i>	ABB70124
CpVTE2-1	<i>Cuphea pulcherrima</i>	ABB70125
OsVTE2-2	<i>Oryza sativa</i>	XP_015646905
OsHGGT	<i>Oryza sativa</i>	AAP43913
HvHGGT	<i>Hordeum vulgare</i>	AAP43911
ZmVTE2-1	<i>Zea mays</i>	ABB70122
GuA6DT	<i>Glycyrrhiza uralensis</i>	KJ123716
LaPT1	<i>Lupinus albus</i>	AER35706
SfILD	<i>Sophora flavescens</i>	BAK52290
SfG6DT	<i>Sophora flavescens</i>	BAK52291
SfN8DT-1	<i>Sophora flavescens</i>	BAG12671
AhR4DT-1	<i>Arachis hypogaea</i>	AQM74172
AhR3'DT-1	<i>Arachis hypogaea</i>	AQM74173
AhR3'DT-2	<i>Arachis hypogaea</i>	AQM74174

Abréviation	Organisme	Numéro d'accession GenBank
AhR3'DT-4	<i>Arachis hypogaea</i>	AQM74176
CIPT1	<i>Citrus limon</i>	BAP27988
PcPT	<i>Petroselinum crispum</i>	BAO31627
PsPT1	<i>Pastinca sativa</i>	AJW31563
PsPT2	<i>Pastinca sativa</i>	AJW31564
MaIDT	<i>Morus alba</i>	AJD80982
CtIDT	<i>Cudrania tricuspidata</i>	AJD80983

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] J. Holland, *The pot book: A complete guide to cannabis*, Simon and Schuster, 2010.
- [2] S.A. Bonini, M. Premoli, S. Tambaro, A. Kumar, G. Maccarinelli, M. Memo, A. Mastinu, *Cannabis sativa: A comprehensive ethnopharmacological review of a medicinal plant with a long history*, *Journal of ethnopharmacology* 227 (2018) 300-315.
- [3] E.B. Russo, H.-E. Jiang, X. Li, A. Sutton, A. Carboni, F. Del Bianco, G. Mandolino, D.J. Potter, Y.-X. Zhao, S. Bera, *Phytochemical and genetic analyses of ancient cannabis from Central Asia*, *Journal of experimental botany* 59 (2008) 4171-4182.
- [4] A. Rai, K. Saito, M. Yamazaki, *Integrated omics analysis of specialized metabolism in medicinal plants*, Wiley Online Library, 2017, pp. 764-787.
- [5] M. Ben Amar, *Le cannabis: une substance controversée*, *Drogues, santé et société* 2 (2004).
- [6] G.d. Canada, *Règlement sur le chanvre industriel (DORS/2018-145)*, in: S.W.d.l.l. (Justice) (Ed.), 2018.
- [7] E.d. de Meijer, H. Van der Kamp, F. Van Eeuwijk, *Characterisation of Cannabis accessions with regard to cannabinoid content in relation to other plant characters*, *Euphytica* 62 (1992) 187-200.
- [8] L.S.q.d.c. (SQDC), *Au-delà des catégories de cannabis : Indica, sativa, hybride*, <https://www.sqdc.ca/fr-CA/connaitre-le-cannabis/Conseils-pratiques/indica-sativa-hybride%E2%80%AF-au-dela-des-categories-de-cannabis>, 2021 (accessed 10/01/2022).
- [9] J. Tattrie, *Légalisation du cannabis au Canada*, *L'Encyclopédie canadienne*, 2016.
- [10] O. Aizpurua-Olaizola, U. Soydaner, E. Öztürk, D. Schibano, Y. Simsir, P. Navarro, N. Etxebarria, A. Usobiaga, *Evolution of the cannabinoid and terpene content during the growth of Cannabis sativa plants from different chemotypes*, *Journal of natural products* 79 (2016) 324-331.
- [11] Â. Carvalho, E.H. Hansen, O. Kayser, S. Carlsen, F. Stehle, *Designing microorganisms for heterologous biosynthesis of cannabinoids*, *FEMS yeast research* 17 (2017).

- [12] S.J. Gagne, J.M. Stout, E. Liu, Z. Boubakir, S.M. Clark, J.E. Page, Identification of olivetolic acid cyclase from *Cannabis sativa* reveals a unique catalytic route to plant polyketides, *Proceedings of the National Academy of Sciences* 109 (2012) 12811-12816.
- [13] F. Degenhardt, F. Stehle, O. Kayser, The biosynthesis of cannabinoids, *Handbook of cannabis and related pathologies*, Elsevier 2017, pp. 13-23.
- [14] K.A. Rea, J.A. Casaretto, M.S. Al-Abdul-Wahid, A. Sukumaran, J. Geddes-McAlister, S.J. Rothstein, T.A. Akhtar, Biosynthesis of cannflavins A and B from *Cannabis sativa* L, *Phytochemistry* 164 (2019) 162-171.
- [15] C.T. Hammond, P.G. Mahlberg, Morphogenesis of capitate glandular hairs of *Cannabis sativa* (Cannabaceae), *American journal of botany* 64 (1977) 1023-1031.
- [16] S.J. Livingston, T.D. Quilichini, J.K. Booth, D.C. Wong, K.H. Rensing, J. Laflamme-Yonkman, S.D. Castellarin, J. Bohlmann, J.E. Page, A.L. Samuels, *Cannabis* glandular trichomes alter morphology and metabolite content during flower maturation, *The Plant Journal* 101 (2020) 37-56.
- [17] S. Sirikantaramas, F. Taura, Y. Tanaka, Y. Ishikawa, S. Morimoto, Y. Shoyama, Tetrahydrocannabinolic acid synthase, the enzyme controlling marijuana psychoactivity, is secreted into the storage cavity of the glandular trichomes, *Plant and Cell Physiology* 46 (2005) 1578-1582.
- [18] D. Crispim Massuela, J. Hartung, S. Munz, F. Erpenbach, S. Graeff-Hönniger, Impact of Harvest Time and Pruning Technique on Total CBD Concentration and Yield of Medicinal Cannabis, *Plants* 11 (2022) 140.
- [19] M. Bagheri, H. Mansouri, Effect of induced polyploidy on some biochemical parameters in *Cannabis sativa* L, *Applied biochemistry and biotechnology* 175 (2015) 2366-2375.
- [20] S. Suzuki, T. Koeduka, A. Sugiyama, K. Yazaki, T. Umezawa, Microbial production of plant specialized metabolites, *Plant Biotechnology* 31 (2014) 465-482.
- [21] P. Avenas, *La prodigieuse histoire du nom des éléments*, EDP Sciences, [Les Ullis], 2018.
- [22] A. Ludwiczuk, K. Skalicka-Woźniak, M. Georgiev, *Terpenoids*, Pharmacognosy, Elsevier 2017, pp. 233-266.
- [23] G. Ourisson, *Terpènes*, *Encyclopædia Universalis*.
- [24] J.K. Booth, J. Bohlmann, Terpenes in *Cannabis sativa*—From plant genome to humans, *Plant Science* 284 (2019) 67-72.

- [25] E.B. Russo, Taming THC: potential cannabis synergy and phytocannabinoid-terpenoid entourage effects, *British journal of pharmacology* 163 (2011) 1344-1364.
- [26] J.S. Abrams, W.J. Ellyson, V.J. Gomez, J.L. Talleyrand, *Aroma in Cannabis: A Foundation for Chemotype Classification, Recent Advances in the Science of Cannabis*, CRC Press 2021, pp. 31-48.
- [27] A.M. Tomko, E.G. Whynot, L.D. Ellis, D.J. Dupré, Anti-cancer potential of cannabinoids, terpenes, and flavonoids present in cannabis, *Cancers* 12 (2020) 1985.
- [28] I.W. Oswald, M.A. Ojeda, R.J. Pobanz, K.A. Koby, A.J. Buchanan, J. Del Rosso, M.A. Guzman, T.J. Martin, Identification of a New Family of Prenylated Volatile Sulfur Compounds in Cannabis Revealed by Comprehensive Two-Dimensional Gas Chromatography, *ACS omega* 6 (2021) 31667-31676.
- [29] Produit intérieur brut (PIB) aux prix de base, par industries, moyenne annuelle, niveau d'industrie la plus détaillée (x 1 000 000), *Statistique Canada*, 2021.
- [30] G.d. Canada, Cultivateurs, transformateurs et vendeurs de cannabis autorisés en vertu de la Loi sur le cannabis, 2022.
- [31] L.S.q.d. cannabis, *Rapport annuel 2021*, 2021, pp. 110.
- [32] G.d. Canada, *Survol : Données sur le marché du cannabis*, 2021.
- [33] C.B. Times, *State of the cannabis cultivation industry*, 2020.
- [34] E. Mills, The carbon footprint of indoor Cannabis production, *Energy Policy* 46 (2012) 58-67.
- [35] H.M. Summers, E. Sproul, J.C. Quinn, The greenhouse gas emissions of indoor cannabis production in the United States, *Nature Sustainability* (2021) 1-7.
- [36] A. Singh, I. Desgagné-Penix, Biosynthesis of the Amaryllidaceae alkaloids, *Plant Science Today* 1 (2014) 114-120.
- [37] A. Acosta, J. Almirall, Differentiation between hemp-type and marijuana-type cannabis using the Fast Blue BB colorimetric test, *Forensic Chemistry* 26 (2021) 100376.
- [38] A. Janatová, A. Fraňková, P. Tlustoš, K. Hamouz, M. Božik, P. Klouček, Yield and cannabinoid contents in different cannabis (*Cannabis sativa* L.) genotypes for medical use, *Industrial crops and products* 112 (2018) 363-367.
- [39] T. Hussain, B. Plunkett, M. Ejaz, R.V. Espley, O. Kayser, Identification of putative precursor genes for the biosynthesis of cannabinoid-like compound in *Radula marginata*, *Frontiers in plant science* 9 (2018) 537.

- [40] M. Styrzewska, A. Kulma, K. Ratajczak, R. Amarowicz, J. Szopa, Cannabinoid-like anti-inflammatory compounds from flax fiber, *Cellular & molecular biology letters* 17 (2012) 479-499.
- [41] T. Hussain, R.V. Espley, J. Gertsch, T. Whare, F. Stehle, O. Kayser, Demystifying the liverwort *Radula marginata*, a critical review on its taxonomy, genetics, cannabinoid phytochemistry and pharmacology, *Phytochemistry Reviews* 18 (2019) 953-965.
- [42] Y. Asakawa, A. Ludwiczuk, Chemical constituents of bryophytes: structures and biological activity, *Journal of natural products* 81 (2017) 641-660.
- [43] X. He, Y. Sun, R.-L. Zhu, The oil bodies of liverworts: unique and important organelles in land plants, *Critical reviews in plant sciences* 32 (2013) 293-302.
- [44] J. Faubert, *Flore des bryophytes du Québec-Labrador*, Société québécoise de bryologie Saint-Valérien, QU, CA2012.
- [45] M. Toyota, T. Kinugawa, Y. Asakawa, Bibenzyl cannabinoid and bisbibenzyl derivative from the liverwort *Radula perrottetii*, *Phytochemistry* 37 (1994) 859-862.
- [46] A. Chicca, M.A. Schafroth, I. Reynoso-Moreno, R. Erni, V. Petrucci, E.M. Carreira, J. Gertsch, Uncovering the psychoactivity of a cannabinoid from liverworts associated with a legal high, *Science advances* 4 (2018) eaat2166.
- [47] M. Toyota, T. Shimamura, H. Ishii, M. Renner, J. Braggins, Y. Asakawa, New bibenzyl cannabinoid from the New Zealand liverwort *Radula marginata*, *Chemical and pharmaceutical bulletin* 50 (2002) 1390-1392.
- [48] F. Cullmann, H. Becker, Prenylated bibenzyls from the liverwort *Radula laxiramea*, *Zeitschrift für Naturforschung C* 54 (1999) 147-150.
- [49] E. Hodgson, *New Zealand Hepaticae (Liverworts).—IV. A review of the New Zealand species of the genus Radula*, *Transactions of the Royal Society of New Zealand*, 1944, pp. 273-287.
- [50] J.E. Page, Z. Boubakir, Aromatic prenyltransferase from *Cannabis*, Google Patents, 2014.
- [51] M.B. Austin, M.E. Bowman, J.-L. Ferrer, J. Schröder, J.P. Noel, An aldol switch discovered in stilbene synthases mediates cyclization specificity of type III polyketide synthases, *Chemistry & biology* 11 (2004) 1179-1194.
- [52] M.G. Grabherr, B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Full-length transcriptome assembly from RNA-Seq data without a reference genome, *Nature biotechnology* 29 (2011) 644-652.

- [53] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, *Journal of molecular biology* 215 (1990) 403-410.
- [54] E.M. Gertz, Y.-K. Yu, R. Agarwala, A.A. Schäffer, S.F. Altschul, Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST, *BMC biology* 4 (2006) 1-14.
- [55] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic acids research* 22 (1994) 4673-4680.
- [56] J. Castresana, Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis, *Molecular biology and evolution* 17 (2000) 540-552.
- [57] N. Saitou, M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees, *Molecular biology and evolution* 4 (1987) 406-425.
- [58] J. Felsenstein, Confidence-Limits on Phylogenies—an approach using the bootstrap. *evolution* 39: 783-791. doi: 10.2307/2408678 View Article PubMed, NCBI Google Scholar (1985).
- [59] Invitrogen, pYES2/CT, pYES3/CT, and pYC2/CT – Yeast expression vectors with C-terminal tags and auxotrophic selection markers, 2009.
- [60] C.L. Young, Z.T. Britton, A.S. Robinson, Recombinant protein expression and purification: a comprehensive review of affinity tags and microbial applications, *Biotechnology journal* 7 (2012) 620-634.
- [61] A.M. Seddon, P. Curnow, P.J. Booth, Membrane proteins, lipids and detergents: not just a soap opera, *Biochimica et Biophysica Acta (BBA)-Biomembranes* 1666 (2004) 105-117.
- [62] D.G. Gibson, L. Young, R.-Y. Chuang, J.C. Venter, C.A. Hutchison, H.O. Smith, Enzymatic assembly of DNA molecules up to several hundred kilobases, *Nature methods* 6 (2009) 343-345.
- [63] NEB, Gibson Assembly®, <https://www.neb.com/applications/cloning-and-synthetic-biology/dna-assembly-and-cloning/gibson-assembly> (accessed 2022-02-08).
- [64] J.-B.F. Jacques Boidin, Simone Poncet, Levures, *Encyclopædia Universalis*.
- [65] P.N. Lipke, R. Ovalle, Cell wall architecture in yeast: new structure and new challenges, *Journal of bacteriology* 180 (1998) 3735-3740.

- [66] A. Bzducha-Wróbel, S. Błażejczak, K. Tkacz, Cell wall structure of selected yeast species as a factor of magnesium binding ability, *European Food Research and Technology* 235 (2012) 355-366.
- [67] F.M. Klis, P. Mol, K. Hellingwerf, S. Brul, Dynamics of cell wall structure in *Saccharomyces cerevisiae*, *FEMS microbiology reviews* 26 (2002) 239-256.
- [68] B. Aguilar-Uscanga, J. Francois, A study of the yeast cell wall composition and structure in response to growth conditions and mode of cultivation, *Letters in applied microbiology* 37 (2003) 268-274.
- [69] G.H. Fleet, Composition and structure of yeast cell walls, *Current topics in medical mycology* (1985) 24-56.
- [70] M. Anwar, F. Muhammad, M. Awais, M. Akhtar, A review of β -glucans as a growth promoter and antibiotic alternative against enteric pathogens in poultry, *World's Poultry Science Journal* 73 (2017) 651-661.
- [71] H. Huang, E.J. Levin, S. Liu, Y. Bai, S.W. Lockless, M. Zhou, Structure of a membrane-embedded prenyltransferase homologous to UBIAD1, *PLoS biology* 12 (2014) e1001911.
- [72] S. Kawai, W. Hashimoto, K. Murata, Transformation of *Saccharomyces cerevisiae* and other fungi: methods and possible underlying mechanism, *Bioengineered bugs* 1 (2010) 395-403.
- [73] B. Zirpel, F. Degenhardt, C. Martin, O. Kayser, F. Stehle, Engineering yeasts as platform organisms for cannabinoid biosynthesis, *Journal of biotechnology* 259 (2017) 204-212.
- [74] T. Zhang, J. Lei, H. Yang, K. Xu, R. Wang, Z. Zhang, An improved method for whole protein extraction from yeast *Saccharomyces cerevisiae*, *Yeast* 28 (2011) 795-798.
- [75] G.T. James, Inactivation of the protease inhibitor phenylmethylsulfonyl fluoride in buffers, *Analytical biochemistry* 86 (1978) 574-579.
- [76] L.L. Molday, R.S. Molday, 1D4: a versatile epitope tag for the purification and characterization of expressed membrane and soluble proteins, *Protein Affinity Tags*, Springer 2014, pp. 1-15.
- [77] T. Kameyama, H. Suzuki, A. Mayeda, Re-splicing of mature mRNA in cancer cells promotes activation of distant weak alternative splice sites, *Nucleic acids research* 40 (2012) 7896-7906.

- [78] R. Munakata, T. Inoue, T. Koeduka, F. Karamat, A. Olry, A. Sugiyama, K. Takanashi, A. Dugrand, Y. Froelicher, R. Tanaka, Molecular cloning and characterization of a geranyl diphosphate-specific aromatic prenyltransferase from lemon, *Plant Physiology* 166 (2014) 80-90.
- [79] Z. Zhong, W. Zhu, S. Liu, Q. Guan, X. Chen, W. Huang, T. Wang, B. Yang, J. Tian, Molecular characterization of a geranyl diphosphate-specific prenyltransferase catalyzing stilbenoid prenylation from *Morus alba*, *Plant and Cell Physiology* 59 (2018) 2214-2227.
- [80] X. Luo, M.A. Reiter, L. d'Espaux, J. Wong, C.M. Denby, A. Lechner, Y. Zhang, A.T. Grzybowski, S. Harth, W. Lin, Complete biosynthesis of cannabinoids and their unnatural analogues in yeast, *Nature* 567 (2019) 123-126.
- [81] S. Qian, J.M. Clomburg, R. Gonzalez, Engineering *Escherichia coli* as a platform for the in vivo synthesis of prenylated aromatics, *Biotechnology and Bioengineering* 116 (2019) 1116-1127.
- [82] N. Bai, G.-H. Li, S.-L. Luo, L. Du, Q.-Y. Hu, H.-K. Xu, K.-Q. Zhang, P.-J. Zhao, Vib-PT, an aromatic prenyltransferase involved in the biosynthesis of vibrilactone from *Stereum vibrans*, *Applied and environmental microbiology* 86 (2020) e02687-02619.
- [83] H. He, G. Bian, C.J. Herbst-Gervasoni, T. Mori, S.A. Shinsky, A. Hou, X. Mu, M. Huang, S. Cheng, Z. Deng, Discovery of the cryptic function of terpene cyclases as aromatic prenyltransferases, *Nature communications* 11 (2020) 1-13.
- [84] J. Yang, T. Zhou, Y. Jiang, B. Yang, Substrate specificity change of a flavonoid prenyltransferase AhPT1 induced by metal ion, *International Journal of Biological Macromolecules* 153 (2020) 264-275.
- [85] W. Ni, Z. Zheng, H. Liu, P. Wang, L. Wang, H. Wang, X. Sun, Q. Yang, H. Tang, G. Zhao, Synthesis of the carboxymethyl cellulose magnetic nanoparticles for efficient immobilization of prenyltransferase NovQ, *Carbohydrate Polymers* 235 (2020) 115955.
- [86] K. Blatt-Janmaat, Y. Qu, The Biochemistry of Phytocannabinoids and Metabolic Engineering of Their Production in Heterologous Systems, *International Journal of Molecular Sciences* 22 (2021) 2454.
- [87] T. Gulck, J. Booth, Â. Carvalho, B. Khakimov, C. Crocoll, M. Motawia, B. Møller, J. Bohlmann, N. Gallage, Synthetic Biology of Cannabinoids and Cannabinoid Glucosides in *Nicotiana benthamiana* and *Saccharomyces cerevisiae*, *Journal of Natural Products* 83 (2020) 2877-2893.