

THESE DE DOCTORAT DE

L'UNIVERSITE DE NANTES

ECOLE DOCTORALE N° 598

Sciences de la Mer et du littoral

Spécialité : Chimie des substances naturelles marines

Par

Lucie ORY

Du bioguidage à la métabolomique pour l'identification de métabolites fongiques actifs sur des modèles cellulaires du cancer du sein – Etude de l'activité de l'ergostérol.

Thèse présentée et soutenue à l'Université de Nantes, le 28 octobre 2019

Unité de recherche : Laboratoire Mer, Molécules, Santé MMS – EA2160

Rapporteurs avant soutenance :

Anne MURH-TAILLEUX
Joël BOUSTIE

Professeur - Université de Lille
Professeur - Université de Rennes

Composition du Jury :

Président :
Jean-Marie BARD

Professeur- Université de Nantes

Examineurs :
Pierre-Marie ALLARD

Assistant Recherche - Ecole de Pharmacie Genève-Lausanne

Dir. de thèse :
El-Hassan NAZIH

Maître de conférences HDR - Université de Nantes

Co-encadrante :
Catherine ROULIER

Maître de conférences - Université de Nantes

Remerciements

Je tiens tout d'abord à remercier les membres du jury qui me font l'honneur de juger mon travail de thèse malgré leurs obligations professionnelles. Merci aux rapporteurs, le Pr. Anne Muhr-Tailleux et le Pr. Joël Boustie, pour avoir pris le temps d'examiner ce manuscrit. J'aimerais également remercier les autres membres du jury, le Pr. Jean-Marie Bard, le Docteur Allard Pierre-Marie d'avoir accepté d'être mes examinateurs.

J'adresse de chaleureux remerciements à mon directeur de thèse Hassan Nazih, et ma co-encadrante de thèse Catherine Roullier pour la confiance qu'ils m'ont accordée en me permettant de réaliser ce projet. Grâce à vous, et avec vous, j'ai pu vivre cette aventure qu'est « la thèse ». Votre encadrement scientifique, nos réunions quasi mensuelles, votre grande disponibilité m'ont permis de mener à bien ce projet.

Hassan, je tiens à te remercier pour ta patience, ta bienveillance, ta gentillesse et ton optimisme à toute épreuve. Tu as toujours su me soutenir et prendre le temps de répondre à mes interrogations pour me guider dans la réalisation de ce projet. Je n'oublierai jamais tes petits schémas explicatifs griffonnés sur le coin d'une feuille pour m'aider à comprendre les différentes interactions entre les voies de signalisations.

Catherine, je tiens à te remercier pour ta disponibilité, ton écoute, ta joie de vivre, ton optimisme, ton aide et ton soutien du premier au dernier jour de cette thèse. Nous avons effectivement passé un certain nombre d'heures ensemble pour échanger sur le projet, travailler sur l'analyse des données, faire des corrections de publi, j'en passe, mais cela s'est toujours fait dans la bonne humeur. Même après avoir corrigé plus de 20 fois « métabolites bioactives » en « métabolites bioactifs », tu as toujours été très patiente et extrêmement bienveillante. Ton savoir-faire, ta passion pour la recherche, ta pédagogie, ta patience et surtout ta persévérance ont été très bénéfiques dans mon apprentissage du métier d'enseignant chercheur. Au-delà de toute l'aide professionnelle que tu as pu m'apporter, tu possèdes de grandes qualités humaines qui m'ont beaucoup aidée au cours de ces années. Bien que tu aies pu percevoir l'intégralité de ma palette émotionnelle, sans aucun jugement tu m'as toujours soutenue et montré qu'il ne fallait jamais rien lâcher.

J'adresse également mes remerciements à M. Yves François Pouchus pour m'avoir accueillie au sein du laboratoire MMS, et notamment pour vos remarques et conseils avisés lors des réunions d'équipes.

Je remercie également le Pr. Pascal Richomme et le Pr. Khadija Ouguerram, membres de mon comité de suivi de thèse, pour vos conseils scientifiques très précieux au cours du suivi de mon travail.

Je souhaiterais remercier Audrey Aguesse, Stéphanie Crossouard et Mikael Croyal de la plate-forme de Spectrométrie de masse à l'Institut de Recherche en Santé de l'Université de Nantes, pour leur accueil, leur aide et leur disponibilité qui nous ont permis de réaliser la quantification du cholestérol cellulaire par des analyses GC-MS.

Mes remerciements s'adressent également à Yan Guitton et Anne Lise Royer pour les analyses LC-MS effectuées au LABERCA.

Je tiens à remercier le personnel de l'équipe de l'INSERM UMR1235, pour leur aide et la mise en disposition de leur microscopie confocale.

Je remercie également le personnel de l'équipe de l'INSERM UMR1069, notamment Philippe Frank pour son aide dans la réalisation de certains des tests de viabilité cellulaire.

Concernant la réalisation des analyses spectroscopiques RMN, je remercie Arnaud Bondon et Sandrine Pottier de la plateforme PRISM de l'Université de Rennes 1 ainsi que Dimitri Breard et Pascal Richomme du laboratoire SONAS de l'université d'Angers et Aurore Michaud du laboratoire CEISAM. Je tiens également à remercier Mélanie Bourjot de l'UMR7178 de l'Université de Strasbourg pour les analyses spectroscopiques RMN effectuées dans le but de déterminer une empreinte pharmacophorique. Merci d'avoir consacré autant de votre temps à l'acquisition des données spectrales sur mes échantillons, souvent d'une très faible quantité.

Je tiens à remercier le Ministère de l'Education Nationale, le Cancéropôle Grand Ouest ainsi que la Ligue contre le cancer qui ont financé cette thèse.

Enfin merci à tous les membres du laboratoire MMS, et plus particulièrement à ceux de l'équipe de biochimie en médecine, ceux du troisième étage en pharma et ceux du rez-de-chaussée de ISOMER.

Delphine, merci pour ton enthousiasme et ton aide dans les expériences de cytométrie de flux. Jean-Michel et Alexandra, pour leur aide précieuse dans la réalisation de toutes ces manip dont celles de dernières minutes. Jean Michel, je n'oublierai pas les petits mots fléchés du 20 minutes au café. Merci également à Chloé, (ancienne technicienne) pour m'avoir appris les fondements de la culture cellulaire.

Trang et Josiane, pour ces bons moments passés ensemble dans le bureau et ces pauses déjeuner.

Merci à Chantal et Nathalie pour leur aide administrative (ma grande phobie) dans la gestion des OM, des nombreux dossiers dont ceux pour l'encadrement des Tp etc.

Merci à tous les membres de chichamva !, Je suis ravie d'avoir pu faire partie de votre équipe, j'ai passé de très très bons moments en votre compagnie pendant la pause-café ou les repas du midi. Ils étaient souvent très animés mais qu'est-ce que j'ai pu rigoler et décompresser ! Vous allez me manquer.....

Merci à toi Marie-Claude, pour toutes ces petites astuces que tu as pu me montrer pendant les manip. Tu as toujours été là pour répondre à mes questions et m'aider si besoin. Je te remercie énormément et « le labo de Marie-Claude » sans Marie-Claude ça va être étrange...

Merci à mes voisins de bureau, Nico pour ton petit côté parfois « ronchon » qui va me manquer, merci à Thibault pour ces petites phrases : « tu as une sale tête ! » du lundi matin, et le fabuleux surnom de fin de thèse que vous m'avez donné avec : « Miss Kebab », je ne sais plus vraiment qui en est l'auteur. Malgré ce côté taquin qui me faisait vraiment beaucoup rire, vous avez toujours été d'une extrême bienveillance avec moi et je vous en remercie. Merci pour votre aide pour la réalisation des cultures de champignons, et merci à toi Thibault pour l'entretien et le repiquage de cette collection.

Merci à Olivier et Karina pour ces discussions pendant le repas du midi, pour vos récits et petites anecdotes respectives, votre sens du partage ! Karina bien évidemment « girls Tours POWER ! », mais juste pour l'origine géographique ;).

Merci à Gaëtane, Aurélie et Vony, pour votre sympathie et notamment votre aide dans la quantification de l'ergostérol par GC-MS, nous nous sommes finalement que très peu vues au regard de ces quatre ans.

Merci Samuel pour toute l'aide que tu as pu m'apporter durant ma thèse d'un point de vue scientifique, notamment dans la rédaction du script sous R. En effet je suis venue plusieurs fois en toquant à la porte en disant « je ne comprends pas, je n'y arrive pas... » et voici la meilleure des réponses au monde : « Tu veux du chocolat ?? », Suisse en plus... !

Merci à Emmanuel, Samuel et Catherine pour l'utilisation de l'IT-TOF !!! Il n'a pas toujours été très mignon avec nous mais on a finalement réussi...Merci pour le temps passé à l'entretien et la préparation des analyses ! Un grand merci à Emmanuel pour ton savoir-faire !

Je tiens également à remercier Christine pour ta bienveillance, ton aide et les petites attentions à chaque fois que j'ai pu te croiser.

Mes remerciements s'adressent également à tous les stagiaires que j'ai pu encadrer et qui m'ont grandement aidé dans la réalisation de ce projet : Julia ma première stagiaire dont je n'oublierai pas l'humour et le dessin gardé précieusement, Eline, Axel ainsi que Bertrand et Cassilda pour leur aide précieuse dans la purification de molécule.

Enfin le BUREAU DES THESARDS, ou plutôt ma deuxième maison et pour certains le bureau des glands ;) ! Merci à vous tous anciens et nouveaux arrivants, sans vous cela serait juste impossible ! Qu'est-ce que j'ai pu discuter et partager des joies et des peines dans ce bureau avec tout d'abord Thuy, Anne-Isaline et Camilla « Ma crotte d'amour »... Je suis désolée d'avoir laissé le temps filler et de ne pas avoir donné plus de nouvelles et même pour certaines dans le pire des cas d'avoir pu être affreuse pendant la rédaction ! Et bien évidemment un grand merci aux doctorants actuels qui m'ont supporté : Tuyen, Simon, Aurore et ma petite Marjo ! Il est rare que l'on soit réuni au complet mais vous m'avez tous apporté chacun à votre façon un soutien sans faille !

Particulièrement toi ma petite Marjo, merci pour tout ce que tu as fait pour moi, au labo et à l'extérieur pour ma vie mouvementée et ses montagnes russes... Ne t'inquiètes pas Vannes-Nantes ce n'est pas si loin.

Et parce que qu'en dehors de la thèse, on peut avoir une vie sociale bien que fortement limitée pendant la rédaction. Je tiens à remercier ma famille et particulièrement mes parents pour leur aide morale et il faut bien le dire financière depuis le début des études supérieures. Je suis désolée de vous avoir en quelque sorte délaissés. Merci à tous mes amis d'enfance pour leur soutien, leurs visites dans chacune des villes où j'ai pu résider. Merci à mes amis de Nantes (et villes avoisinantes, ils se reconnaîtront) et d'ailleurs notamment toi ma petite mallorquine, vous avez été d'un soutien sans faille alors qu'on peut le dire maintenant, oui j'ai été particulièrement « chiante » ^^.

Pour finir merci à toutes les personnes que j'ai pu croiser pendant l'encadrement de travaux pratiques, les pauses café, etc, que j'aurai malheureusement pu oublier.

Avec tout mon amour, je tiens à te remercier toi « Marsou », pour ce soutien sans faille durant cette thèse. Merci d'avoir absolument tout géré dans notre quotidien pendant cette phase de rédaction, de m'avoir soutenue, épaulée et encouragée sans relâche quel que soit mon état. Merci pour toutes ces attentions. La thèse, c'est une page qui se tourne et une nouvelle vie peut commencer.

SOMMAIRE

Remerciements.....	I
Sommaire.....	V
Valorisation des travaux de thèse	IX
Liste des Abréviations.....	XI
Liste des Figures.....	XV
Liste des Tableaux.....	XIX

Préface.....	1
---------------------	----------

Introduction générale.....	3
-----------------------------------	----------

Chapitre 1 : Criblage et sélection des souches fongiques

1. Introduction bibliographique.....	11
1.1. Les champignons.....	11
1.1.1. Définition générale.....	11
1.1.2. Classification des champignons.....	11
1.2. Que signifie le terme « Champignons marins » ?.....	13
1.2.1. Définition.....	13
1.2.2. Distribution géographique et écologique en milieu marin	14
1.3. Le rôle potentiel des champignons dans la production de molécules bioactives dans le traitement du cancer du sein	20
1.3.1. Le cancer du sein ou plutôt les cancers du sein	20
1.3.2. Les métabolites fongiques d'intérêt dans le traitement du cancer du sein	21
2. Objectifs de l'étude.....	23
3. Résultats et discussion.....	24
3.1. Criblage : évaluation des extraits fongiques marins sur la prolifération des cellules MCF-7	24
3.1.1. Présentation des extraits testés.....	24
3.1.2. Résultats obtenus.....	25
3.2. Sélection des souches fongiques à étudier	28
3.2.1. Présentation et évaluation du potentiel de la souche MMS5 <i>Penicillium Chrysogenum</i>	29
3.2.1.1. Classification taxonomique et distribution de l'espèce <i>P. chrysogenum</i> ..	29
3.2.1.2. Molécules décrites chez <i>P. chrysogenum</i>	30
3.2.2. Présentation et évaluation du potentiel de la souche MMS719 <i>D. americana</i>	30
3.2.2.1. Classification taxonomique et distribution de l'espèce <i>D. americana</i>	30

3.2.2.2. Molécules décrites chez <i>D. americana</i>	31
4. Conclusion.....	32
5. Matériel et Méthodes.....	33
5.1. Description de l'extractothèque	33
5.1.1. Les milieux utilisés pour la culture des souches.....	33
5.1.2. Ensemencement et incubation des cultures fongiques	34
5.1.3. Extraction de la biomasse fongique	34
5.1.4. Profilages HPLC-UV-HRMS	35
5.2. Présentation de la cible biologique étudiée : prolifération des cellules cancéreuses mammaires	35
5.2.1. La lignée cancéreuse mammaire MCF-7	35
5.2.2. Évaluation de l'activité biologique par le test de cytotoxicité MTT	36
5.2.3. Évaluation de l'activation des récepteurs nucléaires LXR et/ou PPAR.....	36

Chapitre 2 : Investigation chimique des souches sélectionnées

1. Introduction bibliographique.....	39
1.1. Définition, intérêt et limites du fractionnement bioguidé.....	39
1.2. Que signifie le terme « Déréplication » ?.....	40
2. Objectifs de l'étude.....	42
3. Résultats discussions	42
3.1. Obtention des extraits fongiques et évaluation de leurs activités cytotoxiques	42
3.2. Étude de la souche <i>Penicillium chrysogenum</i>	43
3.2.1. Fractionnement bioguidé.....	44
3.2.1.1. Isolement et identification d'une molécule bioactive : l'ergostérol	47
3.2.1.2. Description et analyse critique du fractionnement bioguidé	47
3.2.1.3. Evaluation de l'activité biologique et quantification de l'ergostérol	49
3.2.2. Profilages métaboliques, identification des composés majoritaires et purification ciblée	53
3.2.2.1. Profilage métabolique	53
3.2.2.2. Purification ciblée	58
3.2.2.3. Purification ciblée des acides gras	66
3.2.3. Bilan et perspective de l'étude de la souche MMS5 <i>P. chrysogenum</i>	69
3.3. Étude de la souche <i>Didymella americana</i>	74
3.3.1. Fractionnement bioguidé de l'extrait MMS719-527.....	74
3.3.2. Fractionnement bioguidé de l'extrait MMS719-530.....	78

3.3.3.	Profilages métaboliques des composés majoritaires de l'extrait MMS719-527...	82
4.	Conclusion.....	86
5.	Matériel et méthodes	89
5.1.	Culture des deux souches sélectionnées au cours du screening.....	89
5.2.	Extraction de la biomasse fongique	89
5.3.	Évaluation et suivi de l'activité antiproliférative des cellules cancéreuses mammaires	90
5.3.1.	Test MTT.....	90
5.3.2.	Test du crystal violet	90
5.4.	Méthodes de séparation chromatographiques	90
5.4.1.	Chromatographie sur couche mince (CCM)	90
5.4.2.	Chromatographie liquide basse pression.....	90
5.4.2.1.	Fractionnements réalisés sur la souche MMS5 <i>P. chrysogenum</i>	91
5.4.2.2.	Fractionnement réalisé sur la souche MMS719 <i>D. americana</i>	93
5.4.3.	Chromatographie liquide haute performance (CLHP).....	94
5.4.4.	Profilage HPLC-UV-HRMS	96
5.4.5.	Déréplication.....	97
6.	Publication sur la détection de l'ergostérol en electrospray	98

Chapitre 3 : Développement d'un outil bioinformatique dans l'identification précoce de métabolites bioactifs au sein de mélanges complexes : RFiBiBo script

1.	Introduction bibliographique.....	111
1.1.	Qu'est-ce que la « biochimométrie » ?	111
1.2.	L'utilisation de cette approche dans le domaine des produits naturels	112
1.3.	L'acquisition des données	113
2.	Objectifs de l'étude.....	114
3.	Présentation de la publication	115
4.	Conclusion.....	131

Chapitre 4 : Etude de l'ergostérol sur des modèles de cellules du cancer du sein

1.	Introduction bibliographique.....	136
1.1.	Mécanismes impliqués dans la régulation du métabolisme lipidique	136
1.1.1.	Métabolisme des acides gras	136
1.1.1.1.	Synthèse des acides gras	137
1.1.1.2.	Stockage des acides gras.....	138

1.1.1.3.	Bêta-oxydation des acides gras	138
1.1.1.4.	Libération des acides gras à partir des triglycérides	139
1.1.2.	Métabolisme du cholestérol	139
1.1.2.1.	La synthèse endogène du cholestérol	139
1.1.2.2.	La captation du cholestérol via les récepteurs LDLR (Low Density Lipoprotein Receptor).....	141
1.1.2.3.	L'efflux du cholestérol	142
1.2.	Métabolisme lipidique sous le contrôle des facteurs de transcription	142
1.2.1.	Le facteur de transcription SREBP (Sterol Regulatory Element-Binding Protein)	143
1.2.2.	Les récepteurs nucléaires PPAR (Peroxisome Proliferator-Activated Receptor).	144
1.2.3.	Rôle des récepteurs nucléaires PPAR.....	145
1.2.4.	Les récepteurs nucléaires LXR (Liver X Receptor)	146
1.2.4.1.	Les ligands de LXR	147
1.2.4.2.	Le rôle des récepteurs nucléaires LXR	149
1.3.	Les marqueurs de la dérégulation du métabolisme lipidique dans le cancer du sein..	149
1.3.1.	Acides gras et cancer du sein	149
1.3.2.	Cholestérol et cancer du sein	150
2.	Objectifs de l'étude.....	154
3.	Présentation de la publication	155
4.	Résultats complémentaires sur l'effet antiprolifératif de dérivés de l'ergostérol sur nos modèles cancéreux mammaires.	183
5.	Conclusion.....	185
	Conclusions et perspectives	191
	Références Bibliographiques	200
	Annexes.....	213

Valorisation des travaux de thèse

[Articles dans des périodiques internationaux avec comité de lecture](#)

Ory. L, Nazih. H, Daoud. S, Mocquard. J, Bourjot. M, Margueritte. L, Desluc. M-A, Bard. J-M, Pouchus. Y F, Bertrand. S and Roullier. C. (2019) Targeting bioactive compounds in natural extracts - Development of a comprehensive workflow combining chemical and biological data. *Analytica chimica acta*, 1070,29-42.

[Communications orales dans les congrès nationaux](#)

Ory. L, Daoud. S, Franck. P, Huvelin. J-M, Carbonnelle. D, Chaillou. C, Guiterrez-Pajares. J-L, Bard. J-M, Roullier. C, Nazih. H. Régulation du transport lipidique cellulaire par les substances marines ligands de LXR (Liver X Receptor) et développement des tumeurs du sein. (2016) 10ème journée du Cancéropôle Grand Ouest (CGO), Sables d'Olonnes (15min).

Ory. L, Daoud. S, Franck. P, Huvelin. J-M, Carbonnelle. D, Chaillou. C, Guiterrez-Pajares. J-L, Bard. J-M, Roullier. C, Nazih. H. Régulation du transport lipidique cellulaire par les substances marines ligands de LXR (Liver X Receptor) et développement des tumeurs du sein. (2016) Journées de l'axe valorisation des produits de la mer, Nantes (10 min).

[Communications orales dans les congrès internationaux](#)

Ory. L, Nazih. H, Bard. JM, Pouchus. YF, Bertrand. S, Roullier. C. A combination of biochemometrics models to identify bioactives metabolites. (2018) 3^{ème} Symposium International de l'Association Francophone de l'Enseignement et de la Recherche en Pharmacognosie associée à STOLON (AFERP-STOLON), Rennes. **Prix de la meilleure communication orale 20 min**

Ory. L, Huvelin. J-M, Carbonnelle. D, Chaillou. C, Bard. J-M, Nazih. H, Roullier. C. A putative identification of bioactive metabolites against MCF-7 breast cancer cell line from marine derived fungi based on a bioinformatics model. (2017) 2^{ème} Symposium International AFERP STOLON, Angers. (Flash poster communication 3min).

[Communications par poster dans les congrès internationaux](#)

Ory. L, Huvelin. J-M, Carbonnelle. D, Chaillou. C, Bard. J-M, Nazih. H, Roullier. C. A bioassay-guided fractionation complemented by LC-HRMS data: A strategy to identify metabolites against breast cancer cells from marine derived fungi ?. (2017) MafNaP, Kiel-Allemagne.

[Communications par poster dans les congrès nationaux](#)

Ory. L, Huvelin. J-M, Carbonnelle. D, Chaillou. C, Bard. J-M, Nazih. H, Roullier. C. Un challenge dans la découverte de molécules naturelles bioactives : comparaison entre des approches classiques de fractionnement bioguidé et des approches métabolomiques. (2017) Journées Francophones de Nutrition, Nantes.

Ory. L, Huvelin. J-M, Carbonnelle. D, Chaillou. C, Bard. J-M, Nazih. H, Roullier. C. A putative identification of bioactive metabolites against MCF-7 breast cancer cell line from marine-derived fungi based on a bioinformatics model. (2017) Association Francophone pour l'Enseignement et la Recherche en Pharmacognosie, Angers.

Ory. L, Daoud. S, Franck. P, Huvelin. J-M, Carbonnelle. D, Chaillou. C, Guiterrez-Pajares. J-L, Bard. J-M, Roullier. C, Nazih. H. Ergosterol from marine-derived filamentous fungi: Effects on MCF-7 breast cancer cells. (2016) Lipids Nanotechnology and cancer, Tours.

Liste des abréviations

Les abréviations correspondant à des acronymes et termes en langue anglaise sont indiqués avec le symbole suivant : [EN].

22-OH	22-Hydroxy Cholestérol
27-OH	27-Hydroxy Cholestérol
ABC	ATP-Binding Casette transporter [EN]
ABCA1	ATP-Binding Casette sub-family A member 1 [EN]
ABCG1	ATP-Binding Casette sub-family G member 1 [EN]
ABCG5	ATP-Binding Casette sub-family G member 5 [EN]
ABCG8	ATP-Binding Casette sub-family G member 8 [EN]
ACADML	Acyl-CoA Dehydrogenase Medium Long chain [EN]
ACADVL	Acyl-CoA Dehydrogenase Very Long chain [EN]
ACAT	Acyl-CoA Cholesterol Acyl-Transferase [EN]
ACC	Acetyl-CoA Carboxylase [EN]
ACC	Acetyl-CoA Carboxylase [EN]
ACLY	ATP Citrate Lyase [EN]
ACN	Acétonitrile
AcOEt	Acétate d'éthyle
ACP	Analyse en Composantes Principales
ACS	Acyl-CoA Synthetase [EN]
ADN	Acide désoxyribonucléique
AG	Acide Gras
AGMI	Acide Gras monoinsaturés
APOE	Apolipoprotéine E
ARN	Acide RiboNucléique
BCA	Bicinchoninic Acid [EN]
BSA	Bovin Serum Albumin [EN]
CACT	Cartinitine acylcarnitine translocase [EN]
CCM	Chromatographie sur Couche Mince
CH₂Cl₂	Dichlorométhane

CI₅₀	Concentration inhibitrice à 50%
CIRC	Centre International de Recherche sur le Cancer
CPT1	Carnitine Palmitoyltransferase 1 [EN]
CPT2	Carnitine Palmitoyltransferase 2 [EN]
CYA	Czapek Yeast extract Agar [EN]
DAG	Diacylglycerol [EN]
DAM	Déacétyl-Mycoépoxydiène
DCA	Dextrose Caseine Agar [EN]
DMAPP	Diméthyl-allyl pyrophosphate [EN]
DMEM	Dulbecco's Modified Eagle Medium [EN]
DMSO	Diméthylsulfoxyde
ED	Eau Distillée
EDM	Eau De Mer
EDMS	Eau De Mer Synthétique
EGFR	Epidermal Growth Factor Receptor [EN]
ER	Estrogen Receptor [EN]
ESI	ElectroSpray Ionisation [EN]
EtOH	Ethanol
FAD	Flavine Adenine Dinucleotide [EN]
FASN ou FAS	Fatty Acid Synthase [EN]
FDA	Food Drug Administration [EN]
FPP	Farnesyl Pyrophosphate [EN]
GC-MS	Gas Chromatographic Mass Spectrometer [EN]
GPAT	Glycerol-3-Phosphate AcylTransférase [EN]
GPP	Geranyl-PyroPhosphate [EN]
HADHA	L-β-hydroxyacyl-CoA deshydrogenase [EN]
HDL	High Density Lipoprotein [EN]
HER2	Human Epidermal Growth Factor Receptor-2 [EN]
HMG-CoA	3-Hydroxy-3-Méthyl-Glutaryl-Coenzyme A
HMGCR	3-Hydroxy-3-Méthyl-Glutaryl-Coenzyme A Réductase

IDOL	Inducible Degradator Of LDLR [EN]
INCa	Institut National du Cancer [EN]
INSIG	INSulin-Induced Gene [EN]
IPP	Isopentényl-Pyrophosphate
IT-TOF	Ion Trap - Time Of Flight [EN] (= Trappe ionique – Temps de vol)
KMS	Kolhmeyer Medium Solid [EN]
LC-UV-MS	Liquid Chromatography -Ultra Violet-Mass Spectrometry [EN]
LD	Lipid Droplet [EN] (= gouttelette lipidique)
LDL	Low Density Lipoprotein [EN]
LDLR	Low Density Lipoprotein Receptor [EN]
LHS	Lipase HormoSensible [EN]
LXR	Liver X Receptor [EN]
LXRE	Liver X Receptor Element [EN]
MAG	Monoacylglycerol [EN]
MAGL	Monoacylglycérol Lipase [EN]
MAGT	Monoacylglycerol Transferase [EN]
MEA	Malt Extract Agar [EN]
MES	Mold Extract Sucrose [EN]
MTT	Bromure de 3-(4,5-diméthylthiazole-2-yl)-2,5-diphényl de tétrazolium
NADPH₂	Nicotinamide Adénine Dinucléotide
PAF	Paraformaldéhyde
PAP ou LIPIN	Phosphatidic Phosphorylase [EN]
PBS	Phosphate Buffered saline [EN]
PDA	Potato-Dextrose-Agar [EN]
PL	Phospholipides
PLS	Partial Least Square [EN]
PLSDA	Partial Least Square Discriminant Analysis [EN]
PNM	Produit Naturel Marin
PPAR	Peroxisome Proliferator-Activated Receptor [EN]
PR	Progesteron Receptor [EN]

RDX	hexahydro-1,3,5-trinitro-1,3,5-triazine
RMN	Résonance Magnétique Nucléaire
Rps6	Ribosomal protein s6 [EN]
RT-qPCR	Retention Time-quantitative polymerase chain reaction [EN]
RXR	Retinoid X Receptor [EN]
SCAP	SREPB Cleavage-Activating Protein [EN]
SCD	Steraroyl-CoA Desaturase [EN]
SDS	Sodium Dodecyl Sulfate [EN]
SR-BI	Scavenger Receptor class B type I [EN]
SREBP	Sterol Regulatory Element-Binding Protein [EN]
TG	Triglycérides
TGLA	Triglycéride Lipase Adipeuse
TNT	(2, 4, 6-trinitrotoluène)
UV	Ultra-Violet
VLDL	Very Low Density Lipoprotein [EN]
YES	Yest Extract Sucrose [EN]

Liste des figures

Figure 1. Méthodologie en vue de l'isolement de nouveaux produits d'origine naturelle bioactifs	3
Figure 2. Évolution du nombre d'articles référencés sur la période de janvier 1970 à mai 2019 dans la base de données "Pubmed" par recherche des mot clés : « marine-derived fungi » ou « marine fungi ».	9
Figure 3. Distribution des nouveaux produits marins (NMP) en fonction de leur habitat sur la période de janvier à décembre 2017	10
Figure 4. Classification phylogénique.....	12
Figure 5. Classification du règne <i>fongique</i> divisé en 10 sous-embranchements.....	12
Figure 6. Distribution des métabolites isolés de champignons marins en fonction de leurs habitats.....	14
Figure 7. Carte des chaleurs représentant la proportion totale des produits naturels d'origine marine découverts par phylum en fonction des régions géographiques entre 2008 et 2017.....	15
Figure 8. Nouveaux composés isolés de champignons marins (2014-2016) répartis en fonction de leurs classes chimiques (a) ou de leurs activités biologiques (b)	17
Figure 9. Structures des molécules possédant une activité antimicrobienne (1-3) ou une activité antivirale (4 et 5).	17
Figure 10. Structures des molécules possédant une activité anticancéreuse.	18
Figure 11. Molécules d'origine naturelle fongique (A) et leurs dérivés hémisynthétiques (B) présentant une activité biologique améliorée.....	19
Figure 12. Exemple de molécules d'origine fongique marine présentant une activité cytotoxique sur les cellules cancéreuses mammaires.	22
Figure 13. Répartition des espèces fongiques de la collection du laboratoire MMS en fonction de leur genre.....	23
Figure 14. Effet antiprolifératif des extraits issus de 23 souches sélectionnées pour le criblage sur les cellules MCF-7.....	25
Figure 15. Souches dont les extraits obtenus sur différents milieux de culture présentent un effet sur l'augmentation de l'expression de PPAR de cellules Cos-1 transfectées.....	27
Figure 16. Méthodologie du fractionnement bioguidé.	40
Figure 17. Photographies des cultures en Erlenmeyer des souches MMS5 et MMS719.. ..	42
Figure 18. Effet des extraits bruts obtenus pour la souche MMS5 <i>P. chrysogenum</i> et MMS719 <i>D. Americana</i> sur la prolifération des cellules MCF-7.	43
Figure 19. Cytotoxicité des 17 fractions (1-1→1-17) obtenues au cours du fractionnement de l'extrait brut MMS5-526 issu de la souche MMS5 <i>P. chrysogenum</i> sur la prolifération des cellules MCF-7.....	44
Figure 20. Arbre de fractionnement de l'extrait MMS 5-526 issu de la souche <i>P. chrysogenum</i>	46
Figure 21. Structure chimique de l'ergostérol.	47

Figure 22. Cytotoxicité des fractions obtenues au cours du fractionnement numéro 1, 2, 3 et 4 de l'extrait brut MMS5-526 issu de la souche MMS5 <i>P. chrysogenum</i> sur la prolifération des cellules MCF-7.	48
Figure 23. Profils LC-MS des fractions MMS5-526-1-3, MMS5-526-2-3, MMS5-526-3-2.....	49
Figure 24. Cytotoxicité de l'ergostérol sur les cellules MCF-7 à 24H, dans une gamme de concentration de 0,001-50 µM	50
Figure 25. Profils LC-MS de l'extrait brut MMS5-526 obtenus en mode positif (ESI+) et négatif (ESI-).	53
Figure 26. Profils chromatographiques obtenus par CCM pour les fractions issues de la première étape de fractionnement de l'extrait MMS5-526 (Lot 1) et de MMS5-531 (Lot 2), issus de culture de la souche MMS5 <i>P. chrysogenum</i>	59
Figure 27. Structure de la méléagrine	62
Figure 28. Structure de l'anicequol	62
Figure 29. Structure de l'andrastine A	65
Figure 30. Profils CCM des fractions issues de la purification de l'échantillon MMS5-531-1-8..	66
Figure 31. Effet cytotoxique de la fraction MMS5-531-1-8 sur les cellules MCF-7.	67
Figure 32. Arbre de fractionnement de l'extrait MMS 5-531 issu de la souche <i>P. chrysogenum</i>	68
Figure 33. Molécules proposées dans la base de donnée du DNP, après recherche de formule brute C ₂₈ H ₄₄ O ₄ pour les ascomycètes.....	71
Figure 34. Molécules proposées dans la base de donnée du DNP, après recherche de formule brute C ₁₆ H ₃₅ NO ₃ pour les ascomycètes.	72
Figure 35. Molécules proposées dans la base de donnée du DNP, après recherche de formule brute C ₂₈ H ₄₂ O ₂ pour les ascomycètes.....	72
Figure 36. Molécules proposées dans la base de donnée du DNP, après recherche de formule brute C ₂₈ H ₃₆ O ₈ pour les ascomycètes.....	73
Figure 37. Cytotoxicité des 17 fractions (1-1→1-17) obtenues au cours du fractionnement de l'extrait brut MMS719-527 issu de la souche MMS719 <i>D. americana</i> sur la prolifération des cellules MCF-7.....	74
Figure 38. Profils CCM des fractions issues du fractionnement de l'extrait MMS719-527, éluant hexane/ACOET (12 : 8, v/v) (gauche) et CH ₂ Cl ₂ /MeOH (85 : 15, v/v) (droite).....	75
Figure 39. Cytotoxicité des 23 (2-1→2-23) et 26 fractions (3-1→3-23) obtenues au cours du fractionnement de la fraction MMS719-527-14 sur la prolifération des cellules MCF-7.	76
Figure 40. Arbre de fractionnement de l'extrait MMS719-527 issu de la souche <i>D. Americana</i>	77
Figure 41. Structure du cyclo(4-hydroxyprolylphenylalanyl).....	78
Figure 42. Cytotoxicité des 19 fractions (1-1→1-19) obtenues au cours du fractionnement de l'extrait brut MMS719-527 issu de la souche MMS719 <i>D. americana</i> sur la prolifération des cellules MCF-7.....	78

Figure 43. CMM des 19 fractions obtenues après fractionnement de l'extrait brut MMS719-530 issu de la souche <i>D. americana</i> .).....	79
Figure 44. Cytotoxicité des 18 fractions obtenues au cours du fractionnement 2 de l'extrait brut MMS719-530 issu de la souche MMS719 <i>D.americana</i> sur la prolifération des cellules MCF-7.	80
Figure 45. Arbre de fractionnement de l'extrait MMS719-530 issu de la souche <i>D. americana</i> . Les numéros 1 et 2 correspondent aux différentes étapes de fractionnement.....	81
Figure 46. Profils LC-HRMS de l'extrait brut MMS719-527 obtenus en mode positif (ESI+) et négatif (ESI-).	82
Figure 47. Molécules annotées sur le profil LC-HRMS de l'extrait MMS719-527 en comparaison des formules brutes obtenues avec la base donnée du DNP.	83
Figure 48. Régulation du métabolisme des acides gras.	137
Figure 49. Voie de biosynthèse du cholestérol).....	140
Figure 50. Captation du cholestérol par les récepteurs de lipoprotéines de faible densité (LDL-R)	141
Figure 51. Schéma de l'efflux	142
Figure 52. Régulation de l'homéostasie lipidique via le complexe INSIGH/SCAP/SREBP	143
Figure 53. Ligands naturels et synthétiques de PPAR	144
Figure 54. Les Rôles majeurs des différents facteurs de transcriptions nucléaires PPARs (peroxisome proliferator-activated receptor).	146
Figure 55. Mécanisme d'action moléculaire de LXR.	147
Figure 56. Ligands naturels de LXR.....	148
Figure 57. Evaluation de l'activité antiproliférative des stérols et dérivés des stérols fongiques sur les cellules MCF-7.	183

Liste des tableaux

Tableau 1. Composition des différents milieux de cultures.....	34
Tableau 2. Conditions d'élution pour le profilage HPLC-UV-HRMS.....	35
Tableau 3. Quantification de l'ergostérol au sein d'extrait brut et dans la fraction insaponifiable par analyse GC-MS, et évaluation de leur activité sur la prolifération cellulaire des cellules MCF-7.....	52
Tableau 4. Déréplication des ions majoritaires détectés au cours des analyses LC-MS de l'extrait brut MMS5-526.....	54
Tableau 5. Molécules d'intérêts détectées, isolées et identifiées dans les fractions MMS5-531-1-18→MMS5-531-1-20.....	61
Tableau 6. Table de données RMN ¹ H (500 MHz) et ¹³ C (125 MHz) de l'anicequol de formule brute C ₃₀ H ₄₈ O ₆ solubilisé dans le CDCl ₃	63
Tableau 7. Table de données RMN ¹ H (500 MHz), ¹³ C (125MHz), ainsi que RMN2D HMBC et COSY de l'échantillon CTBG-01-593 de formule brute C ₃₆ H ₄₈ O ₇ solubilisé dans le CDCl ₃	64
Tableau 8. Déréplication des molécules détectées dans les fractions les plus pures pour la souche <i>P. chrysogenum</i>	70
Tableau 9. Déréplication des ions majoritaires détectés au cours des analyses LC-MS de l'extrait brut MMS719-527.....	84
Tableau 10. Conditions d'élution du fractionnement N°1 réalisé sur l'extrait MMS5-526.	91
Tableau 11. Conditions d'élution du fractionnement N° 2 réalisée sur la fraction MMS5-5261-3.....	91
Tableau 12. Conditions d'élution du fractionnement N°3 réalisée sur la fraction MMS5-526-2-2.....	92
Tableau 13. Conditions d'élution du fractionnement N°4 réalisée sur la fraction MMS5-526-3-3.....	92
Tableau 14. Conditions d'élution du fractionnement N°1 réalisé sur l'extrait brut MMS719-527.....	93
Tableau 15. Conditions d'élution du fractionnement N°1 réalisé sur l'extrait MMS719-531.	93
Tableau 16. Conditions d'élution pour les analyses LC-UV-HRMS.....	96

Préface

De manière ancestrale, l'Homme utilise des préparations médicinales issues principalement de plantes pour se soigner. Ainsi, les produits naturels occupent une place importante dans la médecine, ils constituent un vaste réservoir de molécules d'intérêt ayant une grande diversité chimique. Des recherches menées sur diverses sources d'origine naturelle ont conduit à l'isolement de molécules antivirales, antibactériennes ou encore anticancéreuses, comme le taxol isolé de l'écorce de *Taxus baccata*. Au sein de la biodiversité existante, les champignons filamenteux se sont montrés comme étant des sources non négligeables dans la découverte de nouveaux candidats médicaments (pénicilline, statines, ciclosporine etc.). Par ailleurs, l'étude de souches isolées du milieu marin est une discipline encore très récente (début des années 90) et en pleine expansion depuis ces quinze dernières années. Ainsi, cette ressource renouvelable, riche, et encore peu explorée s'avère prometteuse quant à la production de molécules structurellement innovantes et biologiquement actives.

Dans ce contexte, a débuté en 2015, le projet de thèse pluridisciplinaire intitulé : « Du bioguidage à la métabolomique pour l'identification de métabolites fongiques actifs sur des modèles cellulaires du cancer du sein- Etude de l'activité de l'ergostérol. » au sein du laboratoire Mer, Molécules, Santé (MMS) de l'Université de Nantes. Ce projet s'inscrit dans l'un des axes de recherche thématique mené au sein du laboratoire qui consiste en l'étude de la chimiodiversité des métabolites produits par des souches fongiques marines pour la mise en évidence de composés potentiellement valorisables en thérapeutique. Il a été initié par l'étude d'une collection d'extraits bruts issus de champignons filamenteux marins récoltés sur la côte Atlantique à partir de prélèvements d'eau de mer, de coquillages ou de sédiments, disponibles au sein même du laboratoire. L'intérêt de cette collection en cancérologie a été démontré au cours d'études antérieures, grâce à l'isolement d'une molécule originale brevetée pour son inhibition sélective d'ostéosarcomes humains : la ligérine (Blanchet *et al.*, 2014; Vansteelandt *et al.*, 2013). Toutefois, jusqu'ici aucune investigation n'a été menée sur le potentiel de cette collection dans le traitement du cancer du sein.

Le cancer du sein, est l'une des principales causes de décès chez les femmes, cette pathologie d'origine multifactorielle qui implique et entraîne des altérations dans plusieurs processus cellulaires tels que l'apoptose, la prolifération, la survie et la migration cellulaire. L'une des caractéristiques des cellules tumorales dont celles du cancer du sein est associée à une dérégulation du métabolisme lipidique qui se traduit par une accumulation de cholestérol cellulaire. En effet, les cellules ont un besoin accru d'énergie et d'éléments constitutifs des membranes cellulaires tel que les lipides pour assurer leur prolifération incontrôlée. Par conséquent, le métabolisme lipidique des cellules cancéreuses particulièrement altéré, constitue une cible thérapeutique dans le traitement du cancer du sein. Cependant, malgré les progrès de la médecine, nombreux sont les médicaments anticancéreux associés à des effets indésirables par manque de sélectivité.

Ainsi, la recherche de nouvelles molécules avec des structures originales et/ou des mécanismes d'action innovant constitue un axe de recherche majeure en cancérologie. C'est pourquoi les champignons filamenteux marins capables de produire des molécules d'une grande diversité chimique, représentent un fort potentiel dans la recherche de molécules possédant une activité sur les cellules cancéreuses mammaires.

Introduction générale

Un des axes de recherche centraux dans le domaine des produits naturels est celui d'isoler des molécules bioactives. Les projets de recherche associés à cette thématique sont toujours initiés par une étape de sélection à la fois de l'organisme et de la cible biologique souhaitant être étudiée, puis finalisés par celle de la valorisation des métabolites isolés. En revanche, à chacune des étapes, des choix stratégiques sont à faire en fonction des résultats attendus, mais également des moyens financiers, humains et technologiques disponibles.

Les grandes lignes de cette méthodologie et des possibilités stratégiques à suivre sont représentées dans la Figure 1.

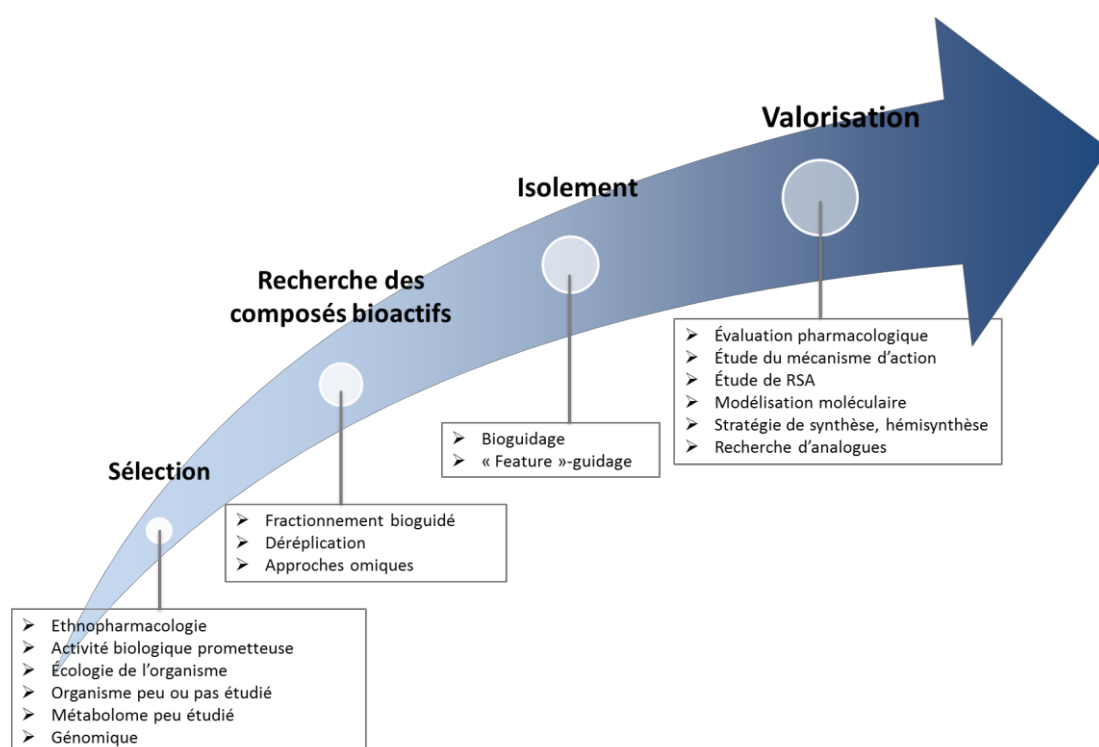


Figure 1. Méthodologie en vue de l'isolement de nouveaux produits d'origine naturelle bioactifs

L'objectif étant d'isoler des métabolites bioactifs, la première étape de sélection de l'organisme ainsi que de la cible biologique peut représenter un élément crucial dans l'obtention et l'originalité des résultats attendus. Pour réaliser cette étape, plusieurs critères peuvent être pris en considération. Le choix peut donc s'effectuer selon des études ethnopharmacologiques, et/ou selon des campagnes de criblages réalisées sur un ou des essais biologiques pertinemment choisis. Le second critère tout aussi important, concerne l'originalité de l'organisme étudié, c'est-à-dire qu'une espèce nouvellement décrite ou n'ayant jamais fait l'objet d'une étude chimique est *a priori* susceptible de produire des molécules n'ayant jamais été référencées. On peut également évoquer l'origine géographique et écologique de l'organisme.

En effet, l'expression du génome et des métabolites produits pour certain sujet d'étude semble être influencée par l'habitat (environnements extrêmes, symbiose, parasitisme, interactions avec d'autres organismes, etc.). Par ailleurs, l'originalité peut résider également dans la faible connaissance de son métabolome. Des études ont rapporté que la production de métabolites spécialisés par de nombreux micro-organismes varie selon les conditions de cultures. Ainsi, une standardisation vers une « monoculture » peut entraîner une surexpression ou une répression de l'activation de certaines voies métaboliques et conduire à une perte d'informations sur l'intégralité du métabolome d'un organisme. Pour améliorer l'accès à celui-ci, des stratégies ont été mises en place telle que l'approche dite « OSMAC » (One Strain MAnny Compounds) (Bode and Bethe, 2002; Romano *et al.*, 2018). Cette dernière stipule qu'une même espèce cultivée sur un grand nombre de milieux de cultures dont les compositions sont variées, permet d'exprimer ou de réprimer des gènes codant pour l'arsenal enzymatique nécessaire à la production des métabolites spécialisés dans les différentes voies de biosynthèse. Ainsi, il est intéressant au cours d'un criblage d'avoir des micro-organismes cultivés dans différentes conditions afin de mettre en évidence la production de métabolites bioactifs dont les voies de biosynthèse sont activées dans des conditions de culture spécifiques. Le dernier choix, plus complexe à mettre en œuvre, est un critère génomique. Une banque d'individus peut être criblée à la recherche de clusters de gènes de biosynthèse d'intérêt (PKS, NRPS, PKS/NRPS, terpènes synthases, prényltransférases, halogénases...) en fonction des molécules recherchées (s'il y a connaissance de la famille structurale responsable de l'activité).

Une fois l'organisme sélectionné pour son activité prometteuse sur une cible biologique définie, la prochaine étape est celle d'identifier les molécules potentiellement responsables de l'activité observée. La recherche de molécules nouvelles et bioactives constitue un véritable challenge dans la découverte de produits naturels et amène à des choix cruciaux dans la stratégie adoptée au cours du projet. Il existe de nombreuses approches plus ou moins classiques, avec depuis quelques années un fort engouement pour celles faisant appel à des compétences pluridisciplinaires dont des compétences informatiques. L'approche la plus classiquement utilisée, le fractionnement bioguidé, consiste à évaluer l'activité des fractions obtenues au cours des différentes étapes de fractionnement dans le but d'orienter la purification vers l'isolement des molécules responsables de l'activité observée. C'est une pratique qui peut se révéler longue et coûteuse sans originalité *in fine*, puisqu'elle ne permet pas d'évaluer l'originalité structurale d'une molécule sans l'avoir purifiée et analysée par spectroscopie RMN. Toutefois, la recrudescence des approches « omiques » et notamment le développement d'outils de métabolomique ouvre aujourd'hui, de nouvelles perspectives dans la recherche de molécules naturelles bioactives. Ainsi, le profilage métabolomique des extraits bruts réalisés par LC-MS⁽ⁿ⁾ ou par RMN permet à partir des données obtenues, d'annoter ou d'identifier des molécules au sein d'un mélange complexe sans les avoir isolées : c'est ce que l'on appelle la déréplication.

Ainsi, dans la recherche de nouvelles molécules cette stratégie peut être réalisée seule, ou en parallèle de celle du fractionnement bioguidé, dans le but d'orienter le fractionnement uniquement vers les molécules nouvelles et bioactives.

L'une des grandes interrogations émanant de la communauté scientifique concerne les moyens à mettre en œuvre pour établir le lien (s'il existe) entre les données chimiques et les données d'activités biologiques. Des approches de biochimométrie utilisant des analyses statistiques ou des modèles mathématiques, peuvent être réalisées dans le but d'établir une relation entre ces deux jeux de données. L'outil métabolomique le plus utilisé et le plus développé ces dernières années est une représentation visuelle : celles des réseaux moléculaires. Cependant, d'autres plus fondamentaux basés sur les analyses statistiques et une classification des métabolites selon leur potentiel bioactif sont en cours de développement.

Quelle que soit la stratégie adoptée pour mettre en évidence les molécules bioactives au sein d'un mélange complexe, celles-ci finissent nécessairement par être isolées dans le but de vérifier l'activité initialement désirée au lancement du projet. Cependant, plusieurs approches peuvent être encore une fois envisagées. En effet, en fonction des choix stratégiques réalisés à la deuxième étape, l'isolement des molécules peut se faire selon l'activité biologique uniquement et/ou selon les données chimiques. La purification est réalisée à l'aide de techniques chromatographiques dont les conditions sont choisies et optimisées en fonction des profils des fractions obtenues. En raison d'une très faible quantité et/ou de la présence de molécules ayant des comportements chromatographiques (élution) très proches des molécules recherchées, cette étape, peut parfois apparaître longue et fastidieuse. La réduction du nombre d'étapes de purification et d'optimisation est alors indispensable à l'exploitation et à la valorisation de l'extrait brut étudié.

Le terme de « valorisation » des produits obtenus peut s'appliquer à divers concepts. Dans le cadre d'un projet ayant pour objectif l'isolement de produits naturel bioactifs, il concerne dans un premier temps la détermination des activités biologiques des molécules isolées en vue de leurs applications pharmacologiques, thérapeutiques, cosmétiques, etc. Ainsi, différents essais biologiques (dont ceux utilisés par exemple au cours du criblage ou du fractionnement bioguidé, suivant la stratégie choisie) peuvent être réalisés dans le but d'évaluer et/ou de vérifier le potentiel pharmacologique d'une molécule. Si les composés les plus actifs sont obtenus en quantité suffisante, des études plus poussées sont envisagées afin de déterminer et d'identifier les cibles pharmacologiques et les mécanismes d'action mis en jeu. Idéalement, les molécules isolées doivent être nouvelles, c'est à dire non rapportées dans la littérature scientifique et avec un mécanisme d'action différent sur la cible étudiée. En réalité, si la structure est déjà connue, la nouveauté résidera dans la compréhension de son mécanisme d'action non élucidé sur une cible d'intérêt ou ses avantages par rapport à d'autres molécules (puissance, stabilité, possibilités de synthèse ou d'hémi-synthèse, etc.) par rapport à des composés structurellement similaires.

Pour aller plus loin, la valorisation peut s'étendre à une famille de molécules avec des études de relation structure-activités pour des molécules obtenues par synthèse, hémisynthèse, recherche d'analogues ou *in silico* par modélisation moléculaire.

C'est donc dans ce contexte, en tenant compte des différentes stratégies et des problématiques associées à l'isolement de nouveaux produits naturels que cette thèse a été menée. Ainsi, Les objectifs de cette thèse ont été les suivants : (1) évaluer le potentiel des souches fongiques issues d'une mycothèque pour leurs activités cytotoxiques sur les cellules cancéreuses mammaires, (2) rechercher par différentes méthodes de bioguidage et de métabolomique les composés responsables de l'activité observée, (3) étudier l'effet de l'ergostérol sur la régulation du métabolisme lipidique dans les cellules cancéreuses du sein.

Les travaux présentés dans ce manuscrit sont divisés en quatre chapitres principaux :

Le premier chapitre, présente tout d'abord un état de l'art sur les champignons filamenteux marins en rappelant les notions nécessaires à l'étude de ces microorganismes. Puis est présenté le criblage ayant permis de sélectionner les deux souches fongiques étudiées au cours de ces travaux pour leur activité cytotoxique sur les cellules cancéreuses mammaires.

Le deuxième chapitre expose les investigations chimiques menées sur les extraits issus des deux souches *P.chrysogenum* et *D. americana* dans le but d'identifier les molécules produites par ces souches et d'isoler celles responsables de l'activité cytotoxique sur les cellules cancéreuses du sein.

Le chapitre trois décrit le « workflow » proposé et développé au cours de ces travaux de thèse dans le but d'identifier précocement les molécules responsables de l'activité au sein d'un mélange complexe selon des approches de biochimométrie.

Enfin, le dernier chapitre approfondit l'effet de l'ergostérol identifié par les différentes approches réalisées sur l'extrait de la souche *Penicillium Chrysogenum* sur la régulation du métabolisme lipidique et plus particulièrement celui du cholestérol, ainsi que l'intégrité des radeaux lipidiques dans les cellules cancéreuses du sein.

Chapitre 1 :

Criblage et sélection des souches fongiques

Au cours des dernières années, l'augmentation significative de différentes études de biodiversité, chimiodiversité ou de biotechnologie réalisées sur les champignons d'origine marine (Cf. Figure 2) révèle un intérêt particulier de la communauté scientifique pour ces organismes associés à un habitat spécifique (Bugni and Ireland, 2004; Imhoff, 2016; Rateb and Ebel, 2011).

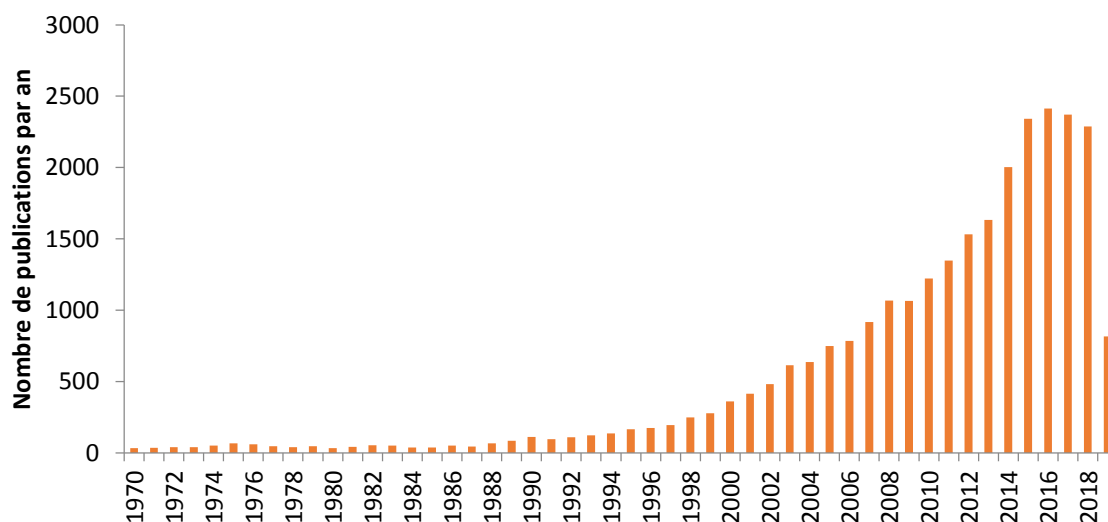


Figure 2. Évolution du nombre d'articles référencés sur la période de janvier 1970 à mai 2019 dans la base de données "Pubmed" par recherche des mot clés : « marine-derived fungi » ou « marine fungi ».

Dans la découverte de nouveaux produits naturels, le nombre de molécules isolées de ces organismes ne cesse d'augmenter et représente à l'heure actuelle environ 12% des molécules d'origine marine (Blunt *et al.*, 2018). Selon le dernier rapport réalisé sur l'année 2016 (Carroll *et al.*, 2019), référençant 1490 nouveaux produits naturels marins (PNM), 30% sont issus de champignons, ce qui en fait le phylum le plus étudié au cours de cette année devant les éponges (16%) et les cnidaires (10%) (Cf. Figure 3). Si l'étude des champignons marins est relativement récente, celle-ci a réellement pris un essor dans les années 2000 avec la prise de conscience que ces organismes étaient assez peu étudiés dans l'environnement marin contrairement à leurs homologues terrestres. Ils représentent ainsi un vaste réservoir de nouvelles molécules originales dont les activités toxiques et/ou biologiques à visées pharmacologiques et médicinales ont déjà pu être démontrées (Elissawy *et al.*, 2015; Roullier *et al.*, 2016; Saleem, 2015; Vansteelandt *et al.*, 2013).

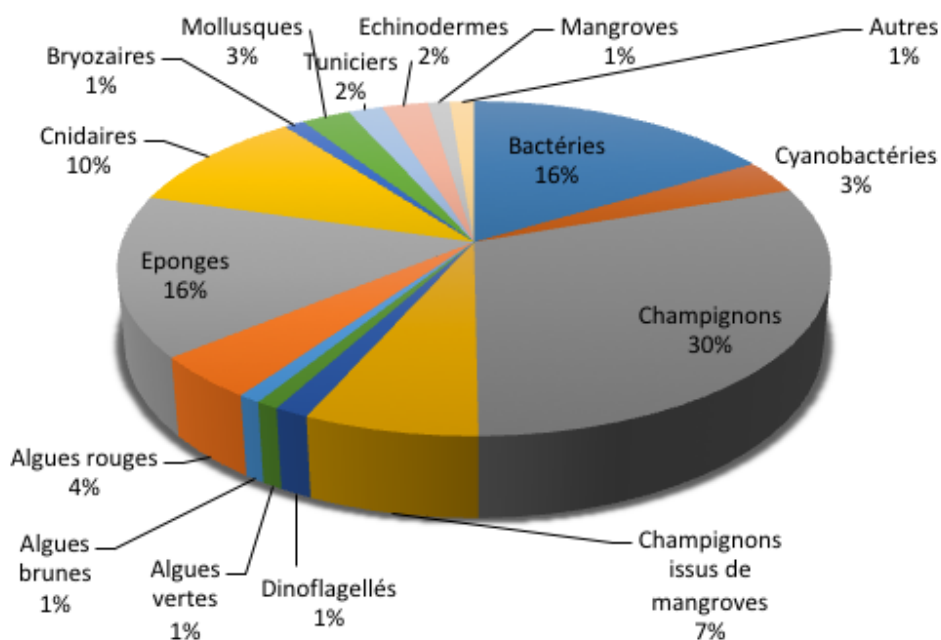


Figure 3. Distribution des nouveaux produits marins (NMP) en fonction de leur habitat sur la période de janvier à décembre 2017 (Carroll *et al.*, 2019).

Dans le cadre de cette thèse, cette ressource prometteuse a donc été investiguée et exploitée au cours d'une campagne de criblage réalisée sur les extraits bruts fongiques issus de l'extractothèque du laboratoire Mer, Molécules, Santé (MMS), afin d'évaluer leur intérêt potentiel dans le traitement du cancer du sein.

Afin de justifier l'étude menée et de la replacer dans son contexte, ce premier chapitre commence par une introduction bibliographique afin de rappeler brièvement la définition générale et la classification actuelle des champignons. Un focus plus spécifique a été réalisé sur le sujet de notre étude, à savoir les champignons marins, leurs métabolites bioactifs et leur potentiel dans le traitement du cancer du sein. Puis, une fois les objectifs de l'étude définis, les résultats obtenus ainsi que les souches sélectionnées pour la suite de la thèse ont été présentés et discutés.

1. Introduction bibliographique

1.1. Les champignons

1.1.1. Définition générale

Les champignons, encore appelés *Fungi* (latin) ou *Mycètes* (grec), correspondent à un groupe diversifié d'organismes. Ce sont des organismes eucaryotes présentant un mode de nutrition caractéristique, basé exclusivement sur l'absorption de nutriments. Ces organismes hétérotrophes nécessitent un apport en carbone qui dépend de la présence de matières organiques produites par d'autres organismes. Ainsi, leur mode de nutrition peut s'apparenter à trois types d'interactions trophiques : saprophytisme (décomposition de la matière morte), parasitisme (aux dépens des organismes hôtes) ou symbiotisme (association à bénéfice réciproque avec d'autres organismes tels que des plantes, des animaux ou d'autres champignons). Les champignons possèdent des caractéristiques similaires aux végétaux comme la possibilité de se reproduire sous forme sexuée et asexuée, mais également des caractéristiques proches de certains animaux comme la présence de chitine et de glucane dans leur paroi cellulaire.

1.1.2. Classification des champignons

Pour comprendre l'évolution de la classification du règne fongique, il est avant tout nécessaire de rappeler de manière générale celle du monde du vivant. Historiquement, dans la classification établie par Haeckel en 1894, le monde du vivant est découpé en trois règnes : Animal, Végétal, et Protozoaires. Les champignons sont classés parmi les végétaux du fait de la présence d'une paroi cellulaire et de plusieurs similitudes entre leurs cycles de reproduction et ceux des algues. Ce n'est qu'en 1969, qu'ils ne constitueront qu'un seul et même règne indépendant de ceux des autres. La classification réalisée par Whittaker (Whittaker, 1969) expose un découpage en cinq règnes : Animaux, Champignons, Végétaux, Protistes, et Monères (les Procaryotes). Par la suite, une subdivision du règne des Monères en Archéobactéries et Eubactéries sera établie par Woese en 1977 (Woese and Fox, 1977). En 1990, ce dernier propose une nouvelle classification simplifiée avec un découpage du monde du vivant en trois domaines : Eucaryotes, Archéobactéries et Eubactéries (ces deux derniers étant regroupés sous le terme de Procaryotes). Depuis, ces regroupements sont toujours discutés, dont la dernière proposition réalisée en 2004 par Cavalier-Smith propose deux empires (eu et procaryotes) avec six règnes : Animaux, Champignons, Végétaux, Chromistes, Protozoaires et Bactéries (Cavalier-Smith, 2004) ; ou plus récemment celle issue d'un travail collaboratif autour du projet "Tree Of Life" présentée à la page suivante (Cf. Figure 4).

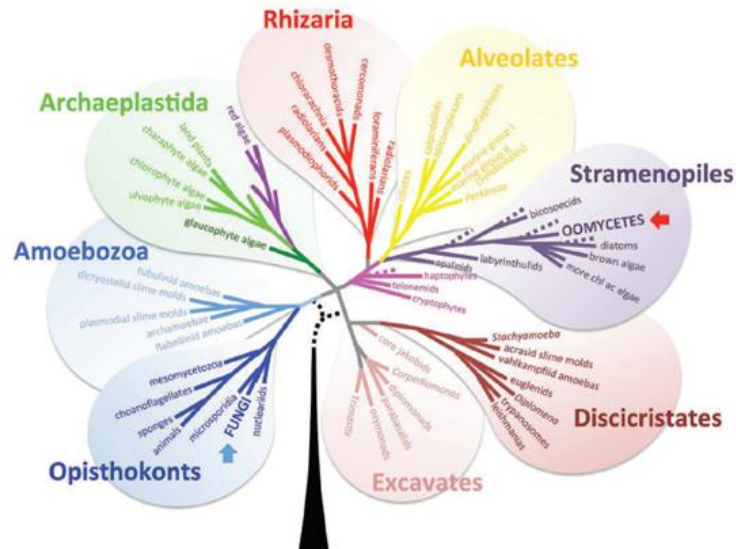


Figure 4. Classification phylogénique (Lee *et al.* 2012)

Actuellement, le règne des champignons appelé eumycètes, ou encore champignons « vrais » est divisé en 10 embranchements (Cf. Figure 5) dont le majoritaire est celui des Ascomycota (ascomycètes) avec environ 64 000 espèces connues (Schoch *et al.*, 2009). Il comprend des champignons filamenteux (ex: *Aspergillus* et *Penicillium*) mais également des levures (ex: *Saccharomyces cerevisiae*) et des lichens résultant de l'association d'une algue et/ou d'une cyanobactérie avec un champignon filamenteux (ex: *Cladonia*, *Lecanora* et *Parmelia*) (Blackwell, 2011).

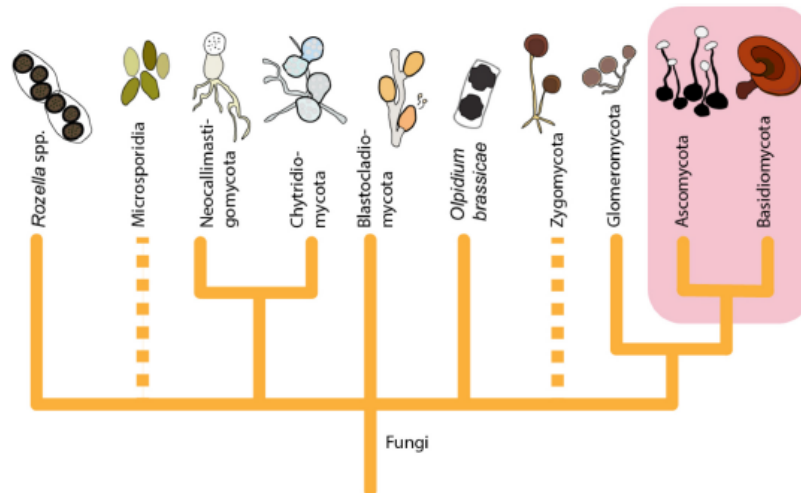


Figure 5. Classification du règne *fongique* divisé en 10 sous-embranchements (Linnakoski *et al.*, 2018).

1.2. Que signifie le terme « Champignons marins » ?

1.2.1. Définition

Pendant longtemps considérés comme ayant une forme de vie fondamentalement terrestre, la présence de champignons dans l'habitat marin n'est que très récemment connue et reconnue par la communauté scientifique. Pourtant, les premières espèces de champignon isolées dans ce milieu (*Phaeosphira typharum* et *Spaeria oceanica*) ont été décrites respectivement par J.B.H.J. Desmazières en 1849 et par C. Durieu de Maisonneuve et J.C.F. Montagne en 1869 (Kohlmeyer and Kohlmeyer, 1979). Ce n'est qu'à partir de 1944, avec la parution des travaux de Barghoorn et Linder regroupant un grand nombre d'espèces, que la mycologie marine prend un réel essor (Barghoorn and Linder, 1944). Cette découverte déclenche une vaste campagne de collecte de champignons dans le monde entier afin d'identifier un maximum des espèces présentes dans l'habitat marin. A l'heure actuelle, alors que le nombre total d'espèces fongiques a été estimé entre 2,2 et 5 millions, il semblerait que seulement 10% aient été recensées (Amend *et al.*, 2019; Hawksworth and Gardens, 2017). Parmi ces espèces, environ 1 000, réparties majoritairement entre les phylums des Ascomycètes et des Basidiomycètes seraient spécifiques du milieu marin (Amend *et al.*, 2019). Néanmoins, cela ne représenterait que 10% de la fonge marine. Par conséquent, ce règne est encore très peu connu de la communauté scientifique.

Aujourd'hui, un grand nombre de publications scientifiques utilisent les termes de «champignon dérivé du milieu marin» ou «champignon d'origine marine» (Pang *et al.*, 2016). Ces termes peuvent être utilisés pour plusieurs situations en faisant référence à l'environnement (emplacement) à partir duquel les champignons sont isolés, ou en se rapportant à leur capacité à se développer en milieu marin d'où les notions de « marins obligatoires » ou « marins facultatifs». Les champignons marins obligatoires sont ceux se développant et sporulant exclusivement dans un habitat marin ou estuarien, tandis que des espèces trouvées en mer et en milieu terrestre sont catégorisées de champignons marins « facultatifs », un terme qui semble plus approprié (Overy *et al.*, 2014).

En 2014, face à ce constat, un consortium regroupant des mycologues et des chimistes des produits naturels fongiques a été établi dans le but de discuter et d'apporter des réponses à l'une des questions majeures toujours non résolue dans le domaine : « **Qu'est-ce qu'un champignon marin ?** ». Des propositions ont été faites sur l'utilisation du terme «champignons d'origine marine» dont la suivante : [tout champignon récupéré à maintes reprises dans des habitats marins pour les raisons suivantes: 1) il est capable de se développer et/ou de sporuler (sur le substrat) dans des environnements marins; 2) il forme des relations symbiotiques avec d'autres organismes marins; ou 3) il est démontré qu'il s'adapte et évolue au niveau génétique ou est métaboliquement actif dans les environnements marins](Pang *et al.*, 2016).

1.2.2. Distribution géographique et écologique en milieu marin

Les champignons sont largement répandus dans tous les habitats marins que ce soit dans les profondeurs de la mer ou dans les eaux de surface, et en présence d'une grande variété de substrats : bois submergé (Rämä *et al.*, 2014), plantes estuariennes (ex mangrove, herbe) (Rämä *et al.*, 2014), les sédiments (Orsi *et al.*, 2013), les invertébrés (Yarden, 2014), les algues (Wainwright *et al.*, 2017) ou les mammifères marins. Bien que la répartition à l'échelle mondiale ne soit pas chiffrée, quelques études ont été réalisées en ce sens sur la base de critères spécifiques. Par exemple, l'analyse réalisée en 2004 sur les publications décrivant l'isolement de métabolites issus de champignons marins (Cf. Figure 6) a permis de mettre en évidence une forte présence des champignons marins dans les éponges (33%) et les algues (24%) (Bugni and Ireland, 2004). Depuis, cette répartition des espèces en fonction des habitats ne cesse d'évoluer en fonction des projets menés et des environnements étudiés, notamment, avec l'étude récente des mangroves ayant permis l'identification d'un grand nombre d'espèces endophytes au sein de cet habitat, jusqu'ici peu exploré (Rateb and Ebel, 2011).

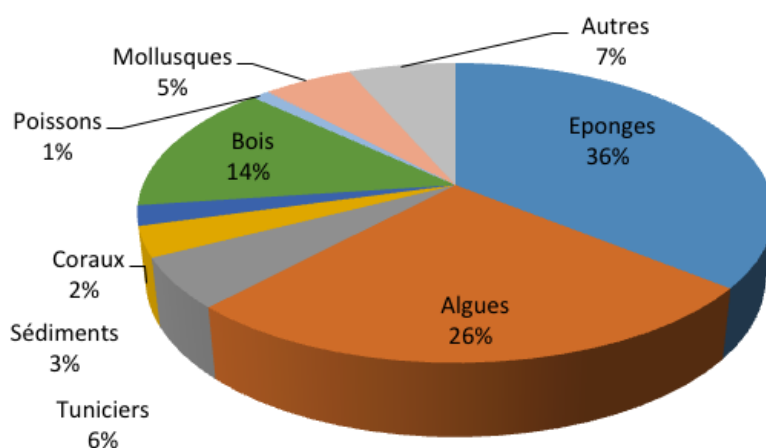


Figure 6. Distribution des métabolites isolés de champignons marins en fonction de leurs habitats. (Bugni and Ireland, 2004).

Au-delà des divers habitats, les champignons marins possèdent une importante capacité physiologique à s'adapter à des conditions environnementales parfois extrêmes liées à une salinité élevée, des modifications de pH, une faible teneur en oxygène, des pressions hydrostatiques fortes liées à la profondeur, ou encore des températures extrêmes (Cantrell *et al.*, 2006; Tisthammer *et al.*, 2016). Cependant, la distribution géographique des microorganismes fongiques marins dans la biosphère n'est que très peu étudiée. En effet, les zones explorées sont souvent inférieures à 1 000 km, elles peuvent être spécifiques d'un climat ou d'une région (ex des climats tropicaux ou des régions polaires et sont focalisées pour la plupart d'entre elles sur des habitats ou substrats spécifiques (ex mangroves, algues, coraux etc.), ou encore se référer uniquement aux espèces dont des molécules ont été publiées dans la littérature (Carroll *et al.*, 2019).

Néanmoins, une récente étude menée au cours des 10 dernières années sur la distribution des produits naturels d'origine marine isolés par phylum en fonction des régions géographiques (Cf. Figure 7) a permis de mettre en évidence l'incroyable diversité géographique du phylum des ascomycètes. Toutefois, certaines régions restent sous explorées comme celles de l'Asie du sud-est, la Nouvelle Zélande, l'Amérique pacifique et Atlantique, les Caraïbes ou encore les régions polaires de l'Arctique et de l'Antarctique (Carroll *et al.*, 2019)

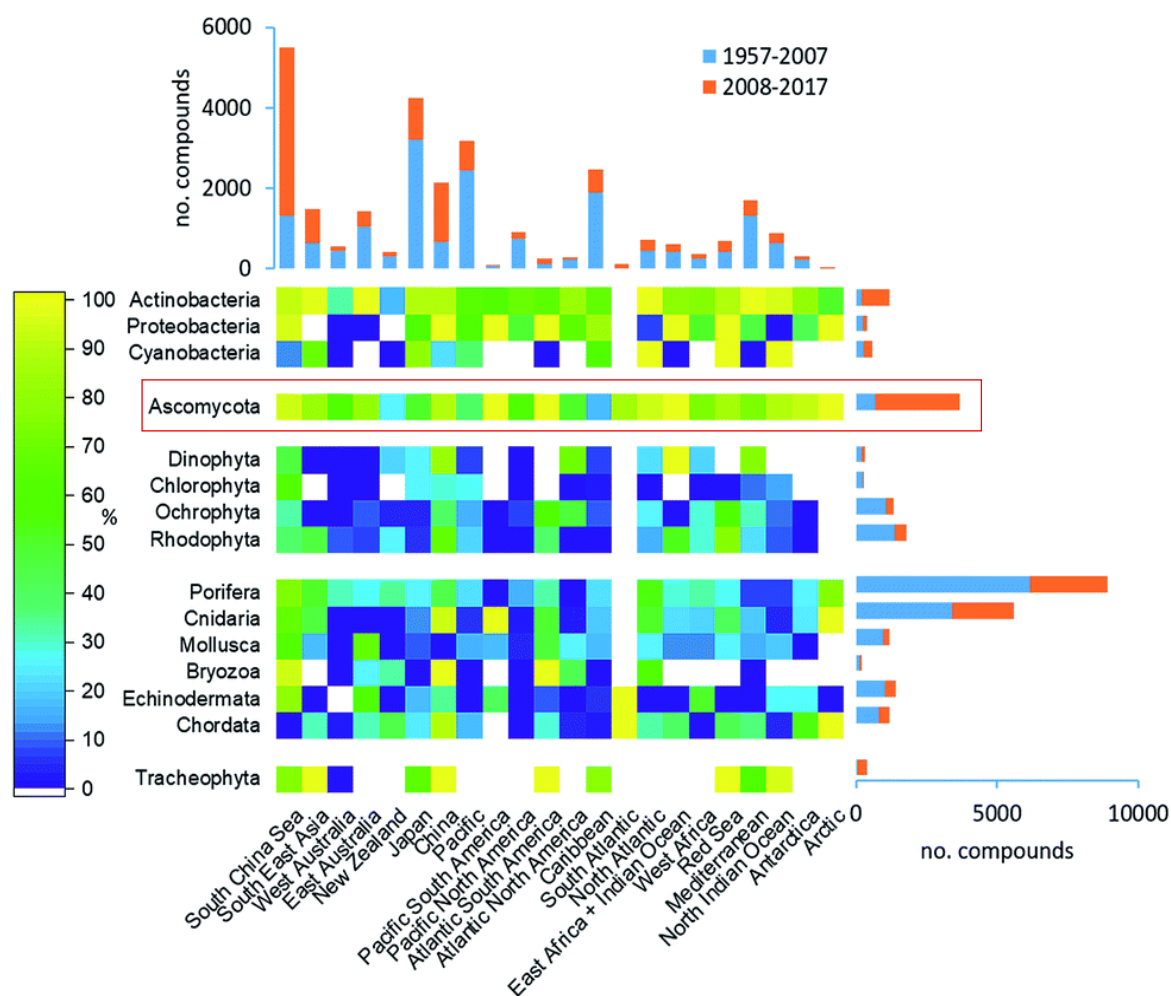


Figure 7. Carte des chaleurs représentant la proportion totale des produits naturels d'origine marine découverts par phylum en fonction des régions géographiques entre 2008 et 2017 (Carroll *et al.*, 2019).

Cependant, le critère écologique des champignons marins facultatifs ou obligatoires dans les études menées sur la découverte de nouvelles molécules d'origine marine n'étant que secondaire, il renforce l'idée selon laquelle la caractérisation génétique et biochimique des espèces nécessite d'être complétée. En effet, pour les produits naturels issus de champignons facultatifs, il est difficile d'évaluer si la même espèce fongique présente dans deux environnements distincts est capable de produire la même molécule, puisqu'il existe peu de comparaison avec les espèces terrestres associées génétiquement.

En revanche, l'importance de la présence de ces microorganismes marins facultatifs ou obligatoires dans les écosystèmes marins, a été soulignée au travers d'études mettant en évidence leur rôle prédominant dans le recyclage de la matière organique (par leur mode de vie saprophyte) (Hyde *et al.*, 1998). D'autres études ont également révélé leur participation active à la régulation du phytoplancton (Gutiérrez *et al.*, 2011) mais également à celle de la formation des sédiments marins (Amend *et al.*, 2019).

Si les champignons terrestres sont connus et étudiés pour leurs rôles dans divers processus tels que celui de la décomposition de la matière organique (comme dans les forêts), de la fermentation dans l'alimentation (ex : vin, bière, pain, produits laitiers) ou encore dans la production de molécules bioactives (ex : pénicilline) (Hyde *et al.*, 2019), on peut s'interroger sur l'intérêt potentiel des champignons d'origine marine.

Tout d'abord, il est à noter que malgré l'impact des progrès technologiques et socio-économiques sur notre environnement, ces organismes sont continuellement présents dans les écosystèmes marins par leur incroyable capacité d'adaptation. Ainsi, le rôle potentiel de ces organismes pour des applications de bioremédiation a été récemment décrit. Des études ont permis de mettre en évidence leur rôle dans la décontamination (transformation ou absorption) des produits accumulés par les environnements marins (Nicoletti and Andolfi, 2018), tels que les produits phytosanitaires (phénanthrène, le benzopyrène, le Dichlorodiphényldichloroethane (DDD) (Ortega *et al.*, 2011), l'esfenvalérate, le parathion méthyl dieldrine) (Nicoletti and Andolfi, 2018) ; les antibiotiques liés à un élevage intensif (l'oxytétracycline) (Ahumada-rudolph *et al.*, 2016) ; des composants aliphatiques du pétrole ; des résidus de 2,4,6-trinitrotoluène (TNT) ou encore de hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX) (Jain *et al.*, 2004); celles de métaux lourds (cadmium, cuivre, plomb, chrome) (Lotlikar *et al.*, 2018; Siddiquee *et al.*, 2015) ; ou encore dans la dégradation des plastiques (Brunner *et al.*, 2018).

En s'intéressant à la production des métabolites produits par les champignons marins, il a été remarqué qu'ils sont capables de synthétiser tout comme ceux d'origine terrestre, un large éventail de composés de classes chimiques diversifiées : terpènes, alcaloïdes, peptides, polycétides, lipides, enzymes, vitamines (Blunt *et al.*, 2018). Ces molécules sont nécessaires à leur croissance mais jouent également un rôle important dans les interactions possibles avec d'autres organismes présents dans leur environnement. En effet, la production des métabolites spécialisés peut à la fois permettre l'établissement de symbioses au sein d'un écosystème, faciliter la conquête du territoire ou résulter de la mise en place de véritables mécanismes de défense pour se protéger des agressions extérieures (parasites, prédateurs, compétiteurs). Les molécules produites peuvent ainsi par leur cytotoxicité présenter un intérêt pharmacologique et/ou thérapeutique.

Ainsi, plusieurs revues ont été réalisées dans le but de classer les molécules bioactives en fonction de leur classe chimique et/ou de leur cible pharmacologique ou thérapeutique (Deshmukh *et al.*, 2018; Elissawy *et al.*, 2015; Linnakoski *et al.*, 2018).

Par exemple, parmi les 153 nouvelles molécules bioactives d'origine fongique marine référencées de 2014 à 2016 (Liming *et al.*, 2016), les alcaloïdes (27,0%) et les polycétides (25,7%) étaient les classes chimiques les plus représentées, suivis des peptides, des terpènes et des stéroïdes. En parallèle, le classement de ces mêmes molécules selon leur cible pharmacologique (Cf. Figure 8) a permis de mettre en évidence le fort potentiel des champignons marins dans la production de métabolites présentant plusieurs activités biologiques, dont la principale observée : la cytotoxicité (37,5%).

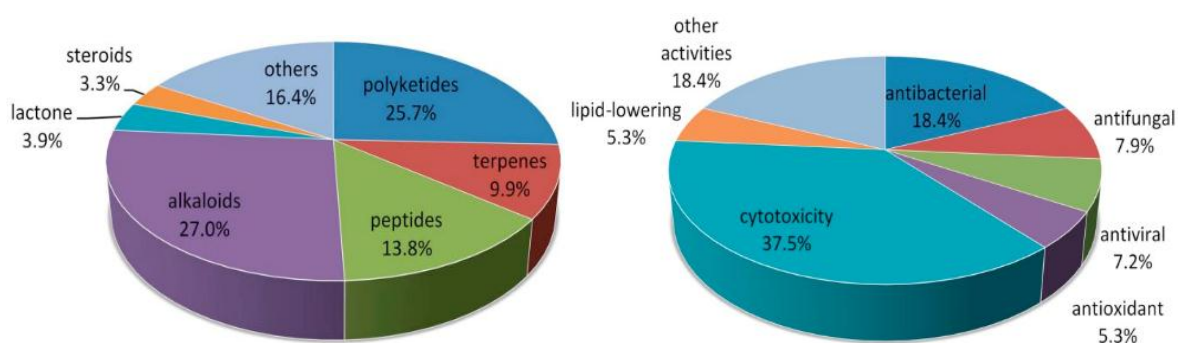


Figure 8. Nouveaux composés isolés de champignons marins (2014-2016) répartis en fonction de leurs classes chimiques (a) ou de leurs activités biologiques (b) (Liming *et al.*, 2016).

Ainsi, selon la cible pharmacologique, il peut être cité comme exemples de molécules isolées de champignons marins possédant une activité antibactérienne : la penicibrocazine B (1), la penicyclone A (2), la desméthylisaridine C1 (3) (Meng *et al.*, 2015; Xu *et al.*, 2015); une activité antivirale: la neoechinuline B (4), la stachybotrine D (5) (Moghadamtousi *et al.*, 2015) et anticancéreuse : l'acaromycine A (6), l'eremofortine C (7), le cinnamolide (8), l'insulicolide A (9), les varioloïdes A et B (10 et 11) (Deshmukh *et al.*, 2018). Les structures des molécules figurent ci-dessous (Cf. Figure 9 et Figure 10).

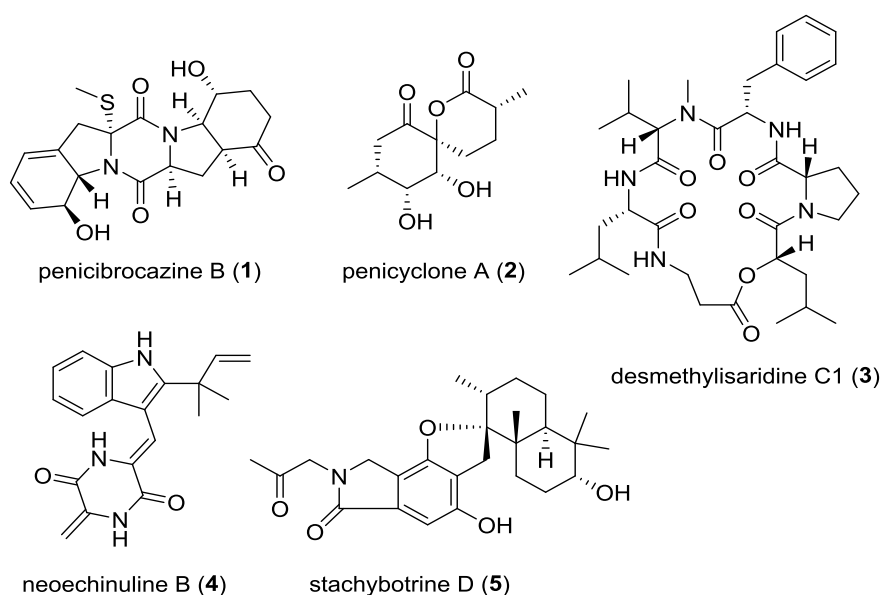


Figure 9. Structures des molécules possédant une activité antimicrobienne (1-3) ou une activité antivirale (4 et 5).

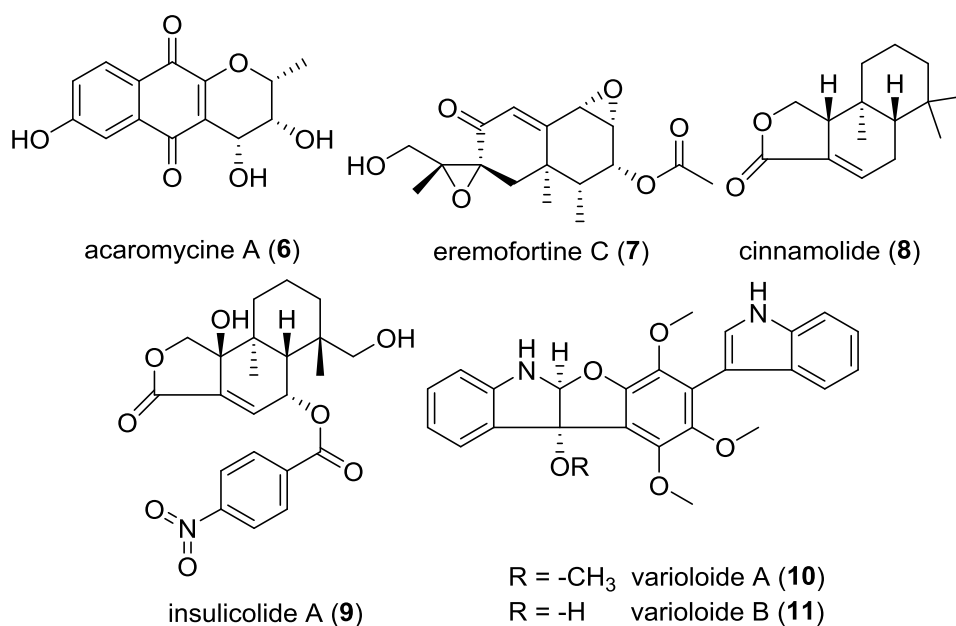


Figure 10. Structures des molécules possédant une activité anticancéreuse.

Bien qu'aucune des molécules isolées d'un champignon marin ne soit encore approuvée en tant que telle par la Food Drug Administration (FDA), celles-ci représentent une source d'inspiration potentielle pour l'obtention de composés hémisynthétiques possédant des activités biologiques améliorées (sélectivité, spécificité ou activité plus élevée). Toutefois, la céphalosporine C (**12**), issue d'une souche de *Cephalosporium acremonium* collectée dans une bouche d'égout donnant sur la mer Méditerranée sur la côte italienne, est souvent considérée comme la première molécule fongique d'origine marine commercialisée. Elle a en effet été approuvée par la FDA en 1964 et est à l'origine d'un important panel de molécules antibactériennes disponibles sur le marché, avec une recherche toujours active sur cette série chimique. En effet, la céphalosporine C est à l'origine d'un important panel de molécules antibactériennes, dont la première a été approuvée par la FDA en 1964. Depuis la commercialisation en 2015, du ceftobiprole (**13**) (MABELIO) indiqué dans le traitement des pneumonies nosocomiales et communautaires, et autorisé pour le moment dans 12 pays européens, la Suisse et le Canada, une cinquième génération d'antibiotiques est en cours de développement. Un autre exemple, est celui de la plinabuline NPI2358 (**14**), un analogue obtenu à partir de l'halimide (**15**) isolé d'*Aspergillus* sp. CNC-139, utilisé actuellement en phase 3 dans le traitement du cancer du poumon non à petites cellules en combinaison avec le docétaxel (Mayer *et al.*, 2010).

Dans certains cas, l'isolement de molécules fongiques d'origine marine a permis d'enrichir et de compléter les études pharmacologiques réalisées sur une série d'analogues. C'est le cas de la ligérine (**16**), isolée d'un *Penicillium* d'origine marine pour laquelle une forte activité sur des lignées d'ostéosarcome humain a pu être démontrée (Blanchet *et al.*, 2014; Vansteelandt *et al.*, 2013).

Cette molécule est un analogue chloré de la fumagiline (**17**), dont la série chimique n'a été étudiée jusqu'ici que pour des activités de type antiparasitaires (van den Heever *et al.*, 2016). Les structures des molécules figurent à la page suivante (Cf. Figure 11).

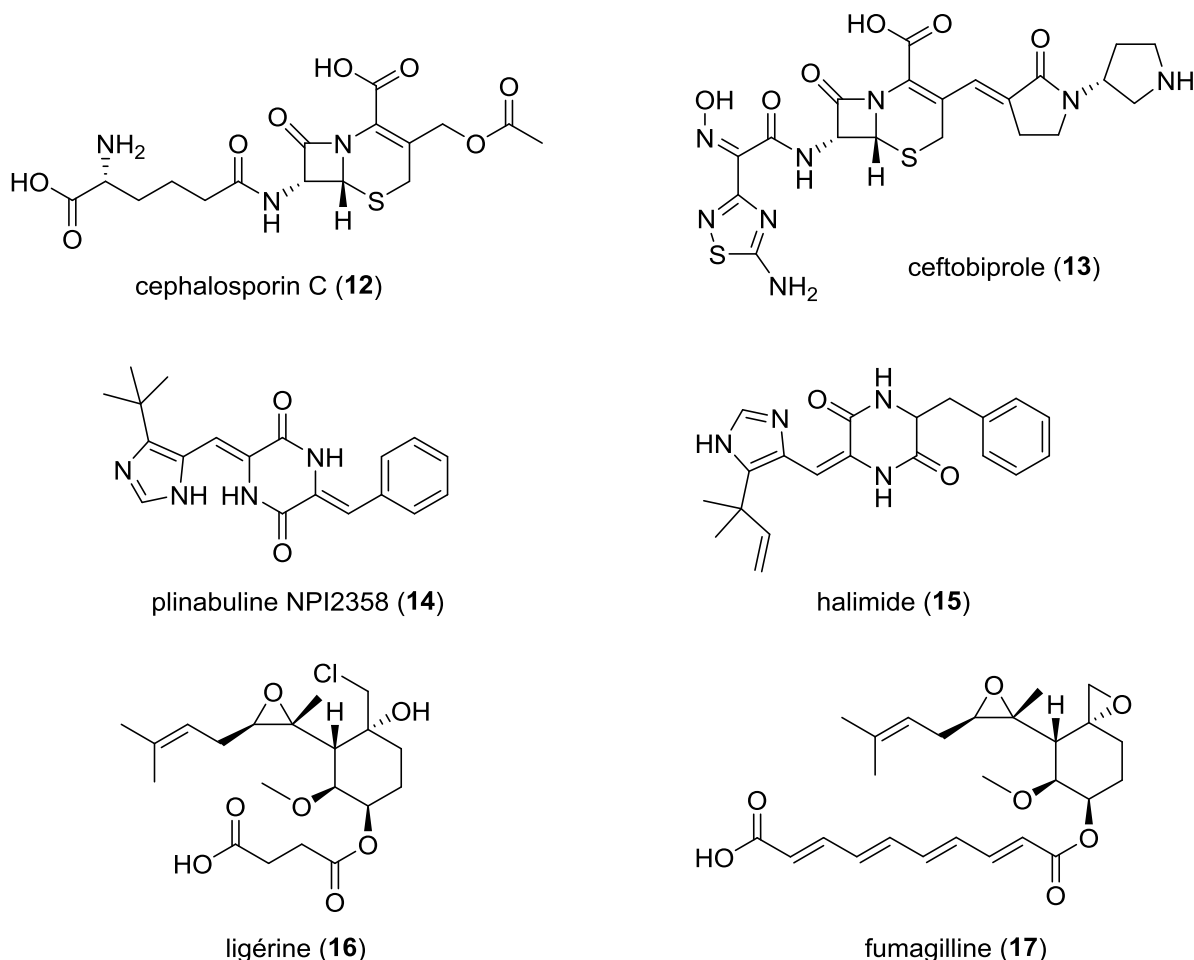


Figure 11. Molécules d'origine naturelle fongique (A) et leurs dérivés hémisynthétiques (B) présentant une activité biologique améliorée.

Parmi les différents exemples cités, les champignons ont montré leur intérêt dans la production de métabolites bioactifs avec des structures de classes chimiques et des cibles biologiques très diversifiées. Cependant, devenu depuis quelques années un enjeu majeur de santé publique mondiale, de nombreuses campagnes de financement ont été et sont réalisées dans la recherche sur le cancer, notamment dans la découverte de nouvelles molécules afin d'améliorer l'efficacité des traitements. Ce phénomène peut ainsi expliquer que la majorité des molécules isolées soient testées pour leur cytotoxicité, ce qui en fait la principale activité observée.

Selon les dernières estimations publiées par le Centre international de recherche sur le cancer (CIRC), 18,1 millions de nouveaux cas de cancer auraient été diagnostiqués dans le monde en 2018 et 9,6 millions de personnes sont décédées d'un cancer.

En France, selon l'Institut national du cancer (INCa), on estime à 382 000 le nombre de nouveaux cas de cancer pour l'année 2018, 204 600 chez l'homme et 177 400 chez la femme. Sachant que le cancer du sein est le cancer le plus fréquent chez la femme, notre intérêt s'est porté sur ce type de cancer.

Dans le but de démontrer l'intérêt de notre étude, une brève présentation permettant de définir non pas le cancer du sein mais les cancers du sein ainsi que les molécules fongiques actives sur cette cible biologique a été réalisée. Des explications plus approfondies sur la compréhension des mécanismes biologiques et physiologiques mis en jeu seront présentés dans le chapitre 4.

1.3. Le rôle potentiel des champignons dans la production de molécules bioactives dans le traitement du cancer du sein

1.3.1. Le cancer du sein ou plutôt les cancers du sein

Le cancer du sein est une tumeur maligne qui se développe à partir des cellules constituant la glande mammaire. Les cellules malignes se multiplient de manière désordonnées jusqu'à créer une tumeur qui s'attaque aux tissus sains avoisinants. Mais sous cette description générale, se cachent plusieurs réalités. En effet, le cancer du sein est une maladie dite hétérogène c'est à dire qu'il existe plusieurs types de tumeurs qui sont d'ailleurs traitées différemment. On parle ainsi de carcinome in situ ou infiltrant suivant la localisation des cellules cancéreuses ; de type luminal A ou B suivant le nombre de récepteurs aux hormones (œstrogènes(ER+) et progestérone (PR+)) ; de type HER2+ (Human Epidermal Growth Factor Receptor-2) si elles expriment le récepteur HER2 à leur surface ; de type basal-like si les cellules ressemblent aux cellules basales ; de type triple-négatif si les cellules cancéreuses n'ont pas de récepteurs aux œstrogènes, ni aux progestérones et qu'elles ne surexpriment pas HER2.

Il existe différents types de traitements pouvant être utilisés pour traiter ces cancers du sein : la chirurgie, la radiothérapie, l'hormonothérapie, la chimiothérapie et les thérapies ciblées. Dans le cas des agents de chimiothérapie destinés à traiter des cancers du sein, on trouve les composés suivants : cyclophosphamide, docétaxel, doxorubicine, épirubicine, fluoro-uracile, méthotrexate, capécitabine, gemcitabine ou paclitaxel. Il est intéressant de noter que pour cette dernière molécule, le paclitaxel, bien que découverte initialement à partir d'écorce d'if (*Taxus baccata*), elle est désormais connue pour être biosynthétisée par de nombreuses espèces fongiques endophytes appartenant au genre *Taxomyces*, *Alternaria*, *Fusarium* et *Aspergillus* (Zhou *et al.*, 2010). Malgré cet arsenal thérapeutique, le cancer du sein est le cancer le plus diagnostiqué et l'une des premières causes de décès chez la femme.

A l'échelle mondiale et pour l'année 2018, il a été reporté 2,1 millions de nouveaux cas et 627 000 décès. Etant donné qu'il n'existe pas un mais des cancers du sein et au vu de ces chiffres alarmant, la recherche de nouveaux agents pour limiter et/ou inverser leur progression reste indispensable.

1.3.2. Les métabolites fongiques d'intérêt dans le traitement du cancer du sein

Dans ce cadre, les métabolites fongiques restent une source prometteuse avec l'isolement de plusieurs composés, dont les activités sur des cibles biologiques en lien avec le traitement du cancer du sein ont pu être démontrées. Une étude bibliographique a été réalisée en ce sens en se focalisant sur les molécules issues d'espèces fongiques marines. D'après la liste de molécules obtenues et figurant en annexe 1, quelques exemples de molécules les plus prometteuses ou encore possédant des structures originales sont présentées ci-dessous.

Parmi ces molécules bioactives (Cf. Figure 12), l'haloroquinone (**18**), une anthraquinone isolée de la souche fongique (No. 1403) *Halorosellinia* sp. collectée d'une mangrove du sud de la Chine en 2010, a démontré *in vitro* et *in vivo* un fort potentiel en tant que principe actif. En effet, son activité antiproliférative a pu être observée sur six lignées cancéreuses, dont deux lignées cancéreuses mammaires avec une CI_{50} comprise entre 3 et 9.6 μ M (Xie *et al.*, 2010). Par ailleurs, l'utilisation de cette molécule *in vivo* a permis de diminuer considérablement la croissance tumorale de modèle murins porteurs de xénogreffes de la lignée MDA-MB-435 ou d'une lignée résistante à l'adriamycine MCF-7 /ADR, et cela sans aucune autre toxicité observée (Zhu *et al.*, 2012). Cette diminution tumorale a également été observée pour l'ophioboline O (**19**), un sesquiterpène isolé de la souche marine *Aspergillus ustus* qui a été rapporté comme ayant un effet pro-apoptotique *in vitro* et *in vivo* (Lv *et al.*, 2015).

Une autre molécule, le déacétyl-mycoepoxydiène (DAM) (**20**) produit par *Phomopsis* sp., testé sur des modèles *in vitro* et *in vivo* a montré un fort potentiel en tant que nouvel inhibiteur des microtubules et de la prolifération cellulaire de la lignée MCF-7 (Shan-Shan *et al.*, 2015). D'autres molécules, tel que l'hirsutanol A (**21**), un sesquiterpène isolé de *Chondrostereum* sp. SF002 ou encore la fumigaclavine C (**22**) isolée de la souche *Aspergillus Fumigatus* ont été rapportées dans la littérature pour leur effet antiprolifératif de la lignée cancéreuses du sein (Li *et al.*, 2013; Yang *et al.*, 2013).

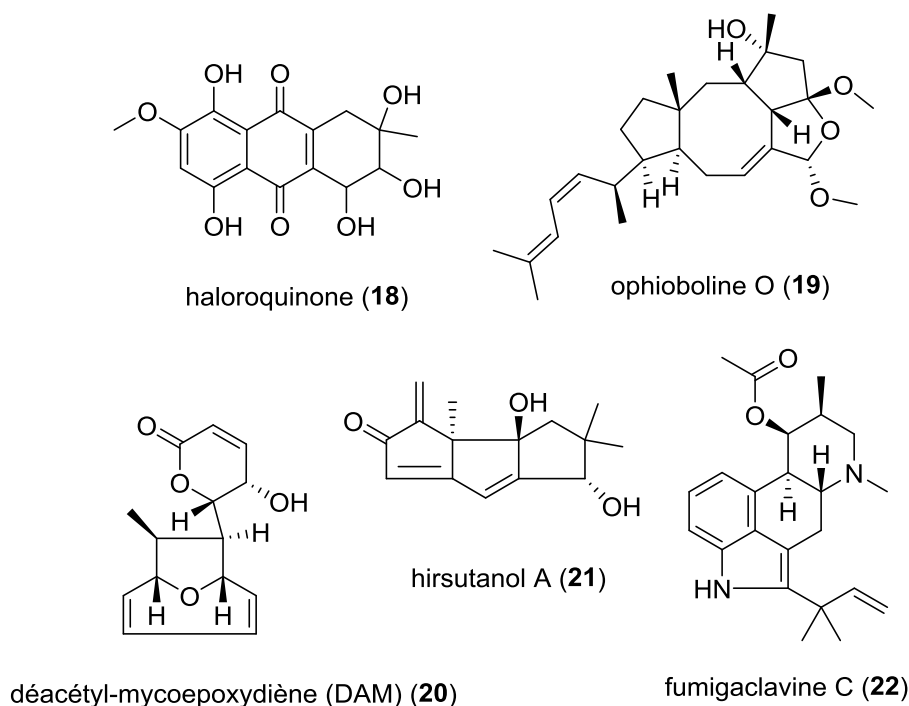


Figure 12. Exemple de molécules d'origine fongique marine présentant une activité cytotoxique sur les cellules cancéreuses mammaires.

L'étude de la fonge et plus particulièrement de celle issue du milieu marin est finalement une thématique de recherche très récente et, par conséquent, peu connue comparée aux autres disciplines. Au regard de la diversité et de l'originalité structurale des molécules produites par ces organismes dont l'intérêt pharmacologique et thérapeutique a pu être démontré pour certaines, leur rôle potentiel dans le traitement des cancers du sein est finalement peu décrit dans la littérature. Ainsi, quelques molécules expérimentées sur des modèles *in vitro* et *in vivo* ont pu démontrer un intérêt potentiel, notamment pour leur capacité à inhiber la prolifération cellulaire de certaines lignées cancéreuses mammaires. Cette ressource semble donc prometteuse et une piste intéressante à creuser pour la découverte de molécules d'intérêt dans cette pathologie qu'est le cancer du sein.

2. Objectifs de l'étude

Le laboratoire Mer Molécules Santé (MMS) de l'Université de Nantes a été pionnier dans la constitution d'une mycothèque d'origine marine. Celle-ci a débuté au cours de campagnes de prélèvements d'échantillons (de coques, de moules, d'eau de mer et de sédiments) réalisés le long de la côte ouest de la France, au nord et au sud de l'estuaire de la Loire, en réponse à des problèmes récurrents de contamination de fruits de mer (Geiger *et al.*, 2013; Grovel *et al.*, 2003; Poirier *et al.*, 2007; Ruiz *et al.*, 2010). En effet, la présence de souches fongiques et plus particulièrement, l'accumulation des métabolites produits par ces organismes a été rapportée comme pouvant être à l'origine de la toxicité et de la mortalité des mollusques observée (Sallenave-Namont *et al.*, 2000). Ainsi, lors des prélèvements d'échantillons réalisés en rapport avec ce précédent projet, il a été isolé 456 souches fongiques permettant d'initier le début de la mycothèque. Puis, celle-ci a été complétée au cours de divers projets permettant de regrouper un total d'environ 1500 souches fongiques d'origine marine réparties en 40 genres, dont le majoritaire est celui des *Penicillium* (Cf. Figure 13).

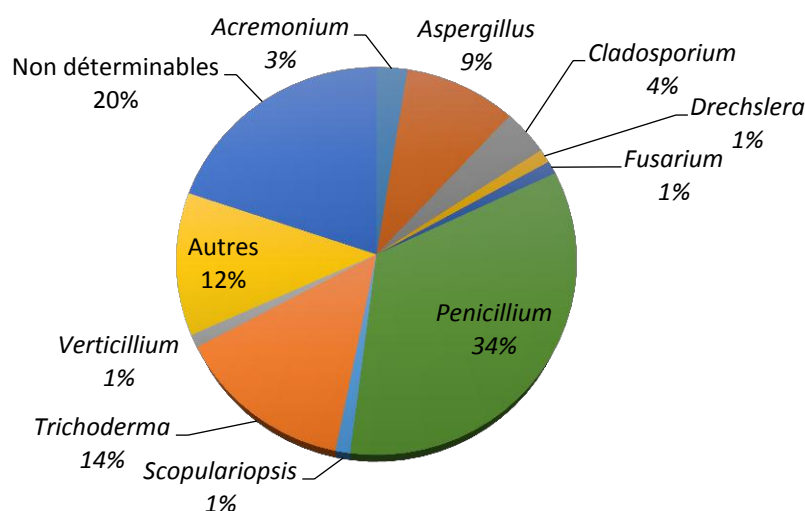


Figure 13. Répartition des espèces fongiques de la collection du laboratoire MMS en fonction de leur genre.

Le premier objectif de cette thèse a donc été d'évaluer le potentiel de la mycothèque du laboratoire MMS sur une cible biologique en lien avec le cancer du sein. Pour ce faire, il a été choisi d'étudier au cours d'un criblage : l'effet de 98 extraits bruts issus de champignons filamenteux marins de cette mycothèque sur la prolifération des cellules cancéreuses mammaires MCF-7 via un test classique et rapide de cytotoxicité nommé MTT en raison du réactif utilisé (le sel de tétrazolium MTT : bromure de 3-(4,5-diméthylthiazol-2-yl)-2,5-diphényl tétrazolium). Le second objectif a été de sélectionner sur la base des résultats obtenus au cours de ce criblage, une ou des souches fongiques présentant le plus grand intérêt dans la recherche de molécules naturelles bioactives.

3. Résultats et discussion

3.1. Criblage : évaluation des extraits fongiques marins sur la prolifération des cellules MCF-7

Afin d'étudier le potentiel de la collection de souches fongiques marines du laboratoire, un criblage a été réalisé sur 98 extraits bruts de champignons filamenteux marins appartenant à l'extractothèque du laboratoire MMS pour leurs activités potentielles dans le traitement du cancer du sein. Ainsi, l'activité des extraits bruts a été évaluée sur la prolifération des cellules cancéreuses mammaires MCF-7 via le test MTT.

3.1.1. Présentation des extraits testés

Les extraits utilisés au cours de ce criblage proviennent d'une collection d'extrait complétée au cours des différents projets menés au sein de notre laboratoire. Sur un total de 382 extraits disponibles, il a été sélectionné 98 échantillons dans le but de constituer un panel au sein duquel la diversité taxonomique des souches et l'originalité chimique des extraits seraient mises en évidence (Cf Annexe 2). Ces échantillons ont été en premier lieu sélectionnés en fonction de la quantité d'extrait déposée dans l'extractothèque. En effet, il était nécessaire d'avoir une quantité suffisante pour réaliser les différents essais biologiques souhaités.

Concernant la diversité taxonomique, les extraits ont été sélectionnés en fonction de l'espèce et/ou du genre de la souche utilisée lors de la culture. Les extraits choisis ont permis d'étudier des souches appartenant à 8 genres différents : *Aspergillus* sp., *Chrysosporium* sp., *Dydimella* sp., *Fusarium* sp., *Penicillium* sp., *Pycnidophora* sp., *Scopulariopsis* sp. et *Trichoderma* sp. L'extractothèque du laboratoire étant constituée principalement d'échantillons issus de souches appartenant au genre *Penicillium* sp., la sélection de 75 échantillons a permis d'étudier 15 espèces différentes au sein de ce même genre. Bien que ce type d'étude n'ait pu être réalisé sur les autres genres en raison d'un nombre d'extraits référencés trop faibles, il a été choisi de favoriser également la diversité chimique des échantillons. Ainsi, il a été sélectionné des extraits obtenus dans des conditions expérimentales différentes pour une seule et même espèce, c'est à dire selon une approche OSMAC (« One Strain Many Compounds ») (Bode and Bethe, 2002; Romano *et al.*, 2018) et/ou avec des techniques d'extractions diverses. L'ensemble des extraits testés figurent en annexe 2.

3.1.2. Résultats obtenus

Au cours de ce criblage sur cellules MCF-7 (Figure 14), nous avons pu montrer que parmi les 95 extraits testés à une concentration de 10 µg/mL pendant 24H sur les cellules cancéreuses MCF-7, 45 extraits présentaient un pourcentage d'inhibition de la croissance des cellules cancéreuses supérieur à 40%, avec une valeur d'inhibition maximale de 78%, obtenue pour deux d'entre eux : MMS231-100 et MM15-53 issus respectivement des souches *Penicillium polonicum* et *Penicillium antarticum*. Ceci peut être considéré comme un bon taux de réponse pour un test biologique réalisé sur un panel d'échantillons de cette envergure. Ce premier résultat représentait déjà une réponse positive quant à l'évaluation du potentiel de la collection du laboratoire sur une cible biologique en lien avec le cancer du sein.

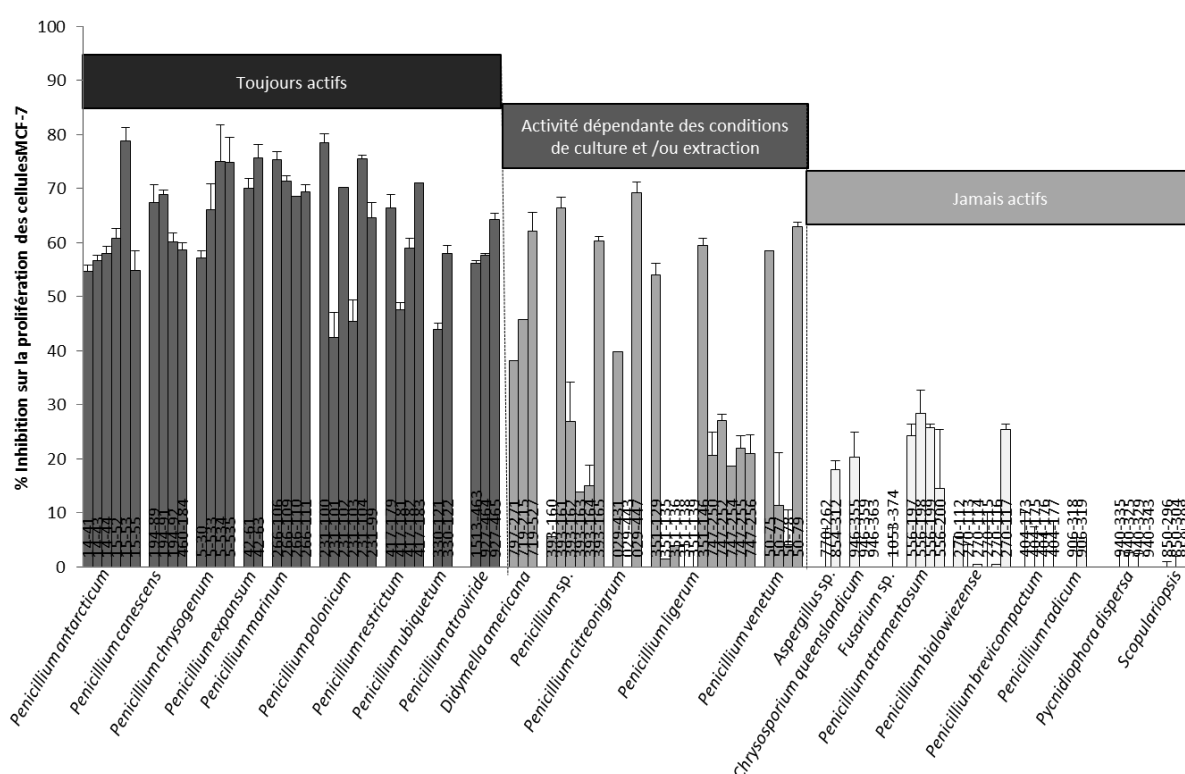


Figure 14. Effet antiprolifératif des extraits issus de 23 souches sélectionnées pour le criblage sur les cellules MCF-7. Les extraits ont été testés à la concentration de 10 µg/mL sur les cellules MCF-7 pendant 24H.

En s'intéressant de façon plus approfondie aux résultats obtenus, il a été constaté au cours de ce criblage que les 23 souches appartenant à la collection du laboratoire MMS pouvaient être classés en trois groupes selon les extraits testés : (1) les souches dont l'ensemble des extraits bruts étaient toujours actifs; (2) les souches dont l'activité des extraits dépendaient des conditions de culture et/ou des méthodes d'extraction; (3) les souches dont les extraits présentaient une inhibition de la prolifération inférieure à 50%.

Etonnamment, il a été observé que les extraits figurant dans le groupe 1, étaient tous issus de souches appartenant au genre *Penicillium*. En effet, quel que soit les conditions expérimentales utilisées pour l'étude des neuf espèces *P. antarcticum*, *P. canescens*, *P. chrysogenum*, *P. expansum*, *P. marinum*, *P. polonicum*, *P. restrictum*, *P. ubiquitum* et *P. atroviride*, les extraits présentaient la même activité. Il semblerait que les voies de biosynthèse conduisant à la production de métabolites bioactifs sur la prolifération des cellules MCF-7 pour ces souches, ne soient que très peu influencées par les milieux de cultures utilisés dans cette étude. Néanmoins, cette observation ne peut être généralisée au genre *Penicillium* sp., puisque les extraits issus des souches de *P. citreonigrum*, *P. ligierum*, *P. venetum* et d'une espèce non identifiée *Penicillium* sp. (appartenant au groupe 2), ont révélé quant à eux, l'existence d'une relation entre l'activité observée et le milieu de culture utilisé. Par exemple, parmi les 4 extraits bruts issus de la souche *P. venetum*, seulement ceux obtenus à partir de cultures réalisées sur milieux CYA et PDA ont démontré un effet antiprolifératif d'une valeur respective de 59% et 63% ; contrairement à ceux obtenus sur milieux DCA et MEA (inactifs). De la même manière, les extraits obtenus pour la souche *citreonigrum* cultivée sur milieux PDA, YES et MEA ont démontré des pourcentages d'inhibition de la prolifération cellulaire très différents (compris entre 0 à 70% d'inhibition). Par ces exemples, il a été démontré que le milieu de culture pouvait fortement influencer la production de métabolites bioactifs pour certaines des espèces fongiques appartenant au genre *Penicillium*. Ce phénomène a également été observé pour une souche appartenant à un genre différent, identifiée comme étant l'espèce *Dydimella americana*, anciennement appelée *Phoma americana*. En effet, seulement un des trois extraits testés, celui obtenu sur un milieu DCA, a révélé un effet antiprolifératif (> 50%) sur les cellules MCF-7 d'une valeur de 62% à 10 µg/mL.

Bien que ce premier essai ait révélé un grand nombre de souches produisant *a priori* des molécules bioactives ayant un effet sur la prolifération des cellules MCF-7, il faut considérer que de nombreuses molécules fongiques peuvent être toxiques et interagir sur différentes cibles biologiques sans aucune sélectivité ou spécificité. Il faut ainsi rester vigilant sur l'utilisation de ce type de test au cours d'un criblage, notamment dans la recherche de molécules anticancéreuses car cette toxicité pourrait s'appliquer aussi à des cellules saines.

Afin de pallier à ce manque de sélectivité du test biologique, il a été décidé d'y adjoindre les résultats d'un autre test biologique qui avait été réalisé dans la cadre d'une collaboration avec le laboratoire UiT The Arctic University of Norway (Tromsø, Norvège) sur l'activation des récepteurs LXR (Liver X Receptors) et/ou PPAR (Peroxisome Proliferator-Activated Receptor). Ces récepteurs nucléaires détaillés plus loin dans le chapitre 4, jouent un rôle important dans l'homéostasie lipidique. Il a également été démontré que leur activation peut avoir un rôle sur la prolifération cellulaire, le développement de la tumeur et l'induction de la mort cellulaire.

Ces derniers sont décrits comme des cibles potentielles pour la prévention et le traitement de plusieurs cancers incluant le cancer du sein (Gou *et al.*, 2017; Hu *et al.*, 2012b).

Par conséquent, l'expression des différents récepteurs nucléaires LXR et/ou PPAR évaluée par bioluminescence sur des cellules Cos-1 transfectées (Moldes-Anaya *et al.*, 2017) a permis de mettre en évidence et de confirmer le potentiel de 15 extraits issus de 6 souches : *Penicillium chrysogenum*, *Penicillium citreonigrum*, *Penicillium ubiquestum*, *Penicillium restrictum*, *Didymella americana* et *Pycnidophora dispersa* (Cf. Figure 15).

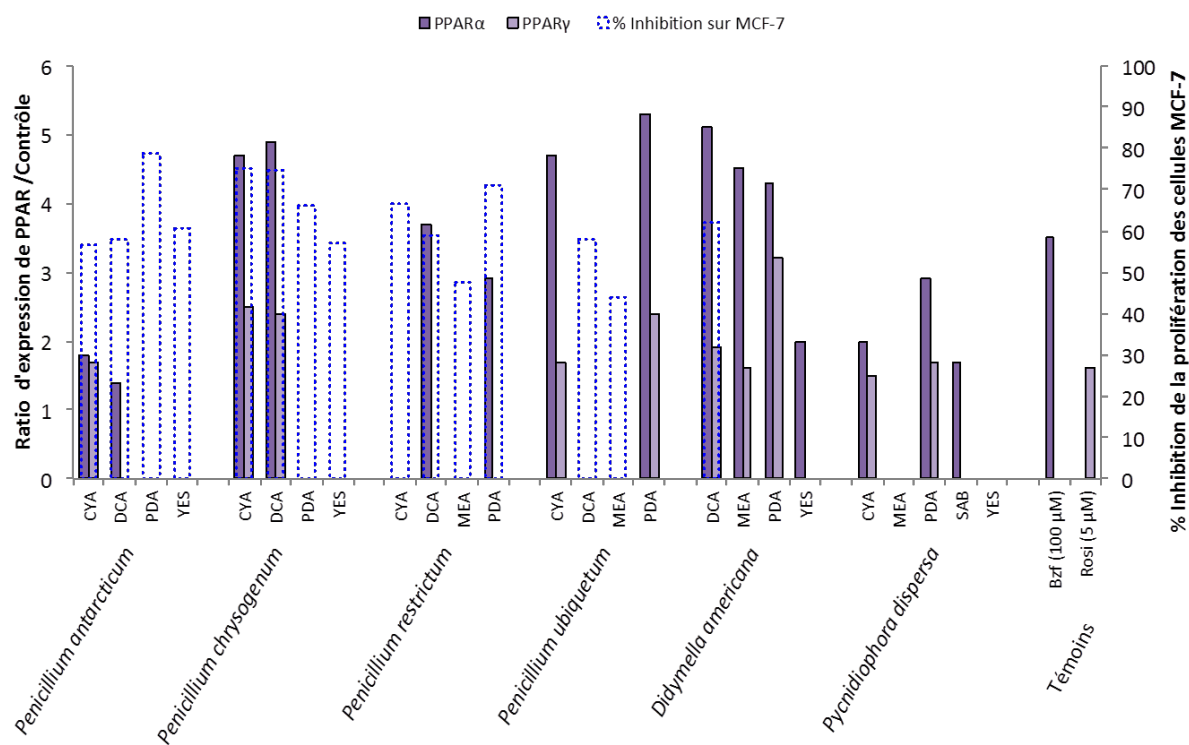


Figure 15. Souches dont les extraits obtenus sur différents milieux de culture présentent un effet sur l'augmentation de l'expression de PPAR de cellules Cos-1 transfectées. Les résultats de l'expression de PPARα et PPARγ obtenus pour les différents extraits testés à la concentration de 50 µg/mL sont exprimées par rapport au contrôle (cellules en présence de DMSO) ; les témoins positifs utilisés sont respectivement le bzf (100 µM) et Rosi (5µM) pour PPARα et PPARγ.

D'après les résultats obtenus, les extraits permettant d'augmenter fortement l'expression de PPAR (multiplié par 5 par rapport au contrôle) sont issus des souches *Penicillium chrysogenum*, *Penicillium ubiquestum* et *Didymella americana*. Pour l'ensemble des extraits issus de souches appartenant au genre *Penicillium* sp., il a pu être mis en évidence une relation entre l'activité observée et le milieu de culture utilisé. En effet, seuls les extraits obtenus sur milieux CYA et DCA sont actifs pour *P. antarcticum* et *P. chrysogenum*, seuls ceux obtenus sur DCA et PDA sont actifs pour *P. restrictum* et seuls ceux obtenus sur CYA et PDA sont actifs pour la souche *P. ubiquestum*.

En revanche, quel que soit le milieu de culture utilisé (DCA, PDA, MEA et YES) pour la souche *Didymella americana*, tous les extraits testés possédaient une activité sur PPAR. Parmi, ces extraits, celui obtenu sur PDA présentait à la fois une forte activité sur les deux isoformes de PPAR, ainsi qu'une inhibition de la prolifération des cellules MCF-7.

De manière surprenante, aucune activité dans les conditions testées n'a été obtenue pour LXR et aucun lien n'a été établi entre l'activité des différents extraits bruts sur les récepteurs nucléaires PPAR et l'inhibition de la prolifération cellulaire. Cependant, la sélection de souches dont les extraits avaient la capacité d'activer à la fois l'expression des récepteurs nucléaires PPAR tout en étant cytotoxique pour les cellules cancéreuses du sein, nous est apparue comme la meilleure stratégie dans le choix de souches d'intérêt pour l'isolement des composés bioactifs.

3.2. Sélection des souches fongiques à étudier

Au-delà des résultats obtenus au cours de ce criblage et de la stratégie adoptée dans la sélection des souches à étudier pour ce projet de thèse, il a également été pris en considération les travaux déjà réalisés au sein du laboratoire. En effet, certaines des souches présentant à la fois une forte activité sur l'inhibition de la prolifération des cellules MCF-7 et une activation sur l'un des récepteurs nucléaires PPAR, telles que les souches *Penicillium antarticum*, *Penicillium restrictum* et *Penicillium ubiquestum* étaient déjà étudiées pour des travaux de thèse (Hoang, 2018) . Ainsi, dans un souci d'originalité et selon la stratégie énoncée précédemment, les deux souches qui ont été sélectionnées pour la suite des travaux ont été : *Penicillium chrysogenum* et *Didymella Americana*, répertoriées respectivement dans la mycothèque sous les codes MMS5 et MMS719. En effet, celles-ci présentaient toutes deux, un fort pourcentage d'inhibition de la prolifération des cellules MCF-7 d'environ 65% à une concentration de 10 µg/mL, ainsi qu'une forte augmentation de l'expression des récepteurs PPAR α (> 4,5) et PPAR γ (> 2).

Parmi les différents extraits testés, il a été mis en évidence une relation entre l'activité et les milieux de cultures utilisés pour les deux souches, mais cette observation ne s'applique pas pour le même essai biologique. En effet, tous les extraits issus de la souche *P. chrysogenum* possédaient une activité sur la prolifération des cellules MCF-7, alors que seulement ceux obtenus sur milieu CYA et DCA permettaient d'augmenter l'expression de PPAR. A l'inverse, tous les extraits obtenus pour la souche *D. americana* possédaient une activité sur l'expression de PPAR mais seulement celui obtenu sur milieu DCA permettait d'inhiber la prolifération des cellules MCF-7. Par conséquent, ces souches ne produisaient respectivement, probablement pas les mêmes métabolites en fonction des différentes conditions de culture. En effet, l'activation ou la répression de certaines voies de biosynthèse pour ces souches semble être régulée par l'utilisation de différents milieux de culture confirmant l'approche OSMAC décrite précédemment.

Ainsi, d'après les résultats obtenus, il a été choisi d'utiliser le milieu de culture DCA, seul milieu ayant permis d'avoir une double activité sur les essais biologiques pour la souche *D. americana*.

En revanche, pour *P. chrysogenum*, bien que les extraits aient été analysés par chromatographie liquide haute performance couplée à un spectromètre de masse, la superposition de leurs profils métaboliques obtenus par analyse LC-MS n'a permis de mettre en évidence une importante variabilité dans leurs compositions chimiques. Il a donc été choisi d'utiliser le milieu de culture CYA étant moins conventionnel au sein du laboratoire.

Un travail bibliographique a été réalisé sur les deux souches afin d'évaluer leur potentiel dans le recherche de métabolite bioactives.

3.2.1. Présentation et évaluation du potentiel de la souche MMS5 *Penicillium Chrysogenum*

3.2.1.1. Classification taxonomique et distribution de l'espèce *P. chrysogenum*

La souche *Penicillium chrysogenum* appartient au genre *Penicillium* sp. et a été créé en 1809 par Link. Son nom issu du latin « penicillus » qui veut dire « pinceau », fait référence à la manière dont s'organisent ses structures végétatives. Ce sont des champignons filamenteux, de type moisissure présents dans tous les environnements (sols, air, denrées alimentaires, matières organiques ou végétales en décomposition).



Synonymes :

- Penicillium brunneorubrum* Dierckx, Annales de la Société Scientifique de Bruxelles 25 (1): 88 (1901) [MB#150025]
- Penicillium griseoroseum* Dierckx, Annales de la Société Scientifique de Bruxelles 25 (1): 86 (1901) [MB#152968]
- Penicillium baculatum* Westling, Svensk Botanisk Tidskrift 4: 139 (1910) [MB#247810]
- Penicillium notatum* Westling, Arkiv för Botanik 11 (1): 95 (1911) [MB#160571]
- Penicillium chlorophaeum* Biourge, La Cellule 33: 271 (1923) [MB#260543]
- Penicillium cyaneofulvum* Biourge, La Cellule 33: 174 (1923) [MB#262580]
- Penicillium meleagrinum* Biourge, La Cellule 33: 147 (1923) [MB#270843]
- Penicillium roseocitreum* Biourge, La Cellule 33: 184 (1923) [MB#276775]
- Penicillium flavidomarginatum* Biourge, La Cellule 33: 150 (1923) [MB#265031]
- Penicillium fluorescens* Laxa, Zentralblatt für Bakteriologie und Parasitenkunde, Abteilung 2 86 (5-7): 160-165 (1932) [MB#492615]
- Penicillium camerunense* R. Heim, Bull. Acad. R. Belg. Cl. Sci.: 42 (1949) [MB#335718]
- Penicillium aromaticum* f. *microsporum* Romankova, Uchen. Zap. Leningr. Univ. Zhadanov: 102 (1955) [MB#347366]
- Penicillium harmonense* Baghd., Novosti Sistematiki Nizshikh Rastenii 5: 102 (1968) [MB#335733]

3.2.1.2. Molécules décrites chez *P. chrysogenum*

Bien que cette espèce soit très étudiée depuis la découverte de la pénicilline G (premier antibiotique à large spectre utilisé en thérapeutique humaine) par Alexander Fleming en 1928, de nouvelles molécules sont encore continuellement isolées de cette espèce. Un recensement des métabolites produits par *Penicillium chrysogenum* a été réalisé dans le cadre de cette thèse afin de prendre connaissance, des métabolites pouvant déjà être décrits dans la littérature. Ainsi, environ 200 molécules d'une grande diversité chimique ont pu être référencées (Cf. Annexe 3) avec pour certaines d'entre elle des activités prometteuses notamment sur différentes lignées cancéreuses. Parmi ces molécules, des lactones (Butyrolactone V), des alcaloïdes (Chaetoglobosine (Huang *et al.*, 2016; Luo *et al.*, 2013)) ainsi que des stérols (ergosterol-5,8-endoperoxyde (Han *et al.*, 2014; He *et al.*, 2018; Russo *et al.*, 2010)). Bien qu'il soit difficile d'évaluer le potentiel biologique de chacune des molécules, les articles décrivant de nouvelles structures et/ou de nouvelles activités témoignent du fort potentiel de cette espèce dans la découverte de nouvelle molécule d'origines naturelles.

3.2.2. Présentation et évaluation du potentiel de la souche MMS719 *D. americana*

3.2.2.1. Classification taxonomique et distribution de l'espèce *D. americana*

La souche *Didymella americana* est référencée comme étant un pathogène de plantes et plus particulièrement d'haricots *phaseolus lunatus* (Gorny *et al.*, 2016) ou de betteraves (Vaghefi *et al.*, 2016).



Synonymes :

Phoma americana Morgan-Jones & J.F. White, Mycotaxon 16 (2): 406 (1983)
[MB#109173]

Peyronellaea americana (Morgan-Jones & J.F. White) Aveskamp, Gruyter & Verkley,
Studies in Mycology 65: 31 (2010) [MB#515596]

3.2.2.2. Molécules décrites chez *D. americana*

Pour l'espèce *Didymella americana*, contrairement à celle de *P. chrysogenum*, aucune étude n'a été rapportée dans la littérature scientifique pour une souche d'origine marine. Ce premier constat, renforcé par l'absence de hit obtenu dans les bases de données de molécules telles que SciFinder (<https://scifinder.cas.org>) ou le DNP (<http://dnp.chemnetbase.com>) pour une recherche effectuée sur cette espèce (quel que soit son origine), a permis de mettre en évidence le fort potentiel de cette espèce dans la découverte de nouvelles molécules. Cependant, en sachant que l'identification des espèces n'étant pas toujours réalisée, le recensement des molécules bioactives a été élargi au genre *Didymella* tout entier ainsi qu'aux différents synonymes pouvant être retrouvés dans la littérature (*Peyronellaea* et *Phoma*). Ainsi, plus de 400 molécules ont été référencées (non présentées dans ce manuscrit), dont certaines ont été reportées pour leur activité (Rai *et al.*, 2018).

4. Conclusion

Cette première étape de sélection des souches réalisée au cours d'un criblage, nous a permis d'évaluer le potentiel de la mycothèque du laboratoire MMS dans la production de métabolites ayant une activité sur une cible biologique en lien avec le cancer du sein. Parmi les souches étudiées, 14 souches ont pu être mises en évidence pour leur effet (correspondant à 73 extraits) sur l'inhibition de la prolifération des cellules cancéreuses MCF-7. Afin d'éviter de sélectionner des souches produisant des métabolites cytotoxiques ne possédant pas de sélectivité sur une cible biologique, il a été choisi d'utiliser les résultats d'un second test biologique. Celui-ci réalisé lors d'une collaboration avec l'Université de Tromsø, a permis d'évaluer l'effet des extraits sur l'activation des récepteurs nucléaires LXR et ou PPAR. Ceci a permis d'affiner le choix dans la sélection des souches, puisque les souches dont les extraits présentaient les deux activités citées précédemment, a été réduit à seulement 6 souches (correspondant à 15 extraits). En s'appuyant sur les travaux réalisés au sein du laboratoire et l'étude bibliographique des espèces, il a donc été choisi de sélectionner pour notre étude les souches *Penicillium chrysogenum* et *Didymella americana*.

Il est à noter que la sélection de ces deux souches a été réalisée selon un positionnement stratégique différent. En effet, la souche *P. chrysogenum*, déjà connue et décrite dans la littérature pour son fort potentiel dans la production de métabolites bioactifs présentait moins de risques à être utilisée dans la découverte de molécules bioactives. Toutefois, le potentiel de nouveauté des molécules identifiées était plus limité, avec tout de même l'avantage de pouvoir bénéficier d'une bonne connaissance des squelettes carbonés existant dans cette espèce en vue de l'élucidation de structures potentiellement nouvelles. Au contraire, la souche *Didymella americana* étant très peu étudiée, a été sélectionnée pour accroître la ou les possibilités de découvrir de nouvelles molécules. C'est donc l'originalité des résultats attendus qui a été promue dans ce cas, avec le risque d'isoler finalement peu de métabolites spécialisés ou difficiles à élucider, mais ce risque est tout de même assez limité au regard des nombreuses molécules rapportées pour l'ensemble du genre.

Après avoir sélectionné les deux souches, un choix a dû être réalisé concernant le milieu de culture à utiliser. Dans le cas de la souche *P. chrysogenum*, les deux milieux de culture ont démontré un intérêt mais nous avons choisi de sélectionner le milieu CYA étant moins conventionnel dans son utilisation au sein du laboratoire.

Pour la souche *Didymella americana*, il a été plus facile de choisir le milieu de culture à utiliser puisque le seul extrait testé démontrant une activité (> 50%) sur les deux essais biologiques correspondait à un extrait obtenu après culture de la souche sur milieu DCA.

5. Matériel et Méthodes

5.1. Description de l'extractothèque

Parmi les 98 extraits bruts testés au cours du screening, la majorité d'entre eux (74) sont issus de souches appartenant au genre *Penicillium*. Au sein de ce genre, 15 espèces fongiques ont été étudiées : *Penicillium antarcticum* (9), *Penicillium atramentosum* (4), *Penicillium bialowiezense* (6), *Penicillium brevicompactum* (4), *Penicillium canescens* (4), *Penicillium chrysogenum* (6), *Penicillium citreonigrum* (3), *Penicillium expansum* (3), *Penicillium ligerum* (11), *Penicillium marinum* (4), *Penicillium polonicum* (6), *Penicillium radicum* (2), *Penicillium restrictum* (4), *Penicillium ubiquetum* (4) et *Penicillium venetum* (4). Le reste des échantillons provenaient de cultures réalisées à partir de souches appartenant aux espèces *Chrysosporium queenslandicum* (3), *Dydimella americana* (5), *Pycnidophora* sp. (6), *Trichoderma atroviride* (3) ou encore d'espèces non identifiées appartenant aux genres *Aspergillus* sp. (2), *Fusarium* sp. (1) et *Scopulariopsis* sp. (3). Toutes ces souches ont été isolées en milieu marin et l'ensemble des extraits obtenus à partir de celles-ci ainsi que les conditions opératoires utilisées lors des cultures et l'extraction figurent dans l'annexe 2.

5.1.1. Les milieux utilisés pour la culture des souches

Bien que les champignons filamenteux puissent être cultivés en milieu solide ou liquide de manière statique ou agitée, les cultures fongiques étudiées au cours de ces travaux ont toutes été effectuées sur milieu solide gélosé dans des Erlenmeyers ou des boîtes de pétri, en conditions statiques. Différents milieux gélosés décrits dans le tableau ci-dessous (Cf. Tableau 1) ont été utilisés. Parmi ces milieux, 6 sont classiquement utilisés (KMS, MEA, CYA, PDA, YES, MES) et un milieu plus original (MES à base d'extrait de moules *Mytilus edulis*) développé au laboratoire afin d'étudier plus particulièrement l'influence de métabolites de moules sur la production de composés bioactifs par des souches isolées de coquillages (Geiger *et al.* 2013). Deux conditions osmotiques pour les milieux ont été utilisées : l'une utilisant de l'eau distillée (ED) et l'autre de l'eau de mer synthétique (EDMS) ou naturelle (EDM).

Tableau 1. Composition des différents milieux de cultures

	KMS Kohlmeyer modifié solide	MEA solideExtrait de malt - agar	PDA Pomme de terre - dextrose - agar	DCA Dextrose - extrait de caséine - agar	CYA Czapek-extrait de levure-agar	YES Extrait de levure - sucrose - agar	MES Extrait de moule - sucrose - agar
CuSO ₄ ·5H ₂ O		0,005g	0,005g		0,005g	0,005g	0,005g
FeSO ₄ ·7H ₂ O					0,010g		
MgSO ₄ ·7H ₂ O	2,400g				0,500g	0,500g	0,500g
ZnSO ₄ ·7H ₂ O		0,010g	0,010g		0,010g	0,010g	0,010g
KCl					0,500g		
K ₂ HPO ₄					1g		
NaNO ₃					3g		
NH ₄ NO ₃	2,400g						
Tampon Tris (R-NH ₂)	1,210g						
Glucose (=dextrose) C ₆	5g	20g	20g	40g			
Saccharose (= sucrose) C ₁₂					30g	150g	150g
Peptone		1g					
Digestion enzymatique de caséine				10g			
Extrait de levure					5g	20g	
Extrait de moule							20g
Extrait de malt		20g					
Extrait de pomme de terre			4g				
Agar	20g	20g	15g	15g	15g	20g	20g
Eau	1L	1L	1L	1L	1L	1L	1L

Afin de faciliter la préparation de ces milieux, les mélanges ont été solubilisés à l'aide d'un autoclave (1 min à 105 °C) avant d'être coulés à chaud dans des Erlenmeyer de 250 mL à raison de 50 mL de milieu par Erlenmeyer. Ils ont ensuite été bouchés à l'aide de cotons cardés puis recouverts d'aluminium avant d'être stérilisés par autoclavage (20 min à 121 °C).

5.1.2. Ensemencement et incubation des cultures fongiques

A partir de prélèvements de fragments de mycélium ou de spores issus de souches ayant été ensemencées quelques jours sur milieu DCA en boîte de Pétri, des inoculations ont été réalisées dans des Erlenmeyers ou des boîtes de pétri dont le nombre dépend des études menées. Les cultures ont ensuite été incubées dans une étuve à 27 °C pour une période dont la durée correspond au temps nécessaire de développement de la souche sur l'ensemble de surface de la gélose, favorisant la production d'une biomasse suffisante.

5.1.3. Extraction de la biomasse fongique

La biomasse fongique est classiquement extraite au laboratoire selon une technique d'extraction solide-liquide réalisée sur la totalité de la culture fongique (Gélose et mycélium), en présence d'un mélange de solvants AcOEt/CH₂Cl₂ (Cf. Chapitre 2§5.2). Cependant, pour certains projets, l'extraction a été réalisée sur une partie spécifique (mycélium ou gélose) et/ou selon d'autres méthodes (microextraction en présence de CH₂Cl₂ (Smedsgaard), extraction réalisée en présence de MeOH puis CH₂Cl₂, ou uniquement avec le solvant AcOEt).

5.1.4. Profilages HPLC-UV-HRMS

Les analyses des extraits bruts ont été réalisées à l'aide d'une HPLC couplée à un spectromètre de masse haute résolution (HRMS). Pour cette étude l'appareil utilisé était une LC-MS-IT-TOF (Ion Trap - Time of Flight) de la marque Shimadzu, avec une source d'ionisation de type electrospray (ESI). L'appareil de chromatographie liquide est composé de deux pompes LC-20ADxr, d'un passeur d'échantillons SIL-20ACxr, d'un four à colonne CTO-20AC et d'un contrôleur MBC-système 20A. En plus de la spectrométrie de masse, l'absorption UV-visible en sortie de colonne est également mesurée à l'aide d'un détecteur à barrettes de diodes SPD-M20A. Les analyses de chromatographie liquide haute performance ont été effectuées sur une colonne Kinetex™ C₁₈ (100 x 2,1 mm, 2,6 µm), chauffée à 40 °C, avec une phase mobile constituée d'un mélange CH₃CN + 0,1% AF/H₂O + 0,1% AF selon le gradient d'élution suivant présenté dans le tableau ci-contre (Cf. Tableau 2).

Tableau 2. Conditions d'élution pour le profilage HPLC-UV-HRMS

Temps (min)	% H ₂ O + 0,1% AF	% ACN + 0,1% AF
0	85	15
2	85	15
25	0	100
30	0	100
31	85	15
35	85	15

Pour chaque analyse d'une durée de 35 min avec un débit de 0,3 mL/min, un volume d'injection de 5 µL des différents échantillons préparés dans du méthanol de grade UPLC-MS (Biosolve) à la concentration de 0,4 mg/mL pour les extraits bruts a été utilisé. Les échantillons ont été maintenus à 4 °C durant la totalité de l'analyse.

5.2. Présentation de la cible biologique étudiée : prolifération des cellules cancéreuses mammaires

5.2.1. La lignée cancéreuse mammaire MCF-7

La lignée cancéreuse mammaire MCF-7 exprime les récepteurs aux œstrogènes ce qui entraîne une réponse de prolifération aux œstrogènes. On parle alors d'une lignée ER+ (Estrogen Receptors positive) estrogéno-dépendante. *In vitro*, ces cellules sont adhérentes et se présentent phénotypiquement sous forme de fuseau. Les cellules MCF-7 ont été maintenues dans le milieu de culture Dulbecco's Modified Eagle's Medium (DMEM) supplémenté avec du sérum de veau fœtal 10% (SVF), 1% de L-glutamine et 1% des antibiotiques Pénicilline/Streptomycine sous une atmosphère humide de 37°C avec 5% de CO₂.

5.2.2. Évaluation de l'activité biologique par le test de cytotoxicité MTT

Le test colorimétrique du MTT, dont le nom provient du réactif utilisé, à savoir le sel de tétrazolium (ou bromure de 3-(4,5-diméthylthiazol-2-yl)-2,5-diphenyl tetrazolium (réf : M5655-16, Sigma), a été utilisé pour étudier les effets de différents traitements sur la croissance des cellules. Le sel de tétrazolium (couleur jaune) réduit par la succinate déshydrogénase mitochondriale des cellules vivantes actives en formazan conduit à la formation d'un précipité (couleur violette) dont la quantité est proportionnelle à celle de cellules vivantes (mais également à l'activité métabolique de chaque cellule).

Protocole

Les cellules MCF-7 ont étéensemencées dans une plaque 96 puits avec une densité de 10 000 cellules par puits. Après 24H d'incubation, les milieux de cultures pour les lignées MCF-7 ont été remplacés par les différents traitements préparés dans un milieu DMEM supplémenté (0,1% BSA, 1% Pénicilline, 1% Streptomycine, 1% Glutamine). Pour chaque concentration, trois puits ont été réalisés. Passé la période d'incubation de 24H à 37 °C, 100 µL de milieu sont retirés, puis 50 µL de MTT (2,5 mg/mL dilué dans le PBS) sont ajoutés dans chacun des puits. Après une période d'incubation de 4H à 37 °C, les puits sont vidés et 200 µL de DMSO (Réf : D8418, Sigma) sont ajoutés dans chaque puits afin de solubiliser les cristaux violets de formazan. La lecture se fait à 570 nm à l'aide d'un spectrophotomètre Spectra MAX 190, Molecular Devices. L'inhibition de la prolifération des cellules traitées est exprimée en pourcentage par rapport à celle des cellules sans traitement (contrôle) selon la formule suivante : $\% \text{ Inhibition} = 100 - \left(\frac{DO_{\text{Traitement}}}{DO_{\text{Contrôle}}} \right)$.

5.2.3. Évaluation de l'activation des récepteurs nucléaires LXR et/ou PPAR

L'effet des différents extraits a été évalué à une concentration de 50 µg/mL sur l'expression des récepteurs nucléaires LXR et PPAR. Les essais ont été réalisés en collaboration avec le laboratoire UiT The Arctic University of Norway (Tromsø, Norvège) sur des cellules COS-1 (cellules de foie issu de singe) transfectées de manière transitoire avec les plasmides d'expressions comprenant le domaine de liaison du ligand (LBD) de PPAR (α, β et γ) ou LXR (α et β) couplé au domaine de liaison à l'ADN (DBD) du facteur de transcription de levure GA ; ainsi que le gène rapporteur de la luciférase. Ce qui a permis par la réaction de bioluminescence catalysée par la luciférase de mesurer directement l'activité liée à l'expression des gènes d'intérêts (Moldes-Anaya *et al.*, 2017).

Chapitre 2 :

Investigation chimique des souches
sélectionnées

Suite au criblage mené dans le chapitre précédent, nous nous sommes focalisés sur l'étude des métabolites produits par les deux espèces fongiques : MMS5 *P. chrysogenum* et MMS719 *D. americana*. Un des challenges dans le domaine des produits naturels est celui d'isoler des molécules bioactives issues de mélanges complexes. Ainsi, plusieurs stratégies ont été utilisées et comparées lors de cette étude telle que le fractionnement bioguidé ou des techniques plus récentes comme les approches de déréplication. Cette seconde approche est basée sur l'acquisition de données chimiques obtenues dans notre cas par chromatographie liquide couplée à un détecteur UV et un spectromètre de masse haute résolution. Une introduction bibliographique a été réalisée dans le but de rappeler brièvement les différentes approches utilisées dans le cadre de cette étude.

1. Introduction bibliographique

1.1. Définition, intérêt et limites du fractionnement bioguidé

L'approche la plus classiquement utilisée dans l'isolement de molécules bioactives à partir d'un extrait d'origine naturelle est celle du fractionnement bioguidé (Ferchichi *et al.*, 2012; Pandey *et al.*, 2018; Pieters and Vlietinck, 2005; Weller, 2012). Elle consiste à évaluer l'activité des fractions obtenues au cours des différentes étapes de fractionnement de cet extrait dans le but d'orienter la purification vers l'isolement des composés d'intérêts, c'est à dire ceux responsables de l'activité observée (Cf. Figure 16). Par conséquent, il est nécessaire de choisir une cible biologique dont l'évaluation expérimentale de l'activité soit facile et rapide à déterminer, ainsi que suffisamment sensible pour ne pas utiliser une quantité trop importante de produit (Beutler *et al.*, 1990; Cos *et al.*, 2006; Suffness *et al.*, 1989).

Cette approche permet d'orienter l'isolement vers une ou plusieurs molécules dont l'activité sur la cible choisie est certaine. Toutefois, c'est une pratique qui peut se révéler longue et coûteuse si le nombre d'étapes de fractionnement et d'essais biologiques devient trop important (Weller, 2012). De plus, avec un nombre croissant de produits naturels décrits au fil du temps, cette méthode a de plus en plus de chances de conduire à l'isolement de molécules sans originalité *in fine*, c'est à dire dont la structure et/ou l'activité biologique sont déjà décrites dans la littérature. En effet, cette approche ne permet en aucun cas d'évaluer l'originalité structurale d'une molécule sans l'avoir purifiée puis identifiée par analyse spectrale RMN et/ou MS, ce qui peut entraîner un repositionnement stratégique au cours de l'étude. Par ailleurs, certaines limitations ont pu être rapportées dans la littérature, telles que l'observation d'une perte d'activité observée au cours du fractionnement bioguidé pouvant s'expliquer par une dégradation, une instabilité des molécules ou encore la présence d'effets synergiques au sein de l'extrait (qui peuvent être perdus au cours du fractionnement).

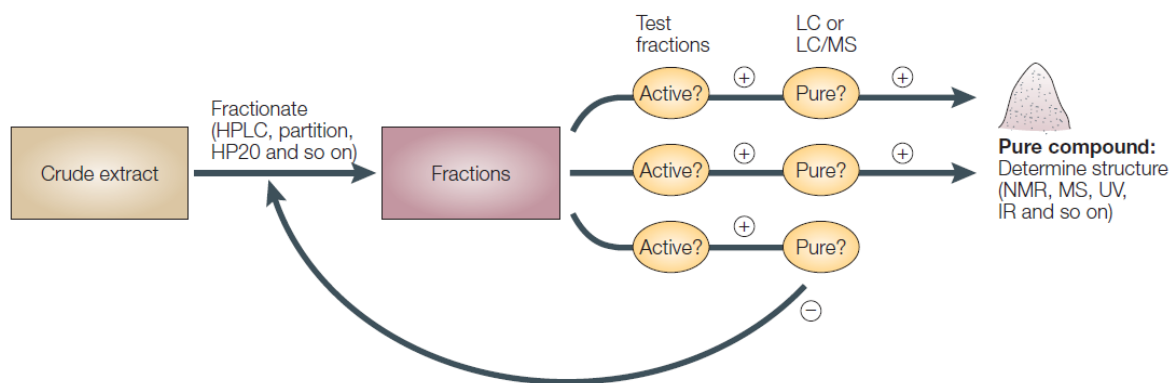


Figure 16. Méthodologie du fractionnement bioguidé (Koehn and Carter, 2005).

Ces limites du fractionnement bioguidé ont conduit plusieurs équipes à chercher des méthodes alternatives et complémentaires pour accéder à des résultats innovants dans le domaine de la chimie des produits naturels (Van Beek *et al.*, 2009). C'est dans ce contexte que sont développées des méthodes d'évaluation précoce du potentiel d'extraits bioactifs, avant même de purifier leurs constituants, plus largement regroupées sous le concept de « déréplication ».

1.2. Que signifie le terme « Déréplication » ?

La déréplication est une approche assez récente dans la recherche de nouveaux produits d'origine naturelle. Ce concept apparu au début des années 90, est défini à l'origine comme le fait d'éviter la réplication des résultats, notamment une sur une cible biologique. Par exemple, elle peut être utilisée dans la recherche de produits naturels bioactifs afin d'éviter de ré-isoler deux fois la même molécule responsable de l'activité recherchée et évaluée, ou d'isoler une molécule dont l'activité était déjà décrite dans la littérature (Beutler *et al.*, 1990; Cardellina *et al.*, 1993; Yarbrough *et al.*, 1992). Il faut donc être en capacité d'identifier que les résultats obtenus au cours d'un projet sont identiques ou similaires à des résultats qui ont déjà été obtenus (intra et/ou inter projet). Cette similitude peut correspondre par exemple à l'isolement d'une molécule appartenant à une classe chimique déjà connue pour une activité biologique donnée.

Ainsi, en fonction des objectifs définis, différentes approches de déréplication ont pu être proposées : 1) soit par étude des propriétés physico-chimiques des molécules présentes, repérage de chimiotypes particuliers (connus pour être actifs sur la cible choisie), et comparaison à celles déjà connues, 2) soit par comparaison de profils d'activités biologiques, 3) soit par combinaison des deux. Cependant, avec le développement des outils chromatographiques et d'analyses conduisant à l'acquisition de données de plus en plus précises et avec une sensibilité accrue, l'utilisation de ce terme a quelque peu évolué vers le concept « d'annotation ». En effet, il est aujourd'hui possible de caractériser ou d'identifier les différents composés présents au sein d'un mélange complexe, à partir de données spectrales obtenues principalement par LC-MS ou RMN.

Ces données sont comparées à celles de bases de données répertoriant l'ensemble des molécules décrites dans la littérature, afin d'identifier les molécules connues en présence. Cependant, il peut y avoir dans certains cas plusieurs correspondances dans les structures proposées (plusieurs « hits »).

Afin d'accélérer le processus de déréplication, de nombreuses bases de données spectrales ou de formules brutes sur les molécules ont été développées telles que PubChem (environ 96 millions de composés) (Kim *et al.*, 2019), ChemSpider (75 millions) (Pence and Williams, 2010), ChEMBL (1,9 millions) (Mendez *et al.*, 2019), ChEBI (56 000) (Hastings *et al.*, 2016), DNP (300 000)(<http://www.dnp.chemnetdatabase.com>), KEGG (18 600) (Kanehisa *et al.*, 2017), MetaCyc (15 655) (Caspi *et al.*, 2018), NuBBE_{DB} (2500) (Pilon *et al.*, 2017), MIBiG (1727) (Chem and Author, 2015), DrugBank (13 344) (Wishart *et al.*, 2018), et NORINE (1 191 composés) (Flissi *et al.*, 2016), mzCloud (8 839 composés et 5 millions de spectres LC-MS) (<http://www.mzcloud.org>), GNPS (74 000 spectres MS²) (Wang *et al.*, 2016).

2. Objectifs de l'étude

Après une étape de criblage ayant permis de sélectionner les souches *Penicillium chrysogenum* (MMS5) et *Dydimella americana* (MMS719) pour leur activité cytotoxique sur la lignée cancéreuse mammaire MCF-7, celles-ci ont été cultivées à grande échelle afin de pouvoir mener les investigations chimiques nécessaires à l'identification des molécules responsables de l'activité observée. Ainsi, après vérification de l'effet des nouveaux extraits bruts obtenus sur l'inhibition de la prolifération des cellules cancéreuses mammaires MCF-7, la recherche des métabolites bioactifs pour ces deux souches a été réalisée selon différentes approches : fractionnement bioguidé et déréplication. Les différentes approches ayant mené à la purification ou à l'annotation de molécules produites pour ces deux souches, sont alors décrites dans ce chapitre. Par ailleurs, l'effet des molécules isolées sur la prolifération des cellules cancéreuses mammaires a également été évalué.

3. Résultats discussions

3.1. Obtention des extraits fongiques et évaluation de leurs activités cytotoxiques

Afin de réaliser une investigation plus approfondie des métabolites produits par les deux souches sélectionnées : *Penicillium chrysogenum* et *Didymella americana*, celles-ci ont été cultivées à grande échelle sur les milieux respectifs CYA et DCA (Cf.§5.1). Après une période d'incubation de 12 jours de la souche *P. chrysogenum*, l'extraction des métabolites a permis d'obtenir deux extraits bruts d'une masse de 818,4 mg et 484,9 mg, référencés sous les codes MMS5-526 et MMS5-531 dans l'extractothèque du laboratoire. Pour la souche *D. americana*, l'extraction réalisée après une période d'incubation de 21 jours, a permis d'obtenir trois extraits MMS719-527 (1,1644 g), MMS719-529 (659,9 mg) et MMS719-530 (345,5 mg).

On remarque que la période d'incubation entre les deux souches diffère en raison d'une vitesse de croissance beaucoup plus lente pour la souche *D. americana*. L'aspect macroscopique des deux souches figure ci-dessous (Cf. Figure 17).

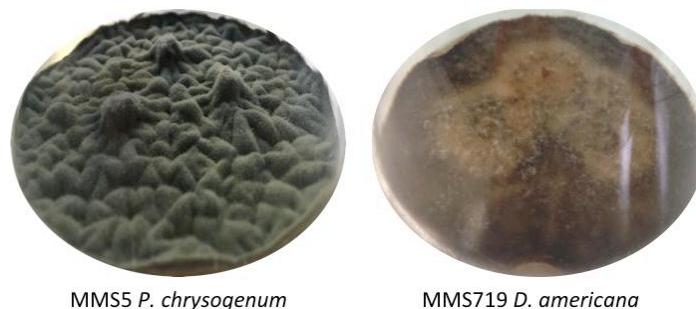


Figure 17. Photographie des cultures en Erlenmeyer des souches MMS5 et MMS719. Ces photographies ont été réalisées après une période d'incubation de 12 et 21 jours, respectivement.

Dans le but de vérifier l'activité antiproliférative de ces nouveaux extraits bruts obtenus, un traitement de 24H dans la gamme de concentration de 0,001 à 10 µg/mL a été réalisé sur les cellules cancéreuses mammaires MCF-7. D'après les résultats obtenus selon un test MTT (Cf. Figure 18), les deux extraits issus de la souche MMS5 *P. chrysogenum* présentent tous deux une activité sur l'inhibition de la prolifération cellulaire avec une valeur respective de 75 et 61% à la concentration de 10 µg/mL. Un effet dose-réponse a été déterminé pour l'extrait MMS5-526 avec une $CI_{50} = 0,071 \pm 0,022$ µg/mL. Pour la souche MMS719 *D. americana*, les résultats obtenus pour les trois extraits testés ne présentent pas la même activité. En effet, seul les extraits MMS719-527 et MMS719-530 sont actifs sur le modèle cellulaire MCF-7 avec une valeur d'inhibition respective de 60 et 53% à la concentration de 10 µg/mL. La différence d'activité observée entre les trois lots souligne ainsi la nécessité de procéder à la vérification de l'activité d'un nouvel extrait obtenu pour une même souche. En effet, malgré une standardisation des procédés de culture, les organismes fongiques restent des organismes vivants, dont la production de métabolites peut varier et évoluer sous l'influence de facteurs qui ne sont pas toujours contrôlables et identifiés. Par ailleurs, la courbe obtenue pour l'extrait MMS719-527 (Cf. Figure 18) ne permet pas de déterminer une valeur de CI_{50} unique.

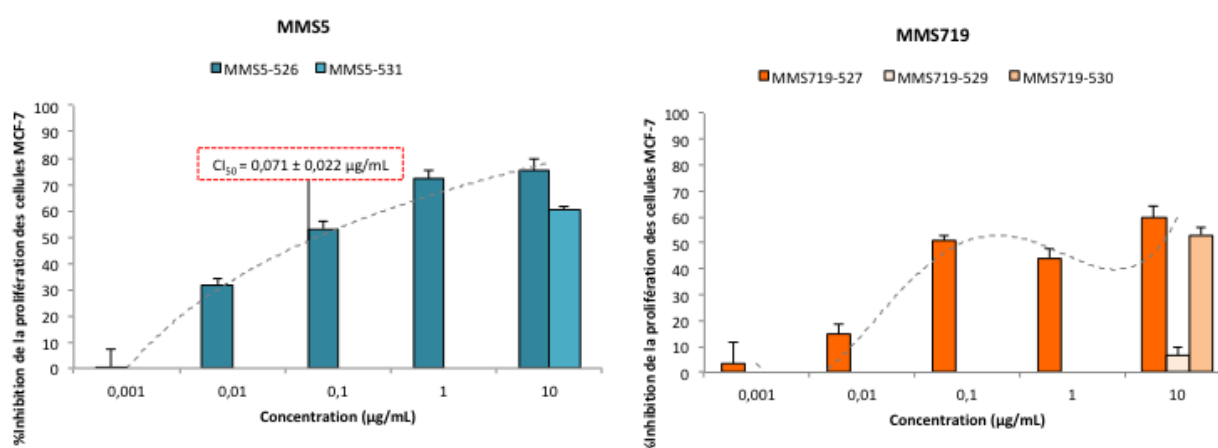


Figure 18. Effet des extraits bruts obtenus pour la souche MMS5 *P. chrysogenum* et MMS719 *D. Americana* sur la prolifération des cellules MCF-7.

3.2. Étude de la souche *Penicillium chrysogenum*

Dans le cadre de ce projet, la souche MMS5 *P. chrysogenum* a tout d'abord été étudiée dans une approche classique de fractionnement bioguidé afin d'isoler le ou les produits responsables de l'activité. Par la suite, grâce à un profilage LC-HRMS des extraits et un travail de déréplication, certaines des molécules présentes dans l'extrait ont pu être annotées. Dans certains cas, en l'absence de correspondance dans les bases de données, de nouvelles molécules potentielles ont pu être mises en évidence. Elles ont fait l'objet d'une purification ciblée guidée par la spectrométrie de masse.

Les différentes étapes de fractionnement ayant mené à l'identification de molécules obtenues par fractionnement bioguidé et/ou ciblée ont été décrites dans ce chapitre. Par ailleurs, l'effet des molécules isolées sur la prolifération des cellules cancéreuses mammaires a également été évalué.

3.2.1. Fractionnement bioguidé

Dans la recherche des molécules responsables de l'activité par fractionnement bioguidé, différentes étapes de purification ont été réalisées sur l'extrait MMS5-526 issu de la souche *P. chrysogenum*.

Dans un premier temps, un fractionnement de l'extrait brut MMS5-526 réalisé à l'aide d'un système de chromatographie liquide basse pression (flash chromatographie) a permis d'obtenir 17 fractions. L'ensemble de ces fractions (1-1 → 1-17) a été testé sur notre modèle cellulaire MCF-7 via un test MTT. En parallèle, ces mêmes échantillons ont été évalués en collaboration avec le laboratoire INSERM UMR 1069, Nutrition Croissance Cancer, selon une autre méthode, celle du Crystal Violet, sur deux lignées cellulaires : MCF-7 dépendante des estrogènes (ER+) et MDA-MB-231 non dépendante (ER) (Cf. Annexe 4).

D'après les résultats d'activités biologiques obtenus selon la méthode du MTT pour les 17 fractions testées sur la lignée MCF-7 (Cf. Figure 19), il a été montré que la fraction MMS5-526-1-3 possédait la plus forte inhibition (75% d'inhibition à une concentration de 10 µg/ml), équivalente à celle de l'extrait brut. De plus, des résultats similaires ont été obtenus pour cette fraction avec le test au Crystal Violet sur cette même lignée cellulaire mais également sur la lignée non estrogéno-dépendante MDA-MB-231, avec un effet dose-réponse notable (Cf Annexe 4). Par conséquent, c'est cette fraction qui a par la suite été purifiée.

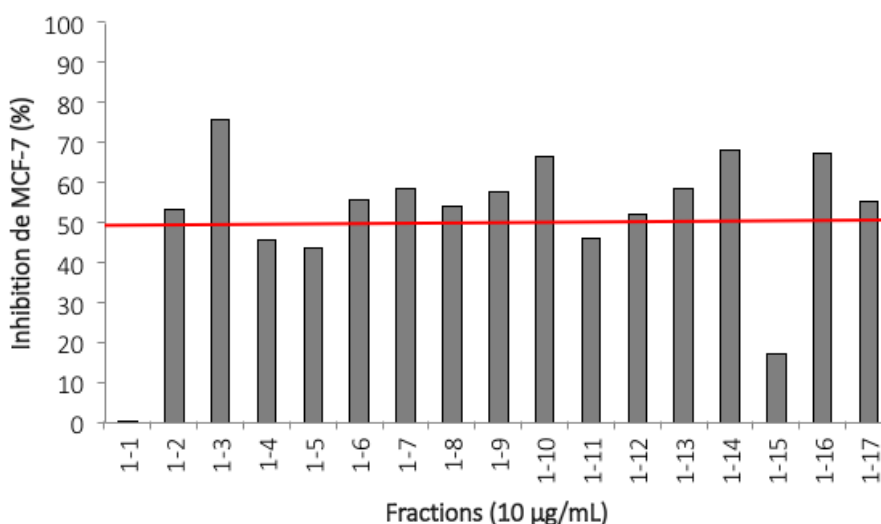


Figure 19. Cytotoxicité des 17 fractions (1-1→1-17) obtenues au cours du fractionnement de l'extrait brut MMS5-526 issu de la souche MMS5 *P. chrysogenum* sur la prolifération des cellules MCF-7. Pour chaque fraction testée à la concentration de 10 µg/mL, le pourcentage d'inhibition de la prolifération des cellules MCF-7 a été déterminé après un traitement de 24H.

Puisqu'aucune différence de sélectivité sur la viabilité cellulaire des deux lignées cellulaires étudiées selon la méthode du crystal violet n'a pu être observée, le suivi de l'activité pour le reste de l'étude a donc été déterminé uniquement sur la lignée MCF-7 à l'aide du test MTT, utilisé classiquement dans le laboratoire et ayant servi au cours du criblage initial.

Pour la suite du fractionnement, il a été choisi de manière générale, de fractionner certains échantillons en raison de plusieurs facteurs, à savoir : 1) un pourcentage d'inhibition de la prolifération des cellules MCF-7 le plus important pour un traitement de 24H réalisé à une concentration de 10 µg/mL ; et/ou 2) une masse suffisante pour des profils chromatographiques sur couche mince (CCM) peu complexes. Ces critères ont été définis afin de réaliser une séparation dans les meilleures conditions possibles pour espérer obtenir un ou plusieurs produits purs en quantité suffisante. De cette façon, 7 étapes de fractionnement ont été réalisées sur l'extrait MMS5-526, permettant d'obtenir un total de 79 échantillons. L'ensemble des travaux de purification a été représenté schématiquement sous la forme d'un « arbre de fractionnement » (Cf. Figure 20).

Parmi les échantillons obtenus, une molécule pure responsable de l'activité observée a pu être isolée et identifiée comme étant l'ergostérol. Ainsi, une première partie a été consacrée à l'isolement et l'identification de cette molécule, suivie d'une discussion critique sur les choix stratégiques et les observations ayant pu être effectuées au cours de ce fractionnement bioguidé.

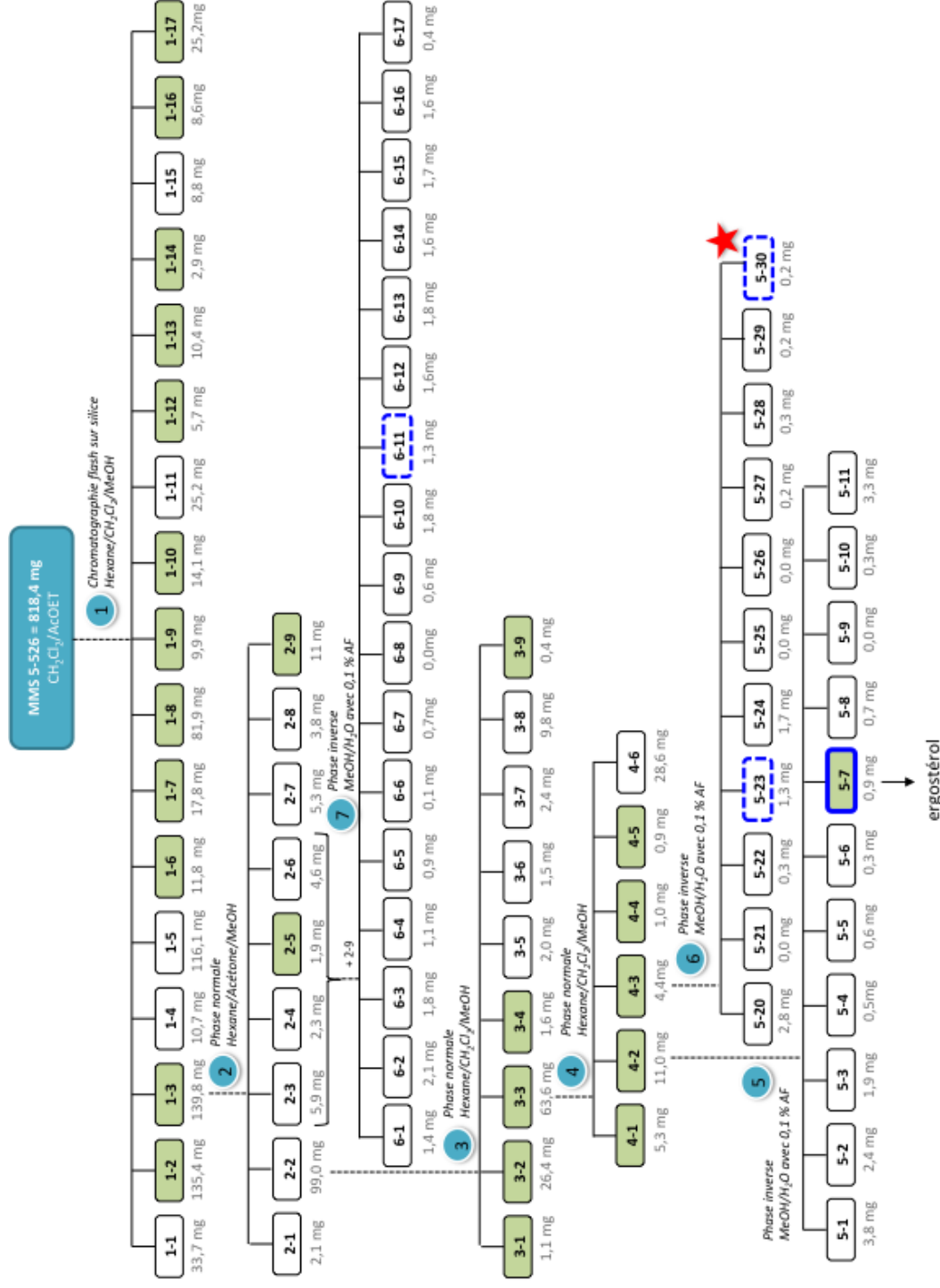
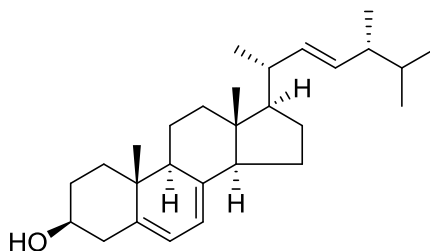


Figure 20. Arbre de fractionnement de l'extrait MMS 5-526 issu de la souche *P. chrysogenum*. Les numéros de 1 à 7 correspondent aux différentes étapes de fractionnement. Pour les fractionnements numéro 1 à 4, les fractions colorées en vert sont celles ayant la meilleure activité sur l'inhibition de la prolifération des cellules MCF-7 (> 50% pour une concentration testée de 10 µg/mL). Les fractions encadrées en bleu (trait plein), sont celles correspondant à des molécules pures isolées et identifiées par RMN. Les fractions encadrées en pointillées sont celles présentant un bon degré de pureté et pour lesquelles des propositions de formules brutes des molécules présentes ont pu être réalisées à l'aide de la LC-HRMS. L'étoile rouge correspond à la détection d'une molécule potentiellement nouvelle (0 hit dans le DNP). Voir les Annexes correspondantes : 6 (MMS5-5-7), 16 (MMS5-5-23), 17 (MMS5-5-30) et 18 (MMS5-6-11).

3.2.1.1. Isolement et identification d'une molécule bioactive : l'ergostérol

Au cours de ce fractionnement bioguidé de l'extrait MMS5-526, seule une molécule a ainsi pu être isolée et identifiée avec certitude dans la fraction MMS5-526-5-7. Il s'agit de l'ergostérol (Cf. Figure 21). Les analyses spectroscopiques RMN (Cf. Annexe 5) réalisées sur l'échantillon MMS5-526-5-7 (0,9 mg) ont permis en comparaison avec la littérature de confirmer son identification comme étant : l'ergostérol (Alexandre *et al.*, 2017). De plus, le spectre d'absorption UV de la molécule correspond parfaitement à celui de l'ergostérol avec 3 pics d'absorption UV à 272, 277 et 282 nm, respectivement (Cf. Annexe 6).



ergostérol

Figure 21. Structure chimique de l'ergostérol.

Etonnamment, il est à noter que le profil LC-MS de cette molécule n'est pas apparu comme correspondant à un produit pur (Cf. Annexe 6). Cela peut s'expliquer par une capacité d'ionisation de la molécule extrêmement faible par électrospray, favorisant la détection d'autres ions correspondant plutôt à du bruit d'analyse ou à des impuretés mineures. De plus, les ions semblant caractéristiques de la molécule (tels que m/z 481.2905) ne correspondaient pas à des adduits communément observés en électrospray ($[M+H]^+$, $[M+Na]^+$ ou $(M+K)^+$). Cette dernière observation nous a conduit à rédiger un article qui sera soumis à « Rapid Communications in Mass Spectrometry » et qui est présenté à la fin de ce chapitre (Cf. §6).

3.2.1.2. Description et analyse critique du fractionnement bioguidé

D'après la Figure 20 et Figure 22, il a été remarqué au cours de la première étape de fractionnement que plusieurs échantillons (colorés en vert) présentaient une activité sur l'inhibition de la prolifération des cellules MCF-7 supérieure à 50%. Il a tout d'abord été choisi de fractionner l'échantillon possédant la meilleure activité, soit : MMS5-526-1-3 (75% d'inhibition), tel qu'explicité précédemment.

Parmi les différentes activités biologiques déterminées pour les fractions suivantes MMS5-526-2-1 à MMS5-526-2-9, la fraction possédant la plus forte inhibition sur la prolifération de MCF-7 a été observée pour la fraction MMS5-526-2-5 (57% d'inhibition). Cependant, avec une masse de seulement 1,9 mg, il a été choisi par comparaison des profils chromatographiques obtenus pour les différentes fractions de regrouper les échantillons suivant : MMS5-526-2-3, 2-4, 2-5 et 2-9, permettant ainsi d'obtenir une masse plus importante d'une valeur de 25,7 mg.

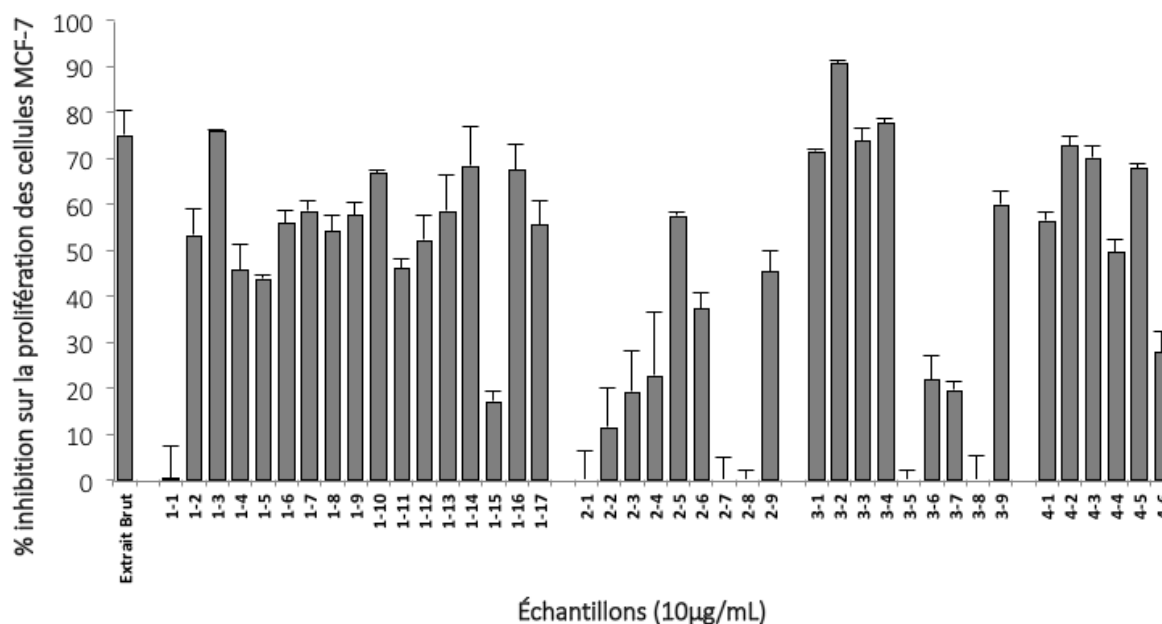


Figure 22. Cytotoxicité des fractions obtenues au cours du fractionnement numéro 1, 2, 3 et 4 de l'extrait brut MMS5-526 issu de la souche MMS5 *P. chrysogenum* sur la prolifération des cellules MCF-7. Pour chaque fraction testée à la concentration de 10 µg/mL, le pourcentage d'inhibition de la prolifération des cellules MCF-7 a été déterminé après un traitement de 24H.

A ce stade, aucune molécule responsable de l'activité observée n'a donc pu être isolée. Les quantités de chacune des fractions obtenues étant trop faibles, aucune purification n'a pu être poursuivie. Il a donc été choisi de mener un fractionnement sur la fraction MMS5-526-2-2, présentant la masse la plus importante (99,0 mg). Parmi les fractions obtenues (MMS5-526-3-1 → MMS5-526-3-9) le fractionnement a été poursuivi sur la fraction 3-3 présentant à la fois une forte inhibition sur la prolifération des cellules MCF-7 ainsi qu'une masse suffisante de 63,9 mg pour un profil CCM peu complexe. Puis les investigations ont été menées sur les fractions MMS5-526-4-2 et MMS5-526-4-3. Seul le fractionnement de MMS5-526-4-3 a permis d'isoler une molécule pure identifiée par RMN comme étant l'ergostérol (code **MMS5-526-5-7**).

Cependant, un point critique a été mis en évidence au cours du fractionnement bioguidé correspondant à la perte de l'activité de la fraction MMS5-526-2-2 au cours de la première étape. En effet, d'après la forte inhibition obtenue pour les échantillons 3-1 à 3-9 issu de cette fraction, qui ne présentait pourtant pas d'activité à l'origine (11%), des investigations plus poussées ont été réalisées dans le but de comprendre le phénomène observé. Bien que dans certains cas, ceci puisse s'expliquer par un effet de concentration des molécules bioactives au sein d'un échantillon, nous pensons qu'il s'agit ici très certainement d'un problème de solubilité lors de la préparation des échantillons d'analyses. Nous avons pu d'ailleurs le confirmer en comparant les profils LC-MS de ces fractions (Cf. Figure 23).

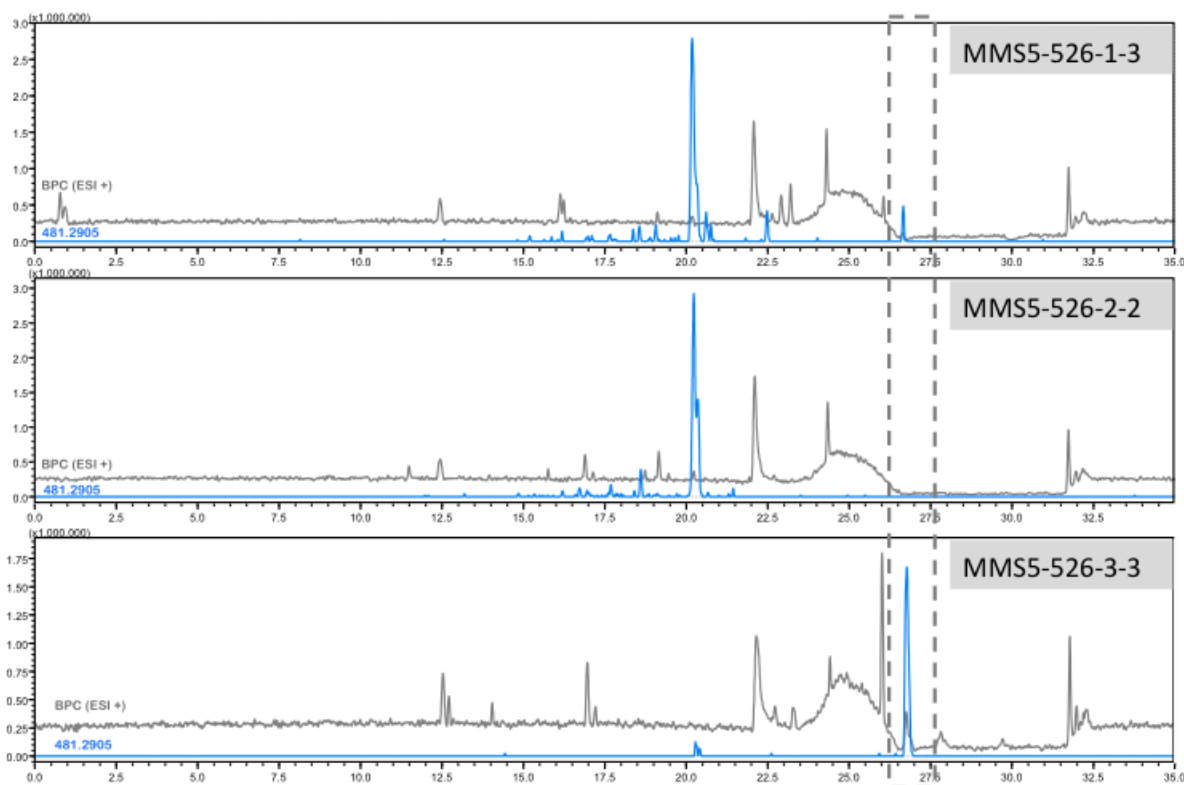


Figure 23. Profils LC-MS des fractions MMS5-526-1-3, MMS5-526-2-3, MMS5-526-3-2. Tracé BPC (Base peak chromatogram) obtenu en mode positif et tracé EIC (Extracted Ion chromatogram) de l'ion m/z 481.2927 correspondant à l'ergostérol.

Au final au cours de ce fractionnement bioguidé, la seule molécule pure isolée a été identifiée par RMN comme étant l'ergostérol, son activité sur la prolifération des cellules MCF-7 a donc été évaluée dans le but de confirmer son implication dans l'activité initialement observée. Sachant que l'ergostérol est un composé majoritairement produit par de nombreuses espèces fongiques, nous nous sommes interrogés sur l'activité des autres extraits de la collection, à savoir, les activités observées au cours du criblage sont-elles uniquement dues à l'ergostérol ou d'autres molécules actives peuvent-elles être présentes ?.

3.2.1.3. Evaluation de l'activité biologique et quantification de l'ergostérol

D'après les résultats obtenus via le test MTT (Cf. Figure 24), une valeur de CI_{50} de $0.10 \pm 8 \mu M$ a été déterminée pour la gamme de concentration testée (0,001-50 μM) avec une durée d'incubation de 24H. Cette activité de l'ergostérol sur les cellules MCF-7 est en accord avec une étude récente relatant de résultats similaires obtenus sur des modèles cancéreux du sein dont celui de la lignée MCF-7 (Li *et al.*, 2015). De plus, la valeur de CI_{50} obtenue de l'ordre du nanomolaire semble assez intéressante en comparaison avec d'autres molécules cytotoxiques d'intérêt dans le traitement du cancer du sein (Ling *et al.*, 2019).

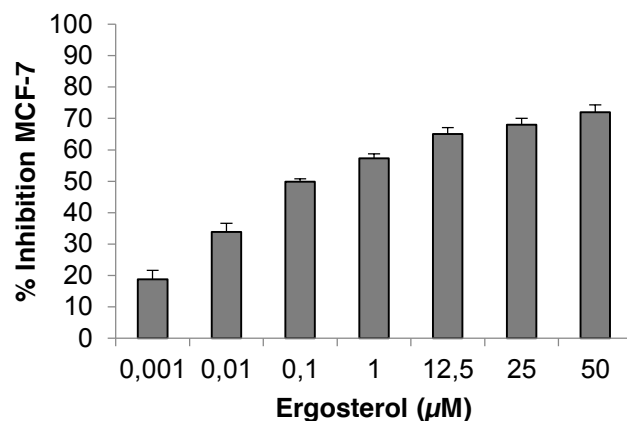


Figure 24. Cytotoxicité de l'ergostérol sur les cellules MCF-7 à 24H, dans une gamme de concentration de 0,001-50 µM

Afin d'évaluer de façon préliminaire la capacité d'autres souches de la collection à produire des molécules bioactives autres que l'ergostérol, nous avons choisi de quantifier l'ergostérol dans plusieurs des extraits disponibles et mis en regard ces résultats avec les activités mesurées sur MCF-7.

La production d'ergostérol au sein des extraits bruts a donc été quantifiée par analyse GC-MS, après une étape de saponification dans le but de séparer les composés en deux classes : saponifiables (ex : acides gras, triglycérides, phosphoglycérides) et insaponifiables (terpènes et stéroïdes). Les essais étant réalisés en triplicata, il a été choisi parmi les différents extraits étudiés au cours de criblage de mener des investigations sur ceux ayant une masse suffisante. Ainsi, 5 extraits issus des souches *Pycnidiophora dispersa* (MMS940-335 et MMS940-331), *Didymella americana* (MMS719-215 et MMS719-271) et *Chrysosporium queenslandicum* (MMS946-355) ont été étudiés (Cf. Tableau 3).

D'après les analyses GC-MS réalisées sur les extraits bruts et les fractions insaponifiables contenant les stérols et par conséquent l'ergostérol, il a pu être remarqué que contrairement aux échantillons issus de *Pycnidiophora dispersa* et *Chrysosporium queenslandicum*, celui-ci n'a pas été détecté dans ceux issus de la souche *Didymella americana*. Ainsi, il a pu être supposé que les techniques d'extraction utilisées ne sont pas optimisées pour ce genre de composé présent dans la membrane cellulaire.

Et pourtant, les deux extraits bruts testés présentaient une inhibition de la prolifération de 45,8 et de 38,2%. On peut donc en conclure que dans certains des extraits criblés, l'activité observée n'est pas uniquement due à l'ergostérol. En l'occurrence, il a pu être remarqué que les fractions saponifiées de ces extraits présentaient pour l'une d'entre elle une légère activité d'une valeur de 31,5%. Ainsi il a été suggéré que cette souche produit des molécules insaponifiables dont des lipides de types acides gras, triglycérides ou phospholipides pouvant présenter une activité potentielle.

En s'intéressant aux souches productrices d'ergostérol, les analyses GC-MS ont montré pour la souche *Pycnidiophora dispersa* que la quantité d'ergostérol était plus importante dans les échantillons obtenus à partir d'une culture réalisée sur milieu CYA (avec des valeurs de 8,8% pour l'extrait brut et 62,8% pour la fraction insaponifiable), que celle réalisée sur milieu DCA (3,0% et 39,7%). D'après ces résultats, il a été suggéré que le milieu DCA permet une production de stérols et de terpènes de structures plus diversifiée que le milieu CYA. Concernant la souche *Chrysosporium queenslandicum*, les analyses GC-MS réalisées sur l'extrait brut et la fraction insaponifiable ont permis de quantifier l'ergostérol avec des valeurs respectives de 5,3% et 51,8%, intermédiaires aux précédentes.

Sur l'ensemble des échantillons pour lesquels l'ergostérol a pu être détecté, les forts pourcentages d'ergostérol obtenus dans les fractions insaponifiables ont permis de mettre en évidence que celui-ci représentait tout de même au moins 40% de l'ensemble des terpènes et stérols produits.

Concernant l'activité de ces extraits, l'activité la plus importante (81,5%) correspond bien à la teneur d'ergostérol la plus importante (8,8%). Toutefois, la corrélation ne semble pas linéaire car de bien plus faibles pourcentages d'inhibition mais proches (17,3% et 20,3%) sont observés pour des teneurs en ergostérol de 3,0% et 5,3% pour MMS940-335 et MMS 946-355, respectivement. En revanche pour chacune des fractions insaponifiables contenant au minimum 40% d'ergostérol, celles-ci ont toutes présenté une inhibition de la prolifération des cellules MCF-7 supérieure à 50%. Ce qui est cohérent avec l'activité biologique obtenue pour cette molécule pure.

En conséquence, il nous semble important pour la suite des études menées sur la collection de champignons marins du laboratoire, de vérifier la teneur en ergostérol d'extraits actifs sur MCF-7, avant de poursuivre les investigations.

Tableau 3. Quantification de l'ergostérol au sein d'extrait brut et dans la fraction insaponifiable par analyse GC-MS, et évaluation de leur activité sur la prolifération cellulaire des cellules MCF-7.

Espèces fongiques	Code extrait MMS	Milieu	Quantitification de l'ergostérol (%)		Inhibition de la prolifération des cellules MCF-7 à 10 µg/mL (%)			
			Extrait brut	Fraction Insaponifiable	Fraction insaponifiable	Fraction saponifiable	Extrait brut	Extrait brut
<i>Pycnidophora dispersa</i>	940-335	DCA	3,0 ± 3,7	39,7 ± 1,00	59,9 ± 1,0	0,0 ± 7,2	17,3 ± 3,0	
<i>Pycnidophora dispersa</i>	940-331	CYA	8,8 ± 5,9	62,8 ± 1,3	62,6 ± 2,4	5,6 ± 3,1	81,5 ± 0,6	
<i>Didymella americana</i>	719-215	YES	ND*	ND	0,0 ± 6,7	31,5 ± 2,32	45,8 ± 0,3	
<i>Didymella americana</i>	797-271	YES	ND	ND	17,0 ± 1,42	11,5 ± 5,0	38,2 ± 1,2	
<i>Chrysosporium queenslandicum</i>	946-355	DCA	5,3 ± 3,7	51,8 ± 7,87	62,3 ± 1,4	1,0 ± 14,0	20,3 ± 4,6	

* ND : Non Détecté

3.2.2. Profilages métaboliques, identification des composés majoritaires et purification ciblée

3.2.2.1. Profilage métabolique

Au cours de ces travaux de thèse, afin de mieux évaluer la chimiodiversité produite par la souche MMS5 *P. chrysogenum* et de compléter son étude phytochimique, un profilage métabolique a été réalisé sur l'extrait MMS5-526. Ce dernier a permis d'annoter les molécules majoritairement produites et/ou de mettre en évidence de nouvelles molécules au sein du mélange complexe obtenu initialement sans que celles-ci ne soient purifiées. D'après les profils LC-MS obtenus pour l'extrait MMS5-526 en mode positif (ESI+) et négatif (ESI-), 21 pics chromatographiques ont pu être mis en évidence sur le chromatogramme (Cf. Figure 25).

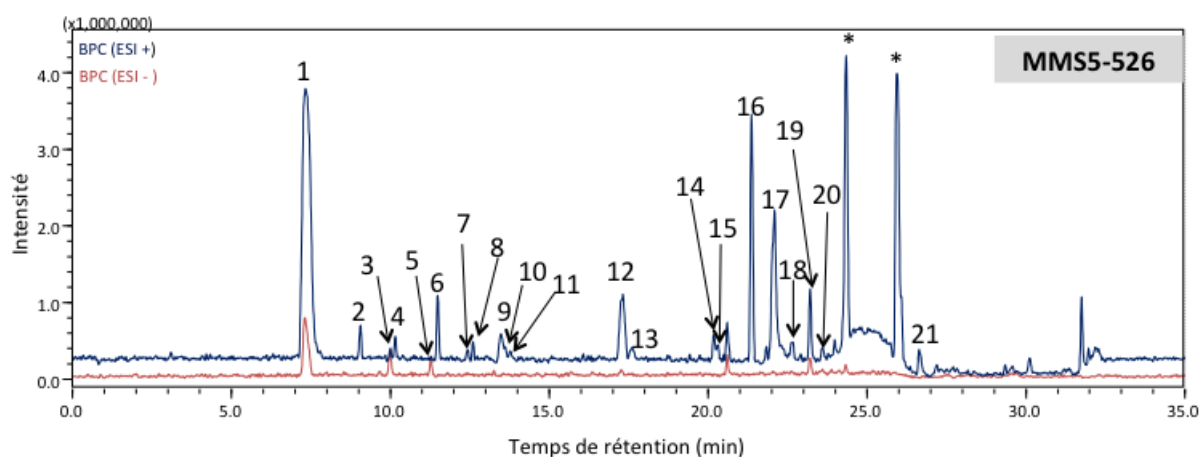


Figure 25. Profils LC-MS de l'extrait brut MMS5-526 obtenus en mode positif (ESI+) et négatif (ESI-). *pics détectés également sur les profils des analyses contrôles de blanc MeOH de grade UPLC.

Pour chacun de ces pics, correspondant à une valeur m/z associée à un temps de rétention, une recherche des adduits et/ou de fragments sur leurs spectres de masse a été réalisée. Puis, à partir de cette première étape, il a été recherché une formule brute associée à ces ions, à l'aide du logiciel Shimadzu LCMS Solution version 3.71. Pour chacune des formules proposées, des paramètres tel que l'iso-score (score de similitude de profils isotopiques entre le profil observé et le profil théorique) et la différence de masse exacte (en ppm) ainsi que certaines « règles chimiques » telles que le ratio C/H ont été utilisés. Enfin, par comparaison de la formule déterminée à celles des molécules référencées dans le dictionnaire des produits naturels (<http://dnp.chemnetbase.com/>), et plus particulièrement celles produites par les ascomycètes, des annotations de structures ont pu être réalisées. Enfin, lorsque cela était possible, nous avons également voulu en comparaison avec la littérature scientifique évaluer l'activité biologique potentielle des molécules annotées sur le cancer du sein. L'ensemble des informations recueillies pour chacun des pics chromatographiques a été présenté sous la forme d'un tableau récapitulatif (Cf. Tableau 4).

Tableau 4. Déréplication des ions majoritaires détectés au cours des analyses LC-MS de l'extrait brut MMSS-526.

Pics	T _R (min)	m/z	Adduit et/ou fragment	Masse monoisotopique	Formule brute	Diff (ppm)	Iso Score *	Hits (DNP)	Structures proposées dans le DNP
1	7,3	456.1559	[M+Na] ⁺	433.1744	C ₂₃ H ₂₃ N ₅ O ₄	-0.23	96.90	1	Mélégrine
		434.1815	[M+H] ⁺						
		403.1674	[M-30] ⁺						
		432.1674	[M-H] ⁻						
2	9,0	445.1730	[M-30] ⁺	475.1918	C ₂₆ H ₂₁ N ₉ O	-3.57	92.30	0	
		476.1925	[M+H] ⁺						
		474.1787	[M-H] ⁻						
		474.1787	[M-H] ⁻						
3	9,9	555.2137	[M+Na] ⁺	532.2339	C ₂₈ H ₃₆ O ₁₀	1.73	71.65	4	Xylarin; 4'-O-De-Me,4'-Ac AustalideD AustalideE Austalide N
		473.2161	[M-59] ⁺						
		429.2267	[M-103] ⁺						
		577.2301	[M+FA-H] ⁻						
4	10,1	645.2238	[M+TFA-H] ⁻	168.0669	C ₁₁ H ₇ NO	0.47	65.31	0	
		211.0867	[M+H+CH ₃ CN] ⁺						
5	11,3	169.0739	[M+H] ⁺	518.2492	C ₂₉ H ₃₄ N ₄ O ₅ C ₂₈ H ₃₈ O ₉	-1.07 1.24	78.96 73.04	0	
		541.2384	[M+Na] ⁺						
		459.2339	[M-59] ⁺						
		415.2655	[M-103] ⁺						
		343.2242	[M-175] ⁺						
6	11,5	563.2511	[M+FA-H] ⁻	271.2667	C ₁₂ H ₂₈ N ₆ O	2.41	73.10	0	
		631.2361	[M+TFA-H] ⁻						
7	12,4	290.2667	[M+NH ₄] ⁺	304.1634	C ₁₃ H ₂₄ N ₂ O ₆	-1.31	73.62	0	
		272.2740	[M+H] ⁺						
		305.1707	[M+H] ⁺ ?						

* Iso Score : score de similitude de profils isotopiques entre le profil observé et le profil théorique (score compris entre 0 et 100)

Pics	T _R (min)	m/z	Adduit et/ou fragment	Masse monoisotopique	Formule brute	Diff (ppm)	Iso Score *	Hits (DNP)	Structures proposées dans le DNP
		465.2468	[M+H] ⁺						
		398.2391	[M-67] ⁺						
8	12,6	376.2629	[M-88] ⁺	464.2398	ND ⁺			1	Asnovolin G
		463.2225	[M-H] ⁻						
		509.2424	[M+FA-H] ⁻						
		509.2895	[M+H] ⁺						CommunesineB
9	13,5	531.2702	[M+Na] ⁺	508.2823	C ₃₂ H ₃₆ N ₄ O ₂	-3.14	100	3	FellutanineC Amauromine
		593.3483	[M+H] ⁺ ?	592.3410	C ₃₇ H ₄₄ N ₄ O ₃	-0.51	65.01	0	
10	13,6				C ₃₆ H ₄₈ O ₇	1.69	65.43		
		511.3045	[M+H] ⁺						
11	13,8	533.2840	[M+Na] ⁺	510.2975	ND ⁺			1	Longirostrerone B
		577.3497	[M+H] ⁺ ?	576.3424	C ₄₃ H ₄₄ O	4.16	85.59	0	
12	17,2				C ₄₀ H ₄₈ OS	-1.73	70.41		
		579.3672	[M+H] ⁺						
13	17,6	601.3485	[M+Na] ⁺	578.3602	C ₃₇ H ₄₆ N ₄ O ₂	-3.80	87.74	0	
		481.2911	[M+Na] ⁺		C ₃₆ H ₅₀ O ₆	-1.38	82.06		
14	20,2	459.3128	[M+H] ⁺	458.3055	C ₂₈ H ₄₂ O ₅	-2.70	79.17	3	Mer-NF8054X ChaxineB Ergokonin B
		497.2691	[M+K] ⁺						
		281.2470	[M+H] ⁺						
15	20,6	279.2322	[M-H] ⁻	280.2398	C ₁₈ H ₃₂ O ₂	-1.78	61.75	1	Acide 4-méthyle-7,11-heptadecadienoïque
		325.2380	[M+FA-H] ⁻						
16	21,3	282.2209	[M+H] ⁺ ?	281.2136	C ₂₀ H ₂₇ N	-2.48	77.70	0	

* Iso Score : score de similitude de profils isotopiques entre le profil observé et le profil théorique

+ ND : Non déterminée car les formules brutes proposées présentaient des scores de profils isotopiques (Iso Score) inférieurs à 50%

Pics	T _R (min)	m/z	Adduit et/ou fragment	Masse monoisotopique	Formule brute	Diff (ppm)	Iso Score *	Hits (DNP)	Structures proposées dans le DNP
17	22	406.3299 381.1762	? (ESI +)	ND ⁺	ND ⁺				
18	22.6	375.3033 393.3180	[M+H] ⁺ [M+K] ⁺	374.2960	C ₂₈ H ₃₈	0.00	93.94	0	
19	23,3	377.3165 395.3339	[M-H ₂ O+H] ⁺ [M+H] ⁺	394.3266	C ₂₈ H ₄₂ O	1.52	55.89	4	1-(10→6)-abeo-Ergosta-5,7,9,22-tetraen-3α-ol 14-Dehydroergostérol (3β,22E) Ergosta-5,7,22,24(28)-tetraen-3-ol (3β,22E) 6-Méthyl-19-norergosta-5,7,9,22-tetraen-3-ol
20	24.3	338.3400 675.6828	[M+H] ⁺ [2M+H] ⁺	337.3327	C ₂₀ H ₄₅ NO	-1.18	74.68	0	
21	26,7	377.3239 481.2827	? (ESI +)	396.3392	C ₂₈ H ₄₄ O				Ergostérol ^a

* Iso Score : score de similitude de profils isotopiques entre le profil observé et le profil théorique

⁺ ND : Non déterminée car les formules brutes proposées présentaient des scores de profils isotopiques (Iso Score) inférieurs à 50%

^a Identification par analyse RMN

Le profilage chimique réalisé sur l'extrait MMS5-526 suivi d'une étape de recherche dans les bases de données a permis d'annoter seulement 9 pics chromatographiques sur les 21 numérotés sur le chromatogramme (Cf. Figure 25). Selon le nombre de hits obtenus dans le DNP à partir des formules brutes déterminées pour ces 9 pics, il a pu être mis en évidence une ou plusieurs molécules.

Au regard du nombre de métabolites référencés dans la littérature scientifique pour cette espèce, ce faible taux d'annotation nous est apparu comme très surprenant. Il démontre que même si une espèce semble très étudiée dans la littérature scientifique, les organismes fongiques sont avant tout des êtres vivants dont l'identification et le référencement de l'ensemble des métabolites qu'ils sont capables de produire est finalement incomplet. Ce phénomène peut s'expliquer d'une part par l'incroyable diversité chimique des molécules produites par ces organismes, mais également par leur forte dépendance aux conditions de cultures utilisées.

Au cours du travail de déréplication réalisé par comparaison des formules brutes dans le DNP, il est arrivé pour certains des pics chromatographiques (2, 4, 5, 6, 13, 16, 18 et 20) qu'aucune correspondance avec des molécules présentes dans la base de données ne soit obtenue. D'après ces résultats, il a été suggéré que ces pics pouvaient correspondre à des molécules potentiellement nouvelles. Cependant, il est nécessaire de rester vigilant car ces pics pourraient également correspondre à des molécules déjà décrites dans la littérature mais qui ne sont pas référencées dans la base de données utilisée.

Pour les autres pics chromatographiques, en raison de facteurs multiples tel que l'absence de détection d'un ion correspondant à l'adduit $[M+H]^+$, d'un profil isotopique au sein duquel le $[M+1]$ n'a pu être détecté, ou encore d'une faible intensité ; aucune formule brute n'a pu être associée. Ainsi, selon la méthodologie exposée qui consiste à identifier les molécules par comparaisons de leurs formules brutes, il a été impossible d'évaluer si certains de ces pics chromatographiques correspondaient à des molécules nouvelles ou déjà décrites dans la littérature. Par conséquent, la comparaison a été faite selon la masse mono-isotopique calculée à partir des adduits identifiés.

D'après la littérature scientifique, l'activité biologique de certaines des molécules annotées a pu être mise en évidence. Parmi l'ensemble des molécules, seulement 2 ont été décrites pour leur activité anti-cancéreuse potentielle sur le cancer du sein : Ainsi, les pics numéros 8 et 11, ont pu être annoté respectivement comme l'Asnovolin G de formule brute $C_{25}H_{36}O_8$ (pour une masse monoisotopique de 464.2398, pic 8) ainsi que la Longirostrerone B, de formule brute $C_{31}H_{42}O_6$ (pour de masse monoisotopique de 510.2975, pic 11). En revanche, pour d'autres pics chromatographiques, aucune annotation d'adduit, de formules brutes et par conséquent de structures n'ont pu être réalisées. Dans ce cas il a été difficile d'évaluer l'intérêt potentiel de ces pics comme correspondant à de potentielles molécules nouvelles car ils pourraient tout à fait correspondre à des pics « parasites ». la méléagrine (Mady *et al.*, 2016) et l'ergostérol (Li *et al.*, 2015).

Cependant, il existe une grande controverse dans les résultats obtenus sur la méléagrine. En effet, des études rapportent son activité cytotoxique sur les lignées cancéreuses mammaires MDA-MB-231 et MCF-7 (Mady *et al.*, 2016); alors que d'autres les réfutent (Yunianto *et al.*, 2014). Par ailleurs, d'autres molécules ont pu montrer des activités biologiques différentes. Ainsi, la longirostrerone B a été décrite pour sa forte toxicité sur les cellules KB ($CI_{50} < 10 \mu M$) (Panthama *et al.*, 2011); la molécule Mer-NF8054X pour son effet antifongique (Sakai *et al.*, 1994) et inhibiteur de croissance de cellules humaines leucémiques HL60 (Takahashi *et al.*, 1998), l'ergokonine B pour une activité antifongique ou encore le 14-déhydro ergostérol pour des propriétés anti-inflammatoires (Ano *et al.*, 2017).

Il est à noter que l'ergostérol comme expliqué précédemment ne s'ionise pas de manière classique et a donc été identifiée en se référant au temps de rétention et au profil LC-UV caractéristique obtenu pour cette molécule pure isolée et identifiée par RMN.

D'après la déréplication réalisée sur les composés majoritaires, seule la méléagrine pourrait présenter une activité sur la cible biologique étudiée. Par ailleurs, il a été mis en évidence que des molécules nouvelles pouvaient être produites par cette souche. Afin de vérifier ces hypothèses une purification ciblée sur la méléagrine ainsi que de nouvelles molécules a été réalisée.

3.2.2.2. Purification ciblée

Afin d'orienter la purification sur d'autres molécules que celles obtenues au cours du fractionnement bioguidé, celle-ci a été réalisée sur un deuxième extrait MMS5-531 (484,9 mg) issue d'une nouvelle culture de la souche MMS5. Ce nouveau lot a été fractionné, selon les mêmes conditions que le lot précédent à l'aide d'un système de chromatographie liquide basse pression. En revanche, afin de limiter le nombre d'étapes de purification et de éviter les risques de dégradations et de réduire au maximum les pertes de quantités liées aux étapes successives, ce deuxième lot a été fractionné de manière beaucoup plus drastique. En effet, contrairement aux 17 fractions obtenues dans le cas précédent, le fractionnement réalisé sur ce nouvel extrait a permis d'obtenir 37 fractions. Ces dernières ont été analysées par chromatographie sur couche mince (CCM) (Cf. Figure 26) et par analyses LC-UV-MS dans le but d'orienter les purifications vers les molécules présentant le plus fort intérêt.

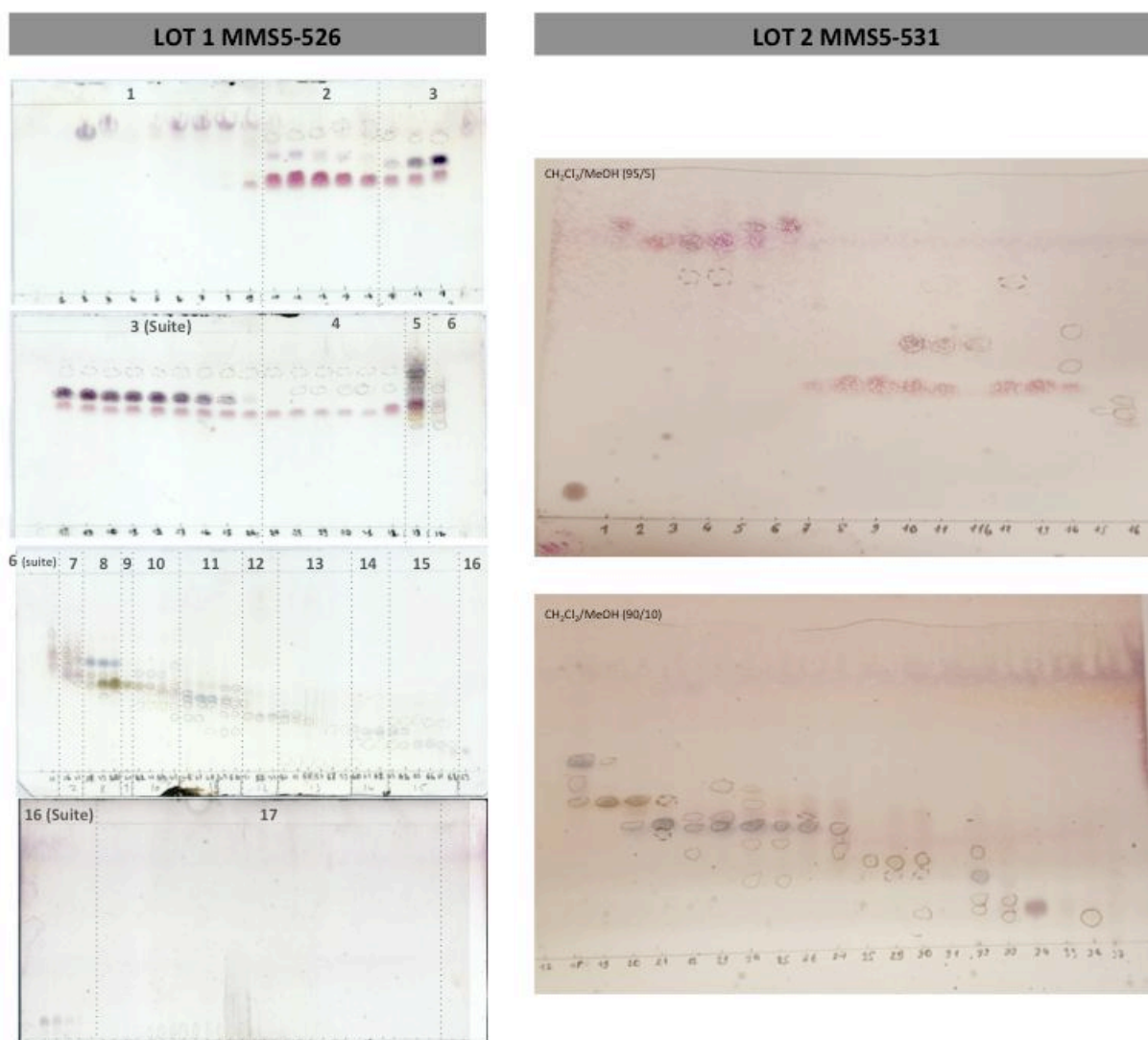


Figure 26. Profils chromatographiques obtenus par CCM pour les fractions issues de la première étape de fractionnement de l'extrait MMS5-526 (Lot 1) et de MMS5-531 (Lot 2), issus de culture de la souche MMS5 *P. chrysogenum*.

Au cours des analyses LC-MS et du travail de déréplication réalisé sur l'extrait brut issu de la souche MMS5, il a été mis en évidence parmi les composés majoritaires, la méléagrine. D'après la controverse existante sur l'activité de la méléagrine dans la littérature, nous avons voulu isoler celle-ci afin d'évaluer son activité potentielle sur notre modèle du cancer du sein. En effectuant une recherche ciblée sur les fractions pour lesquelles la méléagrine a pu être détectée au cours des analyses LC-UV-MS et repérée sur la CCM (tâche marron jaune), il a donc été choisi de réaliser des purifications sur les échantillons du deuxième lot : MMS5-531-18, MMS5-531-19, MMS5-531-20.

Un autre métabolite dérivé de l'ergostérol de m/z 527.3239 correspondant d'après le DNP à l'anicequol a également pu être annotée. L'isolement de cette molécule a donc aussi été entrepris dans le but d'élargir les investigations biologiques menées sur les composés stéroïdiens.

Enfin, d'après les analyses LC-UV-MS et le travail de déréplication réalisé pour ces trois fractions, il a pu être mis en évidence (en plus de la méléagrine) deux molécules ne possédant pas de hit dans le DNP avec des m/z d'une valeur respective de 448.1963 et 593.3488. Ainsi, les purifications ont été ciblées sur l'isolement de ces deux molécules (sans hit dans le dnp) dans le but de rechercher l'originalité.

L'ensemble des molécules pures qui ont été isolées et identifiées par analyse structurale RMN est listé ci-après (Cf. Tableau 5).

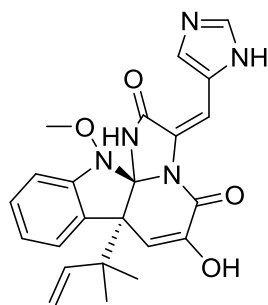
Tableau 5. Molécules d'intérêt détectées, isolées et identifiées dans les fractions MMS5-531-1-18 → MMS5-531-1-20.

Les molécules classées en fonction de leur temps de rétention ont été analysées par LC-MS et analysés par RMN. Les profils LC-UV-HRMS et les spectres de masses sont représentés respectivement, en annexe 7, 11, 12, 13 et 9.

T _R (min)	m/z	Adduit et/ou fragment	UV	Masse monoisotopique	Formule brute	Diff (ppm)	Iso Score *	Hits (DNP)	Structures proposées dans le DNP	Initialement présent dans la fraction MMS	Code échantillon
6,5	434.1815	[M+H] ⁺	202	433.1744	C ₂₃ H ₂₃ N ₅ O ₄	-0.23	96.90	1	Méléagrine	5-531-1-18	CTBG-04-m/z 434
	403.1674	[M-30] ⁺	262							5-531-1-19	MMS5-531-18-2
	432.1674	[M-H] ⁻	329							5-531-1-20	MMS5-531-18-3 MMS5-531-18-4
7,6	448.1966	[M-H ₂ O+H] ⁺	204	465.2013	C ₂₃ H ₃₁ NO ₉ C ₂₄ H ₂₇ N ₅ O ₅	3.00 0.21	91.52 94.20	0		5-531-1-19	CTBG-03-m/z 448
	466.2086	[M+H] ⁺	238							5-531-1-20	
	464.1840	[M-H] ⁻	270								
			308								
13,0	593.3479	[M+H] ⁺	212	592.3406	C ₃₇ H ₄₄ N ₄ O ₃ C ₃₆ H ₄₈ O ₇	1.01 -1.18	92.69 92.89	0		5-531-1-19	CTBG-01-m/z 593
	637.3406	[M+FA-H] ⁻	235								
			270								
15,3	527.3373	[M+Na] ⁺	190	504.3347	C ₃₁ H ₄₄ N ₄ O ₂ C ₃₀ H ₄₈ O ₆	0.19 2.87	98.96 94.72	0 3	Acide 4,7,22,24-Tetrahydroxy-3,4-seco-11,13(18)-oleanadiène-3-oïque Acide 4,15,22,24-Tetrahydroxy-3,4-seco-11,13(18)-oleanadiène-3-oïque Anicequol	5-531-1-19	CTBG-02-m/z 527
	522.3804	[M+NH ₄] ⁺									
	549.3334	[M+FA-H] ⁻									

* Iso Score : score de similitude de profils isotopiques entre le profil observé et le profil théorique (score compris entre 0 et 100)

La méléagrine (Cf. Figure 27) caractérisée par la présence de l'ion m/z 434.1815 $[M+H]^+$ et d'un massif d'absorption UV comprenant 3 pics de longueur d'onde 202, 262 et 329 nm avec un temps de rétention de 6,5 min dans les conditions d'élutions utilisées (Cf. Annexe 7), a pu être isolée plusieurs fois sous les codes échantillons : **CTBG-04- m/z 434**, **MMS5-531-18-2**, **MMS5-531-18-3** et **MMS5-531-18-4**. La confirmation de sa structure obtenue par analyse RMN (Cf. Annexe 8) pour l'échantillon CTBG-04- m/z 434 d'une masse (3,3 mg) a permis en comparaison avec la littérature scientifique de confirmer l'identification de cette molécule.

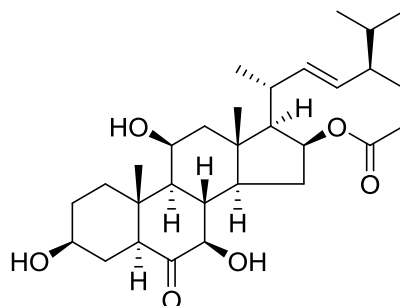


méléagrine

Figure 27. Structure de la méléagrine

Code échantillon : CTBG-04- m/z 434

La purification ciblée menée sur les ions m/z 527.3239 $[M+Na]^+$ et m/z 522.3735 $[M+NH_4]^+$ n'absorbant que très peu en UV (pic d'absorption à 190 nm) avec un temps de rétention de 15,3 min dans les conditions d'élutions utilisées (Cf. Annexe 9), a conduit à l'isolement d'une molécule référencée sous le code échantillons : **CTBG-02- m/z 527**. La confirmation de sa structure obtenue par analyse RMN (Cf. Annexe 10) d'une masse (0,4 mg) a permis en comparaison avec la littérature scientifique (Hua *et al.*, 2005) de confirmer l'identification de cette molécule comme étant l'acinequol ($C_{30}H_{48}O_6$, masse monoisotopique de 504.3447), dont la structure et le table de données figurent ci-dessous (Cf. Figure 28 et Tableau 6).



anicequol

Figure 28. Structure de l'anicequol

Code échantillon : CTBG-02- m/z 527

Tableau 6. Table de données RMN ^1H (500 MHz) et ^{13}C (125 MHz) de l'anicequol de formule brute $\text{C}_{30}\text{H}_{48}\text{O}_7$ solubilisé dans le CDCl_3 .

Position	δ_{C} , type	δ_{H} (J en Hz)	HMBC	COSY
1	36.16, CH2	1.98, dt (12.4, 3.5 et 3.0) 1.37	19	2
2	30.42, CH2	1.88, d (12.5) 1.45		1 et 3
3	70.36, CH-OH	3.60, dt (11.1, 6.1 et 4.6)		2 et 4
4	29.47, CH2	1.88 1.67, q (12.4)		3 et 5
5	54.80, CH	2.27, dd (12.5 et 3.2)	4, 6 et 10	4
6	209.85, C=O	/		
7	79.01, CH-OH	3.71, s	8	8
8	42.60, CH	2.11, m		9
9	55.43, CH	1.34, tt (11.9, 11.4, 3.9 et 3.3)		8 et 11
10	40.85, C q	/		
11	68.50, CH-OH	4.39, m	8	12 et 9
12	49.02, CH2	2.17, dd (14.4 et 2.8)	9,11, 13 et 18	11
13	43.05, C	/		
14	56.23, CH	1.31, td (6.9, 5.9 et 2.9)		15
15	36.61, CH2	2.61, dt (14.8 et 7.7) 1.53		14 et 16
16	74.78, CH-O	5.10, td (8.0, 4.5)		15 et 17
17	60.29, CH	1.22, dd (10.9 et 7.8)	13, 12, 20 et 22	16 et 20
18	25.59, CH3	1.55, m	13, 17 et 12	
19	15.79, CH3	0.96, s	10 et 5	
20	34.45, CH	2.54, td (11.1 et 7.3)		17, 21 et 22
21	21.29, CH3	1.07, d (6.8)	20, 22 et 17	20
22	135.05, CH	5.16, m	21 et 23	20 et 23
23	133.17, CH	5.16	22 et 24	22 et 24
24	43.42, CH3	1.77, sex (6.7)	23	23, 28 et 26
25	33.20, CH	1.41, t (6.6)	23	26
26	20.25, CH3	0.81, d (6.7 et 3)	24	26 et 24
27	19.81, CH3	0.80, d (6.6 et 3)	24	
28	18.22, CH3	0.86, d (6.8)	25	25
	/	3.72, s		
29	170.24, C=O	/		
30	21.66, CH3	1.98, s		

La purification ciblée menée sur la molécule correspondant aux ions m/z 448.1963 $[\text{M}-\text{H}_2\text{O}+\text{H}]^+$, m/z 466.2086 $[\text{M}+\text{H}]^+$ et m/z 464.1868 $[\text{M}-\text{H}]^-$ à un temps de rétention de 7,6 min a permis d'isoler une molécule référencée sous le code **CTBG-03- m/z 448** (1,2 mg) ne présentant pas de hit dans le DNP. Cette molécule possède un massif d'absorption UV caractéristique comprenant 3 pics aux longueurs d'onde 204, 238 et 308 nm (Cf. Annexe 11). Les analyses RMN réalisées sur cette molécule sont actuellement en cours de traitement afin d'élucider la structure (Cf. Annexe 12).

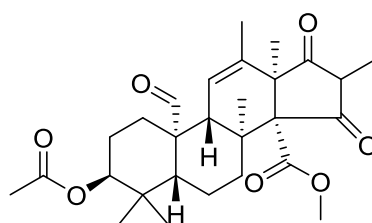
Enfin, la dernière purification a été réalisée sur le produit correspondant aux ions m/z 593.3488 $[M+H]^+$ et 637.3406 $[M+FA-H]^-$ correspondant à la forme ionisée d'une molécule (Tr = 13,0 min) non référencée dans le DNP. Les analyses LC-MS-UV de cette molécule portant le code **CTBG-01- m/z 593 (0.7mg)** ont permis de mettre en évidence la présence d'un massif d'absorption UV caractérisé par deux pics de longueur d'onde 212 et 270 nm (Cf. Annexe 13). La structure de cette molécule est en cours d'élucidation à partir des analyses RMN (Cf. Annexe 14 et Tableau 7) et de donnée de spectrométrie de masse haute résolution ($C_{38}H_{48}O_7$, masse monoisotopique de 504.3447).

Tableau 7. Table de données RMN 1H (500 MHz), ^{13}C (125MHz), ainsi que RMN2D HMBC et COSY de l'échantillon CTBG-01-593 de formule brute $C_{36}H_{48}O_7$ solubilisé dans le $CDCl_3$.

Position	δ_C , type	δ_H (J en Hz)	HMBC	COSY
1	142.37, CH	7.31, dd (15.1 et 10.6)	27	7 et 8
2	125.54, CH	7.14, d (7.6)	6 et 30	5
3	129.16 CH	6.88, td (7.6 et 3.2)	9 et 28	10 et 11
4	134.85, CH	6.72, dd (15.8 et 4.9)	2, 6, 22, 12	8 et 6
5	120.49, CH	6.65, t ou q (7.75)	2 et 32	
6	123.03, CH	6.58, (15.8 et 4.1)	32, 35, 4, 2 et 1	
7	121.50, CH	6.54, d (15.0)	8 et 27	
8	130.91, CH	6.18, dd (15.6 et 10.7)		
9	137.73, CH	6.12, q (6.5)	1	23
10	113.55, CH	6.09, d (7.9)	11, 32 et 15	
11	102.09, CH	5.94, dd (7.6 et 3.3)		
12	118.08, CH ₂	5.1, d (8.5)		
13	79.49, CH-O	5.11		20 ou 22
14	82.54, CH-O	4.75, d (8.2)	36, 32, 28, 21 et 1	
15	65.90, CH-O	4.17, d (9.1)	18, 17 et 31	18
16 a	44.50, CH ₂	3.87, td (8.2 et 4.1)	35	20 et 16b
16 b	44.50, CH ₂	3.07, m		20
17	36.16, CH ₂	3.44, m	15	
18	64.18, CH-O	2.90, dd (9.2 et 5.1)	34 et 15	
19	29.84, CH ₃	2.83, s	28 et 14	
20	30.69, CH ₂	2.73, m 2.01, m	16, 32 et 35	21
21	38.22, CH ₂	2.36, m 2.28, m	35	
22	18.88, CH ₃	2.00, s	1, 12 et 4	
23	19.19, CH ₃	1.84, d (6.3)	31, 8 et 1	
24	20.87, CH ₃	1.65, s	18, 26 et 34	
25	30.45, CH ₃	1.44, d (1.68)	33, 29 et 18	
26	25.20, CH ₃	1.41, s	34, 24 et 18	
27	168.85, C			

Position	δ_C , type	δ_H (J en Hz)	HMBC	COSY
28	150.57, C			
29	141.26, C			
30	139.98, C			
31	137.00, C			
32	132.77, C			
33	71.49, C			
34	60.02, C			
35	52.22, C			
36	51.51, C			

Au cours de la purification ciblée de ces molécules il a également pu être isolé et identifié une molécule déjà référencée dans la littérature et connue pour être produite par *P. chrysogenum* : l'andrastine A (C₂₈H₃₈O₇) (Matsuda *et al.*, 2013), référencée sous le code échantillon **MMS5-531-18-10** (0,5 mg). Au cours des analyses LC-UV-MS (Cf. Annexe 15), ce meroterpénoïde (Cf. Figure 29) a pu être détecté par la présence des adduits [M+Na]⁺, [M+H]⁺ et [M-H]⁻ ainsi que d'un fragment [M-59]⁺, associé à un temps de rétention de 12,9 min. Les analyses RMN (Cf. Annexe 16) ont confirmé en comparaison avec la littérature (Hoai Trinh *et al.*, 2018), la structure de ces molécules avec la présence de signaux caractéristiques obtenus sur le spectre ¹H. Malheureusement aucun signal de qualité n'a pu être obtenu pendant l'acquisition du spectre carbone ¹³C. Par ailleurs, le massif d'absorption de cette molécule caractérisé par la présence des deux pics d'absorption aux longueurs d'onde de 220 et 260 nm correspondait également aux données obtenues dans la littérature.



andrastine A

Figure 29. Structure de l'andrastine A**Code échantillon : MMS5-531-18-10**

Parmi les différentes molécules isolées, aucune n'a présenté de cytotoxicité sur notre modèle cancéreux mammaire. Cependant, ces molécules bien que peu étudiées dans la littérature ont été décrites pour leurs activités sur d'autres cibles biologiques. L'anicequol et l'andrastine A connus pour être produits par des espèces appartenant au genre *Penicillium* sp. ont été rapportés respectivement pour leur activité antineurotrophique (Hua *et al.*, 2005) et inhibitrice des protéines farnesyl transférase (une cible thérapeutique impliquée dans le cancer) (Omura *et al.*, 1996).

3.2.2.3. Purification ciblée des acides gras

Sachant que le fractionnement bioguidé a conduit à l'isolement de l'ergostérol et qu'un nombre important d'acides gras et de stérols ont été mis en évidence dans la littérature pour leurs activités cytotoxiques sur les lignées cancéreuses du sein, il a donc été choisi de purifier également ce type de composés. Ces derniers étant plutôt apolaires et par conséquent très peu retenus sur une colonne de silice, ils sont alors élués rapidement. Ainsi, en s'appuyant sur les analyses CCM réalisées sur les fractions issues du lot 1 (MMS5-526), il a été supposé en fonction de la polarité que les trois premières fractions étaient composées principalement de triglycérides et de phospholipides (fraction 1), suivie d'acides gras (fraction 2), puis de stérols (fraction 3). En prenant également en considération les résultats obtenus au cours du fractionnement bioguidé réalisé sur la fraction 1-3, il a donc été choisi de se focaliser sur la fraction 1-2 possédant également une forte activité sur l'inhibition de la prolifération des cellules MCF-7 (%inhibition > 50% pour une concentration de 10 µg/mL). Par comparaison des profils CCM obtenus pour les deux lots (Cf. Figure 26), celui de la fraction 1-8 (lot 2 : MMS5-531) est apparu semblable à celui de la fraction 1-2 (lot 1 : MMS5-526) contenant les acides gras (tâche rose sur la CCM) ainsi que très peu complexe avec une masse de 41,5 mg (lot 2).

Une purification réalisée sur cet échantillon a permis d'obtenir 20 fractions, dont les profils chromatographiques obtenus par CCM sont présentés ci-dessous (Cf. Figure 30).

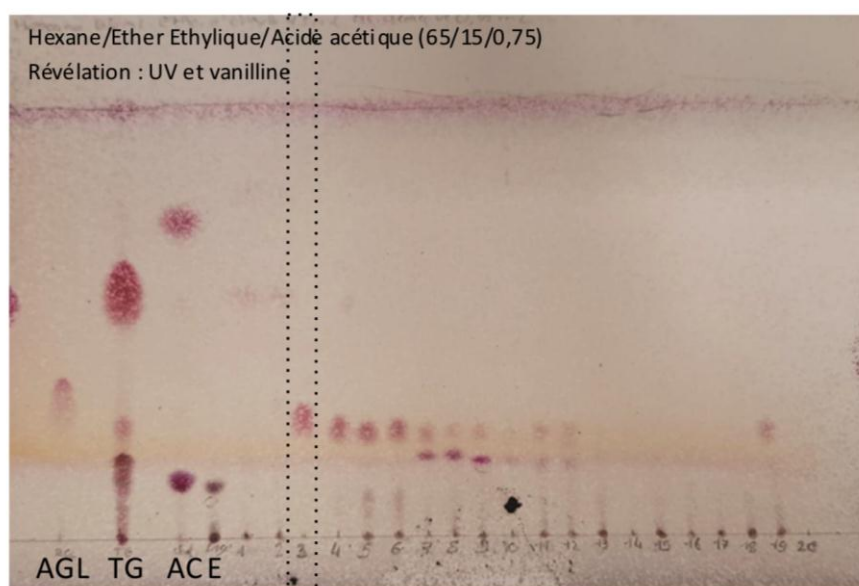


Figure 30. Profils CCM des fractions issues de la purification de l'échantillon MMS5-531-1-8. Comparaison des profils avec les témoins suivant : AGL (Acides gras libres), TG (Tryglycérides), AC (Acétate de cholestérol) et E (Ergostérol).

D'après la Figure 30, il semblerait que la fraction 3 (MMS5-531-8-3) présente un profil chimique peu complexe avec la présence d'une seule tâche correspondant à un acide gras pour une masse de 0,3 mg. Malheureusement l'analyse de cette fraction par GC-MS a révélé que celle-ci était constituée en réalité d'un mélange d'acides gras de composition suivante : acide oléique (43,2%), acide palmitique (26,1%), un C18:0 non identifié (16,5%), un C7:0 non identifié (3,8%), acide linoléique (2,2%) et un C24:0 non identifié (2,0%). En raison, de la masse de cette fraction de seulement 0,3 mg et des faibles pourcentages des autres acides gras composant ce mélange, aucune purification n'a pu être poursuivie.

Néanmoins, l'activité de cette fraction évaluée sur les cellules MCF-7 dans la gamme de concentration de 0,1 à 50 $\mu\text{g}/\text{mL}$ a permis d'obtenir un pourcentage d'inhibition de 54,1% (Cf. Figure 31). Bien que l'activité de cette fraction ne soit pas aussi importante que celle obtenue avec l'ergostérol, il a été impossible de déterminer si l'activité observée pour cette fraction étant dépendante d'un ou plusieurs acides gras.

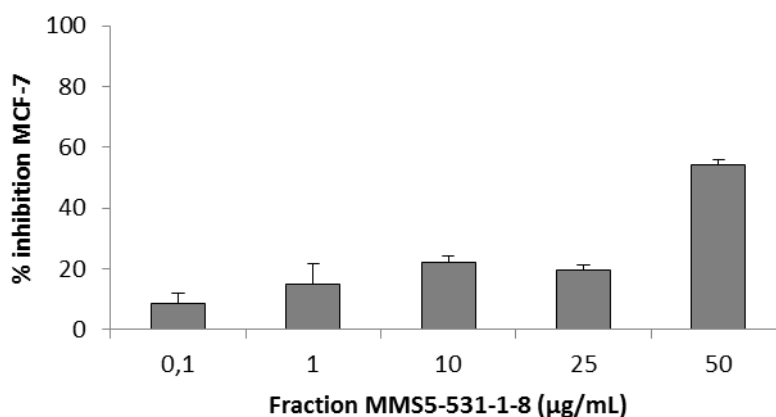


Figure 31. Effet cytotoxique de la fraction MMS5-531-1-8 sur les cellules MCF-7. Les cellules MCF-7 ont été traitées par la fraction MMS5-531-1-8 (0,001 à 50 $\mu\text{g}/\text{mL}$) pendant 24H. Les essais ont été réalisés en triplicata et les valeurs moyennes sont exprimées avec les écarts types associés.

L'ensemble des travaux de purification a été représenté schématiquement sous la forme d'un « arbre de fractionnement » (Cf. Figure 32).

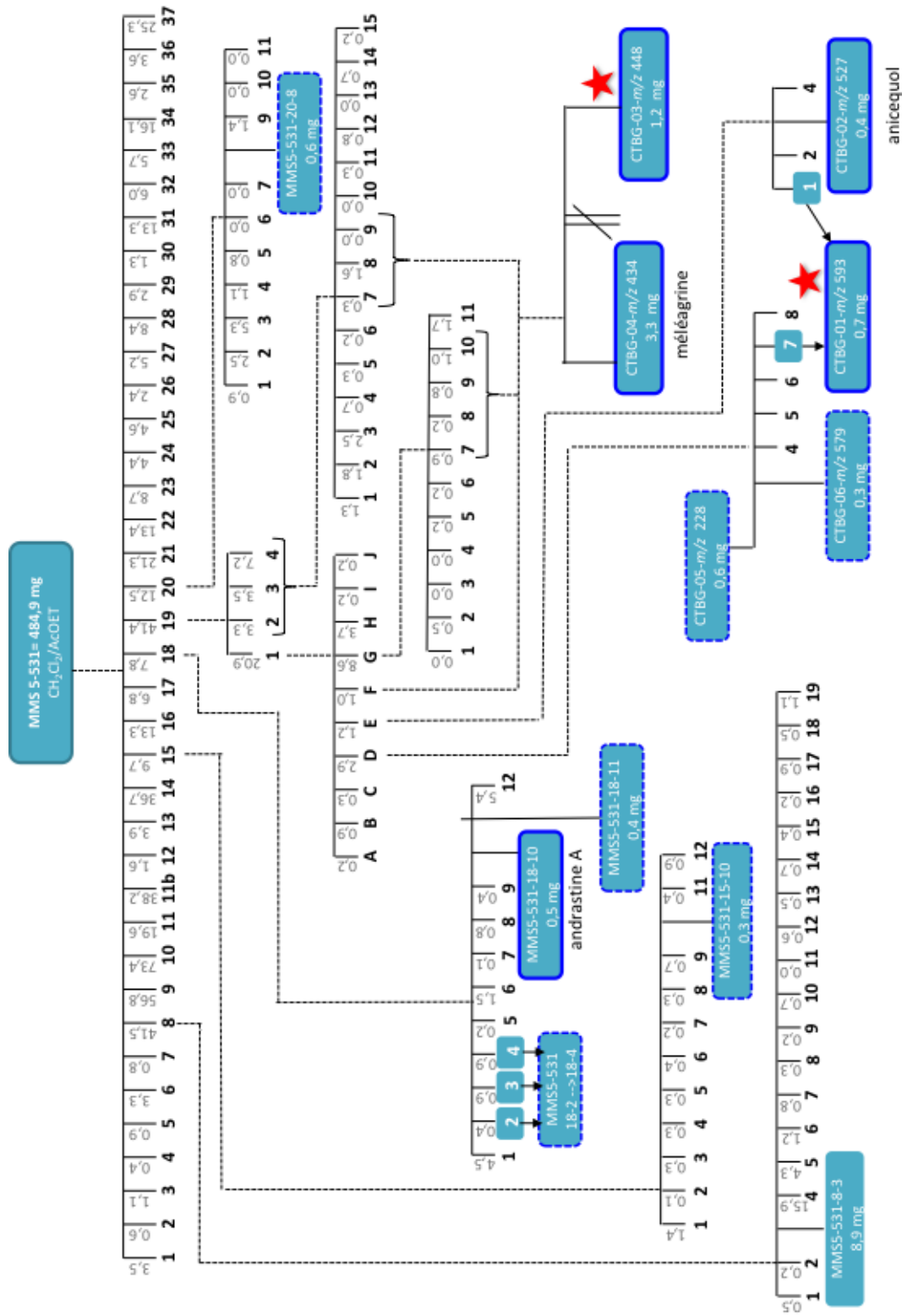


Figure 32. Arbre de fractionnement de l'extrait MMS 5-531 issu de la souche *P. chrysogenum*. Les fractions encadrées en bleu (trait plein), sont celles correspondant à des molécules pures isolées et identifiées par RMN. Les fractions encadrées en pointillées sont celles présentant un bon degré de pureté et pour lesquelles des propositions de formules brutes des molécules présentes ont pu être réalisées à l'aide de la LC-HRMS. L'étoile rouge correspond à la présence d'une molécule potentiellement nouvelle (0 hit dans le DNP). Voir les Annexes correspondantes 15 (MMS5-531-18-10), 14 (CTBG-01-*m/z* 593), 9 (CTBG-02-*m/z* 527), 11 (CTBG-03-*m/z* 448), 7 (CTBG-04-*m/z* 434).

3.2.3. Bilan et perspective de l'étude de la souche MMS5 *P. chrysogenum*

Les différentes approches réalisées ont permis d'isoler une molécule pure responsable de l'activité observée : l'ergostérol. Cinq autres molécules ont également pu être isolées : la méléagrine, l'anicequol et l'andrastine A ainsi que deux potentielles nouvelles molécules dont les élucidations structurales sont en cours. Bien que ces nouvelles molécules ne présentent pas de cytotoxicité sur notre modèle cellulaire cancéreux mammaire, celles-ci pourraient être testées sur d'autres cibles biologiques.

Afin de vérifier la pureté des différentes fractions, l'ensemble des échantillons a été analysé par LC-UV-MS. Pour ceux dont le profil LC-UV et/ou LC-MS semblait présenter un fort potentiel dans la purification de molécules (bon degré de pureté), les données de LC-MS ont été étudiées afin de proposer des formules brutes et des annotations potentielles. Parmi ces molécules détectées dans les fractions les plus pures, il a pu être mis en évidence 4 molécules présentant des hits dans la base donnée du DNP mais également 3 nouvelles molécules potentielles (absence de hit).

Tableau 8. Déréplication des molécules détectées dans les fractions les plus pures pour la souche *P. chrysogenum*

T _R (min)	m/z	Adduit et/ou fragment	UV	Masse monoisotopique	Formule brute	Diff (ppm)	Iso Score	Hits (DNP)	Structures proposées dans le DNP	Initialement présent dans la fraction MIMS5-	Code MIMS5																																																																																																																																
19,8	445.3310	[M+H] ⁺	/	44.3237	C ₂₈ H ₄₄ O ₄	-3.37	85.99	3	Citréospirostéroïde 3β,5α,9α-trihydroxyergosta-7,22-dien-6-one MER-NF-8054A	526-4-3	526-5-23																																																																																																																																
	443.3165	[M-H] ⁻										11,7	290.2685	[M+H] ⁺ ?	/	289.2612	C ₁₆ H ₃₅ NO ₃	-3.10	96.95	2	2-Amino-1,3,4-hexadecanetriol 2-Amino-14-méthyle-1,3,4-pentadecanetriol	526-4-3	526-5-30	395.3302	[M+H] ⁺							23,0	377.3215	[M+H-H ₂ O] ⁺	/	394.3229	C ₂₈ H ₄₂ O	-0.25	91.85	3	1-(10→6)-abeo-ergosta-5,7,9,22-tetraen-3α-ol 14-déhydroergostérol (3β,22E) 6-méthyl-19-norergosta-5,7,9,22-tetraen-3-ol	526-1-3	526-6-11	399.3043	[M+Na] ⁺	/	376.3174	C ₁₇ H ₄₀ N ₆ O ₃	-0.75	100	0	19,2	377.3247	[M+H] ⁺	/	227.1888	C ₁₃ H ₂₅ NO ₂	1.31	78.24	0	Citréohybridonol	526-1-15	531-15-10	228.1961	[M+H] ⁺							10,2	250.1771	[M+Na] ⁺	/	500.2397	C ₂₈ H ₃₆ O ₈	-2.00	88.15	5	Tetrahydrotrichodimérol Tropolactone B Acide confluentique Bisvertinolone; 3-deoxo,2',3'-dihydro	531-1-19	CTBG-05- m/z 228	523.2291	[M+Na] ⁺	210						11,4	501.2497	[M+H] ⁺	236	578.3268	C ₃₄ H ₄₆ O ₇	4.32	89.65	0	CTBG-06- m/z 579	531-1-20	531-20-8	441.2252	[M-59] ⁺	256						12,9	579.3341	[M+H] ⁺	/	578.3268	C ₃₆ H ₄₆ N ₄ O ₃	1.90	90.39	0		531-1-19	CTBG-06- m/z 579	623.3124	[M+FA-H] ⁻								577.3080	[M-H] ⁻					
11,7	290.2685	[M+H] ⁺ ?	/	289.2612	C ₁₆ H ₃₅ NO ₃	-3.10	96.95	2	2-Amino-1,3,4-hexadecanetriol 2-Amino-14-méthyle-1,3,4-pentadecanetriol	526-4-3	526-5-30																																																																																																																																
	395.3302	[M+H] ⁺										23,0	377.3215	[M+H-H ₂ O] ⁺	/	394.3229	C ₂₈ H ₄₂ O	-0.25	91.85	3	1-(10→6)-abeo-ergosta-5,7,9,22-tetraen-3α-ol 14-déhydroergostérol (3β,22E) 6-méthyl-19-norergosta-5,7,9,22-tetraen-3-ol	526-1-3	526-6-11	399.3043	[M+Na] ⁺	/	376.3174	C ₁₇ H ₄₀ N ₆ O ₃	-0.75	100	0	19,2	377.3247	[M+H] ⁺	/	227.1888	C ₁₃ H ₂₅ NO ₂	1.31	78.24	0	Citréohybridonol	526-1-15	531-15-10	228.1961	[M+H] ⁺							10,2	250.1771	[M+Na] ⁺	/	500.2397	C ₂₈ H ₃₆ O ₈	-2.00	88.15	5	Tetrahydrotrichodimérol Tropolactone B Acide confluentique Bisvertinolone; 3-deoxo,2',3'-dihydro	531-1-19	CTBG-05- m/z 228	523.2291	[M+Na] ⁺	210						11,4	501.2497	[M+H] ⁺	236	578.3268	C ₃₄ H ₄₆ O ₇	4.32	89.65	0	CTBG-06- m/z 579	531-1-20	531-20-8	441.2252	[M-59] ⁺	256						12,9	579.3341	[M+H] ⁺	/	578.3268	C ₃₆ H ₄₆ N ₄ O ₃	1.90	90.39	0		531-1-19	CTBG-06- m/z 579	623.3124	[M+FA-H] ⁻								577.3080	[M-H] ⁻																									
23,0	377.3215	[M+H-H ₂ O] ⁺	/	394.3229	C ₂₈ H ₄₂ O	-0.25	91.85	3	1-(10→6)-abeo-ergosta-5,7,9,22-tetraen-3α-ol 14-déhydroergostérol (3β,22E) 6-méthyl-19-norergosta-5,7,9,22-tetraen-3-ol	526-1-3	526-6-11																																																																																																																																
	399.3043	[M+Na] ⁺	/	376.3174	C ₁₇ H ₄₀ N ₆ O ₃	-0.75	100	0				19,2	377.3247	[M+H] ⁺	/	227.1888	C ₁₃ H ₂₅ NO ₂	1.31	78.24	0	Citréohybridonol	526-1-15	531-15-10	228.1961	[M+H] ⁺							10,2	250.1771	[M+Na] ⁺	/	500.2397	C ₂₈ H ₃₆ O ₈	-2.00	88.15	5	Tetrahydrotrichodimérol Tropolactone B Acide confluentique Bisvertinolone; 3-deoxo,2',3'-dihydro	531-1-19	CTBG-05- m/z 228	523.2291	[M+Na] ⁺	210						11,4	501.2497	[M+H] ⁺	236	578.3268	C ₃₄ H ₄₆ O ₇	4.32	89.65	0	CTBG-06- m/z 579	531-1-20	531-20-8	441.2252	[M-59] ⁺	256						12,9	579.3341	[M+H] ⁺	/	578.3268	C ₃₆ H ₄₆ N ₄ O ₃	1.90	90.39	0		531-1-19	CTBG-06- m/z 579	623.3124	[M+FA-H] ⁻								577.3080	[M-H] ⁻																																													
19,2	377.3247	[M+H] ⁺	/	227.1888	C ₁₃ H ₂₅ NO ₂	1.31	78.24	0	Citréohybridonol	526-1-15	531-15-10																																																																																																																																
	228.1961	[M+H] ⁺										10,2	250.1771	[M+Na] ⁺	/	500.2397	C ₂₈ H ₃₆ O ₈	-2.00	88.15	5	Tetrahydrotrichodimérol Tropolactone B Acide confluentique Bisvertinolone; 3-deoxo,2',3'-dihydro	531-1-19	CTBG-05- m/z 228	523.2291	[M+Na] ⁺	210						11,4	501.2497	[M+H] ⁺	236	578.3268	C ₃₄ H ₄₆ O ₇	4.32	89.65	0	CTBG-06- m/z 579	531-1-20	531-20-8	441.2252	[M-59] ⁺	256						12,9	579.3341	[M+H] ⁺	/	578.3268	C ₃₆ H ₄₆ N ₄ O ₃	1.90	90.39	0		531-1-19	CTBG-06- m/z 579	623.3124	[M+FA-H] ⁻								577.3080	[M-H] ⁻																																																																	
10,2	250.1771	[M+Na] ⁺	/	500.2397	C ₂₈ H ₃₆ O ₈	-2.00	88.15	5	Tetrahydrotrichodimérol Tropolactone B Acide confluentique Bisvertinolone; 3-deoxo,2',3'-dihydro	531-1-19	CTBG-05- m/z 228																																																																																																																																
	523.2291	[M+Na] ⁺	210									11,4	501.2497	[M+H] ⁺	236	578.3268	C ₃₄ H ₄₆ O ₇	4.32	89.65	0	CTBG-06- m/z 579	531-1-20	531-20-8	441.2252	[M-59] ⁺	256						12,9	579.3341	[M+H] ⁺	/	578.3268	C ₃₆ H ₄₆ N ₄ O ₃	1.90	90.39	0		531-1-19	CTBG-06- m/z 579	623.3124	[M+FA-H] ⁻								577.3080	[M-H] ⁻																																																																																					
11,4	501.2497	[M+H] ⁺	236	578.3268	C ₃₄ H ₄₆ O ₇	4.32	89.65	0	CTBG-06- m/z 579	531-1-20	531-20-8																																																																																																																																
	441.2252	[M-59] ⁺	256									12,9	579.3341	[M+H] ⁺	/	578.3268	C ₃₆ H ₄₆ N ₄ O ₃	1.90	90.39	0		531-1-19	CTBG-06- m/z 579	623.3124	[M+FA-H] ⁻								577.3080	[M-H] ⁻																																																																																																									
12,9	579.3341	[M+H] ⁺	/	578.3268	C ₃₆ H ₄₆ N ₄ O ₃	1.90	90.39	0		531-1-19	CTBG-06- m/z 579																																																																																																																																
	623.3124	[M+FA-H] ⁻											577.3080	[M-H] ⁻																																																																																																																													
	577.3080	[M-H] ⁻																																																																																																																																									

Les analyses LC-UV-MS (Cf. Annexe 17) réalisées sur l'échantillon **MMS5-526-5-23** (2,3 mg) ont permis de mettre en évidence deux molécules dont l'ergostérol. En effet, les caractéristiques UV décrites précédemment pour cette molécule ont permis rapidement d'identifier sa présence au sein de cette fraction. Cependant, une seconde molécule avec un temps de rétention de 20 min de formule brute $C_{28}H_{44}O_4$ sans absorption UV a pu être repérée avec la présence des ions m/z 445.3310 et m/z 443.3165 correspondant respectivement aux espèces moléculaires $[M+H]^+$ et $[M-H]^-$. Des données supplémentaires de fragmentation de l'ion le plus intense soit m/z 445.3310 ont également permis de mettre en évidence la présence d'un fragment m/z 427.3290. Après une recherche dans les bases de données réalisée à partir de la formule brute, cette molécule semble correspondre à trois molécules potentielles (Cf. Figure 33) déjà isolées dans des ascomycètes : le citréospirostéroïde, la $3\beta,5\alpha,9\alpha$ -trihydroxyergosta-7,22-dien-6-one et le $3',4',5',17$ -tetrahydro-4'-(1-hydroxy-2,3-diméthylbutyl)-3'-méthylcyclopenta[13,17]-18-norandrost-6,8(14)-diène-3,5,9-triol (autrement appelé Mer-NF-8054A). En combinant les résultats obtenus, il semblerait que la première proposition de molécule soit la plus probable en raison de l'absence d'absorption UV. Cette molécule isolée à partir d'une espèce fongique appartenant au genre *Penicillium* sp. (Kosemura and Yamamura, 1997) n'a jamais été évaluée pour son effet sur une cible en lien avec le cancer du sein.

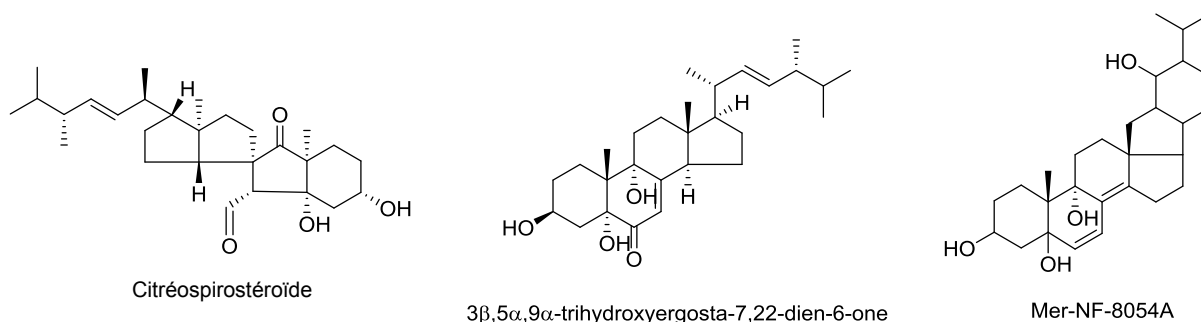
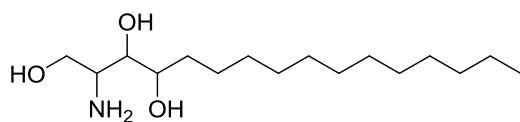
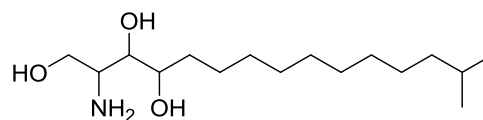


Figure 33. Molécules proposées dans la base de donnée du DNP, après recherche de formule brute $C_{28}H_{44}O_4$ pour les ascomycètes.

De la même façon pour l'échantillon **MMS5-526-5-30** (0,2 mg) il a pu être détecté au cours des analyses LC-MS (Cf. Annexe 18) la présence de l'ion m/z 290.2685 $[M+H]^+$, correspondant probablement à la forme ionisée d'une molécule de formule brute $C_{16}H_{35}NO_3$ dont le temps de rétention est de 11,3 min. Les données de fragmentation acquises pour cet ion, ont également permis de mettre en évidence la présence d'un fragment de m/z 242.2490 $[M-47]^+$. En comparaison avec le DNP, deux hits ont été obtenus : le 2-amino-1,3,4-hexadecanetriol et 2-amino-14-méthyle-1,3,4-pentadecanetriol (Cf. Figure 34).



2-amino-1,3,4-hexadecanetriol

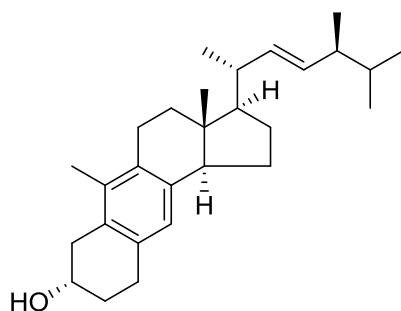


2-amino-14-méthyle-1,3,4-pentadecanetriol

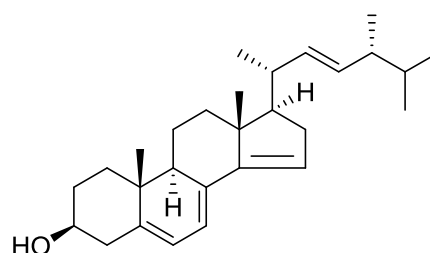
Figure 34. Molécules proposées dans la base de donnée du DNP, après recherche de formule brute $C_{16}H_{35}NO_3$ pour les ascomycètes.

Pour l'échantillon **MMS5-526-6-11** (1,3 mg), une molécule avec un temps de rétention de 23,0 min et de formule brute obtenue $C_{28}H_{42}O$ a pu être repérée au cours des analyses LC-UV-MS (Cf. Annexe 19), avec la présence des ions m/z 395.3302 et m/z 377.3215 correspondant respectivement à la molécule protonnée $[M+H]^+$ et au fragment $[M-H_2O+H]^+$. Après une recherche dans les bases de données réalisée à partir de la masse monoisotopique, cette molécule semble correspondre à quatre molécules potentielles (Cf.

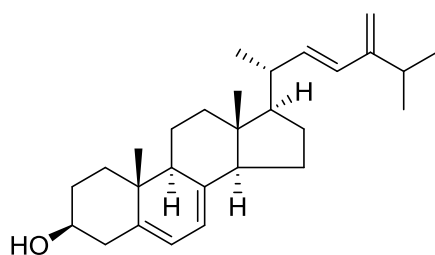
Figure 35) déjà isolées dans des ascomycètes : le 1-(10→6)-abeo-ergosta-5,7,9,22-tetraen-3 α -ol, le 14-déhydroergostérol, le (3 β ,22E) ergosta-5,7,22,24(28)-tetraen-3-ol et le (3 β ,22E) 6-méthyl-19-norergosta-5,7,9,22-tetraen-3-ol.



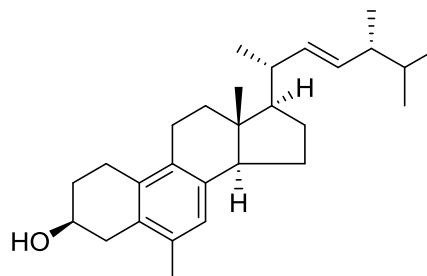
1-(10→6)-abeo-ergosta-5,7,9,22-tetraen-3 α -ol



14-déhydroergostérol



(3 β ,22E) ergosta-5,7,22,24(28)-tetraen-3-ol



(3 β ,22E) 6-méthyl-19-norergosta-5,7,9,22-tetraen-3-ol

Figure 35. Molécules proposées dans la base de donnée du DNP, après recherche de formule brute $C_{28}H_{42}O$ pour les ascomycètes.

En s'appuyant sur les données UV obtenues, celles-ci ont montré la présence d'une absorption à la longueur d'onde de 190 nm, aucune des structures proposées ne peut correspondre.

Néanmoins, cette molécule de pureté insuffisante pour être analysée en RMN nécessite des étapes de purifications supplémentaires qui n'ont pu être réalisées dans le cadre de ce projet.

Il a pu être remarqué que la molécule détectée dans l'échantillon **MMS5-531-20-8** présentait la même ionisation que celle d'une molécule pure isolée et identifiée comme étant l'andrastine A. En effet, la présence des adduits similaires tels que $[M+Na]^+$, $[M+H]^+$, $[M-59]^+$, $[M-H]^-$ a été observée (Cf. Annexe 19). Par conséquent, il est probable que cette molécule soit un analogue de l'andrastine A. Par comparaison avec le DNP, cinq propositions de structures ont été faites (Cf. Figure 36).

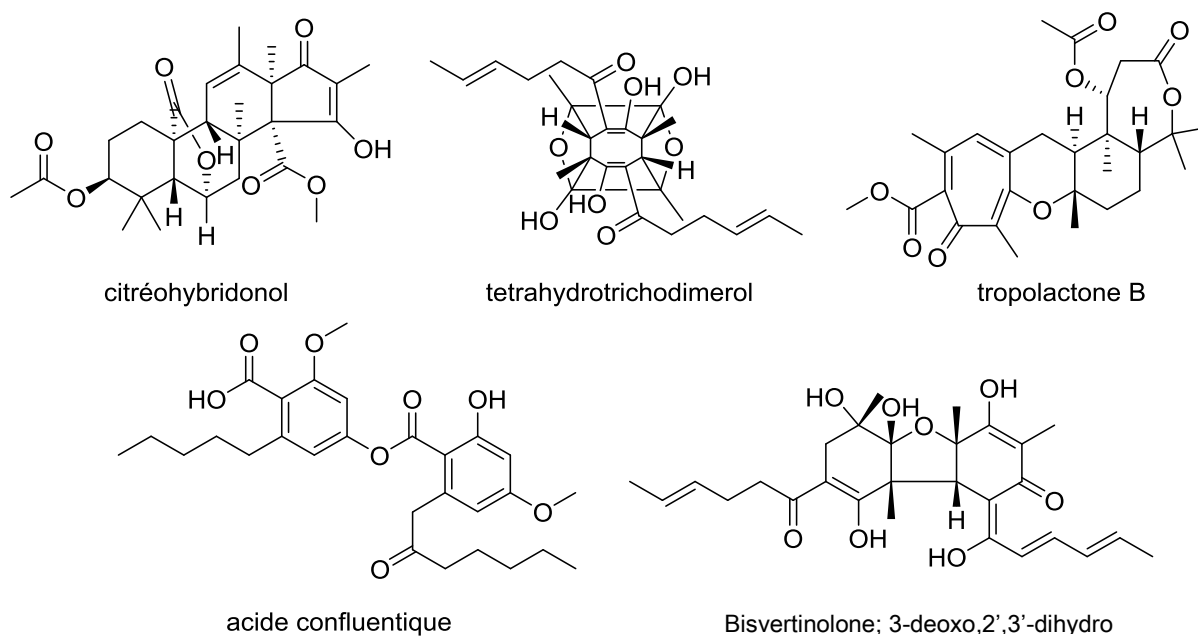


Figure 36. Molécules proposées dans la base de donnée du DNP, après recherche de formule brute $C_{28}H_{36}O_8$ pour les ascomycètes.

Enfin, trois molécules potentiellement nouvelles ont pu être détectées au cours des analyses dans les échantillons **MMS5-531-15-10**, **CTBG-05-*m/z* 228** et **CTBG-06-*m/z* 579**. Cependant, en raison d'une masse insuffisante aucune purification ciblée sur ces molécules n'a pu être réalisée.

3.3. Étude de la souche *Didymella americana*

Pour la souche *D. americana*, trois lots de cultures ont été réalisés, mais comme il a pu être exposé au début de ce chapitre, seul les extraits bruts MMS719-527 (Lot1) et MMS719-530 (Lot2) ont montré une activité sur la prolifération des cellules MCF-7. Par ailleurs, la superposition des profils chromatographiques (Cf. Annexe 20) a permis de mettre en évidence une certaine hétérogénéité entre les profils. Ainsi, il a été choisi de mener séparément des investigations chimiques sur ces deux lots issus de la souche MMS719 *Didymella americana* ayant présente une activité. Ces investigations chimiques ont été réalisées selon une approche de fractionnement bioguidé et de profilage métabolique.

3.3.1. Fractionnement bioguidé de l'extrait MMS719-527

Un fractionnement réalisé à l'aide de la chromatographie flash sur l'extrait brut MMS719-527 a permis d'obtenir 20 fractions regroupées selon une analyse par CCM. Pour un suivi de l'activité au cours du fractionnement, la cytotoxicité a été déterminée sur le modèle cellulaire cancéreux mammaire de la lignée MCF-7 à l'aide du test MTT. D'après les résultats obtenus sur l'ensemble de ce premier fractionnement, étonnamment seule la fraction MMS719-527-1-14 (parmi les 17 testées) a présenté une activité intéressante avec une valeur d'inhibition sur la prolifération des cellules MCF-7 de 54% à la concentration de 10 µg/mL (Cf. Figure 37).

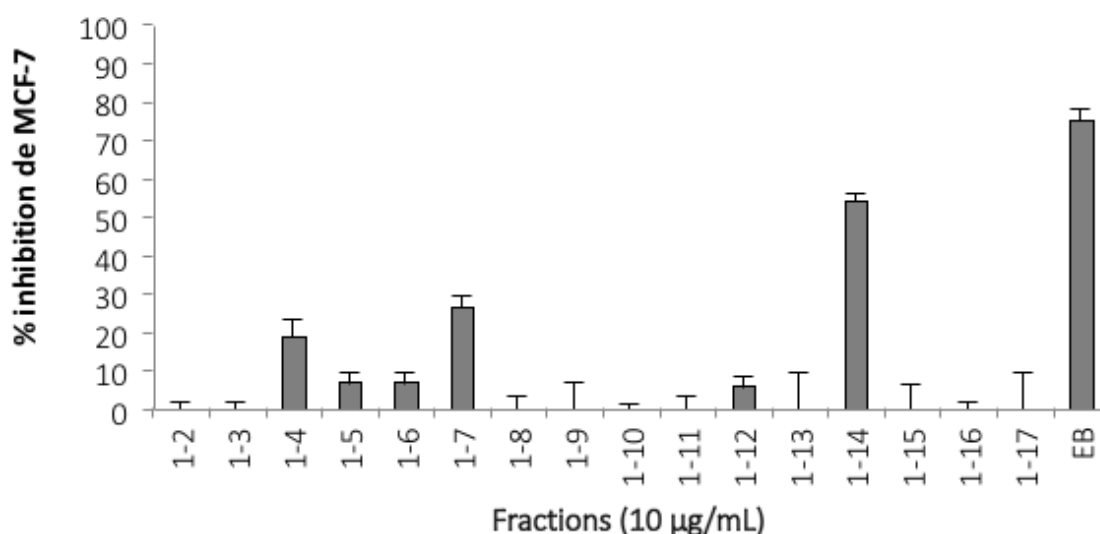


Figure 37. Cytotoxicité des 17 fractions (1-1→1-17) obtenues au cours du fractionnement de l'extrait brut MMS719-527 issu de la souche MMS719 *D. americana* sur la prolifération des cellules MCF-7. Pour chaque fraction testée à la concentration de 10 µg/mL, le pourcentage d'inhibition de la prolifération des cellules MCF-7 a été déterminé après un traitement de 24H.

En parallèle, afin d'évaluer rapidement et facilement la complexité chimique des différentes fractions et en particulier celle de la fraction MMS719-527-14, des analyses par CCM ont été réalisées (Cf. Figure 38).

D'après les résultats de la CCM, il a pu être observé des profils chromatographiques très différents pour les différentes fractions analysées, suggérant ainsi que le fractionnement de l'extrait brut avait permis de réaliser une bonne séparation des composés de polarité croissante. En s'intéressant au profil chromatographique de la fraction 14, celui-ci est apparu comme peu complexe (peu de spots sur la CCM). Par ailleurs, il a tout de suite été remarqué que cette fraction comportait des molécules d'une classe chimique bien différente de celle des stéroïdes ou des acides gras. En effet, d'après les travaux réalisés sur la souche précédente, il a pu être mis en évidence que les stéroïdes et les acides gras (caractérisés par la présence de larges spots de couleur rose après révélation chimique à la vanilline) sont plutôt localisés en début d'élution et par conséquent dans les premières fractions collectées, lors d'un fractionnement réalisé sur colonne de silice. De la même façon il a pu être observé dans les fractions 1 à 5 correspondant aux lipides, l'absence d'ergostérol (caractérisé par la présence d'un spot de couleur violet par révélation à la vanilline), en accord avec les résultats précédents de quantification par GC-MS d'extraits issus de cette même souche (cf § 1.1.1.3).

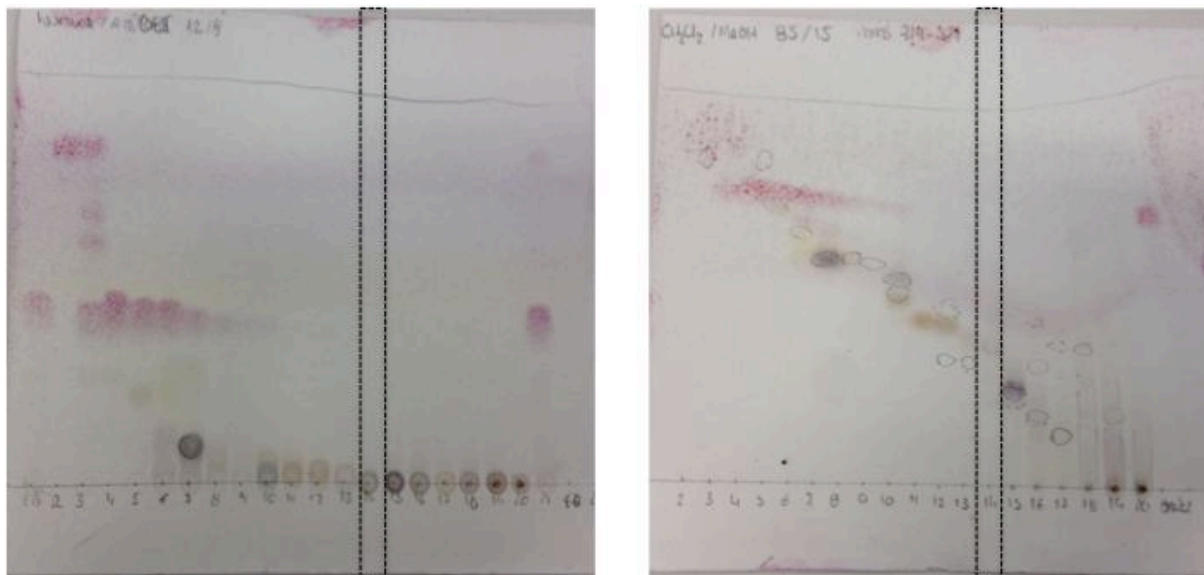


Figure 38. Profils CCM des fractions issues du fractionnement de l'extrait MMS719-527, éluant hexane/AcOEt (12 : 8, v/v) (gauche) et $\text{CH}_2\text{Cl}_2/\text{MeOH}$ (85 : 15, v/v) (droite).

Dans le but d'isoler la molécule responsable de l'activité observée, un fractionnement a donc été réalisé ensuite sur la fraction MMS719-527-1-14. Au cours de la préparation de l'échantillon solubilisé dans le solvant MeOH, des phénomènes de précipitations ont conduit à la séparation de cette fraction en deux « sous fractions ». Ainsi, le fractionnement réalisé sur chacune d'elle, selon le même gradient d'élution mais des collectes différentes réalisées en mode manuelle et dans l'autre cas en mode automatique, ont permis d'obtenir respectivement 23 et 26 fractions. L'activité de ces fractions a également été déterminée sur le modèle cellulaire MCF-7 (Cf. Figure 39).

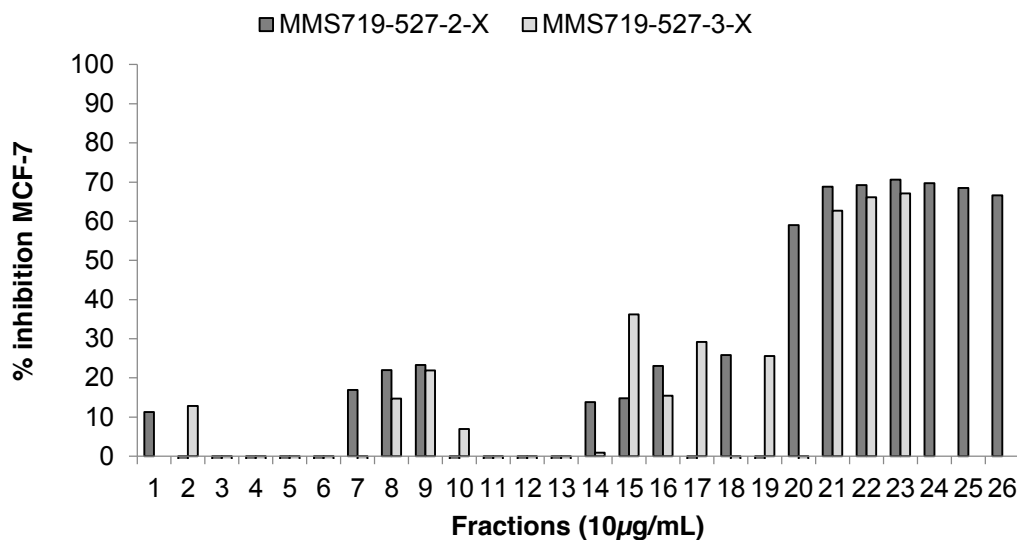


Figure 39. Cytotoxicité des 23 (2-1→2-23) et 26 fractions (3-1→3-23) obtenues au cours du fractionnement de la fraction MMS719-527-14 sur la prolifération des cellules MCF-7. Pour chaque fraction testée à la concentration de 10 µg/mL, le pourcentage d'inhibition de la prolifération des cellules MCF-7 a été déterminé après un traitement de 24H.

D'après les résultats obtenus, une inhibition de la prolifération des cellules MCF-7 commune aux deux fractionnements a pu être observée pour les fractions 21, 22, et 23 issus des deux fractionnements. Cependant, les investigations chimiques n'ont pu être poursuivies en raison de problèmes d'appareillage et par manque de temps.

Pour faire le bilan : 2 étapes de fractionnement ont été réalisées sur l'extrait MMS719-527, permettant d'obtenir un total de 69 échantillons. L'ensemble des travaux de purification a été représenté schématiquement sous la forme d'un « arbre de fractionnement » (Cf. Figure 40)

Néanmoins aucune molécule bioactive et de pureté suffisante n'a pu être isolée. Cependant, l'observation de cristaux dans la fraction MMS527-2-9 obtenue après fractionnement de la fraction MMS719-527-1-14 a suscité notre intérêt.

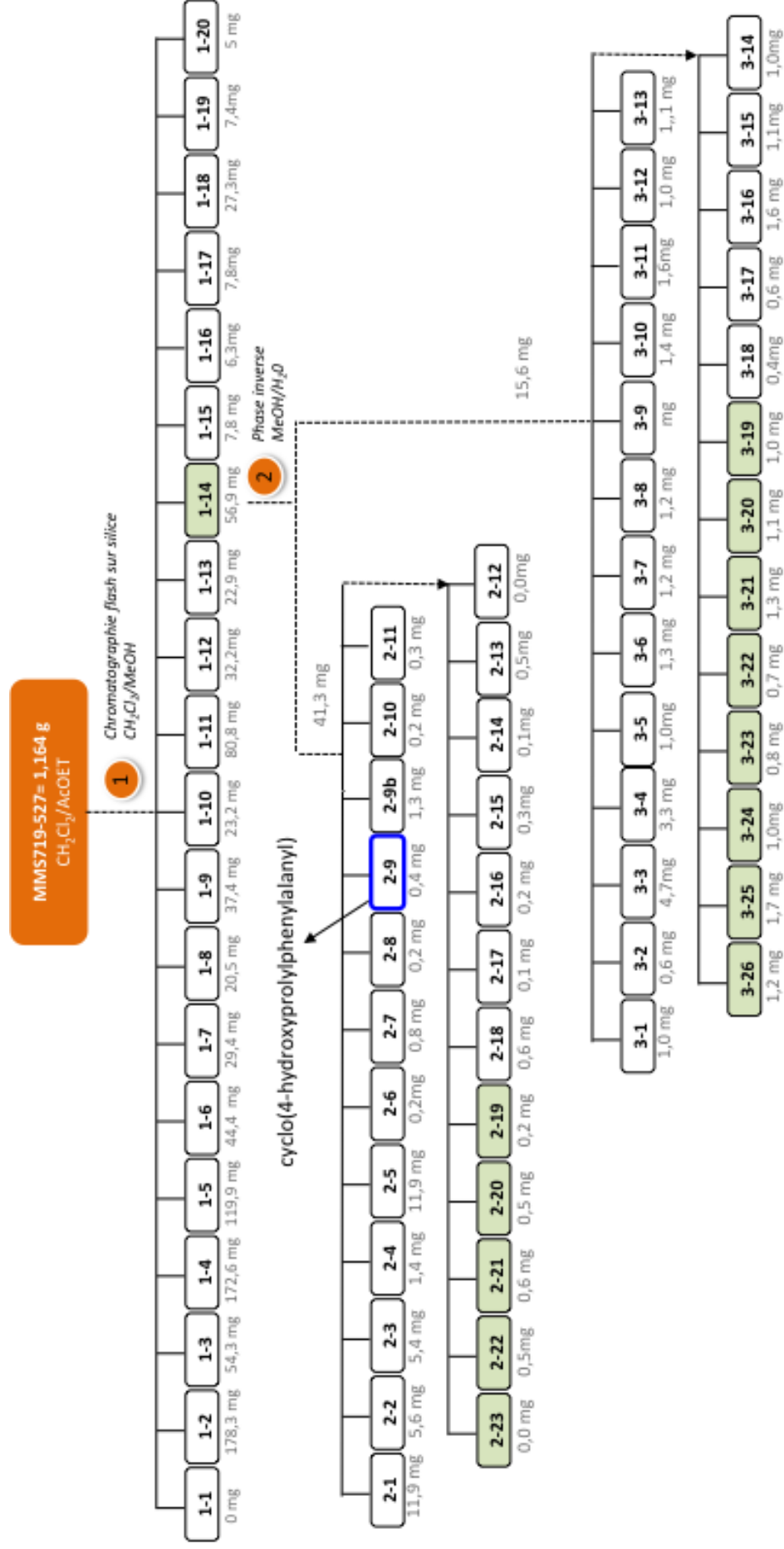


Figure 40. Arbre de fractionnement de l'extrait MMS719-527 issu de la souche *D. Americana*. Les numéros de 1 et 2 correspondent aux différentes étapes de fractionnement. La fraction 1-14 colorée en vert est la seule fraction de prolifération des cellules MCF-7 est supérieure à 50%, après un traitement de 24H réalisé à la concentration de 10 µg/mL. Les fractions encadrées en bleu (trait plein), sont celles correspondant à des molécules pures isolées et identifiées par RMN.

En effet, il arrive au cours d'un fractionnement bioguidé que des molécules présentes dans les échantillons collectés puissent re-cristalliser si les conditions le permettent (pureté, température etc.). Ainsi, au cours de la deuxième étape de fractionnement réalisée sur l'extrait MMS719-527 (lot 1), il a pu être observé pour la fraction MMS719-527-2-9 (0,4 mg) la présence de cristaux. Ce phénomène a suscité notre intérêt et nous avons donc choisi d'identifier cette molécule par analyse par RMN. Ainsi, il a pu être mis en évidence la présence d'au moins deux molécules dont une a été identifiée en comparaison avec la littérature comme étant le cyclo(4-hydroxypropylphenylalanyl) ($C_{14}H_{16}N_2O_3$), les signaux obtenus sur les spectre RMN 1H et ^{13}C (Cf. Figure 41 et Cf. Annexe 20). Néanmoins, cette fraction n'a présenté aucune activité sur le modèle de cellules cancéreuses mammaires MCF-7.

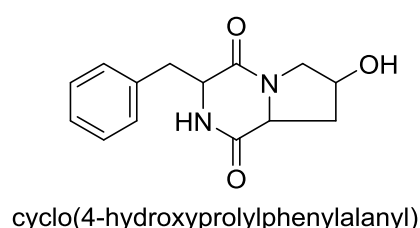


Figure 41. Structure du cyclo(4-hydroxypropylphenylalanyl)

3.3.2. Fractionnement bioguidé de l'extrait MMS719-530

Comme il a pu être énoncé précédemment, un fractionnement bioguidé a également été réalisé sur un deuxième extrait brut, MMS719-530 (lot 3, 298 mg fractionnés sur les 345,5 mg d'extrait). Ce dernier fractionnement, réalisé à l'aide d'une chromatographie liquide basse pression a permis d'obtenir 20 fractions dont l'activité sur la prolifération des cellules MCF-7 a été évaluée à la concentration de 100 $\mu g/mL$.

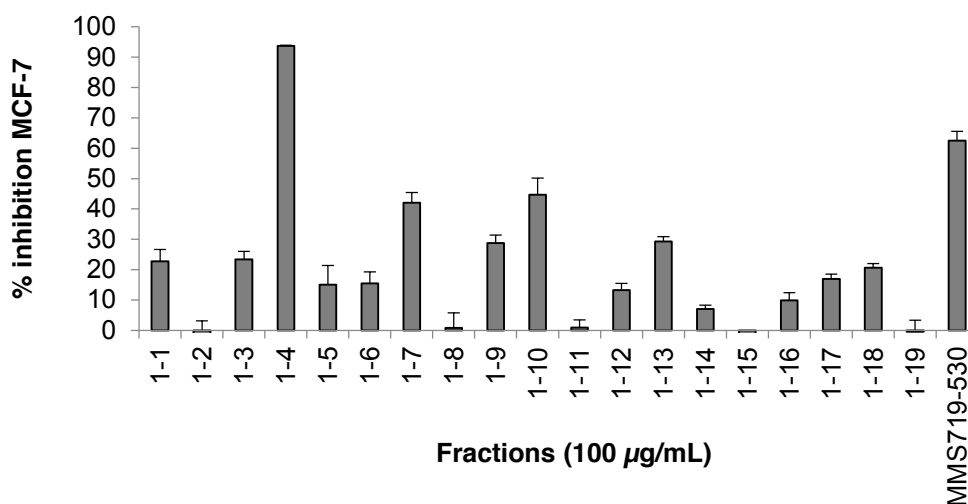


Figure 42. Cytotoxicité des 19 fractions (1-1→1-19) obtenues au cours du fractionnement de l'extrait brut MMS719-527 issu de la souche MMS719 *D. americana* sur la prolifération des cellules MCF-7. Pour chaque fraction testée à la concentration de 100 $\mu g/mL$, le pourcentage d'inhibition de la prolifération des cellules MCF-7 a été déterminé après un traitement de 24H.

D'après les résultats obtenus (Cf. Figure 42), la plus forte inhibition de la prolifération des cellules MCF-7 avec une valeur supérieure à celle de l'extrait brut a été obtenue pour la fraction MMS719-527-1-4 (93% d'inhibition à la concentration de 100 µg/mL). En comparaison avec le lot précédent, dont les conditions d'élution lors du fractionnement étaient sensiblement les mêmes (Cf. §5.4.2), il a été remarqué que l'activité observée n'était pas localisée dans les mêmes fractions (MMS719-527-1-14). Ainsi, il a été suggéré que la ou les molécules responsable(s) de l'activité n'étaient probablement pas les mêmes.

En parallèle, afin d'évaluer rapidement et facilement la complexité chimique des différentes fractions et en particulier celle de la fraction MMS719-530-1-4, des analyses par CCM ont été réalisées (Cf. Figure 43)

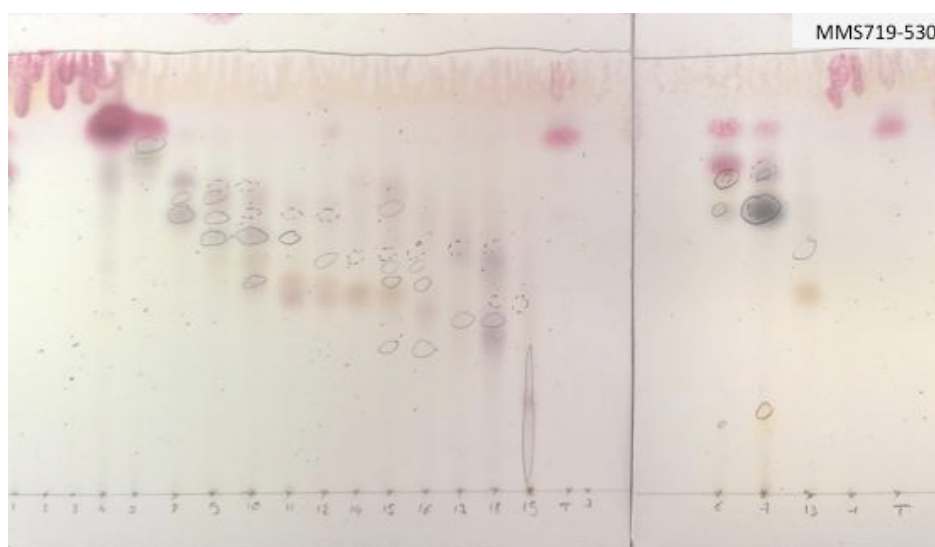


Figure 43. CCM des 19 fractions obtenues après fractionnement de l'extrait brut MMS719-530 issu de la souche *D. americana*. Dépôt de 3µL (5mg/mL), Eluant CH₂CL₂/MeOH (9:1), Révélation : UV (254 et 360 nm) et chimique (vanilline).

D'après les résultats obtenus pour la CCM (attention aux numéros de fractions 6-7-13 et 1 qui n'ont pas été déposées dans l'ordre d'élution), il semblerait que cette fraction MMS719-527-1-4 soit composée de molécules lipidiques. Dans le but d'isoler la ou les molécules responsables de l'activité observée, un fractionnement a donc été réalisé sur celle-ci (116,1 mg). Puis, selon le schéma classique de fractionnement bioguidé, l'activité des différentes fractions obtenues (numérotées de MMS719-2-1 à MMS719-2-18) a été déterminée.

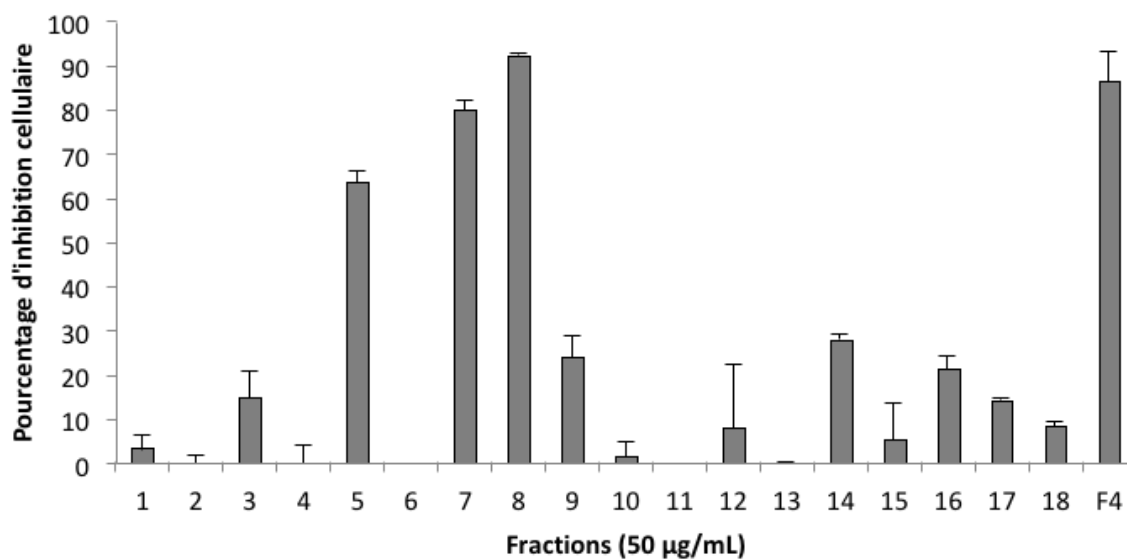


Figure 44. Cytotoxicité des 18 fractions obtenues au cours du fractionnement 2 de l'extrait brut MMS719-530 issu de la souche MMS719 *D.americana* sur la prolifération des cellules MCF-7. Les fractions ont été testées à la concentration de 50 µg/mL et le pourcentage d'inhibition de la prolifération des cellules MCF-7 a été déterminé après un traitement de 24H.

Selon le test MTT réalisé sur cette deuxième étape de fractionnement, il a pu être observé une inhibition de la prolifération des cellules MCF-7 d'une valeur supérieure à 50% pour les fractions MMS719-530-2-5, MMS719-530-2-7 et MMS719-530-2-8.

Afin d'évaluer la pureté des fractions obtenues, l'ensemble des échantillons a été analysé par LC-UV-MS. Malheureusement aucun échantillon n'a présenté de profils LC-MS ou LC-UV avec un seul pic chromatographique laissant supposer la présence d'une seule molécule.

Pour expliquer ces différences d'activités observées entre les deux lots, il a été émis différentes hypothèses, telles qu'une composition chimique différente entre les deux extraits (des molécules différentes et/ou en concentrations différentes) ou encore la présence potentielle d'un effet synergique. En effet, il est probable au cours des fractionnements respectifs réalisés sur les deux lots que les molécules n'aient pas été collectées de la même façon, entraînant ainsi des compositions chimiques différentes de chacune des fractions. Il a donc été choisi de réaliser un profilage métabolique des deux extraits brut afin d'identifier les métabolites majoritairement produits par cette souche.

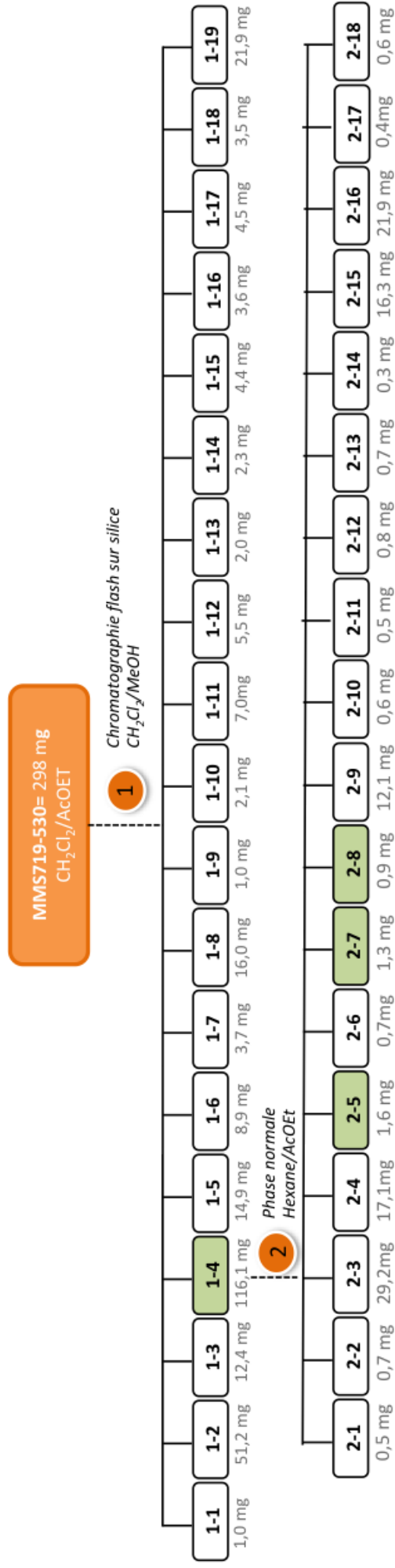


Figure 45. Arbre de fractionnement de l'extrait MMS719-530 issu de la souche *D. americana*. Les numéros 1 et 2 correspondent aux différentes étapes de fractionnement. Les fractions colorées en vert sont celles ayant la meilleure activité sur l'inhibition de la prolifération des cellules MCF-7 (> 50% pour une concentration testée de 50 µg/mL pour le fractionnement 1 et 100 µg/mL pour le fractionnement 2).

3.3.3. Profilages métaboliques des composés majoritaires de l'extrait MMS719-527

Afin de mieux évaluer la chimiodiversité produite par la souche MMS719 *D. americana*, un profilage métabolique a été réalisé sur l'extrait MMS719-527 (lot1). Ce dernier a permis d'annoter les molécules et/ou de mettre en évidence de nouvelles molécules au sein de ce mélange complexe sans que celles-ci ne soient purifiées. D'après les profils LC-MS obtenus pour cet extrait en mode positif (ESI+) et négatif (ESI-), 24 pics chromatographiques majoritaires ont pu être mis en évidence sur le chromatogramme (Cf. Figure 46).

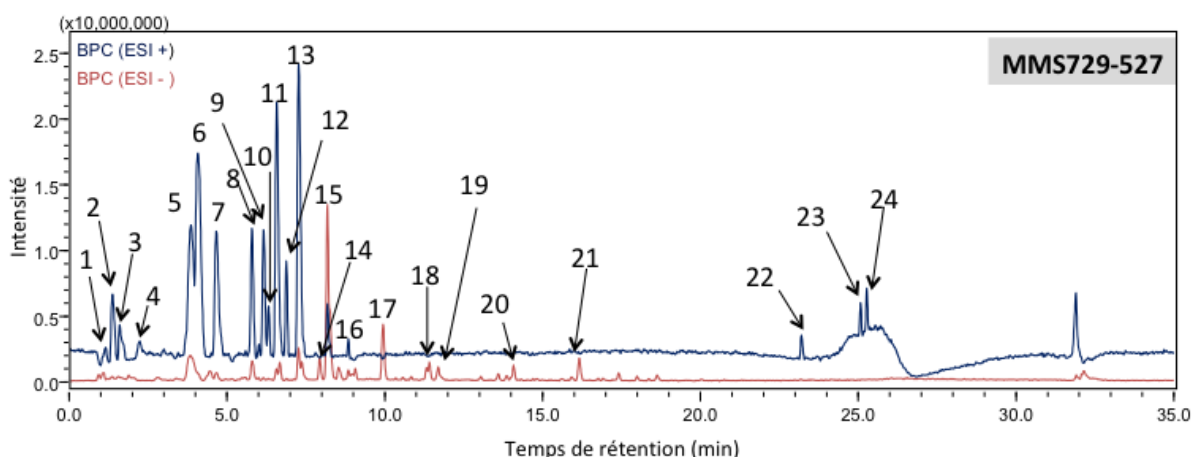


Figure 46. Profils LC-HRMS de l'extrait brut MMS719-527 obtenus en mode positif (ESI+) et négatif (ESI-).

Le profilage chimique réalisé sur l'extrait MMS719-527 suivi d'une étape de déréplication a permis d'annoter seulement 6 pics chromatographiques sur les 24 numérotés sur le chromatogramme (Cf. Figure 46). D'après ce premier résultat et l'absence de données dans la littérature scientifique, celui-ci semble confirmer le potentiel de cette souche dans la production de molécules nouvelles.

Selon le nombre de hit obtenus dans le DNP à partir des formules brutes déterminées pour ces 6 pics, il a pu être mis en évidence une ou plusieurs molécules, dont l'ensemble des structures chimiques sont présentées ci-après (Cf. Figure 47).

Les ions majoritairement détectés au cours des analyses LC-MS associés à des masses monoisotopiques très faibles (environ 200) ont tous été par comparaison des formules brutes associées dans la base de données du DNP annotés comme étant la forme ionisée de composés principalement identifiés comme étant des dicétopiperazines et pipérazines.

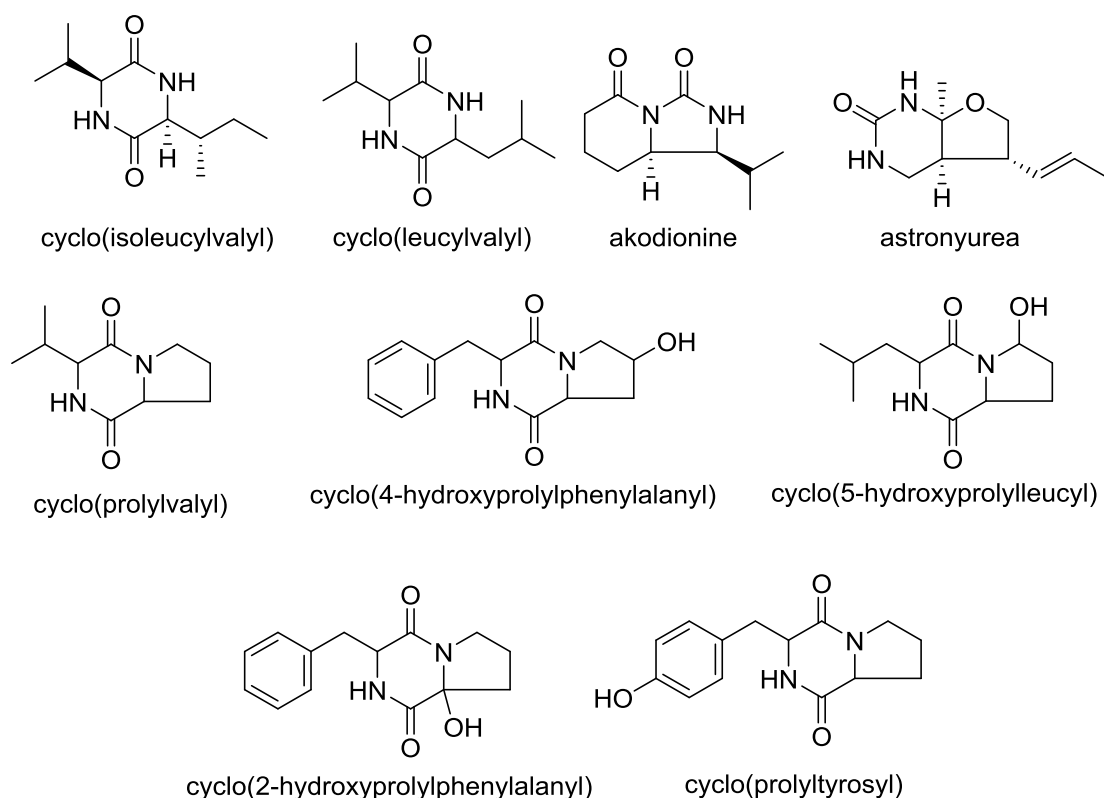


Figure 47. Molécules annotées sur le profil LC-HRMS de l'extrait MMS719-527 en comparaison des formules brutes obtenues avec la base donnée du DNP.

D'après des recherches bibliographiques, aucune des molécules identifiées ou annotées n'a semblé être rapportée dans la littérature pour une activité antitumorale sur des lignées du cancer du sein. Pourtant de nombreuses activités biologiques ont pu être démontrées pour les cyclopeptides tel qu'une activité cytotoxique, antitumorale, antivirale, antifongique, antioxidante, antibiofilm (Wang *et al.*, 2017).

Cependant, le fractionnement bioguidée n'ayant permis d'obtenir aucune molécule pure responsable de l'activité pour cet extrait ni pour celui du deuxième lot MMS719-530 et par manque de temps, les investigations chimiques sur cette souche n'ont pas été poursuivies.

Tableau 9. Déréplication des ions majoritaires détectés au cours des analyses LC-MS de l'extrait brut MMS719-527.

Pics	T _R (min)	m/z	Adduit et/ou fragment	Masse monoisotopique	Formule brute	Diff (ppm)	Iso Score*	Hits (DNP)	Structures proposées dans le DNP
1	1,1	235.1206	[M+Na] ⁺ ? [M+K] ⁺ ?	212.1314 196.1575	C ₁₄ H ₁₆ N ₂ C ₁₁ H ₂₀ N ₂ O	0.00 -0.43	80.72 48.34	0 2	Cyclo(isoleucylvalyl) Cyclo(leucylvalyl)
2	1,4	155.0825 319.0495	? (ESI+) ? (ESI-)	ND ⁺					
3	1,6	169.0972	? (ESI+)	ND ⁺					
4	2,2	171.1138 195.1144	? (ESI+)	ND ⁺					
5	3,8	261.1247 283.1071 248.9609	[M+H] ⁺ [M+Na] ⁺ ? (ESI-)	ND ⁺					
6	4,1	197.1292	[M+H] ⁺ ?	196.1219	C ₁₀ H ₁₆ N ₂ O ₂	3.55	85.06	4	Akodonine Astronyurea Cyclo(propylvalyl)
7	4,6	227.1395	[M+H] ⁺ ?	226.1322	C ₁₁ H ₁₈ N ₂ O ₃	2.20	84.79	3	Cyclo(4-hydroxypropylleucyl) Cyclo(5-hydroxypropylleucyl)

* Iso Score : score de similitude de profils isotopiques entre le profil observé et le profil théorique (score compris entre 0 et 100)

+ ND : Non déterminée car les formules brutes proposées présenteraient des scores de profils isotopiques (Iso Score) inférieurs à 50%

Pics	T _R (min)	m/z	Adduit et/ou fragment	Masse monoisotopique	Formule brute	Diff (ppm)	Iso Score*	Hits (DNP)	Structures proposées dans le DNP
8	5,8	261.1247 283.1066	[M+H] ⁺ [M+Na] ⁺	260.1174	C ₁₄ H ₁₆ N ₂ O ₃	4.98	47.35	6	Cyclo(2-hydroxypropylphenylalanyl)
9	6,1	211.1448 352.0995	[M+H] ⁺ ? [M-H] ⁻ ?	?					Cyclo(4-hydroxypropylphenylalanyl)
10	6,3	211.1445	[M+H] ⁺ ?	210.1372					Cyclo(propyltyrosyl)
11	6,5	211.1443	[M+H] ⁺ ?	210.1370	C ₁₁ H ₁₈ N ₂ O ₂	-0.95	66.93	0	
17	9,9	291.0208 337.0273 293.0352	[M-H] ⁻ [M+FA-H] ⁻ [M+H] ⁺	292.0279	C ₁₃ H ₁₀ N ₄ Cl ₂	7.85	66.36	0	
18	11,4	242.1757 244.1889	[M-H] ⁻ [M+H] ⁺	243.1829	ND ⁺				
19	11,7	329.2345	[M-H] ⁻ ?	328.2272	ND ⁺				
20	14,1	293.1792	[M-H] ⁻ ?	294.1864	ND ⁺				
21	16,1	265.1479	[M-H] ⁻ ?	266.1551	C ₁₃ H ₂₂ N ₄ S	-1.89	70.84	0	
22	23,2	406.3288	[M+H] ⁺ ?	405.3215	ND ⁺				
23	25,1	413.2650	[M+H] ⁺ ?	412.2577	ND ⁺				
24	25,3	338.3444	[M+H] ⁺ ?	337.3436	ND ⁺				

* Iso Score : score de similitude de profils isotopiques entre le profil observé et le profil théorique (score compris entre 0 et 100)

+ ND : Non déterminée car les formules brutes proposées présenteraient des scores de profils isotopiques (Iso Score) inférieurs à 50%.

4. Conclusion

Les investigations chimiques menées sur les souches fongiques MMS5 *P. chrysogenum* et MMS719 *D. americana*, sélectionnées pour leurs activités cytotoxiques sur une lignée cancéreuse mammaire au cours de ce projet, ont été réalisées selon différentes approches de fractionnement bioguidé et de profilage métabolique. Un travail de déréplication par comparaison des données chimiques (principalement de formule brute ou de masse monoisotopique) a été réalisé au cours des deux approches dans le but d'annoter les molécules produites par ces souches tout en évaluant leur intérêt potentiel sur la cible biologique étudiée.

Les études menées sur la souche MMS5 *P. chrysogenum* par fractionnement bioguidé ont permis d'isoler après 7 étapes de fractionnement, une seule molécule pure bioactive identifiée par RMN comme étant : l'ergostérol. L'isolement de ce stérol majoritairement produit par un grand nombre d'espèces fongiques, a alors permis de mettre en évidence les limitations associées à l'utilisation de cette stratégie dans le recherche de molécules naturelles bioactives. En effet, c'est un processus long pouvant conduire comme dans notre cas avec l'ergostérol, à l'isolement de molécule bioactive sans originalité chimique *in fine*. Cependant, les mécanismes d'actions de cette molécule sur l'inhibition de la prolifération des cellules cancéreuses mammaires n'étant pas décrits, des investigations ont été menées et font l'objet du chapitre 4 de ce manuscrit.

Le profilage métabolique et la déréplication réalisés sur cette souche ont permis de mettre en évidence, malgré le nombre important de publications, que l'ensemble de son métabolome n'est pas décrit dans la littérature (ou du moins référencé dans le DNP). La purification ciblée à la fois sur la recherche de molécules connues : la méléagrine et un dérivé de l'ergostérol l'anicequol, mais également sur celles ne présentant pas de correspondance dans le DNP a conduit au final à l'isolement de 5 molécules pures. Ces molécules identifiées par analyses spectrales RMN ont permis de confirmer les annotations réalisées pour la méléagrine, l'andrastine A et l'anicequol. Pour les deux autres nouvelles molécules potentielles, les élucidations structurales respectives sont en cours et elles n'ont malheureusement pas pu être identifiées pour la rédaction de ce manuscrit.

Enfin, les travaux de purification menés lors du fractionnement bioguidé ainsi que ceux réalisés lors de la purification ciblée ont permis d'obtenir un certain nombre de fractions. Ces fractions analysées par LC-UV-MS ont montré que la combinaison des deux approches pouvait être une alternative intéressante dans la recherche et l'identification de molécules. Basée sur l'acquisition de ces profils LC-UV-MS, des études de déréplication ont permis en comparaison avec le DNP de mettre en évidence pour les fractions les plus pures, la présence de 7 molécules. Parmi ces molécules, quatre ont pu être annotées comme étant des stérols avec des propositions tel que le 1-(10→6)-abeo-ergosta-5,7,9,22-tetraen-3 α -ol ou encore le citréospirostéroïde.

De la même façon, un fractionnement bioguidé a été réalisé sur deux extraits bruts issus de la souche MMS719. Malheureusement pour cette souche, par manque de temps, les fractionnements nécessaires à la purification de la ou des molécules responsables de l'activité n'ont pu être achevés. Néanmoins, une molécule a pu être isolée et identifiée par RMN comme étant le cyclo(4-hydroxypropylphenylalanyl), une dicétopipérazine. Cependant, cette molécule n'a démontré aucune activité sur la cible biologique étudiée.

Selon la deuxième approche de profilage métabolique réalisée sur l'extrait MMS719-527, la déréplication réalisée sur les ions majoritairement détectés au cours des analyses LC-UV-MS a révélé que beaucoup d'entre eux ne correspondaient à aucune des molécules référencées dans le DNP. Parmi les quelques annotations effectuées, celles-ci correspondaient principalement à des dicétopipérazines. Ces résultats ont renforcé l'idée selon laquelle cette souche peu décrite dans la littérature présente un fort potentiel dans la découverte de nouvelle molécule.

Par conséquent, les investigations chimiques menées sur ces deux souches *P. chrysogenum* et *D. americana* selon deux approches différentes, ont permis de révéler et d'exposer les avantages et limites de l'utilisation de chacune. Le fractionnement bioguidé très utilisé dans la recherche des produits naturels a montré au cours de cette étude qu'il était d'une grande efficacité dans l'isolement de molécule bioactive. En revanche, comme dans notre cas avec l'isolement de l'ergostérol, celui-ci peut se révéler long et fastidieux sans originalité chimique *in fine*. Ainsi, le manque de connaissance sur l'originalité structurale des molécules ainsi que sur leurs « abondances » au sein d'un extrait peut représenter un frein dans la découverte de nouvelles molécules.

Dans le cas de la seconde approche, la déréplication réalisée sur les profils métaboliques est apparue comme un outil essentiel à la réalisation de ce projet. En effet, elle a permis par comparaison avec des bases de données : (1) d'annoter les composés majoritairement produits par ces souches, (2) de mettre en évidence des molécules potentiellement nouvelles en l'absence de hits dans les bases de données ; (3) ou encore au cours d'un fractionnement bioguidé d'annoter les molécules et d'évaluer ainsi en amont leurs potentiels (chimiques et biologiques) avant de les purifier. Cependant, contrairement à d'autres techniques utilisées en métabolomique comme la RMN, l'usage de la LC-UV-HRMS ne permet pas de quantifier de manière absolue l'abondance des molécules. En effet, la performance de l'analyse est basée sur la capacité des molécules à s'ioniser. En revanche, elle présente l'avantage d'accéder à certaines caractéristiques chimiques (temps de rétention, l'UV, masse monoisotopique et intensité) qui permettent de comparer un même pic chromatographique (correspondant à une molécule) entre différents échantillons. Malheureusement dans certains cas, comme pour l'ergostérol, des phénomènes d'ionisation inexpliquée ou peu communs entraînent des erreurs d'annotations.

D'après l'ensemble de ces travaux, pour orienter la purification uniquement sur les molécules potentiellement nouvelles et bioactives, il faudrait être en mesure d'évaluer le potentiel de chacune d'elle sur la cible biologique étudiée. Une interprétation plus approfondie des données chimiques et biologiques est alors nécessaire. C'est pourquoi, il a été choisi de poursuivre cette étude en proposant une méthodologie permettant d'identifier précocement les métabolites responsables de l'activité observée au sein d'un mélange complexe

5. Matériel et méthodes

5.1. Culture des deux souches sélectionnées au cours du screening

A partir de prélèvements de fragments de mycélium issus des souches ayant été ensemencées quelques jours sur milieu DCA en boîte de Pétri, des inoculations ont été réalisées en trois points dans des Erlenmeyer sur les milieux gélosés CYA pour *P. chrysogenum* et DCA pour *D. americana*. Deux cultures ont été réalisées pour la souche MMS5 avec un nombre d'Erlenmeyer de 62 (lot 1) et 50 (lot 2), et trois cultures ont été réalisées pour la souche MMS719 avec un nombre d'Erlenmeyer de 100 (lot 1) et 30 (lots 2 et 3). Les cultures ont ensuite été incubées dans une étuve à 27 °C pour une période dont la durée correspond au temps nécessaire de développement de la souche sur l'ensemble de surface de la gélose afin de favoriser la production d'une biomasse suffisante. Ainsi des périodes d'incubation de 12 jours pour la souche *P. chrysogenum* et de 21 jours pour *D. americana* ont été réalisées.

5.2. Extraction de la biomasse fongique

La procédure d'extraction des cultures fongiques employée pour ces travaux correspond à une méthode couramment utilisée au laboratoire (Geiger *et al.*, 2013). Par ailleurs, l'ensemble de la culture fongique (biomasse fongique et gélose) est extrait afin d'obtenir aussi bien les métabolites excrétés dans la gélose que ceux contenus dans le mycélium et les spores. Le choix du système de solvants d'extraction $\text{CH}_2\text{Cl}_2/\text{AcOEt}$ (1:1, v/v) permet d'accéder à une large gamme de polarité au niveau des métabolites extraits, tout en restant non miscible à l'eau pour éviter l'extraction de l'agar. Ainsi, après la fin de période d'incubation, 100 mL d'un mélange $\text{CH}_2\text{Cl}_2/\text{AcOEt}$ (1 :1, v/v) ont été ajoutés dans chaque Erlenmeyer. Une fois l'ensemble broyé (gélose + mycélium + solvant) par une hélice rotative motorisée, les différents Erlenmeyers ont été placés 30 minutes dans un bain à ultrasons. Les rétentats obtenus après filtration sur Büchner ont été regroupés afin de réaliser une deuxième étape d'extraction de la biomasse selon les mêmes conditions que celles énoncées précédemment. Ainsi, les filtrats issus des deux extractions ont été regroupés, déshydratés par ajout de sulfate de sodium anhydre (Na_2SO_4) et filtrés (papier filtre de porosité 1.20 μm puis membrane de cellulose régénérée de porosité 0,45 μm , Sartorius Stedim Biotech, Goettingen, Allemagne) afin d'éliminer les morceaux de gélose, de mycélium et enfin les spores. Ils ont ensuite été évaporés à l'aide d'un évaporateur rotatif jusqu'à siccité de l'extrait brut. Pour la souche MMS5 deux extraits ont été obtenus : MMS5-526 (818,4 mg) et MMS5-531 (484,9 mg) et trois extraits ont été obtenus pour la souche MMS719 : MMS719-527 (1,164 mg), MMS719-529 (659,9 mg) et MMS719-530 (345,5 mg).

5.3. Évaluation et suivi de l'activité antiproliférative des cellules cancéreuses mammaires

5.3.1. Test MTT

Les conditions expérimentales de ce test sont les mêmes que celles décrites dans le chapitre 1 §5.2.2. Par ailleurs, l'effet de dose réponse des extraits bruts MMS5-526 et MMS719-527 reconstitués dans le DMSO et dilués dans le milieu DMEM contenant 0.1% de BSA ont été évalués dans la gamme de concentration suivante : 0,001-0,001-0,01-0,1-1-10 µg/mL. Pour les composés purs, le solvant de reconstitution utilisée était le DMSO, à l'exception de l'ergostérol reconstitué dans l'éthanol. Ces solvants ont également été utilisés en tant que contrôles négatifs. La durée d'incubation suivant les essais a été de 24H, 48H ou 72H sous atmosphère humide de 37 °C avec 5% de CO₂.

5.3.2. Test du crystal violet

Les cellules MCF-7 et MDA-MB-231 ont étéensemencées dans des plaques à 96 puits. Après une période d'incubation de 24H, le milieu de culture a été retiré et une solution de coloration au cristal violet (CV) a été ajoutée pour fixer et colorer les cellules pendant 1H. Après l'élimination de la solution de coloration, les plaques ont été lavées plusieurs fois avec de l'eau distillée pour éliminer l'excès de colorant. Les plaques ont ensuite été séchées à 37°C pendant au moins 3H. Après ajout d'une solution d'acide acétique à 10% la coloration a été quantifiée par spectrophotométrie.

5.4. Méthodes de séparation chromatographiques

5.4.1. Chromatographie sur couche mince (CCM)

Les chromatographies sur couche mince ont été réalisées sur des plaques d'aluminium recouverte de gel de silice en présence d'un indicateur de fluorescence UV₂₅₄ (porosité de 60 Å, épaisseur 0,2 mm, Macherey-Nagel, Düren, Allemagne). Les solvants de migration ont été choisis selon les profils chimiques des échantillons. La lecture des résultats s'est faite par observation dans le visible, sous lampe UV (aux longueurs d'onde $\lambda = 254$ nm et 365 nm) et après révélation chimique (vanilline sulfurique).

5.4.2. Chromatographie liquide basse pression

Au cours des étapes de séparation, des chromatographies préparatives ont été réalisées avec un système PuriFlashTM 430, Interchim[®] (Interchim, Montluçon, France) équipé d'un détecteur UV-DAD, ainsi que d'un collecteur automatique.

5.4.2.1. Fractionnements réalisés sur la souche MMS5 *P. chrysogenum*

Pour chaque étape de fractionnement ou purification réalisées sur la souche MMS 5, celle-ci est numérotée en faisant référence à l'arbre de fractionnement de l'extrait MMS5-526 (Cf. Figure 20) ou sur l'extrait MMS5-531 (Cf. Figure 32).

Pour l'extrait MMS5-526, les différents échantillons préalablement adsorbés sur d'un mélange de silice et de sable (50/50) ont été déposés pour MMS5-526 et MMS5-526-3-1 sur une colonne de phase normale de 25 g de silice (cartouche Chromabond® Flash RS SiOH 25, Macherey-Nagel). La séparation a été réalisée avec un débit de 20 mL/min et une collecte de 20 mL/tube suivie par les signaux UV obtenus à la longueur d'onde de 254 nm et sur un scan de 210-400 nm. Pour les échantillons MMS5-526-2-2 et MMS5-526-3-3, ces derniers ont été déposés respectivement sous forme solide et liquide (solubilisé dans le CH₂Cl₂) sur une colonne de (Puriflash, 4 g silica 15µ, Interchim) avec un débit de 5 mL/min.

Pour le second lot, la première étape de fractionnement réalisé sur l'extrait brut MMS5-531 (masse = 484,9 mg) a été réalisée selon les mêmes conditions que celles utilisées pour l'extrait MMS5-526.

Tableau 10. Conditions d'élution du fractionnement N°1 réalisé sur l'extrait MMS5-526.

Fractionnement réalisé sur l'échantillon MMS5-526 (masse = 818,4 mg)

min	% Hexane	% CH ₂ Cl ₂	% CH ₂ Cl ₂ /MeOH (90/10)	% CH ₂ Cl ₂ /MeOH (50/50)
0	50	50	0	0
15	0	100	0	0
20	0	100	0	0
50	0	40	60	0
70	0	0	0	100
85	0	0	0	100

Tableau 11. Conditions d'élution du fractionnement N°2 réalisée sur la fraction MMS5-526-1-3.

Fractionnement réalisé sur l'échantillon MMS5-526-1-3 (masse = 139,8 mg)

min	% Hexane	% Acétone	% MeOH
0	86	14	0
20	83	17	0
25	83	75	0
30	75	25	0
35	75	25	0
45	0	100	0
55	0	70	30
60	0	70	30

Tableau 12. Conditions d'élution du fractionnement N°3 réalisée sur la fraction MMS5-526-2-2.

Fractionnement réalisé sur l'échantillon MMS5-526-2-2 (masse = 99,0 mg)

min	% Hexane	% CH ₂ Cl ₂	% MeOH
0	100	0	0
3	50	50	0
6	50	50	0
9	50	49	1
12	49	49	2
18	49	49	3
25	48	48	4
28	45	45	4
50	45	45	10
60	45	45	10

Tableau 13. Conditions d'élution du fractionnement N°4 réalisée sur la fraction MMS5-526-3-3.

Fractionnement réalisé sur l'échantillon MMS5-526-3-3 (masse = 63,6 mg)

min	% Hexane	% CH ₂ Cl ₂	% MeOH
0	50	50	0
5	50	50	0
8	0	99	1
10	0	99	1
13	0	98	2
15	0	98	2
18	0	97	3
20	0	97	3
23	0	96	4
25	0	96	4
30	0	95	5
35	0	95	5
38	0	94	6
48	0	*	20

* 40/40, v/v, CH₂Cl₂/Acétone

5.4.2.2. Fractionnement réalisé sur la souche MMS719 *D. americana*

L'extrait brut MMS719-527 (masse = 1,164 g), préalablement adsorbé sur 4 g d'un mélange de silice et de sable (50/50) a été déposé sur une colonne de phase normale de 80 g de silice (cartouche Chromabond® Flash RS SiOH 80, Macherey-Nagel, Lot : 0715115). L'élution a été réalisée par un gradient CH₂Cl₂/MeOH de polarité croissante dont les conditions figurent dans le tableau ci-dessous (Cf. Tableau 14):

Tableau 14. Conditions d'élutions du fractionnement N°1 réalisé sur l'extrait MMS719-527.

Fractionnement réalisé sur l'échantillon MMS719-527 (masse = 1,164 g)

Volume colonne	% CH ₂ Cl ₂	% CH ₂ Cl ₂ /MeOH (90/10)	% MeOH
0	50	50	0
6	50	50	0
7	0	100	0
10	0	100	0
12	0	75	25
15	0	75	25

La séparation a été réalisée avec un débit de 40 mL/min et une collecte automatique de 20 mL/tube. L'élution a été suivie aux longueurs d'onde de 254 nm et avec un scan de 210-400 nm.

L'extrait brut MMS719-530 (masse = 298 mg), préalablement adsorbé sur 1,5 g d'un mélange de silice et de sable (50/50) a été déposé sur une colonne de phase normale de 25 g de silice (cartouche Chromabond® Flash RS SiOH 25, Macherey-Nagel, Lot : 0715115). L'élution a été réalisée par un gradient CH₂Cl₂/MeOH de polarité croissante dont les conditions figurent dans le tableau ci dessous :

Tableau 15. Conditions d'élutions du fractionnement N°1 réalisé sur l'extrait MMS719-531.

Fractionnement réalisé sur l'extrait brut MMS719-531 (masse = 298 mg)

Volume colonne	% CH ₂ Cl ₂	% CH ₂ Cl ₂ /MeOH (90/10)	% MeOH
0	50	50	0
6	50	50	0
7	0	100	0
10	0	100	0
12	0	75	25
15	0	75	25

La séparation a été réalisée avec un débit de 15 mL/min et un collecte automatique de 15 mL/tube. L'élution a été suivie aux longueurs d'onde de 254 nm et avec un scan de 210-400 nm.

5.4.3. Chromatographie liquide haute performance (CLHP)

Les séparations ont été réalisées sur une chaîne CLHP de la série Agilent 1200 (Agilent, Waldbronn, Allemagne) constituée d'un dégazeur (Agilent Technologies série 1200 G1322A, d'une pompe quaternaire (Agilent Technologie série 1200 G1311A), d'un injecteur automatique (Hewlett Packard série 1100), d'un détecteur UV-DAD (Agilent Technologie série 1200 G1315D) et d'un collecteur automatique Pep FC (Agilent Technologies série 1200 G1364B). Les conditions d'élutions ont été choisies selon les caractéristiques des échantillons et après optimisation sur CCM. L'élution a été suivi avec un scan de 200-600 nm.

Les caractéristiques des colonnes utilisées figurent ci-dessous :

- Pour les Colonnes analytiques phase C₁₈ :
 - AC18-8 : 254 mm x 4,6 mm x 0,5 μm (Interchim®), numéro de série 73281
 - AC18.23 : Luna® 5 μM C18 (2) 110 A, Size : 254 mm x 4,6 mm x 0,5 μm (Interchim®), numéro de série : H15-118776.
- Pour les Colonnes analytiques phase Silice :
 - ASI.12 : 254 mm x 4,6 mm x 0,5 μm (Interchim®), numéro de série : 309713.
- Pour les colonnes préparatives phase C₁₈:
 - PC18.6 : 250 mm x 10,0 mm x 5 μm (Interchim®)
 - PC18.8 : Kinetex 5u C18 100A, size : 250 x 10.0 mm x 5 μm (Phenomenex®), numéro de série : 742816-1
- Pour les colonnes préparatives phase Silice:
 - Colonne préparative PSI.7 : Luna® 5 μM Silica (2) 100 A, Size : 250 mm x 10,0 mm x 5 μm (Phenomenex®), numéro de série : H15-182495.

D'après l'arbre de fractionnement, pour l'extrait MMS5-526 :

- le fractionnement numéro 5 (MMS5-526-4-2) a été réalisé en phase inverse sur colonne C₁₈ (AC18.8) avec un gradient MeOH+0.1 AF%/eau+0.1AF% de 5% à 100% de MeOH en 30 min suivit d'un lavage de 15 min, avec un débit de 1 mL/min
- le fractionnement numéro 6 (MMS5-526-4-3) a été réalisé en phase normale sur colonne de silice (ASI.12) mode isocratique Hexane/AcOET (60 min), avec un débit de 1mL/min.
- fractionnement numéro 7 a été réalisé en phase inverse sur colonne C₁₈ (AC18.8) avec un gradient MeOH+0.1 AF%/Eau+0.1 AF% de 65% à 100% de MeOH en 60 min suivit d'un lavage de 15 min, avec un débit de 1mL/min.

D'après l'arbre de fractionnement, pour l'extrait MMS5-531 :

- le fractionnement réalisé sur la fraction **MMS5-531-1-8** a été effectué en phase normale sur une colonne de Silice (PSI.7) avec une élution en mode isocratique Hexane/AcOEt (87 :13, v/v), avec un débit de 4 mL/min.

- le fractionnement réalisé sur la fraction **MMS5-531-1-15** a été effectué en phase normale sur une colonne de Silice (PSI.7) avec une élution en mode isocratique CH₂Cl₂/MeOH (98:2, v/v), avec un débit de 2,5 mL/min.
 - le fractionnement de **MMS5-531-1-18** a été réalisée sur colonne de C₁₈ (PC18.6) avec un gradient d'élution ACN+0.1 AF%/Eau+0.1 AF% de 30 à 40% d'ACN en 25 min, puis de 40 à 100% d'ACN en 35 min, avec un débit de 1,5 mL/min.
 - le fractionnement de **MMS5-531-1-20** a été réalisée sur colonne de C₁₈ (PC18.6) avec un gradient d'élution ACN+0.1 AF%/Eau+0.1 AF% de 30 à 40% d'ACN en 20 min, puis de 40 à 100% d'ACN en 35 min, avec un débit de 1,5 mL/min
- la fraction **MMS5-531-1-19** a tout d'abord été fractionnée selon des essais de solubilité en deux sous fractions (A et B) par ajout d'acétate d'éthyle, le surnageant (A) une fois séché à été de nouveau fractionné en deux échantillons (**MMS5-531-1-19-1** et **MMS5-531-1-19-3**) par ajout de CH₂Cl₂. La sous-fraction (B) a également été séparée en deux échantillons (**MMS5-531-1-19-2** et **MMS5-531-1-19-4**) par ajout de MeOH. Puis, la fraction **MMS5-531-1-19-1** a été fractionnée sur colonne de Silice (PSI-7) avec une élution réalisée en mode isocratique CH₂Cl₂/MeOH (95 :5, v/v). Les fractions MMS5-531-2→4 regroupées ont été fractionnées sur colonne C₁₈ (PC18.6) avec une élution en mode isocratique ACN/Eau (65 :35, v/v), avec un débit de 1.5 ml/min, pendant 60 min.
- la fraction **MMS5-531-19-G** a été fractionnée sur colonne C₁₈ (PC18.6) avec une élution en mode isocratique MeOH/Eau (67 : 33, v/v), avec un débit de 1.5 ml/min.
- Les fractions **MMS5-531-19-D** et **MMS5-531-19-E** ont été fractionnées en phase inverse sur colonne C₁₈ (AC18.23) avec une élution en mode isocratique ACN+0.1 AF%/Eau+0.1AF% (48/52, v/v) avec un débit de 1mL/min.
 - la fraction **MMS5-531-19-F** et **MMS5-531-19-G7→G8** ont été fractionnées en phase inverse sur colonne C₁₈ (AC18.23) avec une élution en mode isocratique ACN+0.1 AF%/Eau+0.1AF% (25/75, v/v) avec un débit de 4mL/min.

Pour l'extrait MMS719-527 :

- fractionnement de la fraction MMS719-527-1-14 a été réalisé en phase inverse sur colonne C₁₈ (PC18.8) avec un gradient MeOH/eau de 10% à 100% de MeOH en 40 min suivi d'un lavage de 15 min, avec un débit de 2,5mL/min.
- fractionnement de la fraction MMS719-530-1-4 a été réalisé en phase inverse sur colonne C₁₈ (PC18.8) avec un gradient MeOH/eau de 10% à 100% de MeOH en 40 min suivi d'un lavage de 15 min, avec un débit de 2,5mL/min.

5.4.4. Profilage HPLC-UV-HRMS

Les analyses des extraits bruts et des fractions ont été réalisées en haute résolution sur un appareil de chromatographie liquide haute performance couplé à un spectromètre de masse de type IT-TOF, c'est à dire un analyseur à trappe ionique couplé à un analyseur à temps de vol, avec une source d'ionisation de type electrospray (ESI). L'appareil de chromatographie liquide est composé de deux pompes LC-20ADxr, d'un passeur d'échantillons SIL-20ACxr, d'un four à colonne CTO-20AC, d'un détecteur SPD-M20A PDA et d'un contrôleur MBC-système 20A. Les analyses de chromatographie liquide haute performance ont été effectuées sur une colonne Kinetex™ C18 (100 x 2,1 mm, 2,6 µm), chauffée à l'aide d'un four colonne à 40 °C, avec une phase mobile CH₃CN + 0,1% AF/H₂O + 0,1% AF selon le gradient d'élution suivant (Cf. Tableau 16) :

Tableau 16. Conditions d'élutions pour les analyses LC-UV-HRMS

Temps (min)	% H ₂ O + 0,1% AF	% ACN + 0,1% AF
0	85	15
2	85	15
25	0	100
30	0	100
31	85	15
35	85	15

Pour chaque analyse d'une durée de 35 min avec un débit de 0,3 mL/min, un volume d'injection de 5 µL des différents échantillons préparés dans du méthanol de grade UPLC aux concentrations de 0,5 mg/mL pour les extraits bruts et 0,2 mg/mL pour les fractions a été utilisé. Les échantillons ont été maintenus à 4 °C.

Les analyses MS ont été réalisées en mode positif et négatif pour une gamme de masse *m/z* comprise entre 100-1000 avec un temps accumulation des ions dans la trappe de 20 ms en mode positif et de 5 ms en mode négatif. La référence interne utilisée au cours des analyses est le NaTFA. Les paramètres suivants ont été utilisés : une température du bloc de chauffage et de la ligne de désolvatation à 200 °C; un débit en azote à 1,5 L/min ; une tension au niveau de l'interface à 4,5 kV et -3,5 kV pour les modes positif et négatif respectivement ; une tension au niveau du détecteur de type TOF de 1,6 kV. Les

fragmentations MS/MS ont été obtenues en utilisant les paramètres suivants : énergie 50%, gaz de collision 50%, q (fréquence) 0,251 (45,0 Hz). Les ions précurseurs ont été sélectionnés dans la gamme de valeurs de m/z comprises entre 150 et 1000, avec un déclenchement de la fragmentation pour des valeurs d'intensités supérieures à 10^6 et un temps d'exclusion dynamique de 3 secondes. L'absorption UV a également été détectée en mode PDA dans la gamme des longueurs d'ondes : 190-600 nm.

5.4.5. Déréplication

Enfin, pour les fractions d'intérêt analysées par LC-HRMS, les différents ions détectés, la recherche de présence d'adduits par étude des différences de masses entre les pics ($[M+H]^+$, $[M+Na]^+$, $[M+K]^+$, $[2M+H]^+$, $[2M+Na]^+$,...) a permis de déduire l'ion moléculaire $[M+H]^+$. A partir de la masse monoisotopique M déduite, et des profils isotopiques, une ou plusieurs formules brutes probables ont pu être calculées et proposées à l'aide du logiciel Shimadzu LCMS Solution version 3.71. Ces dernières ont été recherchées dans la base de données « Dictionary of Natural Products » pour annoter les molécules connues et effectuer l'étape de déréplication (<http://dnp.chemnetbase.com>). Pour, chaque pic étudié, ces recherches ont été répertoriées dans des tableaux dans lesquels figurent le rapport m/z des ions détectés, le temps de rétention associé, l'annotation des adduits et/ou fragments, les formules brutes proposées et l'écart calculé masse théorique/masse observée en ppm pour chacune, le nombre de hits dans le DNP et les noms des structures proposées.

6. Publication sur la détection de l'ergostérol en electrospray

L'article présenté à la page suivante est en cours de préparation avec une soumission prévue dans le journal « Rapid Communications in Mass Spectrometry ».

Detection of ergosterol by liquid chromatography/electrospray mass spectrometry – investigation of unusual ionization

Lucie Ory, Emmanuel Gentil and Catherine Roullier*

EA2160 – MMS, Nantes University, 9 rue Bias, 44035 Nantes cedex, France.

* Corresponding author.

Email: Catherine.roullier@univ-nantes.fr; Phone: + 33 (0)2 51 12 56 86

Introduction

In the past decades, mass spectrometry has become increasingly and routinely used for automated dereplication purposes in the field of natural product chemistry. To avoid re-isolation of known compounds or to investigate new natural resources, it now becomes necessary to identify, at an early stage, compounds that have already been described and studied in the literature. One of the most popular methods for annotation is based on HPLC/MS profiling of mixtures. In this field, high resolution mass spectrometers, which have become more and more accessible, allow accurate annotation of compounds in complex mixture such as crude extracts. In fact, based on the minimum standards defined by the Metabolomics Society and other members of the field for compound identification,¹⁻³ comparison of the data obtained with different databases usually allows to annotate known compounds with quite a good level of confidence.

Most studies are performed with HPLC coupled to ESI-MS, as electrospray ionization provides many advantages. It is well adapted to natural products as it allows the analysis of metabolites in a broad range of polarity, with few degradation and in-source fragmentation. In most cases, the protonated ($[M+H]^+$) or deprotonated ($[M-H]^-$) molecule is conserved allowing easy accurate mass deduction and molecular formula prediction for low MW molecules ($M < 1000$). Moreover, most common ESI-MS adducts have been described and can be easily calculated by developed automated tools for dereplication.⁴⁻⁶

However, in some cases, electrospray ionization can unusually be leading to non-reported adducts or fragments.^{7,8} This becomes detrimental to annotation, because these are not included in automated pipelines for compound identification. Moreover, mechanisms underlying these phenomena are not always well understood. In the present study, unusual ionization by electrospray was observed for ergosterol, a major sterol produced by fungi. While, indeed, ESI-MS does not constitute the best method to analyse and quantify ergosterol in samples because of low ionization efficiency (better ionized by APCI-MS or EI-MS), the present paper reports ionization of this compound by ESI-MS as repeated observations of major ions for this compound with none of them corresponding to $[M+H]^+$, $[M+Na]^+$ or $[M+K]^+$ attracted our attention. In the context of dereplication studies by HPLC/ESI-MS, this report would help to improve further annotations of this common compound in fungi. Moreover, attempts to understand and explain the reactions occurring in the source leading to the ions observed are presented and discussed.

Experimental

Chemicals and sample preparation

ULC-MS grade methanol, acetonitrile, water and formic acid were purchased from Biosolve (Dieuze, France). Ergosterol standard was purchased from Sigma (Code: 45480). Samples were prepared just prior to the analyses at a concentration of 0.5 mg/mL in methanol.

Mass spectrometry

The mass spectrometry experiments were carried out on two different mass spectrometer instruments, both fitted with electrospray ionization source. One corresponded to an ion trap coupled to a time-of-flight analyser equipped with a heated ESI source operating in both positive and negative ion modes: LCMS-IT-TOF (Shimadzu, Kyoto, Japan). The second one was an Exactive Orbitrap mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) equipped with a heated ESI source operating in positive ion mode. Both instruments were used for HPLC/HRMS analyses. Additionally, the IT-TOF was used for flow injection analyses (FIA).

On the IT-TOF spectrometer, MS data were recorded in the range of m/z 100-1000 with a mass accuracy of 5 ppm and a resolution of 10,000 at m/z 500, using the following parameters: heat block and curved desolvation line temperatures at 200 °C; nebulizing nitrogen gas flow at 1.5 L min⁻¹; interface voltage at (+) 4.5 kV and detector voltage of the TOF analyser at 1.6 kV.

On the Orbitrap spectrometer, MS data were recorded in the range of m/z 64-1000 with capillary temperature set at 300 °C, sheath gas flow rate at 45 a.u. (arbitrary units), auxiliary gas flow rate at 5 a.u. and heater temperature at 300 °C.

High performance liquid chromatography

High performance liquid chromatography analyses were performed on a Shimadzu Prominence UFLC system (Shimadzu, Kyoto, Japan) for the coupling with the IT-TOF-MS and an Agilent 1200 Infinity Series system (Agilent Technologies, Santa Clara, California, USA) for the coupling with the orbitrap-MS. The elution was performed on a Kinetex™ C₁₈ column (100 × 2.1 mm, 2.6 μm, Phenomenex, Torrance, CA, USA) heated in an oven equilibrated at 40 °C. A mobile phase consisting of CH₃CN/H₂O (acidified with 0.1% formic acid) was used. Gradient 1 started with 15% CH₃CN during 1 min, then increased linearly to 35% CH₃CN within 2 min, increased again to 100% CH₃CN within 14 min, hold at 100% CH₃CN for another 4 min, then returned to the initial conditions within 0.5 min, and hold for 5.5 min, for a total run time of 27 min. Gradient 2 started with 75% CH₃CN during 1 min, then increased linearly to 100% CH₃CN within 16 min, hold at 100% CH₃CN for another 4 min, then returned to the initial conditions within 0.5 min, and hold for 5.5 min, for a total run time of 27 min. The separation was performed on 5 μL of ergosterol solution with a flow rate of 0.3 mL/min. Both mass spectrometers were operated in full-scan mode. On the Shimadzu instrument, UV-VIS spectra were also detected and collected from 190 to 600 nm.

Results and Discussion

Observation of unusual ionization through HPLC/HRMS analyses

In the course of our ongoing research on natural products produced by marine fungi, repeated observations of the same peak eluting late in the HPLC/MS analyses performed on

fungus extracts, attracted our attention. These analyses were generally conducted using a large gradient (gradient 1) to access a broad range of metabolite polarities from the studied organisms. While the signal was not very intense, many ions were observed but none could be attributed to a single monoisotopic mass, by performing calculations from the commonly accepted list of classical ions observed in ESI-MS.⁹ It was found after comparison with standards that it in fact corresponded to ergosterol, which is the major sterol in fungi. The ergosterol monoisotopic mass is 396.3392 Da and corresponds to the molecular formula $C_{28}H_{44}O$. The predominantly observed peaks in the mass spectrum of the compound eluting at 19.7 min (Shimadzu HPLC/IT-TOF-MS) under gradient 1 conditions were m/z 481.2920 and m/z 379.3345 with the following minor peaks m/z values as 531.2943, 513.2820, 497.2856, 465.2982, 451.3158, 395.3311 and 355.2237 (Figure 1A). Although, the perfect overlap of all EICs showed they all came from the same compound, the classical adducts identification as $[M+H]^+$, $[M+Na]^+$ and $[M+CH_3OH]^+$ were not observed. Analyses on a different instrument provided similar results (Figure 1C).

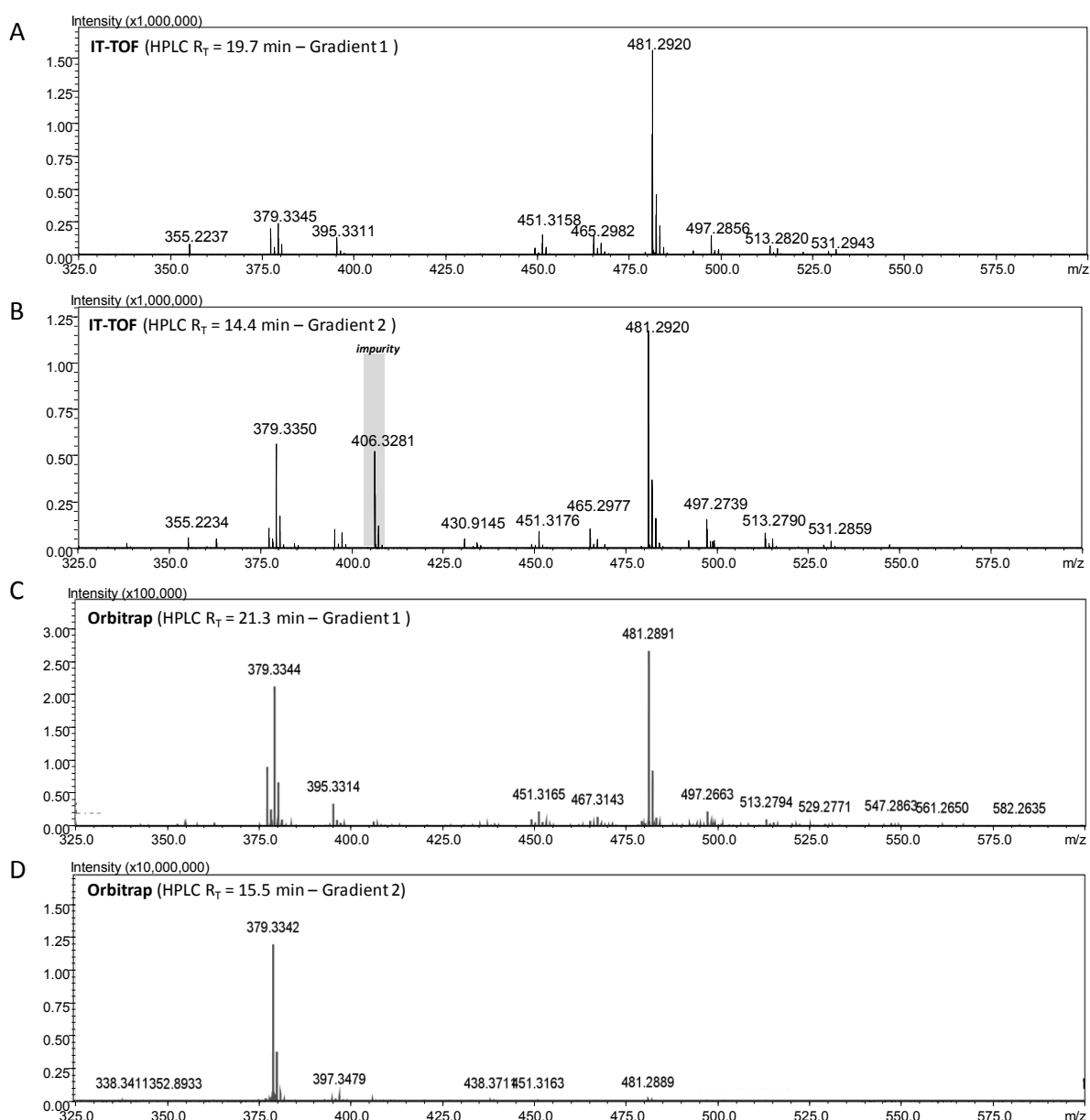


Figure 1. Observed mass spectra of ergosterol after HPLC through two different instruments, a Shimadzu UFLC/IT-TOF-MS (A, B) and a Thermo HPLC/Orbitrap-MS (C,D) and two different gradients: gradient 1 (A,C) and gradient 2 (B,D). Impurity

After thorough examination of the literature, only one paper was found showing the mass spectrum of ergosterol obtained by ESI-MS, with the observation of the same ions. However, no tentative explanation of the observed ions other than $[M+H-H_2O]^+$ was given.¹⁰ Besides, this fragment at m/z 379.4 was already reported with APCI-MS.^{11–15} In the present work, the effects of different parameters such as the elution solvents or the formic acid concentration were then investigated to better understand the ions observed.

Effect of elution solvent

With a second gradient (gradient 2) carried out on both instruments the mass spectrum of ergosterol (eluting at 14.4 and 15.5 min on the HPLC/IT-TOF-MS and the HPLC/Orbitrap-MS respectively) was shown to differ from gradient 1 (Figure 1B and 1D). In both cases, the intensity of the fragment $[M + H - H_2O]^+$ at m/z 379.4 was increased with gradient 2 while the ion at m/z 481.2920 was decreased with a more drastic effect on the Orbitrap-MS instrument.

Whatever the gradient used on the column, HPLC-UV profiles showed the same area under the curve (AUC) for ergosterol with the corresponding characteristic UV absorption spectrum, showing that no substantial degradation of the compound occurred on the column.

As modification of the HPLC gradient was shown to impact ergosterol ionisation, effect of the elution solvent composition was suspected. To carefully study this effect, flow injection analyses (FIA) were then further carried out on the IT-TOF instrument with different proportions of H_2O /Acetonitrile (+ 0.1% formic acid), ranging from. Acetonitrile percentages were ranging from 50% to 100% (Figure 2). These experiments demonstrated the influence of acetonitrile percentage on the ionisation pattern of ergosterol by electrospray. The more important the proportion of acetonitrile, the more important the m/z 481, which was the most intense ion observed with gradient 1.

Effect of formic acid

As acidic conditions have been described to impact the stability of ergosterol leading to its degradation,¹⁶ experiments were conducted in order to compare the influence of formic acid on the ions observed (Supplementary Information S1). In fact, no clear influence of the percentage of formic acid was noted in the range 0.1-0.4%.

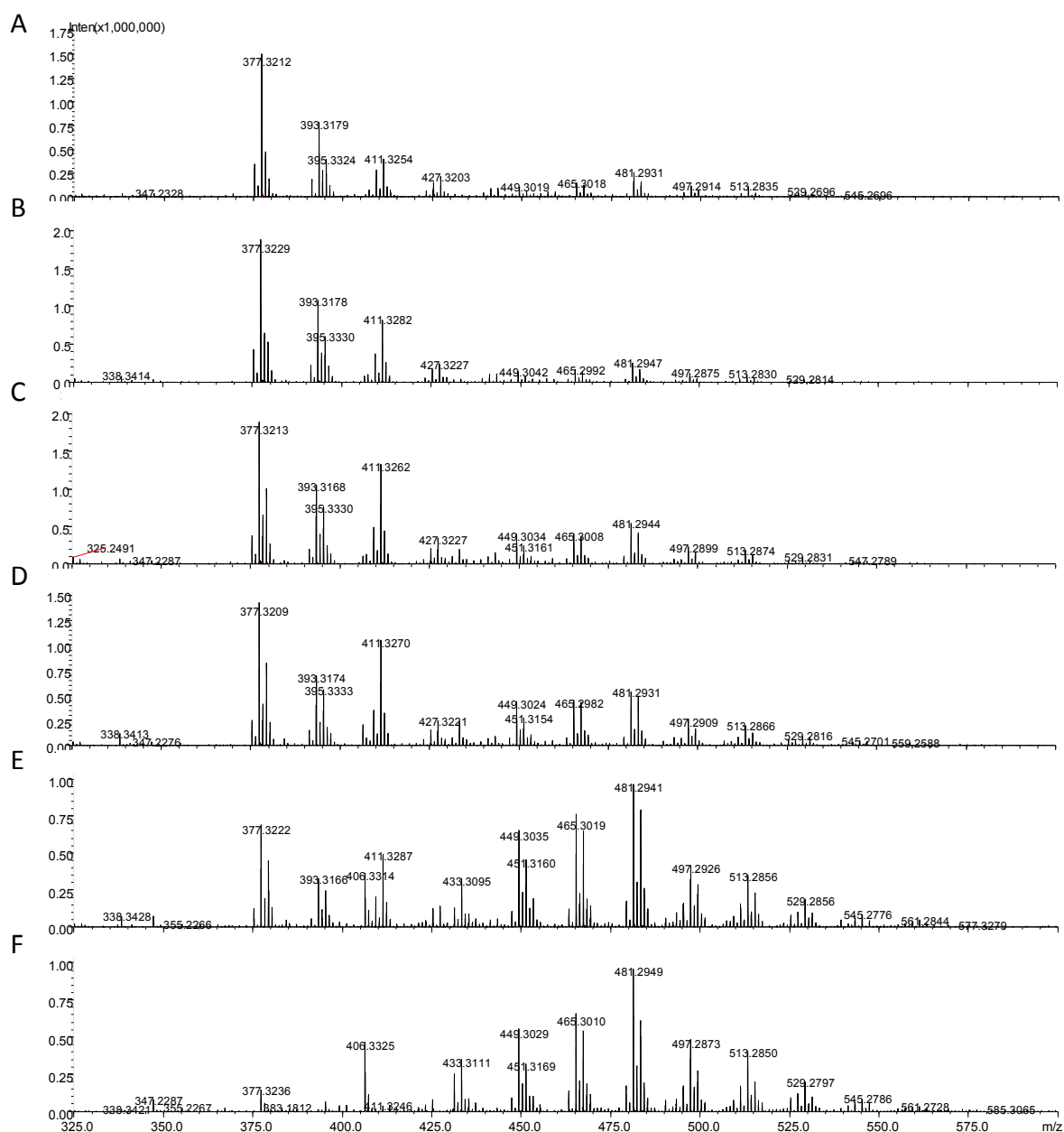


Figure 2. Influence of acetonitrile (+ 0.1% formic acid) on the ionisation pattern of ergosterol by electrospray, observed after flow injection analyses (FIA) with acetonitrile percentages of 50% (A), 60% (B), 70% (C), 80% (D), 90% (E) and 100% (F).

Hypothesis on the structures of the ions observed

Searching the literature did not provide any explanation on the ions observed, especially as ESI-MS is not the method of choice to detect and analyse ergosterol. Other MS methods describe the use of APCI-MS¹¹⁻¹⁵ and APPI-MS¹⁷ sources coupled to HPLC or GC/MS.^{18,19} For conventional LC-APCI-MS the predominant adducts obtained have been reported to be $[M + H]^+$ at m/z 397.4 and $[M + H - H_2O]^+$ at m/z 379.4.¹¹⁻¹⁵ Under low energy collision induced dissociation (CID) conditions, the major product-ion formed from the precursor ion m/z 379.4 $[M + H - H_2O]^+$ was described to be m/z 69.4 $[(CH_3)_2CHCH=CH]^+$.¹¹

Consequently, molecular formula prediction was performed on all major peaks observed in our case and their MS² collision induced fragmentation was investigated, as reported in Table 1. Regarding the molecular formula predictions of the different ions, it was hypothesized that dehydrogenations of the molecule and oxygen addition might be the key. In fact, annotation of the ion at m/z 393.3160 as [M-2H₂+H]⁺ was consistent with the same ion reported in the literature after APPI-MS ionisation of ergosterol, together with its major fragment at 268.1852.¹⁷ As it is well described that ergosterol is affected by light and air, easily suffering photooxidation,²⁰ it was proposed that most of the different ions observed corresponded to oxidations of the molecule.

Table 1: Summary of ions observed after LC/ESI-MS of ergosterol and their corresponding attribution based on molecular formula prediction and MS² fragmentation.

m/z observed	Molecular formula	Error (in ppm)	Attribution	Fragments m/z
377.3212	C ₂₈ H ₄₀	2.65	[M-H ₂ O-H ₂ +H] ⁺	251.1799; 293.2266 ; 307.2459; 332.2912; 376.3064
379.3350	C ₂₈ H ₄₂	2.25	[M-H ₂ O+H] ⁺	253.1964; 295.2408 ; 309.2581; 378.33232
393.3160	C ₂₈ H ₄₀ O	2.25	[M-2H ₂ +H] ⁺	251.1787; 268.1852 ; 291.2106; 309.2215; 335.2763; 375.3058; 392.3014
395.3324	C ₂₈ H ₄₂ O	4.1	[M-H ₂ +H] ⁺	251.1796; 269.1890; 293.2287; 311.2360; 337.2919; 377.3287 ; 394.3209
409.3123	C ₂₈ H ₄₀ O ₂	5.37	[M-2H ₂ +O+H] ⁺	
411.3254	C ₂₈ H ₄₂ O ₂	-0.68	[M-H ₂ +O+H] ⁺	251.1817; 269.1901 ; 285.1834; 309.2234; 327.2319; 341.2485; 393.3174 ; 410.3153
425.3060	C ₂₈ H ₄₀ O ₃	2.35	[M-2H ₂ +O ₂ +H] ⁺	
427.3227	C ₂₈ H ₄₂ O ₃	4.94	[M-H ₂ +O ₂ +H] ⁺	300.1585; 381.3229; 391.3122; 409.3140 ; 426.3127
433.3095	C ₂₈ H ₄₂ O ₂	4.15	[M-H ₂ +O+Na] ⁺	
449.3029	C ₂₈ H ₄₂ O ₃	0.46	[M-H ₂ +O ₂ +Na] ⁺	230.9073; 431.2935; 448.2945
451.3169	C ₂₈ H ₄₄ O ₃	-4.31	[M+O ₂ +Na] ⁺	
465.3010	C ₂₈ H ₄₂ O ₄	6.76	[M-H ₂ +O ₃ +Na] ⁺	282.3065; 339.2508; 379.2637; 433.3075 ; 447.2862
467.3154	C ₂₈ H ₄₄ O ₄	3.92	[M+O ₃ +Na] ⁺	
481.2949	C ₂₈ H ₄₂ O ₅	4.31	[M-H ₂ +O ₄ +Na] ⁺	355.2245 ; 421.2307; 449.2974; 463.2855
483.3118	C ₂₈ H ₄₄ O ₅	7.01	[M+O ₄ +Na] ⁺	423.2526; 451.3190 ; 465.2962; 482.2952
497.2873	C ₂₈ H ₄₂ O ₆	-1.14	[M-H ₂ +O ₅ +Na] ⁺	353.2121; 355.2282; 371.2196 ; 413.1864; 437.2310; 451.3108; 465.2968; 479.2821; 496.2715
499.3011	C ₂₈ H ₄₄ O ₆	-3.81	[M+O ₅ +Na] ⁺	
513.2850	C ₂₈ H ₄₂ O ₇	4.58	[M-H ₂ +O ₆ +Na] ⁺	
515.2994	C ₂₈ H ₄₄ O ₇	2.91	[M+O ₆ +Na] ⁺	
529.2797	C ₂₈ H ₄₂ O ₈	4.01	[M-H ₂ +O ₇ +Na] ⁺	
505.3175	C ₂₈ H ₄₄ O ₅	0.79	[M+O ₄ +HCOO] ⁺	

Previous studies have shown that ergosterol can suffer from regioselective and stereoselective attack of singlet oxygen. The authors stated that the efficiency of the different processes involved depended on the polarity and the proton affinity of the medium, which affect the stability of the adducts formed when singlet oxygen attacks the molecule, and that it also depends on the lifetime of the singlet oxygen.²¹ In our case, it could be deduced that singlet oxygen is generated in the electrospray ionisation source, which could be affected by the solvent composition. This would explain the different ions observed for ergosterol.

According to the literature, electrospray ionisation source can be considered as an electrochemical cell, and oxidation reactions can occur.²² Reactions such as

dehydrogenations have then been described in that context.²³ Losses of H₂ and oxygen addition in the case of ergosterol could then be attributed to electrochemical reactions due to electrospray.

Following this, the structural hypothesis raised for the most abundant detected ion (*m/z* 481.2922), was that ergosterol suffered of an oxidative degradation leading to a double peroxidation of the compound during the analysis, leading to a diepoxyergosterol.

Conclusions

In summary, ergosterol was found to ionize by electrospray in an unpredictable way. This was hypothesized to be due to electrochemical reaction on the source involving singlet oxygen formation and oxidation of the compound. These reactions were influenced by the composition of the solvent leading to increased oxygenation when acetonitrile percentage was increased. While automated annotation of compounds becomes more and more accessible, these observations show that they might be impaired by such unusual ionisations. However, reporting of this phenomenon for ergosterol should help further annotation of that compound through LC/ESI-MS experiments in fungal extracts.

Acknowledgements

The authors would like to thank the French Ministry and La Ligue contre le Cancer for the PhD grant for LO. They also thank Yann Guitton and Anne-Lise Royer (LABERCA, Nantes, France) for the experiments performed on the Orbitrap-MS. Finally, they also thank the ThalassOMICS platform for all the experiments carried out on the IT-TOF-MS.

References

1. Sumner LW, Amberg A, Barrett D, et al. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics Off J Metabolomic Soc.* 2007;3(3):211-221. doi:10.1007/s11306-007-0082-2
2. Creek DJ, Dunn WB, Fiehn O, et al. Metabolite identification: are you sure? And how do your peers gauge your confidence? *Metabolomics.* 2014;10(3):350-353. doi:10.1007/s11306-014-0656-8
3. Schymanski EL, Jeon J, Gulde R, et al. Identifying small molecules via high resolution mass spectrometry: Communicating confidence. *Environ Sci Technol.* 2014;48(4):2097-2098. doi:10.1021/es5002105
4. Nelson H, Siegel MM, Kruppa GH, Laukien FH. Automation of a fourier transform ion cyclotron resonance mass spectrometer for acquisition, analysis, and e-mailing of high-

- resolution exact-mass electrospray ionization mass spectral data. *J Am Soc Mass Spectrom.* 1999;10(11):1166-1173. doi:10.1016/S1044-0305(99)00089-6
5. Kind T, Fiehn O. Advances in structure elucidation of small molecules using mass spectrometry. *Bioanal Rev.* 2010;2(1-4):23-60. doi:10.1007/s12566-010-0015-9
 6. Bertrand, S., Guitton, Y., Roullier, C. Successes and pitfalls in automated dereplication strategy using liquid chromatography coupled to mass spectrometry data: A CASMI 2016 experience. *Phytochem Lett.* 2017;21:297-305. doi:10.1016/j.phytol.2016.12.025
 7. Fang L, Dong C, Guo C, et al. Two competing ionization processes in ESI-MS analysis of N-(1,3-diphenylallyl)benzenamines: formation of the unusual $[M-H]^+$ ion versus the regular $[M+H]^+$ ion. *Eur J Mass Spectrom.* 2018;24(3):251-260. doi:10.1177/1469066717717228
 8. Sheng H, Lexa KW, Zhang L-K, et al. Unusual (+/-)-electrospray ionization induced fragmentation: Structural elucidation of an in-process synthetic intermediate of doravirine (MK-1439) using liquid chromatography/high-resolution tandem mass spectrometry and two-dimensional nuclear magnetic resonance. *Rapid Commun Mass Spectrom.* 2017;31(8):719-727. doi:10.1002/rcm.7833
 9. Huang N, Siegel MM, Kruppa GH, Laukien FH. Automation of a fourier transform ion cyclotron resonance mass spectrometer for acquisition, analysis, and e-mailing of high-resolution exact-mass electrospray ionization mass spectral data. *J Am Soc Mass Spectrom.* 1999;10(11):1166-1173. doi:10.1016/S1044-0305(99)00089-6
 10. Magalhães PJ, Carvalho DO, Guido LF, Barros AA. Detection and quantification of provitamin D₂ and vitamin D₂ in hop (*Humulus lupulus L.*) by liquid chromatography–diode array detection–electrospray ionization tandem mass spectrometry. *J Agric Food Chem.* 2007;55(20):7995-8002. doi:10.1021/jf071308d
 11. Headley JV, Peru KM, Verma B, Roberts RD. Mass spectrometric determination of ergosterol in a prairie natural wetland. *J Chromatogr A.* 2002;958(1-2):149-156. doi:10.1016/S0021-9673(02)00326-6
 12. Henderson CM, Lozada-Contreras M, Naravane Y, Longo ML, Block DE. Analysis of major phospholipid species and ergosterol in fermenting industrial yeast strains using atmospheric pressure ionization ion-trap mass spectrometry. *J Agric Food Chem.* 2011;59(24):12761-12770. doi:10.1021/jf203203h
 13. McEwen C, Gutteridge S. Analysis of the inhibition of the ergosterol pathway in fungi using the atmospheric solids analysis probe (ASAP) method. *J Am Soc Mass Spectrom.* 2007;18(7):1274-1278. doi:10.1016/j.jasms.2007.03.032
 14. Toh TH, Prior BA, van der Merwe MJ. Quantification of plasma membrane ergosterol of *Saccharomyces cerevisiae* by direct-injection atmospheric pressure chemical ionization/tandem mass spectrometry. *Anal Biochem.* 2001;288(1):44-51. doi:10.1006/abio.2000.4877

15. Hailat IA. Mass spectrometric methods for analysis of sterols and sterol esters in biological samples. 2014, Doctoral dissertation. Memorial University of Newfoundland.
16. Gessner MO, Schmitt AL. Use of solid-phase extraction to determine ergosterol concentrations in plant tissue colonized by fungi. *Appl Env Microbiol.* 1996;62(2):415-419.
17. Varga M, Bartók T, Mesterházy Á. Determination of ergosterol in *Fusarium*-infected wheat by liquid chromatography–atmospheric pressure photoionization mass spectrometry. *J Chromatogr A.* 2006;1103(2):278-283. doi:10.1016/j.chroma.2005.11.051
18. Řezanka T. Analysis of sterol esters from alga and yeast by high-performance liquid chromatography and capillary gas chromatography-mass spectrometry with chemical ionization. *J Chromatogr A.* 1992;598(2):219-226. doi:10.1016/0021-9673(92)85051-T
19. Nielsen KF, Madsen J. Determination of ergosterol on mouldy building materials using isotope dilution and gas chromatography–tandem mass spectrometry. *J Chromatogr A.* 2000;898(2):227-234. doi:10.1016/S0021-9673(00)00873-6
20. Ergosterol. In: *The Merck Index. 9th Ed. Rahway.* New Jersey: Merck & Co., Inc.; 1976:478.
21. Ponce MA, Ramirez JA, Galagovsky LR, Gros EG, Erra-Balsells* R. A new look into the reaction between ergosterol and singlet oxygen in vitro. In memory of Dr Eduardo G. Gros. *Photochem Photobiol Sci.* 2002;1(10):749-756. doi:10.1039/b204452h
22. Van Berkel GJ, Kertesz V. Using the electrochemistry of the electrospray ion source. *Anal Chem.* 2007;79(15):5510-5520. doi:10.1021/ac071944a
23. Xu X, He G, Xu X, Wu Z, Cai T. Investigation of the electrochemical oxidation of 2,3'-bisindolylmethanes in positive-ion electrospray ionization mass spectrometry. *RSC Adv.* 2019;9(19):10727-10732. doi:10.1039/C9RA00348G

Chapitre 3 :

Développement d'un outil
bioinformatique dans l'identification
précoce de métabolites bioactifs au
sein de mélanges complexes :

RFiBiBo script

Jusqu'ici, les investigations chimiques menées sur la souche *P. chrysogenum* ont permis d'isoler et d'identifier une seule molécule bioactive : l'ergostérol. L'utilisation d'un profilage métabolique mené en parallèle d'un fractionnement bioguidé est apparue comme une stratégie pertinente. En effet, l'acquisition simultanée de données chimiques et biologiques a permis d'évaluer la complexité chimique des échantillons et d'orienter la purification vers les molécules bioactives. De plus, la déréplication effectuée sur l'ensemble des ions détectés au sein des profils LC-UV-HRMS a conduit à l'annotation : (1) de molécules connues, (2) de molécules pouvant être rapportées pour leurs activités biologiques sur la cible étudiée ou (3) de molécules potentiellement nouvelles en l'absence de références dans les bases de données. En revanche, il a été impossible de mettre en évidence les ions correspondant aux molécules potentiellement responsables de l'activité observée, d'où la problématique suivante: « **Comment établir un lien (s'il existe) entre les données chimiques et biologiques ?** ». Face à ce constat, l'utilisation d'une approche de biochimométrie correspondant à l'analyse de ces deux jeux de données par des analyses statistiques ou des modèles mathématiques s'est révélée prometteuse. Ainsi, une brève introduction bibliographique a été réalisée dans le but de définir et de réaliser un état de l'art sur l'utilisation de cette approche émergente au sein de la communauté scientifique des produits naturels.

1. Introduction bibliographique

La recrudescence des approches « omiques » et notamment le développement d'outils métabolomiques ouvre aujourd'hui de nouvelles perspectives dans la recherche de molécules bioactives. En effet, l'optimisation des méthodes chromatographiques d'analyses spectrales et / ou structurales ainsi que l'apparition de la haute résolution génère une quantité de données chimiques conséquente d'une haute précision. Afin d'établir une relation entre des données chimiques et biologiques, une approche de biochimométrie peut être utilisée.

1.1. Qu'est-ce que la « biochimométrie » ?

Le terme « biochimométrie » vient de l'anglais « biochemometrics », une discipline récente issue à l'origine de celle de la chimométrie apparue dans les années 70 (Geladi and Esbensen, 1990). Cette dernière consiste en l'application de modèles mathématiques notamment l'utilisation d'outils statistiques dans le but d'analyser et d'interpréter le maximum de résultats à partir de données chimiques. Ces données peuvent être de types spectrales, spectroscopiques ou physicochimiques.

Depuis les années 2000, avec le fort développement des outils analytiques et informatiques dans le domaine de la biologie (ex : séquençage AND) ou de la chimie (ex : profilage Métaboliques), de plus en plus de données sont générées par les techniques utilisées. Dans cette ère du « Big data », le traitement et l'interprétation des données deviennent de plus en plus compliqués. Ainsi, l'utilisation de modèles statistiques est devenue nécessaire.

Par ailleurs, la multidisciplinarité des projets notamment dans le domaine des produits naturels a conduit à l'utilisation de cette approche dans le but de mettre en relation non pas une typologie de données mais celle de plusieurs disciplines telles que la biologie et la chimie, d'où la notion de biochimométrie.

1.2. L'utilisation de cette approche dans le domaine des produits naturels

L'un des enjeux majeurs dans le domaine des produits naturels est celui d'identifier de nouvelles molécules bioactives au sein de mélanges complexes. Pour compléter les approches existantes de fractionnement bioguidée et de profilage métabolique, ces nouvelles approches de biochimométrie sont utilisées et développées pour identifier précocement les métabolites bioactifs afin de prioriser leur purification au sein d'un mélange complexe.

En réalisant une recherche rapide dans deux moteurs de recherche, par les mots clés suivants : « Biochemometrics » et « Natural Product », il a pu être observé seulement 7 publications référencées dans Pubmed et 169 dans Google scholar. En revanche plus de la moitié ont été référencées après 2015, démontrant ainsi le récent potentiel de ces approches dans la recherche de molécules bioactives. Ainsi, l'idée générale de ces articles consiste, après définition de l'organisme vivant et de la cible biologique à étudier, à réaliser une empreinte chimique de l'organisme (ex : RMN, LC-MS et Infra-Rouge) tout en évaluant son impact sur une cible biologique (séquençage ADN, cytotoxicité etc.). Ces données sont ensuite traitées à l'aide d'outils statistiques de façon à faire ressortir les informations pertinentes, c'est-à-dire l'impact de la chimie sur la biologie. A l'heure actuelle ; les modèles classiquement utilisés sont ceux de l'analyse en composante principale (ACP) et de la régression linéaire « partial least square » (PLS). Leur intérêt a été démontré dans l'identification précoce de métabolites antimicrobiens (Britton *et al.*, 2018; Kellogg *et al.*, 2016).

1.3. L'acquisition des données

Au regard des investigations menées jusqu'ici et de la littérature, pour maximiser l'accès à la composition chimique d'un échantillon, l'étude de fractions plutôt que celle d'un extrait brut directement est préférable. Pour réaliser cet enrichissement, des extractions successives, des fractionnements grossiers à l'aide de chromatographie semi-préparative, ou encore des microfractionnements réalisés sur colonne de phase solide (SPE) peuvent être utilisés. En effet, lors du profilage d'un extrait brut par LC-MS par exemple, des produits minoritaires peuvent être masqués par la présence des produits majoritaires et même dans certains cas par le bruit du fond. Le fractionnement permettant d'augmenter la concentration des métabolites minoritaires, et donc leurs détections au cours des analyses LC-UV-HRMS. Ainsi il a été rapporté dans la littérature que le fractionnement permettait d'effectuer une meilleure déréplication. De plus, lors d'un fractionnement bioguidé, si ces métabolites sont bioactifs, ce fractionnement permet d'accroître la probabilité d'observer leur activité antiproliférative. Par ailleurs, des effets synergiques ou antagonistes peuvent également être rapidement mis en évidence.

Les analyses statistiques mettant en relation deux types de données, celles-ci nécessitent pour obtenir des résultats d'une significativité suffisante d'avoir plusieurs points de comparaison. Par conséquent, un microfractionnement de l'extrait brut réalisé sur plusieurs colonnes de phases différentes semble tout à fait adapté. En effet, il permet d'utiliser peu de quantité d'extrait brut au départ et d'obtenir des fractions de composition chimique variée, parmi lesquelles la même molécule initialement présente dans l'extrait peut être détectée dans plusieurs fractions (minimum une fraction par colonne).

2. Objectifs de l'étude

Dans le cadre de ce projet, il a été choisi de proposer et de développer une méthode permettant d'identifier précocement les métabolites potentiellement bioactifs au sein d'un mélange complexe. Afin d'évaluer l'activité potentielle de chacune des molécules produites et présentes au sein d'un extrait brut, il est alors nécessaire d'établir un lien (s'il existe) entre les données chimiques et biologiques. Pour répondre à cette problématique, il a donc été envisagé de combiner les approches de bioguidage et de profilage métabolique utilisés jusqu'ici, mais en complétant le traitement des données obtenues selon une approche de biochimométrie.

Une méthode a alors été développée en se basant sur un cas d'étude réel, soit sur notre extrait de la souche MMS5 *P. chrysogenum*. Puisque cet extrait a pu être mis en évidence pour son activité antiproliférative sur les cellules cancéreuses du sein due à la présence en grande quantité de l'ergostérol, nous nous sommes interrogés sur la capacité d'outils de biochimométrie à repérer ce métabolite d'intérêt. Ainsi, l'approche globale proposée pour l'évaluation du potentiel chimique d'un extrait sur une cible biologique choisie, s'appuie sur un fractionnement réalisé sur une série de quatre micro colonnes de phase solide (SPE) et la combinaison de quatre outils statistiques. En effet, pour modéliser de manière différente la relation entre : (1) la présence (aire sous pic) d'un ion au sein des profils chromatographiques LC-MS d'un échantillon et (2) l'activité de cet échantillon il a été choisi de combiner les résultats de quatre modèles statistiques (Spearman, F-PCA, PLS, PLS-DA). Le premier permet de déterminer une corrélation de type non linéaire entre deux variables X et Y, contrairement à l'ACP-F basée sur l'utilisation du coefficient de Pearson (corrélation linéaire). Les deux autres sont des modèles de régression linéaires avec une particularité pour la PLS-DA qui est une analyse discriminante permettant de comparer des groupes. Pour réaliser cette combinaison de traitements de données qui n'existe pas au sein de la communauté, un programme informatique nommé « R FIBICO » a été créé sous le logiciel de programmation R. Ainsi, en organisant les données chimiques et biologiques au sein d'une même matrice, celle-ci est traitée par le script informatique dans le but de générer *in fine* une liste d'ions et de représentations graphiques pour chacun des modèles mais également pour la combinaison des quatre. Au sein de cette « Super-liste », les ions sont classés par niveau d'importance selon un score établi d'après les quatre modèles. Différents critères ont été définis lors de sa rédaction pour qu'il puisse s'adapter à divers projets et qu'il soit le plus facile d'utilisation possible.

Ainsi, cette méthode basée sur l'association d'un microfractionnement exploratoire réalisée sur des micro-colonnes de phase solide (SPE) et de biochimométrie combinant les résultats de quatre modèles statistiques (Spearman, F-PCA, PLS et PLS-DA) a été développée et a fait l'objet d'une publication parue dans le journal *Analytica Chimica Acta* d'un impact factor de 5.25 dont le « supplementary material » et le code (programmation sous R) figurent en annexe 22 et 23.

3. Présentation de la publication

L'article paru dans le journal *Analytica Chimica Acta* est présenté sous sa forme publiée.



Contents lists available at ScienceDirect

Analytica Chimica Acta

journal homepage: www.elsevier.com/locate/aca

Targeting bioactive compounds in natural extracts - Development of a comprehensive workflow combining chemical and biological data

Lucie Ory ^a, El-Hassane Nazih ^a, Sahar Daoud ^a, Julia Mocquard ^a, Mélanie Bourjot ^b, Laure Margueritte ^c, Marc-André Delsuc ^d, Jean-Marie Bard ^a, Yves François Pouchus ^a, Samuel Bertrand ^{a,e}, Catherine Roullier ^{a,e,*}

^a Université de Nantes, Mer Molécules Santé, MMS EA 2160, F-44000, Nantes, France

^b UMR 7178 CNRS, Institut Pluridisciplinaire Hubert Curien, Faculté de Pharmacie, Université de Strasbourg, F-67401, Illkirch, France

^c UMR 7200 CNRS, Laboratoire d'Innovation Thérapeutique, Faculté de Pharmacie, Université de Strasbourg, F-67401, Illkirch, France

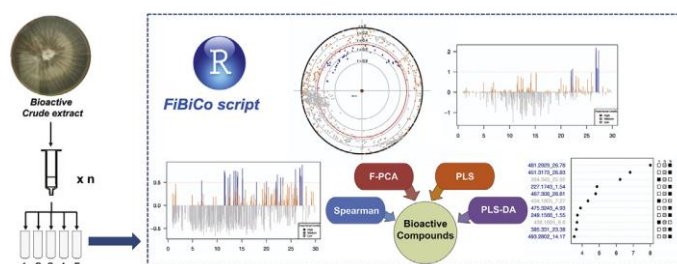
^d INSERM U596, UMR 7104 CNRS, Institut de Génétique et de Biologie Moléculaire et Cellulaire, Université de Strasbourg, F-67401, Illkirch, France

^e Biogenouest, Université de Nantes, Corsaire-ThalassOMICS, F-44000, Nantes, France

HIGHLIGHTS

- Fractionation coupled to biochemometrics is an interesting strategy for drug discovery.
- The workflow proposed provides a new tool for the detection of bioactive compounds.
- R-FiBiCo script proved to be efficient in detecting bioactive compounds from a mixture.
- Performance evaluation of the script showed its ability to detect minor compounds.
- Mass spectrometry- and NMR-based biochemometrics approaches can be complementary.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 20 December 2018

Received in revised form

19 March 2019

Accepted 18 April 2019

Available online xxx

Keywords:

Metabolomics

Biochemometrics

Natural products

Liquid chromatography

Mass spectrometry

R script

ABSTRACT

In natural product drug discovery, several strategies have emerged to highlight specifically bioactive compound(s) within complex mixtures (fractions or crude extracts) using metabolomics tools. In this area, a great deal of interest has raised among the scientific community on strategies to link chemical profiles and associated biological data, leading to the new field called "biochemometrics". This article falls into this emerging research by proposing a complete workflow, which was divided into three major steps. The first one consists in the fractionation of the same extract using four different chromatographic stationary phases and appropriated elution conditions to obtain five fractions for each column. The second step corresponds to the acquisition of chemical profiles using HPLC-HRMS analysis, and the biological evaluation of each fraction. The last step evaluates the links between the relative abundances of molecules present in fractions (peak area) and the global bioactivity level observed for each fraction. To this purpose, an original bioinformatics script (encoded with R Studio software) using the combination of four statistical models (Spearman, F-PCA, PLS, PLS-DA) was here developed leading to the generation of a "Super list" of potential bioactive compounds together with a predictive score. This strategy was validated by its application on a marine-derived *Penicillium chrysogenum* extract exhibiting

* Corresponding author. Université de Nantes, Faculté de Pharmacie, MMS, BP 61112, F-44035, Nantes, France.

E-mail address: catherine.roullier@univ-nantes.fr (C. Roullier).

<https://doi.org/10.1016/j.aca.2019.04.038>

0003-2670/© 2019 Elsevier B.V. All rights reserved.

Please cite this article as: L. Ory et al., Targeting bioactive compounds in natural extracts - Development of a comprehensive workflow combining chemical and biological data, *Analytica Chimica Acta*, <https://doi.org/10.1016/j.aca.2019.04.038>

antiproliferative activity on breast cancer cells (MCF-7 cells). After the three steps of the workflow, one main compound was highlighted as responsible for the bioactivity and identified as ergosterol. Its antiproliferative activity was confirmed with an IC_{50} of 0.10 μ M on MCF-7 cells. The script efficiency was further demonstrated by comparing the results obtained with a different recently described approach based on NMR profiling and by virtually modifying the data to evaluate the computational tool behaviour. This approach represents a new and efficient tool to tackle some of the bottlenecks in natural product drug discovery programs.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

In Natural Product (NP) research, bioassay-guided fractionation approach is mostly used to isolate bioactive metabolites from a crude extract [1–3]. Despite the fact that this method has proved to be efficient for the discovery of many active compounds including taxol, artemisinin or vinblastine [4], it is now often considered by industrials as a time-consuming, costly and risky investment [5]. In some cases, the activity originally observed on a mixture can be lost due to irreversible binding of the components to chromatographic resins, degradation, chemical modification, antagonistic or synergistic effects. Therefore, the need to improve productivity and efficiency in the discovery of new bioactive NPs to address limitations of bioassay-guided fractionation has resulted in the past few years in the emergence of several novel strategies.

Even if such strategies are considered as recent, initial strategies were started in the 80's. Pouchus et al. [6], Samuelsson et al. [7] and Cardellina et al. [8], were among the first people interested in improving the classical bioassay-guided fractionation method for active NP extracts. In 1989, Pouchus developed a mathematical script allowing a better purification of active compounds by calculating their relative quantities in each fraction. The program was also able to detect potential synergistic effects or the presence of several active compounds in an active mixture. In 1985 and 1993, respectively, Samuelsson et al. and Cardellina et al. studied the chemical nature of bioactive compounds in NP extracts. Based on a combination of several extraction solvents and several Solid Phase Extraction (SPE) columns, they managed to deduce the chemical properties of active compounds (size, polarity, stability, acido-basic properties ...) and therefore develop appropriate and more focused purification strategies. Interestingly, the authors represented their results in an elution matrix with active fractions being highlighted, thus revealing the chemical profiles of the active compounds. Therefore, this approach also allowed them to better select their active extracts for subsequent investigation by avoiding probable synergistic effects and false positives (when activity was lost with fractionation) but also by performing preliminary dereplication (eliminating from further consideration) when activity of the fractions was observed with similar chemical profiles. This approach was further developed by Månsson et al. [9] as the Explorative Solid-Phase Extraction protocol (E-SPE), by implementing the approach with the acquisition of LC-UV-MS profiles for active fractions. These analyses provided additional chemical properties for the active compounds (MS and UV spectra), allowing better dereplication of the active constituents. However, the authors especially focused on recurrent peaks in active fractions to manually detect and highlight compounds responsible for the activity.

In the past few decades, metabolomics has appeared as a rapidly emerging and developing field in NP chemistry. Extracts analyses mainly by HPLC-MS⁽ⁿ⁾ and NMR spectroscopy has been employed for different purposes such as chemical profiling and dereplication

[10–13], biomarker characterization [14,15], quality control [16–18] and also for bioactive drug discovery [19–22]. Metabolomics studies generate a huge amount of data related to all detected compounds in a sample and need the use of bioinformatics tools to highlight the information of interest within the collected data. To this purpose many software have been developed such as MZmine [23], R Cran packages [24] or Python packages [25] and are now available for the community. Nevertheless, effective strategies for identification of compounds present in small amounts and their associated biological effect from a complex mixture are crucially lacking. While the most commonly used statistical analysis to compare the chemical composition of different mixtures remains principal component analysis (PCA), this approach is generally not sufficient in NPs drug discovery programs because it does not take into consideration bioactivity data. Consequently, a great deal of interest has risen among the scientific community to link chemical fingerprints to bioactivity data using statistical methods, leading to the new term “biochemometrics.” Such approaches would overcome some bioassay-guided fractionation limitations and provide a more comprehensive insight of compounds responsible for the activity. According to the literature, several statistical models were used for detection of active compounds such as Pearson correlation [26–29], partial least squares (PLS) [30–35], discriminant analysis (PCA-DA, PLS-DA, OPLS-DA) [21,28,36–41] and hierarchical cluster analysis (HCA) [42]. However, most of these studies use only one statistical model for interpretation, which may impair exhaustiveness and accuracy of highlighted features mainly because all of these models are not well adapted to delineate chemical fingerprints and bioactivity data relationships. A combination of them appears as an interesting solution to overcome the limitations of each model independently and increase performance of biochemometrics.

In this study, we propose a new workflow based on the association of both E-SPE and biochemometrics using the combination of multiple statistical models (PCA, Spearman, F-PCA, PLS, PLS-DA) to target bioactive compounds from extracts. This workflow was developed and applied in a real-case study, with an extract of a marine-derived fungal strain: *Penicillium chrysogenum* MMS5, presenting high antiproliferative activity on breast cancer cells (MCF-7 cell line), which was mainly due to the presence of high amounts of ergosterol [43]. The script, written using the open-source R Cran software, allowed to combine all data-mining strategies and was made accessible for the whole NP community (Supplementary information S1). This approach represents a new approach to tackle some of the bottlenecks currently existing in NP drug discovery programs.

2. Material and methods

2.1. Fungal strain, culture and extraction

Penicillium chrysogenum (MMS5) was collected in November

1994 from cockles in Le Croisic (France) and identified by ITS sequencing (Genbank accession number MK015724). This marine-derived fungal strain was 3-point inoculated on PDA medium (Potato Dextrose Agar) in 50 Erlenmeyer flasks (50 mL of media per flask) and incubated for 12 days at 27 °C. The crushed gelose and mycelium were extracted twice with 100 mL of a CH₂Cl₂/EtOAc (50:50, v/v) solvent mixture after 30 min ultrasound treatment for the first step extraction and after a night at room temperature incubation for the second. The supernatants were combined and filtered over Büchner and reextracted. The filtrates were dehydrated by Na₂SO₄ and filtered over filter paper and over a 0.45 µm regenerated cellulose membrane (Sartorius Stedim Biotech) to remove spores. Solvent was evaporated using a rotary evaporator to obtain a crude extract (710 mg).

2.2. Fractionation for E-SPE

Fractionations were performed on four columns of solid phase extraction (SPE) using Silica gel (Chromabond[®], 6 mL, 1000 mg, pore size 60 Å, particle size 45 µm, from Macherey-Nagel), C₁₈ (Chromabond[®], 6 mL, 1000 mg, pore size 60 Å, particle size 45 µm from Macherey-Nagel), Sephadex[™] LH20 (GE Healthcare Biosciences AB) and Strata[™]-X phase (6 mL, 100 mg, Phenomenex[®]). These phases are abbreviated in the following manuscript as SiOH, C₁₈, LH20 and SX, respectively. For separation using Sephadex[™] LH20, the dry powder was swelled in MeOH and manually packed (1.5 mL in 6 mL empty cartridges from Macherey-Nagel).

The different mobile phases used were composed of CH₂Cl₂/MeOH mixtures (from 1:0 to 1:1) for the normal phase Silica gel, MeOH/H₂O mixtures (from 0:1 to 1:0) for the reverse phase C₁₈, MeOH/H₂O mixtures (from 0:1 to 1:0 and additional MeOH + 1% formic acid) for the polymeric reverse phase Strata[™]-X, and only MeOH for LH20. Fractionation was performed in multiple sub-fractions until no more measurable mass was recovered from the column. The sub-fractions were then pooled in 5 fractions according to mass amounts. In our study, after placing on top of each cartridge a frit with 50 mg of the dried crude extract obtained from MMS5 *P. chrysogenum*, nineteen fractions were obtained by the following successive mobile phases: SiOH-1 (6 mL CH₂Cl₂ and 3 mL CH₂Cl₂/MeOH (1:9)), SiOH-2 (3 mL CH₂Cl₂/MeOH (3:7)), SiOH-3 (9 mL CH₂Cl₂/MeOH (1:1)), SiOH-4 (9 mL CH₂Cl₂/MeOH (1:1)), SiOH-5 (9 mL CH₂Cl₂/MeOH (1:1)), C₁₈-1 (3 mL H₂O, 3 mL MeOH/H₂O (1:2), 3 mL MeOH/H₂O (2:1) and 4 mL MeOH), C₁₈-2 (4 mL MeOH), C₁₈-3 (3 mL MeOH), C₁₈-4 (6 mL MeOH), C₁₈-5 (56 mL MeOH), LH20-1 (2.75 mL MeOH), LH20-2 (0.75 mL MeOH), LH20-3 (0.75 mL MeOH), LH20-4 (4 mL MeOH); LH20-5 (34 mL MeOH), SX-1 (4 mL H₂O, 3 mL MeOH/H₂O (1:2) and 3 mL MeOH/H₂O (2:1)), SX-2 (3 mL de MeOH), SX-3 (4 mL de MeOH), SX-4 (6 mL MeOH), SX-5 (21 mL MeOH and 59 mL MeOH + 1% formic acid). All fractions were dried under reduced pressure at room temperature.

2.3. Controls preparation

For all four columns, a blank sample was prepared by extracting the PDA medium used for cultivation in the same conditions as described above and eluting the extract with the same mobile phases on the four columns. All PDA medium sub-fractions were pooled to obtain a negative control (BM) for cytotoxic assay and chromatographic analyses. Similarly, a part of the crude extract from MMS5 previously spotted on a frit was eluted on a column without any phase (but equipped with a frit, top and bottom) with a mix of the different solvents to obtain a positive control (CE). Additionally, to make sure the activity was recovered in the fractions, a reconstituted crude extract (RCE) was obtained for all four columns by mixing the corresponding fractions in proportional

amounts. A quality control (QC) for subsequent HPLC-MS analyses was also prepared by mixing all the samples (fractions and controls) in equal quantities.

2.4. HPLC-MS analyses

Analyses of fractions and controls were performed on a Shimadzu instrument consisting of an Ultra-Fast Liquid Chromatography coupled to UV detection and High Resolution Electrospray Ionization Mass Spectrometry combining Ion trap and Time of Flight analysers (UFLC-UV-ESI-IT-TOFMS). The unit consists of two LC-20ADxr pumps, a SIL-20ACxr autosampler, a CTO-20AC column oven, an SPD-M20A PDA detector and a MBC-20A system controller. High performance liquid chromatography analyses were performed on a Kinetex[™] C₁₈ column (100 × 2.1 mm, 2.6 µm, Phenomenex) heated in an oven equilibrated at 40 °C. A mobile phase consisting of CH₃CN/H₂O (acidified with 0.1% formic acid) was used, starting with 15% CH₃CN during 2 min, then increasing linearly to 100% CH₃CN within 23 min, holding at 100% CH₃CN for another 5 min, then returning to the initial conditions within 1 min, and holding for 4 min, for a total run time of 35 min at a flow rate of 0.3 mL/min. The mass spectrometer was operated in full-scan mode. MS data were recorded in the ESI positive mode in the mass range of *m/z* 100–1000 with a mass accuracy of 7 ppm and a resolution of 10,000 at *m/z* 500, using the following parameters: heat block and curved desolvation line temperatures at 200 °C; nebulizing nitrogen gas flow at 1.5 L min⁻¹; interface voltage at (+) 4.5 kV and detector voltage of the TOF analyser at 1.6 kV. UV-VIS spectra were detected and collected from 190 to 600 nm.

The samples were prepared in MeOH (UPLC/MS grade) at concentrations of 0.4 mg mL⁻¹, stored at 4 °C before injection of 5 µL for each. The analyses were performed randomly and included solvent blank samples (pure MeOH) and QCs injected regularly throughout the sequence.

2.5. Cytotoxicity assays

Human breast cancer MCF-7 cells were purchased from the European Collection of Animal Cell Cultures (ECACC, Salisbury, UK). 3-(4,5 Dimethyl-2-thiazolyl)-2,5-diphenyltetrazolium bromide (MTT) was purchased from Sigma Aldrich (Saint Quentin Fallavier, France). MCF-7 cells were cultured at 37 °C in a humidified incubator with 5% CO₂ in DMEM medium supplemented with 10% fetal bovine serum (FBS), 1% glutamine and 1% penicillin-streptomycin. Viability of MCF-7 was tested in 96-well plate at a density of 10 000 cells per well in 200 µL of culture medium and allowed to adhere overnight. Then the seeding medium was removed before cell treatment. Crude Extract (CE), fractions and Reconstituted Crude Extracts (RCE) were dissolved in DMSO and tested at a unique concentration of 50 µg mL⁻¹. Pure ergosterol and ergosterol-5,8-endoperoxyde (purchased from Sigma-Aldrich and from in-house library respectively) were dissolved in EtOH and then diluted in 0.1% BSA containing-medium in order to obtain concentrations of 0.001-0.01-0.1-1-12,5-25-50 µM. After 24 h of incubation, MTT assay was performed by removing 100 µL of the medium and adding 50 µL MTT (at 2.5 mg mL⁻¹) to each well. The mixture was further incubated for 4 h, and the liquid in the wells was removed thereafter. Dimethyl sulfoxide (DMSO 200 µL) was then added to each well to solubilize the formazan product and the absorbance was read at 570 nm. The relative inhibition was expressed as a percentage of the non-treated control, which corresponded to medium supplemented with the same final concentration of DMSO or EtOH.

2.6. LC-MS data treatment

LC-MS data obtained for studied mixtures were treated. Automatic feature detection between 0 and 30 min in positive mode in MZmine2.31 software [23] was achieved using the parameters selected according to the TOF-MS detector. Peak detection was performed with the “mass detection” algorithm with a noise level of “1.6E4” in centroid mode. Then, chromatograms were built for all detected ions with a minimum time span of 0.1 min and a minimum intensity of “8.0E4” counts in positive mode, allowing 80 ppm tolerance on m/z values. Peak deconvolution was applied to the generated chromatograms with the “baseline cut-off” algorithm using a min peak height of “8.0E4”, a peak duration range of 0.1–5 min and a baseline level of 1.6E4. Deisotoping filter was applied using the “isotopic peaks grouper” module with tolerance parameters adjusted to 0.1 min and 0.01 on m/z values. Feature alignment was achieved with the “Join Aligner” module with a m/z tolerance of 0.01 and a retention time tolerance of 0.3 min followed by gap filling using the “Gapfiller” module, yielding a combined dataset. The features detected from blank MeOH and non-inoculated culture medium samples (BM) were removed from the generated matrix to focus on the features really corresponding to the fungus production.

2.7. Biochemometric analyses

To link chemical profiles and associated biological data of all fractions obtained, biochemometrics analyses were performed. The spectral data matrix obtained from Mzmine software [23] including m/z , retention time, and peak area for each detected ions was imported to Excel Microsoft office 2011 version 14.7.7 (2010 Microsoft Corporation) and merged with the bioactivity dataset (inhibition percentages at $50 \mu\text{g mL}^{-1}$) to form the final matrix. The R FiBiCo script developed in the present work is presented in details in the “Results and Discussion” section. All statistical models used and included in the script such as Principal Component Analysis (PCA), Spearman, Focused- Principal Component Analysis (F-PCA), Partial Least Squares regression (PLS) and Partial Least Squares Discriminant Analysis (PLS-DA) were computed using the open source software Rstudio version 1.1.447 (2009–2018 RStudio, Inc) and the following packages: “MetaboAnalystR” [44], “psy” [45], “mdatools” [46], “readr” [47] and “psych” [48].

2.8. Virtual evaluation of the script efficiency

The initial data matrix was manually modified to evaluate the ability of the R FiBiCo script to detect potential bioactive compounds. A total of 65 matrices were designed by adding a virtual feature which peak area values in each fraction were defined according to different models of relationships between peak area values and biological activity (linear, exponential, logarithmic and two other relationships with monotonic and non-monotonic “S-shaped” curves). For all relationships peak area values in each fraction for the virtual feature were given in accordance to the antiproliferative activity on MCF-7 (in %) with lowest peak area value for the less active fraction C18-2 (except for the non-monotonic model) and the highest for the most active fraction SX-5. Based on the initial matrix, where peak area values ranged from “8.96E3” to “7.68E8”, different matrices were then constructed with the additional virtual feature having peak area values either in the same range (from “8.96E3” to “7.68E8”) or 1000 times less intense (from “8.96E3” to “7.68E5”). The different models were here obtained by using the following equations for peak area value range from “8.96E3” to “7.68E5”: Eq. (1) for linear, Eq. (2) for exponential and Eq. (3) for logarithmic, with A corresponding to

biological activity values (inhibition percentages as obtained on MTT assays) and P_{area} to calculated peak area values. For monotonic and non-monotonic “S-shaped” curves, peak area values were assigned manually to fit this type of curve.

$$A = 1.00 \text{ E7} \times P_{area} - 18.94 \quad \text{Eq. (1)}$$

$$A = 0.95 \exp. (5.56 \text{ E} - 6 \times P_{area}) \quad \text{Eq. (2)}$$

$$A = 16.44 \ln(P_{area}) - 168.53 \quad \text{Eq. (3)}$$

Moreover, a random variation of 10, 30 and 50% (using “randbetween” function in excel) on each peak area value defined for each model was applied. Additionally, three additional matrices were obtained with virtual peak area values defined according to activity groups. This time, random values were assigned (using “randbetween” function) to the virtual feature based on the following conditions: (1) in accordance with biological activity groups, with peak area values in group 1 (the less active) ranging from “1.00E4” to “5.00E5”, in group 2 from “5.00E5” to “2.50E7” and in group 3 from “2.50E7” to “1.25 E9”, (2) not in accordance with biological activities with values in group 3 < group 2 < group 1 or with values in group 2 < group 1 < group 3. Finally, a last assay was performed with completely random peak area values assigned to the virtual feature based on the initial matrix range (from “8.96E3” to “7.68E8”) or on a smaller range (from “8.96E3” to “7.68E5”) for the 19 samples.

These experiments were repeated three times to obtain three matrices for each condition (except for matrices where the peak area value of virtual feature range from “8.96E3” to “7.68E8” with a random variation of 10%), which were all analysed by the FiBiCo script. The results generated were compared, especially the position of the virtual feature in the “super list” and which models allowed its selection.

2.9. NMR pharmacophoric deconvolution approach

Fractions C18-1 to C18-5 were dissolved in deuterated methanol ($10 \times 0.75 \text{ mL}$) from Eurisotop, (Saint-Aubin, France) at a concentration of 5 mg mL^{-1} . Acquisitions were performed on a Bruker NMR spectrometer operating at 500 MHz. COSY spectra were acquired on 2 scans, 4134 points on F2 axis and 512 points on F1 axis. Spectral processing was performed using the Plasmodesma program, written in Python and based on the SPIKE library [49,50].

3. Results and Discussion

As identifying relevant active molecules from complex mixtures represents a major challenge in NP drug discovery, this study here proposes a complete workflow to target bioactive compounds from a crude extract, based on a combination of fractionation, metabolomics and chemometrics. It is divided into three steps, namely the Explorative-Solid Phase Extraction (E-SPE) followed by the acquisition of biological and chemical data and the biochemometrics analysis with the combination of four statistical models (Fig. 1). The last step required the development of the new FiBiCo script as described below and available in Supplementary information S1.

3.1. Development of the bioinformatics tool to Find Bioactive Compounds: “FiBiCo” script

3.1.1. Statistical models

So far, different models have been described in the field to link biological to chemical data, mostly linear Pearson correlation, PLS

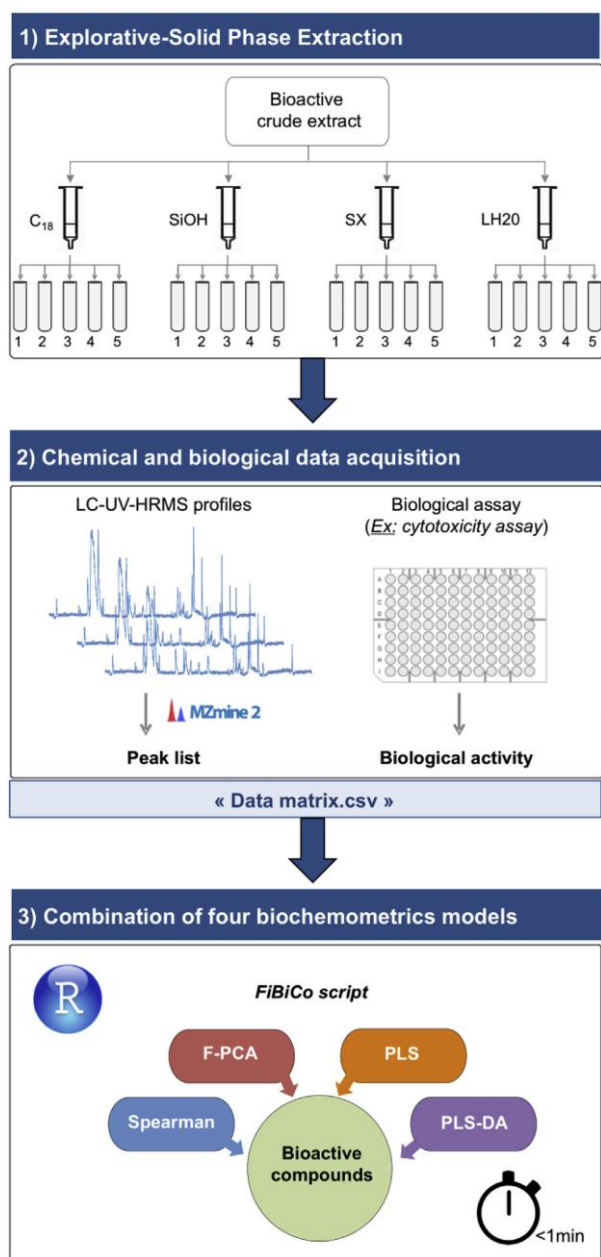


Fig. 1. Bioactive natural products discovery workflow based on a combination of Explorative-Solid Phase Extraction (E-SPE) and four biochemometrics approaches (Spearman, F-PCA, PLS, PLS-DA).

and PLS-DA. However, not many compare the results obtained from several analyses [37]. This is the reason why, in the present study, it was chosen to combine results from complementary univariate and multivariate models, namely Spearman correlation, focalised-PCA, PLS and PLS-DA. This should provide a more comprehensive view for the identification of bioactive compounds taking advantages of the differences between all those statistical approaches [51]. Spearman correlation is a univariate statistical approach, which measures the strength of correlation between two variables (*i.e.* peak area and biological activity) by evaluating their monotonic relationships (rank correlation coefficient). In this study, positive

Spearman correlation scores (r) closer to 1 represented features, which peak area values seemed to increase along with biological activity. F-PCA, which was developed by Falissard *et al.* [52], while never used in the field of NP until now, revealed to be an interesting feature selection strategy. In fact, this model relying on Pearson correlation (linear correlation) allows to visualize simultaneously on a graphical representation, both correlations between a variable of interest Y (*i.e.* biological activity) and a set of variables X (*i.e.* feature peak areas) and correlations between X variables themselves. Like PCA, it corresponds to the projection of a correlation matrix in a two-dimensional plane, but unlike PCA, it accurately represents correlations of a given variable with the others (represented by the radius of the concentric circles) and even to test the statistical significance at the 5% level (represented by the red circle). The F-PCA plot center then corresponds to the variable of interest Y (biological activity) directing the analysis, whereas other variables (X) are represented by dots (*i.e.* features). Their colour, localisation and organization around this central point allow to define the two types of correlations previously described. Consequently, the closer the feature is to the center, the closer to 1 (or -1) is the correlation coefficient. The red circle delimits statistical significance at the 5% level and allows to highlight dots with significant positive or negative correlation. Moreover, two dots close to one another indicate a strong positive correlation between these features themselves, while two diametrically opposed dots indicate a strong negative correlation between them. Partial Least Square (PLS) and Partial Least-Squares with Discriminant Analysis (PLS-DA) models are multivariate linear regression methods, commonly used in biochemometrics [53]. These supervised methods differ from PCA, because they allow to maximize the covariance of independent variables (peak area in our case) with a dependent variable (*i.e.* biological activity). PLS-DA allows to sharpen the separation between defined groups (*i.e.* nonactive/moderately or active fractions) of observations, by rotating PCA components such that a maximum separation among classes is obtained, and to understand which features carry the class separating information. Many values are employed in PLS for feature selection, including PLS loadings, weights, variable importance on projection (VIP), regression coefficients (RC), target projections, and selectivity ratio (SR) [54–56]. To date, VIP and selectivity ratio are the most popular ones in metabolomics [57].

The combination of all those models appeared as a new useful tool to comprehensively highlight features of interest. Metabolites appearing with a high score on several statistical models strengthen their importance to explain biological activity observed in the extract and fractions, and thus confirming their potential effect on the biological target studied.

3.1.2. Design of the script

This script was written using R Software (CRAN) to perform an automated processing of biochemometrics analyses on MacOS or Windows systems. It was designed to allow any user to perform the analyses on his/her own dataset. Therefore, it was very important that the operator could define and modify some parameters to improve methodology efficiency in accordance with his/her own data. In the same way, the file containing LC-MS and biological activity data to be read by the script was arranged as a simple matrix in a comma separated values (.csv) file. In this matrix, columns contained in the following order: sample names, groups, biological activity and peak areas or intensities of the different features “*m/z*_RT” detected. The script was divided into five steps (Fig. 2) and used “MetaboAnalystR” [44], “psy” [45], “mdatools” [46], “readr” [47], “psych” [48] packages.

The first step consists in filling information in the Rscript section “Define important information” with (1) general parameters

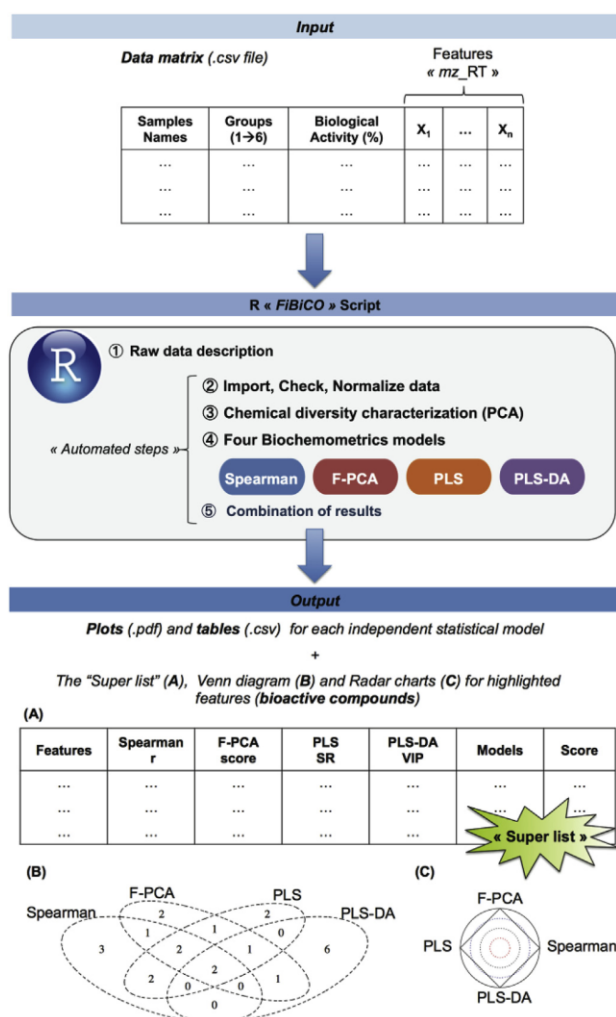


Fig. 2. Design of the "FiBiCo" script with its 5 major steps and the different results generated. At the end of the script, one bioactive compound can correspond to several highlighted features on (A), (B) and (C), as the same compound can generate several ions by electrospray ionization.

(operating system, file path to download the data matrix and working directory to export the results); (2) specific parameters according to the biological assay (biological activity, range and colours for groups ranked by ascending order); (3) parameters for normalization of raw data matrix (choosing options for sample normalization, data transformation and data scaling) [58] and (4) parameters for statistical analyses and graphical representations (as xy axes used for PCA and PLS with the associated titles and legends, etc.). After this first unique required contribution, the following 4 steps can be run successively without any more input from the operator. **The second step** of the script consists in importing, checking and normalizing the data matrix. **The third step** proceeds to the chemical diversity characterization of the fractions with PCA analysis. One output of this third step includes a graphical representation of the PCA score plot allowing the operator to assess if the best components to describe his/her data have been defined appropriately, and to highlight potential outliers. **The fourth step** is devised to perform biochemometrics analyses with

the four chosen statistical models: Spearman, F-PCA, PLS and PLS-DA. Each model allows to determine the link between peak area (or intensities) and biological activity. Finally, **the last and most important step** in the biochemometrics tool designed in the present study, is the combination of all statistical results generating a "Super list" of features presenting a high score in the four models. It then highlights the metabolites for which the presence most probably explains the activity observed for the initial extract.

For the fourth step performing all biochemometrics analyses, a first output devised in the script was the generation and exportation in the working directory of one table (in a.csv file) per model with all the results obtained, together with the associated graphs (.pdf) as "Spearman score plot", "F-PCA score plot", "PLS score plot" and "PLS-DA score plot" (Supplementary information S2.1). To get a first overview of all detected peaks and all highlighted ones according to each model independently, each table was implemented with an additional classification column with importance levels as "High/Medium or Low". A combined graph representing the results of the four models was then created (with the colour code blue for "high", orange for "medium" and grey for "low"). To perform the classification, it was necessary to define conditions for each model (Table 1). While Spearman resulted in only one score of correlation ("r" coefficient) and was easier to classify, PLS and PLS-DA statistical models generated several important scores. Therefore, only few of them were selected for PLS and PLS-DA. So far, selectivity ratio for PLS and VIP score for PLS-DA have been reported to be the most important scores for interpretation of differences between groups or samples [33,35,59]. However, they did not reflect if the correlation was positive (activity was due to the presence of peak) or negative (activity was due to the absence of peak). Consequently, it was required to also take into account the sign value of regression coefficient for PLS model, and the peak area concentration of features in the groups for PLS-DA model. For F-PCA, the classification was carried out based on the graphical results, taking into account the three parameters defined as the correlation score (corresponding to the distance for each feature to the center of the graph), the sign of the correlation and the significance of the score (corresponding to features localised inside the red circle of the graph). This latter score was calculated as reported by Falissard et al. [52]. Some other parameters were calculated and reported in the table during the script (cf NA in Table 1). While they were not used for subsequent classification, they were kept available in the final table, as they could be useful for further in-depth interpretation.

For the final step of the script combining all the results from each independent model, it was chosen to allow the operator to define a maximal number of interesting ions for each model to obtain a suitable and easily interpretable list. It's important to note that the resulting final number of interesting features can be less if a smaller number of features (or even none) is in accordance with the defined limits in each model. For example, if the operator decided to choose a maximum of ten potential bioactive features per model, the "Super list" would contain a maximum of forty features (from 0 to 40). To focus on metabolites having a high score on several statistical results, the output table was designed to contain annotations identifying which model(s) allowed the selection of metabolites. Moreover, the distribution of this model-dependent selection was graphically represented on a Venn diagram. Additionally, after normalization of each model score (range 0–1), a global score for each feature was then calculated either by using the sum (range 0–4), the mean (range 0–1) or the Euclidean distance to 0 (range 0–2). Graphical representation of the "Super list" was finally generated as radar charts for each feature ordered by the calculated global score.

Table 1
Different values recovered from the statistical analyses performed by the FiBiCo script and conditions defined for correlation levels classification.

Models	Values generated and exported in the final tables	Classification of importance levels		
		High	Medium	Low
Spearman	Spearman coefficient score "r"	$r \geq 0.5$	$0.5 > r > 0$	$r \leq 0$
F-PCA	F-PCA score (combining distance and correlation sign)	F-PCA score > 0 & Significant ^a	F-PCA score > 0 & Non-significant	F-PCA score ≤ 0
PLS	Selectivity Ratio (SR)	$SR \geq 1$	$SR < 1$	NA ^b
	Regression Coefficients (RC)	& RC > 0	& RC > 0	RC ≤ 0
	Variable of Importance in the projection (VIP)	NA	NA	NA
PLS-DA	Distances 1 and 2 ^c	NA	NA	NA
	Variable of Importance in the projection (VIP)	$VIP \geq 1$	$1 > VIP \geq 0.5$	$VIP < 0.5$
	Mean peak areas or intensities for each group C ₁ , C ₂ , ..., C _x ^d	$C_x > C_1$ & $C_x > C_2$ & ... & $C_x > C_{(x-1)}$	$C_x > C_1$ & $C_x > C_2$ & ... & $C_x > C_{(x-1)}$	NA
	Regression Coefficients (RC)	NA	NA	NA

^a F-PCA score (dot on the score plot) localised in the red significant circle (meaning F-PCA score > limit value of significance at 5% level calculated by the "psy" package and depending on the data, especially the number of features).

^b NA (not applicable) is noted when these values calculated by the script were not used for the classification of importance levels.

^c On the PLS loadings plot, distances 1 and 2 correspond to the distance to the point of biological activity, and to the regression line, respectively.

^d The number of defined groups being "x" (maximum x value = 6) and the xth group being the most active; NA (not applicable) is noted when these values calculated by the script were not used for the classification of importance levels.

3.2. Application to the marine-derived *P. chrysogenum* MMS5 extract

Following an initial screening of fungal extracts from our in-house library on MCF-7 breast cancer cell proliferation, extracts from the marine-derived *Penicillium chrysogenum* MMS5 isolated from cockles displayed strong inhibition whatever the growth medium used with 57–66% inhibition on CYA, YES and PDA media at $10 \mu\text{g mL}^{-1}$ (data not shown). It was found that the activity of this extract was related to high amounts of ergosterol, which was recently described as cytotoxic on MCF-7 cells [43] (Supplementary information S2.2). The proposed workflow was then applied to this MMS5 extract to evaluate its ability to highlight this bioactive compound.

3.2.1. Explorative-solid phase extraction: bioactivity elution matrix

Explorative-solid phase extraction on the crude extract was carried out with three stationary phases classically used in the NP field (SiOH, C₁₈, LH20) and a polymeric phase (StrataX). Those phases allow the fractionation of chemical constituents of the extract according to their physicochemical properties, such as polarity and hydrophobicity. As one fraction from SiOH (SiOH-5) did not present any measurable mass, a total of 19 fractions (4, 5, 5 and 5 for SiOH, C₁₈, LH20 and StrataX, respectively) were submitted to biological assay (MCF-7 proliferation inhibition), along with the positive control corresponding to the crude extract eluted without any stationary phase (CE) and a reconstructed crude extract (RCE). These latter RCE were obtained by mixing the fractions recovered from each column in the same proportions. This allowed evaluation of potential degradation of compounds present in the CE. For an easier visual representation, data results were sorted in an elution matrix (Fig. 3), where fractions were coloured in three groups according to their proliferation inhibition activity: red (0–20%), orange (20–40%) and green (40–60%). The activities ranged from 0 to 54% and the amount of fraction collected suggested a homogeneous separation of NPs present in the crude extract among them. As an example, all fractions from C₁₈ column presented similar amounts (around 7 mg), while their corresponding antiproliferative activities were different, from no inhibition (C₁₈-1, C₁₈-2 and C₁₈-3) to higher inhibition (43% and 50% for C₁₈-4 and C₁₈-5, respectively). SiOH column was the less efficient separation strategy as most of the weight was recovered in fraction SiOH-2 along with most of the activity, contrarily to SiOH-5 which was not tested due to the absence of mass.

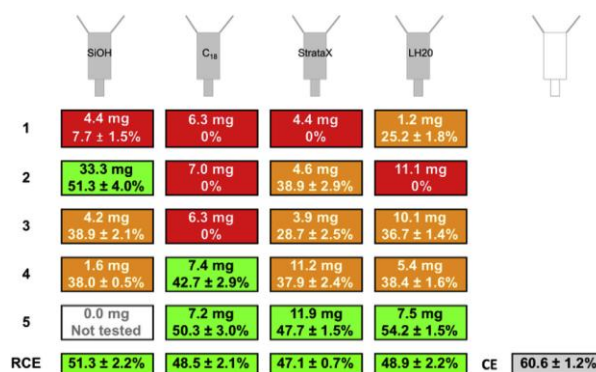


Fig. 3. Bioactivity elution matrix of the active crude extract indicating for each fraction the corresponding dry weight and proliferation inhibition percentage obtained at $50 \mu\text{g mL}^{-1}$ on MCF-7 cells (mean of triplicates \pm SD), with the 3-group colour code defined as red (0–20%) for inactive, orange (20–40%) for moderately active and green (40–60%) for highly active samples. RCE and CE correspond to recombined crude extracts with fractions mixed in the same proportions and crude extract, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Interestingly, from this elution matrix, a first insight into the chemical nature of bioactive compounds could be hypothesized as being rather non-polar. Indeed, highest activity levels were observed in the last fractions of reversed phase and in the second fractions of normal phase columns. From this perspective, LH20 and StrataX columns also seemed to release the most active compounds in the last fractions meaning hydrophobic interactions may play an important role in these columns [60]. Interestingly, sufficiently distinct activity profiles were observed for C₁₈, LH20 and StrataX columns, allowing further comparison of their respective fraction compositions. It is important to note that several compounds should be responsible for the activity as a slight inhibition was observed for LH20-1 while this activity was lost for the following fraction and recovered afterwards. Finally, the maximum of proliferation inhibition obtained for individual fractions appeared at similar levels as RCE, meaning there was no loss of activity during fractionation and that activity was not depending on synergistic effect between compounds from different fractions. However, activities of the four RCE (48–51%) while close to the initial CE activity (61%), were slightly lower, probably in relation to the mass loss on

the column.

3.2.2. Chemical profiling and samples characterization

All fractions obtained from E-SPE were analysed with liquid chromatography coupled to high-resolution mass spectrometry (LC-HRMS) to compare their chemical profiles. Peaks only appearing in active fractions and not in inactive fractions were potential candidates for the observed bioactivity. However, no differences were revealed by manual comparison of the chromatograms, as their noticeable differences (corresponding to major peaks) were not linked to activity modulation (Supplementary information S2.3). Moreover, this manual inspection tends to be a very time-consuming task, in particular when differences are not obvious and if the active compounds correspond to low intensity peaks. Therefore, in this configuration, a metabolomics strategy appears more appropriate for the analysis of LC-MS profiles in regards to measured cell proliferation inhibition. This allows a more comprehensive description of the chemical composition of each fraction. Thus, automatic peak detection was here performed on raw chemical profiles by using Mzmine2 software [20], which allowed to obtain a list of 795 individual features among all samples, corresponding to peaks characterised by their m/z values and associated retention times (referred as “ m/z -RT”), with their respective peak areas. The aligned data matrix was amended with biological data (inhibition percentages and groups as defined previously) and applied to the newly developed R FiBiCo script for statistical analyses.

After the first two steps consisting of filling needed information, defining appropriate parameters, uploading and normalizing the data, the script proceeded to the third step. The chemical diversity of fractions obtained from the E-SPE was then investigated using principal component analysis (PCA), an unsupervised multivariate statistical approach. PCA is based on a linear transformation that preserves as much of the variance in the original data as possible in a lower dimensional space (usually two- or three-dimensions). The first three components of the PCA explained 78.6% of the total variability of the model (with 41.7%, 28.3% and 8.3% for components 1, 2 and 3 respectively). PCA score plot (Fig. 4A) obtained for the components 1 versus 2, revealed that the chemical composition of the fractions from fractionation using SiOH column were very different from the other fractions. One exception was SiOH-2, which corresponded to the most active fraction of this specific fractionation. This confirms that separation on this phase was orthogonal to the other E-SPE phases. Interestingly, the PCA seemed to show that bioactive samples stood out from the others according to component 1. However, no feature on the loadings plot (Fig. 4B) related to the observed biological activity could be selected from PCA. Therefore, additional specific statistical models provided by the R FiBiCo script were needed to provide accurate insight into the biologically active chemical constituents.

3.2.3. Biochemometrics analysis

After running the fourth step of the script corresponding to the biochemometrics analysis, the four statistical models returned very interesting features “highly” related to the biological activity (Table 2). This observation was consistent with the possibility to find bioactive metabolites produced by *P. chrysogenum*. Differences in features importance levels were observed. For example, among the 795 features detected by LC-MS, the script allowed to highlight 50, 42, 38 and 51 features with “high” importance level for Spearman, F-PCA, PLS and PLS-DA, respectively.

These features of interest were highlighted in blue (“high” importance level) on the combined graph representing the results of the four models (Fig. 5). Spearman score plot (Fig. 5A) displayed two main groups of chromatographic peaks indicating potential

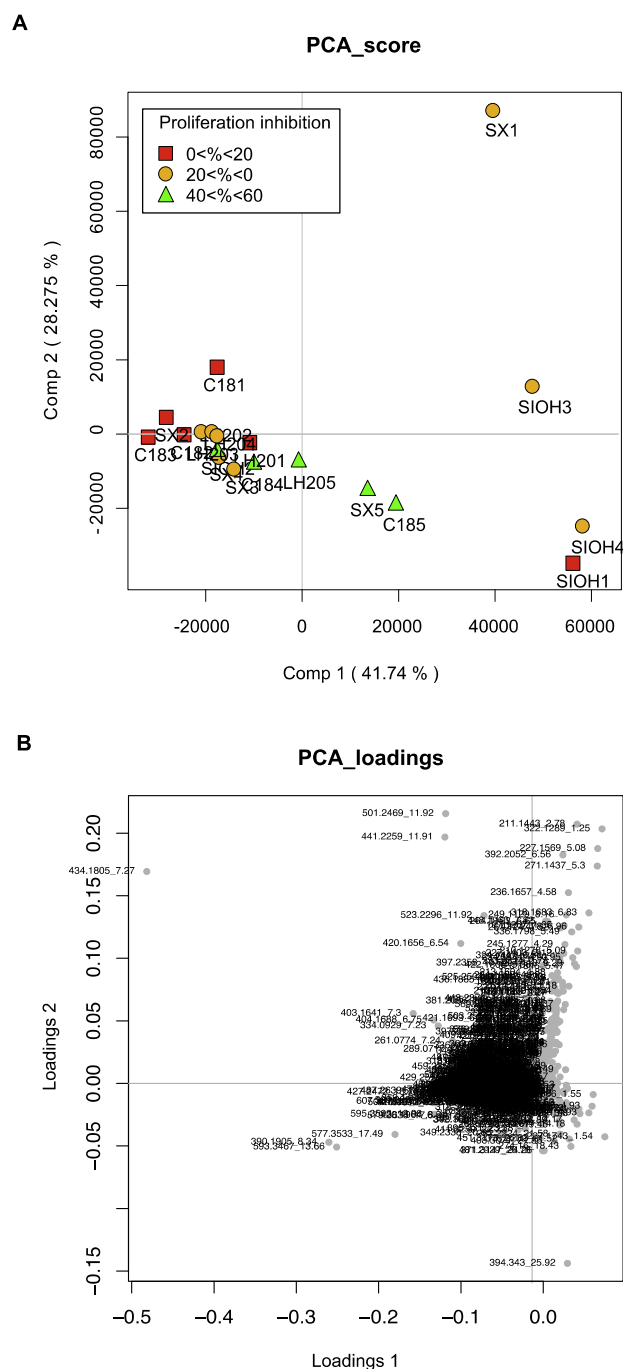


Fig. 4. Principal Component Analysis (PCA) score plot (A) of fractions based on the 795 features (LC-MS peaks), with the 3-group colour code according to proliferation inhibition (MCF-7) defined as red (0–20%), orange (20–40%) and green (40–60%) and the corresponding PCA loadings plot (B). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

bioactive compounds were either moderately polar (RT = 10–15 min) or apolar (RT = 20–28 min). Similarly, F-PCA (Fig. 5B) also revealed two groups of “high” importance features: one in the upper right corner and another in the upper left corner presenting many dots being more significantly correlated to the

Table 2

Number of LC-HRMS features related to proliferation inhibition of MCF-7 cell lines with high, medium or low importance levels in the case of *Penicillium chrysogenum* MMS5 originating data matrix.

Importance Levels	Spearman	F-PCA	PLS	PLS-DA
High	50	42	38	51
Medium	140	164	221	72
Low	605	589	536	672
Total	795	795	795	795

activity (because closer to the central red point representing the biological activity variable) with r coefficients between 0.6 and 0.8. In accordance with the previously defined parameters for the classification, these features (coloured in blue) were highlighted because they had both an absolute F-PCA score value significantly different from 0 at the 5% level (inside the red circle) and a positive correlation sign. Among these highlighted features, F-PCA allowed to obtain additional information of the relationship between features themselves. In fact, many of these dots were closely located on the graph because they also correlated between them, meaning they possessed similar presence pattern across fractions. However, no strong negative correlation was observed between them (absence of diametrically opposed blue dots). The fact that two groups of features were observed in contiguous quadrants could mean that not only one compound (but at least two) has an

influence on the activity observed. Interestingly, PLS model results (Fig. 5C) highlighted similar peaks as Spearman in the RT range 20–28 min. Additionally, the two first VIP from PLS-DA model (Fig. 5D) seemed to highlight features at 26.8 min presenting high area value in the most active group of samples (group 3). The results obtained from PLS and PLS-DA gave satisfying R^2 values (0.82 and 0.96 respectively), meaning the models were correctly fitting the data, allowing to trust their results [61,62]. From these first observations, it appeared interesting to compare highlighted features and especially bring out those which were common in all models. This is where the last step of the FiBiCo script was carried out to provide a “Super list” of most interesting features and their corresponding radar charts (Supplementary information S2.4 and S2.5), based on the combination of the top ten of each model. In the case of *P. chrysogenum* MMS5 a total of 23 highly interesting features was present in the “Super list”. The corresponding Venn diagram (Fig. 6) displayed the importance of combining the results from the four different statistical analyses as they highlight complementarily the features of importance. In fact, among the 23 highlighted in the “Super list”, only 2 were highlighted by all the models, while 13 were selected by only one of them, i.e. 3, 2, 2 and 6 features by Spearman, F-PCA, PLS and PLS-DA respectively (Fig. 6A). It is important to note that complementarity between the models remained even when increasing the number of features to select in each model to the top-20, as in this case, only 6 (out of 41) were

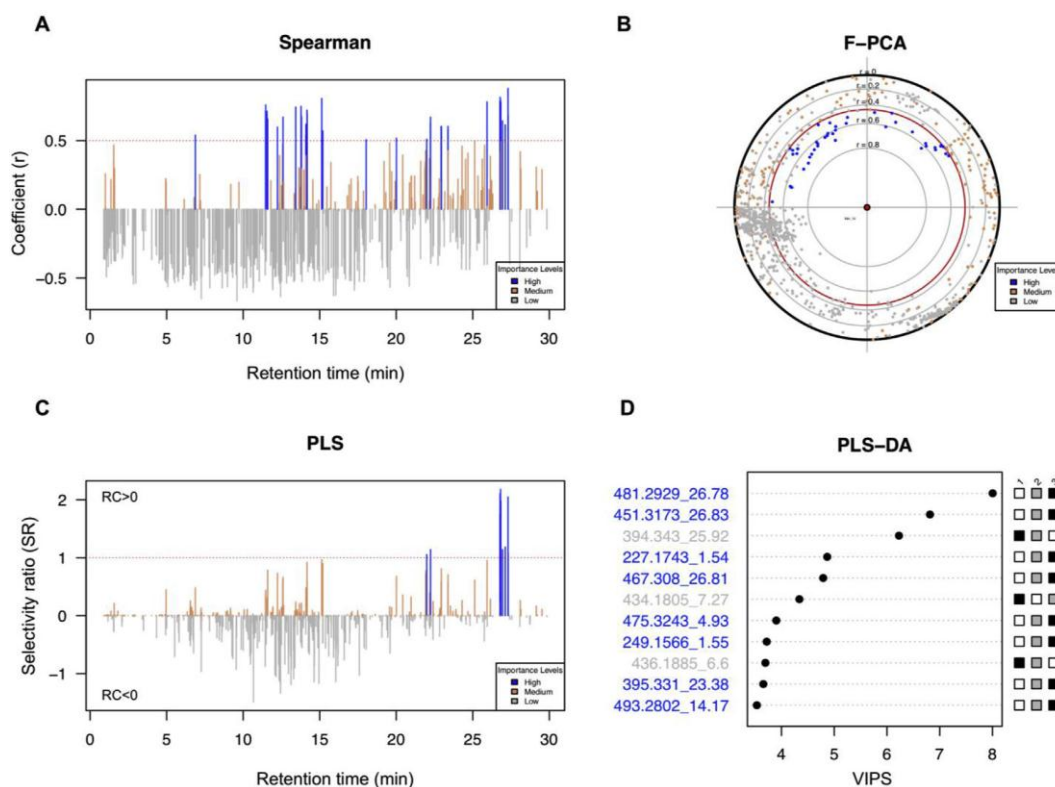


Fig. 5. Graphical representation of results obtained from the four statistical analyses: (A) Spearman, (B) F-PCA, (C) PLS and (D) PLS-DA. A colour-code “blue/orange/grey” was defined according to the feature importance level, with blue corresponding to features with high importance levels, most probably related to bioactive compounds. Orange refers to lower but positive relationships, while grey corresponds to features with low importance levels. The limit value used to select high important features was represented as a red line for Spearman (correlation coefficient (r) ≥ 0.5) and PLS (Selectivity Ratio (SR) ≥ 1 and Regression Coefficient (RC) ≥ 0). For F-PCA model, the red circle corresponded to the limit value of significance for the F-PCA score at the 5% level (features inside this circle have positive or negative correlations significantly different from 0). For PLS-DA, the greyscale on the right represents the mean peak area values in each group for the top-11 features, which are presented by decreasing VIP value (≥ 1) from top to bottom. Highly important features (in blue) correspond to those where the mean peak area value in the most active group is higher than in the others. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Please cite this article as: L. Ory et al., Targeting bioactive compounds in natural extracts - Development of a comprehensive workflow combining chemical and biological data, *Analytica Chimica Acta*, <https://doi.org/10.1016/j.aca.2019.04.038>

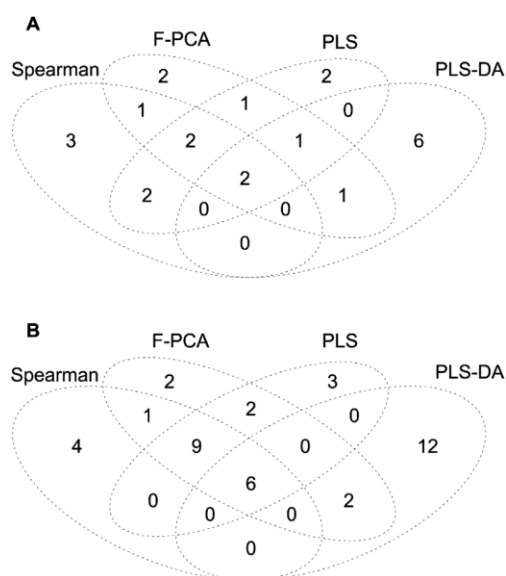


Fig. 6. Venn diagrams representing overlapping interesting features returned by the 4 models of the script (Spearman, F-PCA, PLS and PLS-DA) with the combination of (A) the top-10 features from each model giving a total of 23 features and (B) the top-20 giving 41 features.

highlighted by all the models (Fig. 6B). The user should then be careful on the definition of this number of features (10, 20 or more) to select in each model for the construction of the “Super list”, according to his/her own data, as if too restrictive, it may exclude features of interests in some models.

3.2.4. Identification of highlighted hits and confirmation of biological activity

The “Super list” of 23 features was then thoroughly investigated to annotate and identify the potential bioactive compounds. Overlay of their extracted ion chromatograms (XICs) revealed a perfect overlap for 12 out of the 23 features. Consequently, this overlap suggested that the 12 features arose from the same molecule. It was identified as ergosterol, as expected, by comparison to HPLC-UV-HRMS profiles of a standard (Supplementary information S2.6). The workflow proposed and developed in the present article, combining E-SPE and biochemometrics, was then successful in identifying the main bioactive compound responsible for the initial activity observed from a complex mixture (MMS5 *P. chrysogenum* extract). It is important to note that, while present in large amounts in the extract, ergosterol was not detected at high levels by mass ESI-spectrometry, because of a poor ionization. This means that the script developed allowed to highlight minor peaks as corresponding to potential bioactive compounds.

The other 11 features highlighted by the “Super list” were further investigated to understand if they could also have an influence on the activity observed or be considered as false positives. By grouping them according to peak shapes and retention times, eight other putative molecules could be revealed with the following retention times: 1.5 min, 4.2 min, 14.2 min, 15.1 min, 23.4 min, 25.9 min, 26.9 min and 27.3 min. Among these, annotation was tentatively carried out based on the observed adducts and the predicted molecular formula (Supplementary information S2.7). One of them matched with ergosterol-5,8-endoperoxide, which was further confirmed by the injection of a standard. Its biological activity on MCF-7 breast cancer cell was then confirmed with a proliferation inhibition percentage of 48.7% at 50 μM , which

is much lower than ergosterol but in accordance with our results (lower score: 1.34 compared to 3.51 for ergosterol). Moreover, annotation of chaetoglobosin derivatives could be proposed for the compound at 15.1 min (global score of 1.91). Interestingly, cytotoxicities were reported on breast cancer cells lines for chaetoglobosins A, C and G with IC_{50} of 37.56, 19.97 and 38.77 μM on MDA-MB-231 cells, respectively [63]. The last compound with a good score at 27.3 min had a hit for an ergosterol derivative, which could make sense regarding the activities of both ergosterol and ergosterol endoperoxide in this series. Other annotations included putative diketopiperazine derivatives for compounds at 1.5 and 4.2 min, which biological activities on breast cancer cells were not reported. It is important to note here that these corresponded to signals that were not detected in the crude extract. It could be due to a concentration phenomenon inherent to the fractionation step. For other features, due to the difficulty to predict the molecular formula or a number of matches in databases too important, a good annotation could not be obtained. Further purification work is then currently in progress towards these other features that were not identified. In fact, even if we cannot ascertain that all the features highlighted by the script have an activity on MCF-7 cells, these first results leading to the annotation of two other ergosterol derivatives, together with chaetoglobosin derivatives already reported to inhibit breast cancer cells proliferation, tends to prove our method is quite powerful.

3.3. Performance evaluation of the R FiBiCo script

3.3.1. Estimation of the robustness of the approach

As calculating false positive and false negative rates revealed impossible here on a real case study, because it would require the isolation and testing of all the molecules from the natural crude extract, alternatively, an assessment of the robustness of the approach based on the literature was performed. So far, among the metabolites reported from *P. chrysogenum* species (around 200), only 15 metabolites have been reported to inhibit breast cancer cells proliferation (including MCF-7) (Supplementary information S2.8). Four compounds in this list matched (MS and UV spectra) with features from the peak list (obtained from MMS5 *P. chrysogenum*), namely ergosterol, ergosterol-5,8-endoperoxide, roquefortine C and meleagrins. Additionally, the molecular formula $\text{C}_{32}\text{H}_{36}\text{N}_2\text{O}_5$ (corresponding to chaetoglobosins A, C and G) also appeared to be a good match with one feature at 15.1 min. Its UV spectrum could not be recovered and compared due to low intensity. For the other compounds reported to have antiproliferative activity on breast cancer cells from *P. chrysogenum* species, no matches within the peak list of 795 features were found. This is easily understandable as the conditions used in the present study may have not been appropriate to obtain the biosynthesis of these other metabolites. Among these matches, ergosterol, ergosterol-5,8-endoperoxide and chaetoglobosins corresponding features were effectively highlighted by the R FiBiCo script, while meleagrins and roquefortine C were not. However, the activity of these two latter compounds originating from the same biosynthetic pathway [64], may be questionable as meleagrins has been tested twice on MCF-7 cells with completely different results [65,66]. One reports an IC_{50} of 1.9 μM while the other reports no activity at all. It seems like the MCF-7 cell line used in the present study is not sensitive to this type of molecules. Preliminary results on fractions enriched with meleagrins and roquefortine C (data not shown) tend to prove this absence of inhibition on this cell line.

Given to these different investigations (dereplication on the “Super list” in Table S2.7 and the list of molecules from *P. chrysogenum* reported for their activity on breast cancer cells in

Table S2.8), we can assume the false positive and false negative rates of this method are quite low. Nevertheless, the main purpose of the method described in the present article is to highlight molecules of interest in an extract and orientate further purification work in a more rational way. Even if, there may be false positives, it still helps drastically reducing the number of molecules to be further investigated for their biological activity. Additionally, false positives could correspond in many cases, to molecules presenting the same chromatographic behaviour than the active ones, and then probably to chemically related derivatives, which would also be worth isolating to study structure–activity relationships. However, if no false positive is preferred, one easy option is to focus only on hits returned by the 4 models. In the present case, this would

only return two features only corresponding to ergosterol (See Supplementary information Table S2.7).

3.3.2. Virtual training

While the application of the workflow to a real case proved to be successful in the detection of the main active compound, further evaluation of the FiBiCo script was performed with virtual modifications of the data matrix to verify its ability to detect interesting features and investigate its limits. A virtual feature was then added on the initial MMS5 *P. chrysogenum* data matrix with different peak area values according to proliferation inhibition for each sample. A total of 16 matrices were constructed with one additional feature, which peak area values in the 19 samples, were attributed

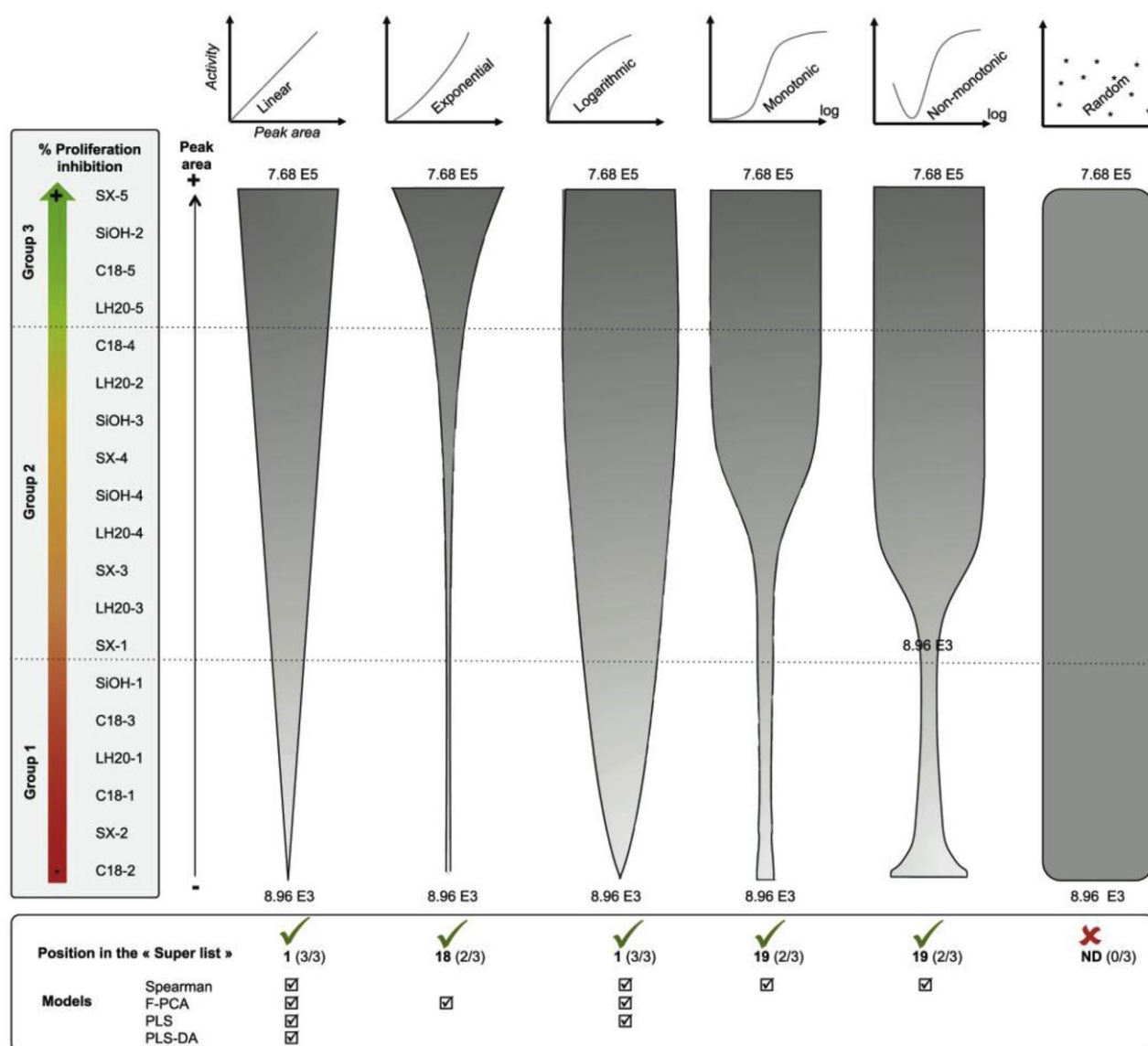


Fig. 7. Virtual evaluation of the FiBiCo script performance with the addition of a virtual feature in the MMS5 *P. chrysogenum* data matrix. The figure presents the conditions tested for the additional feature designing the different relationships between peak area value and biological activity. In accordance with biological activity obtained from the 19 samples, a peak area value was assigned for the additional feature to represent six different relationships: linear, exponential, logarithmic, monotonic, non-monotonic, and random. To be closer to real conditions, a random variation of 30% was applied on the peak area values of the additional feature for each model to obtain a final range from “8.96E3” to “7.68E5”. The R FiBiCo script was applied on the modified initial matrix with this additional feature generating its position in the “Super list” and revealing which models allowed its selection. These experiments were performed in triplicate. Additional conditions tested are presented in Supplementary information S2.9.

according to linear, exponential, logarithmic, monotonic, non-monotonic or random relationships to biological activity. To avoid perfect fit and be closer to real conditions, random variations of 30 and 50% were applied on each peak area value defined for this additional feature in the different matrices. After uploading the matrices on the R FiBiCo script, the results obtained were analysed and reported on a graphical representation (Fig. 7 and Supplementary information S2.9). For each assay, the position of the additional virtual feature in the “Super list” was picked together with the information on which of the biochemometrics models allowed the selection of this feature. Data results obtained with a random variation of 30% performed in triplicate for all relationships (Fig. 7) revealed that R FiBiCo script was able to highlight the additional feature for all meaningful relationships except for the random model, for which no hit was expected. The selection of the additional feature was very good when the relationship was linear or logarithmic, i.e. detected by three or four statistical models (Spearman, F-PCA, PLS and PLS-DA) and with the best global score in the super list (ranking 1st). However, for other activity-to-peak area relationships, some statistical models appeared to be more adapted. For example, for monotonic and non-monotonic relationships, Spearman correlation was the only model able to catch the additional feature. Moreover, while its global score of ranked it in the 19th position, it actually ranked in the two first positions in the Spearman model alone, proving the relevance of Spearman for this type of relationship. Another example is given by the exponential relationship, which was only detected by F-PCA model (in positions 3–5). These results demonstrate the importance to use the combination of the four biochemometrics models to highlight interesting features, because in dose-response curves, different relationships can be observed between the amount of a compound and its biological activity. It is important to note that the previously described experiments were performed on values, which were quite low and in a small range (“8.96E3” – “7.68E5”), with a 30% variation on each. These were chosen to test the model and approach its limits. Other experiments on broader ranges of values (Supplementary information S2.9) showed that increasing the maximal peak area allowed a better detection of the additional feature presenting an interesting relationship with activity. This means that for highly detected compounds, more statistical models may respond. In addition, the smaller the variation of the values (10%), the better the detection as well. Further assays were also performed using groups of biological activity (not the activity for each of the 19 samples). They allowed to confirm the efficiency of the script in the detection, because the additional feature was highlighted only if its peak area values in samples of a group were in accordance with biological activity (i.e. group3>group2>group1 or group2<group3<group1).

3.3.3. Comparison with the NMR pharmacophoric deconvolution approach

The recrudescence of “omics” and the development of new strategies in bioactive NP research allowed to compare our strategy using LC-MS analyses with a recently biochemometrics approach developed by Margueritte et al. [49] using NMR pharmacophoric deconvolution approach. This strategy was based on the use of an NMR fingerprint obtained by automatic differential analysis of 2D NMR data to target and identify bioactive natural product(s) in a complex crude extract. The NMR pharmacophoric fingerprint is obtained through the common cross-peaks in the ^1H – ^1H COSY NMR spectra of the active fractions. Common cross-peaks give partial structural information about the activity-bearing compound because successive fraction obtained by the fractionation should contain shared molecules. Therefore, ^1H – ^1H COSY spectra were recorded for the five C18 fractions obtained from *P. chrysogenum*

crude extract. NMR data was processed by Plasmodesma program. This program written in Python was developed to process automatically NMR data set and provide the best analysis differential of NMR data. To achieve this, Plasmodesma divides NMR data in buckets and measures different variables. Then, the generated bucketlist were used by a Python script on the Jupiter notebook environment. Areas values of buckets were subjected to a cleaning and symmetrization steps to remove a large part of the noise and artefacts. The correlation analysis between spectral features and activity levels was performed with the set of regression tools from the scikit-learn library. A linear regression was applied between areas values and the bioactivity of fractions to produce the pharmacophoric fingerprint (Fig. 8B). Its chemical shifts of correlation peaks suggested a terpenoid skeleton. The overlay of ergosterol spectrum (Fig. 8A) and the pharmacophoric fingerprint (Fig. 8B) allowed to highlight 14 common cross-peaks (Fig. 8C), suggesting ergosterol was the main molecule responsible for the biological activity. This result was in accordance with the previous FiBiCo analysis further confirming its efficiency.

The use of NMR data was of great interest in this presented case, as ergosterol presents an unusual ionization pattern with low signal in ESI-MS, even though it is a major component of the extract. Therefore an analytical method having a more linear link between content and signal should provide an alternate accurate approach.

In fact, while NMR is able to detect all hydrogenated molecules, it suffers from a lack of sensitivity, making minor compounds much

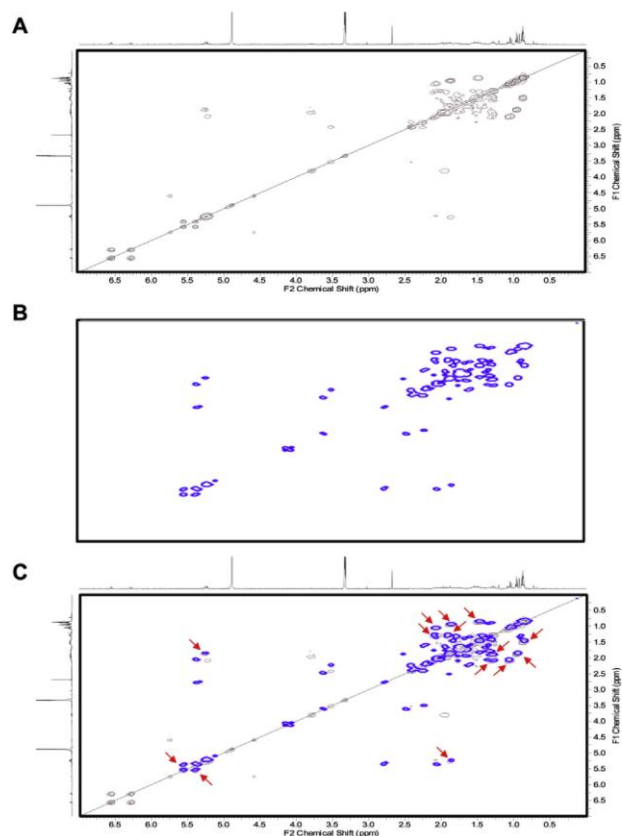


Fig. 8. NMR pharmacophoric approach applied to the five C₁₈ fractions from *P. chrysogenum* MMS5 crude extract. A) ^1H – ^1H COSY NMR spectrum of ergosterol, B) automatic pharmacophoric fingerprint, C) overlay of both with red arrows corresponding to common features. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

more difficult to detect. On the contrary, MS is highly sensitive but is limited by the ability of molecules to ionize depending on the ionization source of the instruments. Therefore, it is important to remember that all strategies reported in the literature have their limits and can be complementary.

4. Conclusion

Currently, the central challenge in NP research is the discovery of new bioactive compounds. With the recrudescence of “omics” studies in this domain, several strategies based on NMR and MS analyses, are currently emerging to highlight and identify bioactive compound from natural complex mixtures at a very early stage. In this study, a complete workflow is proposed, which combines explorative SPE and chemometrics tools using LC-MS data with the integration of four statistical models (Spearman, F-PCA, PLS and PLS-DA) to highlight potential bioactive compounds. Additionally, a script encoded with R software (named FiBiCo) was here developed to automatically highlight the most interesting features by including the complementary results of all statistical analyses, and made available for the scientific community (Supplementary information S1).

This complete workflow applied to a bioactive *P. chrysogenum* MMS5 crude extract, where ergosterol was the main active component, successfully highlighted this compound as responsible for the proliferation inhibition of MCF-7 breast cancer cells. This allowed to prove the efficiency of the developed R FiBiCo script. This is even more interesting as ergosterol was very poorly ionised with electrospray ionization showing that minor compounds can be highlighted, which is essential in bioactive natural product research. The performance of the FiBiCo script was evaluated with virtual modifications of the data matrix, proving its ability to detect different types of relationships between activities and chemical compounds amounts. The complementarity of the statistical models proved to be very important, especially if a low false positive rate is preferred. Moreover, the performance of the script in picking only one feature (among 796), with low values in a small range and with a high variation (30–50%), proved its efficiency to “find the needle in the haystack”. The workflow was also evaluated by comparing the results obtained to those given by another recently described automatic deconvolution strategy using NMR data. The pharmacophoric fingerprint revealed also ergosterol as the bioactive compound, strengthening the validity of the workflow described and developed in the present study.

Perspective of this work will be to apply the methodology to other bioactive extracts, in the context of drug discovery screening programs. It effectively responds to one of the many issues encountered in this field, by rationalizing subsequent purification work to directly focus on compounds responsible for the activity.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Cancéropôle Grand Ouest (CGO), La Ligue Contre le Cancer and the French Ministry of Higher Education and Research (PhD grant). The authors also acknowledge J.M. Huvelin, A. Burghelée and T. Robiou du Pont for their technical participation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.aca.2019.04.038>.

References

- [1] F.E. Koehn, G.T. Carter, The evolving role of natural products in drug discovery, *Nat. Rev. Drug Discov.* 4 (2005) 206–220.
- [2] M.G. Weller, A unifying review of bioassay-guided fractionation, effect-directed analysis and related techniques, *Sensors* 12 (2012) 9181–9209.
- [3] F. Bucar, A. Wube, M. Schmid, Natural product isolation – how to get from biological material to pure compounds, *Nat. Prod. Rep.* 30 (2013) 525–545.
- [4] D.J. Newman, G.M. Cragg, Natural products as sources of new drugs from 1981 to 2014, *J. Nat. Prod.* 79 (2016) 629–661.
- [5] J.W.H. Li, J.C. Vederas, Drug discovery and natural products: end of an era or an endless frontier? *Science* 325 (2009) 161–165.
- [6] Y.F. Pouchus, A.F. Benslimane, J.F. Verbist, SESAME: an expert system for bioassay-directed isolation of active compounds, *Tetrahedron Comput. Methodol.* 2 (1989) 55–64.
- [7] G. Samuelsson, G. Kyerematen, M.H. Farah, Preliminary chemical characterization of pharmacologically active compounds in aqueous plant extracts, *J. Ethnopharmacol.* 14 (1985) 193–201.
- [8] J.H. Cardellina, M.H.G. Munro, R.W. Fuller, K.P. Manfredi, T.C. McKee, M. Tischler, H.R. Bokesch, K.R. Gustafson, J.A. Beutler, M.R. Boyd, A chemical screening strategy for the dereplication and prioritization of HIV-inhibitory aqueous natural products extracts, *J. Nat. Prod.* 56 (1993) 1123–1129.
- [9] M. Månsson, R.K. Phipps, L. Gram, M.H.G. Munro, T.O. Larsen, K.F. Nielsen, Explorative solid-phase extraction (E-SPE) for accelerated microbial natural product discovery, dereplication, and purification, *J. Nat. Prod.* 73 (2010) 1126–1132.
- [10] J.-L. Wolfender, G. Marti, A. Thomas, S. Bertrand, Current approaches and challenges for the metabolite profiling of complex natural extracts, *J. Chromatogr. A* 1382 (2015) 136–164.
- [11] P.-M. Allard, G. Genta-Jouve, J.-L. Wolfender, Deep metabolome annotation in natural products research: towards a virtuous cycle in metabolite identification, *Curr. Opin. Chem. Biol.* 36 (2017) 40–49.
- [12] J.Y. Yang, L.M. Sanchez, C.M. Rath, X. Liu, P.D. Boudreau, N. Bruns, E. Glukhov, A. Wodtke, R. De Felicio, A. Fenner, W.R. Wong, R.G. Linington, L. Zhang, H.M. Debonis, W.H. Gerwick, P.C. Dorrestein, Molecular networking as a dereplication strategy, *J. Nat. Prod.* 76 (2013) 1686–1699.
- [13] J.-L. Wolfender, J.-M. Nuzillard, J.J.J. van der Hooft, J.-H. Renault, S. Bertrand, Accelerating metabolite identification in natural product research: toward an ideal combination of LC-HRMS/MS and NMR profiling, *in silico* databases and chemometrics, *Anal. Chem.* 91 (2019) 704–742.
- [14] D.G. Cox, J. Oh, A. Keasling, K.L. Colson, M.T. Hamann, The utility of metabolomics in natural product and biomarker characterization, *Biochim. Biophys. Acta Gen. Subj.* 1840 (2014) 3460–3474.
- [15] M.D. Luque de Castro, F. Priego-Capote, The analytical process to search for metabolomics biomarkers, *J. Pharm. Biomed. Anal.* 147 (2018) 341–349.
- [16] F. Perrotti, C. Rosa, I. Cicalini, P. Sacchetta, P. Del Boccio, D. Genovesi, D. Pieragostino, Advances in lipidomics for cancer biomarkers discovery, *Int. J. Mol. Sci.* 17 (2016).
- [17] K.M. Lee, J.Y. Jeon, B.J. Lee, H. Lee, H.K. Choi, Application of metabolomics to quality control of natural product derived medicines, *Biomol. Ther.* 25 (2017) 559–568.
- [18] E.O. Olawode, R. Tandlich, C. Garth, 1H-NMR Profiling and chemometric analysis of selected honeys from South Africa, Zambia, and Slovakia, *Molecules* 23 (2018) 2–19.
- [19] S.L. Robinette, R. Brüscheiler, F.C. Schroeder, A.S. Edison, NMR in metabolomics and natural products research: two sides of the same coin, *Acc. Chem. Res.* 45 (2012) 288–297.
- [20] M. Fillet, M. Frédéric, The emergence of metabolomics as a key discipline in the drug discovery process, *Drug Discov. Today Technol.* 13 (2015) 19–24.
- [21] M. Mandrone, A. Coqueiro, F. Poli, F. Antognoni, Y.H. Choi, Identification of a collagenase-inhibiting flavonoid from *Alchemilla vulgaris* using NMR-based metabolomics, *Planta Med.* (2018) 941–946.
- [22] F. Olivon, P.-M. Allard, A. Koval, D. Righi, G. Genta-Jouve, J. Neyts, C. Apel, C. Pannecouque, L.F. Nothias, X. Cachet, L. Marcourt, F. Roussi, V.L. Katanaev, D. Touboul, J.L. Wolfender, M. Litaudon, Bioactive natural products prioritization using massive multi-informational molecular networks, *ACS Chem. Biol.* 12 (2017) 2644–2651.
- [23] T. Pluskal, S. Castillo, A. Villar-Briones, M. Oresic, MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data, *BMC Bioinf.* 11 (2010) 2–11.
- [24] R. R Core Team, A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2018. <https://www.R-project.org/>.
- [25] P.C. Team, Python: a dynamic, open source programming language. <https://www.python.org/>, 2015.
- [26] T. Inui, Y. Wang, S.M. Pro, S.G. Franzblau, G.F. Pauli, Unbiased evaluation of bioactive secondary metabolites in complex matrices, *Fitoterapia* 83 (2012) 1218–1225.

- [27] S. Bertrand, A. Azzollini, A. Nievergelt, J. Boccard, S. Rudaz, M. Cuendet, J.L. Wolfender, Statistical correlations between HPLC activity-based profiling results and NMR/MS microfraction data to deconvolute bioactive compounds in mixtures, *Molecules* 21 (2016) 2–13.
- [28] L.A. Richards, C. Oliveira, L.A. Dyer, A. Rumbaugh, F. Urbano-Muñoz, I.S. Wallace, C.D. Dodson, C.S. Jeffrey, Shedding light on chemically mediated tri-trophic interactions: a 1H-NMR network approach to identify compound structural features and associated biological activity, *Front. Plant Sci.* 9 (2018) 1–12.
- [29] L.-F. Nothias, M. Nothias-Esposito, R. da Silva, M. Wang, I. Protsyuk, Z. Zhang, A. Sarvepalli, P. Leyssen, D. Touboul, J. Costa, J. Paolini, T. Alexandrov, M. Litaudon, P.C. Dorrestein, Bioactivity-based molecular networking for the discovery of drug leads in natural product bioassay-guided fractionation, *J. Nat. Prod.* 81 (2018) 758–767.
- [30] G. D'Urso, C. Pizzi, S. Piacente, P. Montoro, Combination of LC–MS based metabolomics and antioxidant activity for evaluation of bioactive compounds in *Fragaria vesca* leaves from Italy, *J. Pharm. Biomed. Anal.* 150 (2018) 233–240.
- [31] O.M. Kvalheim, H. Yan Chan, I.F.F. Benzie, Y. tong Szeto, A.H. Chung Tzang, D.K. wah Mok, F. Tim Chau, Chromatographic profiling and multivariate analysis for screening and quantifying the contributions from individual components to the bioactive signature in natural products, *Chemometr. Intell. Lab. Syst.* 107 (2011) 98–105.
- [32] J. Xu, Q.S. Xu, C.O. Chan, D.K.W. Mok, L.Z. Yi, F.T. Chau, Identifying bioactive components in natural products through chromatographic fingerprint, *Anal. Chim. Acta* 870 (2015) 45–55.
- [33] J.J. Kellogg, D.A. Todd, J.M. Egan, H.A. Raja, N.H. Oberlies, O.M. Kvalheim, N.B. Cech, Biochemometrics for natural products research: comparison of data analysis approaches and application to identification of bioactive compounds, *J. Nat. Prod.* 79 (2016) 376–386.
- [34] L.K. Caesar, J.J. Kellogg, O.M. Kvalheim, R.A. Cech, N.B. Cech, Integration of biochemometrics and molecular networking to identify antimicrobials in *Angelica keiskei*, *Planta Med.* 84 (2018) 721–728.
- [35] E.R. Britton, J.J. Kellogg, O.M. Kvalheim, N.B. Cech, Biochemometrics to identify synergists and additives from botanical medicines: a case study with *Hydrastis canadensis* (goldenseal), *J. Nat. Prod.* 81 (2018) 484–493.
- [36] G. Xie, M. Ye, Y. Wang, Y. Ni, M. Su, H. Huang, M. Qiu, A. Zhao, X. Zheng, T. Chen, W. Jia, Characterization of pu-erh tea using chemical and metabolic profiling approaches, *J. Agric. Food Chem.* 57 (2009) 3046–3054.
- [37] Y. Fujimura, K. Kurihara, M. Ida, R. Kosaka, D. Miura, H. Wariishi, M. Maeda-Yamamoto, A. Nesumi, T. Saito, T. Kanda, K. Yamada, H. Tachibana, Metabolomics-driven nutraceutical evaluation of diverse green tea cultivars, *PLoS One* 6 (2011) 3–16.
- [38] K.M. Chan, G.G.L. Yue, P. Li, E.C.W. Wong, J.K.M. Lee, E.J. Kennelly, C.B.S. Lau, Screening and analysis of potential anti-tumor components from the stipe of *Ganoderma sinense* using high-performance liquid chromatography/time-of-flight mass spectrometry with multivariate statistical tool, *J. Chromatogr. A* 1487 (2017) 162–167.
- [39] D.M. Kulakowski, S.B. Wu, M.J. Balick, E.J. Kennelly, Merging bioactivity with liquid chromatography-mass spectrometry-based chemometrics to identify minor immunomodulatory compounds from a Micronesian adaptogen, *Phaleria nissidai*, *J. Chromatogr. A* 1364 (2014) 74–82.
- [40] D.A. Chagas-Paula, T. Zhang, F.B. da Costa, R.A. Edrada-Ebel, A metabolomic approach to target compounds from the *Asteraceae* family for dual COX and LOX inhibition, *Metabolites* 5 (2015) 404–430.
- [41] Y. Chen, J. Luo, Q. Zhang, L. Kong, Identification of active substances for dually modulating the renin-angiotensin system in *Bidens pilosa* by liquid chromatography-mass spectrometry-based chemometrics, *J. Funct. Foods* 21 (2016) 201–211.
- [42] A. Patras, N.P. Brunton, G. Downey, A. Rawson, K. Warriner, G. Gernigon, Application of principal component and hierarchical cluster analysis to classify fruits and vegetables commonly consumed in Ireland based on in vitro antioxidant activity, *J. Food Compos. Anal.* 24 (2011) 250–256.
- [43] X. Li, Q. Wu, Y. Xie, Y. Ding, W.W. Du, M. Sdiri, B.B. Yang, Ergosterol purified from medicinal mushroom *Amauroderma rude* inhibits cancer growth in vitro and in vivo by up-regulating multiple tumor suppressors, *Oncotarget* 6 (2015) 17832–17846.
- [44] J. Chong, J. Xia, MetaboAnalystR: an R package for comprehensive analysis of metabolomics data, *Bioinformatics* 34 (2018) 4313–4314.
- [45] B. Falissard, psy: various procedures used in psychology. R package version 1.1. <https://CRAN.R-project.org/package=psy>, 2012.
- [46] S. Kucheryavskiy, mdatools: multivariate data analysis for chemometrics. R package version 0.9.1. <https://CRAN.R-project.org/package=mdatools>, 2018.
- [47] H. Wickham, J. Hester, R. Francois, readr: read rectangular text data. R package version 1.1.1. <https://CRAN.R-project.org/package=readr>, 2017.
- [48] W. Revelle, Psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, 2018. R package version 1.8.10, <https://CRAN.R-project.org/package=psych>.
- [49] L. Margueritte, P. Markov, L. Chiron, J.-P. Starck, C. Vonthron-Sénécheau, M. Bourjot, M.-A. Delsuc, Automatic differential analysis of NMR experiments in complex samples, *Magn. Reson. Chem.* 56 (2018) 469–479.
- [50] L. Chiron, M.-A. Coutouly, J.-P. Starck, C. Rolando, M.-A. Delsuc, SPIKE a Processing Software Dedicated to Fourier Spectroscopies, *ArXiv*, 2016, p. 1608, 06777.
- [51] E. Saccenti, H.C.J. Hoefsloot, A.K. Smilde, J.A. Westerhuis, M.M.W.B. Hendriks, Reflections on univariate and multivariate analysis of metabolomics data, *Metabolomics* 10 (2014) 361–374.
- [52] B. Falissard, Focused principal component analysis: looking at a correlation matrix with a particular interest in a given variable, *J. Comput. Graph. Stat.* 8 (1999) 906–912.
- [53] P.S. Gromski, H. Muhamadali, D.I. Ellis, Y. Xu, E. Correa, M.L. Turner, R. Goodacre, A tutorial review: metabolomics and partial least squares-discriminant analysis – a marriage of convenience or a shotgun wedding, *Anal. Chim. Acta* 879 (2015) 10–23.
- [54] O.M. Kvalheim, T. V. Karstang, Interpretation of latent-variable regression models, *Chemometr. Intell. Lab. Syst.* 7 (1989) 39–51.
- [55] B. Worley, R. Powers, Multivariate analysis in metabolomics, *Curr. Metabolomics* 1 (2015) 92–107.
- [56] Y.H. Yun, B.C. Deng, D.S. Cao, W.T. Wang, Y.Z. Liang, Variable importance analysis based on rank aggregation with applications in metabolomics for biomarker discovery, *Anal. Chim. Acta* 911 (2016) 27–34.
- [57] M. Farrés, S. Platikanov, S. Tsakovski, R. Tauler, Comparison of the variable importance in projection (VIP) and of the selectivity ratio (SR) methods for variable selection and interpretation, *J. Chemom.* 29 (2015) 528–536.
- [58] R.A. Van den Berg, H.C. Hoefsloot, J.A. Westerhuis, A. K. Smilde, M.J. van der Van der Werf, Centering, scaling, and transformations: improving the biological information content of metabolomics data, *BMC Genomics* 7 (2006) 1–15.
- [59] H.-W. Cho, S.B. Kim, M.K. Jeong, Y. Park, N.G. Miller, T.R. Ziegler, D.P. Jones, Discovery of metabolite features for the modelling and analysis of high-resolution NMR spectra, *Int. J. Data Min. Bioinform.* 2 (2008) 176–192.
- [60] V. Prakash, P.K. Nandi, Interaction of amino acids, N-acetyl amino acid esters, thymine and adenine with Sephadex LH-20 gel, *J. Chromatogr.* 106 (1975) 23–31.
- [61] J.A. Westerhuis, H.C.J. Hoefsloot, S. Smit, D.J. Vis, A.K. Smilde, E.J.J. van Velzen, J.P.M. van Duijnoven, F.A. van Dorsten, Assessment of PLS-DA cross validation, *Metabolomics* 4 (2008) 81–89.
- [62] H.Y. Kim, M.Y. Lee, H.M. Park, Y.K. Park, J.C. Shon, K.-H. Liu, C.H. Lee, Urine and serum metabolite profiling of rats fed a high-fat diet and the anti-obesity effects of caffeine consumption, *Molecules* 20 (2015) 3107–3128.
- [63] S. Huang, H. Chen, W. Li, X. Zhu, W. Ding, C. Li, Bioactive chaetoglobosins from the mangrove endophytic fungus *Penicillium chrysogenum*, *Mar. Drugs* 14 (2016) 1–12.
- [64] H. Ali, M. Ries, J. Nijland, P. Lankhorst, T. Hankemeier, R. Bovenberg, R. Vreeken, A. Driessen, A branched biosynthetic pathway is involved in production of roquefortine and related compounds in *Penicillium chrysogenum*, *PLoS One* 8 (2013) 1–12.
- [65] M.S. Mady, M.M. Mohyeldin, H.Y. Ebrahim, H.E. Elsayed, W.E. Houssem, E.G. Haggag, R.F. Soliman, K.A. El Sayed, The indole alkaloid meleagrins, from the olive tree endophytic fungus *Penicillium chrysogenum*, as a novel lead for the control of c-Met-dependent breast cancer proliferation, migration and invasion, *Bioorg. Med. Chem.* 24 (2016) 113–122.
- [66] Z. Shang, X. Li, L. Meng, C. Li, S. Gao, C. Huang, B. Wang, Chemical profile of the secondary metabolites produced by a deep-sea sediment-derived fungus *Penicillium commune* SD-118, *Chin. J. Oceanol. Limnol.* 30 (2012) 305–314.

4. Conclusion

La méthodologie développée et proposée au cours de ce projet de thèse comprenant une étape de microfractionnement exploratoire sur micro colonnes SPE ainsi que la combinaison de quatre modèles statistiques a été validée par son utilisation sur la souche MMS5 *P. chrysogenum*. En effet, selon une approche de biochimométrie, les analyses statistiques réalisées à l'aide du « R FiBiCo » script sur la matrice de données comprenant les aires sous pics de chacun des ions détectés au sein des profils LC-MS et le pourcentage d'inhibition de chacune des fractions obtenues au cours du microfractionnement, a permis une nouvelle fois de mettre en évidence l'activité antiproliférative de l'ergostérol. Ce qui est en accord avec les résultats obtenus, puisque les différentes investigations menées jusqu'ici au cours du fractionnement bioguidé classique ont conduit à l'isolement de cette molécule.

Le microfractionnement de l'extrait brut MMS5-530 réalisé sur quatre colonnes de phases différentes (Silice, C₁₈, StrataX et LH20) a permis d'évaluer la présence potentielle d'effets synergiques. En effet, si les fractions obtenues sont moins actives que l'extrait de départ cela signifie que l'activité résulte probablement de plusieurs molécules. De même, si une fraction est plus active, alors la concentration de molécule bioactive au sein de cette fraction permet d'accroître son activité au cours des essais biologiques. Dans notre cas, aucune fraction ne présentait une inhibition de la prolifération supérieure à celle de l'extrait brut de départ (60%). Néanmoins, il n'a été démontré aucune « perte » d'activité au cours du fractionnement puisque l'extrait brut reconstitué (mélange de chacune des fractions pour une même colonne) présentait une activité similaire à celle de l'extrait brut. De plus, les caractéristiques physicochimiques des molécules bioactives ont pu être mises en évidence. Dans notre cas, les composés responsables de l'activité sont localisés en fin d'élution pour les colonnes C18, LH20 et StrataX, contrairement à la colonne de silice où l'activité a pu être repérée dans les premières fractions. Par conséquent, cette étape de microfractionnement proposé dans ce workflow constitue une première étape indispensable de priorisation dans la recherche de molécules bioactives. De plus, elle permet de démultiplier le nombre d'occurrences des mêmes molécules dans les fractions et ainsi d'améliorer la significativité des corrélations statistiques.

Enfin, la recherche des molécules bioactives réalisée à l'aide du script R FiBiCo sur la matrice de données contenant les aires sous pic de chacun des ions détectés associés à leur temps de rétention (m/z_{R_t}) dans les fractions ainsi que le pourcentage d'inhibition de la prolifération cellulaire des cellules MCF-7 a permis de mettre en évidence la complémentarité des différents modèles sélectionnés. Tout d'abord ; il été relevé par l'analyse en composante principale (ACP) que le fractionnement de l'extrait MMS5-530 sur la colonne de silice a conduit à des fractions dont la composition chimique était bien différente de celles obtenues pour les autres colonnes.

Afin d'identifier les ions responsables de l'activité, quatre modèles statistiques ont été utilisés : Spearman, ACP-F, PLS et PLS-DA. L'interprétation des analyses statistiques n'étant pas aisée, le script a été rédigé de manière à générer des figures présentant un classement de chacun des ions présents au sein des fractions selon leur niveau d'importance, avec en bleu ceux potentiellement responsables de l'activité observée. Mais le plus important pour l'interprétation des résultats est basé sur la « super liste » et la représentation en Diagram de Venn générées par le script. En effet, le script permet, si les différents critères sont respectés pour chacune des analyses statistiques d'obtenir une liste d'ions classés de manière décroissante selon leurs scores obtenus pour la combinaison des quatre modèles. Après un travail de déréplication dans le but d'évaluer l'originalité chimique des molécules correspondantes à ces ions, la purification peut alors être ciblée sur celles dont l'activité n'est pas référencée dans la littérature ou pour lesquelles aucune correspondance n'a pu être déterminée. Cette purification est alors réalisée selon une approche de fractionnement « masse –guidé ».

Pour compléter ce travail, il serait intéressant de l'appliquer à d'autres extraits notamment *Didymella americana* ou ceux de l'extractothèque ayant démontré une activité antiproliférative au cours du criblage initial. Il serait également pertinent de l'utiliser pour cibler des activités biologiques différentes.

Afin d'obtenir une meilleure standardisation de l'étape du microfractionnement, des optimisations pourraient être apportées sur les conditions d'élutions choisies, de même qu'il pourrait être utilisé d'autres colonnes afin d'accroître les différences de compositions chimiques des fractions obtenues.

Au regard des outils développés actuellement au sein de la communauté des PN, ce script pourrait tout à fait être utilisé au cours de projets basés sur une autre approche bioinformatique, celle des réseaux moléculaires. C'est une approche permettant d'organiser et de visualiser des données MS² sous la forme d'une carte de similarité spectrale, mettant en lumière l'existence de groupes spectraux ainsi que leurs degrés de similarité. Ces groupes (ou « clusters ») correspondant à des profils de fragmentation MS similaires sont ainsi attribués à des familles chimiques proches et permettent par similarité à des molécules connues de propager l'annotation et d'identifier plus facilement des molécules. Ainsi, en incluant les scores obtenus par le script pour chacun des ions détectés au cours des analyses LC-MS/MS dans les réseaux moléculaires, cela permettrait d'obtenir une meilleure visualisation des résultats et faciliterait également la déréplication pour la priorisation des extraits. Actuellement, des équipes ont développé ce concept mais en ajoutant les résultats obtenus uniquement par le coefficient de corrélation de Pearson (Britton *et al.*, 2018).

Chapitre 4 :

Etude de l'ergostérol sur des modèles
de cellules du cancer du sein

Dans la recherche de molécules ayant un effet antiprolifératif sur les lignées cancéreuses mammaires, les différentes stratégies réalisées sur la souche *Penicillium chrysogenum* ont toutes menées à l'identification d'une même molécule : l'ergostérol. Cette molécule est essentielle à la composition et au maintien de l'intégrité des membranes des organismes fongiques, comme peut l'être le cholestérol chez l'homme. C'est également un précurseur de la Vitamine D2, qui sous l'effet des UV joue un rôle important dans notre organisme. Bien que cette molécule fut isolée pour la première fois en 1932, son rôle en santé n'est que très peu étudié et connu, comparé à celui d'autres stérols issus de plantes.

Ainsi, l'ergostérol a été rapporté pour la première fois pour son activité anti-inflammatoire et anti-tumorale in vivo en 1994 (Yasukawa *et al.*, 1994). Depuis, des études réalisées in vitro et in vivo en 2015, ont permis de mettre en évidence l'effet de cette molécule sur l'inhibition de la croissance de plusieurs lignées cancéreuses mammaires, dont les lignées MCF-7 (oestrogéno-dépendante) et MDA-MB-231 (oestrogéno-indépendante). Des investigations plus poussées menées sur la lignée MDA-MB-231 ont montré que cet effet antiprolifératif pouvait être en lien avec l'augmentation de l'expression de gènes suppresseurs de tumeurs impliqués dans les mécanismes de survie cellulaire et de l'apoptose.

Dans la recherche sur le cancer, l'étude des différents mécanismes impliqués dans le développement et la progression du cancer a permis de mettre en évidence plusieurs caractéristiques dont l'altération du métabolisme lipidique (Hanahan and Weinberg, 2011; Santos and Schulze, 2012). En effet, il est connu que les cellules cancéreuses présentent une dérégulation de l'homéostasie lipidique, notamment de la lipogenèse *de novo* et de la biosynthèse du cholestérol, ce qui en résulte par une accumulation des lipides (cholestérol et triglycérides (TG)) nécessaires au maintien de leur taux de croissance élevé. Par conséquent, l'ergostérol ayant démontré un effet antiprolifératif des lignées cancéreuses mammaires et appartenant à la classe chimique des stérols, il semble tout à fait légitime d'étudier l'effet de ce métabolite sur la régulation de ce métabolisme lipidique.

Une première partie bibliographique présentant la relation métabolisme lipidique et cancer du sein a été réalisée. Puis, les travaux menés sur l'effet de l'ergostérol sur la régulation du métabolisme lipidique des cellules cancéreuses mammaires, ont été présentés sous forme d'une publication en cours de préparation.

1. Introduction bibliographique

Le cancer du sein est une maladie d'origine chronique et multifactorielle, dont certains facteurs de risques en lien avec le métabolisme lipidique tels que, l'obésité, le surpoids, le mode de vie, ainsi que l'alimentation ont pu être mis en évidence (McDonnell *et al.*, 2014). Il a également été montré que les facteurs hormonaux et reproductifs pouvaient jouer un rôle important (puberté précoce avant 12 ans, ménopause tardive après 55 ans, première grossesse après 30 ans ou allaitement) (Torre *et al.*, 2015). En effet, il a été reporté que le niveau circulant des récepteurs aux œstrogènes (ER) ainsi que le taux d'œstrogènes sont associés à des facteurs de risques du cancer du sein (Travis and Key, 2003).

L'altération du métabolisme des lipides liée à la prolifération excessive et incontrôlée des cellules cancéreuses est l'une des caractéristiques dans le cancer du sein. En effet, les cellules nécessitent une quantité importante de lipides pour générer les membranes cellulaires et assurer leurs besoins énergétiques élevés (Baenke *et al.*, 2013; Mounier *et al.*, 2014). Dans ce contexte, l'étude de l'altération du métabolisme lipidique en tant que potentiel facteur de risque et médiateur du cancer du sein apparait comme une cible thérapeutique d'un grand intérêt. Par conséquent, une brève introduction bibliographique a été réalisée dans le but de rappeler les mécanismes impliqués dans la régulation de l'homéostasie lipidique et notamment celle du cholestérol, avant d'exposer la relation pouvant exister entre le dérèglement du métabolisme lipidique et le risque de développement du cancer du sein.

1.1. Mécanismes impliqués dans la régulation du métabolisme lipidique

Les mécanismes impliqués dans la régulation de l'homéostasie de la lipogenèse (synthèse de *novo* et bêta-oxydation des acides gras) ainsi que celle du cholestérol (synthèse, efflux, captation) sont présentés (Cf. Figure 48).

1.1.1. Métabolisme des acides gras

Les différentes voies de régulation du métabolisme des acides gras à savoir la synthèse, la bêta-oxydation, le stockage et la libération des acides gras stockés sont présentées ci-dessous afin de mettre en évidence les éléments clés de la régulation.

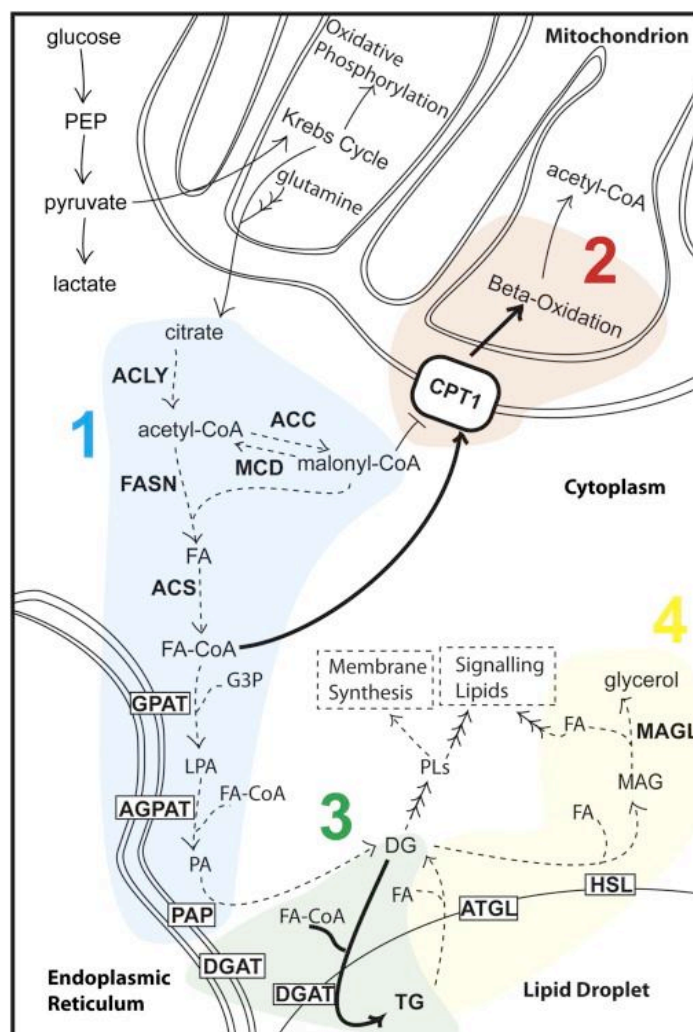


Figure 48. Régulation du métabolisme des acides gras (Currie *et al.*, 2013). (1) synthèse des acides gras (AG), (2) Bêta-oxydation des AG et (3) Stockage des acides gras sous forme de tryglycérides.

1.1.1.1. Synthèse des acides gras

La biosynthèse des acides gras encore appelée lipogenèse *de novo* correspond aux différentes réactions enzymatiques permettant de convertir l'acétyl-CoA en acide gras. Pour ce faire la réaction est initiée par la carboxylation de l'acétyl-CoA en malonyl-CoA par l'acétyl-CoA carboxylase (ACC), suivie par une série de réactions de condensation des acyl-CoA et du malonyl-CoA réalisée par l'enzyme Fatty Acid Synthase (FASN ou FAS). La désaturation des acides gras nouvellement synthétisés est catalysée par la Stéaroyl-CoA Désaturase 1 (SCD) au niveau de la membrane du réticulum endoplasmique (Currie *et al.*, 2013). Une fois les acides gras synthétisés, ces derniers peuvent être stockés sous forme de triglycérides au sein de gouttelettes lipidiques pour être utilisés comme source d'énergie, ou de phospholipides (PLs) pour la synthèse membranaire (Abramson, 2011; Mounier *et al.*, 2014).

1.1.1.2. Stockage des acides gras

La principale voie de synthèse de Triglycérides (TG) est connue sous le nom de voie Kennedy ou glycérol-phosphate. Les TG sont synthétisés à partir d'AGs et de glycérol-3-phosphate (GP3) ou de monoacylglycérol (MAG) au niveau de la membrane du réticulum endoplasmique. Dans la voie où le MAG représente le point de départ, un acyl-CoA est rajouté à ce dernier par la MAG transférase (MAGT). Le produit ainsi formé est le diacylglycérol (DAG) qui sera acylé pour former les TGs. En revanche, dans les adipocytes, la principale voie de synthèse est médiée par le GP3. Ainsi, celui-ci par l'action séquentielle de différentes enzymes : la glycérol-3-phosphate acyltransférase (GPAT), l'acylglycérolphosphate acyltransférase (AGPAT), l'acide phosphatidique phosphorylase (LIPIN ou PAP) permet d'obtenir le DAG qui est lui aussi acylé par la diacylglycérol acyltransférase (DGAT) pour former une molécule de TG.

Ainsi, il a été décrit que les acides gras issus de cette lipolyse jouent un rôle important dans les fonctions cellulaires, telle que la prolifération des cellules cancéreuses (Currie *et al.*, 2013). Ce sont également des précurseurs des glycérophospholipides qui représentent des composants importants des membranes plasmiques.

1.1.1.3. Bêta-oxydation des acides gras

La β -oxydation des acides gras est un processus principalement réalisé dans la matrice mitochondriale mais qui peut également en fonction de la chaîne carbonée, débutée dans les peroxyosomes. Cette dégradation permet d'une part de produire de l'acétyl-CoA mais également le NADH₂ et le FADH₂ nécessaires à la chaîne respiratoire (Michel Houten and A Wanders).

Les acides gras ne pouvant être métabolisés directement sous forme d'acide carboxylique, ils sont tout d'abord activés sous forme d'acyl-CoA par les acyl-CoA synthétase (ACS) dans le cytoplasme. Les acides gras à chaînes longues (C>12) ne pouvant traverser la membrane mitochondriale sont ensuite convertis en acyl-carnitine, au sein de l'espace intermembranaire par la Carnitine Palmityltransférase 1 (CPT1). Cet acyl-carnitine est ensuite transporté par une protéine transmembranaire appelée la cartinitine acylcarnitine translocase (CACT) vers la matrice mitochondriale. Enfin, la Carnitine Palmityltransférase II (CPT2), située du côté de la matrice de la membrane interne de la mitochondrie permet la reconversion de l'acyl-carnitine en acyl-CoA. (Longo *et al.*, 2016)

Une fois l'acyl-CoA formé dans la matrice mitochondriale sa dégradation est initiée par une réaction d'oxydation réalisée par l'acyl-CoA déshydrogénase (ACADVL ou ACADML) en présence du coenzyme FAD (Flavine Adénine Dinucléotide). Cette première étape permet la formation du déhydroacyl-CoA.

Ce dernier est ensuite hydraté par l'énoyl-CoA hydratase pour former le L-3-hydroxyacyl-CoA, puis convertie en 3-cétoacyl-CoA par la L- β -hydroxyacyl-CoA déshydrogénase (HADHA) en présence de NAD (Nicotinamide Adénine Dinucléotide).

La dernière réaction de la β -oxydation est catalysée par la β -cétotliolase, encore appelée Acétyl-coenzyme A acétyltransférase permettant de libérer de l'acétyl-CoA (source d'énergie importante) et d'obtenir l'acyl-CoA comprenant deux carbones de moins que l'acide gras dérivé de départ. Ainsi, cette nouvelle molécule d'acyl-CoA est prête à subir un nouveau cycle de β -oxydation.

1.1.1.4. Libération des acides gras à partir des triglycérides

Une fois que les acides gras sont stockés au sein des gouttelettes lipidiques sous formes d'esters, les Triglycérides (TG), différentes enzymes peuvent permettre leurs libérations. Ainsi, par l'action séquentielle de différentes réactions d'hydrolyses assurées par la triglycéride lipase adipeuse (TGLA), la lipase hormonosensible (LHS) et la monoacylglycérol lipase (MAGL). Ce processus permet de libérer à partir de ces réserves, un pool d'acides gras important pouvant jouer un rôle de précurseur dans la synthèse de lipides de signalisation (Wymann and Schneiter, 2008).

1.1.2. Métabolisme du cholestérol

Toutes les cellules, y compris les cellules du tissu mammaire, sont capables de synthétiser le cholestérol par la voie du mévalonate par une cascade enzymatique dans laquelle l'Hydroxyméthylglutaryl-CoA-Réductase (HMGCR) joue un rôle central. Les cellules peuvent également acquérir du cholestérol par la voie des lipoprotéines. En effet, les lipoprotéines circulantes interviennent dans le transport et la délivrance du cholestérol (provenant de l'alimentation et de la biosynthèse) aux cellules par l'intermédiaire du récepteur aux lipoprotéines de faible densité (LDLR).

1.1.2.1. La synthèse endogène du cholestérol

La biosynthèse du cholestérol (Cf. Figure 49) est initiée par la formation d'un Acéto-AcétylCoA qui est ensuite condensé avec l'acétyl-CoA pour former le 3-Hydroxy-3-Méthyl-Glutaryl-Coenzyme A (HMGCoA). Ce dernier est réduit en mévalonate par l'intermédiaire d'une enzyme transmembranaire du réticulum endoplasmique : l'HMGCR. La double phosphorylation ainsi que la décarboxylation du mévalonate conduisent à la formation de l'unité isoprényle qui est à l'origine de la synthèse de tous les isoprénoïdes : l'Isopentényl-Pyrophosphate (IPP) et de sa forme isomérique : le diméthyl-allyl pyrophosphate (DMAPP). Puis des réactions successives de polymérisations des isoprénoïdes conduisent tout d'abord par la condensation de l'IPP et du DMAPP au géranyl-pyrophosphate (GPP), auquel l'ajout d'un IPP permet d'obtenir le farnésyl pyrophosphate (FPP). Enfin, la condensation de deux FPP permettant de former le squalène, qui est ensuite oxydé puis cyclisé permet d'obtenir un stérol : le lanostérol.

A partir de celui-ci; différentes réactions d'oxydation et de réduction permettent la formation de deux précurseurs du cholestérol : le desmostérol ou de 7-déhydrocholestérol (Ačimovič and Rozman, 2013). Cette voie de biosynthèse est autorégulée puisqu'une accumulation de cholestérol entraîne une inhibition de l' HMG-CoA réductase (Cortes *et al.*, 2014).

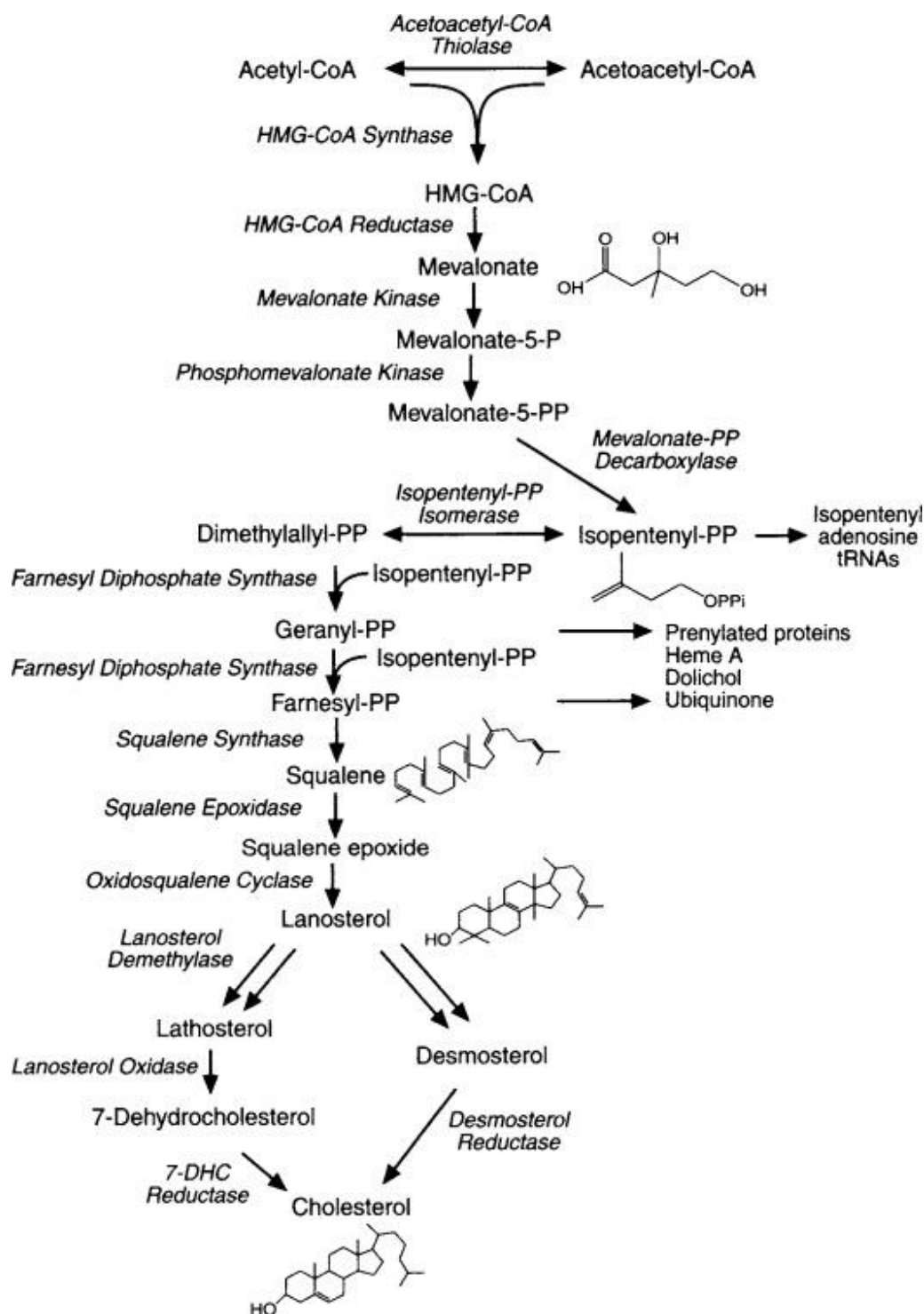


Figure 49. Voie de biosynthèse du cholestérol (Liscum, 2002).

1.1.2.2. La captation du cholestérol via les récepteurs LDLR (Low Density Lipoprotein Receptor)

Le récepteur LDLR est une glycoprotéine membranaire (160 kDa) qui joue un rôle dans la liaison et l'internalisation des LDL circulantes contenant du cholestérol. Ce récepteur est exprimé dans toutes les cellules et constitue un élément de régulation important dans l'homéostasie du cholestérol cellulaire (Goldstein and Brown, 2009).

Le LDLR est synthétisé au niveau des ribosomes du RER (réticulum endoplasmique rugueux), transporté dans l'appareil de Golgi et inséré dans la membrane cellulaire. L'apolipoprotéine B100 des LDL reconnaît le récepteur LDLR. Le complexe formé est ensuite internalisé dans les cellules dans des vésicules entourées de clathrine puis converti en endosome. Ainsi, par une séparation du ligand et du récepteur, le récepteur LDLR est recyclé à la surface de la cellule et les endosomes se combinent aux lysosomes pour dégrader les LDL et libérer le cholestérol sous forme libre.

Le cholestérol, lorsqu'il n'est pas utilisé par la cellule est stocké sous sa forme estérifiée par l'enzyme ACAT (Acyl Coa Cholesterol Acyl-Transférase) (Cf. Figure 50). Un excès du cholestérol intracellulaire entraîne, une répression de l'enzyme HMG-CoA réductase et une inhibition de la synthèse des récepteurs LDLR via la voie de signalisation SREBP.

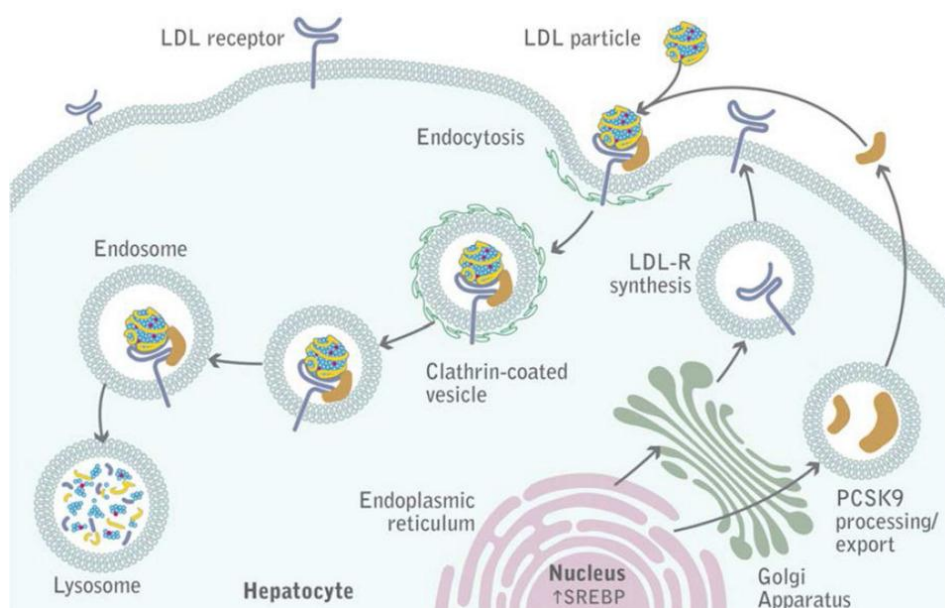


Figure 50. Captation du cholestérol par les récepteurs de lipoprotéines de faible densité (LDL-R) (Stoekenbroek *et al.*, 2015).

En plus des gènes mentionnés ci-dessus, les récepteurs nucléaires Liver X Receptors (LXRs) sont également impliqués dans le maintien de l'homéostasie intracellulaire du cholestérol en contrôlant son efflux et/ou sa captation (voie du LDLR).

1.1.2.3. L'efflux du cholestérol

L'efflux du cholestérol est assuré par des protéines transmembranaires appartenant à la famille des ABC (ATP Binding Casette transporters) et des récepteurs SR-BI (Scavenger Receptor class B type I).

Les protéines ABC divisées en sept sous familles A-G, sont chargées de transporter divers substrats physiologiques tel que les vitamines, les acides aminés ainsi que les stérols à travers la membrane cellulaire (Xiong *et al.*, 2018). Ainsi, les transporteurs ABCA1 et ABCG1 assurent l'efflux de cholestérol cellulaire vers les accepteurs (Cf. Figure 51). Le transporteur ABCA1 est impliqué dans l'efflux de cholestérol des cellules vers des HDL pauvres en lipides, et ABCG1 assure l'efflux du cholestérol vers les HDL matures riches en cholestérol. (Argmann *et al.*, 2005; Gelissen *et al.*, 2006).

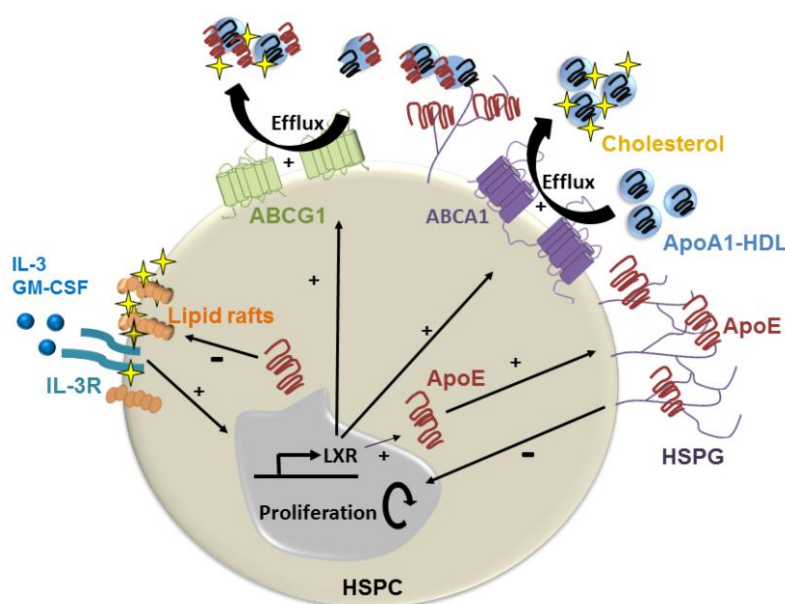


Figure 51. Schéma de l'efflux (Bouchareychas and Raffai, 2018)

1.2. Métabolisme lipidique sous le contrôle des facteurs de transcription

L'ensemble de ces processus permettant de maintenir l'homéostasie lipidique et notamment du cholestérol sont finement régulés par différents modulateurs appartenant à la super-famille des facteurs de transcriptions. Parmi eux, la protéine régulatrice SREBP (Sterol Regulatory Element-Binding Protein) ainsi que deux récepteurs nucléaires : PPAR (Peroxisome Proliferator-Activated Receptor) et LXR (Liver X Receptor).

1.2.1. Le facteur de transcription SREBP (Sterol Regulatory Element-Binding Protein)

Le métabolisme cellulaire des lipides peut être régulé par les facteurs de transcriptions SREBPs (Sterol Regulatory Element-Binding Protein)(Mounier *et al.*, 2014; Santos and Schulze, 2012). Ces facteurs de transcription liés au réticulum endoplasmique existent sous trois isoformes SREBP-1a, SREBP-1c et SREBP-2, dont les rôles sont différents au sein de la cellule. En effet, SREBP-1a contrôle la synthèse des acides gras et du cholestérol ; SREBP-1c régule principalement la synthèse des acides gras ; alors que SREBP-2, régule la synthèse du cholestérol (Eberlé *et al.*, 2004).

Les SREBPs sont synthétisés sous forme de précurseurs inactifs qui interagissent avec les protéines transmembranaires polytopiques SCAP (SREBP Cleavage-Activating Protein), qui se lie elle-même aux protéines INSIG (INSulin-Induced Gene). Le complexe INSIG/SCAP/SREBP (Cf. Figure 52) résultant est conservé dans le RE (réticulum endoplasmique). La dissociation du complexe SCAP/SREBP de la protéine INSIG sous l'effet d'une faible concentration de stérols intracellulaire, suivi d'un changement de conformation de la protéine SCAP, permet une translocation du complexe SCAP/SREBP vers l'appareil de Golgi. Puis, SREBP clivé par deux protéases pénètre dans le noyau sous sa forme active et assure ainsi la transcription des gènes cibles.

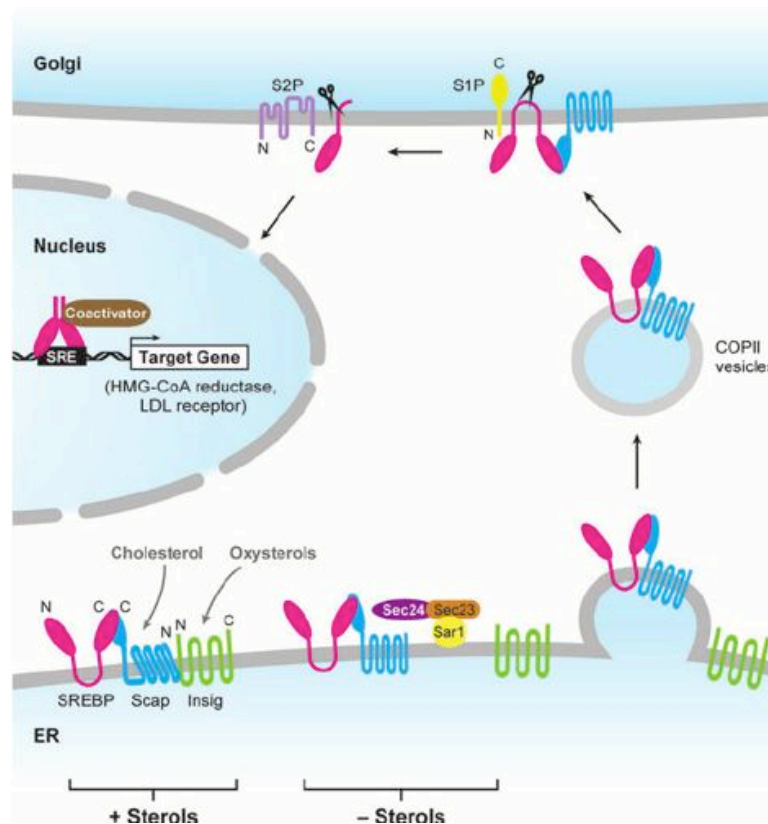


Figure 52. Régulation de l'homéostasie lipidique via le complexe INSIGH/SCAP/SREBP (Espenshade and Hughes, 2007)

Ces derniers codent pour des enzymes clés de la synthèse lipidique telles que ACC, FASN, SCD1, DGAT et l'HMGCR ; mais également des récepteurs LDLR (Baenke *et al.*, 2013; Espenshade and Hughes, 2007; Yoon *et al.*, 2007). Ce processus est autorégulé de manière négative par des stérols qui se lient à la protéine SCAP ou INSIG, favorisant leurs interactions et la formation du complexe INSIG/SCAP/SREBP, ce qui empêche la translocation de SCAP/SREBP vers Golgi et entraîne une réduction de l'activation de SREBP et donc une diminution de la biosynthèse des acides gras et de la captation du cholestérol.

1.2.2. Les récepteurs nucléaires PPAR (Peroxisome Proliferator-Activated Receptor)

Les récepteurs PPAR existent sous trois types : PPAR α , PPAR β/δ et PPAR γ respectivement codés par des gènes distincts (NR1C1, NR1C2, NR1C3) et dont les distributions dans l'organisme diffèrent. Le récepteur PPAR α est principalement exprimé au niveau du foie, des reins et du tissu adipeux ; le récepteur β est ubiquitaire ; et l'isotype γ est exprimé principalement au niveau du tissu adipeux. La stimulation de l'activité transcriptionnelle des récepteurs PPAR dépend de leurs hétérodimérisations avec RXR (Retinoid X Receptor), le partenaire obligatoire de ces récepteurs.

L'hétérodimère PPAR/RXR se lie au niveau de séquences particulières du promoteur de ces gènes cibles, appelé PPRE (éléments de réponse à PPAR) En l'absence de ligands, PPAR est acétylé et se lie aux corépresseurs réprimant ainsi l'expression des gènes cibles. Alors qu'en présence de ligands de PPAR (exemple : les acides gras oméga 3) ou des ligands de RXR (exemple : l'acide 9-cis-rétinoïque), les corépresseurs sont libérés et remplacés par des coactivateurs, induisant ainsi l'expression des gènes cibles (Grygiel-Górniak, 2014).

Les ligands naturels et synthétiques des différents récepteurs de PPAR figurent ci-dessous (Cf. Figure 53).

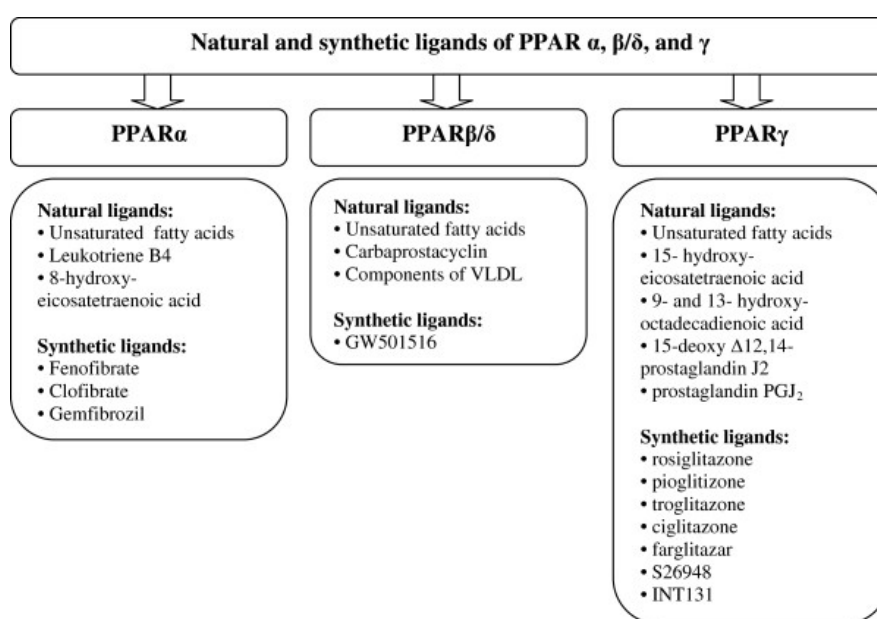


Figure 53. Ligands naturels et synthétiques de PPAR (Grygiel-Górniak, 2014)

1.2.3. Rôle des récepteurs nucléaires PPAR

Les récepteurs PPAR jouent un rôle crucial dans la régulation du métabolisme lipidique (Cf. Figure 54) et particulièrement celle de la synthèse, dégradation et transport des acides gras (Botta *et al.*, 2018; Varga *et al.*, 2011).

En effet, l'activation du récepteur PPAR α permet de réduire la synthèse des acides gras en réprimant l'expression de l'acyl-CoA synthétase nécessaire à la formation des AG-CoA, tout en favorisant leur dégradation par la régulation de l'expression de l' α -carnitine palmitoyltransférase-I et II (CPT1 et CPT2) (Botta *et al.*, 2018). Il permet également de réguler les enzymes impliquées dans la bêta-oxydation tels que l'expression de ACADVL et HADHA.

Parmi les ligands de PPAR α , les fibrates, tels que le fénofibrate ou le gemfibrozil. L'activation de PPAR α par ces molécules est connue pour son effet hypolipidémiant dans les traitements d'hypertriglycéridémie et d'hypoHDLémie, Ce qui se traduit par une diminution des concentrations plasmatiques de triglycérides (TG) et une augmentation des niveaux de cholestérol-HDL (Grygiel-Górniak, 2014).

Les récepteurs PPAR β/δ sont activés par les acides gras saturés et polyinsaturés ainsi que les eicosanoïdes. Parmi les ligands synthétiques qui ont été développés il est possible de nommer les dérivés phénoxyacétiques GW501516 et GW0742. Bien qu'aucunes molécules ne soient approuvées par la FDA, l'activation de PPAR β/δ a été identifiée comme pouvant améliorer le transport et l'oxydation des acides gras tout en diminuant l'inflammation au sein des macrophages. Ainsi, il a été suggéré que ces récepteurs constituent une cible pharmacologique potentielle dans le traitement de la dyslipidémie, de l'obésité et ou de la résistance à l'insuline (Cheang *et al.*, 2015).

Les ligands endogènes des récepteurs PPAR γ sont des composés oxydés dérivant des LDL tel que les acides 9-hydroxy et 13-hydroxy octadécadiénoïques. Ces récepteurs jouant un rôle primordial dans la régulation de l'adipogenèse ont également été identifiés comme étant la cible de molécules synthétiques appartenant à la classe chimique des thiazolinedones (TZDs) (Botta *et al.*, 2018).

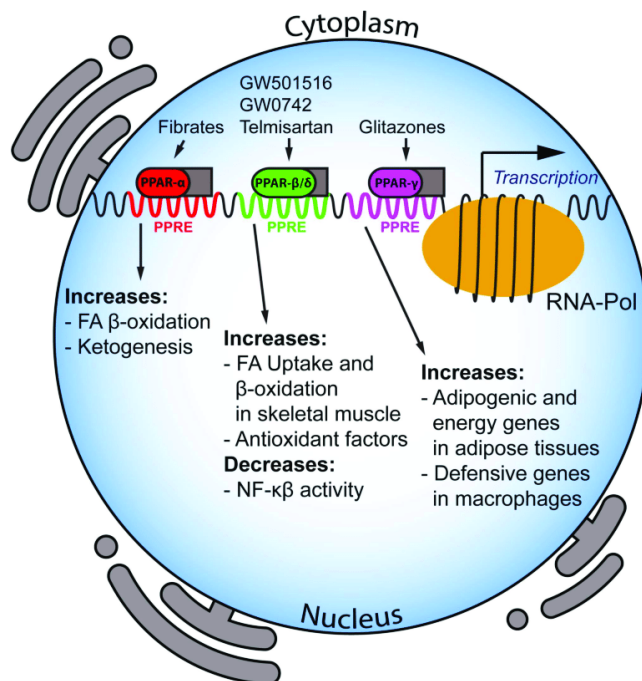


Figure 54. Les Rôles majeurs des différents facteurs de transcriptions nucléaires PPARs (peroxisome proliferator-activated receptor) (Botta *et al.*, 2018). Les facteurs de transcription nucléaires PPAR forment des hétérodimères avec le récepteur RXR (retinoid X receptor apparaissant en gris sur le schéma) dont l'activation par les acides gras ou des ligands synthétiques (ex : fibrates, glitazones etc.) permet de réguler les différentes voies de signalisations tel que celles des acides gras (FA pour Fatty acid) ou celles de l'inflammation NFκB (Nuclear Factor-κB).

1.2.4. Les récepteurs nucléaires LXR (Liver X Receptor)

Le récepteur nucléaire LXR (Liver X Receptor) existe sous deux isoformes : LXRα et LXRβ, dont les séquences sont homologues mais avec des distributions tissulaires différentes. En effet LXRα est exprimé dans les organes présentant un fort métabolisme lipidique tels que le foie, le tissu adipeux, l'intestin, tandis que LXRβ est exprimé de manière ubiquitaire. La stimulation de l'activité transcriptionnelle de LXR dépend également de leur hétérodimérisation avec RXR (Retinoid X Receptor). L'hétérodimère LXR/RXR se lie au niveau de séquences particulières du promoteur de ces gènes cibles, appelé LXRE (éléments de réponse à LXRE) dont le mécanisme d'action moléculaire (Cf. Figure 55) est le même que celui des récepteurs nucléaires PPAR (Wang and Tontonoz, 2018).

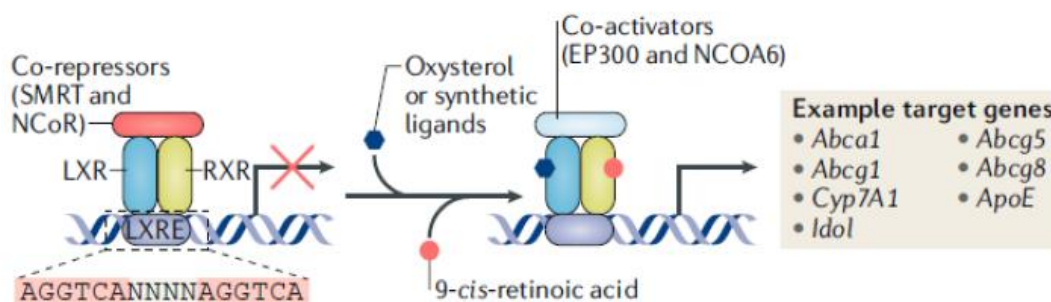


Figure 55. Mécanisme d'action moléculaire de LXR (Wang and Tontonoz, 2018).

1.2.4.1. Les ligands de LXR

La plupart des agonistes naturels des LXRs sont des oxystérols dérivés du cholestérol. (Cf. Figure 56). Ces molécules peuvent avoir deux origines : par voie enzymatique et par voie non-enzymatique. Les oxystérols présentant une forte affinité pour LXR sont les 22(R)-hydroxy-cholestérol, 24(S)-hydroxy-cholestérol, 7 α -hydroxy-cholestérol et 27-hydroxy-cholestérol (Janowski *et al.*, 1996), 24-céto-cholestérol, 25-hydroxy-cholestérol et 24(S),25-époxy-cholestérol. En plus des oxystérols, d'autres molécules dérivées du cholestérol ont été reportées comme des ligands de LXR, tel que le desmostérol, capable d'entrer en compétition avec le ligand 24(S),25-époxycholestérol; ou encore le FF-MAS (follicular fluid meiosis-activating sterol) reporté pour son activité sur LXR α (Janowski *et al.*, 1996). Ils existent également d'autres ligands isolés à partir de plantes ou d'organismes fongiques tels que des stérols : le beta sitostérol, le campestérol, le fucostérol, le sitostanol, le campestanol et l'ergostérol ou le (22E)-ergost-22-ene-1 α ,3 β -diol (produit d'hémisynthèse) ; ou encore des terpènes tels que: l'acide acanthoïc, polycarpol, viperidone (Komati *et al.*, 2017).

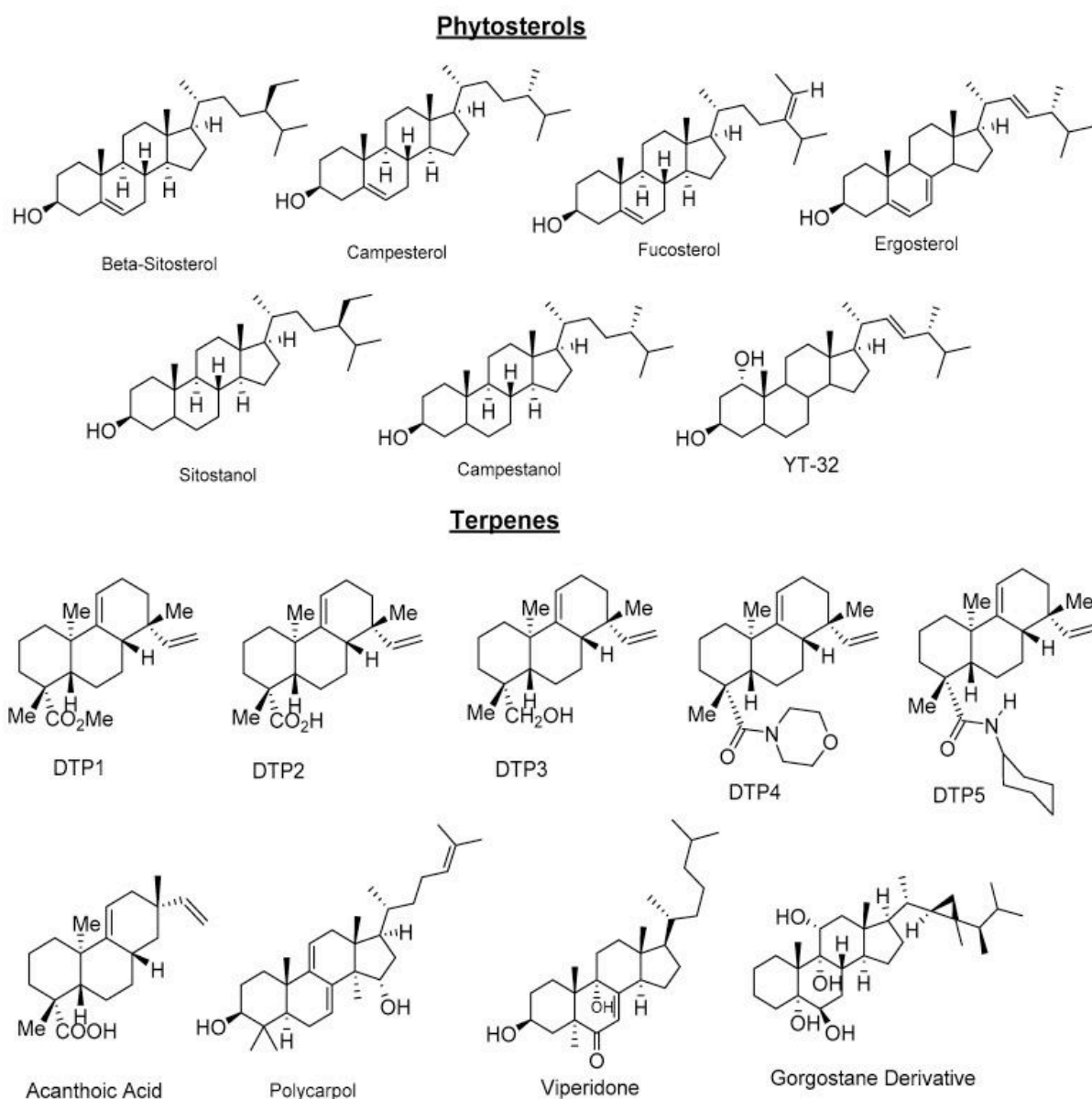


Figure 56. Ligands naturels de LXR (Komati *et al.*, 2017)

De nombreuses compagnies pharmaceutiques se sont intéressées à développer des agonistes synthétiques de LXR, une classe de molécules d'un fort potentiel thérapeutique. Cependant, LXR étant impliqué dans un grand nombre de voies de signalisation, il est difficile d'induire uniquement les voies métaboliques d'intérêt dans le traitement. En effet, il est difficile de favoriser l'efflux sans provoquer l'activation par exemple de la lipogénèse via SREBP1c, un gène sous le contrôle de LXR. Certains oxystérols ont été impliqués dans la formation de tumeurs tandis que d'autres sont considérés comme des agents antitumoraux. Ainsi, parmi les composés les plus étudiés le T0901317 reste celui le plus fréquemment utilisé dans les études fondamentales sur les LXR. D'autres molécules comme le GW3965 ou le GW6340 ont été développées.

1.2.4.2. Le rôle des récepteurs nucléaires LXR

L'activation de LXR régule l'expression des gènes impliqués dans la lipogenèse telle que la synthèse des acides gras. En effet, il a été rapporté que l'agoniste T0901317 induit l'expression de SREBP-1c et de ses gènes cibles tels que FASN, ACC et SCD (Darimont *et al.*, 2006), entraînant une augmentation de des triglycérides dans le plasma et le foie (Schultz *et al.*, 2000). En revanche, l'administration du T0901317 à des souris dépourvues de gènes LXR α et LXR β a entraîné une absence de régulation des gènes de la lipogenèse, ce qui confirme la nécessité de LXR pour l'activation de cette voie métabolique (Joseph *et al.*, 2002).

LXR est également impliqué dans la régulation du métabolisme du cholestérol. En effet, l'augmentation des concentrations de certains stérols active le facteur de transcription LXR et stimule l'efflux du cholestérol via une synthèse accrue des transporteurs ABC et de l'ApoE mais également une diminution de la captation du cholestérol par les LDLR via l'expression de IDOL (Moschetta, 2011).

Bien que l'homéostasie lipidique dont celle du cholestérol soit finement régulée par les différents processus précédemment décrit, ces derniers sont particulièrement altérés dans différentes lignées cancéreuses. Ces dérégulations entraînent une accumulation de lipides et notamment celle du cholestérol intracellulaire nécessaire pour maintenir le taux de croissance élevé et assurer les fonctions des cellules cancéreuses. Par conséquent, le métabolisme des lipides dont celui du cholestérol peut constituer une cible thérapeutique particulièrement intéressante dans le traitement du cancer du sein. Différentes études épidémiologiques cliniques et expérimentales attestent ce dérèglement métabolique et sont exposés par la suite.

1.3. Les marqueurs de la dérégulation du métabolisme lipidique dans le cancer du sein

1.3.1. Acides gras et cancer du sein

Les principaux marqueurs de la dérégulation du métabolisme lipidique sont impliqués dans la lipogenèse de *novo*. En effet, une augmentation importante de l'expression de plusieurs enzymes clés responsable de la synthèse de *novo* telles que ACC1, FASN et SCD1 a pu être démontrée dans plusieurs lignées cancéreuses dont le cancer du sein.

Ainsi, il a été démontré que la répression du gène ACC à la suite d'une interférence de l'ARN entraîne une diminution de la synthèse d'acide palmitique et induit l'apoptose des cellules cancéreuses mammaires MDA-MB-231 (Chajès *et al.*, 2006). Cependant, l'une des plus importantes enzymes impliquées dans la lipogenèse et étudiée en cancérogenèse est la FASN (Menendez and Lupu, 2007). Une surexpression de cette enzyme a été rapportée dans divers cancers, notamment le cancer du sein, prostate, le foie et les ovaires (Phan *et al.*, 2014; Van de Sande *et al.*, 2002; Yoon *et al.*, 2007).

Des études ont permis de mettre en évidence, le rôle oncogène de cette enzyme dans le cancer du sein. En effet, l'inhibition de l'expression de FASN dans les cellules cancéreuses du sein réprime la croissance tumorale et métastatique en interrompant l'inflammation liée à une surexpression de FASN (Liu *et al.*, 2014). L'expression de cette enzyme connue pour être régulée par SREBP a également été reportée pour être sous le contrôle du récepteur de facteur de croissance épidermique HER2 (Human Epidermal Growth Factor Receptor-2) par un mécanisme de phosphorylation directe (Jin *et al.*, 2010). En effet, des analyses transcriptomiques comparant les lignées cellulaires de cancer du sein HER2 positive (HER2+) et HER-2 négative (HER2-) ont montré une surexpression de l'enzyme FASN, ACC1 et SCD1 dans les cellules HER2+(Baumann *et al.*, 2013; Monaco, 2017; Yoon *et al.*, 2007).

La surexpression de SCD1 observée dans différentes cellules cancéreuses notamment du sein, du poumon, du côlon et du rein s'est avérée être associée à une prolifération des cellules cancéreuses et à une diminution de la mort cellulaire (Angelucci *et al.*, 2018). Cette enzyme, pour rappel, est impliquée dans la synthèse des acides gras monoinsaturés (AGMI), décrite précédemment (Cf.§1.1.1.1). Des études ont également démontrés que des œstrogènes nécessaires à la croissance des cellules cancéreuses du sein (ER+), tel que le 17- β -estradiol, augmente l'expression de SCD1 et le contenu en AGMI dans les cellules de cancer du sein ER+, MCF-7 et T-47D (Belkaid *et al.*, 2015). Une étude récente a permis de montrer que l'inhibition pharmacologique de l'activité SCD1 réduit la prolifération cellulaire dans les cellules cancéreuses du sein humain MCF-7, induit l'apoptose et l'arrêt du cycle cellulaire, tout en diminuant la capacité de migration (Zhao *et al.*, 2017).

Par conséquent, le métabolisme des acides gras peut constituer une cible thérapeutique particulièrement intéressante dans le traitement du cancer du sein. Nous nous sommes également intéressés à la relation pouvant exister entre le cholestérol et le cancer du sein.

1.3.2. Cholestérol et cancer du sein

La dérégulation du métabolisme du cholestérol conduisant à son accumulation dans les cellules cancéreuses est connue pour être l'une des caractéristiques de plusieurs cancers dont le cancer du sein. En revanche, bien que la relation entre l'obésité, le syndrome métabolique et le risque de développer un cancer du sein, soit désormais reconnue par la communauté scientifique ; celle entre l'hypercholestérolémie pourtant courante dans l'obésité, le syndrome métabolique et le risque du cancer du sein est très controversée.

Ainsi, d'après les études cliniques épidémiologiques et expérimentales cette controverse est exposée afin de comprendre au mieux l'enjeu potentiel de cette molécule dans le traitement du cancer du sein. Les différences entre les données peuvent s'expliquer par sa distribution entre les différentes grandes classes de lipoprotéines (VLDL, LDL et HDL) et sa modulation selon le mode de vie et le statut ménopausique.

Etudes cliniques et épidémiologiques

Les résultats des études cliniques et épidémiologiques menées sur les associations entre le cholestérol sérique et ses transporteurs lipoprotéiques (VLDL, LDL et HDL), et le cancer du sein demeurent controversés. Certaines études suggèrent que le cholestérol à un effet protecteur, alors que d'autres le présentent comme un facteur de risque, ou même encore dans certains cas sans aucun effet.

Ainsi, d'après l'étude du profil lipidique de patientes et plus particulièrement celui des lipoprotéines, il a été démontré que des taux plus élevés de VLDL-C et des taux plus faibles d'ApoA1, un composant des HDL, étaient significativement associés au cancer du sein (Chang *et al.*, 2007). Au contraire, il a été suggéré dans une autre étude, qu'un taux plus élevé de HDL-C permettrait de réduire le risque de cancer du sein chez les femmes préménopausées (Kim *et al.*, 2009). Dans ce sens, des analyses épidémiologiques réalisées sur la cohorte ARIC ont permis de mettre en évidence qu'un faible taux du HDL-C chez les femmes préménopausées pouvait être un marqueur de risque accru de cancer du sein (Kucharska-Newton *et al.*, 2008).

Ces résultats sont alors en opposition avec ceux obtenus au cours d'une autre étude, pour laquelle les patientes atteintes d'un cancer du sein présentaient une concentration de LDL-C plus élevée et un taux de VLDL-C plus faible (Laisupasin *et al.*, 2013).

Bien qu'aucune association entre le HDL-C ou le cholestérol total et le cancer du sein ne soit évidente, une méta-analyse réalisée en 2015 a permis de confirmer de manière significative la preuve d'une corrélation inverse entre ces deux paramètres (Touvier *et al.*, 2015).

Ainsi, comme il a pu être expliqué précédemment, le cholestérol sérique est issu de l'apport exogène c'est-à-dire par l'alimentation, ou de la synthèse endogène. Par conséquent, l'apport alimentaire en cholestérol ainsi que l'effet de l'utilisation de médicaments hypocholestérolémiants sur l'incidence du cancer du sein sont également des thèmes de recherches étudiées par la communauté scientifique.

Dans une étude de population, il a été décrit que le cholestérol alimentaire était associé positivement au risque de cancer du sein principalement chez les femmes ménopausées (Hu *et al.* 2012). Cette corrélation a également pu être mise en évidence de manière significative lors d'une méta-analyse réalisée en 2016, avec la précision d'un apport en cholestérol alimentaire journalier supérieur à 370mg/jour (Li *et al.*, 2016). Autour de cette relation cholestérol et cancer du sein, l'effet des médicaments hypocholestérolémiants (statines) a également intéressé les chercheurs. Les statines sont reconnues pour inhiber sélectivement l'HMGCoA-réductase, qui est l'enzyme clé de la biosynthèse du cholestérol. Cependant, là encore cette relation est controversée.

Trois méta-analyses ne soutiennent pas l'hypothèse que les statines ont un effet protecteur contre le cancer du sein (Islam *et al.*, 2017; Undela *et al.*, 2012), contrairement à une autre dans laquelle l'utilisation de statines a été associée à une réduction de la récurrence du cancer du sein (Mansourian *et al.*, 2016). Cette méta-analyse comprend les résultats d'une étude clinique menée sur 8 010 patientes ménopausées présentant un cancer du sein invasif à récepteurs hormonaux positifs de stade précoce, dans laquelle il a été démontré que l'ajout hypocholestérolémiants comme adjuvant au cours d'un traitement endocrinien joue un rôle dans la prévention et la récurrence du cancer du sein (Borgquist *et al.*, 2017). Pour appuyer ce résultat, des études expérimentales réalisées récemment ont montré que en plus de cet effet sur la biosynthèse du cholestérol, la simvastatine permettait la réduction de la prolifération cellulaire du cancer du sein et induisait l'apoptose sur des lignées cancéreuses mammaires MCF-7 (ER+) et MDA-MB-231 (ER-) (Bai *et al.*, 2019).

D'après ces études, il est difficile d'identifier le cholestérol comme un facteur de risque à proprement identifié. Il a plutôt été suggéré que 27-hydroxycholestérol (27-OH) pourrait être un candidat plus approprié. Ce composé est un oxystérol endogène avec de multiples fonctions biologiques et notamment une capacité de modulateur sélectif des récepteurs des œstrogènes (SERM). Par conséquent des études ont démontré qu'un taux plus faible de 27-OHC par l'inhibition de l'enzyme CYP27A1, responsable de sa synthèse, entraîne une diminution de la croissance tumorale de plusieurs lignées cancéreuses dont celles du cancer du sein ER+ et ER-. Les mécanismes impliqués dans cette diminution de la croissance tumorale sont en lien avec le système immunitaire. En effet, il a pu être observé qu'aux sites métastatiques, le 27-OHC était associé à une augmentation de la concentration de polynucléaires neutrophiles et des lymphocytes T gamma-delta et à une baisse du nombre de lymphocytes cytotoxiques CD8+ (Baek *et al.*, 2017).

Etudes expérimentales LXR/PPAR régulation du métabolisme lipidique et cancer du sein.

Des études chez l'animal et chez l'humain ont montré que le cholestérol circulant reflète étroitement ceux du métabolite primaire et du 27-OH. De même que l'hypercholestérolémie entraînée des taux élevés de 27-OH (Borgquist *et al.*, 2018; Garcia-Estevez and Moreno-Bueno, 2019; Kimbung *et al.*, 2017) Par ailleurs, certaines ont montré que cet oxystérol fonctionne comme mitogène dans les tumeurs ER-positives et comme ligand du récepteur nucléaire Liver X Receptor (LXR) (Umetani *et al.*, 2007).

Or, les études réalisées sur le récepteur LXR et ses ligands (oxystérols et ligands synthétiques) impliqués dans la tumorigenèse mammaire, ont révélé que tous les oxystérols dérivés du cholestérol ne se comportent pas comme le 27-OH. En effet, les ligands synthétiques et naturels de LXR, T0901317 et 22-OH, ont tous les deux inhibés la prolifération et induit l'apoptose d'une lignée cellulaire modèle de cancer du sein (ER+) (El Roz *et al.*, 2012).

De plus, l'activation de LXR par T0901317 a diminué l'expression de la flottilline-2, un biomarqueur des radeaux lipidiques qui joue un rôle important dans la progression du cancer et la voie de signalisation Akt dans la lignée cellulaire MCF-7 (Carbonnelle *et al.*, 2017). Le 22(R)-hydroxycholestérol et 24(S)-hydroxycholestérol ont été reportés pour inhiber la prolifération des cellules cancéreuses de la prostate et du sein (Fukuchi *et al.*, 2004). Les lignées de cellules cancéreuses dont l'expression de l'ARNm pour le gène LXR α était plus élevée seraient plus sensibles à l'inhibition induite par le 22(R)-hydroxycholestérol (Chuu and Lin, 2010).

Par conséquent, les études montrant à la fois le cholestérol comme étant un facteur de protection et un facteur de risque potentiel de cancer du sein suscite une grande controverse au sein de la communauté (Carmichael, 2006). Néanmoins, cette controverse pourrait s'expliquer par la dépendance des résultats à des critères particuliers en lien avec le statut hormonal bien spécifique. En effet, les dernières méta-analyses réalisées précisent le type de cancer du sein déclaré (ER dépendant ou non dépendant), le statut de ménopause au moment duquel l'obésité et le cancer du sein sont diagnostiqués, mais également la durée d'exposition aux estrogènes qu'a pu avoir une femme au cours de sa vie.

Ainsi, outre la prolifération incontrôlée des cellules cancéreuses mammaires, ces différentes preuves suggèrent que celles-ci sont également caractérisées par une altération du métabolisme des lipides, dont celui du cholestérol (Hanahan and Weinberg, 2011; Santos and Schulze, 2012).

2. Objectifs de l'étude

Les différentes stratégies menées sur la souche *P. chrysogenum* dans la production de métabolites possédant un effet antiprolifératif sur les cellules MCF-7, ont permis de mettre en évidence par fractionnement bioguidée, dérégulation et par analyse biochimométrique le même hit identifié comme étant l'ergostérol.

Au regard de la structure chimique de l'ergostérol s'apparentant à un composé stéroïdien peu étudié en cancérologie, et sachant que le ciblage du métabolisme des lipides dans les cellules cancéreuses constitue une stratégie thérapeutique efficace et spécifique dans le traitement du cancer du sein, nous avons décidé d'étudier et d'évaluer le potentiel de cette molécule sur : 1) l'influence la prolifération des cellules cancéreuses mammaire ; 2) sur la régulation du métabolisme lipidique.

Afin d'établir le lien possible entre cette inhibition de la prolifération et la régulation du métabolisme lipidique, il a été choisi d'étudier l'effet de cette molécule sur les voies de signalisation impliquées dans la régulation de l'homéostasie lipidique de deux lignées cellulaires du cancer du sein MCF-7 (ER+) et MDA-MB-231 (ER-). Ainsi, l'effet de l'ergostérol a été évalué sur l'expression des gènes impliqués dans la lipogenèse et la bêta-oxydation des acides gras ainsi que la synthèse, le stockage et l'efflux du cholestérol. Parmi les investigations menées, les résultats sur l'homéostasie du cholestérol ont été présentés sous la forme d'une publication en cours de préparation. Les résultats sur la lipogenèse et la bêta-oxydation ne sont pas présentés du fait du manque d'effet notable.

Enfin, l'activité cytotoxique de plusieurs dérivés de l'ergostérol a été étudiée sur les cellules cancéreuses mammaires MCF-7, pour identifier si possible les groupements chimiques potentiellement responsables de l'activité antiproliférative observée pour cette classe chimique de molécule.

3. Présentation de la publication

Ergosterol decreases breast cancer cells viability by targeting cholesterol metabolism and rafts membranes

Lucie ORY¹, Delphine Carbonelle¹, Jean-Michel Huvelin¹, Mikael Croyal², Catherine Roullier¹
and Hassan Nazih^{1*}

¹Université de Nantes, Mer Molécules Santé, MMS EA 2160, F-44000, Nantes, France

²Université de Nantes, Nantes Atlantique Université, INRA, UMR1280 Physiologie des Adaptations Nutritionnelles, F-44000 Nantes, France.

* Correspondence: el-hassane.nazih@univ-nantes.fr

Abstract

Background/Aim

Alteration of cholesterol metabolism is common to all cancer cells including mammary cancer cells lines. This alteration consists of an increase in cellular cholesterol. The increase of intracellular cholesterol in cancer cells is required for cells to ensure their rapid proliferation and need for energy. The main purpose of this article is to evaluate the effect of ergosterol on cholesterol metabolism and membrane raft integrity in two breast cancer cell lines: MCF-7 and MDA-MB-231. Its ability to inhibit cell viability and *p*-Akt signaling pathway has also been investigated.

Methods

RT-qPCR was performed to evaluate the expression of genes involved in cholesterol metabolismuptake. Quantification of cellular cholesterol was performed using gas-chromatography coupled to mass spectrometry (GC-MS). Membrane rafts integrity was investigated by studying the protein expression of raft membrane markers, flotillin-2,

caveolin and epidermal growth factor receptor EGFR by western blot. Cell viability and *p*-AKT were studied using the MTT assay and flow cytometry assay, respectively.

Results

Interestingly, we demonstrated that ergosterol treatment reduced total and esterified cholesterol. An increase in the expression of the genes involved in cholesterol efflux, the transporter ABCG1, and in the degradation of LDLR, the ligase IDOL has also been observed. We also demonstrated that ergosterol disrupted membrane rafts by decreasing Flot-2, Caveolin and EGFR in MDA-MB-231 cells. Ergosterol also decreased MCF-7 and MDA-MB-231 cells lines viability and *p*-AKT.

Conclusion

Our results show that ergosterol exerts promising antiproliferative activity by reducing cellular cholesterol content and modifying membrane raft integrity in breast cancer cells.

Keywords: Ergosterol, breast cancer cells, cholesterol homeostasis, membrane raft, ABC transporters

Introduction

Breast cancer is the most commonly diagnosed and the second leading cause of cancer death among women in worldwide. In 2018, approximately 2.1 million of new cases and 627 000 deaths were reported [1]. Cancer is usually considered as uncontrolled replication of abnormal cells, characterised by resistance to apoptosis, cell migration and several metabolic alterations [2]. One of the emerging metabolic hallmarks of cancer cells is the deregulation of lipid metabolism associated with an accumulation of cholesterol [3].

These alterations in lipid metabolism have been reported in different models of cancer cell lines including breast, colorectal, liver and lung [4–6]. Cholesterol is essential and necessary constituent of plasma membrane and especially in the lipid rafts microdomains [7]. These microdomains are signaling platforms involved in many cellular processes such as membrane trafficking and signal transduction including cell growth, apoptosis and migration [8–10]. Several rafts biomarkers are identified in these microdomains such as Flotillin-2 and EGFR (Epidermal growth factor receptor) and are upregulated in various cancer [11], suggesting their involvement in tumorigenesis. Caveolae, other membrane microdomains, are structurally recognized invaginations of the plasma membrane due to the presence of the protein caveolin-1, which plays a critical role in the progression of breast cancer including cell proliferation, apoptosis, autophagy, invasion, migration and breast cancer metastasis [12]. In fact, cancer cells require a constant supply of lipids for membrane biogenesis to ensure the rapid proliferation and need for energy. It has been shown that cancer cells increase their cholesterol biosynthesis via HMGCoA Reductase (3-hydroxy-3-methylglutaryl-CoA reductase) the rate-controlling enzyme of this pathway and that this enzyme is highly expressed in various cancer cells, such as breast and prostate cancer cells [4], and have recently been reported as a target for cancer treatment [13,14]. The increase in cellular cholesterol could also be the result of a decrease in cholesterol efflux via ABC transporters and/or an increase in its uptake via LDL receptors (LDLR) [15–17]. ABC transporters and LDL receptor are under the control of the nuclear receptor LXR (Liver X Receptor) [18]. Indeed, El Roz et al., have demonstrated that LXR activation, by synthetic ligand TO901317 or by the oxysterol 22Rhydroxysterol, in MCF-7 cells deprive cells of cholesterol, required for their growth, by stimulating its efflux through the expression of ABCG1, resulting in the inhibition of cell proliferation and in stimulation of apoptosis [19]. Concerning the mode of action of

the decrease in cholesterol influx via LDLR, LXR stimulates LDL receptor expression through induction of an E3 Ligase that ubiquitinates LDLRs called IDOL (Inducible Degradator of LDLR)[18].

Ergosterol is the most important sterol produced from fungi. Although it was reported for it has anti-inflammatory and antiproliferative effects on breast cancer cells [20–22], no investigations were performed on the effect of this sterol on the cholesterol homeostasis.

In this study, we investigated the ergosterol effect on cellular cholesterol content and the expression of gene involved in cholesterol efflux and uptake in two breast cancer cell lines, MCF-7 and MDA-MB-231. We also evaluated the effect of ergosterol on the integrity of membrane rafts to determine whether targeting of lipid metabolism can disrupts membrane rafts in breast cancer cells. Its ability to inhibit cell growth and *p*-AKt has also been studied.

Materials and methods

Material

Two human breast cancer cell lines, MCF-7 and MDA-MB-231, and human monocyte THP-1 cells were obtained from the European Collection of Animal Cell Cultures (Salisbury, United Kingdom). Ergosterol (ref : 45480), Dulbecco's Modified Eagle's Medium (DMEM), Roswell Park Memorial Institute medium (RPMI 1640), fetal bovine serum (SVF), penicillin-streptomycin, glutamine, bovine serum albumin (BSA), trypsin, 3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide (MTT), dimethyl sulfoxide (DMSO), ethanol (EtOH), primers for qPCR, triton X-100, OptiPrep™ Density Gradient Medium, and antibodies against beta-actin and EGFR were purchased from Sigma Aldrich (Saint-Quentin Fallavier, France). TRIzol reagent for RNA isolation was from Invitrogen (Cergy-Pontoise, France). iScript™ Reverse Transcription Supermix for RT-qPCR and iQTM SYBR Green Supermix for qPCR and SDS page

or nitrocellulose membrane were purchased from Bio-Rad (Marnes-la-Coquette, France). IRDye whole IgG secondary antibodies were purchased from LI-COR Biosciences (Bad Homburg, Germany).

Method

Cell line and culture

Two human breast cancer cell lines (MCF-7 and MDA-MB-231) and human monocyte THP-1 were cultured in DMEM and RPMI 1640, respectively. One mouse macrophage J774 cell line was also used and cultured in DMEM for experiments. Both mediums were supplemented with 10% fetal bovine serum, 1% glutamine and 1% penicillin-streptomycin. Cells were cultured at 37°C under a controlled humidified atmosphere (95%) containing 5% CO₂.

Cell treatment

Ergosterol stock solution dissolved in absolute ethanol (3.78 mM) was diluted in DMEM containing 0.1% bovine serum albumin to prepare cells treatment. The untreated cells (negative controls) were incubated with the same medium containing equals percentage of EtOH.

Cell viability assay

MCF-7 and MDA-MB-231 were plated at a density of 10⁴ cells/well in a 96-well plate and incubated for 24H. Then, cells were treated with various concentrations of ergosterol (0.001-0.01-0.2-1-5-12.5 and 25 µM) for 24, 48 and 72H. For MTT assay, MTT solution (50 µL of 2.5 mg/mL) was added to each well to get formazan crystals and incubated for 4H at 37°C. Then, solution was removed and formazan deposit was solubilized in 200 µL of DMSO. The absorbance of formazan solution at 570 nm was measured using SpectraMax 190.

Cell viability was then calculated as a percentage of the control that was not treated with ergosterol.

Akt Phosphorylation analysis

MCF-7 cells were plated at 4.10^5 cell/well in 6 well plate and treated with ergosterol 25 μ M for 24H. Then cells were trypsinized, and fixed with 4% paraformaldehyde (PAF), permeabilized in 90% MeOH and then incubated with the primary antibody against Phospho-Akt (1:100) for 1H at room temperature. After incubation 1H at room temperature with the fluorescent secondary Alexa Fluor 488 antibody (1:500), the cells were analyzed by flow cytometer (BD Accuri™ C6).

RNA extraction and RT-qPCR (quantitative Reverse Transcription Polymerase Chain Reaction)

MCF-7 and MDA-MB-231 were treated with ergosterol at 12.5 μ M for 24 and 48H. Total RNA was extracted by the TriZol Reagent (Invitrogen, Cergy-Pontoise, France) according to manufacturer's instruction. The mRNA (1 μ g/ μ L) was then reverse-transcribed into complementary DNA using iScript Reverse Transcription Supermix Bio-Rad (Marnes-la-Coquette, France). An initial priming step for 5 min at 25°C was followed by reverse transcription phase of 30 min at 42°C and completed by RT inactivation step of 5 min at 85°C. Quantitative PCR was then performed on a MyiQ Real Time PCR Detection System, using the SYBR Green Supermix (Bio-Rad, Marnes-la-Coquette, France) with 45 cycles of amplification (95°C for 30 s and 60°C for 30 s). The relative quantification was performed using the $2^{-\Delta\Delta CT}$ method, with $\Delta\Delta CT = \Delta CT$ reference gene - ΔCT gene of interest. The primer sequences are shown in Table 1.

Table 1. The primer sequences

Gene	Gene name	Amorce Forward Amorce Reverse
18S rRNA	18S ribosomal RNA	F- GATGCGGCGGCGTTATTCC R- CTCCTGGTGGTGCCCTTCC
Rps6	Ribosomal protein s6	F- AAGCACCCAAGATTCAGCGT R- TAGCCTCCTCATTCTCTTGGC
ABCG1	ATP Binding Cassette Subfamily G Member 1	F- CAGGAAGATTAGACACTGTGG R- GAAAGGGGAATGGAGAGAAGA
ABCA1	ATP Binding Cassette Subfamily A Member 1	F- TCAGTGGGATGGATGGCAAAG R- TCCGACTCCGTCTGGCAATTA
HMGCR	3-Hydroxy-3-Methyl-Glutaryl-Coenzyme A Reductase	F- TAACTCCTCCTTACTCGATAC R- AATAGATACACCACGCTCAT
IDOL	Inducible degrader of the LDLR	F- GGCATATCTTTTTCTTGAC R- AGCTCCTCATAGTTATACTTGG
LDLR	Low-density lipoprotein receptor	R- GAGGACAAAGTATTTTGGACAG F- GTAGGTTTTAGCCAACAAG
APOE	Apolipoprotéine E	R-CTGCGTTGCTGGTCACATTCC F-GTCCTTCAAGAAGAAGGTGATAC
LXR α	Liver X receptor	R-GCTCCCCACCGCTGCTCTC F-TGCCCTTCTCAGTCTGTTCCAC
ABCG5	ATP Binding Cassette Subfamily G Member 5	R- CAAGTTTCTTGTCACCTCTCC F- GAATCTGCCTACAAGAAATCAG
ABCG8	ATP Binding Cassette Subfamily G Member 8	R- CATTGGAGATGACATCCAG F- GTCAGATTTCCAACGACTTC

Biochemical membrane raft isolation

MCF-7 and MDA-MB-231 breast cancer cells were plated with a density of 2×10^6 cells in petri dishes and incubated at 37°C overnight, then cells were treated with ergosterol at 12.5 μ M for 24H. The isolation of the membrane rafts was carried out with protocol optimisation previously described by J.Elia [23]. Cells were washed with ice-cold phosphate-buffered saline (PBS) and lysed with 1 mL of lysis media containing 1% triton X-100 per condition. Next, the lysed cells (0.630 mL) were purified using an OptiPrep™ density gradient (35% (w/v), 0.870 mL) and placed in the bottom of a 10 mL ultracentrifuge tube. A discontinuous density gradient was then prepared by overlaying the three OptiPrep™ gradients (1.5 mL for

each of the following gradients 30, 25 and 20%). The upper gradient layer consisted of 1 mL lysis buffer. Samples were centrifuged at $200,000 \times g$ for 4H at 4°C in a Beckman L8-70M Ultracentrifuge, using TFT 65.13 rotor. Nine 1 mL fractions were collected from top to bottom of the gradient for each condition and then analyzed by western blot. The fractions corresponding to the membrane rafts are fractions 2, 3 and 4 isolated from MCF-7 and MDA-MB-231 cells.

Protein extraction

MCF-7 cells were seeded and treated with ergosterol in a 6-well plate, as described in the previous section on RNA extraction. After washing with fresh PBS, the cells were lysed in 200 μl RIPA buffer [50 mM Tris-HCl pH 8, 150 mM NaCl, 1% Nonidet P-40, 0.5% sodium deoxycholate, 0.1% Sodium Dodecyl Sulfate (SDS)], supplemented with protease inhibitors aprotinin (2 $\mu\text{g}/\text{ml}$) and phenylmethylsulfonyl fluoride (1 mM). Then lysate were incubated at 4°C for 30 min and centrifuged (4°C for 30 min, 12000 rpm). The total cellular protein extract was collected and protein concentration was then determined according to the bicinchoninic acid assay method (BCA) using a bovine serum albumin (BSA) standard.

Western blot analysis

Equal amounts of proteins (20 μg) were separated by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and transferred to a nitrocellulose membrane. Then it was blocked with 5% skim milk prepared in Tris-buffered saline tween 20 solution (TBS-0.1% tween 20) at room temperature for 2H before to be incubated overnight at 4°C with primary antibodies against EGFR(1/1000), FLOTillin-2 (1/500), Caveolin-1 (1/1500). The membrane was then washed and incubated with IRdye whole IgG secondary antibodies (1/15000) at

room temperature for 2H. Finally, proteins bands were visualized and quantified using the B446-LI-COR Odyssey® Infracted Imaging System.

Cellular cholesterol quantification

Cholesterol was extracted from cell lysates and from standard solution (150 μ L) with 400 μ L of a cyclohexane/isopropanol mixture (3/2, v/v) containing D7-cholesterol (10 μ M) as internal standard. After centrifugation (10 min, 14000 g, 15°C), supernatants were split in two vials and dried under nitrogen stream (60° C) to prepare saponification or derivatisation reactions. For saponification, the reaction was performed adding 200 μ L of ethanol 90% and 100 μ L of NaOH 1N (100 °C, 3.5H). Then, 10 μ L of distilled water were added and cholesterol extraction was performed with 1.5 mL of cyclohexane. Extracted sample were spun (3 min, 3000 g, 10°C) and dried under nitrogen (60°C). All Dried samples were then derivate with 30 μ L of N-methyl-N-trimethylsilyl-trifluoroacetamide (40 min, 70 °C) and evaporated to dryness under nitrogen stream (60 °C) before to be analyzed using an Agilent 6890N gas chromatophaphy coupled to a single-quadrupole Agilent 7683B mass spectrometer (Agilent Technologies, Palo Alto, CA). The injection of 5 μ L of samples reconstituted in cyclohexane (100 μ L) were eluted on a HP-1MS capillary column (30 m, ID 250 μ m, film 0.25 μ m) with an injector set at 250 °C and helium gaz vector at a flow rate of 1.2 mL/min. The elution was performed using the following temperature gradient : 50°C/min from 150 °C to 280 °C, increased to 290 °C at 1°C/min and finally increased to 300 °C at 50 °C/min. Peak identification was achieved by comparing the retention time and the characteristic fragments ions observed from standard solution (Table 2).

Tableau 2. Selected reaction monitoring

*major detected peaks

Compound	m/z	R _t (min)
ergosterol	468, 363*, 337	9.00
cholesterol	458, 329*, 301	8.17
D7-cholesterol	465, 336*, 308	8.10

The esterified sterol content from samples was obtained from the subfractions on free sterol to total sterol content. All the concentrations were then normalized to the protein concentration determined according to the bicinchoninic acid assay (BCA).

Lipid staining

MDA-MB-231 and MCF-7 cells were seeded at 2×10^4 cells in 96 microwell plates (Nunc 160376, Nest Biotechnology Co.,Ltd) and treated or not with ergosterol for 24H. Cells were washed with PBS and fixed with cold Paraformaldehyde 4% and then stained with Nile Red (5 μ g/mL) prepared in PBS from a stock solution at 0.5mg/ml in DMSO (Sigma-Aldrich). After incubation (30 min in darkness), cells were washed with 1X PBS and cell nucleus were counterstained (1:200, Life Technologies) from a 5 mg/mL DAPI stock solution diluted in PBS. To remove excess of coloration, cells were washed and fixed using Ibidi aqueous mounting medium (Ibidi cell in focus, Germany). Fluorescence microscopy was done using the INcell Analyzer 2200 (GE Healthcare, Velizy-Villacoublay, France) equipped with Nikon (20X/0.45) focal objective. Nile Red stains the lipid droplets and is observed using the red color channel (Ex/Em : 542/597 nm with an exposure time of 100 ms), whereas DAPI are observed in the blue color channel (Ex/Em : 390/435 nm with an exposure time at 500ms). Images were acquired with the two detection channels and merged using ImageJ software.

Data analysis

Data represent mean values \pm standard deviation of three independent experiments performed in triplicate. The Student's t-test was used and the p-value <0.05 was considered significantly different from the corresponding control using Excel software.

Results

Effect of ergosterol on cellular cholesterol content

The GC-MS quantification of cholesterol performed on MCF-7 cells showed that the amount of total cholesterol was low (29.22 nmol/mg of protein) in cells treated by ergosterol compared to control condition (55.83 nmol/mg of protein). Esterified cholesterol has been also decreased by ergosterol (0.34 compared to 1 nmol/mg of protein for control).

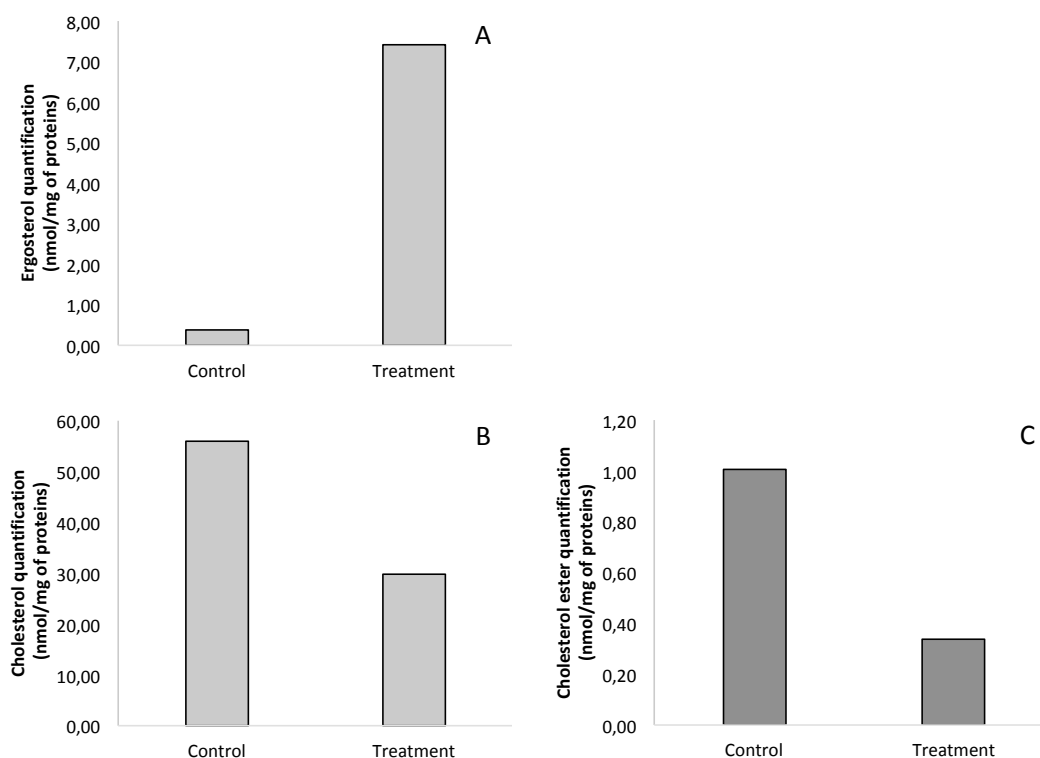


Figure 1. Quantification of cellular ergosterol and cholesterol in MCF-7 cells. After a treatment of ergosterol (12.5 μ M) for 24H, the ergosterol (A) ; cholesterol (B) and esterified cholesterol (C) were quantified from MCF-7 cell lysates by GC-MS. Results are normalized

and expressed according to the amount of protein obtained in cell lysate using the BCA assay (BiCinchoninic Acid).

Effect of ergosterol on lipid droplets

Lipid droplets (LDs) consist of mainly cholesterol ester (CE) and triglycerides (TG) can be stained with Nile Red O. Cells treated with ergosterol (12.5 μ M) have visibly less Nile Red O, consequently less LDs, compared to untreated cells.

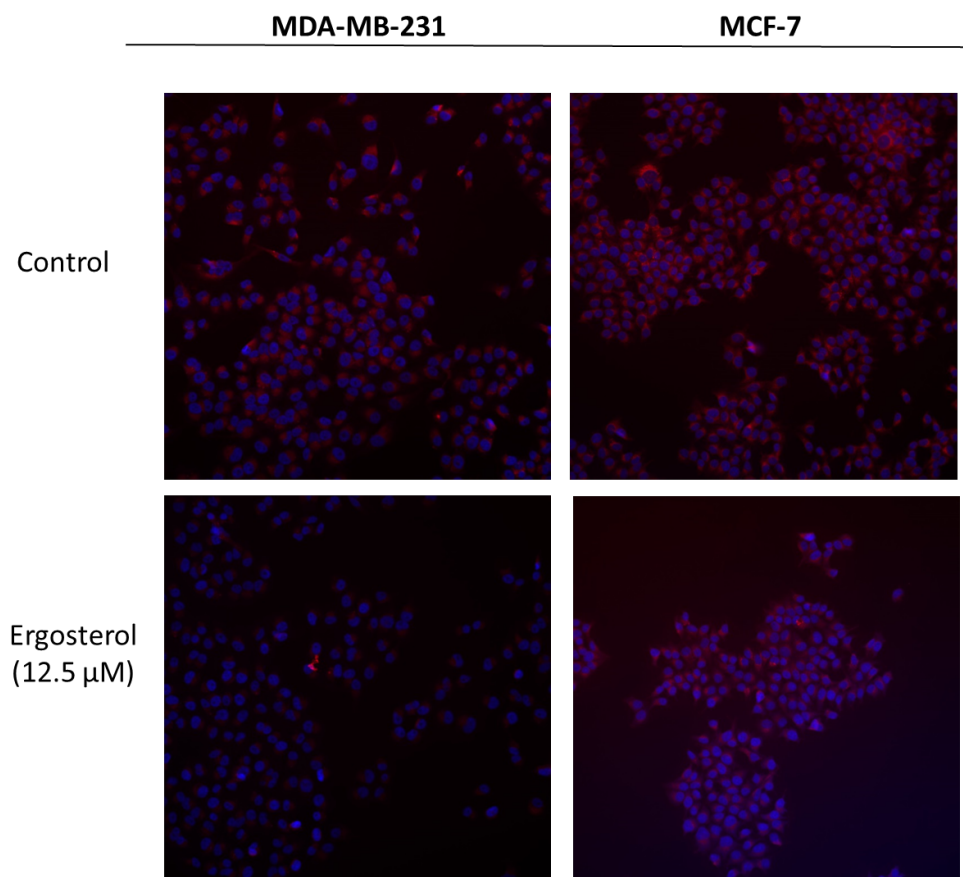


Figure2 Effect of ergosterol on accumulation of lipid droplets in MCF-7 and MDA-MB-231 breast cancer cells. Cells were treated with ergosterol at 12.5 μ M for 24H Nile Red staining corresponding to lipid droplets (Red) and DAPI staining corresponding to nucleus (Blue).

Effect of Ergosterol on ABCG1 cholesterol transporter genes

After treatment of ergosterol (12.5 μ M), the mRNA expression of ABCG1 increased in MCF-7 cells (2.1, 1.7, 3.6) and in MDA-MB-231 (2.4, 3.2, 4.2) for 24, 48 and 72H, compared to untreated cells (Fig. 2).

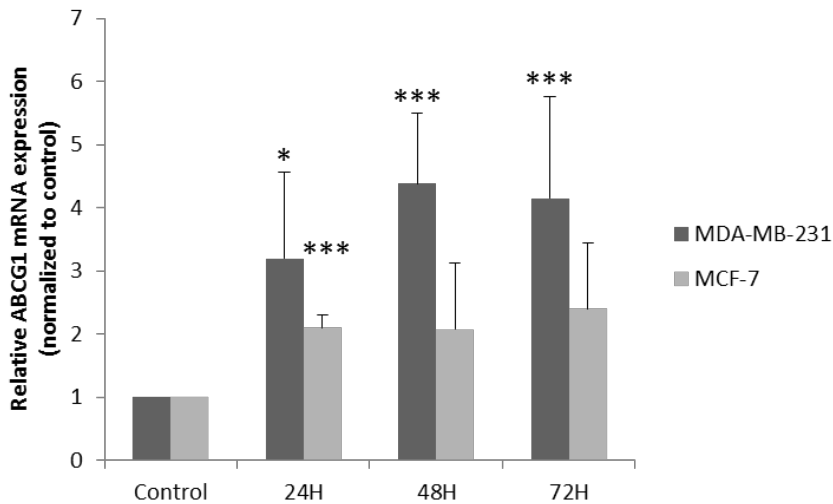


Figure 3. Effect of ergosterol treatment (12.5 μ M) on the ABCG1 mRNA expression in MCF-7 and MDA-MB-231. The mRNA expression was examined by Real-time RT-qPCR. RpS6 rRNA was used as a housekeeping gene. Results are expressed as the means \pm standard deviation with * $p < 0.05$; *** $p < 0.0001$.

Effect of ergosterol on IDOL and the LDLR expression

The effect of ergosterol on gene expression of LDLR and IDOL were determined by RT-qPCR. Data obtained for cells treated with ergosterol (12.5 μ M) suggested an increase of IDOL gene expression in MDA-MB-231 (1.96 and 1.71) and in MCF-7 (2.53 and 1.59) for 24 and 48H.

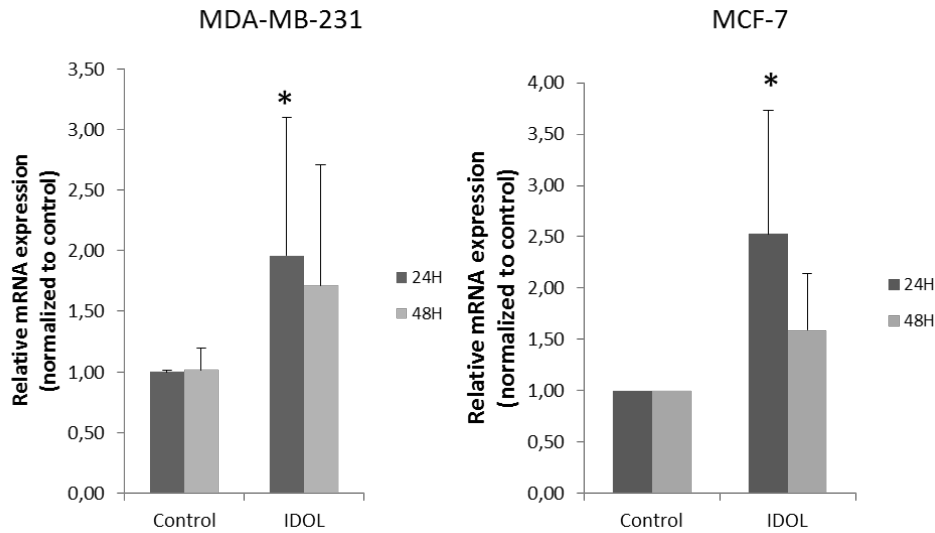


Figure 4. Effect of ergosterol on the mRNA expression of IDOL and LDLR in MDA-MB-231 and MCF-7 cells lines. Cells were incubated with 12.5 μ M of ergosterol for 24H and 48H. Results are expressed as the means \pm standard deviation with * $p < 0.05$.

Effect of Ergosterol on LXR and its target genes in THP-1 cells

The monocytic cell line THP-1 is frequently used as a macrophage cell model. The effect of ergosterol was evaluated on the expression of LXR and its target genes in this model (Fig.5). Treatment of differentiated THP-1 cells performed with 12.5 μ M and 25 μ M of ergosterol allowed a dose-dependant increase in mRNA expression of LXR (2.52 and 2.80), APOE (1.41 and 1.86), ABCA1 (3.20 and 4.12) and ABCG1 (7.39 and 10.23, $p < 0.08$) after 24 h compared to untreated cells.

THP-1

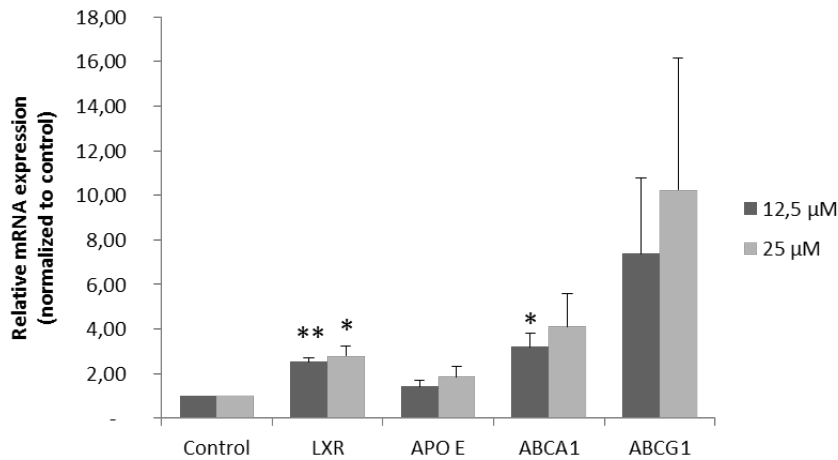


Figure 5. Effect of ergosterol on the mRNA expression of LXR and its target genes in THP-1 macrophages. After a differentiation of THP-1 cells into macrophage-like THP-1 human monocyte cells by exposure to phorbol 12-myristate-13-acetate, the cells were incubated with 12.5 μM or 25 μM of ergosterol for 24 h. The mRNA expression of LXR, APOE, ABCA1 and ABCG1 were determined by Real-time RT-PCR. RpbS6 RNA was used as a housekeeping gene. Results are expressed as the means ± standard deviation with ** $p < 0.01$ and * $p < 0.05$.

Effect of ergosterol on LXR mRNA expression in J774 cells

Treatment of J774 macrophages performed with 12.5 μM and 25 μM of ergosterol for 24H allowed a dose-dependent suggested an increase in mRNA expression of LXRα (1.5 and 1.7) compared to untreated cells, but without significant results.

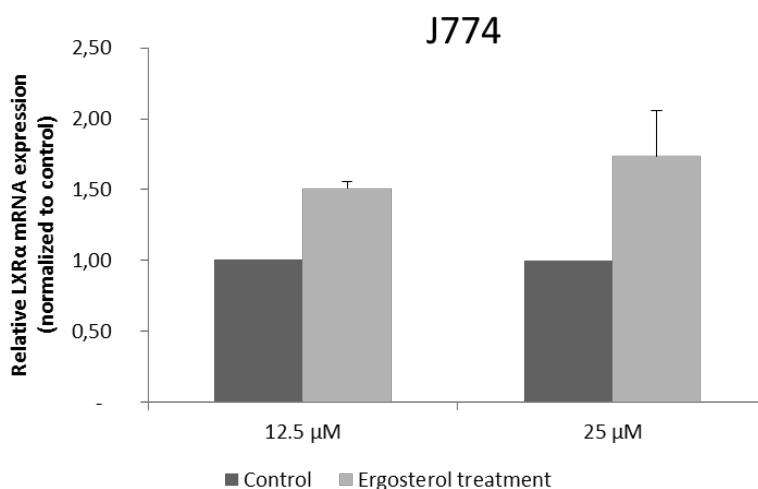


Figure 6. Effect of ergosterol on the mRNA expression of LXR α in J774 macrophages. Cells were treated with 12.5 μ M or 25 μ M of ergosterol for 24 h. The mRNA expression of LXR α was determined by Real-time RT-PCR. Actin rRNA was used as a housekeeping gene. Results are expressed as the means \pm standard deviation with * $p < 0.05$.

Effect of ergosterol on membrane rafts disruption of breast cancer cells

To evaluate the ergosterol effect on the membrane rafts of breast cancer cells, we studied the protein expression of flotillin-2 in MCF-7 cells and EGFR, flotillin-2 and caveolin-1 for MDA-MB-231. Western blot analysis showed that ergosterol treatment resulted in modifications of protein expression in fractions corresponding to membrane rafts (Fig. 7). It was observed an increase in the level of flotillin-2 protein expression in membrane rafts fractions F2 isolated from MCF-7 cells treated with ergosterol as compared with those of untreated cells. The data obtained for MDA-MB-231 revealed an important decrease of flotillin-2 and Caveolin 1 expression in fraction 2. However, western blot analysis shown that treatment of MDA-MB-231 with ergosterol resulted in a decrease of EGFR in Fraction 2 and an increase in fraction 4.

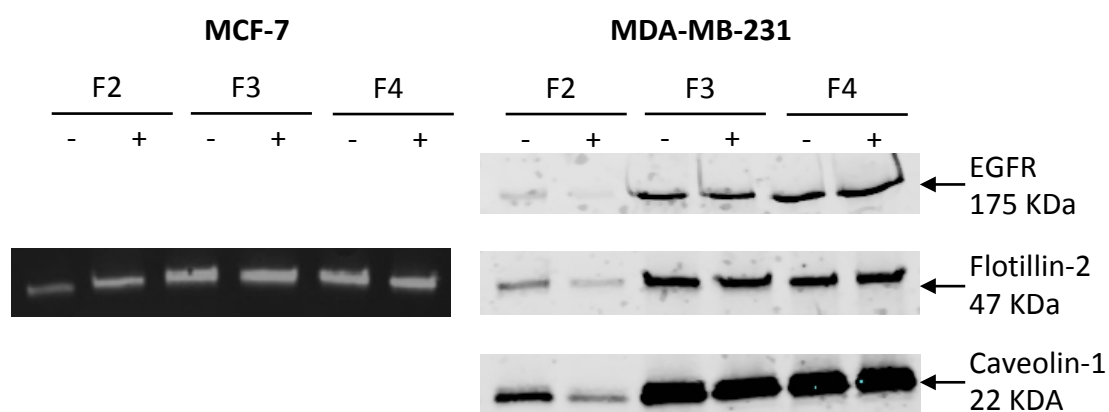


Figure 7. Effect of ergosterol on the expression of EGFR, flotillin-2 and Caveolin-1 in membrane raft fractions. Membrane rafts were isolated from MCF-7 and MDA-MB-231 using density-gradient ultracentrifugation. Then membrane raft fraction were separated by SDS-PAGE and analyzed by western blot using anti-flotillin-2, anti-EGFR and anti-caveolin-1

antibodies. A representative western blot showing protein expression in cancer cells treated 24H (+) or not treated (-) with ergosterol (12.5 μ M).

Effect of ergosterol on breast cancer cell viability

MCF-7 and MDA-MB-231 cells were treated with increasing concentrations (0.001-12.5 μ M) for 24, 48 and 72H and cell viability was measured by MTT assay. As shown in Fig.8, ergosterol reduced MCF-7 and MDA-MB-231 cancer cell viability in a dose-and time-dependent manner. The IC₅₀ values obtained for 24, 48 and 72H were 6.3 \pm 1 nM, 66.8 \pm 1 nM and 8.4 \pm 4 nM for MDA-MB-231 and 10.3 \pm 8 nM, 160.0 \pm 8 0.67 \pm 3 nM for MCF-7, respectively.

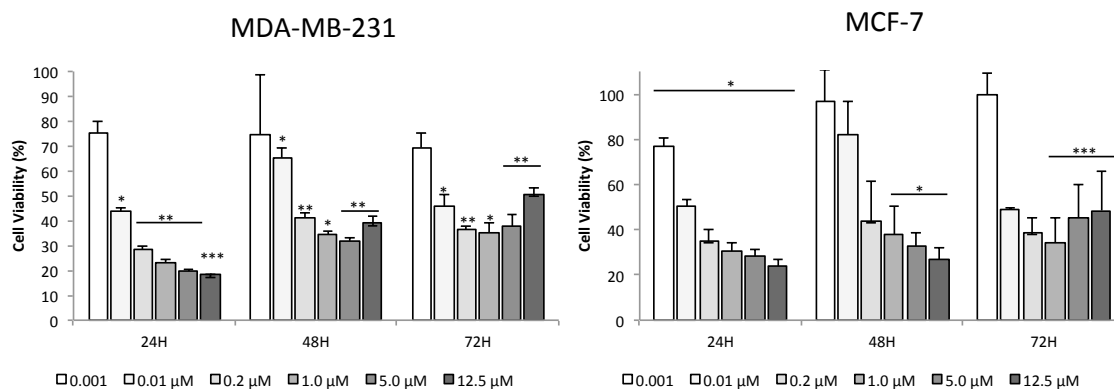


Figure 8. Cell viability of MDA-MB-231 and MCF-7 breast cancer cells after exposure to ergosterol. Cells were treated with various concentrations range from 0.001 to 12.5 μ M for 24, 48 and 72H. The cell viability was determined by MTT assay. Data are represented as mean \pm standard deviation with * p < 0.05, ** p < 0.01 and *** p < 0.001.

Effect of Ergosterol on AKT phosphorylation

Labeling MCF-7 cells by the antibody against Phospho-Akt allowed to evaluate the Akt phosphorylation of status. Results obtained from analysis performed using a flow cytometer, showed a decrease of Akt protein phosphorylation in MCF-7 cells treated with ergosterol (16%) compared to the control. (Fig.9).

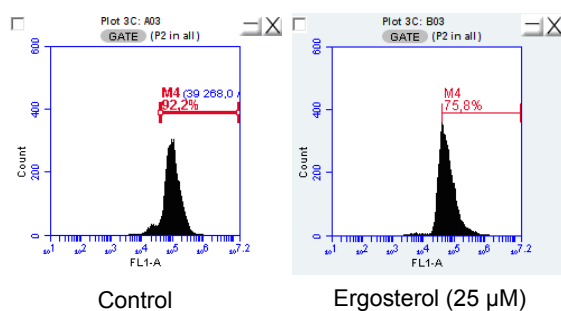


Figure 9. Effect of ergosterol on *p*-akt in MCF-7 breast cancer cells

Discussion

Ergosterol, is the major sterol produced by fungi. Although several ergosterol derivatives such as ergosterol peroxides [24–28], were evaluated as potential drugs in breast cancer treatment, ergosterol has rarely been valorised in cancer research studies. A recent study has shown *in-vitro* and *in-vivo* activities of ergosterol in breast cancer model. It was demonstrated that ergosterol reduced cell viability, induced apoptosis and up regulated expression of tumor suppressor in MDA-MB-231 cell line [20]. Deregulation of lipid metabolism associated with an accumulation of cellular cholesterol is a recognised hallmark in breast [29,30]. Cholesterol is needed for the biogenesis and maintenance of fluidity of cell membranes [31]. It is also a central component of lipid rafts, specialized microdomains of the plasma membrane that serve as organizing centers for the assembly of signalling molecules [9]. Therefore, rapidly proliferating cancer cells including breast cancer cell lines, with highly activated signal transductions have an enhanced requirement for cholesterol. The accumulation of cellular cholesterol can be explained by several modifications such as upregulation of its biosynthesis, downregulation of transporters involved in its efflux or increase of its uptake [32]. In this study, we aimed to evaluate whether the effect of

ergosterol on cholesterol metabolism in breast cancer cells has an impact on mammary cancer cells survival and on the organisation of rafts membranes.

Results obtained with GC-MS quantification and lipid droplets experiments revealed that ergosterol could reduce the cellular total and esterified cholesterol content in MCF-7 cell line. Lipid droplets (LDs), composed esterified cholesterol and other neutral lipids, are the sites where cells store excessive lipids, to ensure diverse biological functions [33]. High levels of LDs in cancer cells were demonstrated to be related with increased aggressiveness of cancer including breast cancer [34,35]. Here, we have shown that ergosterol reduces LDs. To provide an explanation for the cholesterol decrease in cells, caused by ergosterol, we studied the 3 major pathways involved in cellular cholesterol metabolism.

In our case, this decrease in cellular cholesterol content could be the consequence of an increase in the expression of ABC carriers, which could result in an increase in cholesterol efflux; and/or an increase in IDOL (E3 ubiquitin ligase inducible degrader of LDLR), a key regulator of LDL receptor[36]. In accordance with these hypotheses, we have demonstrated that ergosterol stimulates ABCG1 and IDOL in both cell lines. The effects of these two proteins on the proliferation of cancer cells are now well documented. Indeed, their genes are targets of the nuclear receptor LXR whose stimulation by some oxysterols (e.g. 22ROH-Cholesterol) or by synthetic molecules (e.g. TO901317) reduces cell proliferation [19]. Since ergosterol has stimulated the expressions of ABCG1 and IDOL, we have investigated whether it behaves as an LXR activator. For this purpose we have studied the effect of ergosterol in two macrophages cell lines THP1 and J774 (human and mouse), widely used cells to study LXR pathways and its target genes.

The increase of LXR, ABCG1, ABCA1 and ApoE genes expression in macrophage-differentiated THP-1 cells and ABCA1 in J774 cells suggest that ergosterol could be an activator of LXR.

Further studies are needed to confirm the interaction of ergosterol with LXR. Since ergosterol stimulated genes expression of ABC transporters and IDOL, we have hypothesized whether it would have an effect on membrane rafts. These latter are described as detergent-resistant microdomains of the cell membrane enriched in cholesterol and are involved in various cellular functions, including cell survival, proliferation and migration, which play an important role in cancer development and progression [9,37]. Flotillin-2 and Caveolin-1 are two membrane raft markers, they can interact with cholesterol as well as multiple signalling proteins involved in cancer progression [12,38].

Immunoblot performed on the membrane rafts isolated from the cell lysates by ultracentrifugation process, allowed to highlight a decrease of caveolin-1 and flotillin-2 proteins expression in membrane raft fraction of MDA-MB-231 cells, suggesting a disruption of membrane raft integrity.

The EGF receptor (EGFR), a transmembrane protein with tyrosine kinase activity [39], has been shown to be enriched in lipid rafts and its overexpressed in many cancer types including approximately 30% of breast cancer [40]. Localization of the EGF receptor to rafts appears to modulate its tyrosine kinase activity. Upon stimulation, EGFR can activate several signaling intermediates, including Akt, one of the most important signalling pathways activated in human cancer, leading to control of cell survival, proliferation, migration, angiogenesis and metastasis [41]. Consequently, the hypothesis that ergosterol could have an effect on membrane raft EGFR has also been tested. Our results show that ergosterol

reduced EGFR in lipid rafts in MDA-MB231, which could suggest a negative impact on cell survival signalling by reducing the relevant signalling pathways, for example Akt. As reported in the literature, EGFR expressing breast cancers are typically characterized as triple-negative breast cancers, which lack expression of estrogen receptor and progesterone receptor and do not contain HER2 amplification [41]. We also did not detect EGFR in our model MCF-7, which is an ER+ cancer cell line. Moreover, we did not detect Caveolin in MCF-7. Expression of caveolin-1 in the different breast cancer cell lines was controversial. with recent study suggested it was more related to basaloid cancers than estrogen or progesterone negative receptor or HER-2–expressing cancers [42].

Since some LXR activators have anti-proliferative effects cancer cells [43] and since ergosterol has stimulated IDOL and ABC transporters and disrupted rafts membranes, we have hypothesized that it could reduce cell survival. We showed that ergosterol exhibited a dose- and time-dependent cytotoxic effect on MCF-7 and MDA-MB-231 breast cancer cells with IC_{50} values for 10.3 ± 8 and 6.3 ± 1 nM. Moreover, we found that ergosterol decreased *p-Akt*

Conclusion

In conclusion, the fungal metabolite, ergosterol, can exert an anti-proliferative activity of mammary cancer cells by reducing the accumulation of intracellular cholesterol by stimulating the genes of the cholesterol efflux (ABC transporters) and reducing the genes of cholesterol uptake via IDOL. We have shown that the consequences of these ergosterol effects on cellular cholesterol metabolism could affect membrane rafts by reducing the EGFR, flotillin-2 and caveolin-2 markers and *p-AKT*, the key factor in cellular proliferation. This study suggests that breast cancer cell viability could be regulated by cholesterol

metabolism pathway and membrane raft integrity under the control of ergosterol. We thus suggest a potential therapeutic role of ergosterol in breast cancer cells.

Acknowledgment

The authors would like to thank the French Ministry and La Ligue contre le Cancer for the PhD grant for LO. They also thank Stéphanie Crossouard and Audrey Aguesse for the experiments performed on GC-MS and Alexandra Burghilea for technical participation

References

- [1] F. Bray, J. Ferlay, I. Soerjomataram, R.L. Siegel, L.A. Torre, A. Jemal, Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA. Cancer J. Clin.* 68 (2018) 394–424. doi:10.3322/caac.21492.
- [2] M. Sciacovelli, E. Gaude, M. Hilvo, C. Frezza, *The metabolic alterations of cancer cells*, 1st ed., Elsevier Inc., 2014. doi:10.1016/B978-0-12-416618-9.00001-7.
- [3] S. Beloribi-djefa, S. Vasseur, F. Guillaumond, Lipid metabolic reprogramming in cancer cells, *Oncogenesis*. 5 (2016) 1–10. doi:10.1038/oncsis.2015.49.
- [4] C.R. Santos, A. Schulze, Lipid metabolism in cancer: Lipid metabolism in cancer, *FEBS J.* 279 (2012) 2610–2623. doi:10.1111/j.1742-4658.2012.08644.x.
- [5] J. Long, C.-J. Zhang, N. Zhu, K. Du, Y.-F. Yin, X. Tan, D.-F. Liao, L. Qin, Lipid metabolism and carcinogenesis, cancer development., *Am. J. Cancer Res.* 8 (2018) 778–791. <http://www.ncbi.nlm.nih.gov/pubmed/29888102><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5992506>.

- [6] A. Pakiet, J. Kobiela, P. Stepnowski, T. Sledzinski, A. Mika, Changes in lipids composition and metabolism in colorectal cancer: A review, *Lipids Health Dis.* 18 (2019) 1–21. doi:10.1186/s12944-019-0977-8.
- [7] K. Simons, R. Ehehalt, Cholesterol, lipid rafts, and disease, *Biol. Biochem. Cholest.* 110 (2002) 597–603. doi:10.1172/JCI200216390.Lipid.
- [8] S. Munro, Lipid Rafts: Elusive or Illusive? Review, *Cell Press.* 115 (2003) 377–388.
- [9] F. Mollinedo, C. Gajate, Lipid rafts as major platforms for signaling regulation in cancer, *Adv. Biol. Regul.* 57 (2015) 130–146. doi:10.1016/j.jbior.2014.10.003.
- [10] I.S. Babina, S. Donatello, I.R. Nabi, A.H. Hopkins, Lipid Rafts as master regulators of breast cancer cell function, 2011. doi:10.5772/21324.
- [11] X. Wang, Q. Yang, L. Guo, X.H. Li, X.H. Zhao, L.B. Song, H.X. Lin, Flotillin-2 is associated with breast cancer progression and poor survival outcomes, *J. Transl. Med.* 11 (2013) 1–10. doi:10.1186/1479-5876-11-190.
- [12] X.L. Qian, Y.H. Pan, Q.Y. Huang, Y.B. Shi, Q.Y. Huang, Z.Z. Hu, L.X. Xiong, Caveolin-1: A multifaceted driver of breast cancer progression and its application in clinical treatment, *Onco. Targets. Ther.* 12 (2019) 1539–1552. doi:10.2147/OTT.S191317.
- [13] J.W. Clendening, A. Pandyra, P.C. Boutros, S. El Ghamrasni, F. Khosravi, G.A. Trentin, A. Martirosyan, A. Hakem, R. Hakem, I. Jurisica, L.Z. Penn, Dysregulation of the mevalonate pathway promotes transformation, *Proc. Natl. Acad. Sci. U. S. A.* 107 (2010) 15051–15056. doi:10.1073/pnas.0910258107.
- [14] S. Borgquist, O. Bjarnadottir, S. Kimbung, T.P. Ahern, Statins: a role in breast cancer therapy?, *J. Intern. Med.* 284 (2018) 346–357. doi:10.1111/joim.12806.

- [15] J. Neumann, D. Rose-Sperling, U.A. Hellmich, Diverse relations between ABC transporters and lipids: An overview, *Biochim. Biophys. Acta - Biomembr.* 1859 (2017) 605–618. doi:10.1016/j.bbamem.2016.09.023.
- [16] L. Yvan-charvet, N. Wang, A.R. Tall, The role of HDL, ABCA1 and ABCG1 transporters in cholesterol, *Arter. Thromb Vasc Biol.* 30 (2010) 139–143. doi:10.1161/ATVBAHA.108.179283.The.
- [17] J.L. Goldstein, M.S. Brown, The LDL receptor, *Arterioscler. Thromb. Vasc. Biol.* 29 (2009) 431–438. doi:10.1161/ATVBAHA.108.179564.
- [18] N. Zelcer, C. Hong, R. Boyadjian, P. Tontonoz, LXR regulates cholesterol uptake through Idol-dependant ubiquitination of the LDL receptor, *Science* (80-.). 325 (2009) 100–104. doi:10.1038/jid.2014.371.
- [19] A. El Roz, J.-M. Bard, J.-M. Huvelin, H. Nazih, LXR agonists and ABCG1-dependent cholesterol efflux in MCF-7 breast cancer cells: relation to proliferation and apoptosis, *Anticancer Res.* 32 (2012) 3007–3013. <http://ar.iiarjournals.org/content/32/7/3007.short> (accessed October 7, 2015).
- [20] X. Li, Q. Wu, Y. Xie, Y. Ding, W.W. Du, M. Sdiri, B.B. Yang, Ergosterol purified from medicinal mushroom *Amauroderma rude* inhibits cancer growth in vitro and in vivo by up-regulating multiple tumor suppressors, *Oncotarget.* 6 (2015) 17832–17846.
- [21] M.T.R. Subbiah, W. Abplanalp, Ergosterol (major sterol of baker's and brewer's yeast extracts) Inhibits the growth of human breast cancer cells in vitro and the potential role of its oxidation products, *Int. J. Vitam. Nutr. Res.* 73 (2003) 19–27.
- [22] A. Gil-Ramírez, A. Ruiz-Rodríguez, F.R. Marín, G. Reglero, C. Soler-Rivas, Effect of

- ergosterol-enriched extracts obtained from *Agaricus bisporus* on cholesterol absorption using an in vitro digestion model, *J. Funct. Foods.* 11 (2014) 589–597. doi:10.1016/j.jff.2014.08.025.
- [23] J. Elia, D. Carbonnelle, C. Logé, L. Ory, J. Huvelin, M. Tannoury, M. Diab-assaf, K. Petit, H. Nazih, 4-cholesten-3-one decreases breast cancer cell viability and alters membrane raft-localized EGFR expression by reducing lipogenesis and enhancing LXR-dependent cholesterol transporters, *Lipids Health Dis.* 18 (2019) 1–15.
- [24] M. Govindharaj, S. Arumugam, G. Nirmala, M. Bharadwaj, K. Murugiyar, Effect of marine basidiomycetes *fulvifomes* sp.-derived ergosterol peroxide on cytotoxicity and apoptosis induction in MCF-7 cell line, *J. Fungi.* 5 (2019). doi:10.3390/jof5010016.
- [25] B.J. Grattan, Plant sterols as anticancer nutrients: Evidence for their role in breast cancer, *Nutrients.* 5 (2013) 359–387. doi:10.3390/nu5020359.
- [26] T.A. Woyengo, V.R. Ramprasath, P.J.H. Jones, Anticancer effects of phytosterols, *Eur. J. Clin. Nutr.* 63 (2009) 813–820. doi:10.1038/ejcn.2009.29.
- [27] M. Shahdaat, B. Sayeed, S.S. Ameen, M. Shahdaat, B. Sayeed, S.S. Ameen, Beta-Sitosterol : A promising but orphan nutraceutical to fight against cancer, *Nutr. Cancer.* 0 (2015) 1–7. doi:10.1080/01635581.2015.1087042.
- [28] M.M. Martínez-Montemayor, T. Ling, I.J. Suárez-Arroyo, G. Ortiz-Soto, C.L. Santiago-Negrón, M.Y. Lacourt-Ventura, A. Valentín-Acevedo, W.H. Lang, F. Rivas, Identification of biologically active *ganoderma lucidum* compounds and synthesis of improved derivatives that confer anti-cancer activities in vitro, *Front. Pharmacol.* 10 (2019). doi:10.3389/fphar.2019.00115.

- [29] D. Gonzalo-calvo, D. De Gonzalo-calvo, L. López-vilaró, L. Nasarre, M. Perez-olabarria, T. Vázquez, Intratumor cholesteryl ester accumulation is associated with human breast cancer proliferation and aggressive potential: a molecular and clinicopathological study Intratumor cholesteryl ester accumulation is associated with human breast cancer proliferat, *BMC Cancer*. (2015). doi:10.1186/s12885-015-1469-5.
- [30] J. Baumann, C. Sevinsky, D.S. Conklin, Lipid biology of breast cancer, *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids*. 1831 (2013) 1509–1517. doi:10.1016/j.bbalip.2013.03.011.
- [31] O.F. Kuzu, M.A. Noory, G.P. Robertson, The role of cholesterol in cancer, *Cancer Res*. 76 (2016) 2063–2070. doi:10.1158/0008-5472.CAN-15-2613.
- [32] P.M.R. Cruz, H. Mo, W.J. McConathy, N. Sabnis, A.G. Lacko, The role of cholesterol metabolism and cholesterol transport in carcinogenesis: A review of scientific findings, relevant to future cancer therapeutics, *Front. Pharmacol*. 4 (2013) 1–7. doi:10.3389/fphar.2013.00119.
- [33] C.M. Bunney, P. E., Zink, A. N., Holm, A. A., Billington, C. J., & Kotz, Lipid droplet functions beyond energy storage, *Biochim Biophys Acta*. 1862 (2017) 1260–1272. doi:10.1016/j.bbalip.2017.07.006.
- [34] T. Petan, E. Jarc, M. Jusović, Lipid droplets in cancer: Guardians of fat in a stressful world, *Molecules*. 23 (2018). doi:10.3390/molecules23081941.
- [35] C. Blücher, C. Zilberfain, T. Venus, N. Spindler, A. Dietrich, R. Burkhardt, S.C. Stadler, I. Estrela-Lopis, Single cell study of adipose tissue mediated lipid droplet formation and biochemical alterations in breast cancer cells, *Analyst*. (2019) 1–13.

doi:10.1039/c9an00816k.

- [36] A.J. Brown, J. Hsieh, Foiling IDOL to help control cholesterol Andrew, *Circ.* 176 (2017) 139–148. doi:10.1016/j.physbeh.2017.03.040.
- [37] T. Murai, The role of lipid rafts in cancer cell adhesion and migration, *Int. J. Cell Biol.* 2012 (2012). doi:10.1155/2012/763283.
- [38] X. Wang, Q. Yang, L. Guo, X.H. Li, X.H. Zhao, L.B. Song, H.X. Lin, Flotillin-2 is associated with breast cancer progression and poor survival outcomes, *J. Transl. Med.* 11 (2013) 1. doi:10.1186/1479-5876-11-190.
- [39] Z. Weihua, R. Tsan, W.C. Huang, Q. Wu, C.H. Chiu, I.J. Fidler, M.C. Hung, Survival of cancer cells is maintained by EGFR independent of its kinase activity, *Cancer Cell.* 13 (2008) 385–393. doi:10.1016/j.ccr.2008.03.015.
- [40] M.E. Irwin, H.L. Mueller, N. Bohin, Y. Ge, Lipid raft localization of EGFR alters the response of cancer cells to the EGFR tyrosine kinase inhibitor gefitinib, *J. Cell. Physiol.* 226 (2008) 2310–2328. doi:10.1002/jcp.22570.
- [41] P. Wee, Z. Wang, Epidermal growth factor receptor cell proliferation signaling pathways, *Cancers (Basel).* 9 (2017) 1–45. doi:10.3390/cancers9050052.
- [42] Y. Liu, R. Sun, W. Wan, J. Wang, J.J. Oppenheim, L. Chen, N. Zhang, The involvement of lipid rafts in epidermal growth factor-induced chemotaxis of breast cancer cells, *Mol. Membr. Biol.* 24 (2007) 91–101. doi:10.1080/10929080600990500.
- [43] C.-P. Chuu, H.-P. Lin, Antiproliferative effect of LXR agonists T0901317 and 22 (R)-hydroxycholesterol on multiple human cancer cell lines, *Anticancer Res.* 30 (2010) 3643–3648.

4. Résultats complémentaires sur l'effet antiprolifératif de dérivés de l'ergostérol sur nos modèles cancéreux mammaires.

Afin de mettre en évidence l'activité potentielle de stérols issus de champignons filamenteux marins autres que la molécule d'ergostérol, une étude comparative a été réalisée sur l'effet antiprolifératif de différents dérivés de cette molécule. Ainsi, en plus de l'anicequol (**7**) isolé de la souche *P. chrysogenum*, cinq composés obtenus lors d'une collaboration avec un partenaire portugais (Institut des sciences biomédicales d'Abel Salazar (ICBAS)) ont été testés sur la prolifération de MCF-7 (0,001 -50 μ M) avec le test MTT: Ergosta-4,6,8 (**14**), 22-tétraén-3-one (**2**), Ergostérol-5,8 endopéroxyde (**3**), Cyathistérone (ergosta-7,22-diène-3,6-dione) (**4**), (3 β , 5 α , 22E) -3,5-Dihydroxyergosta-7,22-dien-6-one (**5**), l'acide helvolique (**6**) ; avec le 22-hydroxycholestérol (**8**) comme composé de référence (Cf. Figure 57).

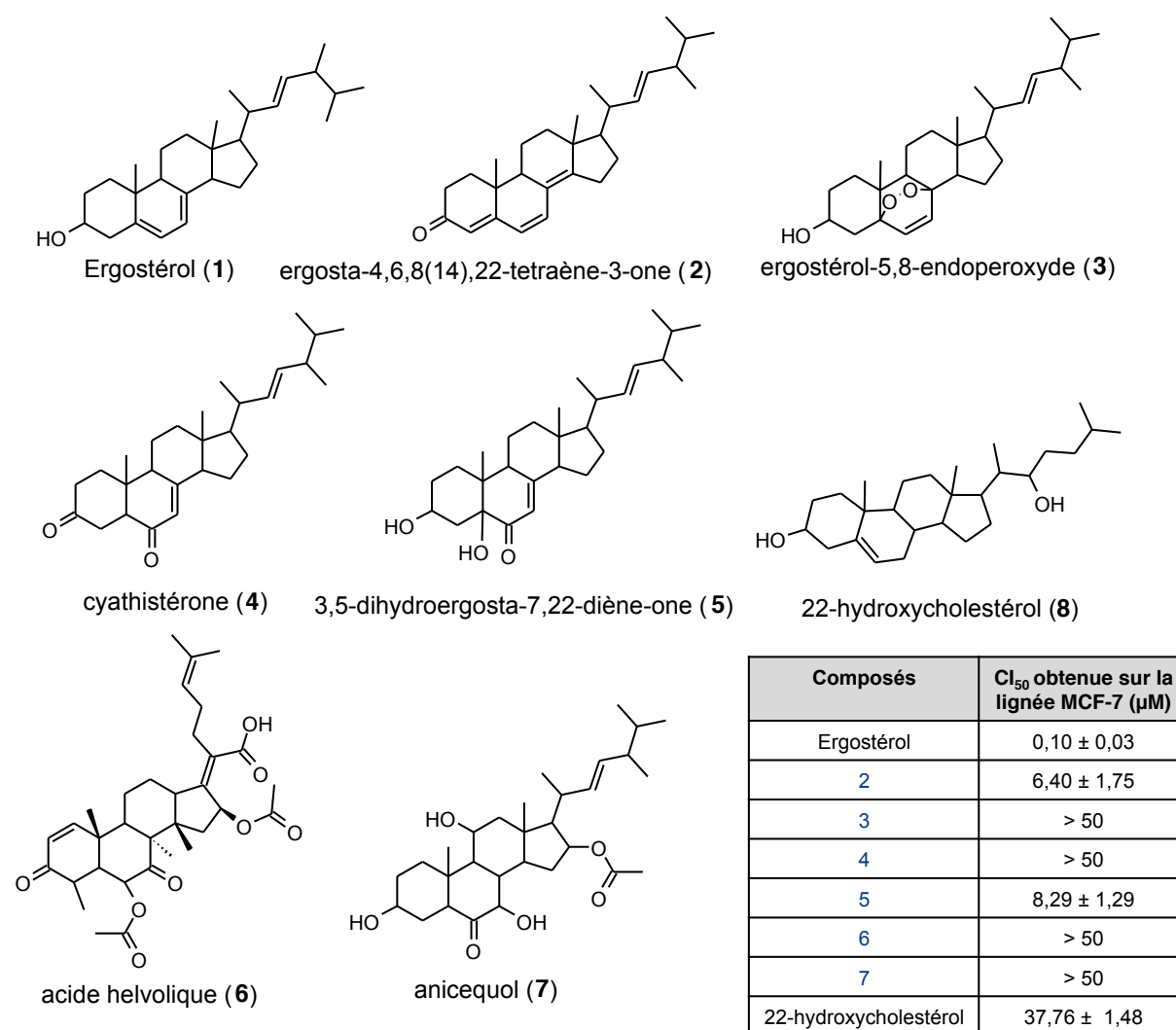


Figure 57. Evaluation de l'activité antiproliférative des stérols et dérivés des stérols fongiques sur les cellules MCF-7.

Un effet dose-dépendant a été observé pour les dérivés **1** à **5** de l'ergostérol avec une CI_{50} de 6,40 μ M pour le composé (**2**) et 8,29 μ M pour le composé (**5**), mais leurs activités étaient inférieures à celles de l'ergostérol (CI_{50} de 0,10 μ M). Le composé (**2**) a été rapporté dans la littérature pour son activité diurétique, son activité inhibitrice dans la production d'oxyde nitrique, immunosuppressive et récemment pour sa capacité à induire l'arrêt du cycle G2/M ainsi que l'apoptose sur des cellules de carcinome hépatocellulaire humain (HepG213).

Pour les composés (**3**) et (**4**), il a été suggéré une inhibition de la prolifération des cellules, mais la gamme de concentration testée était trop basse pour déterminer une valeur d' IC_{50} . Cependant, l'activité de l'Ergostérol-5,8 endopéroxyde a déjà été reportée dans la littérature sur différentes lignées cellulaires : HepG2, SK-Hep1, MCF-7 et MDA-MB-231 avec une CI_{50} de 3,15-18,34-17,50 et 15,24 μ M, respectivement. En revanche, l'acide helvolique (**6**) n'a montré aucune activité sur la lignée cellulaire MCF-7.

D'après ces premiers résultats, il semble difficile d'établir des relations structure-activité. Bien que les produits **1-5** et **7** soient tous des dérivés de l'ergostérol, il ne semble pas y avoir de consensus sur l'activité de ce type de molécules sur la lignée MCF-7. Si on compare les molécules **4** et **5**, il semble que la présence d'un hydroxyle en position 3 (au lieu d'une cétone) améliore l'activité. Toutefois la molécule **2** présentant une cétone en position 3 est quant-à elle bien active avec une CI_{50} à 6,40 μ M. Il serait intéressant de tester cette molécules après réduction de la cétone en 3 en hydroxyle. L'absence d'activité pour l'anicequol (**6**) et l'acide helvolique (**7**) semblent montrer que la présence de l'acétate en 16 n'amène rien de plus en termes d'inhibition de prolifération. Au regard des 3 molécules les plus actives (**1**, **2** et **5**), on peut aussi se poser la question de l'importance de la présence de double liaisons conjuguées sur les positions 3, 4, 5, 6, 7, 8 et 14. Une investigation plus approfondie d'autres dérivés de l'ergostérol, complétée par des études de docking *in silico* serait nécessaire afin de mieux comprendre les relations structure-activité pour ce type de molécules en rapport la prolifération des cellules cancéreuses mammaires.

5. Conclusion

L'ergostérol est majoritairement produit par les espèces fongiques et est le stérol le plus abondant dans les membranes de leurs cellules. Ce composé, isolé et identifié à maintes reprises par différentes approches de fractionnement bioguidé et de métabolomique dans la souche MMS5 *Penicillium chrysogenum* a déjà été reporté dans la littérature pour son effet cytotoxique sur les cellules cancéreuses mammaires (Li *et al.*, 2015). En revanche, l'étude de celui-ci sur la régulation du métabolisme lipidique des cellules cancéreuses n'a jamais été réalisée.

Les altérations du métabolisme des lipides, particulièrement celui du cholestérol, sont communes à toutes les cellules cancéreuses y compris les cellules cancéreuses du sein. Cette altération touche l'ensemble des voies de régulation assurant le maintien de l'homéostasie lipidique à savoir la synthèse (Baenke *et al.*, 2013; Santos and Schulze, 2012), l'efflux, la dégradation ou encore le stockage des lipides. Le métabolisme lipidique constitue alors une cible thérapeutique d'intérêt dans le traitement du cancer du sein.

Bien que la souche *P. chrysogenum* ait été sélectionnée au départ pour son activité antiproliférative sur les cellules MCF-7 et son activité sur l'expression de PPAR, les premières investigations menées sur l'effet de l'ergostérol sur les voies de signalisation impliquant les acides gras, tel que la synthèse ou la bêta-oxydation n'ont révélé aucune modification majeure. Ainsi, il a été suggéré que l'ergostérol ne fait pas partie des molécules produites par la souche *P. chrysogenum* qui sont capables d'activer l'expression des récepteurs nucléaires PPAR. Nous nous sommes donc intéressés à d'autres voies de régulation du métabolisme lipidique telles que la lipogenèse, la biosynthèse du cholestérol ainsi que son transport (efflux et captation).

Dans cette étude, notre objectif était d'évaluer si l'effet de l'ergostérol sur le métabolisme du cholestérol dans les cellules cancéreuses du sein avait un impact sur la survie des cellules cancéreuses mammaires et sur l'organisation des radeaux des membranes. Ainsi, il a pu être démontré par analyse GC-MS et des expériences de marquage de gouttelette lipidique qu'un traitement de nos modèles de cellules cancéreuses mammaires réalisé avec l'ergostérol entraînait une réduction du contenu cellulaire en cholestérol total et estérifié.

Pour expliquer la diminution du cholestérol dans les cellules, causée par l'ergostérol, nous avons étudié les 3 principales voies impliquées dans le métabolisme du cholestérol cellulaire. D'après les résultats obtenus, seules des modifications sur l'expression d'un transporteur membranaire impliqué dans l'efflux du cholestérol et d'un inducteur de dégradation du récepteur responsable de l'influx du cholestérol ont pu être soulignées. Cependant, aucune corrélation directe n'a pu être effectuée entre l'expression des ARNm d'ABCG1 et d'IDOL avec la diminution du cholestérol observé dans nos modèles cellulaires.

Il peut simplement être noté que ces deux phénomènes coexistent lorsque les cellules ont été incubées avec l'ergostérol. Ainsi, des études complémentaires devraient être réalisées sur d'autres modèles cellulaires, telles que les cellules épithéliales (ex : Caco-2) ou endothéliales (ex : HepG2) exprimant ces deux protéines à l'état basal afin d'évaluer leurs expressions membranaires, par immunofluorescence ou par immunoblot, en présence de l'ergostérol. Le mécanisme d'action de cette molécule sur la régulation du métabolisme lipidique n'est donc pas complètement défini.

Le transporteur ABCG1 et IDOL étant sous le contrôle de LXR, une autre piste a pu être investiguée en parallèle. L'augmentation de l'expression d'ARNm de ce récepteur nucléaire ainsi que de ces gènes cibles ABCG1, ABCA1 et ApoE (tous impliqués dans l'efflux du cholestérol) dans le modèle cellulaire THP-1 a suggéré que l'ergostérol pouvait être un activateur de LXR.

Afin d'apporter des preuves supplémentaires sur le rôle potentiel de l'ergostérol dans la régulation du métabolisme lipidique via une activation du récepteur nucléaire LXR, il a été envisagé d'étudier l'expression de deux stérolines ABCG5 et ABCG8. Ces deux semi-transporteurs impliqués dans la régulation de l'absorption du cholestérol en induisant son excrétion des cellules cibles, ont été rapportés pour être stimulés suite à une activation du récepteur nucléaire LXR par certains stérols (phytostérols et oxystérols). D'après cette observation, si l'ergostérol est un activateur de LXR, la diminution du contenu cellulaire en cholestérol pourrait aussi s'expliquer par l'augmentation des stérolines, ABCG5 et ABCG8. Or, ces deux protéines ne sont pas bien exprimées à l'état basal dans nos modèles cellulaires MCF-7 et MDA-MB-231 (résultats non montrés). En effet, aucune influence de l'ergostérol n'a pu être montrée sur l'expression de ces deux transporteurs, probablement en raison de leur faible expression. Il est indispensable de réaliser d'autres expériences sur des modèles exprimant ces deux protéines (Caco-2 et HepG2 par exemple) pour évaluer l'effet de l'ergostérol sur leur expression.

Ainsi, pour valider que l'effet de l'ergostérol est dépendant de LXR, il serait intéressant de bloquer ce récepteur par un antagoniste (Elia *et al.*, 2019) ou d'inhiber son expression par un siRNA (Griffett *et al.*, 2015) spécifique dans les lignées cellulaires d'intérêt et d'évaluer par la suite les effets provoqués par ce stérol.

Des études mécanistiques ont également été réalisées dans le but d'évaluer l'effet de l'ergostérol sur l'intégrité des radeaux lipides. Ces microdomaines résistants aux détergents de la membrane cellulaire particulièrement enrichie en cholestérol sont impliqués dans diverses fonctions cellulaires jouant un rôle important dans le développement et la progression du cancer (Mollinedo and Gajate, 2015). Parmi les protéines exprimées au sein de ces radeaux lipidiques il a été choisi d'étudier l'expression de la flotillin-2, de la caveolin-1 et de l'EGFR, toutes rapportées pour être impliquées dans la progression tumorale du cancer du sein (Liu *et al.*, 2007; Qian *et al.*, 2019; Wang *et al.*, 2013; De Weille *et al.*, 2013).

D'après nos travaux, la diminution de l'expression protéique des biomarqueurs flotilline-2, caveolin-1 et EGFR observée dans les cellules MDA-MB-231 a suggéré que l'ergostérol entraîne une dérégulation de ces plateformes de signalisation.

Le récepteur de l'EGF (EGFR) à activité tyrosine kinase (Weihua *et al.*, 2008), étant décrit pour son implication dans différentes voies de signalisation, dont celle de l'Akt, impliquée dans la prolifération des cellules cancéreuses. Une légère diminution d'Akt phosphorylée (16%) a été montrée sur les lysats totaux du modèle cellulaire MCF-7 traité par l'ergostérol. Il serait intéressant d'évaluer l'effet de ce stérol sur Akt totale et Akt phosphorylée. Suivant les résultats les investigations pourraient être poursuivies sur les voies de signalisation des kinases comme celle des PI3K/AKT, une cible actuelle dans le traitement du cancer du sein.

Cette étude suggère que l'ergostérol pourrait réduire la viabilité des cellules cancéreuses du sein par son action sur la voie du métabolisme du cholestérol et sur l'intégrité des radeaux membranaire. D'autres études sont nécessaires pour élucider les mécanismes par lesquels ce stérol pourrait influencer l'efflux et l'influx du cholestérol et le lien de ces deux phénomènes avec les protéines des rafts (EGFR, flotilline-2 et Cavéoline-1) et la survie cellulaire. Bien que ces premiers résultats soient préliminaires, il est tentant de proposer des études *in-vivo* chez l'animal afin de suggérer à l'ergostérol un rôle thérapeutique anti-tumoral potentiel de l'ergostérol

D'après la littérature, les études réalisées sur des modèles *in vivo* de tumeurs mammaires ont montré des effets bénéfiques de l'administration pharmacologique de l'ergostérol (Li *et al.*, 2015). Néanmoins, aucune étude *in vivo* réalisée avec ce stérol fongique sur le métabolisme lipidique et cancer. Nous sommes convaincus que des expérimentations *in vivo* sur des modèles murins de xéno greffes de cellules cancéreuses mammaires sont nécessaires pour valider le potentiel préventif et/ou curatif de l'ergostérol. Ceci pourrait se faire de deux manières, une administration orale ((Li *et al.*, 2015) ou une supplémentation à un régime alimentaire.

Conclusions générales et perspectives

Le projet de recherche mené au cours de cette thèse a été réalisé dans le but d'identifier des produits naturels fongiques marins bioactifs sur des modèles cellulaires du cancer du sein. Comme il a pu être exposé à la présentation de ce projet, il existe quatre grandes étapes dans la méthodologie à suivre dans la découverte de nouveaux produits naturels : La sélection de l'organisme et de la cible biologique, la recherche des composés bioactifs, l'isolement et enfin la valorisation. Les travaux ont donc été réalisés en suivant cette méthodologie, selon trois axes majeurs dans le but de répondre à diverses interrogations qui sont à l'origine de la création de ce projet :

- 1) La collection de champignons marins filamenteux du laboratoire a-t-elle du potentiel dans le traitement du cancer du sein ? Si oui, quelle(s) souche(s) étudier ?
- 2) Quelle semble être la meilleure stratégie dans l'isolement de nouvelles molécules bioactives ?
- 3) Parmi les molécules cytotoxiques isolées, est ce qu'il existe un lien entre l'inhibition de la prolifération cellulaire observée et la régulation du métabolisme lipidique des cellules cancéreuses mammaires ? Si oui, quels sont les mécanismes impliqués ?

Ainsi ces travaux ont débuté par une première étape de criblage réalisé selon deux essais biologiques dans le but d'évaluer le potentiel de la collection et de sélectionner les souches à étudier dans le cadre de ce projet. Les résultats obtenus au cours de celui-ci ont révélé l'intérêt potentiel de 73 extraits (correspondant à 14 souches) possédant une activité intéressante avec un pourcentage d'inhibition de la prolifération des cellules supérieur à 50%. Ces résultats, incriminés de ceux d'un second test biologique permettant d'évaluer l'expression de deux récepteurs nucléaires ciblés dans le traitement du cancer du sein LXR et PPAR; ont permis d'affiner la sélection en ne choisissant que les extraits possédant les deux activités. D'après l'ensemble des résultats, il a ensuite été choisi d'étudier les souches *P. chrysogenum* et *D. americana* possédant toutes deux un effet antiprolifératif d'environ 65% à une concentration de 10 µg/mL, ainsi qu'une forte augmentation de l'expression des récepteurs PPARα (> 4,5) et PPARγ (> 2) et qui n'ont pas encore fait l'objet de travaux de recherche au sein du laboratoire.

Ce criblage, a permis d'une part de souligner l'intérêt de réaliser une approche OSMAC lors d'une étude de ce type, afin d'accroître la diversité chimique des métabolites produits par ces organismes vivants et de favoriser la production de molécules originales et bioactives. D'autre part, celui-ci a démontré l'intérêt potentiel de la collection du laboratoire dans le traitement du cancer du sein. Toutefois, ces résultats sont à modérer au vu de l'activité importante de l'ergostérol sur la cible choisie. Il semblerait donc intéressant dans les perspectives de ce travail d'affiner ce criblage par une quantification de l'ergostérol dans l'ensemble des extraits afin d'orienter la sélection des sujets d'étude vers des extraits pauvres en cette molécule. En effet, la valorisation de la collection d'extraits sur cette cible sera plus facile si on focalise sur des molécules non décrites dans la littérature.

Un deuxième volet a été consacré à la réalisation d'investigations chimiques sur les extraits des deux souches sélectionnées dans le but d'identifier les métabolites responsables de l'activité observée. Pour ce faire, différentes stratégies ont été mises en place et comparées afin d'évaluer leur efficacité dans l'identification de nouvelles molécules bioactives. L'approche du fractionnement bioguidé réalisée sur la souche *P. chrysogenum* a conduit à l'isolement de l'ergostérol comme molécule responsable de l'activité observée. C'est d'ailleurs la seule molécule présentant une activité qui a pu être isolée pour cette souche, ce qui prouve l'efficacité de cette approche. En parallèle, le profilage métabolique et la déréplication réalisés sur cette même souche ont permis d'annoter différents métabolites et de mettre en évidence l'originalité potentielle de certaines molécules en l'absence de hit avec la base de données du DNP. La purification ciblée a conduit au final à l'isolement de 5 molécules pures dont trois ont pu être identifiées par analyses spectrales RMN : la méléagrine, l'andrastine A, l'anicequol ; et deux nouvelles molécules potentielles dont les élucidations structurales respectives sont en cours. Concernant les investigations chimiques menées sur la seconde souche *D. americana*, il n'a malheureusement pas pu être identifié de molécules responsables de l'activité.

D'après ces travaux il a pu être mis en évidence un certain nombre d'avantages et de limitations liés à l'utilisation de ces approches. En effet, le fractionnement bioguidé a pu mener comme dans notre cas à l'isolement d'une molécule bioactive mais sans originalité *in fine*. D'autre part, le profilage métabolique LC-UV-HRMS et l'étude déréplicative ont permis d'accéder rapidement aux caractéristiques chimiques des molécules détectées sous leurs formes ionisées (avec leur temps de rétention, leur absorption UV, leur masse monoisotopique et l'intensité associée) mais sans idées sur leur activité biologique.

L'utilisation de ces deux approches menées en parallèle, a conduit à l'annotation de molécules qui ont pu être évaluées en comparaison avec la littérature pour leur potentiel d'un point de vue chimique (structure nouvelle) et/ou biologique (activité non décrite).

La complémentarité évidente de ces approches nous a conduit à développer et proposer une nouvelle méthodologie alliant fractionnement, mesure de l'activité des fractions et profilage métabolique par LC-MS afin de permettre *in fine* d'évaluer simultanément les potentiels biologiques et chimiques d'un extrait actif. Cette méthodologie comprend une étape de microfractionnement exploratoire sur micro colonne SPE ainsi que la combinaison de quatre modèles statistiques. Selon, cette approche de biochimométrie de plus en plus développée au sein de la communauté des produits naturels, un script « R FiBiCo » a été conçu sous le logiciel de programmation R. Ce dernier appliqué à la matrice de données comprenant les aires sous pics de chacun des ions détectés au sein des profils LC-MS et le pourcentage d'inhibition de chacune des fractions obtenues au cours du microfractionnement, a prouvé son efficacité. Il a en effet mis en évidence des ions correspondant à l'ergostérol et donc son implication dans l'activité antiproliférative de l'extrait initial.

Cette méthode appliquée à d'autres extraits actifs de la collection devrait donc permettre d'identifier plus rapidement les molécules bioactives et, après annotation, de cibler les isollements à venir sur des molécules bioactives et nouvelles. Toutefois, cette approche reste dépendante de la capacité d'ionisation des molécules et de leur ionisation en elle-même au sein de la source. En effet, dans le cas de l'extrait de MMS5 *P. chrysogenum*, si cette méthode avait été appliquée dès le départ, elle n'aurait pas permis d'anticiper la présence de l'ergostérol pourtant majoritairement présent au sein de l'extrait brut, puisque les ions observés pour cette molécule après ionisation par electrospray ne correspondent pas à des adduits classiquement observés rendant l'annotation plus difficile. Néanmoins, il est à noter que ces phénomènes observés pour l'ergostérol ne sont pas si fréquents.

Quelques optimisations de la méthodologie proposée pourraient être envisagées dans les perspectives de ce travail. D'une part, l'étape du microfractionnement pourrait être standardisée afin de s'appliquer à toutes sortes d'extraits. D'autre part, afin de faciliter la visualisation des résultats et la déréduplication, il pourrait être envisagé d'inclure les scores des analyses statistiques au sein d'un réseau moléculaire (carte de similarité de spectres MS²). De cette façon, sur le même graphe, on pourrait à la fois visualiser pour chaque pic son importance dans l'activité de l'extrait et sa similarité spectrale avec des produits connus et donc déduire son potentiel de nouveauté chimique.

Pour compléter ce travail, il serait intéressant d'appliquer la méthodologie proposée à d'autres extraits notamment ceux de la deuxième souche *Dydimella americana*, finalement peu étudiée au sein de ce projet, et/ou ceux ayant démontré une activité antiproliférative au cours du criblage initial après avoir vérifié qu'ils ne contenaient pas d'ergostérol.

Enfin, l'ensemble des approches réalisées sur l'extrait de la souche MMS5 *P. chrysogenum* ayant permis de mettre en évidence l'ergostérol comme métabolite bioactif sur la prolifération des cellules cancéreuses mammaires de la lignée MCF-7, des investigations mécanistiques ont été menées sur sa valorisation potentielle comme agent thérapeutique dans le traitement du cancer du sein. Il est à noter qu'aucun des autres dérivés stéroïdiques isolés et testés dans le cadre de cette étude n'a montré une activité aussi importante que celle de l'ergostérol.

La prolifération incontrôlée des cellules cancéreuses étant caractérisée par un dérèglement du métabolisme lipidique notamment celui du cholestérol, la recherche de composés ciblant celui-ci constitue alors une cible thérapeutique prometteuse. Au regard des effets notables des stérols (oxystérols et phytostérols) rapportés sur les voies de régulation du métabolisme lipidique, il a donc été envisagé naturellement d'évaluer l'effet potentiel de l'ergostérol sur celui-ci. Parmi les différentes investigations menées *in vitro*, il semblerait que l'ergostérol en plus de son effet antiprolifératif, ait une action sur la régulation de l'homéostasie du cholestérol.

En effet, une diminution du cholestérol cellulaire total et estérifié a pu être observée. D'après les résultats préliminaires, il semblerait que cette diminution soit associée à une augmentation de l'expression d'un transporteur de l'efflux ABC et d'un inducteur de dégradation du récepteur LDLR responsable de l'influx du cholestérol l'IDOL. Ces résultats obtenus pour ces deux gènes sous le contrôle de LXR, implémentés de ceux réalisés sur le modèle THP-1 montrant une augmentation de l'expression de ces gènes cibles (ABCA1, ABCG1, et ApoE) ont suggéré que l'ergostérol pouvait être un activateur de LXR. Cette hypothèse est bien évidemment à confirmer en réalisant des études supplémentaires avec un antagoniste de LXR par exemple ou en réprimant l'expression de celui-ci avec des SiRNA.

Par ailleurs, il a pu être montré que cette altération de la teneur en cholestérol intracellulaire observée dans notre modèle cellulaire après un traitement réalisé avec l'ergostérol avait une influence sur l'intégrité des radeaux lipidiques. En effet, une diminution de l'expression des protéines flottilin-2, caveolin-1 et EGFR impliquées dans de nombreuses voies de signalisation jouant un rôle important dans la progression tumorale a été mise en évidence. Il serait donc judicieux d'étudier l'effet de l'ergostérol sur les voies de signalisation telles que PI3K/AKT et MAPK impliquées dans la survie et la prolifération cellulaire.

Bien que les mécanismes impliqués n'aient pu être entièrement décrits, l'ensemble des effets exposés pourrait après leur confirmation par la réalisation de travaux supplémentaires expliquer la diminution de prolifération cellulaire observée dans nos lignées cancéreuses mammaires.

Enfin, pour évaluer le potentiel préventif et / ou curatif de l'ergostérol des expérimentations *in vivo* sur des modèles murins de xénogreffes de cellules cancéreuses mammaires sont nécessaires. Ceci pourrait se faire de deux manières, une administration comme dans l'étude (Li *et al.*, 2015) ou supplémentée à un régime alimentaire.

Par conséquent, ce projet de thèse pluridisciplinaire a permis au travers des différents travaux réalisés de souligner les différentes étapes et outils nécessaires à la recherche et à l'identification de métabolites fongiques bioactifs possédant une activité antiproliférative sur le cancer du sein. Ces travaux ont pu être valorisés par une publication concernant la méthodologie à suivre dans l'identification précoce de molécules bioactives, parue dans le journal *Analytica Chimica Acta*. Deux autres publications sont en cours de préparation, l'une concernant la détection de l'ergostérol en spectrométrie de masse après ionisation par électrospray et la seconde décrivant l'effet biologique de l'ergostérol sur la régulation du métabolisme lipidique des cellules cancéreuses mammaires.

Références bibliographiques

- Abramson, H.N. (2011). The Lipogenesis Pathway as a Cancer Target. *J. Med. Chem* *54*, 5615–5638.
- Ačimovič, J. and Rozman, D. (2013). Steroidal triterpenes of cholesterol synthesis. *Molecules* *18*, 4002–4017.
- Ahumada-rudolph, R. *et al.* (2016). Marine fungi isolated from Chilean fjord sediments can degrade oxytetracycline. *Environ. Monit. Assess.* *188*, 468–478.
- Alexandre, T.R. *et al.* (2017). Ergosterol isolated from the basidiomycete *Pleurotus salmoneostramineus* affects *Trypanosoma cruzi* plasma membrane and mitochondria. *J. Venom. Anim. Toxins Incl. Trop. Dis.* *23*, 30–40.
- Amend, A. *et al.* (2019). Fungi in the marine environment: Open questions and unsolved problems. *Ecol. Evol. Sci.* *10*, e01189–e01207.
- Angelucci, C. *et al.* (2018). Pivotal role of human stearoyl-CoA desaturases (SCD1 and 5) in breast cancer progression: Oleic acid-based effect of SCD1 on cell migration and a novel pro-cell survival role for SCD5. *Oncotarget* *9*, 24364–24380.
- Ano, Y. *et al.* (2017). Identification of 14-dehydroergosterol as a novel anti-inflammatory compound inducing tolerogenic dendritic cells. *Sci. Rep.* *7*, 2–8.
- Argmann, C.A. *et al.* (2005). Regulation of macrophage cholesterol efflux through HMG-CoA reductase inhibition: A role for RhoA in ABCA1-mediated cholesterol efflux. *J. Biol. Chem.* *23*, 22212–22221.
- Baek, A.E. *et al.* (2017). The cholesterol metabolite 27 hydroxycholesterol facilitates breast cancer metastasis through its actions on immune cells. *Nat. Commun.* *8*, 864–874.
- Baenke, F. *et al.* (2013). Hooked on fat: The role of lipid synthesis in cancer metabolism and tumour development. *Dis. Model. Mech.* *6*, 1353–1363.
- Bai, F. *et al.* (2019). Simvastatin induces breast cancer cell death through oxidative stress up-regulating miR-140-5p. *Aging (Albany, NY)*. *11*, 3198–3219.
- Barghoorn, E.S., and Linder, D.H. (1944). Marine fungi: their taxonomy and biology. *Farlowia* *1*, 395–467.
- Baumann, J. *et al.* (2013). Lipid biology of breast cancer. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids* *1831*, 1509–1517.
- Van Beek, T.A. *et al.* (2009). Recent developments in the rapid analysis of plants and tracking their bioactive constituents. *Phytochem. Rev.* *8*, 387–399.
- Belkaid, A. *et al.* (2015). 17 β -estradiol induces stearoyl-CoA desaturase-1 expression in estrogen receptor-positive breast cancer cells. *BMC Cancer* *15*, 440–454.
- Beutler, J.A. *et al.* (1990). Dereplication of phorbol bioactives: *Lyngbya majuscula* and *Croton cuneatus*. *J. Nat. Prod.* *53*, 867–874.
- Blackwell, M. (2011). The fungi : 1, 2, 3 ... 5.1 millions species ? *Am. J. Bot.* *98*, 426–438.
- Blanchet, E. *et al.* (2014). Synthesis and antiproliferative activity of ligerin and new fumagillin analogs against osteosarcoma. *Eur. J. Med. Chem.* *79*, 244–250.

- Blunt, J.W. *et al.* (2018). Natural product reports. *Nat. Prod. Rep.* *35*, 8–53.
- Bode, H.B., and Bethe, B. (2002). Big effects from small changes: Possible ways to explore nature's chemical diversity. *ChemBioChem* *7*, 619–627.
- Borgquist, S. *et al.* (2017). Cholesterol, cholesterol-lowering medication use, and breast cancer outcome in the BIG 1-98 study. *J. Clin. Oncol.* *35*, 1179–1188.
- Borgquist, S. *et al.* (2018). Statins: a role in breast cancer therapy? *J. Intern. Med.* *284*, 346–357.
- Botta, M. *et al.* (2018). PPAR agonists and metabolic syndrome: An established role? *Int. J. Mol. Sci.* *19*, 1197–1218.
- Bouchareychas, L., and Raffai, R. (2018). Apolipoprotein E and atherosclerosis: from lipoprotein metabolism to microRNA control of inflammation. *J. Cardiovasc. Dev. Dis.* *5*, 30.
- Britton, E.R. *et al.* (2018). Biochemometrics to identify synergists and additives from botanical medicines: A case study with *Hydrastis canadensis* (Goldenseal). *J. Nat. Prod.* *81*, 484–493.
- Brunner, I. *et al.* (2018). Ability of fungi isolated from plastic debris floating in the shoreline of a lake to degrade plastics. *PLoS One* *13*, e0202047–e0202061.
- Bugni, T.S., and Ireland, C.M. (2004). Marine-derived fungi: a chemically and biologically diverse group of microorganisms. *Nat. Prod. Rep.* *21*, 143–163.
- Cantrell, S.A. *et al.* (2006). Characterization of fungi from hypersaline environments of solar salterns using morphological and molecular techniques. *Mycol. Res.* *110*, 962–970.
- Carbonnelle, D. *et al.* (2017). LXR activation down-regulates lipid raft markers FLOT2 and DHHC5 in MC-7 breast cancer cells. *Anticancer Res.* *37*, 4067–4073.
- Cardellina, J.H. *et al.* (1993). A chemical screening strategy for the dereplication and prioritization of HIV-inhibitory aqueous natural products extracts. *J. Nat. Prod.* *56*, 1123–1129.
- Carmichael, A.R. (2006). Obesity and prognosis of breast cancer. *Obes. Rev.* *7*, 333–340.
- Carroll, A.R. *et al.* (2019). Marine natural products. *Nat. Prod. Rep.* *36*, 122–173.
- Caspi, R. *et al.* (2018). The MetaCyc database of metabolic pathways and enzymes. *Nucleic Acids Res.* *46*, D633–D639.
- Cavalier-Smith, T. (2004). Only six kingdoms of life. *Proc. R. Soc. B Biol. Sci.* *271*, 1251–1262.
- Chajès, V. *et al.* (2006). Acetyl-CoA carboxylase α is essential to breast cancer cell survival. *Cancer Res.* *66*, 5287–5294.
- Chang, S. *et al.* (2007). The association between lipid profiles and breast cancer among Taiwanese women. *Clin. Chem. Lab. Med.* *45*, 1219–1223.
- Cheang, W.S. *et al.* (2015). The peroxisome proliferator-activated receptors in cardiovascular diseases: Experimental benefits and clinical challenges. *Br. J. Pharmacol.* *172*, 5512–5522.

- Chem, N., and Author, B. (2015). Minimum Information about a biosynthetic gene cluster. *Nat Chem Biol* *11*, 625–631.
- Chuu, C.-P., and Lin, H.-P. (2010). Antiproliferative effect of LXR agonists T0901317 and 22 (R)-hydroxycholesterol on multiple human cancer cell lines. *Anticancer Res.* *30*, 3643–3648.
- Cortes, V.A. *et al.* (2014). Physiological and pathological implications of cholesterol. *Front. Biosci., Landmark Ed.* *19*, 416–428.
- Cos, P. *et al.* (2006). Anti-infective potential of natural products: How to develop a stronger in vitro “proof-of-concept.” *J. Ethnopharmacol.* *106*, 290–302.
- Currie, E. *et al.* (2013). Cellular fatty acid metabolism and cancer. *Cell Metab.* *18*, 153–161.
- Darimont, C. *et al.* (2006). Liver X receptor preferentially activates de novo lipogenesis in human preadipocytes. *Biochimie* *88*, 309–318.
- Deshmukh, S.K. *et al.* (2018). Marine fungi: A source of potential anticancer compounds. *Front. Microbiol.* *8*, 1–24.
- Eberlé, D. *et al.* (2004). SREBP transcription factors: Master regulators of lipid homeostasis. *Biochimie* *86*, 839–848.
- Elia, J. *et al.* (2019). 4-cholesten-3-one decreases breast cancer cell viability and alters membrane raft-localized EGFR expression by reducing lipogenesis and enhancing LXR-dependent cholesterol transporters. *Lipids Health Dis.* *18*, 1–15.
- Elissawy, A.M. *et al.* (2015). Bioactive terpenes from marine-derived fungi. *Mar. Drugs* *13*, 1966–1992.
- Espenshade, P.J., and Hughes, A.L. (2007). Regulation of sterol synthesis in eukaryotes. *Annu. Rev. Genet.* *41*, 401–427.
- Ferchichi, L. *et al.* (2012). Bioguided fractionation and isolation of natural inhibitors of advanced glycation end-products (AGEs) from *Calophyllum flavoramulum*. *Phytochemistry* *78*, 98–106.
- Flissi, A. *et al.* (2016). Norine, the knowledgebase dedicated to non-ribosomal peptides, is now open to crowdsourcing. *Nucleic Acids Res.* *44*, D1113–D1118.
- Fukuchi, J. *et al.* (2004). Antiproliferative effect of liver X receptor agonists on LNCaP human prostate cancer cells. *Cancer Res.* *64*, 7686–7689.
- Garcia-Estevez, L., and Moreno-Bueno, G. (2019). Updating the role of obesity and cholesterol in breast cancer. *Breast Cancer Res.* *21*, 35–42.
- Geiger, M. *et al.* (2013). Cytotoxicity and mycotoxin production of shellfish-derived *Penicillium* spp., a risk for shellfish consumers. *Lett. Appl Microbiol.* *57*, 385–392.
- Geladi, P., and Esbensen, K. (1990). The start and early history of chemometrics: Selected interviews. Part 1. *J. Chemom.* *4*, 337–354.
- Gelissen, I.C. *et al.* (2006). ABCA1 and ABCG1 synergize to mediate cholesterol export to ApoA-I. *Arterioscler., Thromb., Vasc Biol.* *26*, 534–540.

- Goldstein, J.L., and Brown, M.S. (2009). The LDL receptor. *Arterioscler., Thromb., Vasc. Biol.* **29**, 431–438.
- Gorny, A.M. *et al.* (2016). First report of *Didymella americana* on baby lima bean (*Phaseolus lunatus*). *Can. J. Plant Pathol.* **1**, 389–394.
- Gou, Q. *et al.* (2017). Peroxisome proliferator-activated receptors (PPARs) are potential drug targets for cancer therapy. *Oncotarget* **8**, 60704–60709.
- Griffett, K. *et al.* (2015). The LXR inverse agonist SR9238 suppresses fibrosis in a model of non-alcoholic steatohepatitis. *Mol. Metab.* **4**, 353–357.
- Grovel, O. *et al.* (2003). Accumulation of gliotoxin, a cytotoxic mycotoxin from *Aspergillus fumigatus*, in blue mussel (*Mytilus edulis*). *Toxicon* **42**, 297–300.
- Grygiel-Górniak, B. (2014). Peroxisome proliferator-activated receptors and their ligands: Nutritional and clinical implications - A review. *Nutr. J.* **13**, 1–10.
- Gutiérrez, M.H. *et al.* (2011). The role of fungi in processing marine organic matter in the upwelling ecosystem off Chile. *Mar. Biol.* **158**, 205–219.
- Han, J. *et al.* (2014). Upregulation of death receptor 5 and activation of caspase 8/3 play a critical role in ergosterol peroxide induced apoptosis in DU 145 prostate cancer cells. *Cancer Cell Int.* **14**, 117–125.
- Hanahan, D. and Weinberg, R.A. (2011). Hallmarks of cancer: The next generation. *Cell* **144**, 646–674.
- Hastings, J. *et al.* (2016). ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic Acids Res.* **44**, D1214–D1219.
- Hawksworth, D. and Gardens, R.B. (2017). Fungal diversity revisited: 2.2 to 3.8 million species. *Microbiol. Spectr* **5**, 79–95.
- He, L. *et al.* (2018). Anticancer action and mechanism of ergosterol peroxide from *paecilomyces cicadae* fermentation broth. *Int. J. Mol. Sci.* **19**, 3935–3947.
- Van den Heever, J.P. *et al.* (2016). The effect of dicyclohexylamine and fumagillin on *Nosema ceranae*-infected honey bee (*Apis mellifera*) mortality in cage trial assays. *Apidologie* **47**, 663–670.
- Hoai Trinh, P.T. *et al.* (2018). Secondary metabolites from *Penicillium chrysogenum* 045-357-2. *J. Sci. Technol.* **55**, 65–72.
- Hoang, T.P.T. (2018). De l'analyse métabolomique par LC-HRMS/MS de sources fongiques marines à la production, l'isolement et l'hémisynthèse de produits naturels bioactifs. Thèse de doctorat, Université de Nantes, soutenue le 16/10/2018 à Nantes.
- Hu, J. *et al.* (2012a). Dietary cholesterol intake and cancer. *Ann. Oncol.* **23**, 491–500.
- Hu, X. *et al.* (2012b). LXR β activation increases intestinal cholesterol absorption, leading to an atherogenic lipoprotein profile. *J. Intern. Med.* **272**, 452–464.
- Hua, Z. *et al.* (2005). The synthesis and preliminary biological evaluation of a novel steroid with neurotrophic activity: NGA0187. *J. Org. Chem.* **70**, 9849–9856.

- Huang, S. *et al.* (2016). Bioactive chaetoglobosins from the mangrove endophytic fungus *Penicillium chrysogenum*. *Mar. Drugs* *14*, 172–184.
- Hyde, K.D. *et al.* (1998). Role of fungi in marine ecosystems. *Biodivers. Conserv.* *7*, 1147–1161.
- Hyde, K.D. *et al.* (2019). The amazing potential of fungi: 50 ways we can exploit fungi industrially. *Fungal Divers.* *97*, 1–136.
- Imhoff, J.F. (2016). Natural products from marine fungi - Still an underrepresented resource. *Mar. Drugs* *14*, 1–19.
- Islam, M.M. *et al.* (2017). Exploring association between statin use and breast cancer risk: an updated meta-analysis. *Arch. Gynecol. Obstet.* *296*, 1043–1053.
- Jain, M.R. *et al.* (2004). 2,4,6-Trinitrotoluene transformation by a tropical marine yeast, *Yarrowia lipolytica* NCIM 3589. *Mar. Pollut. Bull.* *49*, 783–788.
- Janowski, B.A. *et al.* (1996). An oxysterol signalling pathway mediated by the nuclear receptor LXR alpha. *Lett. to Nat.* *96*, 728–731.
- Jin, Q. *et al.* (2010). Fatty acid synthase phosphorylation: A novel therapeutic target in HER2-overexpressing breast cancer cells. *Breast Cancer Res.* *12*, R96.
- Joseph, S.B. *et al.* (2002). Synthetic LXR ligand inhibits the development of atherosclerosis in mice. *Proc. Natl. Acad. Sci. U. S. A.* *99*, 7604–7609.
- Kanehisa, M. *et al.* (2017). KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* *45*, D353–D361.
- Kellogg, J.J. *et al.* (2016). Biochemometrics for natural products research: Comparison of data analysis approaches and application to identification of bioactive compounds. *J. Nat. Prod.* *79*, 376–386.
- Kim, S. *et al.* (2019). PubChem 2019 update: Improved access to chemical data. *Nucleic Acids Res.* *47*, D1102–D1109.
- Kim, Y. *et al.* (2009). Serum high-density lipoprotein cholesterol and breast cancer risk by menopausal status, body mass index, and hormonal receptor in Korea. *Cancer Epidemiol. Biomarkers Prev.* *18*, 508–515.
- Kimbung, S. *et al.* (2017). Impact of 27-hydroxylase (CYP27A1) and 27-hydroxycholesterol in breast cancer. *Endocr. Relat. Cancer* *24*, 339–349.
- Koehn, F.E., and Carter, G.T. (2005). The evolving role of natural products in drug discovery. *Nat. Rev. Drug Discov.* *4*, 206–220.
- Kohlmeyer, J., and Kohlmeyer, E. (1979). *Marine mycology: the higher fungi*, Academic press New York, 1–690.
- Komati, R. *et al.* (2017). Ligands of therapeutic utility for the liver X Receptors. *Molecules* *22*, 88–112.
- Kosemura, S., and Yamamura, S. (1997). Citreospirosteroid, a novel metabolite of a hybrid strain KO 0011 derived from *Penicillium citreo-viride* B. IFO 6200 and 4692. *Tetrahedron* *202*

Lett. 38, 7373–7374.

Kucharska-Newton, A.M. *et al.* (2008). HDL-cholesterol and incidence of breast cancer in the ARIC cohort. *Ann Epidemiol.* 18, 671–677.

Laisupasin, P. *et al.* (2013). Comparison of serum lipid profiles between normal controls and breast cancer patients. *J. Lab. Physicians* 5, 38–41.

Li, C. *et al.* (2016). Systematic review and meta-analysis suggest that dietary cholesterol intake increases risk of breast cancer. *Nutr. Res.* 36, 627–635.

Li, X. *et al.* (2015). Ergosterol purified from medicinal mushroom *Amauroderma rude* inhibits cancer growth *in vitro* and *in vivo* by up-regulating multiple tumor suppressors. *Oncotarget* 6, 17832–17846.

Li, Y.-X. *et al.* (2013). Fumigaclavine C from a Marine-Derived Fungus *Aspergillus Fumigatus* Induces Apoptosis in MCF-7 Breast Cancer Cells. *Mar. Drugs* 11, 5063–5086.

Liming, J. *et al.* (2016). Potential pharmacological resources: Natural bioactive compounds from marine-derived fungi. *Mar. Drugs* 14, 76–126.

Ling, T. *et al.* (2019). Cytostatic and cytotoxic natural products against cancer cell models. *Molecules* 24, 2012–2027.

Linnakoski, R. *et al.* (2018). Antiviral agents from fungi: Diversity, mechanisms and potential applications. *Front. Microbiol.* 9, 2325–2344.

Liscum, L. (2002). Cholesterol biosynthesis. In *biochemistry of lipids, lipoproteins and membranes*, Elsevier, ed. pp. 409–430.

Liu, Q. *et al.* (2014). Blockade of fas signaling in breast cancer cells suppresses tumor growth and metastasis via disruption of fas signaling-initiated cancer-related inflammation. *J. Biol. Chem.* 289, 11522–11535.

Liu, Y. *et al.* (2007). The involvement of lipid rafts in epidermal growth factor-induced chemotaxis of breast cancer cells. *Mol. Membr. Biol.* 24, 91–101.

Longo, N. *et al.* (2016). Carnitine transport and fatty acid oxidation. *Biochim Biophys Acta* 1863, 2422–2435.

Lotlikar, N.P. *et al.* (2018). Potential of marine-derived fungi to remove hexavalent chromium pollutant from culture broth. *Indian J. Microbiol.* 58, 182–192.

Luo, H. *et al.* (2013). Chaetoglobosin K inhibits tumor angiogenesis through downregulation of vascular epithelial growth factor-binding hypoxia-inducible factor 1 α . *Anticancer Drugs* 24, 715–724.

Lv, C. *et al.* (2015). Ophiobolin O isolated from *aspergillus ustus* induces G1 arrest of MCF-7 cells through interaction with AKT/GSK3 β /cyclin D1 signaling. *Mar. Drugs* 13, 431–443.

Mady, M.S. *et al.* (2016). The indole alkaloid meleagrins, from the olive tree endophytic fungus *Penicillium chrysogenum*, as a novel lead for the control of c-Met-dependent breast cancer proliferation, migration and invasion. *Bioorg. Med. Chem.* 24, 113–122.

Mansourian, M. *et al.* (2016). Statins use and risk of breast cancer recurrence and death: A

- systematic review and meta-analysis of observational studies. *J. Pharm. Pharm. Sci.* **19**, 72–81.
- Matsuda, Y. *et al.* (2013). Reconstituted biosynthesis of fungal meroterpenoid andrastin A. *Tetrahedron* **69**, 8199–8204.
- Mayer, A.M.S. *et al.* (2010). The odyssey of marine pharmaceuticals: a current pipeline perspective. *Trends Pharmacol. Sci.* **31**, 255–265.
- McDonnell, D.P. *et al.* (2014). Obesity, cholesterol metabolism and breast cancer pathogenesis. *Cancer Res.* **74**, 4976–4982.
- Mendez, D. *et al.* (2019). ChEMBL: Towards direct deposition of bioassay data. *Nucleic Acids Res.* **47**, D930–D940.
- Menendez, J.A., and Lupu, R. (2007). Fatty acid synthase and the lipogenic phenotype in cancer pathogenesis. *Nat. Rev. Cancer* **7**, 763–777.
- Meng, L.H. *et al.* (2015). Penicibrocazines A-E, five new sulfide diketopiperazines from the marine-derived endophytic fungus *Penicillium brocae*. *Mar. Drugs* **13**, 276–287.
- Michel Houten, S, and Wanders, R.J. A general introduction to the biochemistry of mitochondrial fatty acid β -oxidation. *J. Inherit. Metab. Dis.* **33**, 469-477.
- Moghadamtousi, S.Z. *et al.* (2015). Potential antiviral agents from marine fungi: An overview. *Mar. Drugs* **13**, 4520–4538.
- Moldes-Anaya, A. *et al.* (2017). Two isomeric C16 oxo-fatty acids from the diatom *Chaetoceros karianus* show dual agonist activity towards human peroxisome proliferator-activated receptors (PPARs) α/γ . *Mar. Drugs* **15**, 148-164.
- Mollinedo, F. and Gajate, C. (2015). Lipid rafts as major platforms for signaling regulation in cancer. *Adv. Biol. Regul.* **57**, 130–146.
- Monaco, M.E. (2017). Fatty acid metabolism in breast cancer subtypes. *Oncotarget* **8**, 29487–29500.
- Moschetta, A. (2011). Nuclear receptor LXR as a novel therapeutic antitumoral target in glioblastoma. *Cancer Discov.* **1**, 381–382.
- Mounier, C. *et al.* (2014). Lipogenesis in cancer progression (review). *Int. J. Oncol.* **45**, 485–492.
- Nicoletti, R., and Andolfi, A. (2018). The marine-derived filamentous fungi in biotechnology, "Grand challenge in marine biotechnology", Springer, Cham, 157–189.
- Omura, S. *et al.* (1996). Andrastins A-C, new protein farnesyltransferase inhibitors. Produced by *Penicillium* sp. FO-3929. *J. Antibiot. (Tokyo)*. **49**, 414–417.
- Orsi, W. *et al.* (2013). Deep sequencing of subseafloor eukaryotic rRNA reveals active fungi across marine subsurface provinces. *PLoS One* **8**, e56335–e56345.
- Ortega, S.N. *et al.* (2011). Isolation of Brazilian marine fungi capable of growing on DDD pesticide. *Biodegradation* **22**, 43–50.

- Overy, D.P. *et al.* (2014). An assessment of natural product discovery from marine (*sensu strictu*) and marine-derived fungi. *Mycology* 5, 145–167.
- Pandey, S. *et al.* (2018). Bio-guided fractionation of papaya leaf juice for delineating the components responsible for the selective anti-proliferative effects on prostate cancer cells. *Front. Pharmacol.* 9, 1319–1329.
- Pang, K.L. *et al.* (2016). ‘Marine fungi’ and ‘marine-derived fungi’ in natural product chemistry research: Toward a new consensual definition. *Fungal Biol. Rev.* 30, 163–175.
- Panthama, N. *et al.* (2011). Cytotoxic and antimalarial azaphilones from *Chaetomium longirostre*. *J. Nat. Prod.* 74, 2395–2399.
- Pence, H.E., and Williams, A. (2010). Chemspider: An online chemical information resource. *J. Chem. Educ.* 87, 1123–1124.
- Phan, L.M. *et al.* (2014). Cancer metabolic reprogramming: importance, main features, and potentials for precise targeted anti-cancer therapies. *Cancer Biol. Med.* 11, 1–19.
- Pieters, L., and Vlietinck, A.J. (2005). Bioguided isolation of pharmacologically active plant components, still a valuable strategy for the finding of new lead compounds? *J. Ethnopharmacol.* 100, 57–60.
- Pilon, A.C. *et al.* (2017). NuBBEDB: An updated database to uncover chemical and biological information from Brazilian biodiversity. *Sci. Rep.* 7, 7215–7227.
- Poirier, L. *et al.* (2007). Toxicity assessment of peptaibols and contaminated sediments on *Crassostrea gigas* embryos. *Aquat. Toxicol.* 83, 254–262.
- Qian, X.L. *et al.* (2019). Caveolin-1: A multifaceted driver of breast cancer progression and its application in clinical treatment. *Onco. Targets. Ther.* 12, 1539–1552.
- Rai, M. *et al.* (2018). Marine-derived Phoma—the gold mine of bioactive compounds. *Appl. Microbiol. Biotechnol.* 102, 9053–9066.
- Rämä, T. *et al.* (2014). Fungi ahoy! Diversity on marine wooden substrata in the high North. *Fungal Ecol.* 8, 46–58.
- Rateb, M.E., and Ebel, R. (2011). Secondary metabolites of fungi from marine habitats. *Nat. Prod. Rep.* 28, 290–344.
- Romano, S. *et al.* (2018). Extending the “one strain many compounds” (OSMAC) principle to marine microorganisms. *Mar. Drugs* 16, 244–273.
- Roullier, C. *et al.* (2016). Automated detection of natural halogenated compounds from LC-MS profiles—application to the isolation of bioactive chlorinated compounds from marine-derived fungi. *Anal. Chem.* 88, 9143–9150.
- El Roz, A. *et al.* (2012). LXR agonists and ABCG1-dependent cholesterol efflux in MCF-7 breast cancer cells: relation to proliferation and apoptosis. *Anticancer Res.* 32, 3007–3013.
- Ruiz, N. *et al.* (2010). Enhancement of domoic acid neurotoxicity on *Diptera larvae* bioassay by marine fungal metabolites. *Toxicon* 55, 805–810.

- Russo, A. *et al.* (2010). Pro-apoptotic activity of ergosterol peroxide and (22E)-ergosta-7,22-dien-5 α -hydroxy-3,6-dione in human prostate cancer cells. *Chem. Biol. Interact.* *184*, 352–358.
- Sakai, K. *et al.* (1994). Mer-NF8054A and X, novel antifungal steroids, isolated from *Aspergillus* sp. *J. Antibiot. (Tokyo)*. *47*, 591–594.
- Saleem, M. (2015). Chapter 9 – Bioactive natural products from marine-derived fungi: An update. "Studies in Natural Product Chemistry", Elsevier, *45*, 297–361.
- Sallenave-Namont, C. *et al.* (2000). Toxigenic saprophytic fungi in marine shellfish farming areas. *Mycopathologia* *149*, 21–25.
- Van de Sande, T. *et al.* (2002). Role of the phosphatidylinositol 3'-kinase/PTEN/Akt kinase pathway in the overexpression of fatty acid synthase in LNCaP prostate cancer cells. *Cancer Res.* *62*, 642–646.
- Santos, C.R. and Schulze, A. (2012). Lipid metabolism in cancer: Lipid metabolism in cancer. *FEBS J.* *279*, 2610–2623.
- Schoch, C.L. *et al.* (2009). The ascomycota tree of life: A phylum-wide phylogeny clarifies the origin and evolution of fundamental reproductive and ecological traits. *Syst. Biol.* *58*, 224–239.
- Schultz, J.R. *et al.* (2000). Role of LXRs in control of lipogenesis. *Genes Dev.* *14*, 2831–2838.
- Shan-Shan, Z. *et al.* (2015). Deacetyl-mycoepoxydiene, isolated from plant endophytic fungi *Phomopsis* sp. demonstrates anti-microtubule activity in MCF-7 cells. *Biomed. Pharmacother.* *69*, 82–89.
- Siddiquee, S. *et al.* (2015). Heavy metal contaminants removal from wastewater using the potential filamentous fungi biomass: A review. *J. Microb. Biochem. Technol.* *7*, 384–393.
- Stoekenbroek, R.M. *et al.* (2015). PCSK9 inhibition: The way forward in the treatment of dyslipidemia. *BMC Med.* *13*, 258–264.
- Suffness, M. *et al.* (1989). Discovery and development of antineoplastic agents from natural. In *Bioorganic Marine Chemistry*, "Bioorganic Marine Chemistry", Springer, Berlin, *3*, 131–168.
- Takahashi, N. *et al.* (1998). Induction of differentiation and apoptosis in human promyelocytic leukemia HL60 cell line by a new type of steroids. *Exp. Cell Res.* *245*, 313–320.
- Tisthammer, K.H. *et al.* (2016). Global biogeography of marine fungi is shaped by the environment. *Fungal Ecol.* *19*, 39–46.
- Torre, L. *et al.* (2015). Global Cancer Statistics, 2002. *CA. Cancer J. Clin.* *65*, 87–108.
- Touvier, M. *et al.* (2015). Cholesterol and breast cancer risk: a systematic review and meta-analysis of prospective studies. *Br. J. Nutr.* *114*, 347–357.
- Travis, R.C., and Key, T.J. (2003). Oestrogen exposure and breast cancer risk. *Breast Cancer Res.* *5*, 239–247.

- Umetani, M. *et al.* (2007). 27-Hydroxycholesterol is an endogenous SERM that inhibits the cardiovascular effects of estrogen. *Nat. Med.* *13*, 1185–1192.
- Undela, K. *et al.* (2012). Statin use and risk of breast cancer: A meta-analysis of observational studies. *Breast Cancer Res. Treat.* *135*, 261–269.
- Vaghefi, N. *et al.* (2016). Incidence and pathogenicity of *Didymella americana* on table beet in new york. *Plant Heal. Prog.* *17*, 270–272.
- Vansteelandt, M. *et al.* (2013). Ligerin, an antiproliferative chlorinated sesquiterpenoid from a marine-derived *penicillium* strain. *J. Nat. Prod.* *76*, 297–301.
- Varga, T. *et al.* (2011). PPARs are a unique set of fatty acid regulated transcription factors controlling both lipid metabolism and inflammation. *Biochim. Biophys. Acta - Mol. Basis Dis.* *1812*, 1007–1022.
- Wainwright, B.J. *et al.* (2017). Fungi associated with mesophotic macroalgae from the 'Au'au Channel, west Maui are differentiated by host and overlap terrestrial communities. *PeerJ.* *5*, e3532–e3559.
- Wang, B., and Tontonoz, P. (2018). Liver X receptors in lipid signalling and membrane homeostasis. *Nat. Rev. Endocrinol.* *14*, 452–463.
- Wang, X. *et al.* (2013). Flotillin-2 is associated with breast cancer progression and poor survival outcomes. *J. Transl. Med.* *11*, 190–211.
- Wang, X. *et al.* (2017). Structural diversity and biological activities of the cyclodipeptides from Fungi. *Molecules* *22*, 2026–2085.
- Wang *et al.* (2016). Sharing and community curation of mass spectrometry data with GNPS. *Nat Biotechnol* *34*, 352–359.
- Weihua, Z. *et al.* (2008). Survival of cancer cells is maintained by EGFR independent of its kinase activity. *Cancer Cell* *13*, 385–393.
- De Weille, J. *et al.* (2013). Oxysterols in cancer cell proliferation and death. *Biochem. Pharmacol.* *86*, 154–160.
- Weller, M.G. (2012). A unifying review of bioassay-guided fractionation, effect-directed analysis and related techniques. *Sensors* *12*, 9181–9209.
- Whittaker, R.H. (1969). New concepts of kingdoms of organisms. *Science.* *163*, 150–160.
- Wishart, D.S. *et al.* (2018). DrugBank 5.0: A major update to the drugbank database for 2018. *Nucleic Acids Res.* *46*, D1074–D1082.
- Woese, C.R., and Fox, G.E. (1977). Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc. Natl. Acad. Sci.* *74*, 5088–5090.
- Wymann, M.P., and Schneider, R. (2008). Lipid signalling in disease. *Nat. Rev. Mol. Cell Biol.* *9*, 162–176.
- Xie, G. *et al.* (2010). SZ-685C, a marine anthraquinone, is a potent inducer of apoptosis with anticancer activity by suppression of the Akt / FOXO pathway. *Br J Pharmacol.* *159*, 689–697.

- Xiong, T. *et al.* (2018). ATP- binding cassette transporter A1: A promising therapy target for prostate cancer (Review). *Mol. Clin. Oncol.* *8*, 9–14.
- Xu, L. *et al.* (2015). Antibacterial and antifungal compounds from marine fungi. *Mar. Drugs* *13*, 3479–3513.
- Yang, F. *et al.* (2013). Hirsutanol A, a novel sesquiterpene compound from fungus *Chondrostereum* sp., induces apoptosis and inhibits tumor growth through mitochondrial-independent ROS production: Hirsutanol A inhibits tumor growth through ROS production. *J. Transl. Med.* *11*, 32-42.
- Yarbrough, G.G. *et al.* (1992). Screening microbial metabolites for new drugs- Theoretical and practical issues. *J. Antibiot. (Tokyo)*. *46*, 535–545.
- Yarden, O. (2014). Fungal association with sessile marine invertebrates. *Front. Microbiol.* *5*, 228–254.
- Yasukawa, K. *et al.* (1994). Inhibitory effects of ergosterol isolated from the edible mushroom *Hypsizigus marmoreus* on TPA-induced inflammatory ear oedema and tumour promotion in mice. *Phyther. Res.* *8*, 10–13.
- Yoon, S. *et al.* (2007). Up-regulation of acetyl-CoA carboxylase α and fatty acid synthase by human epidermal growth factor receptor 2 at the translational level in breast cancer cells. *J. Biol. Chem.* *282*, 26122–26131.
- Yunianto, P. *et al.* (2014). Alkaloid (meleagrine and chrysogine) from endophytic fungi (*Penicillium* sp.) of *Annona squamosa* L . *Pakistan J. Biol. Sci.* *17*, 667–674.
- Zhao, J. *et al.* (2017). Exogenous lipids promote the growth of breast cancer cells via CD36. *Oncol. Rep.* *38*, 2105–2115.
- Zhou, X. *et al.* (2010). A review: Recent advances and future prospects of taxol-producing endophytic fungi. *Appl. Microbiol. Biotechnol.* *86*, 1707–1717.
- Zhu, X. *et al.* (2012). A marine anthraquinone SZ-685C overrides adriamycin-resistance in breast cancer cells through suppressing akt signaling. *Mar. Drugs* *10*, 694–711.

ANNEXES

Annexe 1. Molécules d'origine fongique marines cytotoxiques sur des lignées cellulaires cancéreuses mammaires.

Composés	Formule brute	Masse monoisotopique	Exemple de genre ou d'espèce fongique produisant cette molécule	Origine (si connue)	Activité antiproliférative (CI ₅₀ µM ou µg/mL)
isosclerone	C ₁₀ H ₁₀ O ₃	178.0630	<i>Aspergillus fumigatus</i>	algue verte (Corée)	MCF-7 (63 µM)
acaromycin a	C ₁₄ H ₁₂ O ₆	298.0463	<i>Acaromyces ingoldii</i> FS121	Mer (Sud de la Chine)	MCF-7 (6,7 µM)
(+)-cryptosporin	C ₁₄ H ₁₂ O ₆	276.0630	<i>Acaromyces ingoldii</i> FS121	Mer (Sud de la Chine)	MCF-7 (4,13 µM)
4-(3-hydroxyphenyl)-3-methoxyquinolin-2(1H)-one	C ₁₅ H ₁₀ NO ₃	252.0661	<i>Aspergillus versicolor</i> Y31-2,		MCF-7 (16,6 µM)
asperolide b	C ₁₆ H ₁₆ O ₅	304.0946	<i>Aspergillus wentii</i> EN-48	<i>Sargassum</i> sp., algue	MCF-7 (63 µM) MDA-MB-231 (35 µM)
wentilactone a	C ₁₆ H ₁₆ O ₆	304.0947	<i>Aspergillus versicolor</i> en-298	<i>ceramium japonicum</i> , algue	MCF-7 (69 µM)
rhytidchromone a	C ₁₆ H ₁₆ O ₇	320.0896	<i>Rhytidhysteron rufulum</i> BG2-Y	<i>A. ilicifolii</i> , Mangrove (Chine)	MCF-7 (19,3 µM)
asperolide e	C ₁₆ H ₁₈ O ₅	290.1154	<i>Aspergillus wentii</i> SD-310	Sédiment	MCF-7 (11 µM)
wentilactone b	C ₁₆ H ₁₈ O ₅	304.0947	<i>Aspergillus versicolor</i> en-298	<i>ceramium japonicum</i> , algue	MCF-7 (56 µM)
asperolide e	C ₁₆ H ₁₈ O ₆	290.1154	<i>Aspergillus wentii</i> SD-310	Sédiment	MCF-7 (11 µM)
botryosphaerin f	C ₁₆ H ₂₀ O ₅	292.1311	<i>Aspergillus terreus</i> (No. GX7-3B)	Mer (Sud de la Chine)	MCF-7 (4,4 µM)
xanthocillin x	C ₁₈ H ₁₂ N ₂ O ₂	288.0898	<i>Penicillium commune</i> SD-118		MCF-7 (12 µg/mL) MDA-MB- 231 (8 µg/mL)
rhytidchromone c	C ₁₈ H ₂₂ O ₈	366.1315	<i>Rhytidhysteron rufulum</i> BG2-Y	<i>Bruguiera gymnorrhiza</i> , Mangrove	MCF-7 (17,7 µM)
ethyl-2,4-dihydroxy-6-(8'-hydroxyonyl)- benzoate	C ₁₈ H ₂₈ O ₅	324.1937	<i>Lasiodiplodia</i> sp. 318#		MDA-MB-235 (10,1 µM)
chromosulfine	C ₁₉ H ₂₀ O ₉ S	424.0828	<i>Penicillium purpurogenum</i> G59		MCF-7 (5,2 µM)
preussomerin g	C ₂₀ H ₁₀ O ₇	362.0426	<i>Lasiodiplodia theobromae</i> ZJ-HQ1	<i>A. ilicifolii</i> , Mangrove (Chine)	MCF-7 (4,2 µM)
preussomerin h	C ₂₀ H ₁₂ O ₇	364.0583	<i>Lasiodiplodia theobromae</i> ZJ-HQ1	<i>A. ilicifolii</i> , Mangrove (Chine)	MCF-7 (2,6 µM)

Composés	Formule brute	Masse monoisotopique	Exemple de genre ou d'espèce fongique produisant cette molécule	Origine (si connue)	Activité antiproliférative (CI ₅₀ µM ou µg/mL)
preussomerin d	C ₂₀ H ₁₂ O ₇	364.0583	<i>Lasiodiplodia theobromae</i> ZJ-HQ1	<i>A. ilicifoliu</i> , Mangrove (Chine)	MCF-7 (9,8 µM)
preussomerin k	C ₂₀ H ₁₂ O ₈	380.0532	<i>Lasiodiplodia theobromae</i> ZJ-HQ1	<i>A. ilicifoliu</i> , Mangrove (Chine)	MCF-7 (2,5 µM)
preussomerin f	C ₂₀ H ₁₄ O ₇	366.0739	<i>Lasiodiplodia theobromae</i> ZJ-HQ1	<i>A. ilicifoliu</i> , Mangrove (Chine)	MCF-7 (3,1 µM)
scopararane d	C ₂₀ H ₂₈ O ₄	332.1988	<i>Eutypella scoparia</i> FS26	Sediment (Chine)	MCF-7 (25,6 µM)
libertellenone a	C ₂₀ H ₂₈ O ₄	332.1988	<i>Eutypella scoparia</i> FS26	Sediment (Chine)	MCF-7 (12,0 µM)
diaporthin b	C ₂₀ H ₂₈ O ₆	364.1886	<i>Eutypella scoparia</i> FS26	Sediment (Chine)	MCF-7 (4,4 µM)
chloropreussomerin a	C ₂₁ H ₁₅ ClO ₈	430.0455	<i>Lasiodiplodia theobromae</i> ZJ-HQ1	<i>A. ilicifoliu</i> , Mangrove (Chine)	MCF-7 (5,9 µM)
preussomerin m	C ₂₁ H ₁₆ O ₈	396.0845	<i>Lasiodiplodia theobromae</i> ZJ-HQ1	<i>A. ilicifoliu</i> , Mangrove (Chine)	MCF-7 (13 µM)
chloropreussomerin b	C ₂₂ H ₁₇ ClO ₈	444.0612	<i>Lasiodiplodia theobromae</i> ZJ-HQ1	<i>A. ilicifoliu</i> , Mangrove (Chine)	MCF-7 (6,2 µM)
luteoalbusin a	C ₂₃ H ₂₀ N ₄ O ₃ S ₂	464.0977	<i>Acrostalagmus luteoalbus</i> SCSIO F457	Mer (Sud de la Chine)	MCF-7 (0,233 µM)
gliocladine c	C ₂₃ H ₂₀ N ₄ O ₃ S ₂	464.0977	<i>Acrostalagmus luteoalbus</i> SCSIO F457	Mer (Sud de la Chine)	MCF-7 (0,91 - 17,78 µM)
luteoalbusin b	C ₂₃ H ₂₀ N ₄ O ₃ S ₃	496.0697	<i>Acrostalagmus luteoalbus</i> SCSIO F457	Mer (Sud de la Chine)	MCF-7 (0,25 µM)
gliocladine d	C ₂₃ H ₂₀ N ₄ O ₃ S ₃	496.0698	<i>Acrostalagmus luteoalbus</i> SCSIO F457	Mer (Sud de la Chine)	MCF-7 (0,91 - 17,78 µM)
meleagrín	C ₂₃ H ₂₃ N ₅ O ₄	433.1750	<i>Penicillium commune</i> SD-118		MDA-MB-231 (1,0 µg/mL)
breviones a	C ₂₇ H ₃₄ O ₄	422.2457	<i>Penicillium</i> sp	Mer profonde (Chine)	MCF-7 (28,4 µM)

Composés	Formule brute	Masse monoisotopique	Exemple de genre ou d'espèce fongique produisant cette molécule	Origine (si connue)	Activité antiproliférative (CI ₅₀ µM ou µg/mL)
brevione i	C ₂₇ H ₃₄ O ₅	438.2406	<i>Penicillium</i> sp	Mer profonde (Chine)	MCF-7 (7,44 µM)
2,3-didehydro-19α-hydroxy-14-epicochiloquinone b	C ₂₈ H ₃₈ O ₇	486.2617	<i>Nigrospora</i> sp. MA75	semi-mangrove plant <i>Pongamia pinnata</i>	MCF-7 (4 µg/mL)
aszonapyrone a	C ₂₈ H ₄₀ O ₅	456.28758	<i>Neosartorya laciniosa</i> (KUFC 7896)	Sol forestier côtier (Thaïlande)	MCF-7 (13,6 µM)
ergosterol peroxide	C ₂₈ H ₄₄ O ₃	428.3290	<i>Paecilomyces lilacinus</i> ZBY-1	Eaux profondes	MCF-7 (40 µg/mL)
(22e,24r)-5α,6α-epoxy-3β-hydroxyergosta-22-ene-7-one	C ₂₈ H ₄₄ O ₃	428.3290	<i>Paecilomyces lilacinus</i> ZBY-1	Eaux profondes	MCF-7 (9,5-59,6 mg/L)
5α,6α-epoxy-3β-hydroxyergost-22-ene-7-one	C ₂₈ H ₄₄ O ₃	428.3290	<i>Aspergillus terreus</i> (No. GX7-3B)	Mer (Sud de la Chine)	MCF-7 (4,98 µM)
3β,5α-dihydroxy-(22e,24r)-ergosta-7,22-dien-6-one	C ₃₀ H ₄₂ O ₇	514.2931	<i>Neosartorya laciniosa</i> (KUFC 7896)	Sol forestier côtier (Thaïlande)	MCF-7 (115 µM)
sartorypyrone a	C ₃₂ H ₃₇ NO ₆	531.2621	<i>Arthrinium arundinis</i> ZSDS1-F3	éponge <i>Phakellia fusca</i> (Chine)	MCF-7 (11,1 µM)
cytochalasin k	C ₄₅ H ₅₇ N ₃ O ₉	783.4095	<i>Aspergillus terreus</i> (No. GX7-3B)	Mer (Sud de la Chine)	MCF-7 (2,02 µM)
beauvericin	C ₁₆ H ₁₆ O ₇	320.0900	<i>Nigrospora</i> sp. No. 1403 as		MDA-MB-435 (3,1 µM)
deoxybostrycin	C ₁₅ H ₂₄ O ₅	284.1624	<i>Pseudolagarobasidium acaciicola</i>	<i>Bruguiera gymnorrhiz</i> , mangrove	MDA-MB-231 (11,41-49,08 µg/mL)
merulin b	C ₁₄ H ₂₂ O ₄	254.1518	<i>Pseudolagarobasidium acaciicola</i>	<i>Bruguiera gymnorrhiz</i> , mangrove	MDA-MB-231 (0,68-3,71 µg/mL)
3-epi-steperoxide a	C ₁₄ H ₂₂ O ₄	254.1518	<i>Pseudolagarobasidium acaciicola</i>	<i>Bruguiera gymnorrhiz</i> , mangrove	MDA-MB-231 (0,67-5,25 µg/mL)
steperoxide a	C ₃₀ H ₂₀ O ₁₀	540.10565	<i>Acremonium</i> sp.	<i>Sonneratia casolaris</i> , mangrove	MDA-MB-231 (0,3-1,5 µM)
torribiellin b	C ₂₀ H ₃₂ O	288.2453	<i>Trichoderma</i> sp. Xy24	<i>Xylocarpus granatum</i> , mangrove	MCF-7 (30,1 µM)

Composés	Formule brute	Masse monoisotopique	Exemple de genre ou d'espèce fongique produisant cette molécule	Origine (si connue)	Activité antiproliférative (CI ₅₀ µM ou µg/mL)
harzianolactone	C ₂₀ H ₃₀ O ₂	302.2246	<i>Trichoderma</i> sp. Xy24	<i>Xylocarpus granatum</i> , mangrove	MCF-7 (30,7 µM)
sumalarin a	C ₂₀ H ₂₆ O ₈ S	426.1348	<i>Penicillium sumatrense</i> MA-92	<i>Lumnitzera racemos</i> , rhizosphère	MCF-7 (3,8-10 µM)
sumalarin b	C ₂₅ H ₃₂ O ₁₁ S	540.1665	<i>Penicillium sumatrense</i> MA-92	<i>Lumnitzera racemos</i> , rhizosphère	MCF-7 (3,8-10 µM)
sumalarin c	C ₁₉ H ₂₄ O ₈ S	400.1192	<i>Penicillium sumatrense</i> MA-92	<i>Lumnitzera racemos</i> , rhizosphère	MCF-7 (3,8 -10 µM)
dehydrocurvularin	C ₁₆ H ₁₈ O ₅	290.1154	<i>Penicillium sumatrense</i> MA-92	<i>Lumnitzera racemos</i> , rhizosphère	MCF-7 (3,8 - 10 µM)
tmc-264	C ₁₆ H ₁₃ ClO ₇	352.0350	<i>Penicillium brocae</i> MA-231	<i>Avicennia marin</i> , mangrove	MDA-MB-231 (2,81 µM)
aspergisocoumrin a	C ₁₄ H ₁₂ O ₆	276.0634	<i>Aspergillus</i> sp. HN15-5D	Feuilles d' <i>Acanthus ilicifolius</i> , mangrove	MDA-MB-435 (5,08 µM)
aspergisocoumrin b	C ₁₄ H ₁₂ O ₆	276.0634	<i>Aspergillus</i> sp. HN15-5D	Feuilles d' <i>Acanthus ilicifolius</i> , mangrove	MDA-MB-435 (4,98 µM)
mycoepoxydiene	C ₁₆ H ₁₈ O ₅	290.1154	<i>Phomopsis</i> sp. A818	<i>Kandelia candel</i>	MDA-MB-435 (7,85 µM)
deacetylmycoepoxydiene	C ₁₄ H ₁₆ O ₄	248.10486	<i>Phomopsis</i> sp. A818	Foliage of <i>Kandelia candel</i>	MDA-MB-435 (14,61 µM)
fusaristatin a	C ₃₆ H ₅₈ N ₄ O ₇	658.4305	<i>Diaporthe phaseolorum</i> SKS019	Feuilles d' <i>Acanthus ilicifolius</i> , mangrove	MDA-MB-435 (8,15 µM)
5-deoxybostrycoidin	C ₁₅ H ₁₁ NO ₄	178.0477	<i>Diaporthe phaseolorum</i> SKS019	Feuilles d' <i>Acanthus ilicifolius</i> , mangrove	MDA-MB-435 (5,32 µM)
deacetylphomoxanthone b	C ₃₄ H ₃₄ O ₁₄	666.1949	<i>Phomopsis</i> sp. HNY29-2B	Feuilles d' <i>Acanthus ilicifolius</i> , mangrove	MDA-MB-435 (14,40 µM)
dicerandrol c	C ₃₈ H ₃₈ O ₁₆	750.2160	<i>Phomopsis</i> sp. HNY29-2B	Feuilles d' <i>Acanthus ilicifolius</i> , mangrove	MDA-MB-435 (44,10 µM)
dicerandrol b	C ₃₆ H ₃₆ O ₁₅	708.2054	<i>Phomopsis</i> sp. HNY29-2B	Feuilles d' <i>Acanthus ilicifolius</i> , mangrove	MDA-MB-435<10 µM
dicerandrol a	C ₃₄ H ₃₄ O ₁₄	666.1949	<i>Phomopsis</i> sp. HNY29-2B	Feuilles d' <i>Acanthus ilicifolius</i> , mangrove	MDA-MB-435<10 µM

Composés	Formule brute	Masse monoisotopique	Exemple de genre ou d'espèce fongique produisant cette molécule	Origine (si connue)	Activité antiproliférative (CI ₅₀ µM ou µg/mL)
phomopsichalasin g	C ₂₈ H ₃₅ NO ₅	465.2515	<i>Phomopsis</i> spp. Xy21	<i>Xylocarpus granatum</i> , mangrove	MDA-MB-231 (3,4 µM)
pestalamine a	C ₂₁ H ₂₅ NO ₆	387.1682	<i>Pestalotiopsis vaccinia</i>	<i>Kandelia candel</i>	MCF-7(40,3 µM)
sz-685c	C ₁₄ H ₁₆ O ₈	312.0845	<i>Aspergillus ustus</i>	<i>Halorosellinia</i> sp.	MCF-7 (7,38 µM) MCF-7/ADR (4,17 µM) MCF-7/Akt (3,36 µM)
ophiobolin z	C ₂₇ H ₄₂ O ₄	430.3083	<i>Aspergillus ustus</i>	<i>Halorosellinia</i> sp., mangrove	MDA-MB-231 (18,5 µM) MCF-7 (13,2 µM) MCF/ADR (5,4 µM)
21- <i>epi</i> -ophiobolin z	C ₂₇ H ₄₂ O ₄	430.3083	<i>Aspergillus ustus</i>	<i>Halorosellinia</i> sp., mangrove	MDA-MB-231 (21,4µM) MCF-7 (9,4 µM) MCF/ADR (7,9µM)
ophiobolin k	C ₂₅ H ₃₆ O ₃	384.2664	<i>Aspergillus ustus</i>	<i>Halorosellinia</i> sp., mangrove	MDA-MB-231 (5,9 µM) MCF-7 (5,1 µM) MCF/ADR (4,1 µM)
6- <i>epi</i> -ophiobolin k	C ₂₅ H ₃₆ O ₃	384.2664	<i>Aspergillus ustus</i>	<i>Halorosellinia</i> sp., mangrove	MDA-MB-231 (4,1µM) MCF-7 (12,8 µM) MCF/ADR (26,1 µM)
ophiobolin g	C ₂₅ H ₃₄ O ₂	366.2559	<i>Aspergillus ustus</i>	<i>Halorosellinia</i> sp., mangrove	MDA-MB-231 (11,6 µM) MCF-7 (15,4 µM) MCF/ADR (35,4 µM)
6- <i>epi</i> -ophiobolin g	C ₂₅ H ₃₄ O ₂	366.2559	<i>Aspergillus ustus</i>	<i>Halorosellinia</i> sp., mangrove	MDA-MB-231 (18,5 µM) MCF-7 (13,2 µM) MCF/ADR (5,4 µM)
fumiglacavine c	C ₂₃ H ₃₀ N ₂ O ₂	366.2307	<i>Aspergillus Fumigatus</i>		MCF-7 (93% d'inhibition à 20 µM)
hirsutanol a	C ₁₅ H ₁₈ O ₃	246.1256	<i>Chondrostereum</i> sp.	<i>Sarcophyton tortuosum</i>	MDA-MB-231 (0,48 µM)

Composés	Formule brute	Masse monoisotopique	Exemple de genre ou d'espèce fongique produisant cette molécule	Origine (si connue)	Activité antiproliférative (CI ₅₀ µM ou µg/mL)
emodine	C ₁₅ H ₁₀ O ₅	270.0528	<i>Aspergillus</i> sp.		ZR-75-30 (< 40 µM) Bcap-37 (< 40 µM) SKBR-3 (25 µM)
penicillivinacine	C ₂₂ H ₂₆ N ₂ O ₄	382.1893	<i>Penicillium vinaceum</i>		MDA-MB-231 (18,4 µM)
terretrione a	C ₁₈ H ₂₄ N ₂ O ₃	316.1787	<i>Penicillium vinaceum</i>		MDA-MB-231 (17,7 µM)
penitrem a	C ₃₇ H ₄₄ ClNO ₆	633.2857	<i>Penicillium commune</i> GS20.		MCF-7 (11,9 µM) MDA-MB-231 (9,8 µM)
penitrem b	C ₃₇ H ₄₅ NO ₅	583.3298	<i>Penicillium commune</i> GS20.		MCF-7 (5,5 µM) MDA-MB-231 (13,7 µM)
penitrem d	C ₃₇ H ₄₅ NO ₄	567.3349	<i>Penicillium commune</i> GS20.		MCF-7 (8,3 µM) MDA-MB-231 (29,7µM)
penitrem e	C ₃₇ H ₄₅ NO ₆	599.3247	<i>Penicillium commune</i> GS20.		MCF-7 (17,5µM) MDA-MB-231 (25,4µM)
penitrem f	C ₃₇ H ₄₄ ClNO ₅	617.2908	<i>Penicillium commune</i> GS20.		MCF-7 (15,0 µM) MDA-MB-231 (13,8 µM)
paspaline	C ₂₈ H ₃₉ NO ₂	421.2981	<i>Penicillium commune</i> GS20.		MCF-7 (12,8 µM) MDA-MB-231 (12,4 µM)
emindole sb	C ₂₈ H ₃₉ NO	405.3032	<i>Penicillium commune</i> GS20.		MCF-7 (10,1 µM) MDA-MB-231 (21,3 µM)
6-bromopenitrem b	C ₃₇ H ₄₄ BrNO ₅	661.2403	<i>Penicillium commune</i> GS20.		MCF-7 (19,3 µM) MDA-MB-231 (18,8 µM)
6-bromopenitrem e	C ₃₇ H ₄₄ BrNO ₆	677.2352	<i>Penicillium commune</i> GS20.		MCF-7 (16,7 µM) MDA-MB-231 (8,5 µM)
aszonapyrone a	C ₂₈ H ₄₀ O ₅	456.2876	<i>Neosartorya fischeri</i> (KUFC 6344)		MCF-7 (13,6 µM)

Composés	Formule brute	Masse monoisotopique	Exemple de genre ou d'espèce fongique produisant cette molécule	Origine (si connue)	Activité antiproliférative (CI ₅₀ µM ou µg/mL)
sartorypyrone b	C ₃₀ H ₄₂ O ₇	514.2931	<i>Neosartorya tsunodae</i> (KUFC 9213)		MCF-7 (17,8 µM)

Réalisé à partir de :

Deshmukh, K *et al*, Marine fungi: A source of potential anticancer compounds, *Frontiers in Microbiology*. 8 (2018). doi:10.3389/fmicb.2017.02536.

Deshmukh, K *et al*. Mangrove-Associated Fungi: A Novel Source of Potential Anticancer Compounds, *Journal of fungi* 4,3, 101-128 (2018). doi:10.3390/jof4030101

Annexe 2. Présentation des extraits issus de l'extractothèque du laboratoire MMS

Extractothèque	Souche						Culture		Extraction		Essais biologiques		
	CODE EXTRAIT	Poids (mg)	Genre	espèce	Collecte	Site de prélèvement	Code Genbank Séquence ITS	milieu utilisé	type eau EDM = Mer EDMurelle EDMS = Mer Synthét. ED = Eau Distillée autre : préciser	Contenant	Matrice T = totalité G = Géluse M = Mycélium S = Spores	Techniques	% d'inhibition des cellules MCF-7 (10 µg/mL, 24H)
029-431	522,2	<i>Penicillium</i>	<i>citreonigrum</i>	Sédiments de surface	Le Croisic		YES	EDM	Petri 16 cm	M	Smedsgaard puis CH ₂ Cl ₂	39,8 ± 1,1	1,5 (PPARα)
029-443	60,8	<i>Penicillium</i>	<i>citreonigrum</i>	Sédiments de surface	Le Croisic		PDA	EDM	Petri 16 cm	M	Smedsgaard puis CH ₂ Cl ₂	-37,89 ± 7,6	
029-447	40,3	<i>Penicillium</i>	<i>citreonigrum</i>	Sédiments de surface	Le Croisic		MEA	EDM	Petri 16 cm	M	Smedsgaard puis CH ₂ Cl ₂	69,23 ± 1,89	
1053-374	22,5	<i>Fusarium</i>		Sédiments de surface	Le Croisic		DCA	EDM	Erlen	T	WSLE	-9,42 ± 7,2	
14-41	2,4	<i>Penicillium</i>	<i>antarcticum</i>	Coques	Le Croisic		YES	EDM	Erlen	T	EtOAc/CH ₂ Cl ₂	54,69 ± 1,06	
14-43	2,9	<i>Penicillium</i>	<i>antarcticum</i>	Coques	Le Croisic		CYA	EDM	Erlen	T	EtOAc/CH ₂ Cl ₂	56,67 ± 1	1,8 (PPARα)
14-44	2,8	<i>Penicillium</i>	<i>antarcticum</i>	Coques	Le Croisic		DCA	EDM	Erlen	T	EtOAc/CH ₂ Cl ₂	58,04 ± 1,29	1,4 (PPARα)
15-52	4,8	<i>Penicillium</i>	<i>antarcticum</i>	Coques	Le Croisic		YES	EDM	Erlen	T	EtOAc/CH ₂ Cl ₂	60,79 ± 1,85	
15-53	4,5	<i>Penicillium</i>	<i>antarcticum</i>	Coques	Le Croisic		PDA	EDM	Erlen	T	EtOAc/CH ₂ Cl ₂	78,72 ± 2,47	
15-54	7,2	<i>Penicillium</i>	<i>antarcticum</i>	Coques	Le Croisic		CYA	EDM	Erlen	T	EtOAc/CH ₂ Cl ₂	ND	1,8 (PPARα) 1,7 (PPARγ)
15-55	3,4	<i>Penicillium</i>	<i>antarcticum</i>	Coques	Le Croisic		DCA	EDM	Erlen	T	EtOAc/CH ₂ Cl ₂	54,85 ± 3,56	
1513-463	11,6	<i>Trichoderma</i>	<i>atroviride</i>	Moules	Bonne Anse	JQ653085.1	DCA	EDMS	Pétri	M	MeOH puis CH ₂ Cl ₂	56,15 ± 0,55	

Extractothèque		Souche							Culture			Extraction		Essais biologiques	
		Penicillium	canescens	Eau de mer	La Baule	MN133997.1	MEA	EDM	Erlen	T	EtOAc	67,46 ± 3,23			
194-89	82,7	<i>Penicillium</i>	<i>canescens</i>	Eau de mer	La Baule	MN133997.1	MEA	EDM	Erlen	T	EtOAc	67,46 ± 3,23			
194-91	5,9	<i>Penicillium</i>	<i>canescens</i>	Eau de mer	La Baule	MN133997.1	MEA	EDM	Erlen	T	EtOAc/CH ₂ Cl ₂	68,81 ± 0,92			
194-92	9,5	<i>Penicillium</i>	<i>canescens</i>	Eau de mer	La Baule	MN133997.1	YES	EDM	Erlen	T	EtOAc/CH ₂ Cl ₂	60,09 ± 1,71			
231-100	8,0	<i>Penicillium</i>	<i>polonicum</i>	Moules	La Baule	KU720402.1	CYA	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	78,46 ± 1,65			
231-101	6,3	<i>Penicillium</i>	<i>polonicum</i>	Moules	La Baule	KU720402.1	MES	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	42,45 ± 4,68			
231-102	4,8	<i>Penicillium</i>	<i>polonicum</i>	Moules	La Baule	KU720402.1	MEA	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	70,2 ± 1,29			
231-103	3,9	<i>Penicillium</i>	<i>polonicum</i>	Moules	La Baule	KU720402.1	DCA	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	45,38 ± 4,08			
231-104	7,6	<i>Penicillium</i>	<i>polonicum</i>	Moules	La Baule	KU720402.1	PDA	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	75,51 ± 0,70			
231-99	4,7	<i>Penicillium</i>	<i>polonicum</i>	Moules	La Baule	KU720402.1	YES	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	64,59 ± 2,72			
266-106	8,6	<i>Penicillium</i>	<i>marinum</i>	Moules	La Baule		MEA	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	75,27 ± 1,59			
266-109	2,3	<i>Penicillium</i>	<i>marinum</i>	Moules	La Baule		CYA	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	71,39 ± 0,94			
266-110	2,3	<i>Penicillium</i>	<i>marinum</i>	Moules	La Baule		PDA	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	68,5 ± 1,05			
266-111	6,8	<i>Penicillium</i>	<i>marinum</i>	Moules	La Baule		MES	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	69,44 ± 1,28			
270-112	7,1	<i>Penicillium</i>	<i>bialowiezense</i>	Coques	La Baule		YES	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	5,58 ± 4,51			
270-113	7,9	<i>Penicillium</i>	<i>bialowiezense</i>	Coques	La Baule		CYA	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	-1,78 ± 4,18			
270-114	5,7	<i>Penicillium</i>	<i>bialowiezense</i>	Coques	La Baule		MES	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	0,53 ± 9,87			
270-115	6,1	<i>Penicillium</i>	<i>bialowiezense</i>	Coques	La Baule		MEA	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	-2,66 ± 10,12			
270-116	6,7	<i>Penicillium</i>	<i>bialowiezense</i>	Coques	La Baule		DCA	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	0,42 ± 5,26			
270-117	7,6	<i>Penicillium</i>	<i>bialowiezense</i>	Coques	La Baule		PDA	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	25,35 ± 1,15			
330-119	8,3	<i>Penicillium</i>	<i>ubiquetum</i>	Moules	Port Giraud	MH172367.1	CYA	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	ND	1,7 (PPARV) 4,7 (PPARα)		
330-121	9,2	<i>Penicillium</i>	<i>ubiquetum</i>	Moules	Port Giraud	MH172367.1	MEA	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	43,95 ± 1,20			
330-122	9,6	<i>Penicillium</i>	<i>ubiquetum</i>	Moules	Port Giraud	MH172367.1	DCA	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	57,92 ± 1,60			

Extractothèque		Souche						Culture			Extraction		Essais biologiques	
		Penicillium	ubiquetum	Moules	Port Giraud	MH172367.1	PDA	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	ND	2,4 (PPARV) 5,3 (PPARα)	
330-123	7,5	Penicillium	ubiquetum	Moules	Port Giraud	MH172367.1	PDA	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	ND	2,4 (PPARV) 5,3 (PPARα)	
351-129	9,5	Penicillium	ligerum	Eau de mer	La Prée	JN676192	YES	EDM	Erlen	T	EtOAc	54,02 ± 2,15		
351-135	3,9	Penicillium	ligerum	Eau de mer	La Prée	JN676192	DCA	ED	Petri	T	Micro Smedsgaard	1,53 ± 7,15		
351-136	2,8	Penicillium	ligerum	Eau de mer	La Prée	JN676192	MEA	ED	Petri	T	Micro Smedsgaard	3,94 ± 6,21		
351-138	3,4	Penicillium	ligerum	Eau de mer	La Prée	JN676192	PDA	ED	Petri	T	Micro Smedsgaard	-5,81 ± 4,06		
351-139	8,0	Penicillium	ligerum	Eau de mer	La Prée	JN676192	YES	ED	Petri	T	Micro Smedsgaard	-3,57 ± 3,0		
351-146	10,7	Penicillium	ligerum	Eau de mer	La Prée	JN676192	MES	EDMS	Erlen	T	EtOAc	59,43 ± 1,41		
393-160	8,1	Penicillium		Moules	La Prée		YES	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	5,69 ± 2,27		
393-161	18,3	Penicillium		Moules	La Prée		CYA	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	66,32 ± 2,1		
393-162	7,9	Penicillium		Moules	La Prée		MES	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	26,97 ± 7,28		
393-163	9,1	Penicillium		Moules	La Prée		MEA	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	13,8 ± 12,72		
393-164	8,4	Penicillium		Moules	La Prée		DCA	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	15,05 ± 3,7		
393-165	3,0	Penicillium		Moules	La Prée		PDA	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	60,34 ± 0,72		
404-173	7,7	Penicillium	brevicompactum	Coques	Le Croisic		CYA	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	-2,53 ± 10,15		
404-175	6,0	Penicillium	brevicompactum	Coques	Le Croisic		MEA	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	-1,81 ± 7,36		
404-176	7,0	Penicillium	brevicompactum	Coques	Le Croisic		DCA	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	-1,66 ± 5,99		
404-177	5,2	Penicillium	brevicompactum	Coques	Le Croisic		PDA	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	-6,41 ± 8,65		
417-179	4,2	Penicillium	restrictum	Coques	Le Croisic	KU720404.1	CYA	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	66,47 ± 2,38		
417-181	3,5	Penicillium	restrictum	Coques	Le Croisic	KU720404.1	MEA	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	47,51 ± 1,36		
417-182	5,8	Penicillium	restrictum	Coques	Le Croisic	KU720404.1	DCA	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	58,96 ± 1,88	3,7 (PPARα)	
417-183	4,8	Penicillium	restrictum	Coques	Le Croisic	KU720404.1	PDA	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	70,99 ± 0,03	2,9 (PPARα)	
42-60	2,9	Penicillium	expansum	Sédiments	Le Croisic	JN794527	MEA	EDM	Erlen	T	EtOAc	69,94 ± 1,89		

Extractothèque		Souche						Culture			Extraction		Essais biologiques	
		Penicillium	expansum	Sédiments	Le Croisic	JN794527	MEA	EDM	Erlen	T	EtOAc/CH ₂ Cl ₂	70,08 ± 1,78		
42-61	6,2	Penicillium	expansum	Sédiments	Le Croisic	JN794527	MEA	EDM	Erlen	T	EtOAc/CH ₂ Cl ₂	70,08 ± 1,78		
42-63	3,9	Penicillium	expansum	Sédiments	Le Croisic	JN794527	PDA	EDM	Erlen	T	EtOAc/CH ₂ Cl ₂	75,69 ± 2,40		
460-184	7,5	Penicillium	canescens	Sédiments	Le Croisic	KU720405.1	MEA	EDM	Erlen	T	EtOAc/CH ₂ Cl ₂	58,67 ± 1,28		
5-30	5,1	Penicillium	chrysogenum	Coques	Le Croisic	MK015724.1	YES	EDM	Erlen	T	EtOAc/CH ₂ Cl ₂	57,23 ± 1,29		
5-CYA	7,3	Penicillium	chrysogenum	Coques	Le Croisic	MK015724.1	CYA	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	66,14 ± 4,70	ND	
5-DCA	17,1	Penicillium	chrysogenum	Coques	Le Croisic	MK015724.1	DCA	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	75,05 ± 6,72	ND	
5-PDA	8,3	Penicillium	chrysogenum	Coques	Le Croisic	MK015724.1	PDA	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	74,84 ± 4,65	ND	
5-32	3,3	Penicillium	chrysogenum	Coques	Le Croisic	MK015724.1	CYA	EDM	Erlen	T	EtOAc/CH ₂ Cl ₃	ND	2,5 (PPARY) 4,7 (PPARα)	
5-33	8,2	Penicillium	chrysogenum	Coques	Le Croisic	MK015724.1	DCA	EDM	Erlen	T	EtOAc/CH ₂ Cl ₄	ND	2,4 (PPARY) 4,9 (PPARα)	
50-75	14,4	Penicillium	venetum	Moules	La Tara		CYA	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	58,5 ± 1,39		
50-77	7,9	Penicillium	venetum	Moules	La Tara		MEA	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	11,46 ± 9,7		
50-78	7,9	Penicillium	venetum	Moules	La Tara		DCA	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	4,2 ± 6,35		
50-79	7,7	Penicillium	venetum	Moules	La Tara		PDA	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	63,01 ± 0,73		
556-197	56,1	Penicillium	atramentosum	Sédiments	La Louippe		MEA	EDM	Erlen	T	EtOAc	24,31 ± 2,11		
556-198	193,8	Penicillium	atramentosum	Sédiments	La Louippe		YES	EDM	Erlen	T	EtOAc	28,44 ± 4,29		
556-199	7,1	Penicillium	atramentosum	Sédiments	La Louippe		MEA	EDM	Erlen	T	EtOAc/CH ₂ Cl ₂	25,81 ± 0,56		
556-200	4,9	Penicillium	atramentosum	Sédiments	La Louippe		YES	EDM	Erlen	T	EtOAc/CH ₂ Cl ₂	14,46 ± 11,03		
719-215	139,5	Didymella	americana	Sédiments de surface	La Couplasse	MN133961.1	YES	EDM	Petri 16 cm	M	Smedsgaard puis CH ₂ Cl ₂	45,8 ± 0,24	2 (PPARα)	
719-227	7,3	Didymella	americana	Sédiments de surface	La Couplasse	MN133961.1	PDA	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	ND	4,3 (PPARα) 3,2 (PPARY)	
719-231	76,0	Didymella	americana	Sédiments de surface	La Couplasse	MN133961.1	MEA	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₂	ND	4,5 (PPARα) 1,6 (PPARY)	
719-527	1,0	Didymella	americana	Sédiments de surface	La Couplasse	MN133961.1	DCA	EDMS	Erlen	T	EtOAc/CH ₂ Cl ₃	62,08 ± 3,54	ND	

Extractothèque		Souche						Culture			Extraction		Essais biologiques	
719-223	118,7	<i>Didymella</i>	<i>americana</i>	Sédiments de surface	La Couplasse	MN133961.1	DCA	EDMS	Petri 16 cm	M	Smedsgaard puis CH ₂ Cl ₂	ND	5,1 (PPARα) 1,9 (PPARY)	
747-250	8,3	<i>Penicillium</i>	<i>ligerum</i>	Sédiments	La Couplasse	JN794528.1	MEA	EDMs	Erlen	T	EtOAc	20,71 ± 4,22		
747-252	9,1	<i>Penicillium</i>	<i>ligerum</i>	Sédiments	La Couplasse	JN794528.1	CYA	EDMs	Erlen	T	EtOAc	27,05 ± 1,16		
747-254	3,4	<i>Penicillium</i>	<i>ligerum</i>	Sédiments	La Couplasse	JN794528.1	KMS	EDMs	Erlen	T	EtOAc	18,6 ± 1,24		
747-255	18,8	<i>Penicillium</i>	<i>ligerum</i>	Sédiments	La Couplasse	JN794528.1	DCA	EDMs	Erlen	T	EtOAc	22,02 ± 2,23		
747-256	7,4	<i>Penicillium</i>	<i>ligerum</i>	Sédiments	La Couplasse	JN794528.1	MES	EDMs	Erlen	T	EtOAc	20,92 ± 3,48		
770-262	119,5	<i>Aspergillus</i>		Sédiments	Port du Collet		DCA	EDM	Erlen	T	WSLE	-7,99 ± 7,09		
797-271	710,4	<i>Phoma</i>	<i>exigua</i> Desm. Var. <i>exigua</i>	Sédiments de surface	Port Giraud		YES	EDM	Petri 16 cm	M	Smedsgaard puis CH ₂ Cl ₂	38,2 ± 1,18		
850-296	151,2	<i>Scopulariopsis</i>		Sédiments de surface	Port Giraud		YES	EDM	Petri 16 cm	M	Smedsgaard puis CH ₂ Cl ₂	-9,57 ± 0,97		
850-304	173,3	<i>Scopulariopsis</i>		Sédiments de surface	Port Giraud		DCA	EDM	Petri 16 cm	M	Smedsgaard puis CH ₂ Cl ₂	-0,72 ± 2,04	2,1 (PPARα)	
850-308	26,8	<i>Scopulariopsis</i>		Sédiments de surface	Port Giraud		PDA	EDM	Petri 16 cm	M	Smedsgaard puis CH ₂ Cl ₂	-10,92 ± 7,01		
854-312	12,0	<i>Aspergillus</i>		Sédiments de surface	Port Giraud		DCA	EDM	Erlen	T	WSLE	18,07 ± 1,65		
906-318	313,3	<i>Penicillium</i>	<i>radicum</i>	Sédiments de surface	Port Giraud	MN133998.1	MEA	EDM	Erlen	T	EtOAc	-6,3 ± 7,10		
906-319	366,8	<i>Penicillium</i>	<i>radicum</i>	Sédiments de surface	Port Giraud	MN133998.1	YES	EDM	Erlen	T	EtOAc	5,47 ± 1,37		
927-464	10,0	<i>Trichoderma</i>	<i>atroviride</i>	Sédiments	Le Croisic	GU947796.1	DCA	EDMS	Pétri	M	MeOH puis CH ₂ Cl ₂	57,64 ± 0,32		
927-465	9,9	<i>Trichoderma</i>	<i>atroviride</i>	Sédiments	Le Croisic	GU947796.1	DCA	EDMS	Pétri	M	MeOH puis CH ₂ Cl ₂	64,26 ± 1,07		
940-331	77,9	<i>Pycnidiothora</i>	<i>dispersa</i>	Sédiments de surface	La Couplasse		CYA	EDM	Petri 16 cm	G	Smedsgaard puis CH ₂ Cl ₂	ND	2,0 (PPARα) 1,5 (PPARY)	
940-325	67,1	<i>Pycnidiothora</i>	<i>dispersa</i>	Sédiments de surface	La Couplasse		YES	EDM	Petri 16 cm	G	Smedsgaard puis CH ₂ Cl ₂	-8,46 ± 2,64	ND	
940-339		<i>Pycnidiothora</i>	<i>dispersa</i>	Sédiments de surface	La Couplasse		PDA	EDM	Petri 16 cm		Smedsgaard puis CH ₂ Cl ₂	ND	2,9 (PPARα) 1,7 (PPARY)	

Extracothèque	Souche						Culture			Extraction		Essais biologiques	
940-329	46,6	<i>Pycnidiophora</i>	<i>dispersa</i>	Sédiments de surface	La Couplasse		CYA	EDM	Petri 16 cm	G	Smedsgaard puis CH ₂ Cl ₂	-9,79 ± 2,48	1,6 (PPARα)
940-335	100,4	<i>Pycnidiophora</i>	<i>dispersa</i>	Sédiments de surface	La Couplasse		DCA	EDM	Petri 16 cm	M	Smedsgaard puis CH ₂ Cl ₂	-17,3 ± 3,02	1,7 (PPARα)
940-343	51,6	<i>Pycnidiophora</i>	<i>dispersa</i>	Sédiments de surface	La Couplasse		MEA	EDM	Petri 16 cm	M	Smedsgaard puis CH ₂ Cl ₂	-22,41 ± 0,31	
946-355	141,7	<i>Chrysosporium</i>	<i>queenslandicum</i>	Sédiments de surface	La Couplasse		DCA	EDM	Petri 16 cm	M	Smedsgaard puis CH ₂ Cl ₂	20,26 ± 4,6	
946-359	67,5	<i>Chrysosporium</i>	<i>queenslandicum</i>	Sédiments de surface	La Couplasse		PDA	EDM	Petri 16 cm	M	Smedsgaard puis CH ₂ Cl ₂	-22,77 ± 0,37	
946-363	35,3	<i>Chrysosporium</i>	<i>queenslandicum</i>	Sédiments de surface	La Couplasse		MEA	EDM	Petri 16 cm	M	Smedsgaard puis CH ₂ Cl ₂	-16,76 ± 3,77	

Annexe 3. Recherche bibliographique sur les métabolites produits par *Penicillium chrysogenum*

Molécule	Classe chimique	Formule brute	Masse mono-isotopique	Isolée de <i>Penicillium chrysogenum</i> [DOI]
versicolorin	anthraquinone	C ₁₈ H ₁₀ O ₇	338.0426	DOI:10.7164/antibiotics.56.459
sterigmatocystine	anthraquinone	C ₁₈ H ₁₂ O ₆	324.0634	DOI:10.7164/antibiotics.56.459
chrysofine/chrysofenine	NRPS/alcaloïde	C ₁₀ H ₁₀ N ₂ O ₂	190.0742	DOI:10.1042/bj0261907
2-pyrrovoylanthranilamide	NRPS ?	C ₁₀ H ₁₀ N ₂ O ₃	206.0691	DOI:10.1039/j39670002240
n-pyrrovoylanthranilamide	NRPS ?	C ₁₀ H ₁₀ N ₂ O ₃	206.0691	DOI:10.21775/9781910190395.09
2-(2-hydroxypropanamido)benzoic acid	NRPS ?	C ₁₀ H ₁₁ NO ₄	209.0688	DOI:10.1007/s10600-014-0971-3
2-(2-hydroxypropionylamino)benzamide	NRPS ?	C ₁₀ H ₁₂ N ₂ O ₃	208.0848	DOI:10.1039/j39670002240
cillifuranone	furanone	C ₁₀ H ₁₂ O ₄	196.0736	DOI:10.3390/md9040561
chrysoedone a	alcaloïde	C ₁₀ H ₁₅ NO ₂	181.1103	DOI:10.1021/np1008976
2(1h)-pyridinone, 5-[(1s)-1-hydroxybutyl]-6-(hydroxyméthyl)-	alcaloïde	C ₁₀ H ₁₅ NO ₃	197.1052	DOI:10.1021/np1008976
maculosin-5 cyclo(pro-val)	dicétopipérazine	C ₁₀ H ₁₆ N ₂ O ₂	196.1211	DOI:10.1515/znc-2013-9-1003
valylvaline	dicétopipérazine	C ₁₀ H ₁₈ N ₂ O ₂	198.1368	DOI:10.1016/j.micres.2015.11.001
biotin	Vitamine	C ₁₀ H ₁₈ N ₂ O ₃	214.1317	DOI:10.1042/bj1010601
desthiobiotin				
dethiobiotin				
vitamine b8				
chrysofenine	NRPS/alcaloïde	C ₁₀ H ₈ N ₂ O ₂	188.0586	DOI:10.1039/j39670002240
(+)-2,3-dihydroxypropyl 2,4-dihydroxy-6-méthylbenzoate	glycérol	C ₁₁ H ₁₄ O ₆	242.0790	DOI:10.3390/md9010059

Molécule	Classe chimique	Formule brute	Masse mono-isotopique	Isolée de <i>Penicillium chrysogenum</i> [DOI]
2-hydroxy-1-(hydroxyméthyl)ethyl 2,4-dihydroxy-6-méthylbenzoate	glycérol	C ₁₁ H ₁₄ O ₆	242.0790	DOI:10.3390/md9010059
phenol a	polycétide	C ₁₁ H ₁₆ O ₃	196.1099	DOI:10.1021/np500586r
cyclo-(pro-leu)	dicétopipérazine	C ₁₁ H ₁₈ N ₂ O ₂	210.1368	DOI:10.1007/s10600-014-0971-3
decarboxyldihydrocitrinone / stoloniferol b	polycétide	C ₁₂ H ₁₄ O ₄	222.0892	DOI:10.1021/np500586r
4-hydroxykigelin	polycétide	C ₁₂ H ₁₄ O ₆	254.0790	DOI:10.1016/j.bmcl.2011.03.076
6-hydroxykigelin	polycétide	C ₁₂ H ₁₄ O ₆	254.0790	DOI:10.1016/j.fitote.2017.01.005
penicimonoterpene	terpene	C ₁₂ H ₂₀ O ₅	244.1310	DOI:10.3390/md9010059
l-leucyl-d-leucine	dipeptide	C ₁₂ H ₂₄ N ₂ O ₃	244.1787	DOI:10.1016/j.micres.2015.11.001
(10s)-diaporthin	polycétide	C ₁₃ H ₁₄ O ₅	250.0841	DOI:10.1016/j.fitote.2017.01.005
sohironne A	phénol	C ₁₃ H ₁₆ O ₃	220.1099	DOI: 10.1021/np040137t
(-)-stoloniferol a	polycétide	C ₁₃ H ₁₆ O ₅	252.0998	DOI:10.1016/j.fitote.2017.01.005
(+)-stoloniferol a	polycétide	C ₁₃ H ₁₆ O ₅	252.0997	DOI:10.1016/j.fitote.2017.01.005
phenol, 3,3'-oxybis[4,6-dibromo-5-méthyl-	polycétide ?	C ₁₄ H ₁₀ Br ₄ O ₃	541.7363	DOI:10.1002/cbdv.201000067
penicitol c	polycétide	C ₁₄ H ₁₄ O ₄	246.0892	DOI:10.1021/np500586r
sorrentanone	benzoquinone	C ₁₄ H ₁₄ O ₄	246.0892	DOI:10.7164/antibiotics.48.520
l-prolyl-l-phenylalanine cyclo(l-prolyl-l-phenylalanyl) cyclo(phe-pro)	dicétopipérazine	C ₁₄ H ₁₆ N ₂ O ₂	244.1212	DOI:10.1007/s10600-014-0971-3
sohironne B	phenol	C ₁₄ H ₁₆ O ₄	248.1049	DOI:10.1021/np040137t

Molécule	Classe chimique	Formule brute	Masse mono-isotopique	Isolée de <i>Penicillium chrysogenum</i> [DOI]
sorbicillinol	sorbicillin	C ₁₄ H ₁₆ O ₄	248.1049	DOI:10.1021/np040137t
oxosorbicillinol	sorbilin	C ₁₄ H ₁₆ O ₅	264.0998	DOI:10.1021/np040137t
2',3'-dihydrosorbicillin	sorbicillin	C ₁₄ H ₁₈ O ₃	234.1255	DOI:10.1021/np040137t
sohirnone c	sorbicillin	C ₁₄ H ₁₈ O ₄	250.1205	DOI:10.1021/np040137t
5,10-dioethoxy-2,3,7,8-tetrahydro-1h,6h-dipyrrolo[1,2-a,1',2'-d]pyrazine		C ₁₄ H ₂₂ N ₂ O ₂	250.1681	DOI:10.1016/j.micres.2015.11.001
phenol, 2,4-bis(1,1-diméthylethyl)-	phénol	C ₁₄ H ₂₂ O	206.1670	DOI:10.1016/j.micres.2015.11.001
α-aminoadipoyl-serinyl-valine	tripeptide	C ₁₄ H ₂₅ N ₃ O ₇	347.1693	DOI:10.1002/hlca.19800630503
phenol, 2,4,6-tribromo-3-(2,4-dibromo-5-hydroxy-3-méthylphenoxy)-5-méthyl-	polycétide ?	C ₁₄ H ₉ Br ₅ O ₃	619.6469	DOI:10.1002/cbdv.201000067
penipanoid c	alcaloïde	C ₁₅ H ₁₀ N ₂ O ₃	266.0691	DOI:10.1002/hlca.201200433
chrysophanol	anthraquinone	C ₁₅ H ₁₀ O ₄	254.0579	DOI:10.1515/znc-2013-9-1003
alatinone	anthraquinone	C ₁₅ H ₁₀ O ₅	270.0528	DOI:10.1515/znc-2013-9-1003
emodin	anthraquinone	C ₁₅ H ₁₀ O ₅	270.0528	DOI:10.1515/znc-2013-9-1003
ω-hydroxyemodin citreorosein	anthraquinone	C ₁₅ H ₁₀ O ₆	270.0528	DOI:10.21775/9781910190395.09
2-(4-hydroxybenzyl) quinazolin-4(3h)-one	alcaloïde	C ₁₅ H ₁₂ N ₂ O ₂	252.0899	DOI:10.1002/hlca.201200433
2-naphthalenecarboxylic acid, 6-acetyl-5,7,8-trihydroxy-4-methoxy-, méthyl ester	polycétide	C ₁₅ H ₁₄ O ₇	306.0740	DOI:10.1016/j.fitote.2017.01.005
penixanacid a	polycétide	C ₁₆ H ₁₀ O ₈	330.0375	DOI:10.1021/np500586r
chrysotriazole b	triazol	C ₁₆ H ₁₅ N ₃ O ₃	297.1113	DOI:10.1002/hlca.201200433

Molécule	Classe chimique	Formule brute	Masse mono-isotopique	Isolée de <i>Penicillium chrysogenum</i> [DOI]
penicitrinol d	polycétide	C ₁₆ H ₂₂ O ₄	278.1518	DOI:10.1021/np500586r
penicitrinol e	polycétide	C ₁₆ H ₂₂ O ₄	278.1518	DOI:10.1021/np500586r
dehydrohistidyltryptophenyl-dicetopiperazine dhtd	alcaloïde	C ₁₇ H ₁₅ N ₅ O ₂	321.1226	DOI:10.1016/j.bmc.2015.11.038
chrysothiazole a	triazol	C ₁₇ H ₁₇ N ₃ O ₃	311.1270	DOI:10.1002/hlca.201200433
rezishanone c	sorbicilin	C ₁₈ H ₂₄ O ₅	320.1623	DOI:10.1021/np040137t
rezishanone d	Sorbicilin	C ₁₈ H ₂₆ O ₅	322.1780	DOI:10.1021/np040137t
galactosyl-lactose	oligosaccharide	C ₁₈ H ₃₂ O ₁₆	504.1690	DOI:10.1016/0040-4020(60)80060-9
acide oléique	acide gras	C ₁₈ H ₃₄ O ₂	282.2558	DOI:10.1515/znc-2013-9-1003
penicitide a	polycétide	C ₁₈ H ₃₄ O ₄	314.2457	DOI:10.3390/md9010059
penicitide b	polycétide	C ₁₈ H ₃₄ O ₅	330.2406	DOI:10.3390/md9010059
acide linoléique	acide gras polyinsaturé	C ₁₈ H ₃₈ O ₂	286.2872	DOI:10.1515/znc-2013-9-1003
5-methoxysterigmatocystin	anthraquinone	C ₁₉ H ₁₄ O ₇	354.0740	DOI:10.7164/antibiotics.56.459
butyrolactone ii	phenolic	C ₁₉ H ₁₆ O ₇	356.0896	DOI:10.1016/j.fitote.2017.01.005
rezishanone a	sorbicilin	C ₁₉ H ₂₂ O ₇	362.1366	DOI:10.1021/np040137t
penicitol a	polycétide	C ₁₉ H ₂₆ O ₄	C ₁₉ H ₂₆ O ₄	DOI:10.1021/np500586r
6,8-o-diméthylversicolorin a	anthraquinone	C ₂₀ H ₁₄ O ₆	350.0790	DOI:10.7164/antibiotics.56.459
versicolorin c	anthraquinone	C ₂₀ H ₁₄ O ₇	366.0739	DOI:10.7164/antibiotics.56.459
(-)-aversin	anthraquinone	C ₂₀ H ₁₆ O ₇	368.0896	DOI:10.7164/antibiotics.56.459
averufin	anthraquinone	C ₂₀ H ₁₆ O ₇	368.0896	DOI:10.7164/antibiotics.56.459

Molécule	Classe chimique	Formule brute	Masse mono-isotopique	Isolée de <i>Penicillium chrysogenum</i> [DOI]
norsolorinic acid	anthraquinone	C ₂₀ H ₁₈ O ₇	366.0740	DOI:10.7164/antibiotics.56.459
averantin	anthraquinone/polycétide	C ₂₀ H ₂₀ O ₇	372.1209	DOI:10.7164/antibiotics.56.459
chrysoenester	meroterpene	C ₂₀ H ₂₈ O ₄	332.1988	DOI:10.1021/acs.jnatprod.7b00846
rezishanone b	sorbicilin	C ₂₀ H ₂₈ O ₅	348.1937	DOI:10.1021/np040137t
conidiogenol b	diterpene / cycloplane	C ₂₀ H ₃₀ O	286.2297	DOI:10.1002/cbdv.201000378
conidiogenone c	diterpene / cycloplane	C ₂₀ H ₃₀ O ₂	302.2246	DOI:10.1002/cbdv.201000378
conidiogenone d	diterpene / cycloplane	C ₂₀ H ₃₀ O ₂	302.2246	DOI:10.1002/cbdv.201000378
conidiogenone f	diterpene / cycloplane	C ₂₀ H ₃₀ O ₂	302.2246	DOI:10.1002/cbdv.201000378
conidiogenone h	diterpene / cycloplane	C ₂₀ H ₃₂ O ₃	320.2351	DOI:10.1002/cbdv.201000378
conidiogenone i	diterpene / cycloplane	C ₂₀ H ₃₂ O ₃	320.2351	DOI:10.1002/cbdv.201000378
conidiogenol	diterpene / cycloplane	C ₂₀ H ₃₄ O ₂	306.2558	DOI:10.1002/cbdv.201000378
8-o-méthylaverufin	anthraquinone	C ₂₁ H ₁₈ O ₇	382.1052	DOI:10.7164/antibiotics.56.459
sorbicillactone a	polycétide/alcaloïde	C ₂₁ H ₂₃ NO ₈	417.1424	DOI:10.1016/j.tet.2005.05.026
sorbicillactone b	polycétide	C ₂₁ H ₂₅ NO ₈	419.1580	DOI:10.1016/j.tet.2005.05.026
6,8-o-diméthylaverufin	anthraquinone	C ₂₂ H ₂₀ O ₇	396.1209	DOI:10.7164/antibiotics.56.459
6,8-diméthylnidurufin	anthraquinone	C ₂₂ H ₂₀ O ₈	412.1158	DOI:10.7164/antibiotics.56.459
11-bromo-17-hydroxyroquefortine c	alcaloïde	C ₂₂ H ₂₃ BrN ₅ O ₃	484.0984	DOI:10.1002/hlca.201300447
11-bromoroquefortine c	alcaloïde	C ₂₂ H ₂₃ BrN ₅ O ₃	484.0984	DOI:10.1002/hlca.201300447
isoroquefortine c	alcaloïde	C ₂₂ H ₂₃ N ₅ O ₂	389.1852	DOI:10.1002/hlca.201300447

Molécule	Classe chimique	Formule brute	Masse mono-isotopique	Isolée de <i>Penicillium chrysogenum</i> [DOI]
roquefortine c	alcaloïde	C ₂₂ H ₂₃ N ₅ O ₂	389.1852	DOI:10.1016/j.bmc.2015.11.038
roquefortine l	alcaloïde	C ₂₂ H ₂₁ N ₅ O ₃	403.1644	DOI:10.1074/jbc.M113.512665
roquefortine m	alcaloïde	C ₂₂ H ₂₃ N ₅ O ₄	421.1750	DOI:10.1186/s12864-015-2154-4
roquefortine n	alcaloïde	C ₂₂ H ₂₃ N ₅ O ₆	453.1648	DOI:10.1186/s12864-015-2154-4
1,8- <i>o</i> -diméthylaverantin	anthraquinone	C ₂₂ H ₂₄ O ₇	400.1522	DOI:10.7164/antibiotics.56.459
bromoroquefortine d	alcaloïde	C ₂₂ H ₂₅ BrN ₅ O ₂	470.1192	DOI:10.1002/hlca.201300447
roquefortine d	alcaloïde	C ₂₂ H ₂₅ N ₅ O ₂	391.2008	DOI:10.1186/s12864-015-2154-4
7-deacetoxyanuthone	hydroquinone meroterpenoid	C ₂₂ H ₃₂ O ₃	344.2351	DOI:10.1021/np040137t
dimerumic acid	dicétopiperazine	C ₂₂ H ₃₆ N ₄ O ₈	484.2533	DOI:10.1016/bs.adgen.2016.01.005
meleagrin a	Indole alcaloïde	C ₂₃ H ₂₃ N ₅ O ₄	433.1750	DOI:10.1016/j.bmc.2015.11.038
penicitrinone a	polycétide	C ₂₃ H ₂₄ O ₅	380.1624	DOI:10.1016/j.fitote.2017.01.005
roquefortine f	alcaloïde/dicétopiperazine	C ₂₃ H ₂₅ N ₅ O ₃	419.1957	DOI:10.1186/s12864-015-2154-4
penicitrinol a	polycétide	C ₂₃ H ₂₆ O ₅	382.1780	DOI:10.1021/np400977e
sorbicatechol a	polycétide	C ₂₃ H ₂₆ O ₆	398.1729	DOI:10.1021/np400977e
sorbicatechol b	polycétide	C ₂₃ H ₂₆ O ₆	398.1729	DOI:10.1021/np400977e
penicitol b	polycétide	C ₂₃ H ₂₆ O ₇	414.1679	DOI:10.1021/np500586r
5-farnesyl-2-méthyl-1- <i>o</i> -méthylhydroquinone	meroterpene	C ₂₃ H ₃₄ O ₂	342.2559	DOI:10.1021/acs.jnatprod.7b00846
norcycloitrinol a	stérol	C ₂₃ H ₃₄ O ₄	374.2457	DOI:10.1002/hlca.201300153
bipenicilisorin	dimeric isocoumarin	C ₂₄ H ₁₈ O ₁₂	498.0798	DOI:10.1016/j.fitote.2017.01.005

Molécule	Classe chimique	Formule brute	Masse mono-isotopique	Isolée de <i>Penicillium chrysogenum</i> [DOI]
terremide d	NRPS?/alcaloïde	C ₂₄ H ₂₁ N ₃ O ₆	447.1430	DOI:10.1016/j.fitote.2017.01.005
méthyl 3,4,5-triméthoxy-2-[[2-[(3-pyridinylcarbonyl)amino]benzoyl]amino]benzoate	NRPS?	C ₂₄ H ₂₃ N ₃ O ₇	465.1536	DOI:10.1016/j.fitote.2017.01.005
butyrolactone i	lactone	C ₂₄ H ₂₄ O ₇	424.1522	DOI:10.1016/j.fitote.2017.01.005
olomoucine	phenolic	C ₂₄ H ₂₄ O ₇	424.1522	DOI:10.1016/j.fitote.2017.01.005
butyrolactone v	phenolic	C ₂₄ H ₂₄ O ₈	440.1471	DOI:10.1016/j.fitote.2017.01.005
penicitrinone f	polycétide ?	C ₂₄ H ₂₆ O ₅	394.1780	DOI:10.1016/j.fitote.2017.01.005
sorbifuranone a	Sorbicillinoids	C ₂₄ H ₂₈ O ₈	444.1784	DOI:10.1016/j.tet.2010.10.057
sorbifuranone c	Sorbicillinoids	C ₂₄ H ₂₈ O ₈	444.1784	DOI:10.1016/j.tet.2010.10.057
sorbifuranone b	Sorbicillinoids	C ₂₄ H ₃₀ O ₈	446.1941	DOI:10.1016/j.tet.2010.10.057
aspernolide c	butyrolactone	C ₂₅ H ₂₆ O ₇	438.1679	DOI:10.1016/j.fitote.2017.01.005
24-epicycloctriinol	stérol	C ₂₅ H ₃₆ O ₄	400.2614	DOI:10.1002/hlca.201300153
cycloctriinol	stérol	C ₂₅ H ₃₆ O ₄	400.2614	DOI:10.1002/hlca.201300153
neocycloctriinol a	stérol	C ₂₅ H ₃₆ O ₄	400.2614	DOI:10.1002/hlca.201300153
neocycloctriinol b	stérol	C ₂₅ H ₃₆ O ₄	400.2614	DOI:10.1002/hlca.201300153
neocycloctriinol c	stérol	C ₂₅ H ₃₆ O ₄	400.2614	DOI:10.1002/hlca.201300153
neocycloctriinol d	stérol	C ₂₅ H ₃₆ O ₄	400.2614	DOI:10.1002/hlca.201300153
pseudocycloctriinol a	stérol	C ₂₅ H ₃₆ O ₄	400.2614	DOI:10.1002/hlca.201300153
pseudocycloctriinol a	stérol	C ₂₅ H ₃₆ O ₄	400.2614	DOI:10.1002/hlca.201300153
erythro-11 α -hydroxyneocycloctriinol	C-25 stérol	C ₂₅ H ₃₆ O ₅	416.2563	DOI:10.1002/hlca.201300153

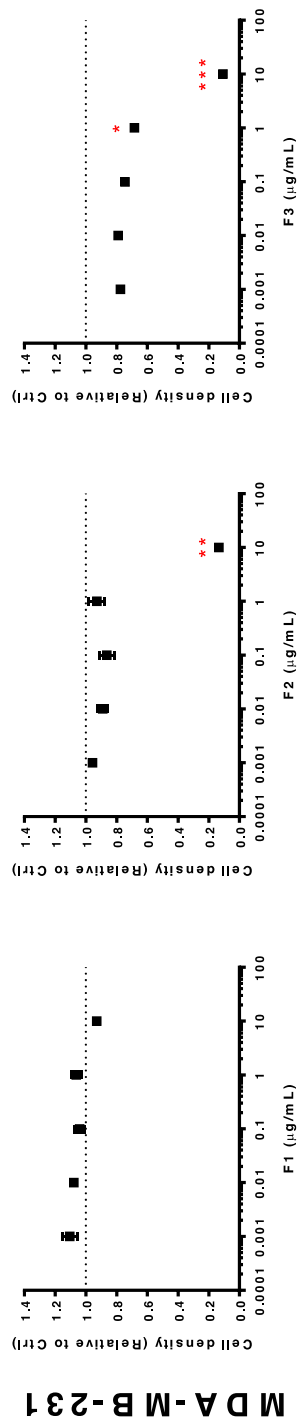
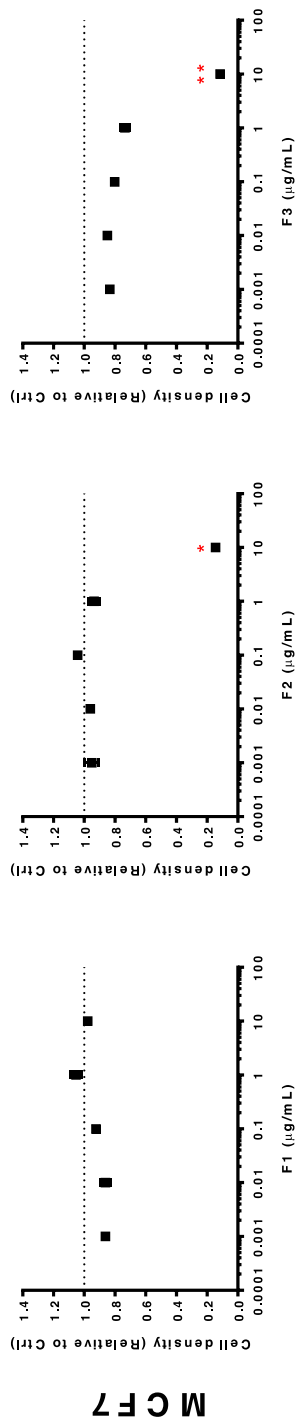
Molécule	Classe chimique	Formule brute	Masse mono-isotopique	Isolée de <i>Penicillium chrysogenum</i> [DOI]
chrysamide C	epoxyamides	C ₂₆ H ₂₆ N ₄ O ₁₀	554.1649	DOI:10.1021/acs.orglett.6b01699
chrysamide A	epoxyamides	C ₂₆ H ₂₆ N ₄ O ₉	538.1699	DOI:10.1021/acs.orglett.6b01699
chrysamide B	epoxyamides	C ₂₆ H ₂₈ N ₄ O ₈	524.1907	DOI:10.1021/acs.orglett.6b01699
xanthoviridicatine E	polycétide	C ₂₇ H ₂₀ O ₉	488.1107	DOI:10.1002/hlca.200390281
xanthoviridicatine F	polycétide	C ₂₇ H ₂₀ O ₉	488.1107	DOI:10.1002/hlca.200390281
bisorbibutenolide	dérivé de sorbicillin	C ₂₈ H ₃₂ O	384.2453	DOI:10.1021/np040137t
bisorbicillinol	bisorbicillinoïdes	C ₂₈ H ₃₂ O ₈	496.2097	DOI:10.21775/9781910190395.09
trichodimérol	hétérotetracyclique	C ₂₈ H ₃₂ O ₈	496.2097	DOI:10.7164/antibiotics.49.234
bisvertinolone	bisorbicillinoïdes	C ₂₈ H ₃₂ O ₉	512.2046	DOI:10.1016/j.tet.2005.05.026
yaminterritrem C	merosquiterpénoïde	C ₂₈ H ₃₄ O ₇	482.2305	DOI:10.1016/j.fitote.2017.01.005
bisvertinoquinol	polycétide	C ₂₈ H ₃₄ O ₈	498.2254	DOI:10.21775/9781910190395.09
citrehybridonol	meroterpénoïde	C ₂₈ H ₃₆ O ₈	500.2410	DOI: 10.15625/2525-2518/55/1A/12383
tetrahydrobisvertinolone	bisorbicillinoïdes	C ₂₈ H ₃₆ O ₉	516.2359	DOI:10.1186/s12864-015-2154-4
chrysoygenamide A	alcaloïde	C ₂₈ H ₃₇ N ₃ O ₂	447.2886	DOI:10.1038/ja.2008.114
andrastine A	meroterpénoïde	C ₂₈ H ₃₈ O ₇	486.2617	DOI:10.15625/2525-2518/55/1A/12383
yanuthone D	meroterpénoïde	C ₂₈ H ₃₈ O ₈	502.2567	DOI:10.3389/fmicb.2018.02768
ergone	stérol	C ₂₈ H ₄₀ O	392.3079	DOI:10.1016/j.bmcl.2011.03.076
andrastine B	meroterpénoïde	C ₂₈ H ₄₀ O ₇	488.2774	DOI:10.5598/ima fungus.2011.02.01.12
cyathisterone	stérol	C ₂₈ H ₄₂ O ₂	410.3185	DOI:10.1016/j.bmcl.2011.03.076

Molécule	Classe chimique	Formule brute	Masse mono-isotopique	Isolée de <i>Penicillium chrysogenum</i> [DOI]
ergostérol	stérol	C ₂₈ H ₄₄ O	396.3392	DOI:10.1007/s10600-014-0971-3
ergosta-5,7,22-triene-3β-ol				
ergosterol 5,8-endoperoxyde (3β,5α,8α,22e)-5,8-epidioxyergosta-6,22-dien-3-ol peroxyergosterol	stérol	C ₂₈ H ₄₄ O ₃	428.3290	DOI:10.1016/j.bmcl.2011.03.076
brassicasterol	stérol	C ₂₈ H ₄₆ O	398.3549	DOI:10.1007/s10600-014-0851-x
cervisterol	stérol	C ₂₈ H ₄₆ O ₃	430.3447	DOI:10.1016/j.bmcl.2011.03.076
ergosta-7,22-diene-3β,5α,6β-triol				
(3β,5α,6β,22e)-ergosta-7,22-diene-3,6-diol	stérol	C ₂₈ H ₄₈ O ₂	416.3654	DOI:10.1016/j.bmcl.2011.03.076
penicisteroid B	stérol	C ₃₀ H ₄₈ O ₄	472.3552	DOI:10.1016/j.bmcl.2011.03.076
anicequol	stérol	C ₃₀ H ₄₈ O ₆	504.3450	DOI:10.1016/j.bmcl.2011.03.076
penicisteroid A	stérol	C ₃₀ H ₅₀ O ₆	506.3607	DOI:10.1016/j.bmcl.2011.03.076
secalonic acid D	polycétide	C ₃₂ H ₁₄ O ₄	462.0892	DOI: 10.5598/imafungus.2011.02.01.12
secalonic acid F	polycétide	C ₃₂ H ₃₀ O ₁₄	638.1636	DOI: 10.5598/imafungus.2011.02.01.12
penochalasin I	alcaloïde	C ₃₂ H ₃₄ N ₂ O ₄	510.2519	DOI:10.1016/j.fitote.2017.09.016
penochalasin K	alcaloïde	C ₃₂ H ₃₄ N ₂ O ₄	510.2519	DOI:10.1016/j.fitote.2017.09.016
chaetoglobosin A	alcaloïde	C ₃₂ H ₃₆ N ₂ O ₅	528.2624	DOI:10.1016/j.fitote.2017.09.016
chaetoglobosin C	alcaloïde	C ₃₂ H ₃₆ N ₂ O ₅	528.2624	DOI:10.1016/j.fitote.2017.09.016
chaetoglobosin G	alcaloïde	C ₃₂ H ₃₆ N ₂ O ₅	528.2624	DOI:10.3390/md14100172
armochaetoglobosin I	alcaloïde	C ₃₂ H ₃₈ N ₂ O ₃	498.2882	DOI:10.3390/md14100172
penochalasin J	alcaloïde	C ₃₂ H ₃₈ N ₂ O ₄	514.2832	DOI:10.3390/md14100172

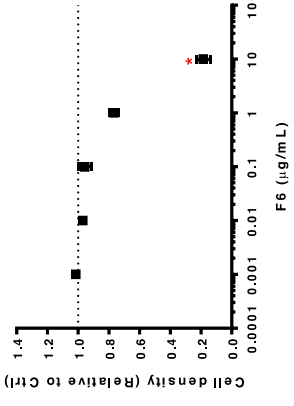
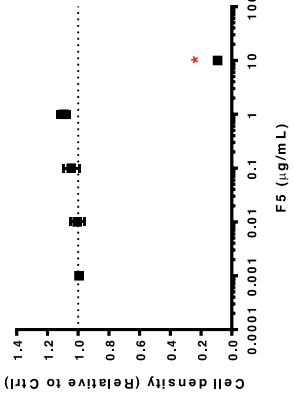
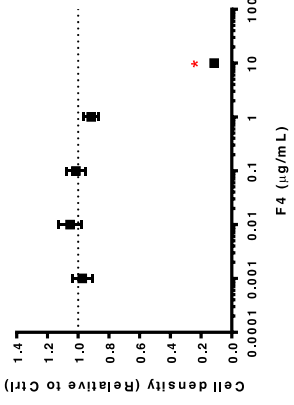
Molécule	Classe chimique	Formule brute	Masse mono-isotopique	Isolée de <i>Penicillium chrysogenum</i> [DOI]
chaetoglobosin E	alcaloïde	C ₃₂ H ₃₈ N ₂ O ₅	530.2781	DOI:10.3390/md14100172
chaetoglobosin F	alcaloïde	C ₃₂ H ₃₈ N ₂ O ₅	530.2781	DOI:10.3390/md14100172
cytoglobosin C	alcaloïde	C ₃₂ H ₃₈ N ₂ O ₅	530.2781	DOI:10.3390/md14100172
chrysoreside A	cérébroside	C ₄₀ H ₇₃ NO ₉	711.5285	DOI:10.1021/np1008976
chrysoreside C	cérébroside	C ₄₀ H ₇₃ NO ₉	711.5285	DOI:10.1021/np1008976
chrysoreside B	cérébroside	C ₄₁ H ₇₅ NO ₉	725.5441	DOI:10.1021/np1008976
chrysoreside E	cérébroside	C ₄₁ H ₇₇ NO ₉	727.5598	DOI:10.1021/np1008976
chrysoreside D	cérébroside	C ₄₃ H ₇₉ NO ₉	753.5754	DOI:10.1021/np1008976
uracil	acide nucléique	C ₄ H ₄ N ₂ O ₂	112.0272	DOI:10.1515/znc-2013-9-1003
ergost-22-ene-3,6,7,11,16-pentol, 16-acetate 3,6-bis[(<i>cr</i>)- <i>α</i> -methoxy- <i>α</i> -(trifluorométhyl)benzeneacetate], (3 β ,5 α ,6 β ,7 β ,11 β ,16 β ,22e)-	stérol	C ₅₀ H ₆₄ F ₆ O ₁₀	938.4403	DOI:10.1016/j.bmcl.2011.03.076
4-hydroxy-6-(méthylamino)-5-nitropyrimidine	pyrimidine	C ₅ H ₆ N ₄ O ₃	170.0440	DOI:10.1016/j.micres.2015.11.001
1-pyrrolidinecarboxaldéhyde	azolidine	C ₅ H ₉ NO	99.0684	DOI:10.1016/j.micres.2015.11.001
1-méthyl-2-pyrrolidone-4-carboxamide		C ₆ H ₁₀ N ₂ O ₂	142.0742	DOI:10.1016/j.micres.2015.11.001
lactone déhydromévalonique	lactone	C ₆ H ₈ O ₂	112.0524	DOI:10.1016/j.micres.2015.11.001
pyrrolo[1,2- <i>a</i>]pyrazine-1,4-dione, hexahydro-		C ₇ H ₁₀ N ₂ O ₂	154.0742	DOI:10.1016/j.micres.2015.11.001
hexahydropyrrolizin-3-one		C ₇ H ₁₁ NO	125.0840	DOI:10.1016/j.micres.2015.11.001
2-propenal, 3-(1-aziridiny)-3-(diméthylamino)-		C ₇ H ₁₂ N ₂ O	140.0949	DOI:10.1016/j.micres.2015.11.001

Molécule	Classe chimique	Formule brute	Masse mono-isotopique	Isolée de <i>Penicillium chrysogenum</i> [DOI]
glycyl-l-proline	dicétopipérazine	C ₇ H ₁₂ N ₂ O ₃	172.0847	DOI:10.1016/j.micres.2015.11.001
acide propionique 3-pyrrolidin-2-y		C ₇ H ₁₃ NO ₂	143.0946	DOI:10.1016/j.micres.2015.11.001
2H-pyran-2-one, 5,6-dihydro-4-(1E)-1-propen-1-yl-	pyrone	C ₈ H ₁₀ O ₂	138.0681	DOI:10.1016/j.fitote.2017.01.005
4(3H)-pyrimidinone, 3-ethyl-2,6-diméthyl-	pyrimidinone	C ₈ H ₁₂ N ₂ O	152.0950	DOI:10.1016/j.micres.2015.11.001
pipéridine, 1-(cyanoacetyl)-	pipéridine	C ₈ H ₁₂ N ₂ O	152.0950	DOI:10.1016/j.micres.2015.11.001
6-oxabicyclo[3.1.0]hexan-3-one, 2,2,4,4-tetraméthyl-		C ₉ H ₁₄ O ₂	154.0994	DOI:10.1016/j.micres.2015.11.001
cyclo-L-ala-L-leu	dicétopipérazine	C ₉ H ₁₆ N ₂ O ₂	184.1211	DOI:10.1515/znc-2013-9-1003
DL-alanyl-L-leucine	dipeptide	C ₉ H ₁₈ N ₂ O ₃	202.1317	DOI:10.1016/j.micres.2015.11.001
n-[(1E)-2-(4-hydroxyphényl)éthényl]formamide	alcaloïde	C ₉ H ₉ NO ₂	163.0633	DOI:10.1002/hlca.201200433
n-[(1Z)-2-(4-hydroxyphényl)éthényl]formamide	alcaloïde	C ₉ H ₉ NO ₂	163.0633	DOI:10.1002/hlca.201200433
n-formyl-4-hydroxybenzèneacetamide	alcaloïde	C ₉ H ₉ NO ₃	179.0582	DOI:10.1002/hlca.201200433
citréoisocoumarine	coumarine	C ₁₄ H ₁₂ O ₆	276.0633	DOI: 10.5598/imafungus.2011.02.01.12
conidiogène B	diterpène tetracyclique	C ₂₀ H ₃₀ O	286.2296	DOI:10.1002/cbdv.201000378
fungisporin	NRPS	C ₂₈ H ₃₆ N ₄ O ₄	492.2736	DOI:10.3389/fmicb.2018.02768
glandicoline a	alcaloïde indolique	C ₂₂ H ₂₁ N ₅ O ₃	403.1644	DOI:10.1186/s12864-015-2154-4
glandicoline b	alcaloïde indolique	C ₂₂ H ₂₁ N ₅ O ₄	419.1593	DOI:10.1186/s12864-015-2154-4
peniméthavone a	polycétide/ flavone	C ₁₄ H ₁₂ O ₆	276.0633	DOI:10.1080/14786419.2016.1163695
pr-toxin	sesquiterpène	C ₁₇ H ₂₀ O ₆	320.1259	DOI:10.21775/9781910190395.09

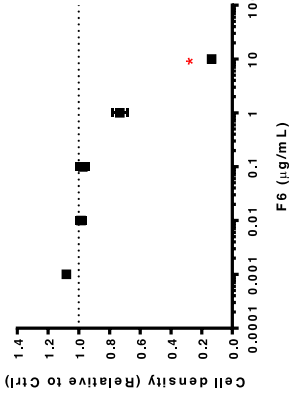
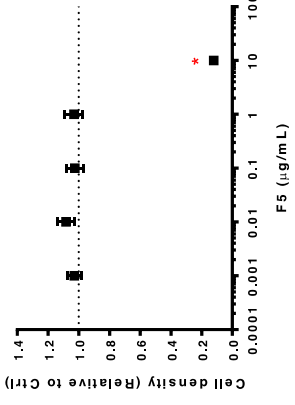
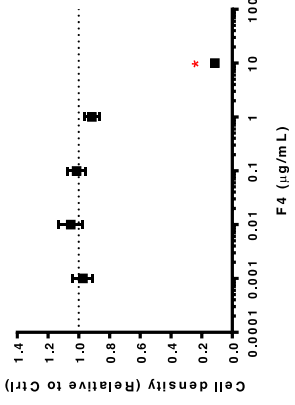
Annexe 4. Résultats de cytotoxicité obtenu pour les fractions MMS5-526-1 → MMS5-526-17 sur les lignées MCF-7 et MDA-MB-231 par la méthode du Crystal Violet



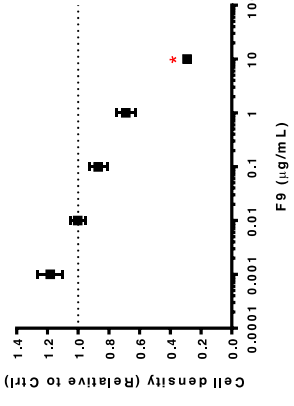
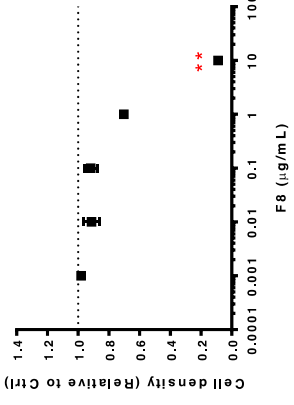
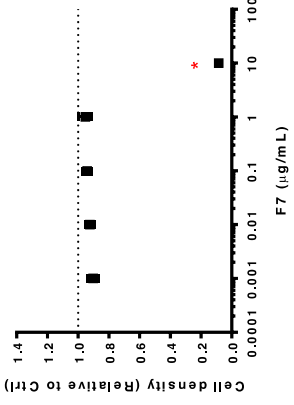
MCF7



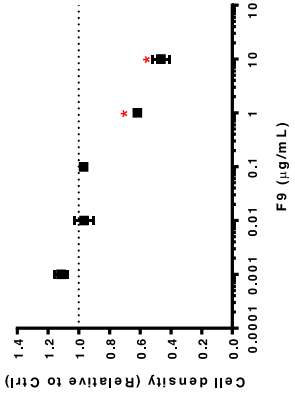
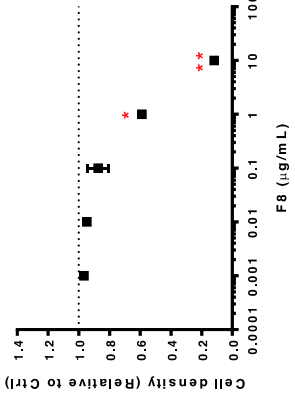
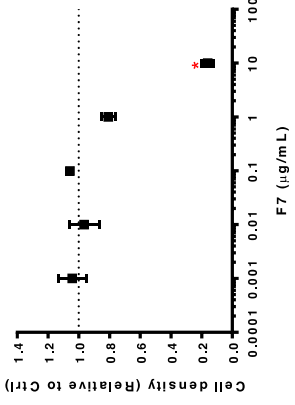
MDA-MB-231



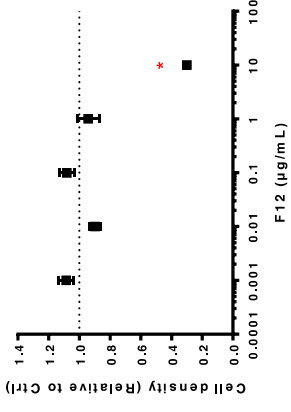
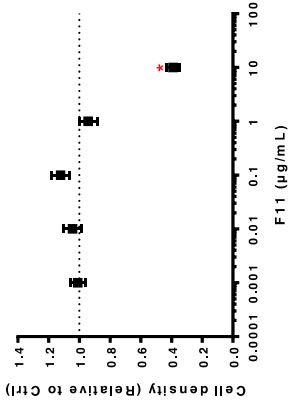
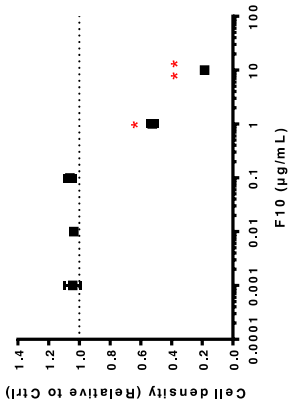
MCF7



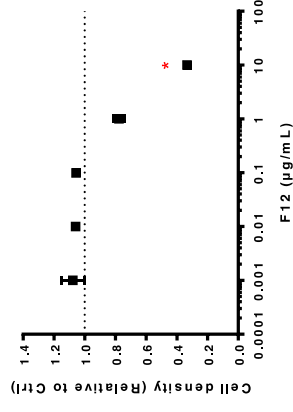
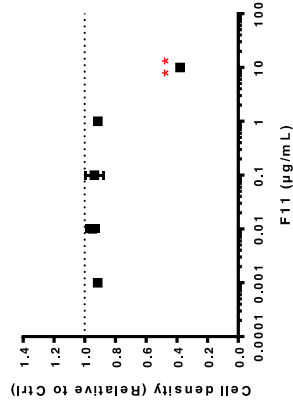
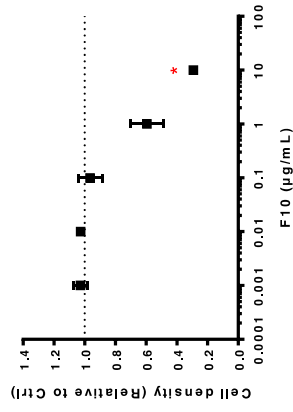
MDA-MB-231



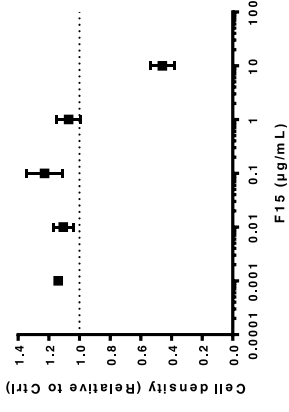
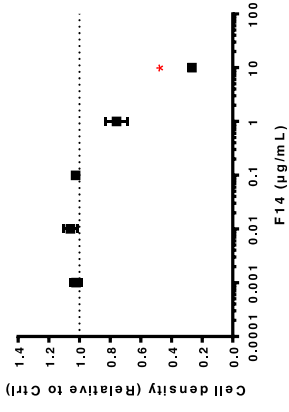
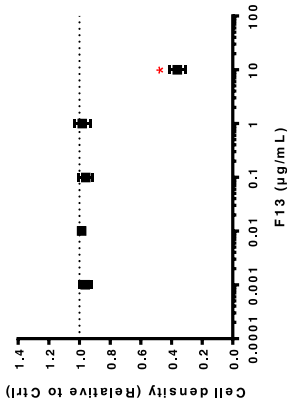
MCF7



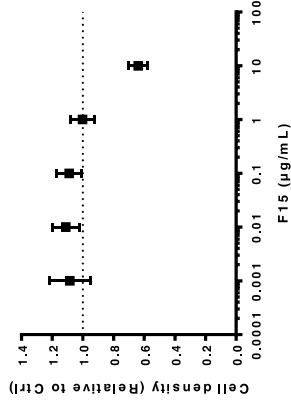
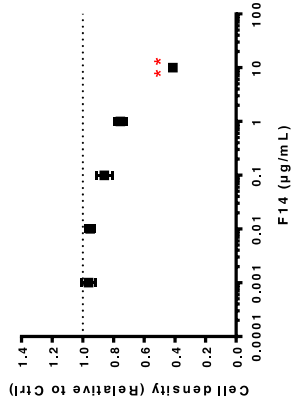
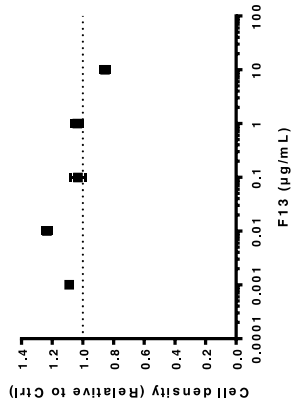
MDA-MB-231



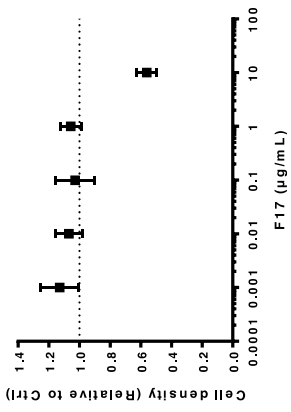
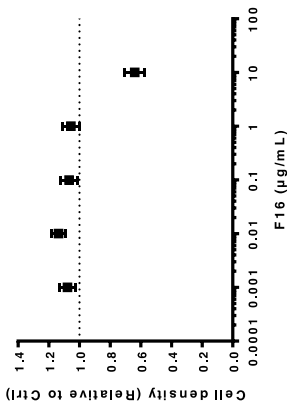
MC F7



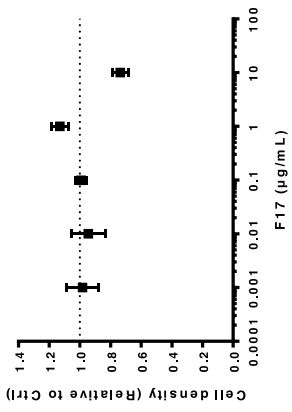
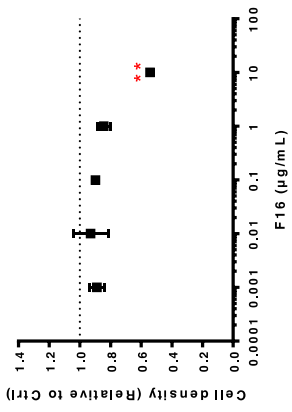
MDA-MB-231



MCF7

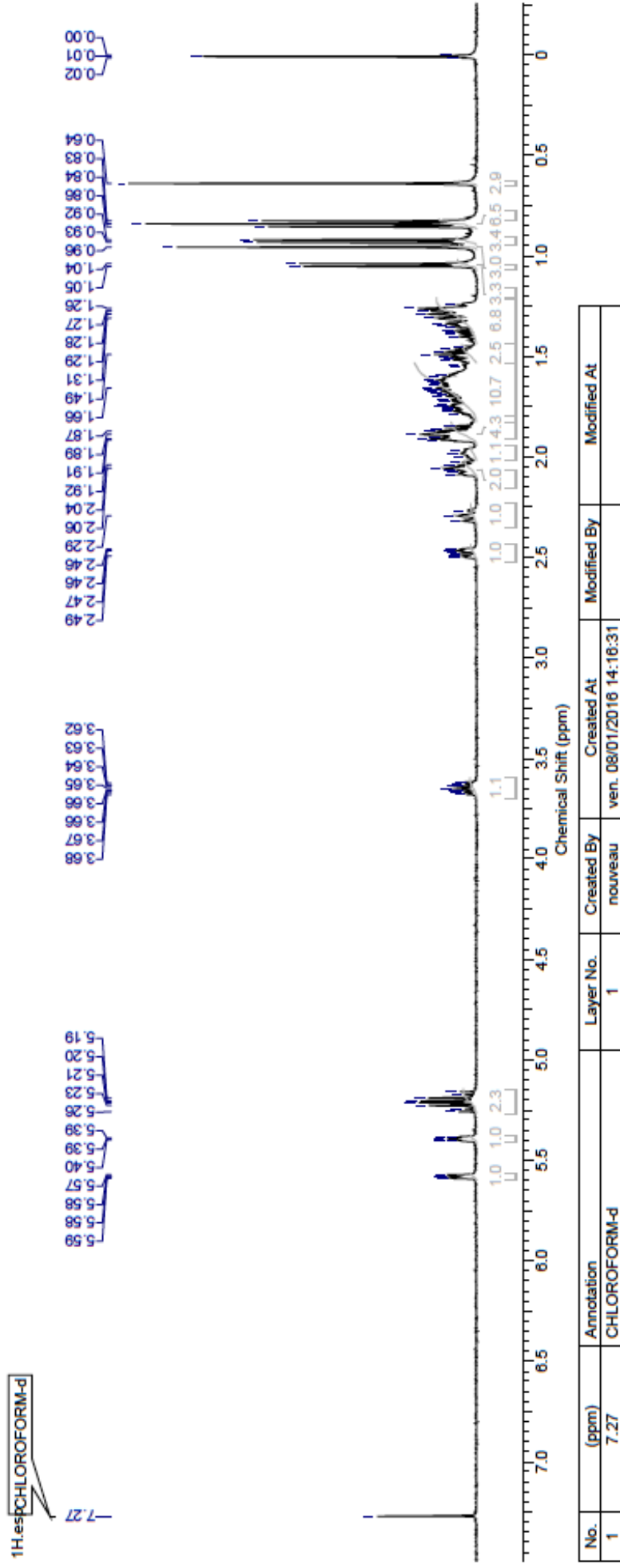


MDA-MB-231



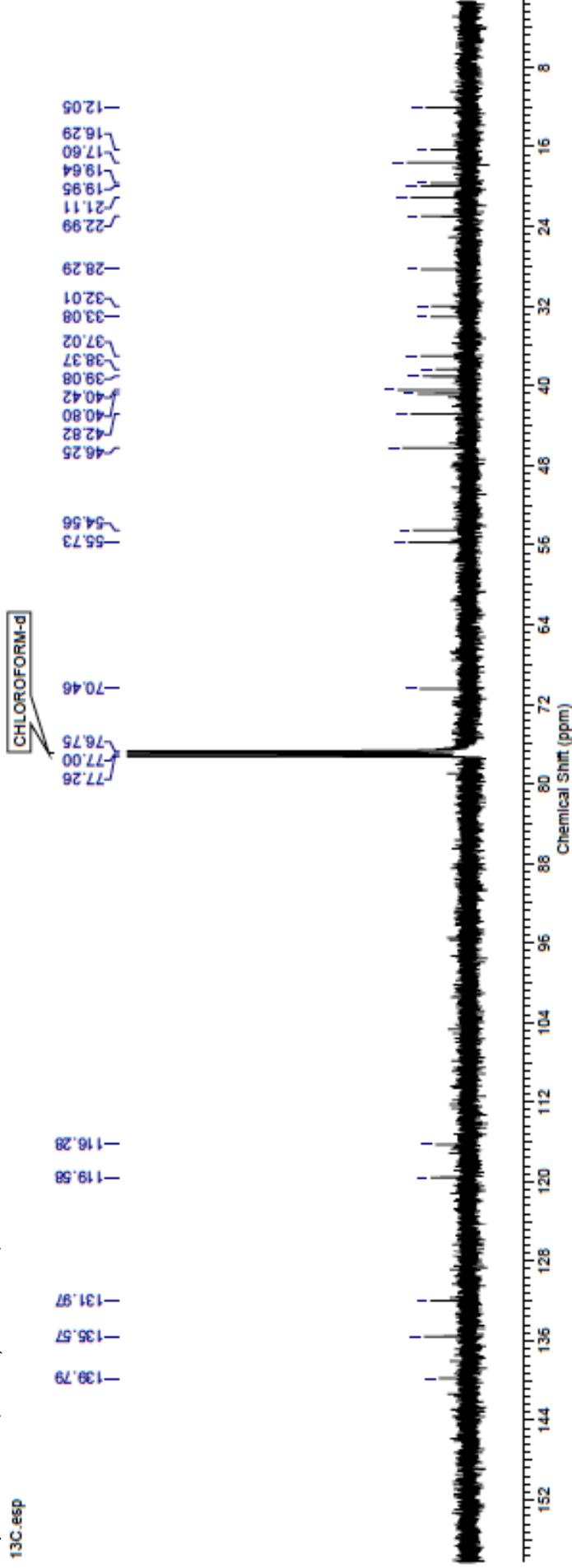
Annexe 5. Spectres RMN de l'ergostérol

- Spectre ^1H , (CDCl_3 , 500 MHz)



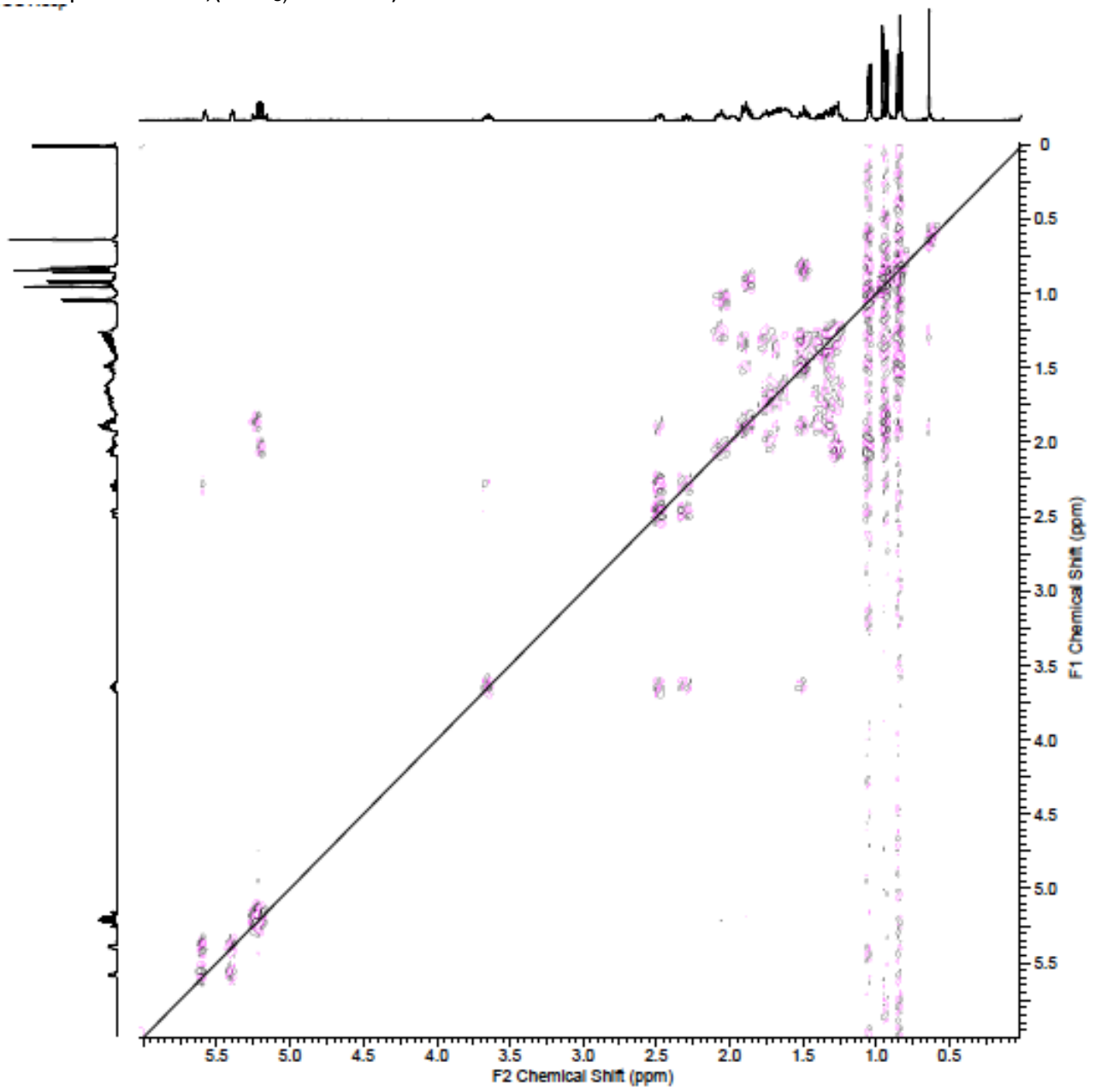
- Spectre ^{13}C , (CDCl_3 , 125 MHz)

13C.esp

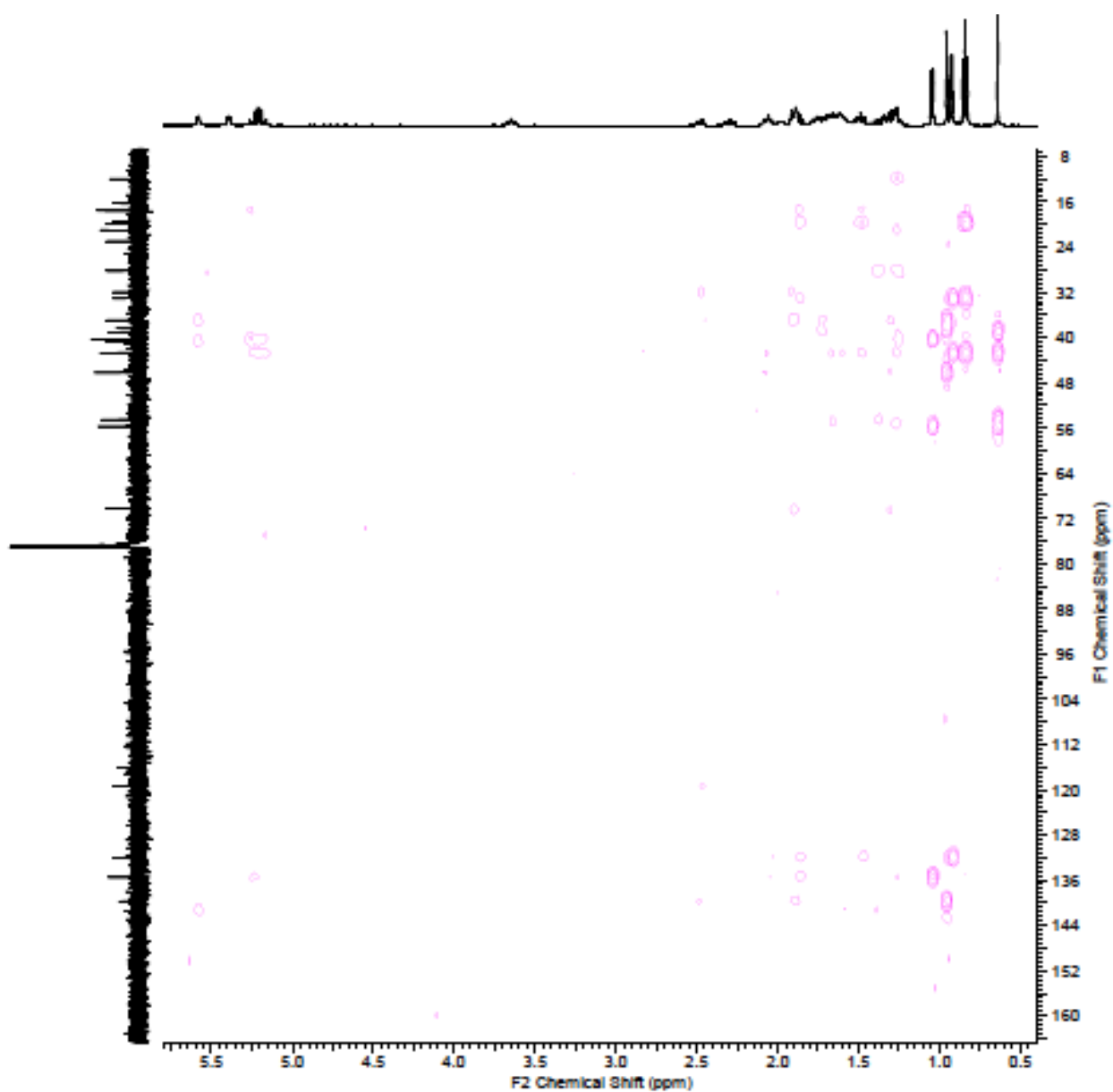


No.	(ppm)	Annotation	Layer No.	Created By	Created At	Modified By	Modified At
1	77.00	CHLOROFORM-d	1	nouveau	ven. 08/01/2016 14:58:57		

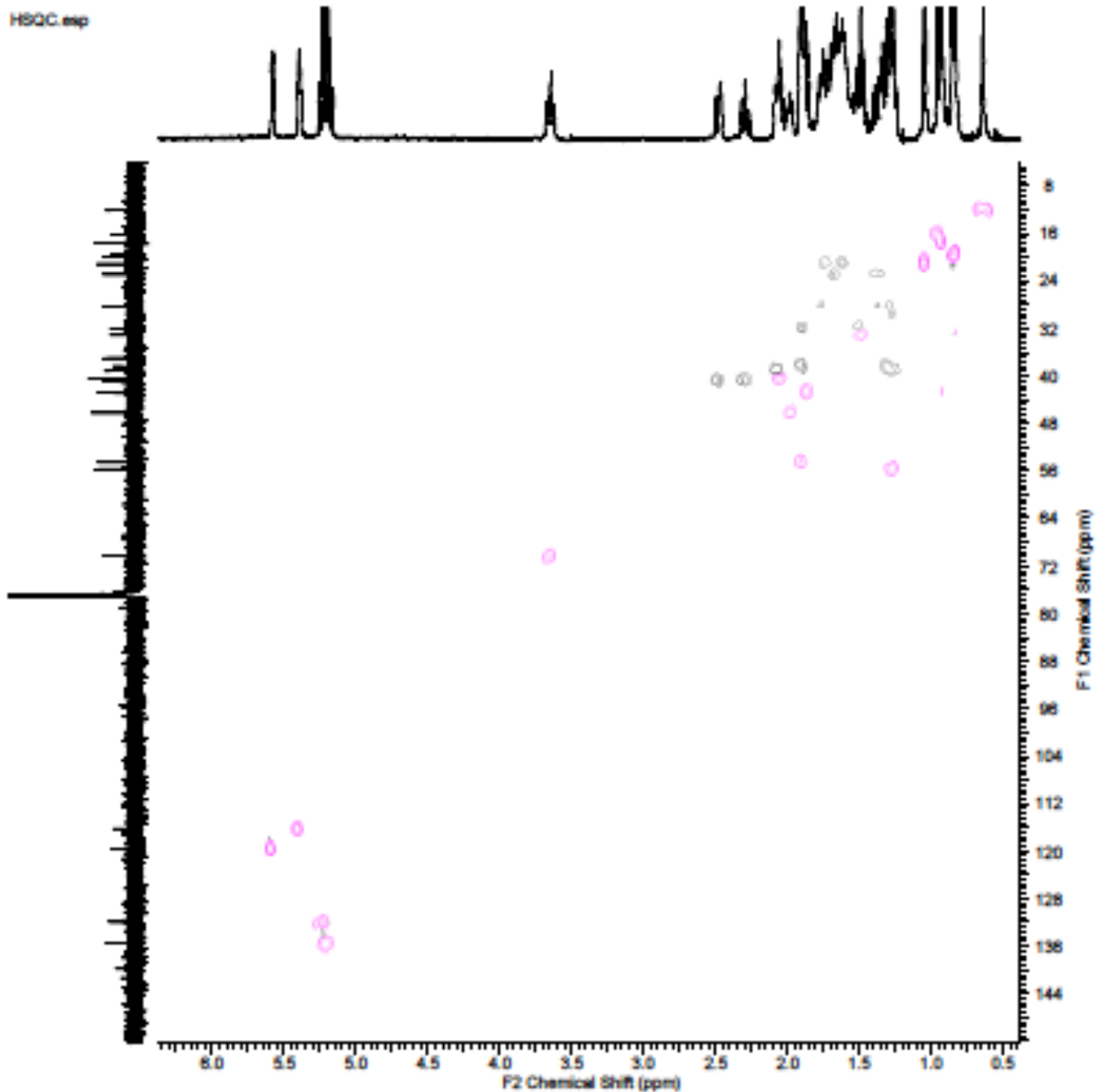
- Spectre COSY, (CDCl₃, 400 MHz)



- Spectre HMBC, (CDCl₃, 400 MHz)



- Spectre HSQC, (CDCl₃, 400 MHz)



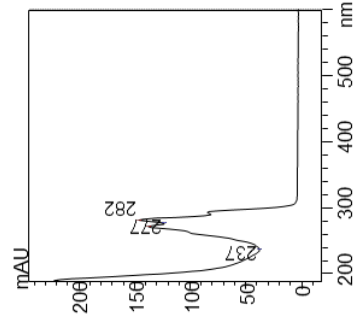
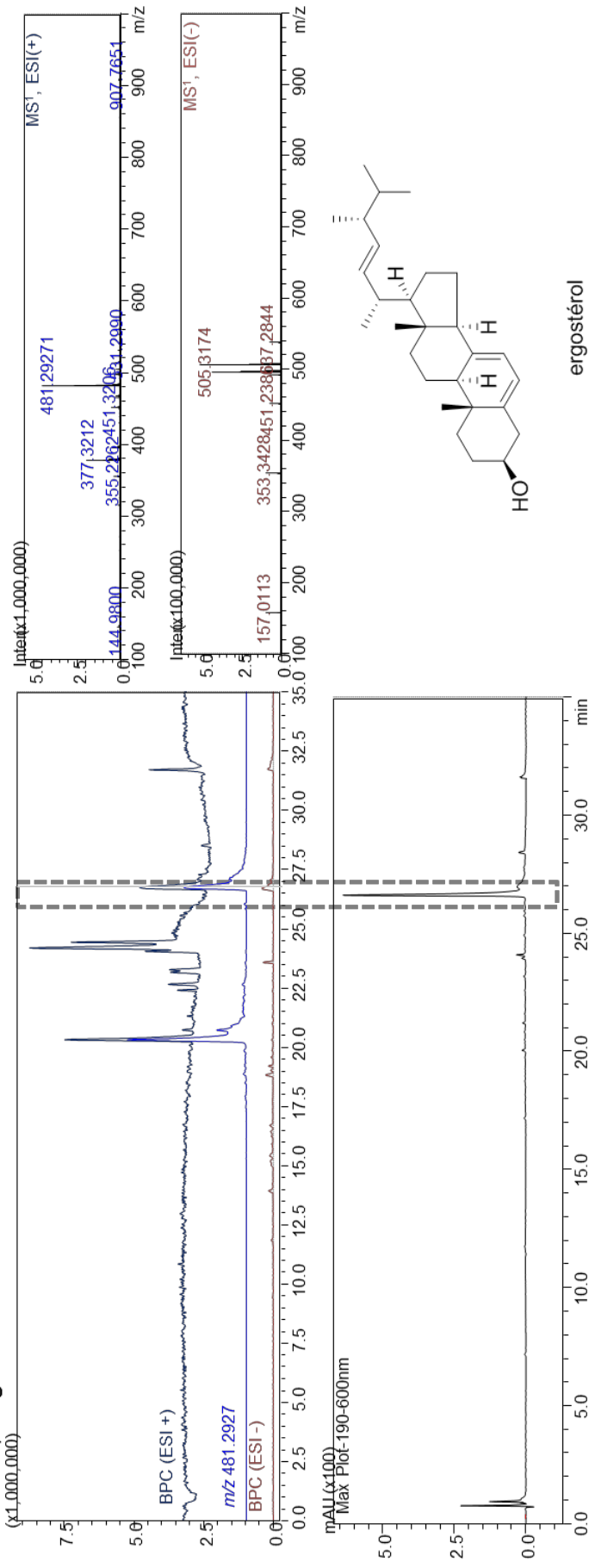
Annexe 6. Les profils LC-MS de l'ergostérol avec le tracé BPC (base peak chromatogram) en mode positif (ESI+) et négatif (ESI-) et les tracés EICs (extracted ion chromatogram) des ions détectés.

Code : MMS 5-526-5-7

Métabolite produit par : *P. chrysogenum*

Masses : 0,9 mg

- Molécule isolée
- Molécule détectée



Proposition(s) de formule(s) brute(s) :

$C_{28}H_{44}O$

Proposition(s) de structure(s) :

Ergostérol

- Annotation via bases de données
- Analyse par spectroscopie RMN

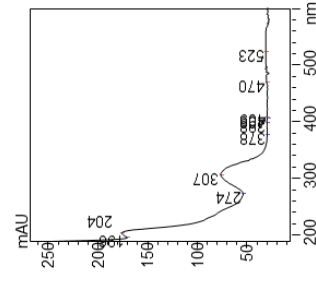
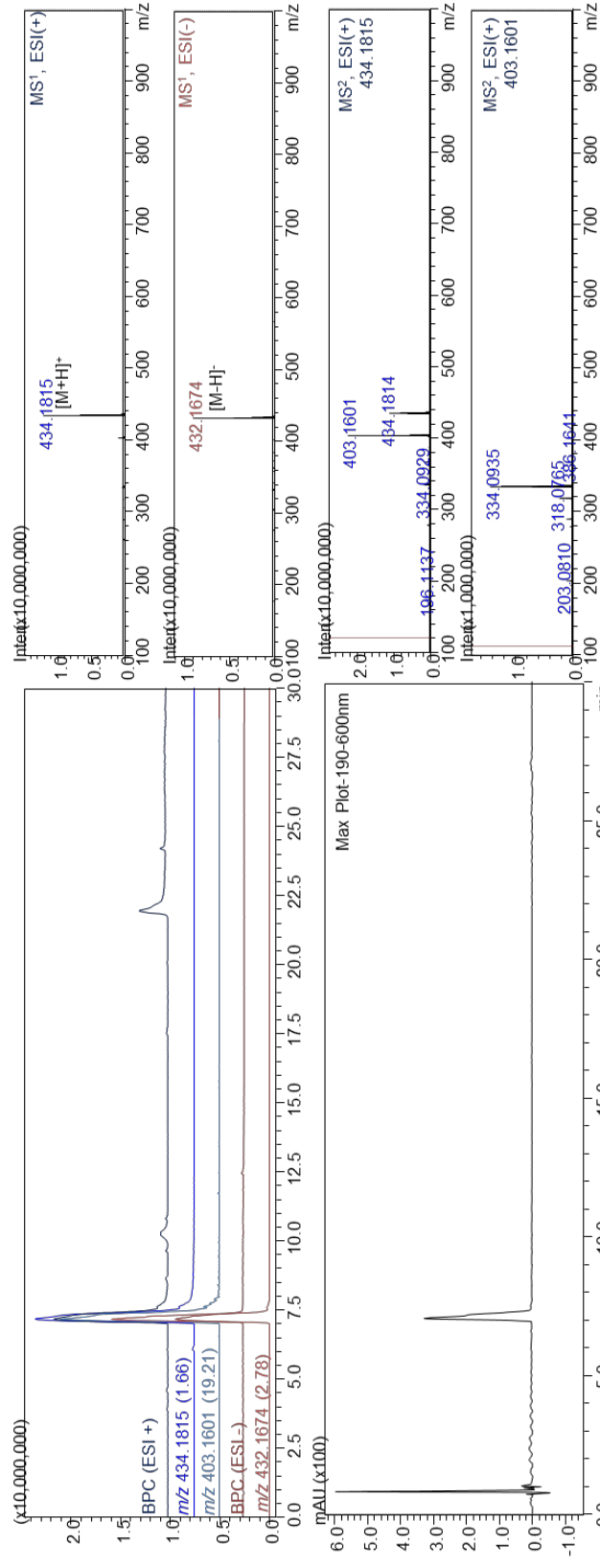
Annexe 7. Les profils LC-MS de la méléagrine avec le tracé BPC (base peak chromatogram) en mode positif (ESI+) et négatif (ESI-) et les tracés EICs (extracted ion chromatogram) des ions détectés.

Code : CTBG-04-*m/z* 434

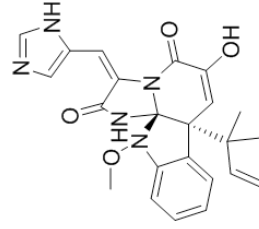
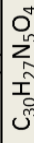
Métabolite produit par : *P. chrysogenum*

Masses : 3,3 mg

- Molécule isolée
- Molécule détectée



Proposition(s) de formule(s) brute(s) :



méléagrine

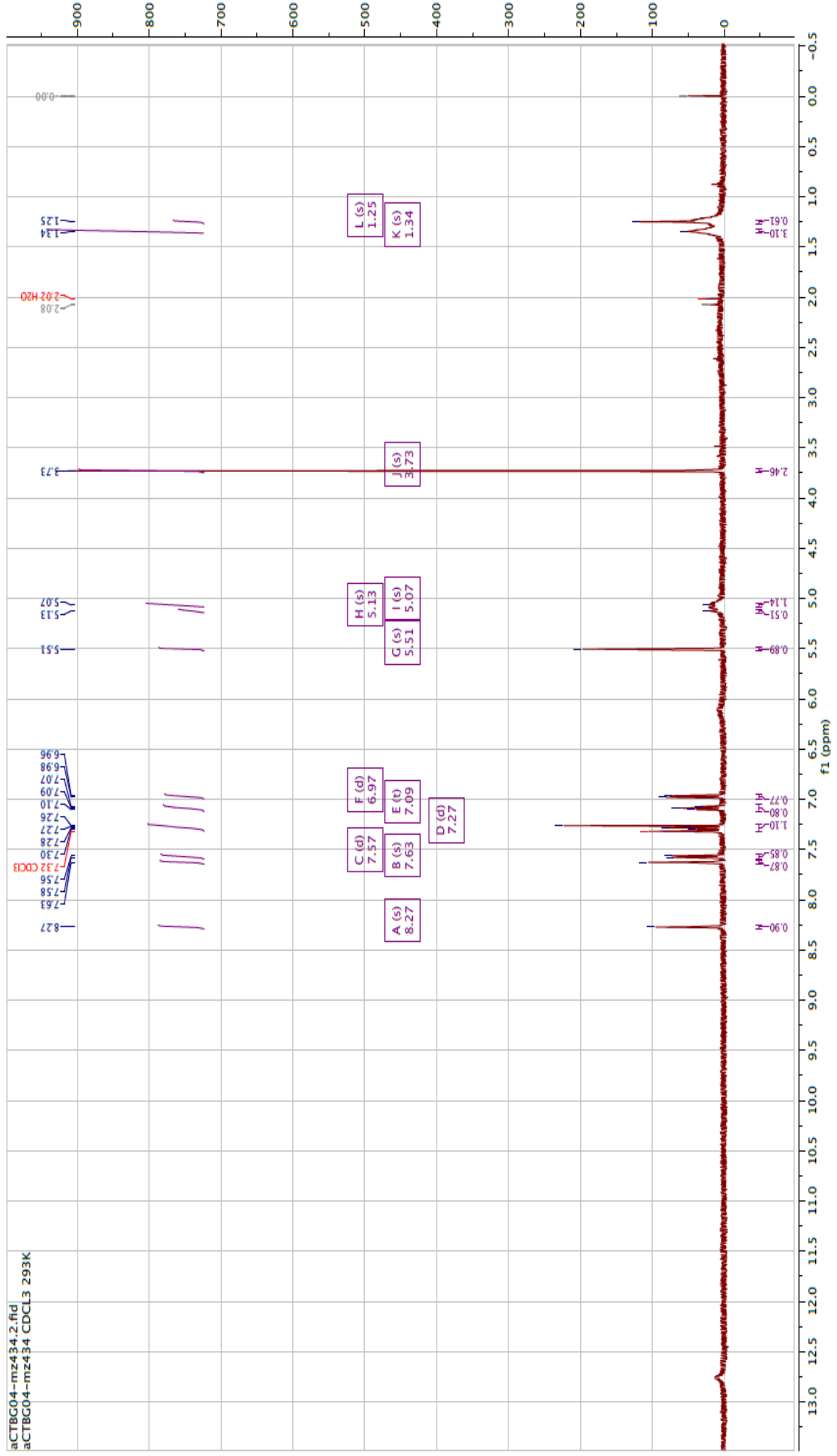
Proposition(s) de structure(s) :

Méléagrine

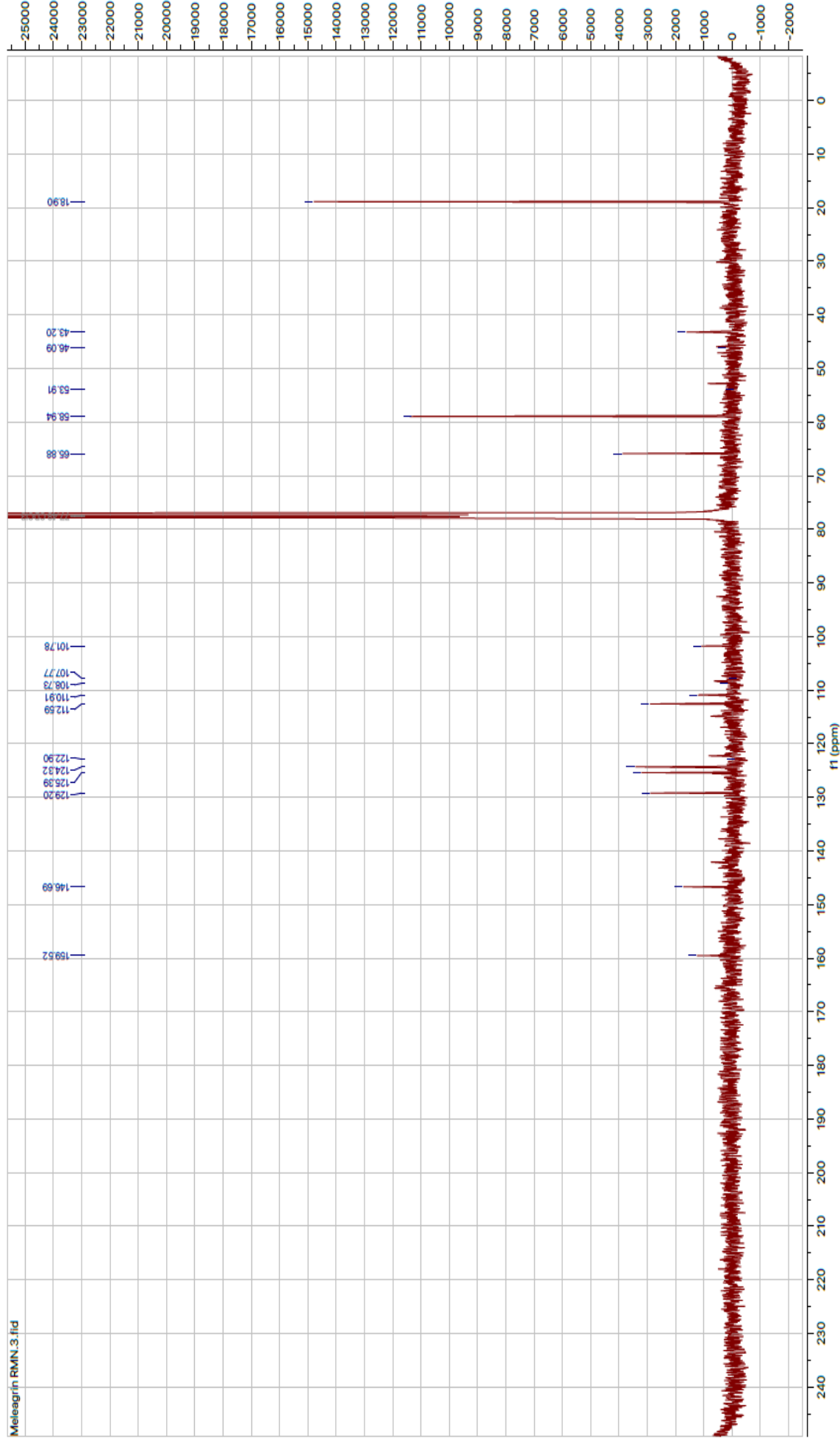
- Annotation via bases de données
- Analyse par spectroscopie RMN

Annexe 8. Spectres RMN de la méléagrine (échantillon CTBG-04-m/z 434)

- Spectre ^1H , (CDCl_3 , 500 MHz)



- Spectre ^{13}C , (CDCl_3 , 125MHz)



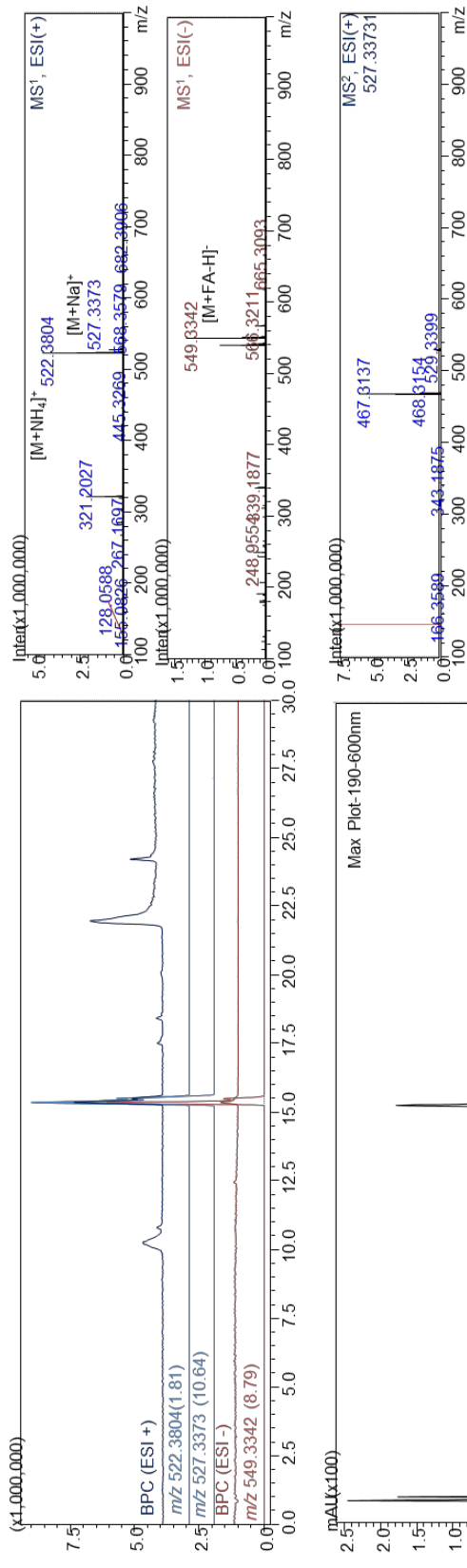
Annexe 9. Les profils LC-MS de l'anicequol avec le tracé BPC (base peak chromatogram) en mode positif (ESI+) et négatif (ESI-) et les tracés EICs (extracted ion chromatogram) des ions détectés.

Code : CTBG-02-*m/z* 527

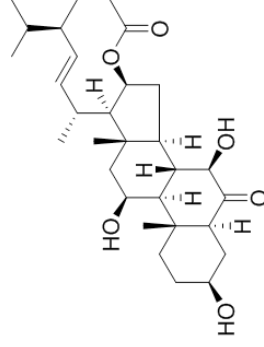
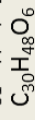
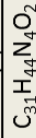
Métabolite produit par : *P. chrysogenum*

Masse : 0,4 mg

- Molécule isolée
- Molécule détectée



Proposition(s) de formule(s) brute(s) :



anicequol

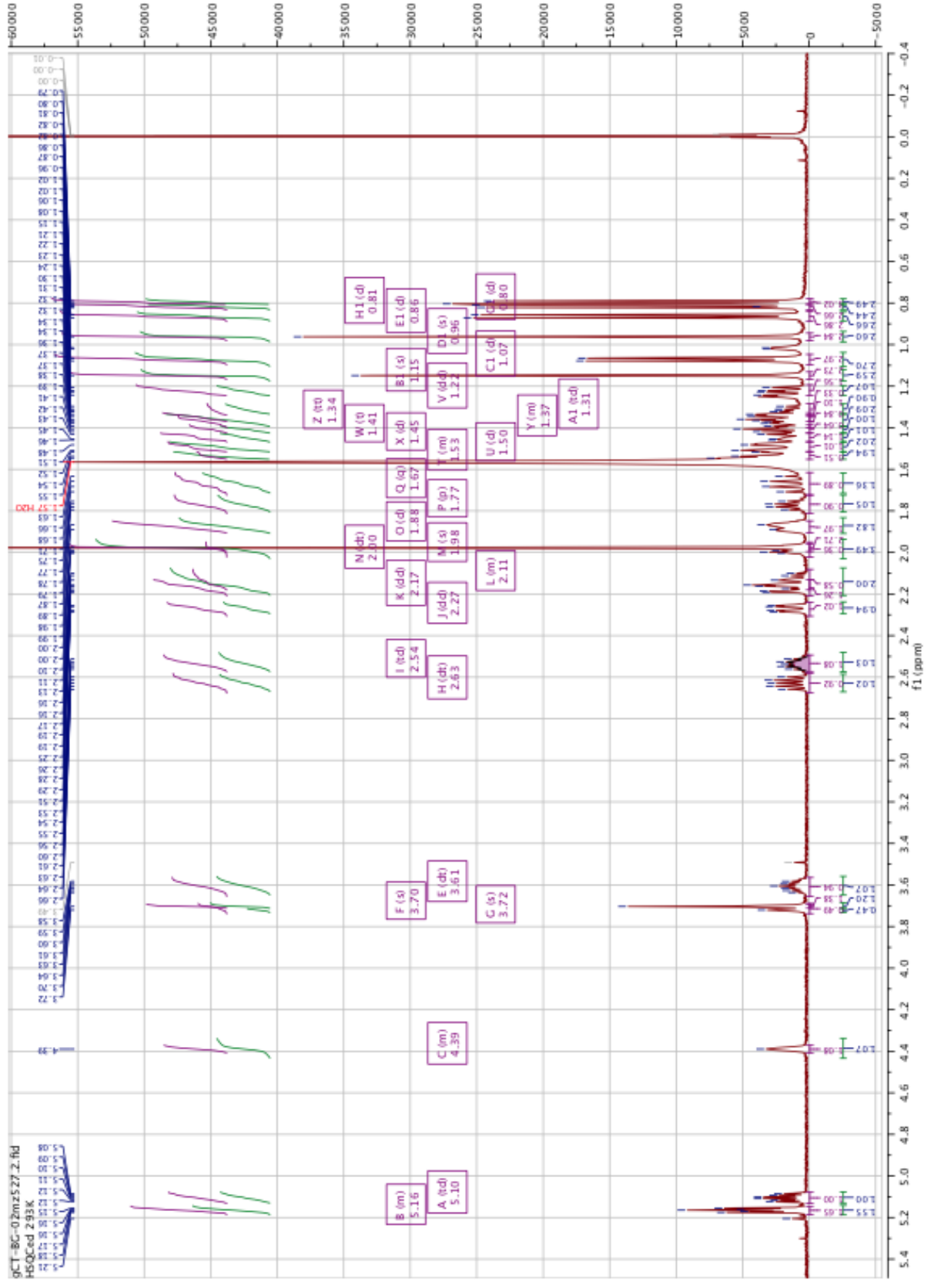
Proposition(s) de structure(s) :

- Acide 4,7,22,24-tetrahydroxy-3,4-seco-11,13(18)-oleanadien-3-oïque
- Acide 4,15,22,24-tetrahydroxy-3,4-seco-11,13(18)-oleanadien-3-oïque
- **Anicequol**

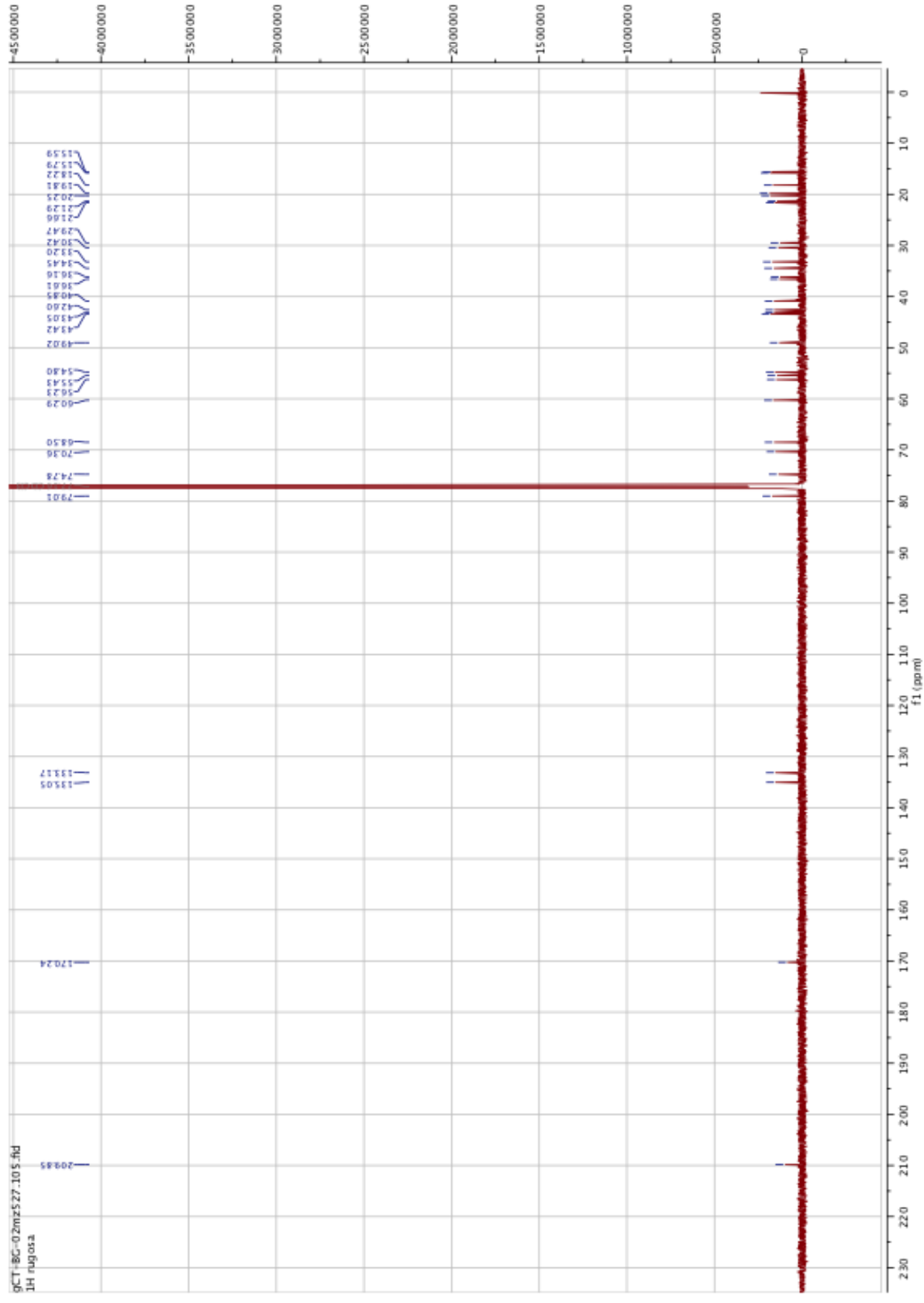
- Annotation via bases de données
- Analyse par spectroscopie RMN

Annexe 10. Spectres RMN de l'anicequol (échantillon CTBG-02-m/z 527)

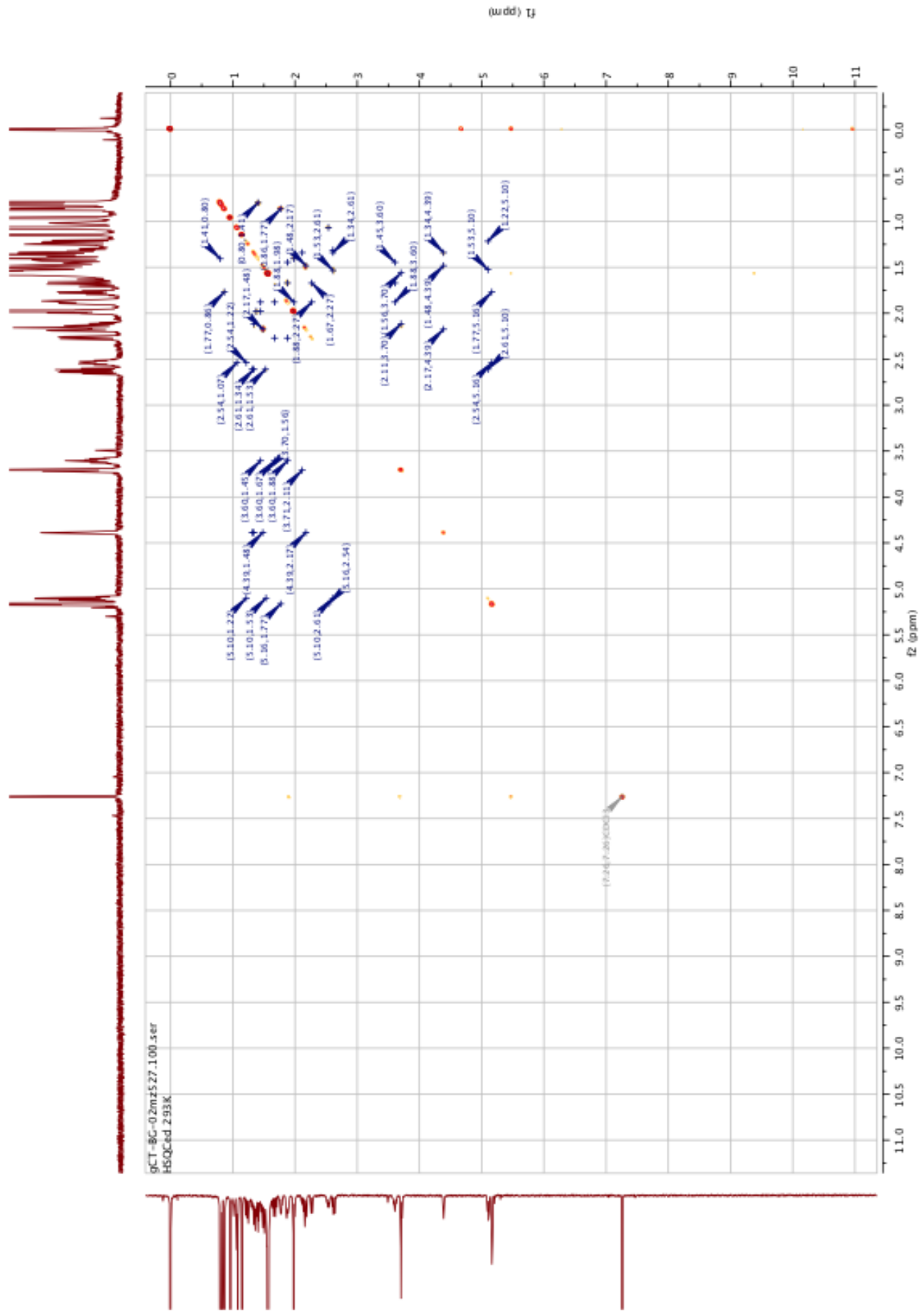
- Spectre ^1H , (CDCl_3 , 500 MHz)



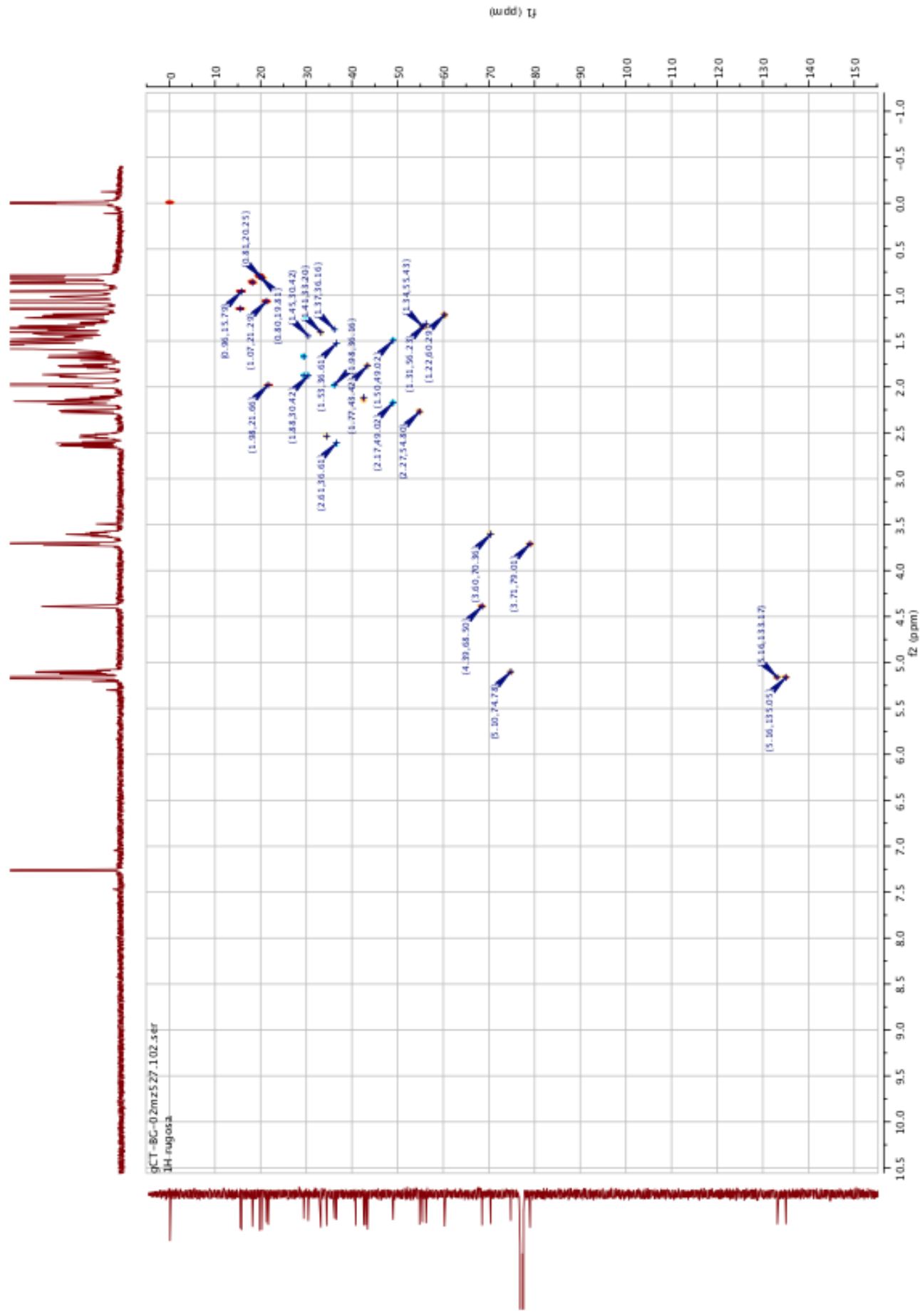
- Spectre ^{13}C , (CDCl_3 , 125 MHz)



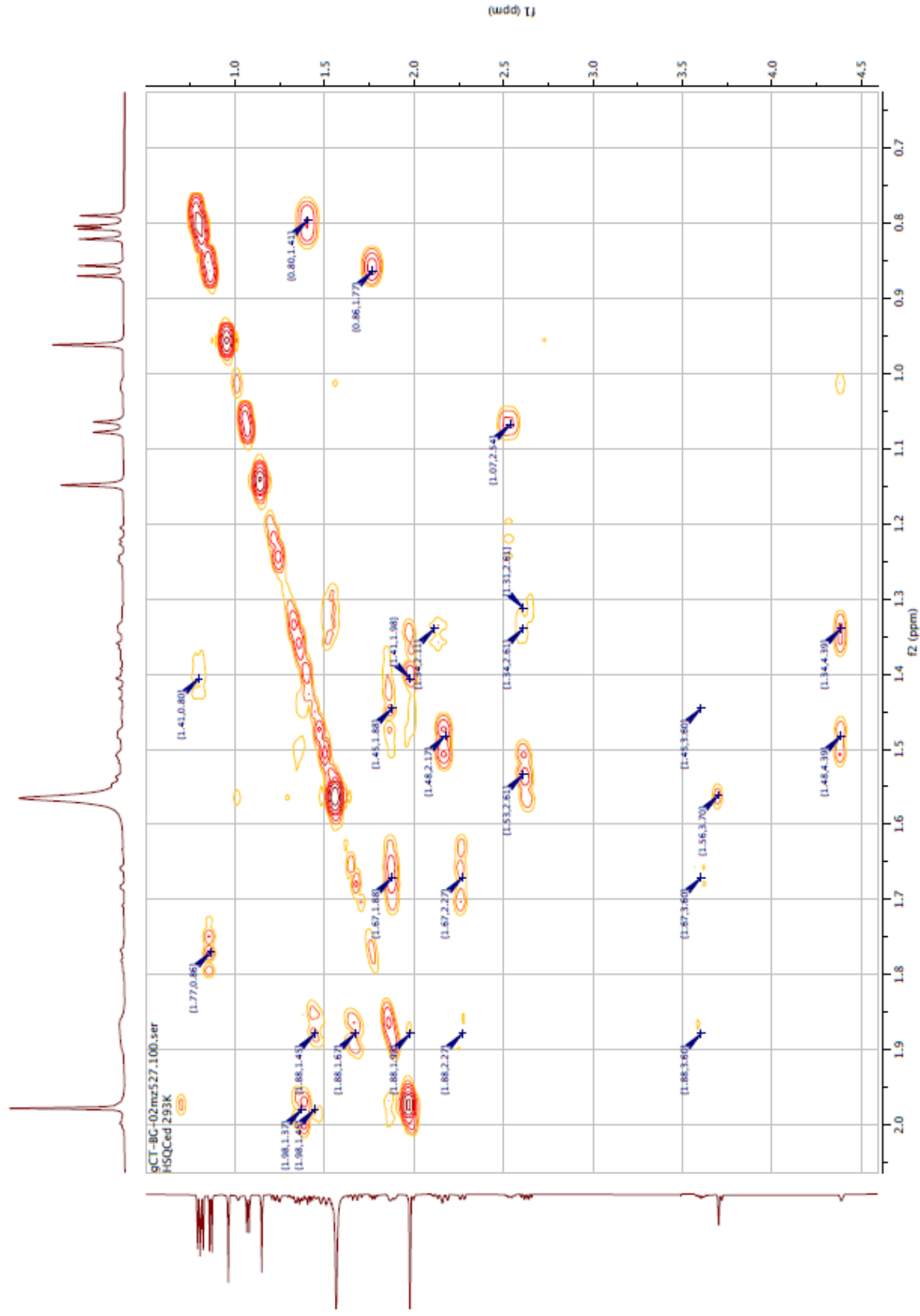
- Spectre RMN 2D HMBC, (CDCl₃, 500 MHz)



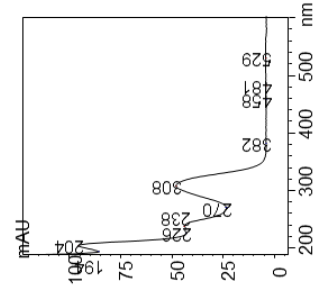
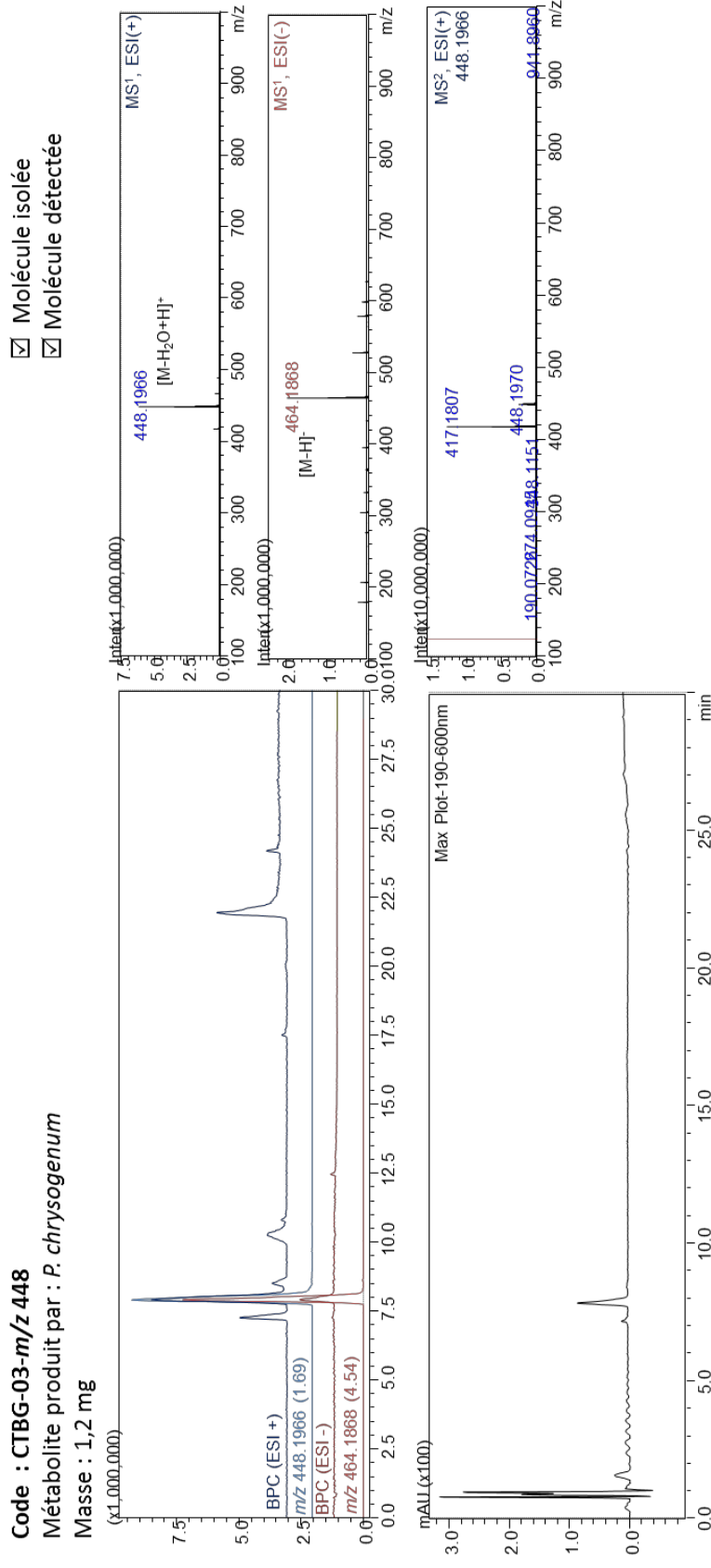
- Spectre RMN 2D HSQC, (CDCl₃, 500 MHz)



- Spectre RMN 2D COSY, (CDCl₃, 500 MHz)



Anexe 11. Les profils LC-MS de l'échantillon CTBG-03-*m/z* 448 avec le tracé BPC (base peak chromatogram) en mode positif (ESI+) et négatif (ESI-) et les tracés EICs (extracted ion chromatogram) des ions détectés.



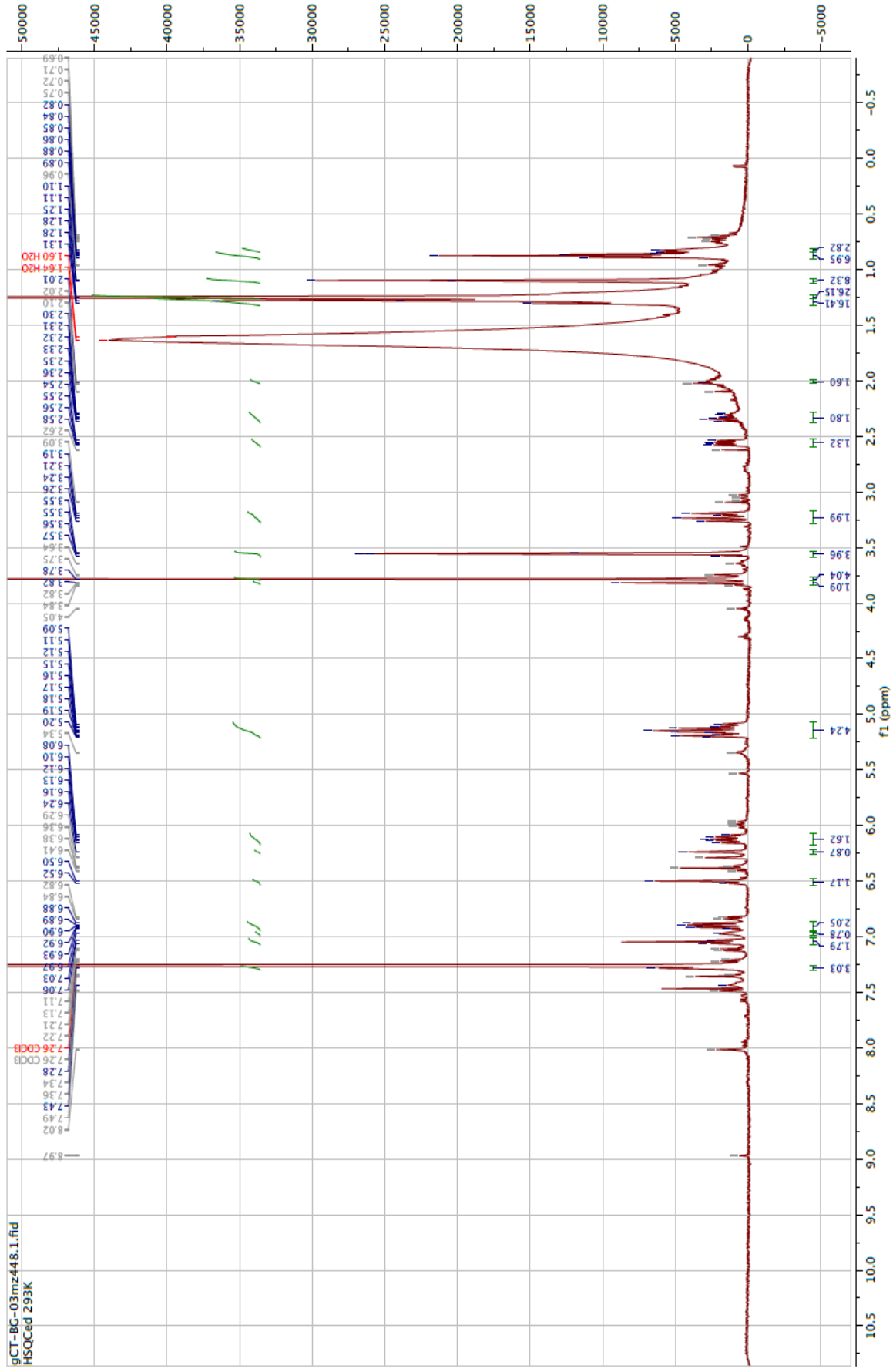
Proposition(s) de formule(s) brute(s) :
C23H31NO9
C24H27N5O5

Proposition(s) de structure(s) :
 → 0 hit dans le DNP

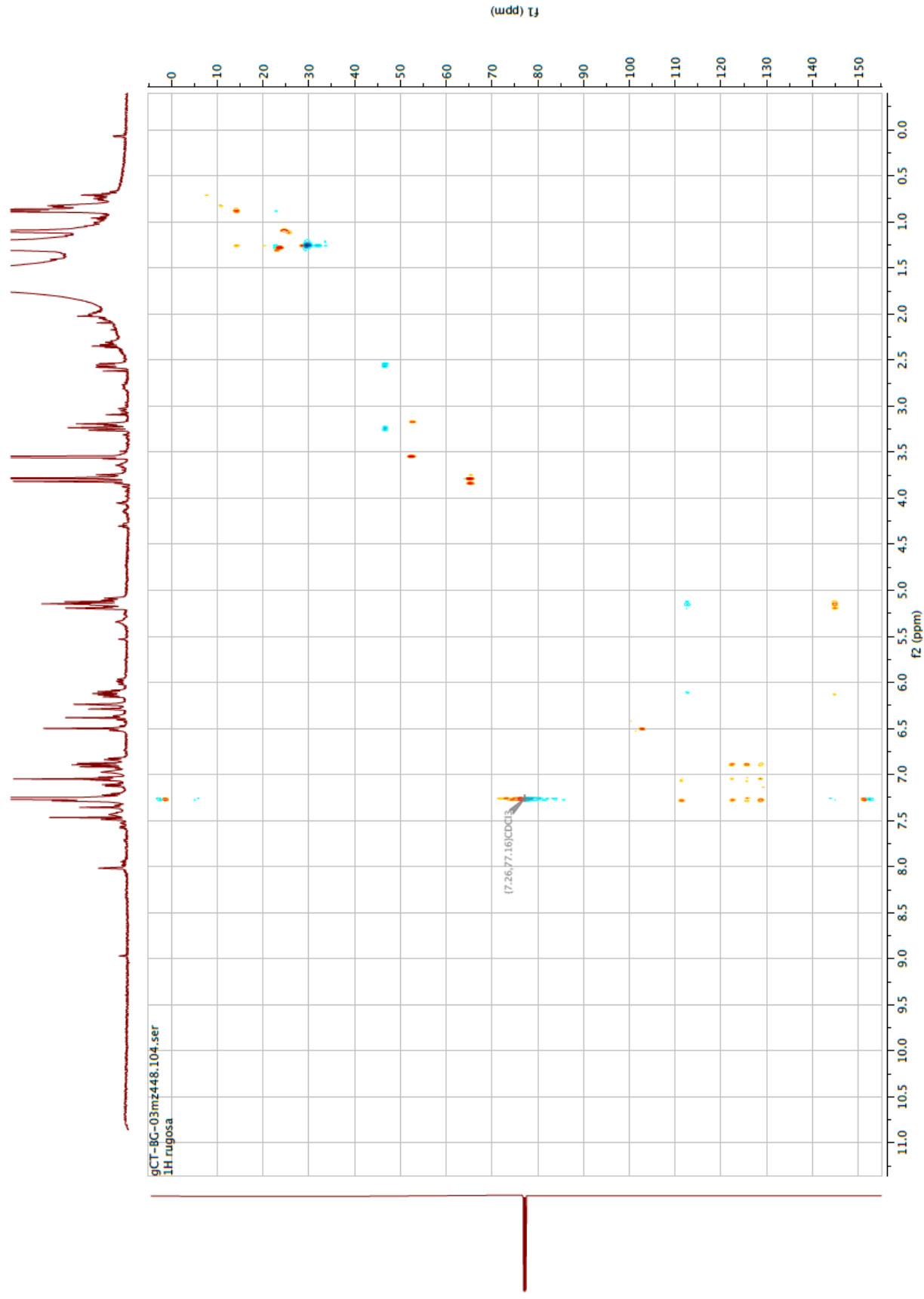
Annotation via bases de données
 Analyse par spectroscopie RMN

Annexe 12. Spectres RMN de l'échantillon CTBG-03-m/z 448

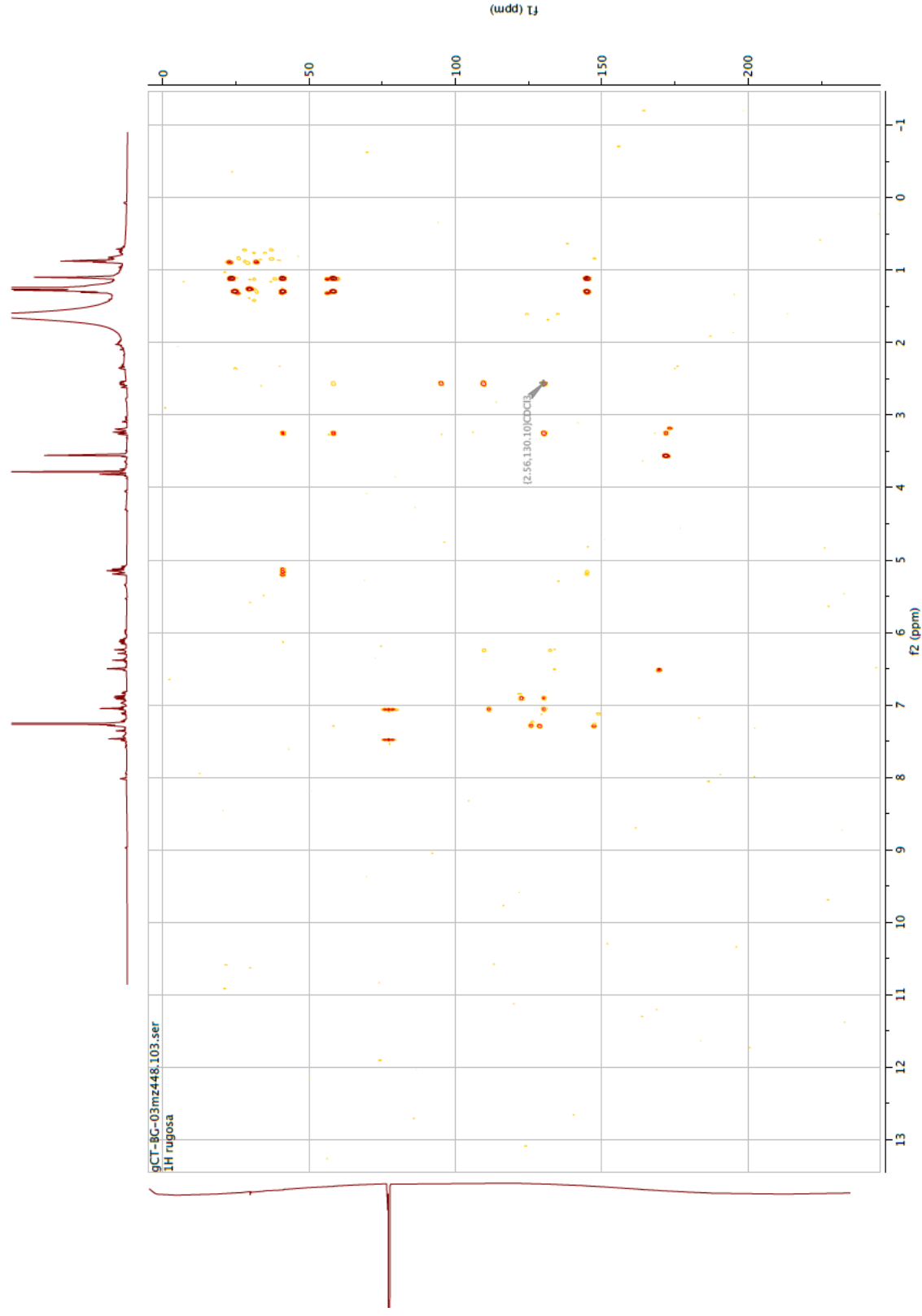
- Spectre ^1H , (CDCl_3 , 500 MHz)



- Spectre RMN 2D HSQC, (CDCl₃, 500 MHz)



- Spectre RMN 2D HMBC, (CDCl₃, 500 MHz)



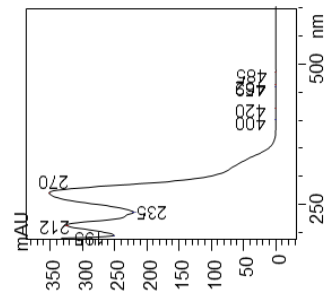
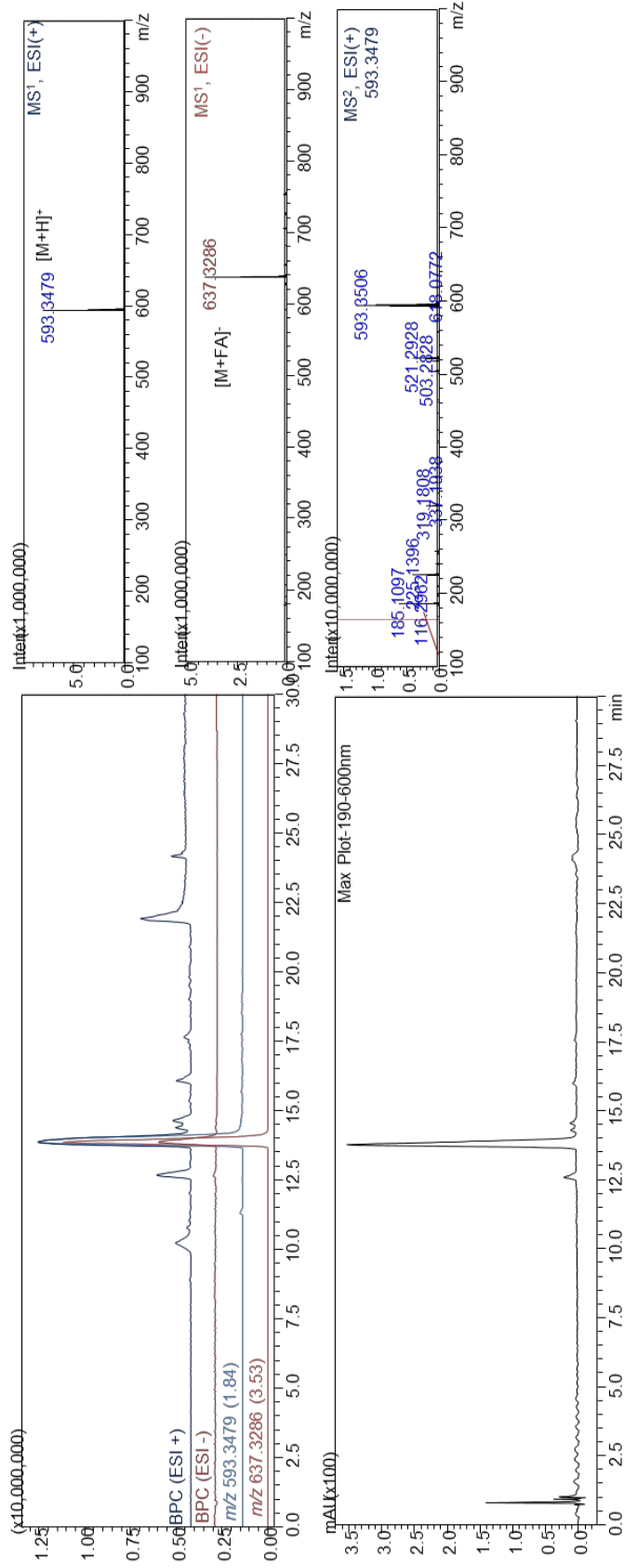
Annexe 13. Les profils LC-MS de l'échantillon CTBG-01-*m/z* 593 avec le tracé BPC (base peak chromatogram) en mode positif (ESI+) et négatif (ESI-) et les tracés EICs (extracted ion chromatogram) des ions détectés

Code : CTBG-01-*m/z* 593

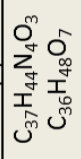
Métabolite produit par : *P. chrysogenum*

Masses : 0,7 mg

- Molécule isolée
- Molécule détectée



Proposition(s) de formule(s) brute(s) :



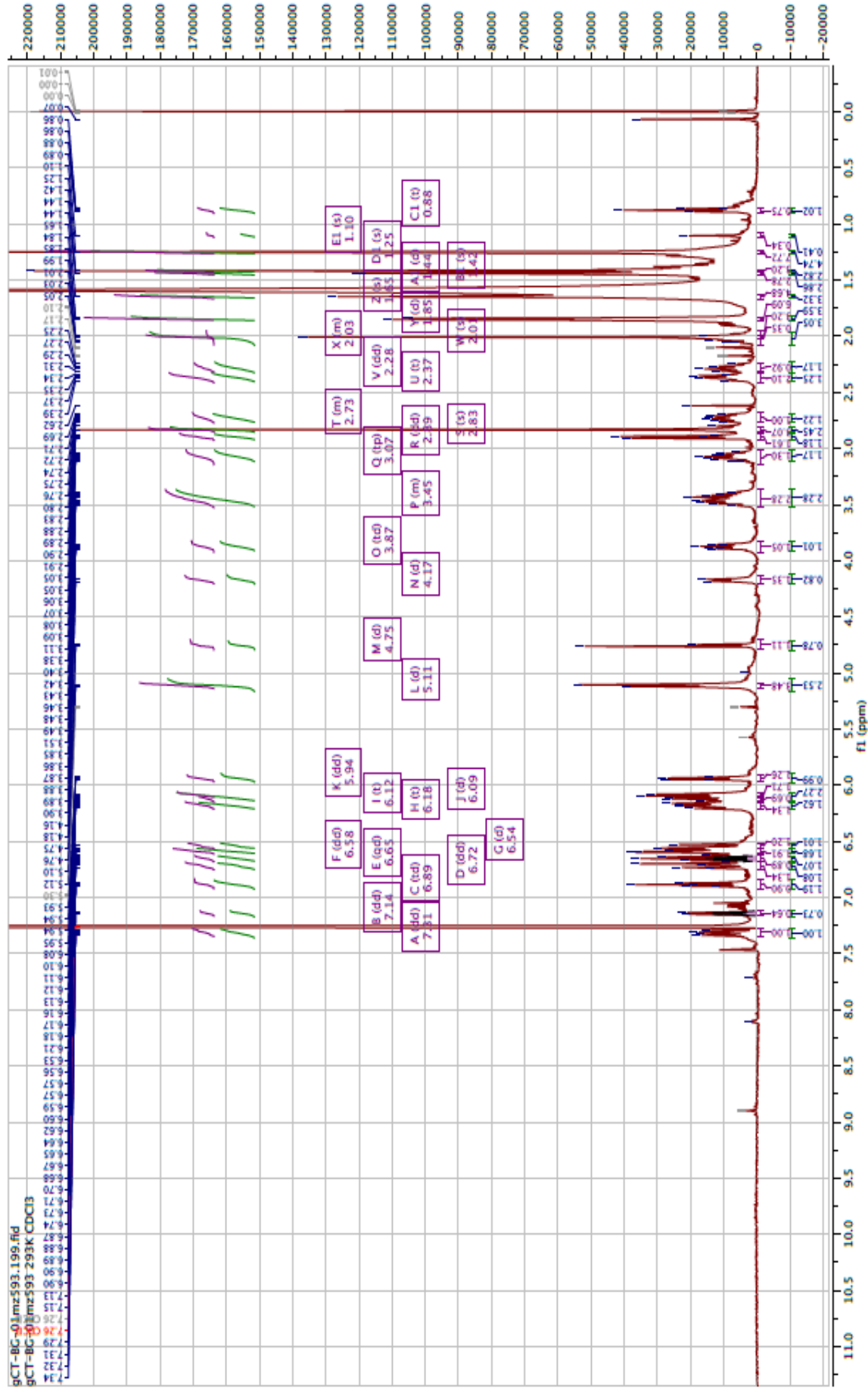
Proposition(s) de structure(s) :

→ 0 Hit dans le DNP pour les deux premières formules brutes

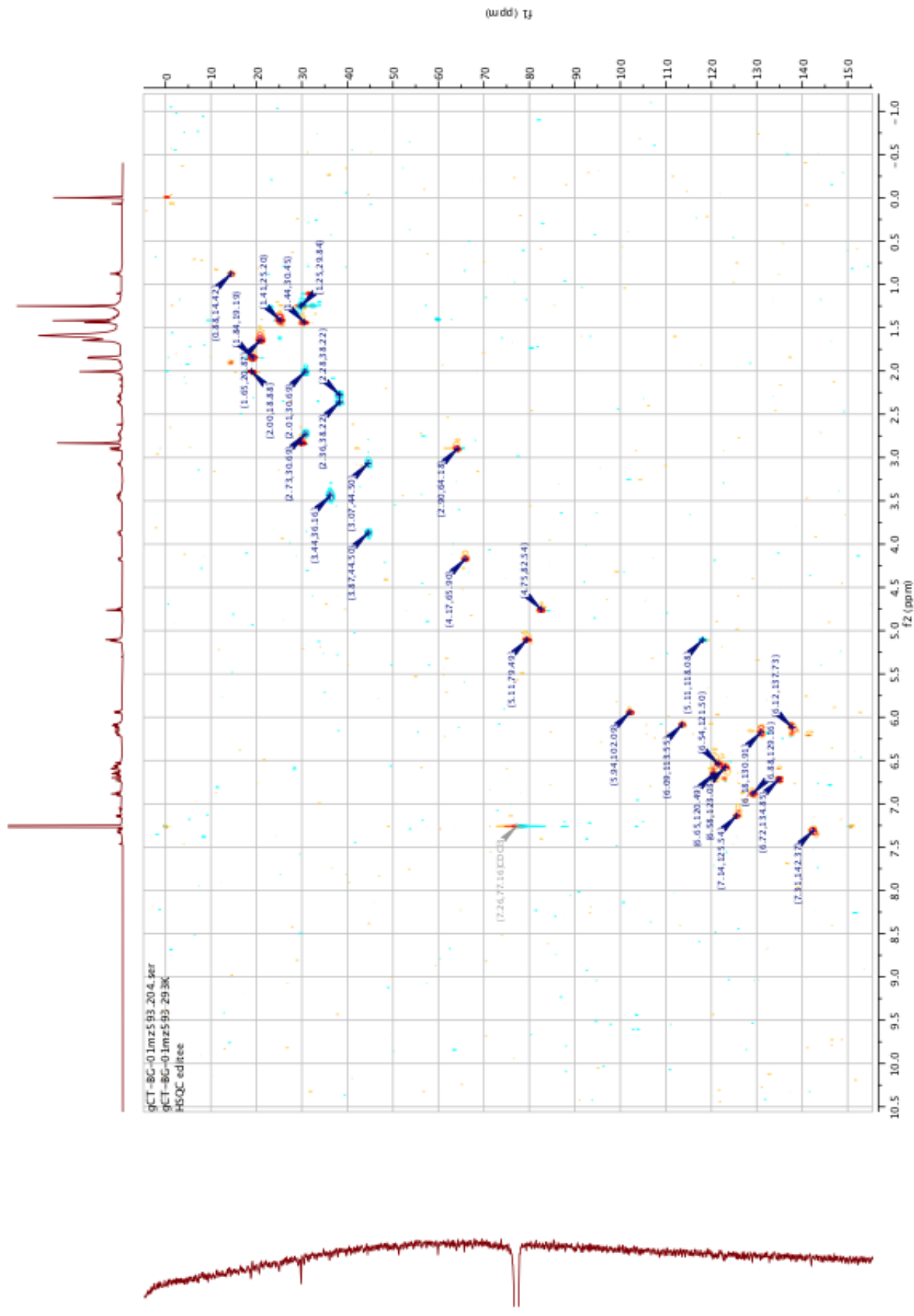
- Annotation via bases de données
- Analyse par spectroscopie RMN

Annexe 14. Spectres RMN de l'échantillon CTBG-01-m/z 593

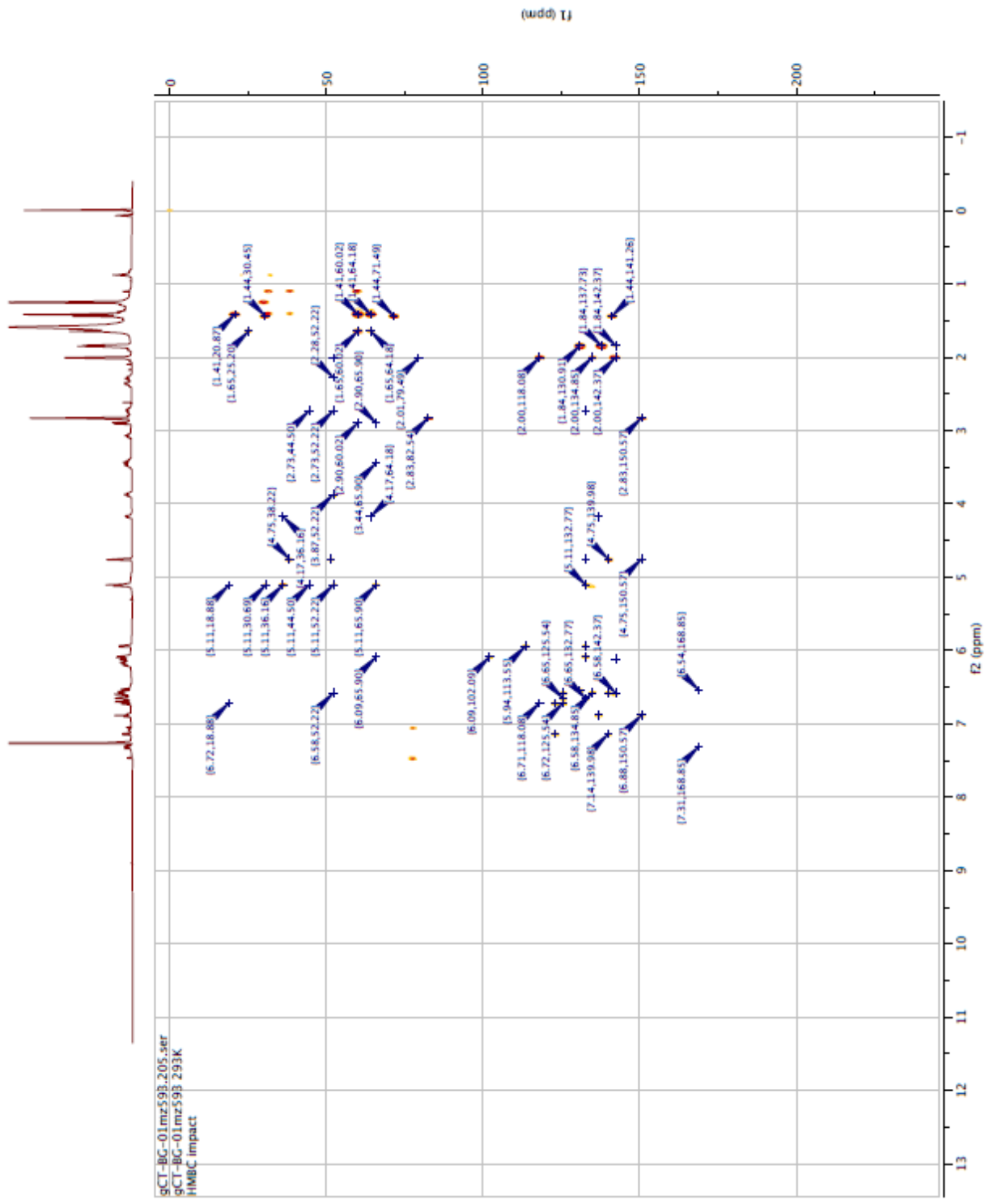
- Spectre ^1H , (CDCl_3 , 500 MHz)



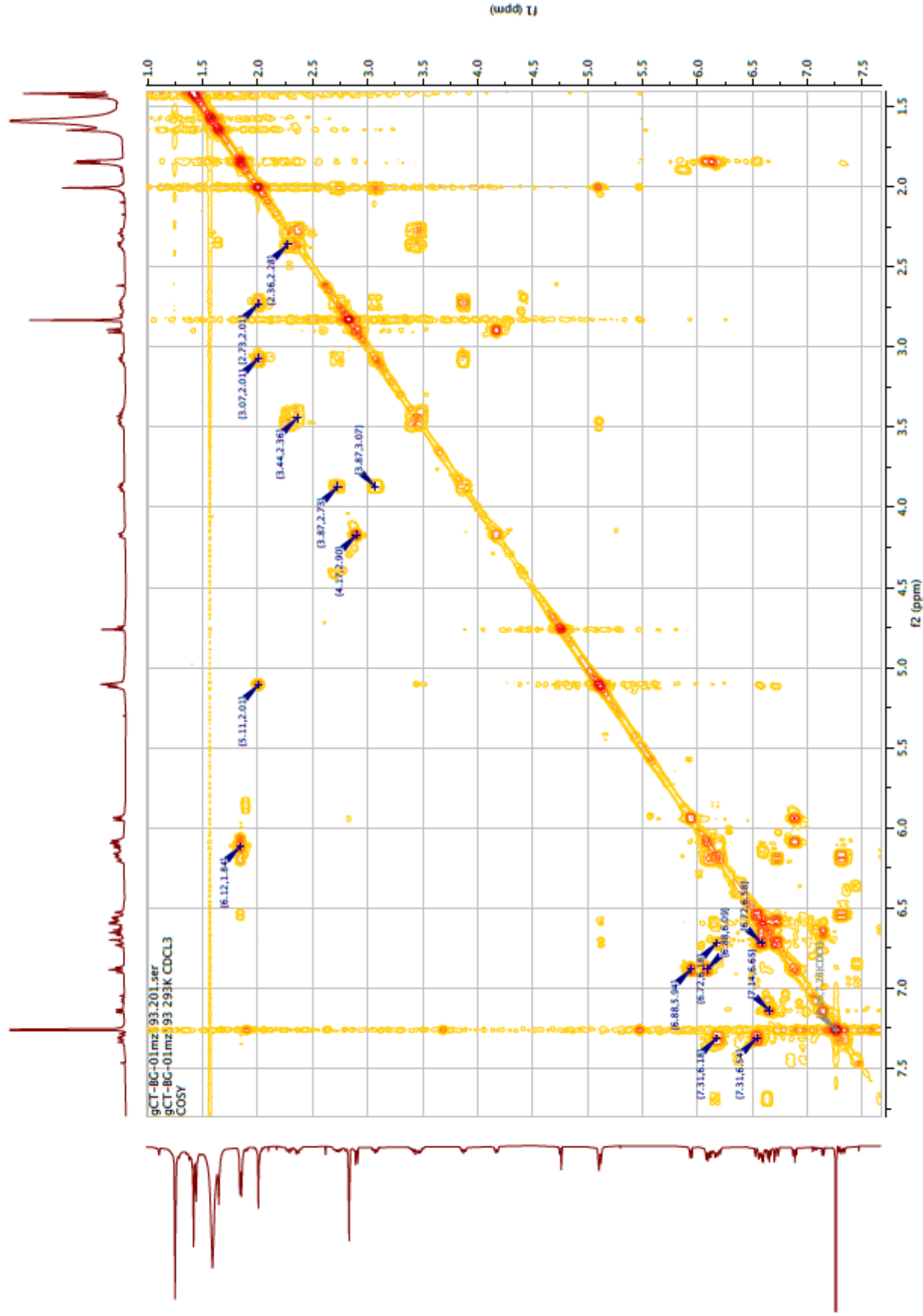
- Spectre RMN 2D HSQC, (CDCl₃, 500 MHz)



- Spectre RMN 2D HMBC, (CDCl₃, 500 MHz)



- Spectre RMN 2D COSY, (CDCl₃ 500 MHz)



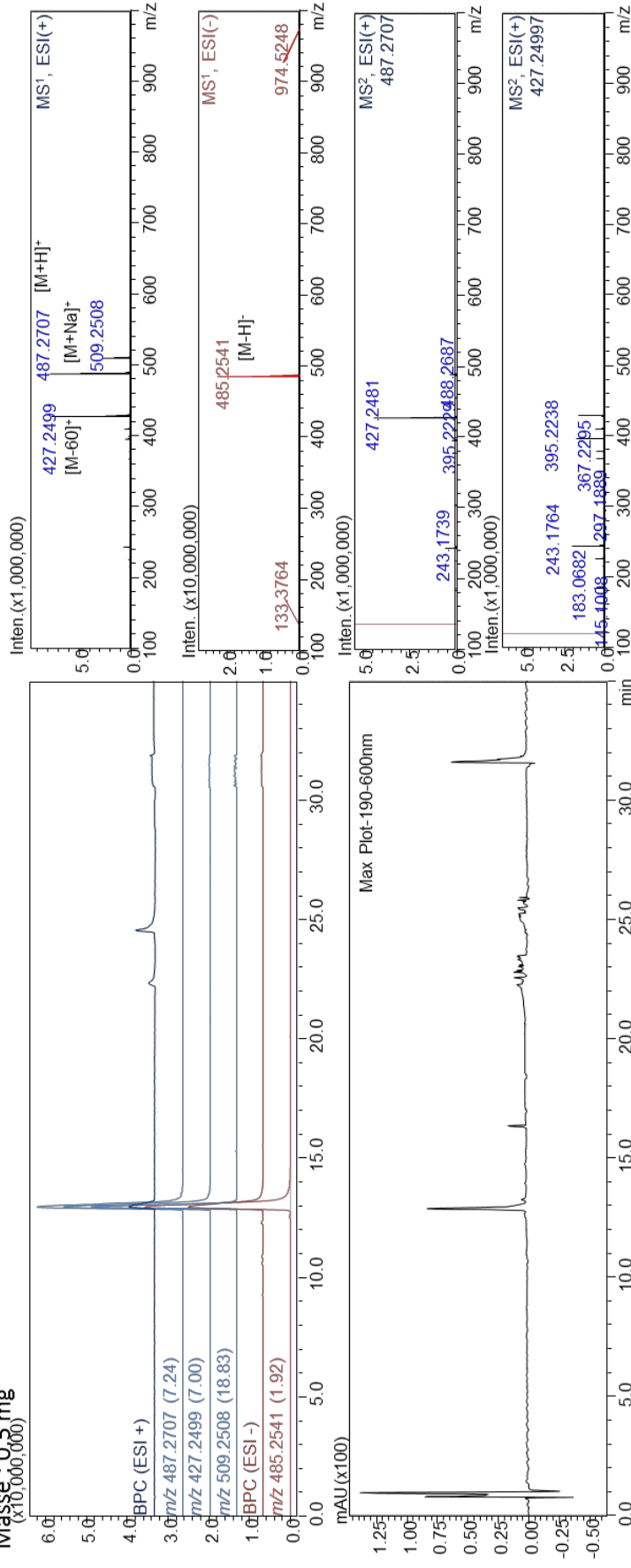
Annexe 15. Les profils LC-MS de l'andrastine A avec le tracé BPC (base peak chromatogram) en mode positif (ESI+) et négatif (ESI-) et les tracés EICS (extracted ion chromatogram) des ions détectés

Code : MMS5-531-18-10

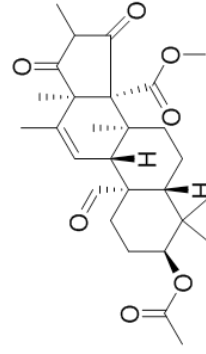
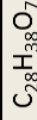
Métabolite produit par : *P. chrysogenum*

Masses : 0.5 mg
(x10,000,000)

- Molécule isolée
- Molécule détectée



Proposition(s) de formule(s) brute(s) :



Andrastine A

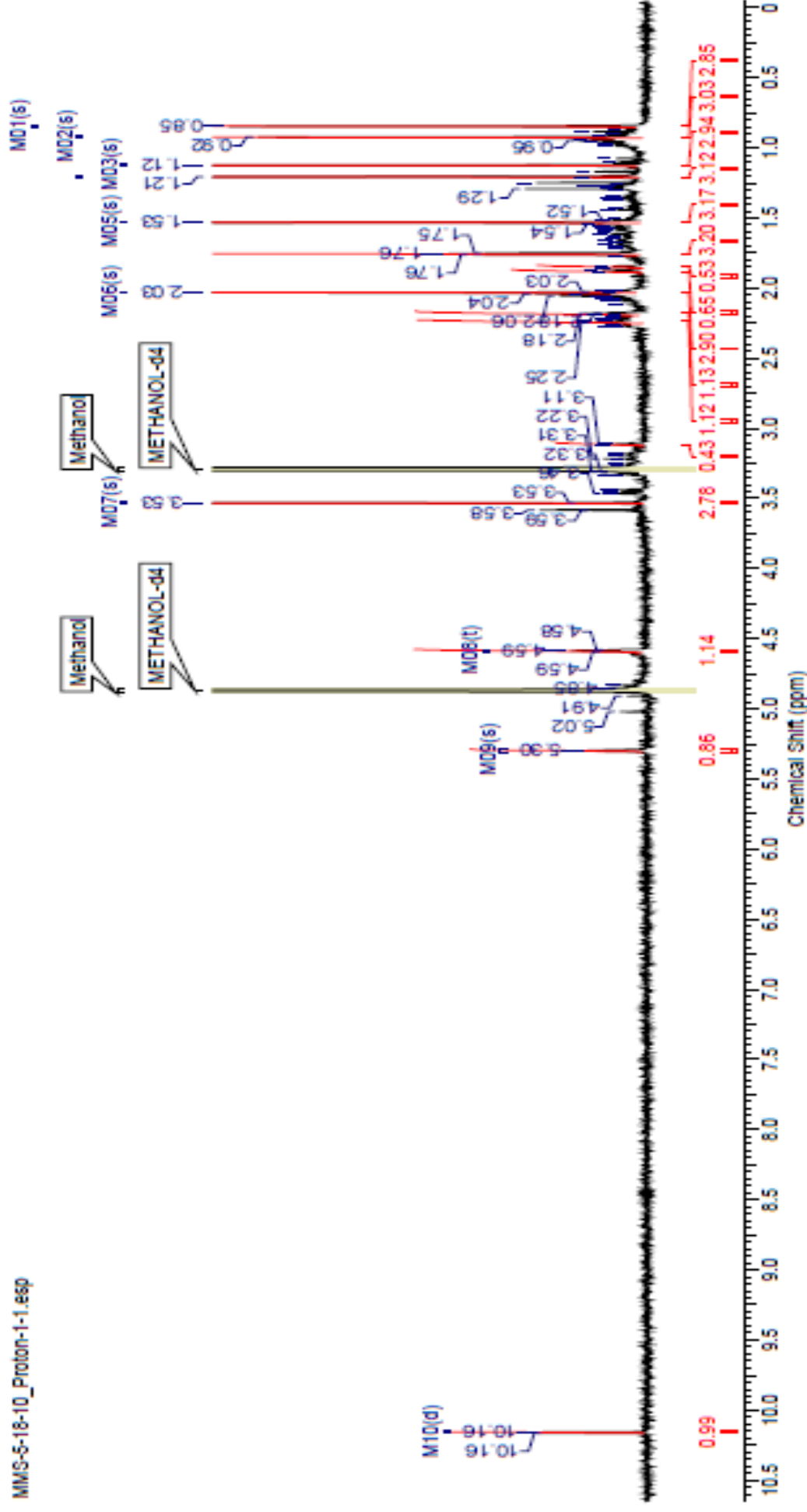
Proposition(s) de structure(s) :

Andrastine A

- Annotation via bases de données
- Analyse par spectroscopie RMN

Annexe 16. Spectre RMN de l'andrastine A (échantillon MMS5-531-18-10)

- Spectre ^1H , (CD_3OD , 400 MHz)



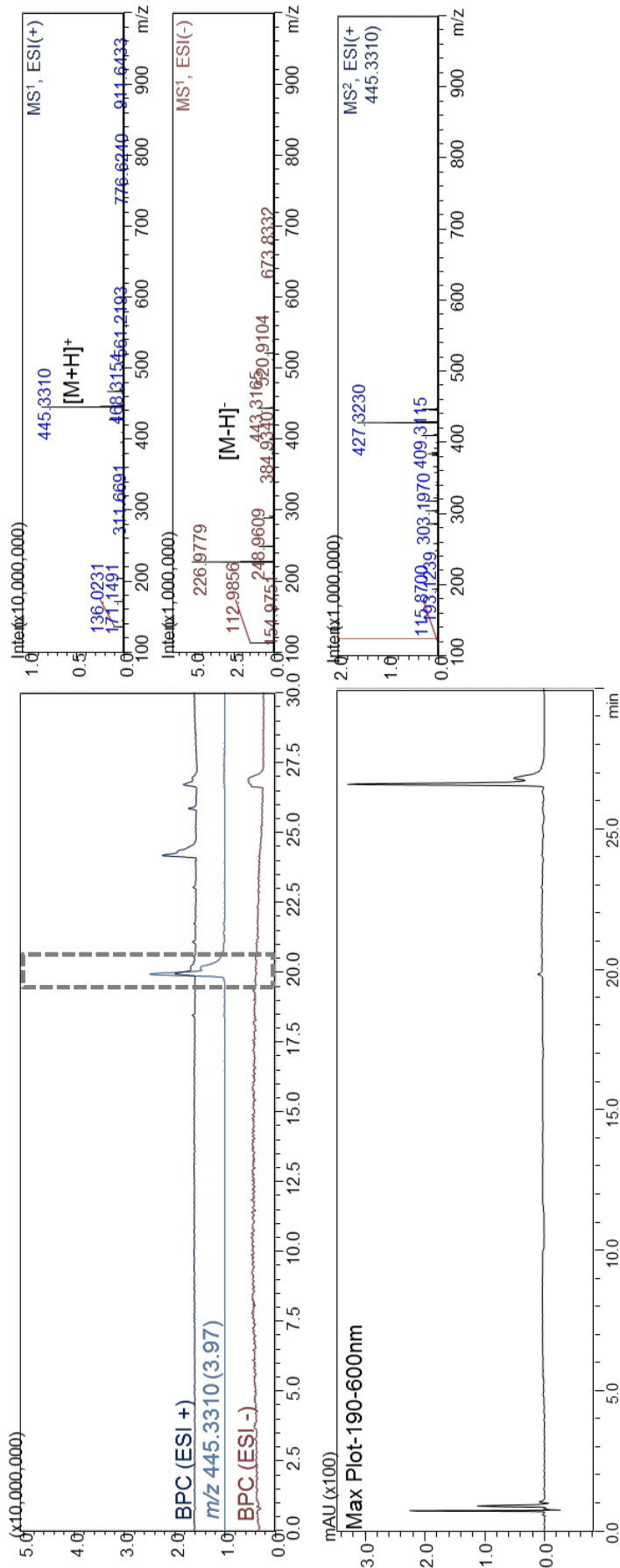
Annexe 17. Les profils LC-MS de l'échantillon MMS5-526-5-23 avec le tracé BPC (base peak chromatogram) en mode positif (ESI+) et négatif (ESI-) et les tracés EICs (extracted ion chromatogram) des ions détectés.

Code : MMS 5-526-5-23

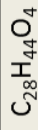
Métabolite produit par : *P. chrysogenum*

Masse : 1,3 mg

- Molécule isolée
 Molécule détectée



Proposition(s) de formule(s) brute(s) :



Proposition(s) de structure(s) :

- Citreospirosteroid
- Ergosta-7,22-dien-6-one-3,5,9-trihydroxy
- Mer-NF 8054A

- Annotation via bases de données
 Analyse par spectroscopie RMN

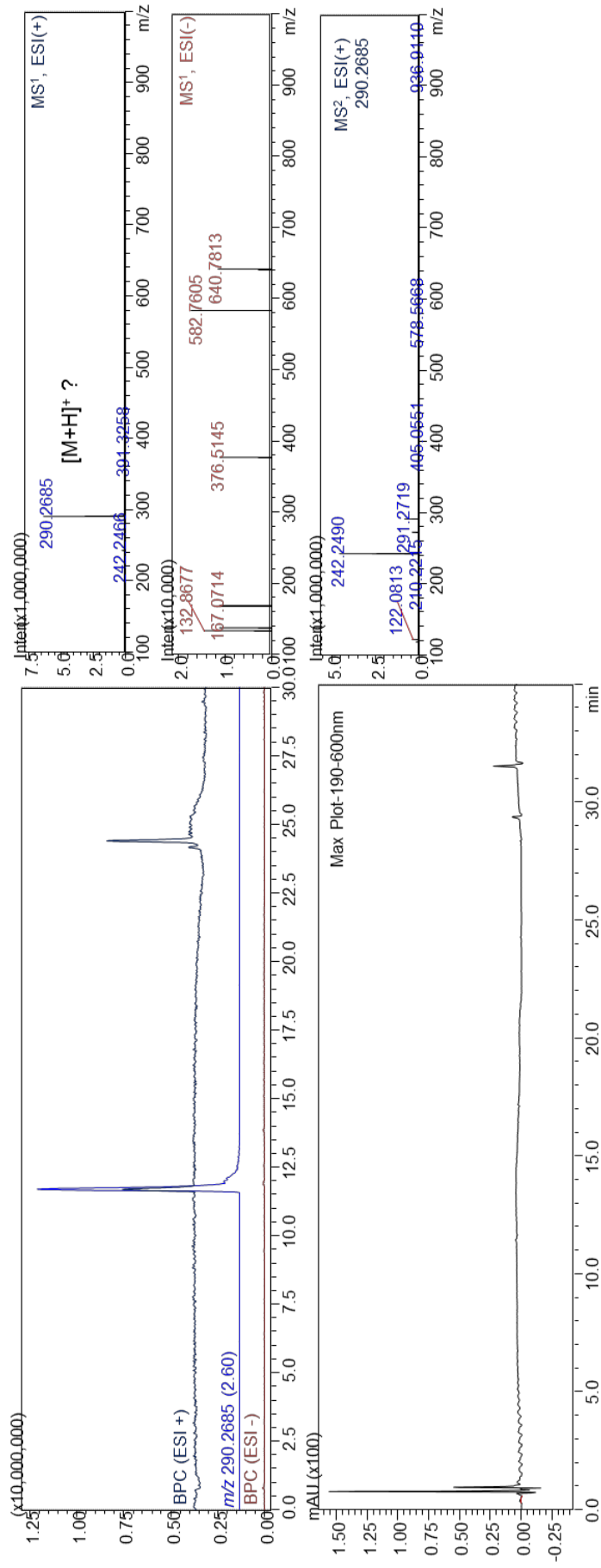
Annexe 18. Les profils LC-MS de l'échantillon MMS5-526-5-30 avec le tracé BPC (base peak chromatogram) en mode positif (ESI+) et négatif (ESI-) et les tracés EICs (extracted ion chromatogram) des ions détectés.

Code : MMS 5-526-5-30

Métabolite produit par : *P. chrysogenum*

Masse : 0,2 mg

- Molécule isolée
 Molécule détectée



Proposition(s) de formule(s) brute(s) :
 $C_{16}H_{35}NO_3$ si $[M+H]^+$

- Proposition(s) de structure(s) :
- 2-Amino-1,3,4-hexadecanetriol
 - 2-Amino-14-méthyle-1,3,4-pentadecanetriol
- Annotation via bases de données
 Analyse par spectroscopie RMN

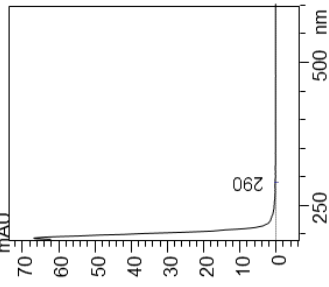
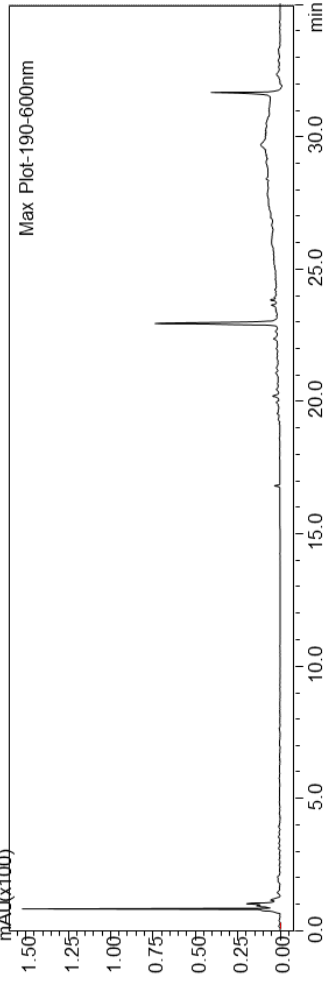
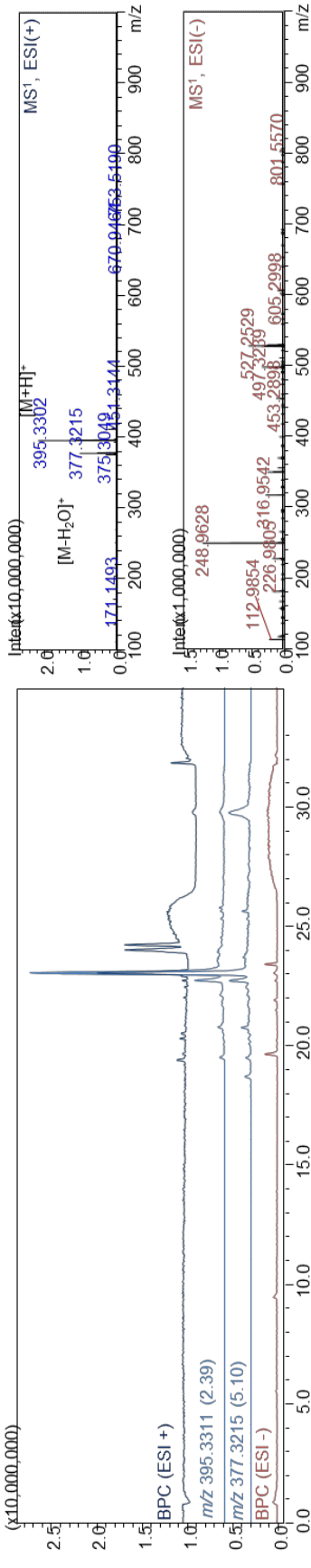
Annexe 19. Les profils LC-MS de l'échantillon MMS5-526-6-11 avec le tracé BPC (base peak chromatogram) en mode positif (ESI+) et négatif (ESI-) et les tracés EICs (extracted ion chromatogram) des ions détectés.

Code : MMS 5-526-6-11

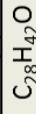
Métabolite produit par : *P. chrysogenum*

Masse : 1,3 mg

- Molécule isolée
 Molécule détectée



Proposition(s) de formule(s) brute(s) :

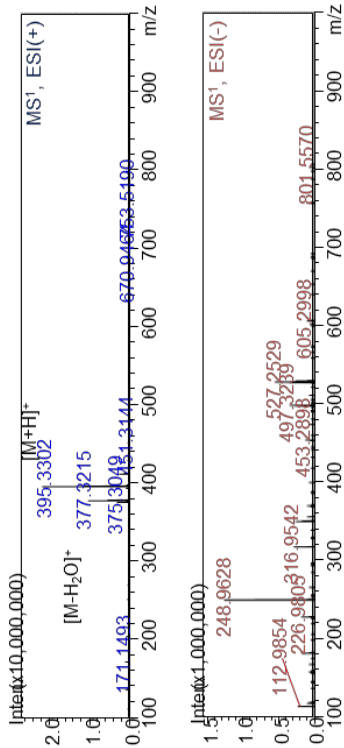


Proposition(s) de structure(s) :

- 1-(10 \rightarrow 6)-abeo-Ergosta-5,7,9,22-tetraen-3 α -ol
- 14-Déhydrocholestérol
- Ergosta-5,7,22,24(28)-tetraen-3-ol
- 6-Methyl-19-norergosta-5,7,9,22-tetraen-3-ol

- Annotation via bases de données
 Analyse par spectroscopie RMN

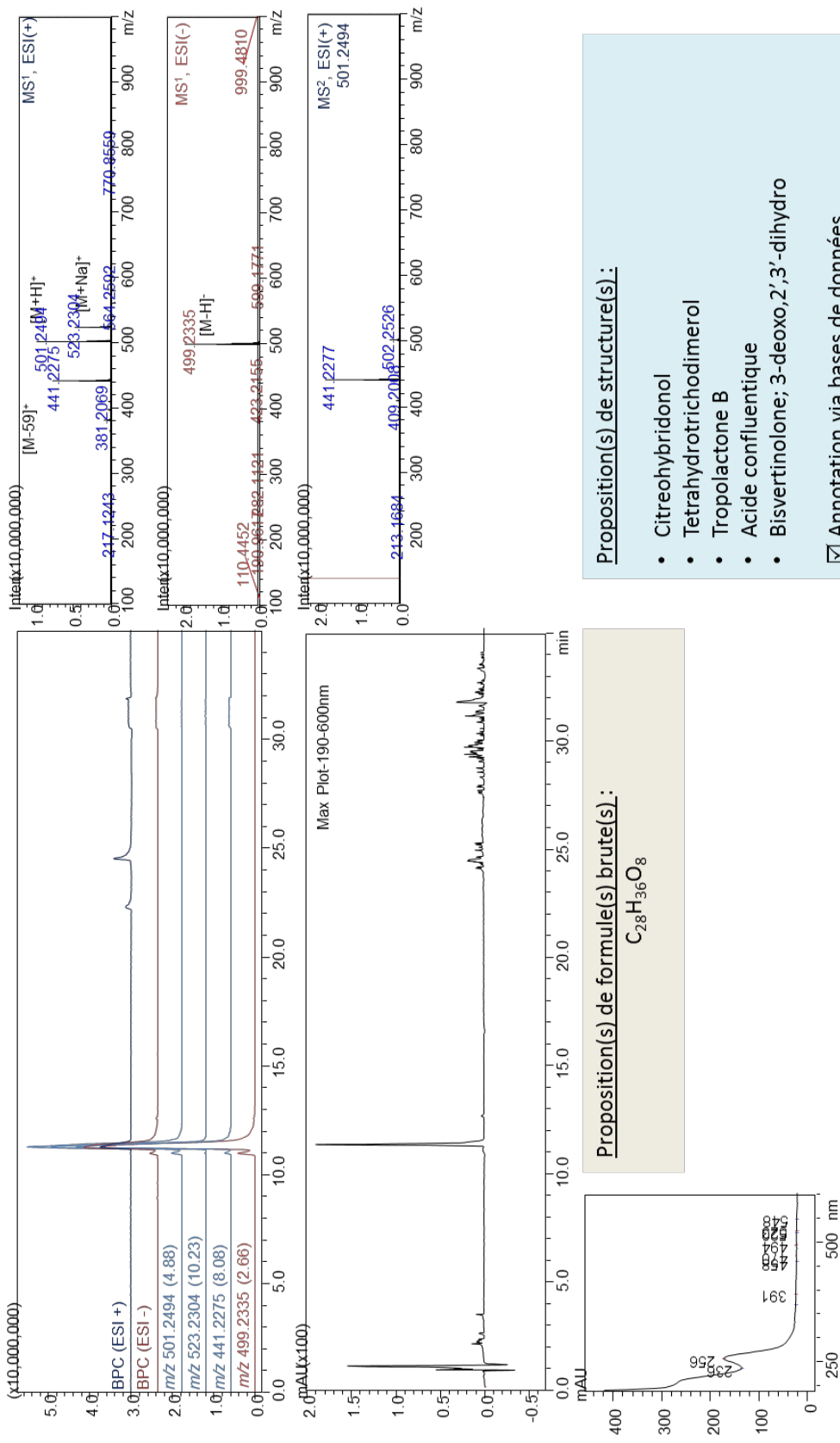
Pas d'acquisition de données MS²



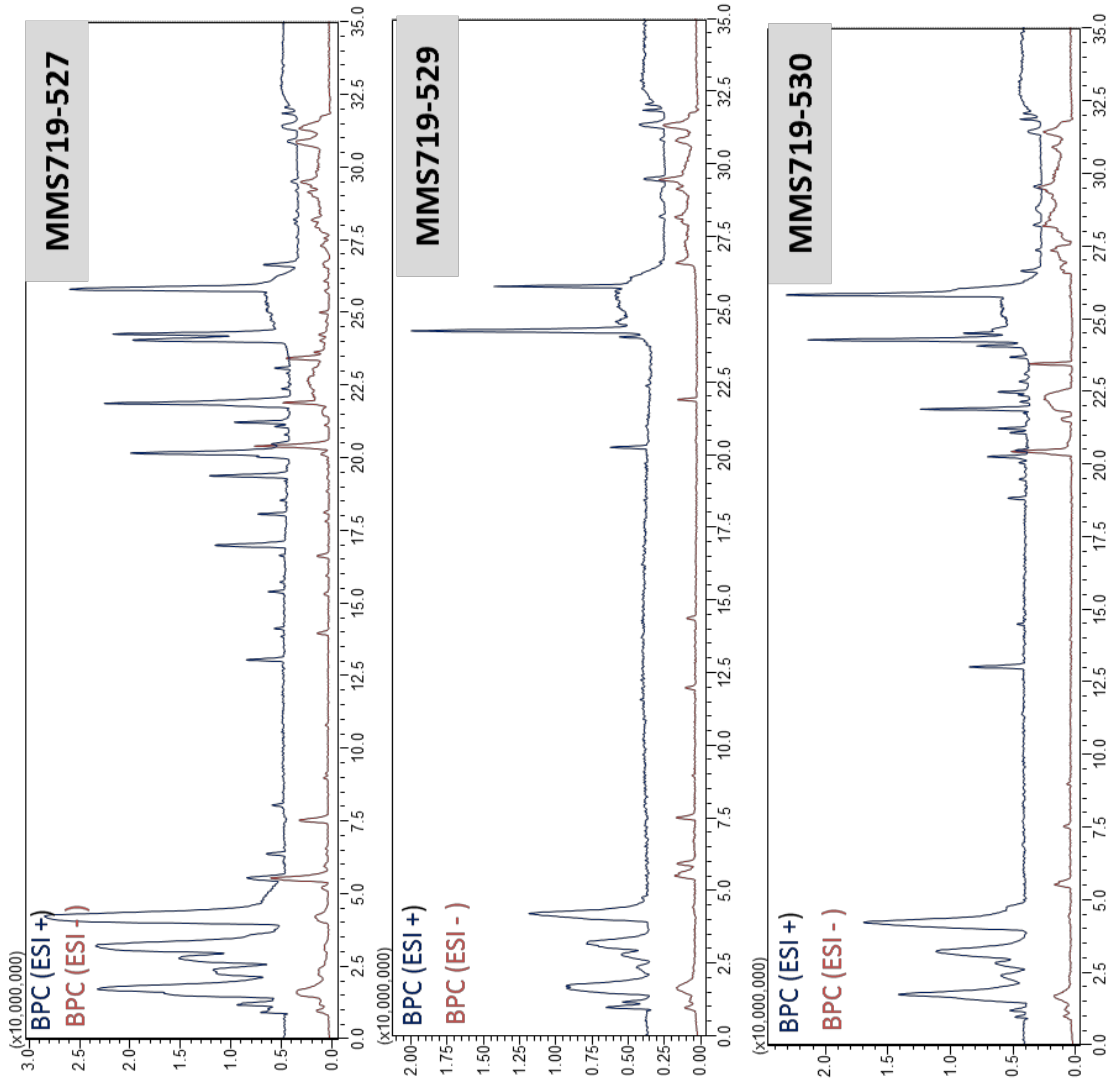
Annexe 20. Les profils LC-MS de l'échantillon MMS5-526-5-20-8 avec le tracé BPC (base peak chromatogram) en mode positif (ESI+) et négatif (ESI-) et les tracés EICs (extracted ion chromatogram) des ions détectés.

Code : MMS5-531-20-8
 Métabolite produit par : *P. chrysogenum*
 Masse : 0,6 mg

- Molécule isolée
 Molécule détectée

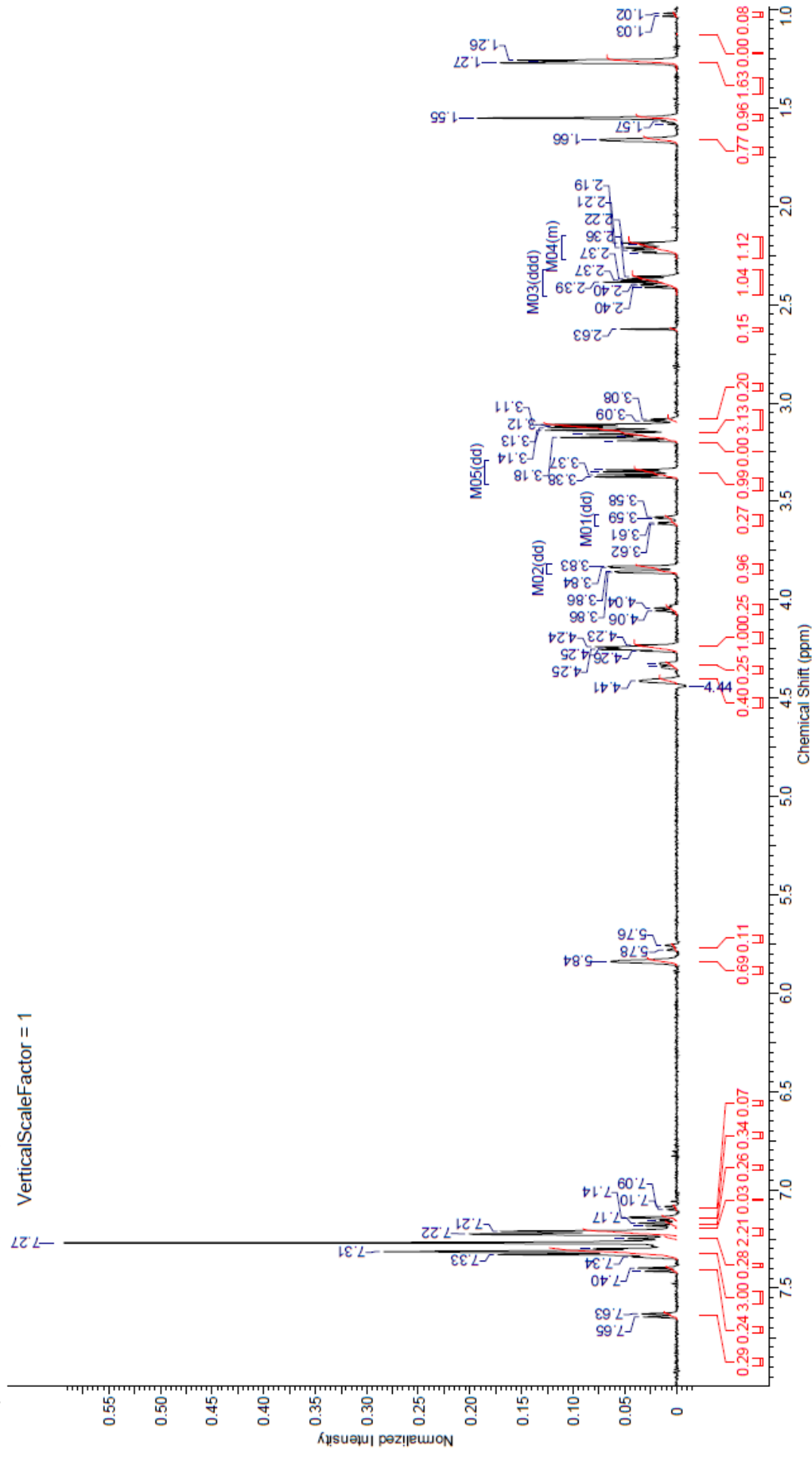


Annexe 21. Superposition des Profils LC-UV-HRMS pour les extraits MS719-527, MMS719-529 et MMS719-530 issus de la souche *Didymella americana*

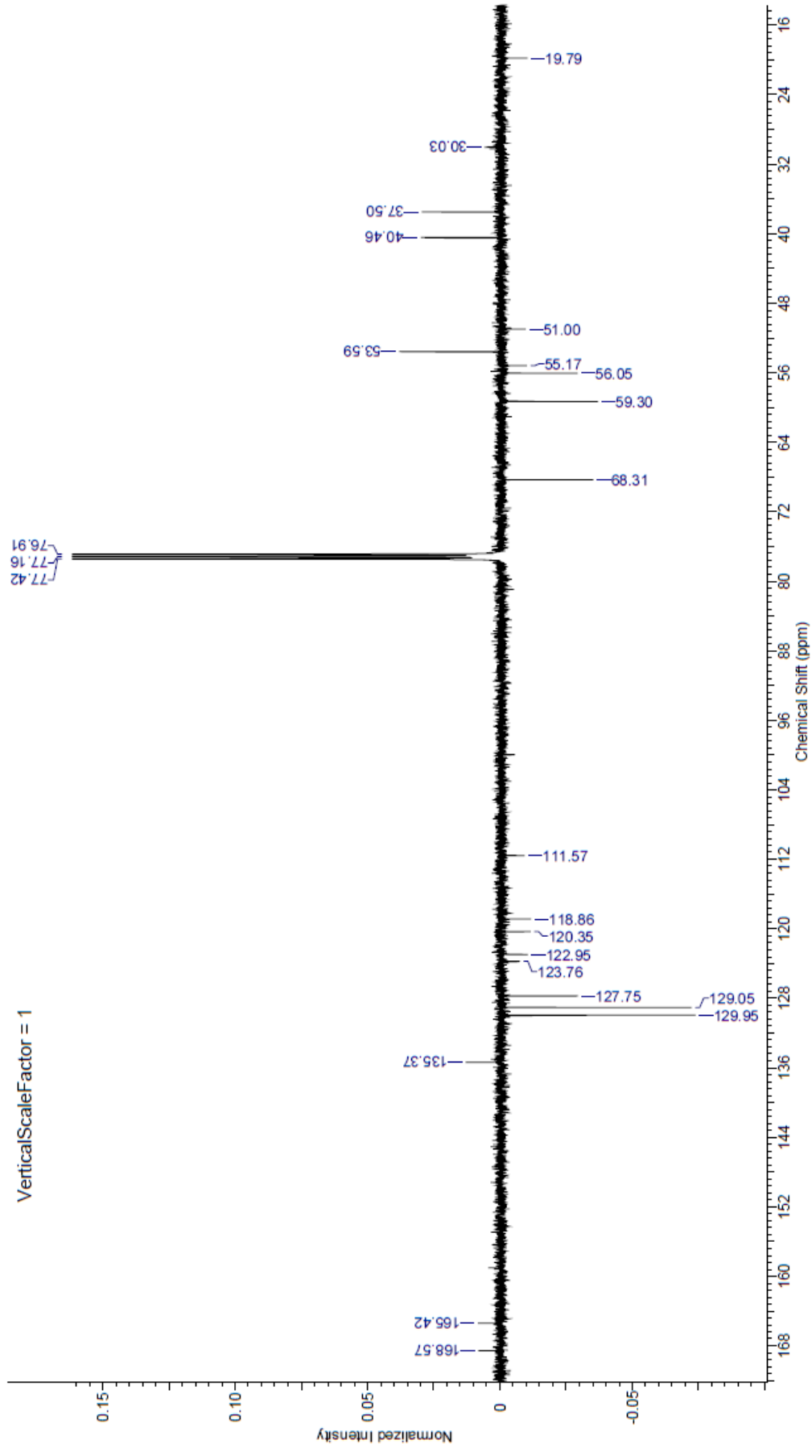


Annexe 22. Spectres RMN du Cyclo(4-hydroxypropylphenylalanyl)

- Spectre ^1H , CDCl_3 (500 MHz)



• Spectre ^{13}C , CDCl_3 (500 MHz)



**Annexe 23. Informations supplémentaires de la publication parue dans le journal
*Analytica Chimica Acta***

**Targeting bioactive compounds in natural extracts - Development of a comprehensive
workflow combining chemical and biological data**

Lucie Ory^a, El-Hassane Nazih^a, Sahar Daoud^a, Julia Mocquard^a, Mélanie Bourjot^b, Laure Margueritte^c, Marc-André Delsuc^d, Jean-Marie Bard^a, Yves François Pouchus^a, Samuel Bertrand^{a,e} and Catherine Roullier^{a,e,*}

^a Université de Nantes, Mer Molécules Santé, MMS EA 2160, F-44000 Nantes, France

^b UMR 7178 CNRS, Institut Pluridisciplinaire Hubert Curien, Faculté de Pharmacie, Université de Strasbourg, 67401 Illkirch, France

^c UMR 7200 CNRS, Laboratoire d'Innovation Thérapeutique, Faculté de Pharmacie, Université de Strasbourg, 67401 Illkirch, France

^d INSERM U596, UMR 7104 CNRS, Institut de Génétique et de Biologie Moléculaire et Cellulaire, Université de Strasbourg, 67401 Illkirch, France

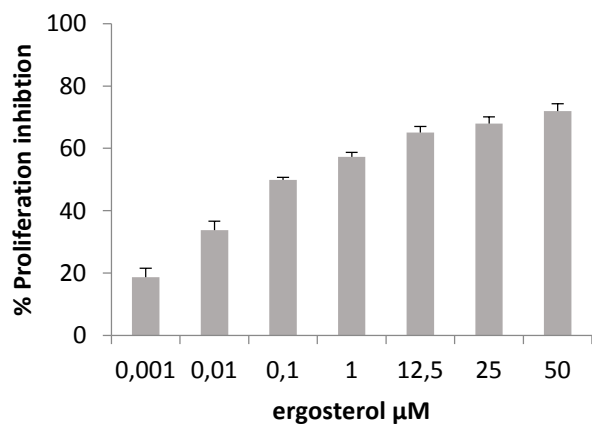
^e Biogenouest, Université de Nantes, Corsaire-ThalassOMICS, F-44000 Nantes, France

Table of contents

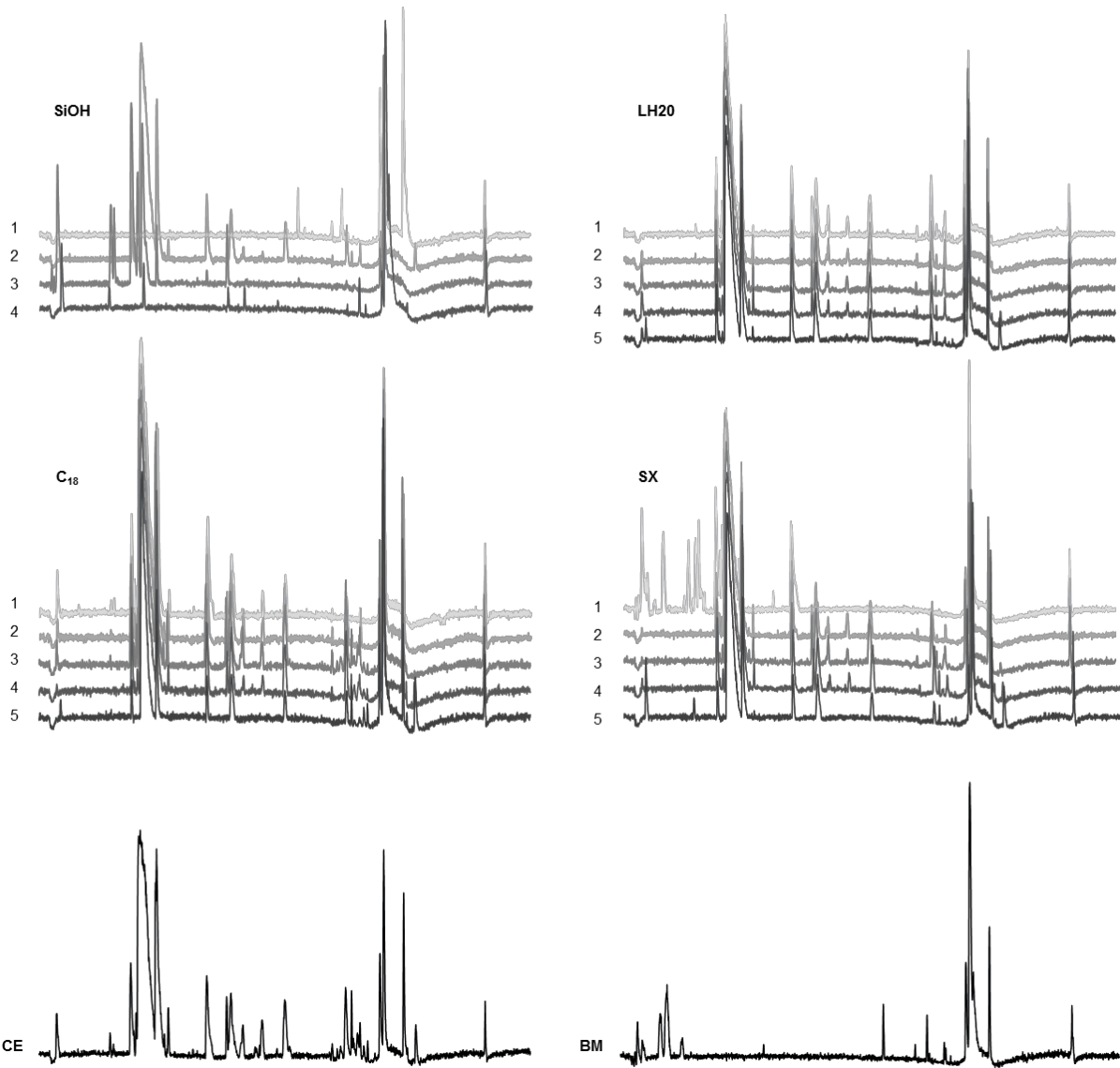
S 2.1: Files automatically generated by the FiBiCo script	2
S 2.2: Proliferation inhibition of MCF-7 breast cancer cells treated by ergosterol at concentrations ranging from 0.001 to 50 µM for 24h. Assays were performed in triplicate and values are expressed as means ± SD.	3
S 2.3: Overlay of LC-MS profiles obtained for MMS 5 fractionation with the four columns (SiOH, LH20, C ₁₈ and SX) and LC-MS profiles for the crude extract (CE) and blank medium (BM)	4
S 2.4: The "super list" generated by the FiBiCo script for MMS5 <i>P. chrysogenum</i> bioactive crude extract investigation	5
S 2.5: Radar charts representing normalized scores on each model for the top-15 features highlighted from the "Super list" and ordered by global score.	279
S 2.6: Identification of highlighted ions as ergosterol by comparison to a standard.....	280
S 2.7: Dereplication of the "Super list" generated by the FiBiCo script for MMS5 <i>P. chrysogenum</i> bioactive crude extract investigation.	8
S 2.8: Metabolites produced by <i>P. chrysogenum</i> strains reported in the literature for cytotoxic activity on breast cancer cells lines.	9
S 2.9: Virtual evaluation of the FiBiCo script performance with the addition of a virtual feature in the MMS5 <i>P. chrysogenum</i> data matrix.....	10

S 2.1: Files automatically generated by the FiBiCo script

Folder	Files
PCA	PCA_summary.csv pca_loadings.csv pca_score.csv PCA_loadings_plot.pdf PCA_score_plot.pdf
Spearman	Spearman.csv Spearman_score_plot.pdf
F-PCA	F-PCA.csv F-PCA_loadings_plot.pdf
PLS	PLS.csv PLS_loadings_plot.pdf PLS_RC_plot.pdf PLS_score_plot.pdf PLS_VIP_plot.pdf PLS_SR_plot.pdf PLS_summary.pdf
PLS-DA	plsda_vip.csv plsda-score.csv plsda.loadings.csv plsda_coef.csv PLS-DA.csv PLS-DA_coef.pdf PLS-DA_VIP.pdf PLS-DA_score.png PLS-DA_Pair_summary.png PLS-DA_loading.png PLS-DA_CV.csv
Summary_Results	Data_result.csv Four_models.pdf Super_list.csv VennDiagram.pdf Box_plot_score_distribution.pdf Radar_chart.pdf



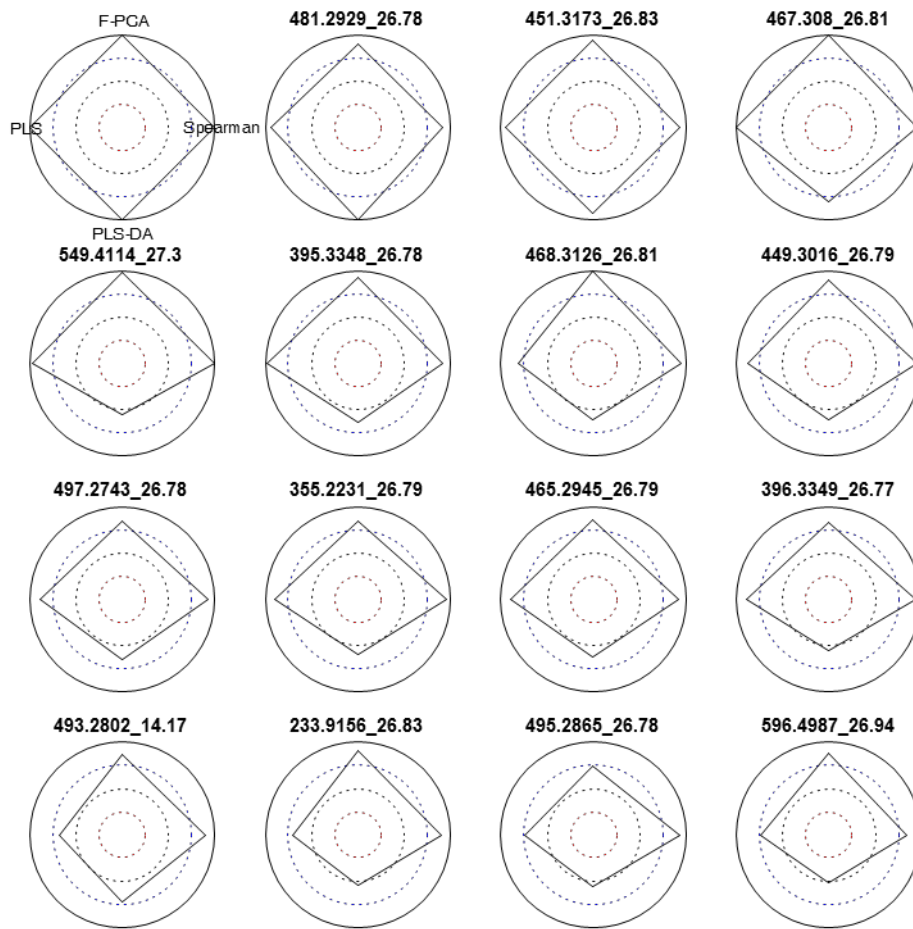
S 2.2: Proliferation inhibition of MCF-7 breast cancer cells treated by ergosterol at concentrations ranging from 0.001 to 50 μM for 24h. Assays were performed in triplicate and values are expressed as means \pm SD.



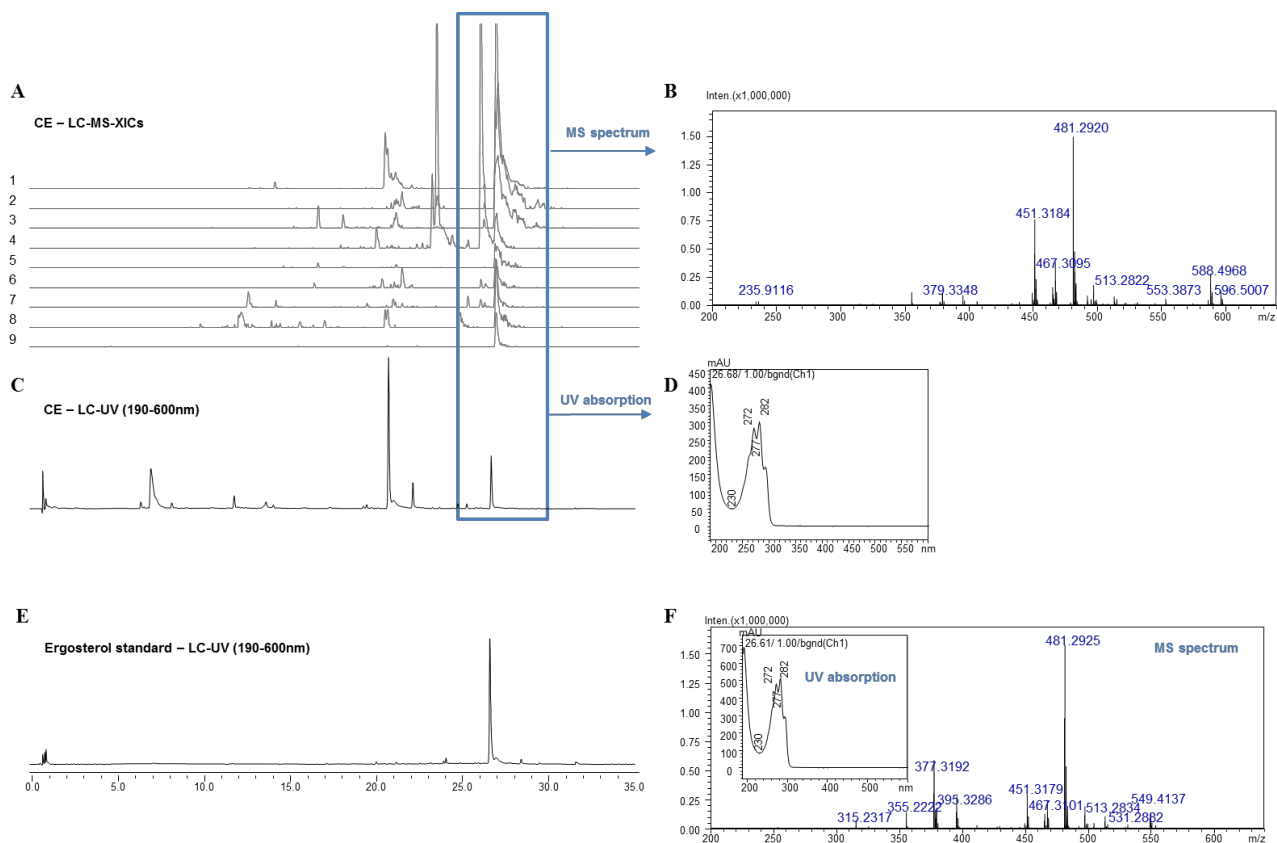
S 2.3: Overlay of LC-MS profiles obtained for MMS 5 fractionation with the four columns (SiOH, LH20, C₁₈ and SX) and LC-MS profiles for the crude extract (CE) and blank medium (BM)

S 2.4: The “super list” generated by the FiBiCo script for MMS5 *P. chrysogenum* bioactive crude extract investigation

Features	Spearman_r	F-PCA_score	PLS_SR	PLS-DA_VIP	Models	SCORE
481.2929_26.78	0.76	1.2	1.93	8.01	F-PCA PLS PLS-DA	3.53
451.3173_26.83	0.78	1.23	1.98	6.82	Spearman F-PCA PLS PLS-DA	3.51
467.3080_26.81	0.8	1.27	2.18	4.79	Spearman F-PCA PLS PLS-DA	3.49
549.4114_27.3	0.88	1.26	2.05	0.82	Spearman F-PCA PLS	3.01
395.3348_26.78	0.76	1.22	2.11	2.24	F-PCA PLS	2.95
468.3126_26.81	0.8	1.27	1.46	1.65	Spearman F-PCA	2.72
449.3016_26.79	0.8	1.19	1.73	1.77	Spearman F-PCA PLS	2.67
497.2743_26.78	0.77	1.15	1.74	2.48	PLS	2.61
355.2231_26.79	0.82	1.15	1.8	1.58	Spearman PLS	2.59
465.2945_26.79	0.76	1.16	1.74	2.15	PLS	2.58
396.3349_26.77	0.79	1.14	1.75	0.91	Spearman PLS	2.42
493.2802_14.17	0.72	1.17	0.92	3.54	F-PCA PLS-DA	2.31
233.9156_26.83	0.73	1.2	1.01	0.63	F-PCA	2.09
495.2865_26.78	0.77	1.08	1.21	0.97	Spearman	1.96
596.4987_26.94	0.64	1.18	1.14	0.37	F-PCA	1.96
529.2772_15.13	0.81	1.13	0.96	0	Spearman	1.89
465.2941_25.93	0.78	0.98	0.96	0.8	Spearman	1.55
395.3310_23.38	0.6	0.95	0.7	3.66	PLS-DA	1.47
227.1743_1.54	0.23	0.93	0.25	4.87	PLS-DA	0.84
515.2614_12.58	0.35	0.9	0.42	3.34	PLS-DA	0.81
453.3399_4.93	0.22	0.94	0.28	3.5	PLS-DA	0.7
475.3243_4.93	0.11	0.95	0.33	3.9	PLS-DA	0.66
249.1566_1.55	0.16	0.93	0.27	3.72	PLS-DA	0.62



S 2.5: Radar charts representing normalized scores on each model for the top-15 features highlighted from the “Super list” and ordered by global score.



S 2.6: Identification of the 12 highlighted ions as ergosterol by comparison to a standard.

Overlay of LC-MS extracted ion chromatograms (A) of the most representative features (9 out of 12) highlighted by the “Super list” from MMS5 *P. chrysogenum* crude extract (CE) with the followings m/z values: 481.2929 (1), 451.3173 (2), 467.3080 (3), 395.3348 (4), 468.3126 (5), 449.3016 (6), 465.2945 (7), 497.2743 (8), 355.2231 (9); (B) the corresponding MS spectrum at 26.8 min for the chromatographic peak; (C) LC-UV chromatographic profile of CE with (D) the corresponding UV absorption spectrum from 190 to 600 nm at 26.8 min; (E) LC-UV chromatographic profile of ergosterol standard injected on the same instrument with the same conditions, with the corresponding (F) MS and UV spectra at 26.8 min. Note: 3 highlighted ions, namely 396.3349; 233.9156 and 495.2865 are not reported here due to lower intensity in CE.

S2.7: Dereplication of the "Super list" generated by the FiBiCo script for MMS5 *P. chrysogenum* bioactive crude extract investigation.

RT (min)	Features "m/z_RT"	Rank	Spearmen	F-PCA	PLS	PLS-DA	Score ^a	Max. intensity (fractions)	Intensity (Crude Extract)	Adducts annotation	Proposed Molecular Formula	Annotation	Described activity on breast cancer cells
1.5	227.1743_1.54	20	x				0.73	5.70E+06 (S10H-4)	ND ^b	[M+H] ⁺	C ₁₂ H ₂₂ N ₂ O ₂	> Cyclo(Leu-Leu) > Cyclo (Ile-Leu) > Cyclo (Ile-Ile)	not reported
1.6	249.1566_1.55	23	x				0.49	2.10E+06 (S10H-4)	ND	[M+Na] ⁺		> 1,8,15,22-tetraaza-cyclooctacosane-2,9,16,23-tetraone > 1,8,15,22-tetraaza-cyclooctacosane-2,7,16,21-tetraone > Cyclo(leucyl-isoleucyl-Heuyl-Hisoleucyl) > Cyclo(leucyl-isoleucyl-Heuyl-Hisoleucyl)	not reported
4.9	453.3395_4.93	21	x				0.58	2.70E+06 (S10H-4)	ND	[M+H] ⁺	C ₂₃ H ₄₄ N ₄ O ₄	> 2S)-2-[(1S,5S,4aS,7S,8S,8aS)-7-(ethylcarbamoyl)amino]-1-hydroxy-4a,8-dimethyloctahydro-1H-naphthalen-2-yl]-N-[2-(morpholin-4-yl)ethyl]propanamide > Cyclo(leucyl-4-hydroxypropyl-Heuyl-4-hydroxypropyl)	not reported
4.9	475.3243_4.93	22	x				0.54	1.80E+06 (S10H-4)	ND	[M+Na] ⁺		> Cyclo(leucyl-4-hydroxypropyl-Heuyl-4-hydroxypropyl)	not reported
14.2	493.2802_14.17	12	x				2.24	8.50E+05 (C18-5)	5.20E+05	[M+H] ⁺	C ₂₉ H ₅₀ N ₂ O ₅	> Myochalasin Y > Sordarin > Sordarin B > Sordarin	not reported
14.2	515.2617_14.18	17	x				1.80	7.40E+05 (S-X-5)	4.50E+05	[M+Na] ⁺	C ₂₇ H ₄₀ O ₈	> g-sordarin	not reported
15.1	529.2772_15.13	15	x				1.91	2.00E+05 (S10H-2)	1.70E+05	[M+H] ⁺ ?	C ₃₂ H ₅₈ N ₂ O ₅ C ₃₀ H ₄₀ O ₈	13 hits > Chaetoglobosin derivatives incl Chaetoglobosin A, Chaetoglobosin C and Chaetoglobosin G > Chaetohydrone derivatives	IC ₅₀ of 37.56, 19.97 and 38.77 μM on MDA-MB-435 cell line, respectively. [1]
23.4	395.3310_23.38	19	x				1.34	2.00E+06 (C18-5)	2.10E+06			Confirmed as ergosterol-5,8-endoperoxide	48.7% inhibition of MCF-7 cells proliferation at 50 μM ^c
25.9	465.2941_25.93	18	x				1.41	2.10E+05 (C18-5)	8.20E+04	[M+H] ⁺ ?	C ₂₈ H ₄₀ O ₇	6 hits > Asnovolin D > Lasiojasmonate B > Lasiojasmonate C > Tandylkisin B > 19-(1α-d-glucopyranosyloxy)sopimara-7,15-dien-3-one > Virescoside C	not reported
26.8	481.2929_26.78	1	x	x	x		3.51	3.40E+06 (C18-5)	3.20E+06			Confirmed as ergosterol standard	IC ₅₀ of 0.1 μM on MCF-7 cell line ^e
26.8	451.3173_26.83	2	x	x	x		3.48	1.10E+06 (LH20-5)	9.50E+05				
26.8	467.3080_26.81	3	x	x	x		3.46	5.40E+05 (C18-5)	4.40E+05				
26.8	395.3348_26.78	5	x	x	x		2.87	3.10E+05 (LH20-5)	2.80E+05				
26.8	468.3126_26.81	6	x	x	x		2.65	1.90E+05 (LH20-5)	1.50E+05				
26.8	449.3016_26.79	7	x	x	x		2.59	2.30E+05 (S-X-5)	2.00E+05				
26.8	497.2743_26.78	8	x	x	x		2.52	5.20E+05 (S-X-5)	4.50E+05				
26.8	355.2231_26.79	9	x	x	x		2.50	3.00E+05 (C18-5)	2.70E+05				
26.8	465.2945_26.79	10	x	x	x		2.49	4.40E+05 (LH20-5)	4.20E+05				
26.8	396.3349_26.77	11	x	x	x		2.31	1.30E+05 (LH20-5)	8.90E+04				
26.8	233.9156_26.83	13	x				1.99	8.20E+04 (LH20-5)	4.30E+04				
26.8	495.2865_26.78	16	x				1.84	1.40E+05 (S-X-5)	1.30E+05				
26.9	596.4987_26.94	14	x				1.97	3.20E+05 (S10H-2)	5.80E+05	[M+H] ⁺ ?	?	?	?
27.3	549.4114_27.30	4	x	x	x		3.07	4.90E+05 (S-X-5)	4.70E+05	[M+H] ⁺ ?	C ₄₀ H ₆₂ O C ₃₃ H ₅₈ O ₆	incl. 3β,25-dihydroxyergost-5-en-3-yl-β-D-Arabinopyranoside	not reported

^a Global Score generated by the R FiBiCo script with the associated colour scale (red to green for the highest)

^b Not detected

^c Determined in the present study

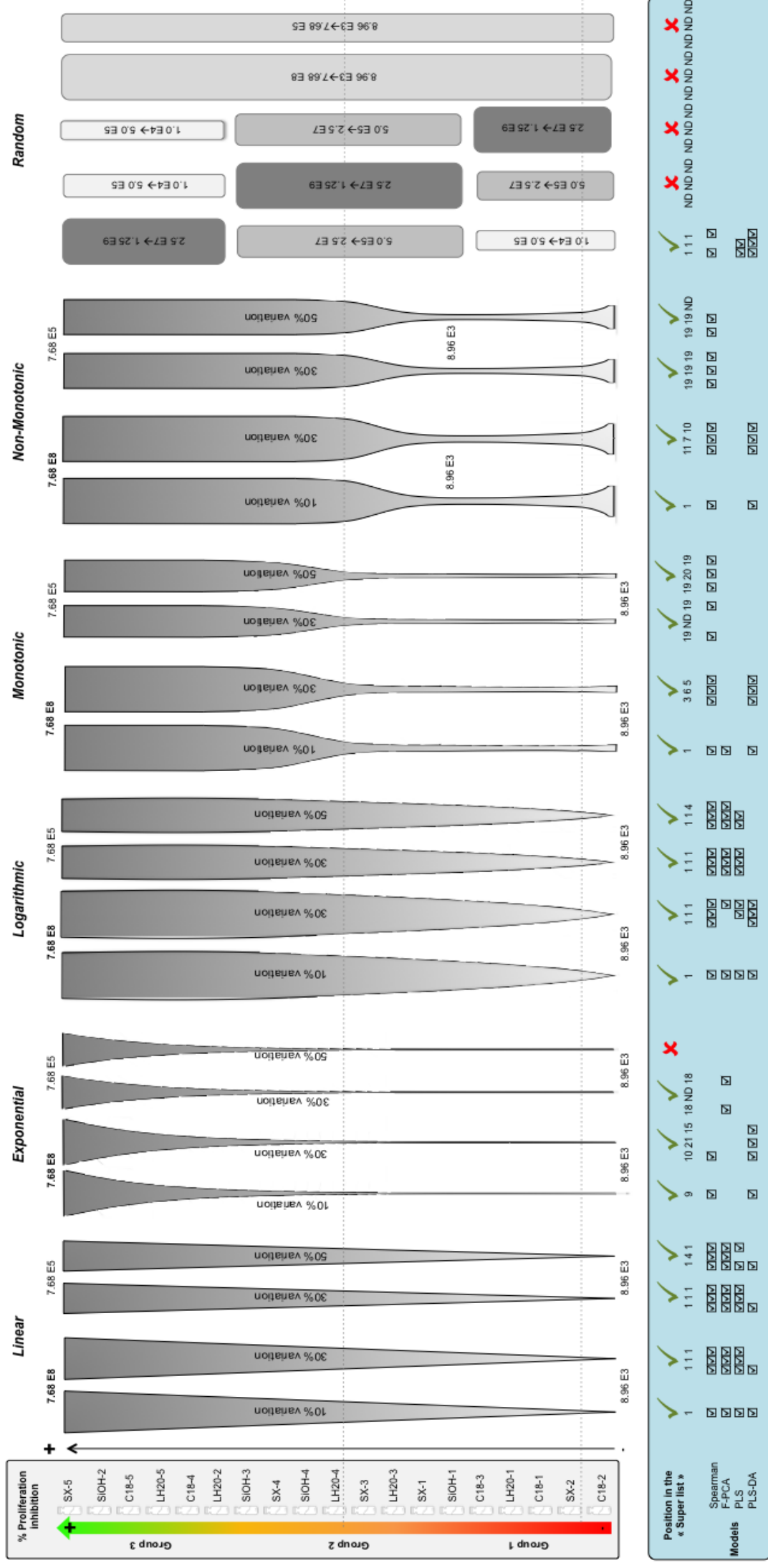
[1] S. Huang, H. Chen, W. Li, X. Zhu, W. Ding, C. Li, Bioactive chaetoglobosins from the mangrove endophytic fungus *Penicillium chrysogenum*. Mar. Drugs. 14 (2016) 1-12.

S 2.8: Metabolites produced by *P. chrysogenum* strains reported in the literature for cytotoxic activity on breast cancer cells lines.

COMPOUND	CHEMICAL CLASS	MF	ISOLATION from <i>P. chrysogenum</i> [Ref]	CYTOTOXIC ACTIVITY on breast cancer cells (target/IC ₅₀)	[Ref]	Corresponding features (m/z_Rt) in the present study
Butyrolactone V	Lactone	C ₂₄ H ₂₄ O ₈	[1]	MCF-7 (31.9 μM) MDA-MB-231 (22.2 μM)	[2]	none for [M+H] ⁺ and [M+Na] ⁺
Chaetoglobosin A	Alkaloid		[3]	MDA-MB-435 (37.56 μM)	[4]	
Chaetoglobosin C	Alkaloid	C ₃₂ H ₃₆ N ₂ O ₅	[3]	MDA-MB-435 (19.97 μM)	[4]	529.2772_15.1*
Chaetoglobosin G	Alkaloid		[4]	MDA-MB-435 (38.77 μM)	[4]	
Penochalasin K	Alkaloid		[3]	MDA-MB-435 (4.65 μM)	[3]	
Penochalasin I	Alkaloid	C ₃₂ H ₃₄ N ₂ O ₄	[3]	MDA-MB-435 (7.55 μM)	[4]	none for [M+H] ⁺ and [M+Na] ⁺
Chaetoglobosin E	Alkaloid		[4]	MDA-MB-435 (>40 μM)	[4]	
Chaetoglobosin F	Alkaloid	C ₃₂ H ₃₈ N ₂ O ₅	[4]	MDA-MB-435 (37.77 μM)	[4]	none for [M+H] ⁺ and [M+Na] ⁺
Cytoglobosin C	Alkaloid		[4]	MDA-MB-435 (12.58 μM)	[4]	
Penochalasin J	Alkaloid		[4]	MDA-MB-435 (36.68 μM)	[4]	none for [M+H] ⁺ and [M+Na] ⁺
Roquefortine C	Alkaloid/ Diketopiperazine	C ₂₂ H ₂₃ N ₅ O ₂	[5]	MCF-7 (21.7 μM) MCF7-dox (24.4 μM) BT-474 (29.6 μM) SKBR-3 (23.4 μM) MDA-MB-231 (23.5 μM) MDA-MB-468 (28.6 μM) T-47D (N.D.)	[5]	390.1905_8.3 min
Meleagrins	Alkaloid	C ₂₃ H ₂₃ N ₅ O ₄	[5]	MCF7 (1.9 μM) MCF7-dox (3.2 μM) BT-474 (2.7 μM) SKBR-3 (2.8 μM) MDA-MB-231 (6.8 μM) MDA-MB-468 (8.9 μM) T-47D (> 40 μM)	[5]	434.1805_7.3 min
				MCF-7 (no activity) MDA-MB-231 (25 μM)	[6]	
				MCF-7 (93 μM)	[8]	
Ergosterol	steroid	C ₂₈ H ₄₄ O	[7]	IC ₅₀ range from 50 to 150 μM for : MDA-MB-231 MDA-MB-468 SK-BR-3 MCF-7 4T1	[9]	Identified with standard at 26.8 min *
Ergosterol 5,8-endoperoxyde	steroid	C ₂₈ H ₄₄ O ₃	[10]	MCF-7 (97 μM)	[8]	Identified with standard at 23.4 min *

* Features highlighted by the R FIBICo script in the present study

- [1] X. Lin, Z. Tu, B. Zhao, Z. Wang, S. Xu, Y. Liu, J. Wang, K. Kaliaperumal, J. Li, X. Liao, S. Chen, Structurally diverse secondary metabolites from a deep-sea-derived fungus *Penicillium chrysogenum* SCSIO 41001 and their biological evaluation, *Fitoterapia*. 117 (2017) 71-78.
- [2] I.P. da Silva, E. Brissow, L.C. Kellner Filho, J. Senabio, K.A. de Siqueira, S. Vandresen Filho, J.L. Damasceno, A.A. Mendes, D.C. Tavares, L.G. Magalhães, P.A. Junior, A.H. Januário, M.A. Soares, Bioactive compounds of *Aspergillus terreus*-F7, an endophytic fungus from *Hyptis suaveolens* (L.) Poit, *World J. Microbiol. Biotechnol.* 33 (2017) 1-10.
- [3] C. Li, X. Zhu, Z. She, F. Liang, Z. Wu, D. Zhou, Penochalasin K, a new unusual chaetoglobosin from the mangrove endophytic fungus *Penicillium chrysogenum* V11 and its effective semi-synthesis, *Fitoterapia*. 123 (2017) 23-28
- [4] S. Huang, H. Chen, W. Li, X. Zhu, W. Ding, C. Li, Bioactive chaetoglobosins from the mangrove endophytic fungus *Penicillium chrysogenum*, *Mar. Drugs*. 14 (2016) 1-12.
- [5] M.S. Mady, M.M. Mohyeldin, H.Y. Ebrahim, H.E. Elsayed, W.E. Houssen, E.G. Haggag, R.F. Soliman, K.A. El Sayed, The indole alkaloid meleagrins, from the olive tree endophytic fungus *Penicillium chrysogenum*, as a novel lead for the control of c-Met-dependent breast cancer proliferation, migration and invasion, *Bioorg. Med. Chem.* 24 (216) 113-122.
- [6] Z. Shang, X. Li, L. Meng, C. Li, S. Gao, C. Huang, B. Wang, Chemical profile of the secondary metabolites produced by a deep-sea sediment-derived fungus *Penicillium commune* SD-118, *Chinese J. Oceanol. Limnol.* 30 (2012) 305-314
- [7] J. Wang, Y. Zhao, L. Men, Y. Zhang, Z. Liu, T. Sun, Y. Geng, Z. Yu, Secondary metabolites of the marine fungus *Penicillium chrysogenum*, *Chem Nat Compd.* 50 (2014) 405-407.
- [8] M.J. Song, H.Y. Bao, T. Bau, Y. Li, Antitumor activity and structure-activity relationship of four steroids from *Fomitiporia ellipsoidea*, *34* (2015) 293-300.
- [9] X. Li, Q. Wu, Y. Xie, Y. Ding, W.W. Du, M. Sdiri, B.B. Yang, Ergosterol purified from medicinal mushroom *Amauroderma rube* inhibits cancer growth *in vitro* and *in vivo* by up-regulating multiple tumor suppressors, *Oncotarget*. 6 (2015) 17832-17846.
- [10] S. Gao, X. Li, C. Li, P. Proksch, B. Wang, Penicilsteroids A and B, antifungal and cytotoxic polyoxygenated steroids from the marine alga-derived endophytic fungus *Penicillium chrysogenum* QEN-245, *Bioorg. Med. Chem. Lett.* 21 (2011) 2894-2897.



S 2.9: Virtual evaluation of the FiBiCo script performance with the addition of a virtual feature in the MMS5 *P.chrysogenum* data matrix.

The figure presents all the conditions tested for the additional feature designing six different relationships between peak area value and biological activity: linear,

exponential, logarithmic, monotonic, non-monotonic, and random. For all relationships peak area values in each fraction for the virtual feature were given in accordance to the antiproliferative activity on MCF-7 (in %) with lowest peak area value for the less active fraction C18-2 (except for the non monotonic model) and the highest for the most active fraction SX-5. Based on the initial matrix, where peak area values ranged from 8.96 E3 to 7.68 E8, different matrices were then constructed with the additional virtual feature having peak area values either in the same range (from 8.96 E3 to 7.68 E8) or 1000 times less intense (from 8.96 E3 to 7.68 E5). A random variation of 10%, 30% and/or 50% was applied (using “randbetween” function in excel) on each virtual peak area value defined for each model. All virtual features were designed in triplicates except for the first virtual features of each model having a range from 8.96 E3 to 7.68 E8 and a random variation of 10%. Consequently a total of 10 matrices for each relationship between peak area value and biological (linear, exponential, logarithmic, monotonic, non-monotonic) were obtained. Additionally, three additional matrices were obtained with virtual peak area values defined according to activity groups. This time, random values were assigned (using “randbetween” function) to the virtual feature based on the following conditions: 1) in accordance with biological activity groups, with peak area values in group 1 (the less active) ranging from 1.00 E4 to 5.00 E5, in group 2 from 5.00 E5 to 2.50 E7 and in group 3 from 2.50 E7 to 1.25 E9, 2) not in accordance with biological activities with values in group 3 < group 2 < group 1 or with values in group 2 < group 1 < group 3. Finally, a last assay was performed with completely random peak area values assigned to the virtual feature based on the initial matrix range (from 8.96 E3 to 7.68 E8) or on a smaller range (from 8.96 E3 to 7.68 E5) for the 19 samples. These experiments were repeated in triplicate. The R FIBiCo script was applied on the modified initial matrix with this additional feature generating its position in the “Super list” and revealing which models allowed its selection.

Annexe 24. R FiBiCo script disponible en version html

1. Raw data description

Define important informations before using the script. **Make changes below (in the grey zones) according to your operating system and your data. Then, select all lines (CTRL + A or CMD + A) and run all lines (CTRL + Enter).**

1.1 General parameters

- Operating System: Windows (win) or MacOS (mac)

```
OpSys = "win"
```

- Directory to upload the data (.CSV file)

```
filePath="C:/Users/data_matrix.csv"
```

- Working Directory to export the results (The working directory must exists)

```
WD="/Users/lucie/Nextcloud/"
```

```
#"C:/Users/Results"
```

```
setwd(WD)
```

1.2. Specific parameters according to biological activity

- Type of biological activity (ex: "Proliferation inhibition")

```
title = "Proliferation inhibition"
```

- Range and colours for groups (maximum 6 groups) replacing aa= "0%<20" by your activity data from less active to best active groups (ex: "0<Activity<20" or "group A"...etc) and choosing a colour for each group, "colour1" being for less active group (ex: "red", "yellow",etc)

```
aa = "0%<20"
```

```
bb = "20%<40"
```

```
cc = "40%<60"
```

```
dd = NULL
```

```
ee = NULL
```

```
ff = NULL
```

```
colour1 = "red"
```

```
colour2 = "orange"
```

```
colour3 = "green"
```

```
colour4 = NULL
```

```
colour5 = NULL
```

```
colour6 = NULL
```

1.3 Parameters for normalization

Normalization can be divided in three categories as sample normalization (rowNorm), data transformation (transNorm) and data scaling (scaleNorm):

- rowNorm:
QuantileNorm, ProbNormT, ProbNormF, CompNorm, SumNorm, MedianNorm, SpecNorm
or NULL

- transNorm: LogNorm,CrNorm or NULL
- scaleNorm: MeanCenter,AutoNorm,Paratonorm,RangeNorm

```
rowNorm = "NULL"
transNorm = "NULL"
scaleNorm = "ParetoNorm"
```

1.4 Parameters for statistical analyses

Principal Component Analysis (PCA)

Define xy components and legends

```
CompX = 1
CompY = 2
xlab= "Comp 1"
ylab= "Comp 2"
```

Partial Least Square Discriminant Analysis (PLS-DA)

- Define parameters for classification and performance of model:
 - methodName= "T" for 10-fold CV or "L" for ("leave one out cross validation")
 - compNumber: Number of components for analysis
 - choice = Q2 or R2 for performance measure

```
methodName = "L"
compNumber= 5
choice = "Q2"
```

- The best component for Very Importance in Projection (VIP):

```
Comp= "Comp. .1"
feat.nm = "Comp. 1"
```

Parameters for the combination of the four models

- Define the number of desired highlighted features in each model

```
NBofInterest<-10
```

- Define the computational mode to obtain a global score from all models:
 - Sum as "SUM"
 - Euclidean distance as "DIST"
 - Mean as "MEAN"

```
SCORE<- "SUM"
```

1.5 Required packages and libraries to run the script

Packages installation

Before using the script for the first time

```
metanr_packages <- function(){
  metr_pkgs <- c("caret", "Rserve", "ellipse", "scatterplot3d", "Cair
```

```

o",
      "randomForest", "caTools", "e1071", "som", "impute"
, "pcaMethods",
      "RJSONIO", "ROCR", "globaltest", "GlobalAncova", "Rg
raphviz",
      "preprocessCore", "genefilter", "pheatmap", "SSPA",
"sva", "Rcpp",
      "pROC", "data.table", "limma", "car", "fitdistrplus
", "lars",
      "Hmisc", "magrittr", "methods", "xtable", "pls", "ig
raph", "gplots",
      "KEGGgraph", "reshape", "RColorBrewer", "tibble", "R
cpp", "siggenes",
      "devtools", "psy", "mdatools", "readr", "psych", "Venn
Diagram")

list_installed <- installed.packages()

new_pkgs <- subset(metr_pkgs, !(metr_pkgs %in% list_installed[, "P
ackage"]))

if(length(new_pkgs)!=0){

  source("http://bioconductor.org/biocLite.R")
  biocLite(new_pkgs, dependencies = TRUE, ask = FALSE)
  print(c(new_pkgs, " packages added..."))
}

if((length(new_pkgs)<1)){
  print("No new packages added...")
}
}

metanr_packages()
if (length(which(search()=="package:MetaboAnalystR"))<0){
  devtools::install_github("xia-lab/MetaboAnalystR", build_vignettes
=TRUE)
}else{
  print ("MetaboAnalystR has been already added...")
}
}

```

In case package installation fails, manually download packages “psy”, “mdatools”, “readr”, “psych”, “VennDiagram” and MetaboAnalyst.tar.gz file (https://github.com/jpsychong/MetaboAnalystR/blob/master/MetaboAnalystR_1.0.0.tar.gz) and run the following command, where ~ correspond to the path:
install.packages("~/MetaboAnalystR_1.0.0.tar.gz", repos = NULL, type = “source”)

Libraries required for each utilization


```
library(devtools)
library(psy)
library(mdatools)
library(readr)
library(psych)
library(MetaboAnalystR)
library(VennDiagram)
```

1.6 Data format

Preprocessing of MS data can be performed with Mzmine software to obtain a list of peaks (features defined by m/z values and retention times) for all samples. Then it can be exported in .CSV and edited with excel.

Please organize your data input as below.

The peak area or intensity table in a .csv file with separator = “,” is organized as:

- Rows: Samples (1 row per sample)
- Columns:
 - 1st: “Samples names”
 - 2nd: “Labels”
 - 3rd: “Biological activity”
 - others: “X” columns for the “X” detected ions associated with their retention time (m/z_RT = features)

In our study:

Preprocessing of MS data was performed with Mzmine software, exported in .CSV and edited with excel to obtain a matrix with:

- 19 samples (19 rows)
- 795 features (detected ions) (columns 4 to 799)
- Biological activity corresponding to % inhibition values (3rd column)
- Labels = Group classification according to the breast cancer cell inhibition of samples: (2nd column)
 - label 1 = $0 < \%inhibition < 20$
 - label 2 = $20 < \%inhibition < 40$
 - label 3 = $40 < \%inhibition < 60$

2. Import, Check and Normalize data

Pay attention to the parameters defined for the normalization at the beginning of the script

2.1 Import data

```
datamatrix<-read_csv(filePath)
```

Recover the features numbers (defined as feat.num.tot), the column number for each features (defined as X) and biological activity denomination (defined as Y)

```
feat.num.tot=ncol(datamatrix)-3
X=3:(feat.num.tot+1)
Y=colnames(datamatrix[,3])
```

Create an object required for storing processed data from statistical analyzes

```
obj<-InitDataObjects("stat","pktable", paired=FALSE)
```

Upload file in a mSet object:

```
mSet<-Read.TextData(mSetObj=obj,filePath, format="rowu", lbl.type="disc")
```

2.2 Check data

```
mSet<-SanityCheckData(mSet)
mSet<-ReplaceMin(mSet)
```

2.3 Normalize data

Normalization can be divided in three categories as sample normalization (rowNorm), data transformation (transNorm) and data scaling (scaleNorm)

```
mSet<-Normalization(mSet, rowNorm=rowNorm, transNorm=transNorm, scaleNorm=scaleNorm, ref=NULL, ratio=FALSE, ratioNum=20)
```

2.3 Recover and save raw data and normalize data

Raw data

```
mSet$dataSet$orig1<-mSet$dataSet$orig
rawdata<-mSet$dataSet$orig1
```

Normalize data

```
mSet$dataSet$norm1<-mSet$dataSet$norm
normdata<-mSet$dataSet$norm1
```

3. Chemical diversity characterization with Principal Component Analysis (PCA)

Chemical investigation with a principal component analysis (PCA) is required to describe chemical diversity among samples and identify possible outliers.

Several results are created and exported in the working directory:

- Tables (.csv)
 - PCA score for 15 components (pca_score.csv)
 - PCA loadings (pca_loadings.csv)
 - A summary table with: Standard deviation, Proportion of Variance, Cumulative Proportion of Variance (PCA_summary.csv)
- Plot (.pdf)

- A PCA score plot with samples coloured according to groups defined by biological activity (PCA_score_plot.pdf)
- A PCA loadings plot (PCA_loadings_plot.pdf)

```

# Define "PCA" working directory
dir.create(file.path(WD, "PCA"))
setwd(file.path(WD, "PCA"))

# The PCA function from "MetaboAnalystR" packages is modified
mSet2<-mSet
mSet2$dataSet$norm<-mSet$dataSet$norm[,2:ncol(mSet$dataSet$norm)]
pca <- prcomp(mSet2$dataSet$norm, center = TRUE, scale = F)

# A summary table is created
sum.pca <- summary(pca)
PCA_Summary<- sum.pca$importance
write.table(PCA_Summary,"PCA_summary.csv",sep=";",dec=".")

imp.pca <- sum.pca$importance
std.pca <- imp.pca[1, ]
var.pca <- imp.pca[2, ]
cum.pca <- imp.pca[3, ]
mSet2$analSet$pca <- append(pca, list(std = std.pca, variance =
var.pca,
    cum.var = cum.pca))

# The two tables "pca_score.csv" and "pca_loadings.csv" are ex
ported on the "PCA" working directory
pca_score<-data.frame(cbind(mSet2$analSet$pca$x,mSet2$dataSet$c1
s))
names(pca_score)[ncol(pca_score)]<-"Group classification"
write.table(signif(pca_score,5),file="pca_score.csv",sep=";",dec
=".",row.names = T)

pca_loadings<-data.frame(mSet2$analSet$pca$rotation,5)
write.table(signif(pca_loadings,5),file="pca_loadings.csv",sep="
;",dec=".",row.names = T)

# PCA score plot
if (OpSys == "mac"){
  quartz(6,6, title="PCA score")
} else {
  X11(6,6, title="PCA score")
}

op<-par(pty="s")
pcx<-pca_score[,CompX]
pcy<-pca_score[,CompY]

# Create the Legend
xlab<-paste(xlab,"(",PCA_Summary[2,CompX]*100,"%",")",sep=' ')

```

```

ylab<-paste(ylab,"(",PCA_Summary[2,Compy]*100,"%"),",",sep=' ')

# Plot
plot(pcx,pcy,main="PCA_score",ylab=ylab,xlab=xlab,cex.axis=1,
     pch=c(15,16,17,18,19,20)[as.factor(pca_score[, "Group classification"])],
     col=c(colour1,colour2,colour3,colour4,colour5,colour6)[as.factor(pca_score[, "Group classification"])])

# Add the axes (x=0 and y=0)
abline(a=0,b=0,h=TRUE,v=TRUE,col="grey")

# Add the samples labels
text(pcx,pcy, label = rownames(pca_score), pos = 1, col = "black",
     xpd = T, cex = 0.5,offset = 0.5)

# Add the legend of groups classification according to biological activity
legend("topleft",inset =.05, title= title, c(aa, bb,cc,dd,ee,ff)
      ,fill = c(colour1,colour2,colour3,colour4,colour5,colour6)
      ), cex=0.5)

# Save the PCA score plot
dev.copy2pdf(file="PCA_score_plot.pdf",device=dev.cur())

# PCA Loadings plot
if (OpSys == "mac"){
  quartz(6,6, title="PCA loadings")
} else {
  X11(6,6, title="PCA loadings")
}
op<-par(pty="s")
Lx<-pca_loadings[,CompX]
Ly<-pca_loadings[,Compy]

# Create the Legend
xlab<-paste("Loadings",CompX,sep=' ')
ylab<-paste("Loadings",Compy,sep=' ')

# Plot
plot(Lx,Ly,main="PCA_loadings",ylab=ylab,xlab=xlab,cex.axis=1,pch=
20,cex=1,col="grey")

# Add the axes (x=0 and y=0)
abline(a=0,b=0,h=TRUE,v=TRUE,col="grey")

# Add the features labels
text(Lx,Ly, label = rownames(pca_loadings), pos = 2, col = "black"
,
     xpd = T, cex = 0.5)

```

```
# Save the PCA Loadings plot
dev.copy2pdf(file="PCA_loadings_plot.pdf", device=dev.cur())
```

4. FOUR BIOCHEMOMETRICS models

This workflow allows to perform statistical analyses of metabolomics data to highlight molecules, which seem to be responsible for biological activity.

Relationship study between LC-MS and biological data is performed with several biochemometrics models:

- Spearman
- Focalised Principal Component Analysis (F-PCA)
- Partial Least Square (PLS)
- Partial Least Square Discriminant Analysis (PLS-DA)

4.1 Spearman analysis

Several results are created in "Spearman" working directory:

- Table (Spearman.csv) Spearman coefficient correlation table with importance levels annotation according to r score:
 - $r \geq 0.5 \rightarrow$ "High"
 - $0 < r < 0.5 \rightarrow$ "Medium"
 - $r < 0 \rightarrow$ "Low"
- Plot (Spearman_score_plot.pdf) Spearman coefficient correlation plot with the colour code is defined according to the importance levels:
 - Spearman correlation score "High" coloured in Blue
 - Spearman correlation score "Medium" coloured in Orange
 - Spearman correlation score "Low" coloured in Grey

```
# Define Working directory
dir.create(file.path(WD, "Spearman"))
setwd(file.path(WD, "Spearman"))

# Apply function
r<-cor(normdata[,2:(feat.num.tot+1)],normdata[,grep(Y,colnames (norm
data))], method="spearman")

# Define the three Importance Levels
Level<-ifelse(r[,1]>=0.5,"High",
              ifelse(r[,1]<0,"Low","Medium"))

# Merge in the same table the "r coefficient score" and "Importance
Levels" for each features
r<-data.frame(r[,1],Level)
colnames(r) <- c("r","Importance Levels")
row.names(r)<-sub("X","",row.names(r))
```

```

r$r<-round(r$r,5)
# Save the table in the "Spearman" working directory
write.table(cbind(features=row.names(r),r),file="Spearman.csv",sep=";",dec=".",row.names = F)

# Graphical representation
# Open the graphical window
if (OpSys == "mac"){
  quartz(6,6, title="Spearman")
} else {
  X11(6,6, title="Spearman")
}
op<-par(pty="s",mar=c(5,4,4,4))

# Define colours according to the importance level obtained for each feature
colour = rep(NA, length=length(r[,"Importance Levels"]))
colour[which(r[,"Importance Levels"]=="High")] = "blue"
colour[which(r[,"Importance Levels"]=="Medium")] = "#FF9966"
colour[which(r[,"Importance Levels"]=="Low")] = "grey"

# Plot function
plot(x=as.numeric(substr(row.names(r),regexr(pattern = "_",text=row.names(r),fixed=TRUE)+1,nchar(row.names(r))))),
      y=r[, 1],
      type='h',pch=19,las=1,cex=0.4, main="Spearman",
      xlab = "Retention time (min)",col.sub="grey",font.sub=4,cex.sub=0.75,
      ylab = "Coefficient (r) ",
      axes=T,col=colour)

# Add a red line to define the threshold for "r" score
abline(h=0.5,col="red",lwd=0.7,lty=3)

# Add the Legend
legend("bottomright",inset =0, title="Importance Levels", c("High", "Medium", "Low"),
      ,fill = c("blue","#FF9966", "grey"), cex=0.5)

# Add the labels of highlighted features having "High" importance level
text(y=r[which(r[,"Importance Levels"]=="High"), 1],x= as.numeric(substr(row.names(r),regexr(pattern = "_",text=row.names(r),fixed=TRUE)+1,nchar(row.names(r))))[which(r[,"Importance Levels"]=="High")],
label = row.names(r)[which(r[,"Importance Levels"]=="High")], pos = 1, col = "black",
      xpd = T, cex = 0.5,offset = -1.3,srt=75)

# Save the plot
dev.copy2pdf(file="Spearman_score_plot.pdf",device=dev.cur())

```

```

# Close the graphical window
dev.off()

# Graphical representation with the four biochemometrics models

if (OpSys == "mac"){
  quartz(11,11)
} else {
  X11(9,9)
}
layout(rbind(c(1,2),c(3,4)))
op<-par(pty="m",mar=c(5,4,4,4))

# Plot
plot(x=as.numeric(substr(row.names(r),regexr(pattern = "_",text=row.names(r),fixed=TRUE)+1,nchar(row.names(r)))),
y=r[, 1],
type='h',pch=19,las=1,cex=0.4, main="Spearman",
xlab = "Retention time (min)",col.sub="grey",font.sub=4,cex.sub=0.75,
ylab = "Coefficient (r) ",
axes=T,col=colour)

# Add a red line to define the threshold for "r" score
abline(h=0.5,col="red",lwd=0.7,lty=3)

# Add the Legend
legend("bottomright",inset =0, title="Importance Levels", c("High",
"Medium", "Low"),
,fill = c("blue", "#FF9966", "grey"), cex=0.5)

```

4.2. Focalised Principal Component Analysis (F-PCA)

Several results are created in "F-PCA" working directory:

- Table (F-PCA.csv) F-PCA scores table with Importance Levels annotation defined according to F-PCA score:
 - F-PCA score with significant positive correlation → "High"
 - F-PCA score with not significant positive correlation → "Medium"
 - F-PCA score with negative correlation → "Low"
- Plot (F_PCA_loadings_plot.pdf) F-PCA correlation plot with the colour code is defined according to the Importance Levels:
 - F-PCA score "High" coloured in blue
 - F-PCA score "Medium" coloured in orange
 - F-PCA score "Low" coloured in grey Moreover, significativity circle appeared in red colour

```

# Create "F-PCA" working directory
dir.create(file.path(WD, "F-PCA"))

```

```

setwd(file.path(WD, "F-PCA"))

# The fpca function from "psy" packages is modified
fpca_lucie <- function (formula = NULL, y = NULL, x = NULL, data=NULL,
  cx = 0.75,
  pvalues = NULL, partial = NULL, input = NULL, contraction = NULL,
  sample.size = 1)
{
  datafile <- data
  if (pvalues[1] != "No")
    partial <- "No"
  p <- ifelse(is.null(y), ncol(x), length(x))
  if (input == "data")
    n <- ifelse(is.null(y), nrow(x), dim(datafile)[1])
  else n <- sample.size
  if (input == "data") {
    if (is.null(y))
      mat <- mf
    else mat <- matrix(ncol = p + 1, nrow = n)
  }
  names <- matrix(nrow = p + 2)
  one2 <- matrix(1, nrow = p)
  load <- matrix(nrow = p, ncol = p)
  norm <- matrix(nrow = p, ncol = p - 1)
  loadx <- matrix(nrow = p + 2, ncol = p - 1)
  loadyp <- matrix(nrow = p + 2, ncol = p - 1)
  loadym <- matrix(nrow = p + 2, ncol = p - 1)
  loady <- matrix(nrow = p + 2, ncol = p - 1)
  q <- 1
  if (input == "data") {
    if (is.null(formula)) {
      q <- min(q, p - 1)
      mat[, 1] <- datafile[, c(y)]
      namey <- ifelse(is.numeric(y), attributes(datafile)$names[y]
,
      y)
      for (i in 1:p) {
        mat[, i + 1] <- datafile[, c(x[i])]
        names[i] <- ifelse(is.numeric(x[i]), attributes(datafile)$
names[x[i]],
        x[i])
      }
      mat <- na.omit(mat)
    }
    n <- dim(mat)[1]
    xv <- matrix(ncol = p, nrow = n)
    yv <- matrix(nrow = n)
    un <- matrix(1, ncol = p, nrow = n)

```



```

one <- matrix(1, nrow = n)
for (i in 1:p) {
  xv[, i] <- (mat[, i + 1] - mean(mat[, i + 1]))/(sqrt(var(mat
[,
i + 1])) * sqrt(n - 1))
}
yv <- mat[, 1]
yv <- (yv - mean(yv))/(sqrt(var(yv)) * sqrt(n - 1))
matcor <- cor(mat)
} else {
  if (is.null(formula)) {
    namey <- ifelse(is.numeric(y), attributes(datafile)$names[y]
,
y)
    for (i in 1:p) {
      names[i] <- ifelse(is.numeric(x[i]), attributes(datafile)$
names[x[i]],
x[i])
    }
    matcor <- matrix(nrow = p + 1, ncol = 1)
    matcor[1, 1] <- 1
    matcor[2:(p + 1), 1] <- datafile[x, y]
    matcorp <- datafile[x, x]
  }
  else {
    namex = names(mf[names(mf) != namey])
    X <- as.data.frame(t(x))[, c(namex)]
    matcor <- t(as.data.frame(t(Y))[, c(namey, namex)])
    matcorp <- X
  }
  decomp <- eigen(matcorp, symmetric = TRUE)
  eigenval <- decomp$values
  eigenvect <- decomp$vectors
  eigenval <- pmax(0 * one2, eigenval)
  load <- eigenvect * sqrt(kronecker(one2, t(eigenval)))
}
if (pvalues[1] == "No") {
  if (input == "data") {
    if (partial == "Yes") {
      scal <- t(t(xv) %*% yv)
      xp <- xv - (un * yv) * kronecker(one, scal)
    }
    else {
      xp <- xv
    }
  }
  matcorp <- var(xp)
  decomp <- eigen(matcorp, symmetric = TRUE)
  eigenval <- decomp$values

```

```

eigenvect <- decomp$eigenvectors
eigenval <- pmax(0 * one2, eigenval)
load <- eigenvect * sqrt(kronecker(one2, t(eigenval)))
}
if (contraction == "No") {
  for (i in 1:p) {
    for (j in 1:p - 1) {
      norm[i, j] <- sqrt(load[i, 1] * load[i, 1] +
                        load[i, j + 1] * load[i, j + 1])
      loadx[i, j] <- load[i, 1] * sqrt(2 - 2 * abs(matcor[i +
1,
1]))/norm[i, j]
      if (matcor[i + 1, 1] > 0)
        loadyp[i, j] <- load[i, j + 1] * sqrt(2 -
2 * matcor[i +
1, 1])/norm[i, j]
      else loadym[i, j] <- load[i, j + 1] * sqrt(2 +
2 * matcor[
i + 1, 1])/norm[i, j]
      loady[i, j] <- load[i, j + 1] * sqrt(2 - 2 *
abs(matcor[i + 1,
1]))/norm[i, j] - 0.05
    }
  }
}
if (contraction != "No") {
  for (i in 1:p) {
    for (j in 1:p - 1) {
      norm[i, j] <- sqrt(load[i, 1] * load[i, 1] +
                        load[i, j + 1] * load[i, j + 1])
      loadx[i, j] <- 1.5 * load[i, 1] * (1 - abs(matcor[i +
1, 1
]))/norm[i, j]
      if (matcor[i + 1, 1] > 0)
        loadyp[i, j] <- 1.5 * load[i, j + 1] * (1 -
matcor[i + 1
, 1])/norm[i, j]
      else loadym[i, j] <- 1.5 * load[i, j + 1] *
(1 + matcor[i + 1, 1])/norm[i, j]
      loady[i, j] <- 1.5 * load[i, j + 1] * (1 -
abs(matcor[i +
1, 1]))/norm[i, j] - 0.05
    }
  }
}
}
if (pvalues[1] != "No") {
  matcorp <- var(xv, na.rm = TRUE)

```

```

decomp <- eigen(matcorp, symmetric = TRUE)
eigenval <- decomp$values
eigenvect <- decomp$vectors
eigenval <- pmax(0 * one2, eigenval)
load <- eigenvect * sqrt(kronecker(one2, t(eigenval)))
pnorm <- matrix(nrow = p)
pvaluesabs <- abs(pvalues)
pvaluesabs <- pmax(pvaluesabs, 0.001)
for (i in 1:p) if (pvaluesabs[i] == 0)
  pnorm[i] <- 0
else pnorm[i] <- pvaluesabs[i]^(log(pvaluesabs[i])/-50)
for (i in 1:p) {
  for (j in 1:p - 1) {
    norm[i, j] <- sqrt(load[i, 1] * load[i, 1] +
                      load[i, j + 1] * load[i, j + 1])
    loadx[i, j] <- 1.5 * load[i, 1] * pnorm[i]/norm[i,
                                                    j]
    if (pvalues[i] > 0)
      loadyp[i, j] <- 1.5 * load[i, j + 1] * pnorm[i]/norm[i,
                                                            j]
    else loadym[i, j] <- 1.5 * load[i, j + 1] * pnorm[i]/norm[
i,
j]
    loady[i, j] <- 1.5 * load[i, j + 1] * pnorm[i]/norm[i,
                                                         j] - 0
.05
  }
}
}
if (is.null(formula)) {
  for (j in 1:p - 1) {
    loadx[p + 1, j] <- 1.5
    loady[p + 1, j] <- 1.5
    loadx[p + 2, j] <- -1.5
    loady[p + 2, j] <- -1.5
  }
  names[p + 1] <- "."
  names[p + 2] <- "."
}
else {
  names = namex
}
j <- 1
{
  # plot function
  plot(x=NULL,y=NULL,type = "l", ylab='',xlab='',axes = FALSE,ce
x.axis=0.1,col.axis="white",
      xlim = c(-1.7, 1.7),

```

```

        ylim = c(-1.7, 1.7),
        col = "grey", main="F-PCA")

# Add the Legend
legend("bottomright",inset = 0.005, title="Importance Levels",
c("High ", "Medium" , "Low")
      ,fill = c("blue", "#FF9966", "grey"), cex=0.5)

# Add axes
lines(x = c(0,0), y = c(-1.6, 1.6), type = "l", col = "grey")
lines(x = c(-1.6,1.6), y = c(0,0), type = "l", col = "grey")
if (pvalues[1] == "No") {
  radius <- matrix(nrow = 5)
  if (contraction == "No") {
    radius[1] <- 1.414
    radius[2] <- 1.265
    radius[3] <- 1.095
    radius[4] <- 0.894
    radius[5] <- 0.632
  }
  else {
    radius[1] <- 1.5
    radius[2] <- 1.2
    radius[3] <- 0.9
    radius[4] <- 0.6
    radius[5] <- 0.3
  }
  symbols(x = 0, y = 0, circles = radius[1], inches = FALSE,
          add = TRUE, lwd = 2)
  symbols(x = c(0, 0, 0, 0), y = c(0, 0, 0, 0), circles = radi
us[2:5],
          inches = FALSE, add = TRUE, lwd = 1, fg = "grey")
}
if (pvalues[1] != "No") {
  symbols(x = 0, y = 0, circles = 1.5, inches = FALSE,
          add = TRUE, lwd = 2)
  symbols(x = c(0, 0, 0), y = c(0, 0, 0), circles = c(1.35,
1, 0.557
), inches = FALSE, add = TRUE, lwd = 1,
          fg = "grey")
  symbols(x = 0, y = 0, circles = 1.254, inches = FALSE,
          add = TRUE, lwd = 1, fg = "red")
}
symbols(x = 0, y = 0, circles = 0.03, bg = "red", inches = FAL
SE,
          add = TRUE, lwd = 1)
if (pvalues[1] == "No") {
  if (input == "data") {
    e <- exp(1.96 * 2/sqrt(n - 3))
  }
}

```

```

else {
  if (sample.size < 4)
    e <- 0
  else e <- exp(1.96 * 2/sqrt(sample.size - 3))
}
if (contraction == "No")
  rayonsign <- sqrt(2 - 2 * (e - 1)/(1 + e))
else rayonsign <- (1 - (e - 1)/(1 + e)) * 1.5
symbols(x = 0, y = 0, circles = rayonsign, inches = FALSE,
        add = TRUE, lwd = 1, fg = "red")
text(x = c(rep(0.01, 5)), y = radius + 0.04, labels = c("r =
0",
                                                         "r =
0.2", "r = 0.4", "r = 0.6", "r = 0.8"),
      cex = 0.5)
}
if (pvalues[1] != "No") {
  text(x = c(rep(0.01, 5)), y = c(0.563, 0.982, 1.239,
1.335, 1.48), labels = c("p
< 0.001", "p = 0.01",
                          "p
= 0.05", "p = 0.1", "p = 1"), cex = cx)
}

# Define colours of features (dots) with positive correlation
symbols(x = loadx[, j], y = loadyp[, j], circles = rep(0.01,
length(
loadyp[, j])), inches = FALSE, add = TRUE,
        fg = "#FF9966", bg = "#FF9966")

# Define colours of features (dots) with negative correlation
symbols(x = loadx[, j], y = loadym[, j], circles = rep(0.01,
length(
loadym[, j])), inches = FALSE, add = TRUE,
        fg = "grey", bg = "grey")

text(x = -0.18, y = -0.12, labels = namey, cex = cx +
0.25)

# Recover the result of F-PCA analysis from the graphical reps
entation
# Features with positive correlation
# (x,y) coordinates
xcor<-loadx[, j]
ycor<-loadyp[, j]
# Calculate the distance of these features to the center o
f the circle correponding to the point of biological activity (dist
1)
dist<-sqrt((xcor^2) + (ycor^2))

```

```

    # Features with negative correlation
    # (x,y) coordinates
    xcor<-loadx[, j]
    y2cor<-loadym[, j]
    # Calculate the distance of these features to the center o
f the circle correponding to the point of biological activity (dist
2)
    dist2<-sqrt((xcor^2) + (y2cor^2))

    # Features with significant positive correlation: F-PCA scor
e < rayonsign
    # (x,y) coordinates
    loadyi<-matrix(nrow = p + 2, ncol = p - 1)
    loadxi<-matrix(nrow = p + 2, ncol = p -1)
    loadyi[which(dist <= rayonsign )] <-loadyp[, j][which(dist <=
rayonsign )]
    loadxi[which(dist <= rayonsign )] <-loadx[, j][which(dist <=
rayonsign )]
    # Define colours of features
    symbols(x = loadxi[, j], y = loadyi[, j], circles = rep(0.01,
length
(loadyi[, j])), inches = FALSE, add = TRUE,
        fg = "blue", bg = "blue")
    # Add the features labels
    text(x = loadxi[, j], y = loadyi[, j], labels = names,
        cex = cx)

    # Recover and merge the names, distance 1 and distance 2 and rayon
sign for each feature
    samples<-names
    fpca<-data.frame(samples,dist,dist2,rayonsign)
}
}

# Apply the modified F-PCA function
# The function allows to obtain automatically a graphical representa
tion
    # F-PCA plot appears with the four biochemometrics models
    dev.set(4)
    op<-par(pty="s",mar=c(0,0,3,0),mai=c(0,0,0.5,0))
    fpca<-fpca_lucie(data=normdata,y=Y,x=c(colnames(normdata[,2:(feat.
num.tot+1)])),pvalues="No",partial="Yes",cx=0.001, contraction="No",
input="data",sample.size=0.0)

    # F-PCA plot appears in a single window
    # Open window
    if (OpSys == "mac"){
    quartz(title="F-PCA")

```

```

op<-par(pty="s",mar=c(0,0,1,0))
} else {
X11(title="F-PCA")
}
# Apply the modified F-PCA function
fpca<-fpca_lucie(data=normdata,y=Y,x=c(colnames(normdata[,2:(feat.
num.tot+1)])),pvalues="No",partial="Yes",cx=0.5, contraction="No",in
put="data",sample.size=0.1)

# Save the F-PCA Loadings plot and close the window
dev.copy2pdf(file="F_PCA_loadings_plot.pdf",device=dev.cur())
dev.off()

# Create a table of F-PCA results
# Remove the biological activity dot (central point)
fpca<-subset(fpca,fpca$samples != ".")
# Define the features labels on the table
rownames(fpca)<-fpca$samples
# Recover the rayonsign, the positives and negatives F-PCA sco
res
rayonsign<-1/fpca$rayonsign[1]
fpca$dist<-(1/fpca$dist)
fpca$dist2<-(1/fpca$dist2)
# Recover the rayonsign, the positives and negatives F-PCA sco
res
fpca$dist[which(is.na(fpca$dist))]<-fpca$dist2[which(is.na(fpc
a$dist))]
sign<-ifelse(is.na(fpca$dist2),"positive","negative")
fpca<-cbind(fpca,sign)
fpca$dist[which(fpca$sign=="negative")]<-(-fpca$dist[which(fpca$si
gn=="negative")])

# Define importance level of each feature
Level<-ifelse(fpca$dist>=rayonsign & fpca$dist>0 , "High",
ifelse(fpca$dist<=0, "Low", "Medium"))

# Merge information
fpca<-cbind(fpca,Level)
fpca<-subset(fpca,select=c(dist,Level))
fpca<-fpca[order(row.names(fpca)),]

# Rename the columns of the table
fpca<-data.frame(fpca)
colnames(fpca)<-c("F-PCA_score", "Importance Levels")

# Round the values
fpca[, "F-PCA_score"]<-round(fpca[, "F-PCA_score"],2))

# Export table in a working directory

```

```
write.table(cbind(features=row.names(fpca), fpca), file="F-PCA.csv",
sep=";", dec=".", row.names = F)
```

4.3 Partial Least Square analysis (PLS)

Several results are created in working directory:

- Table (PLS.csv)
 - Selectivity ratio: “SR”
 - Regression Coefficient: “RC”
 - Very important projection: “VIP”
 - Distances features from the inhibition point: “dist1” in XYloading plot
 - Distances features from the regression line: “dist2” in XYloading plot
 - An information of Importance Levels according to selectivity ratio and the sign of regression coefficient was reported:
SR >=1 & Coef >=1 → “High”, Coef < 0 → “Low”, Others → “Medium”
- Plots (.pdf)
 - PLS summary plot (PLS_summary.pdf)
 - XY Loadings plot with important features appearing in blue colour and biological activity point in red (PLS_loadings_plot.pdf)
 - Score plot with samples coloured according to groups defined by biological activity (PLS_Score_plot.pdf)
 - VIP Score plot (PLS_VIP_plot.pdf)
 - Regression coefficient plot (PLS_RC_plot.pdf)
 - Selectivity Ratio plot (PLS_SR_plot.pdf)

```
# Create "PLS" working directory
dir.create(file.path(WD, "PLS"))
setwd(file.path(WD, "PLS"))

# Apply PLS function
model<-pls(x=normdata[,2:(feat.num.tot+1)], y=(normdata[,grep(Y,colnames(normdata))]), ncomp = 15, center = T, scale = F, cv = 1,
          method = "simpls", alpha = 0.05, coeffs.ci = NULL, coeffs.alpha = 0.1,
          info = "", light = F, ncomp.selcrit = "min")

# Recover the optimal number of component for the model based on results obtain from summary(model) and plot(model) **
selectCompNum(model, ncomp = NULL)
summary(model)

if (OpSys == "mac"){
  quartz()
}else{
  X11()
}
# Save the plot
```



```

dev.copy2pdf(file="PLS_summary.pdf",device=dev.cur())

# Close the graphical window
dev.off()

# Recover statistical results:

# Calculate distances: "dist1" and "dist2"

# xyloadings
featuresload<-as.data.frame(model$xloadings) #Coordinates of Features
Inhload<-as.data.frame(model$yloadings) #Coordinates of Inhibition point
xyloadings<-rbind(featuresload,Inhload) # Merge tables

# Calculate distances features from the inhibition point "dist1"
# using pythagore formula
# Pay attention to the component choosed at the point 1 for this model
#  $a^2 + b^2 = c^2 \rightarrow c = \text{square root } (a^2 + b^2)$ 
# Pay attention to components choosed

dist1<-abs(sqrt((abs((xyloadings[nrow(xyloadings),CompX]-xyloadings[-nrow(xyloadings),CompX]))^2 + (abs((xyloadings[nrow(xyloadings),CompY]-xyloadings[-nrow(xyloadings),CompY]))^2)))

# Calculate distances features from the regression line "dist2"

# Linear equation:  $y = ax + b$ 
# Point distance from regression line:
#  $dist2 = [(yfeature - a * xfeature) - b] / \text{square root } (1/ a^2)$ 

a = xyloadings[nrow(xyloadings),CompY]/xyloadings[nrow(xyloadings),CompX]
dist2 = abs((xyloadings[-nrow(xyloadings),CompY]-a*xyloadings[-nrow(xyloadings),CompX])/(sqrt(1+(xyloadings[nrow(xyloadings),CompY]/xyloadings[nrow(xyloadings),CompX])^2)))

# Very Important Projection (VIP)
VIP<-getVIPScores(model)
VIP<-data.frame(VIP)
VIP[,1]<-(format(VIP[,1],scientific=F,digits=10))

# Regression Coefficient (RC)
RC<-getRegcoeffs(model)
RC<-RC[-1,]
RC<-data.frame(RC)

```

```

# Selectivity Ratio (SR)
SR<-getSelectivityRatio(model)
SR<-data.frame(SR)

# Merge Statistical results in a table:
PLS<-cbind(dist1,dist2,VIP,RC,SR)
colnames(PLS)<-c("dist1","dist2","VIP","RC","SR")

# Define Importance Levels of each features according to the Selectivity Ratio (SR) and regression coefficient (RC)
# Classification of Importance Level
Level<-ifelse(PLS[,"SR"] >=1 & PLS[,"RC"] >=0,"High",
             ifelse(PLS[,"RC"]<0,"Low","Medium"))
# Add this annotation for each fetures in the tables
PLS<-cbind(PLS,Level)
colnames(PLS)<-c("dist1","dist2","VIP","RC","SR","Importance Levels")
# Round the values
PLS1<-as.matrix(PLS)
colnames(PLS1)<-c("dist1","dist2","VIP","RC","SR","Importance Levels")
PLS1[,1:2]<-round(as.numeric(PLS1[,1:2]),2)
PLS1[,3:5]<-format(as.numeric(PLS1[,3:5]),scientific=TRUE,digits=2)

# Export the PLS table
write.table(cbind(features=row.names(PLS1),PLS1),file="PLS.csv",
sep=";",dec=".",row.names = F)

# Graphical representation

# PLS Loadings plot
deltax=0.5*(max(dist1)-min(dist1))
deltay=0.5*(max(dist2)-min(dist2))
group<-ifelse(dist1 < deltax & dist2 < deltax,"good",
             ifelse( dist1 & dist2,""))

loadings<-PLS[,-ncol(PLS)]
loadings<-cbind(loadings,group)
loadings<-data.frame(loadings)

if (OpSys == "mac"){
  quartz()
} else {
  X11()
}

```

```

    PLSplot<-plot(xyloadings[-nrow(xyloadings),CompX],xyloadings[-
nrow(xyloadings),CompY],pch=19,cex=0.6,main="PLS Loadings", xlab=xla
b, ylab=ylab,col=c("grey","blue")[loadings$group])
    points(xyloadings[nrow(xyloadings),CompX],xyloadings[nrow(xylo
adings),CompY],col="red",pch=19,cex=0.9)

    abline(h=0,v=0,col="grey",lty=3,lwd=1)

    # Add the Legend
    legend("topright",inset=.05, title, c("Y variable")
        ,fill = NULL,pch=19,col="red",border="white",cex=0.5)
    # Save the plot
    dev.copy2pdf(file="PLS_loadings.pdf",device=dev.cur())

    # Close the graphical window
    dev.off()

    # PLS Score plot
    PLSXscorecomp<-data.frame(model$calres$xdecomp$expvar)
    xlab<-paste(xlab,"(",round(PLSXscorecomp[CompX,],2),"%",")",se
p=' ')
    ylab<-paste(ylab,"(",round(PLSXscorecomp[CompY,],2),"%",")",se
p=' ')
    PLSXscore<-cbind(model$calres$xdecomp$scores,mSet$dataSet$cIs)
    PLSXscore<-data.frame(PLSXscore)
    names(PLSXscore)[ncol(PLSXscore)]<-"Group"

    if (OpSys == "mac"){
        quartz()
    } else {
        X11()
    }
    PLSplot2<-plot(PLSXscore[,CompX],PLSXscore[,CompY],pch=15,cex=
0.8,
                    main="PLS Xscores",
                    xlab=xlab, ylab=ylab,
                    col=c(colour1,colour2,colour3,colour4,colour5,c
olour6)[as.factor(PLSXscore$Group)])

    text(PLSXscore[,CompX],PLSXscore[,CompY], label = rownames(PLS
Xscore), pos = 1, col = "black",
        xpd = T, cex = 0.8,offset = 0.5)
    # Legend
    legend("topleft",inset=.05, title, c(aa, bb,cc)
        ,fill = c(colour1,colour2,colour3,colour4,colour5,colou
r6), cex=0.5,title= title)
    abline(h=0,v=0,col="grey",lty=3,lwd=1)
    # Save plot
    dev.copy2pdf(file="PLS_Score plot.pdf",device=dev.cur())
    # Close the graphical window

```

```

dev.off()

# PLS VIP plot
if (OpSys == "mac"){
  quartz()
} else {
  X11()
}
plotVIPScores(model,xlab = "Features m/z_RT",col.sub="grey",font.sub=4,cex.sub=0.75)
dev.copy2pdf(file="PLS_VIP_plot.pdf",device=dev.cur())
dev.off()

# PLS RC plot
if (OpSys == "mac"){
  quartz()
} else {
  X11()
}
plotRegcoeffs(model,xlab = "Features",col.sub="grey",font.sub=4,cex.sub=0.75)
dev.copy2pdf(file="PLS_RC_plot.pdf",device=dev.cur())
dev.off()

# PLS plot with Selectivity ratio multiplied by sign value of regression coefficient (SR * sign value of RC)
# Determine the value
selratio<-format(as.numeric(PLS[,"SR"])*as.numeric(PLS[,"RC"])/abs(as.numeric(PLS[,"RC"])),scientific=F,digits=10)
select.ratio<-data.frame(format(as.numeric(PLS[,"SR"])*as.numeric(PLS[,"RC"])/abs(as.numeric(PLS[,"RC"])),scientific=F,digits=10),PLS[,6])
row.names(select.ratio)<-row.names(PLS)
colnames(select.ratio)<-c("SR*RC", "Importance Levels")

# PLS plot appears in the graphical window with the four biochemometrics models
dev.set(4)
op<-par(pty="m",mar=c(5,4,4,4))
# Colours code according to importance levels
colour = rep(NA, length=length(select.ratio[,2]))
colour[which(select.ratio[,2]=="High")] = "blue"
colour[which(select.ratio[,2]=="Medium")] = "#FF9966"
colour[which(select.ratio[,2]=="Low")] = "grey"

# Plot function
plot(x=as.numeric(substr(row.names(select.ratio),regex(pattern = "_",text=row.names(select.ratio),fixed=TRUE)+1,nchar(row.names(select.ratio))))),y=selratio,pch=19,type='h',las=1,cex=0.6, main="PLS

```

```

",
      xlab = "Retention time (min)",col.sub="grey",font.sub=4,c
ex.sub=0.75,
      ylab = "Selectivity ratio (SR) ",
      axes=T,col=colour)
  # Legend
  legend("bottomright",inset =0, title="Importance Levels", c("H
igh", "Medium" , "Low")
      ,fill = c("blue","#FF9966", "grey"), cex=0.5)

  legend("bottomleft",inset =.05, "RC<0",box.lty=0,cex=0.3)
  legend("topleft",inset =.05, "RC>0",box.lty=0, cex=0.3)

  # TreshoLd (SR*RC >= 1)
  abline(h=1,col="red",lwd=0.7,lty=3)

  if (OpSys == "mac"){
    quartz(title="PLS")
  } else {
    X11(title="PLS")
  }

  # PLS plot
  plot(x=as.numeric(substr(row.names(select.ratio),regexr(patte
rn = "_",text=row.names(select.ratio),fixed=TRUE)+1,nchar(row.names(
select.ratio))))),y=selratio,pch=19,type='h',las=1,cex=0.6, main="PLS
",
      xlab = "Retention time (min)",col.sub="grey",font.sub=4,c
ex.sub=0.75,
      ylab = "Selectivity ratio (SR) ",
      axes=T,col=colour)
  # Legend
  legend("bottomright",inset =0, title="Importance Levels", c("H
igh ", "Medium" , "Low")
      ,fill = c("blue","#FF9966", "grey"), cex=0.5)
  legend("bottomleft",inset =.05, "RC<0",box.lty=0,cex=0.4,horiz=
FALSE)
  legend("topleft",inset =.05, "RC>0",box.lty=0,cex=0.4,horiz=FAL
SE)

  # Treshold (SR*RC >= 1)
  abline(h=1,col="red",lwd=0.7,lty=3)

  # Labels of highlited features (features with "high" importanc
e levels)
  text(y=as.numeric(select.ratio[which(select.ratio[,2]=="High")
,2]),x= as.numeric(substr(row.names(select.ratio),regexr(pattern =
"_",text=row.names(select.ratio),fixed=TRUE)+1,nchar(row.names(selec
t.ratio))))[which(select.ratio[,2]=="High")], label = row.names(sele

```

```

ct.ratio)[which(select.ratio[,2]=="High")], pos = 1, col = "black",
      xpd = T, cex = 0.5,offset = -1.3,srt=75)
# Save plot
dev.copy2pdf(file="PLS_score.pdf",device=dev.cur())
# Close the graphical window
dev.off()

```

4.4 Partial Least Square Discriminant analysis (PLS-DA)

Several results are created in working directory:

- Tables (.csv)
 - PLS-DA classification (CV)
 - PLS-DA scores and PLS-DA loadings
 - PLS-DA with:
 - Very Important Projection (VIP)
 - Weighted sum of absolute Regression Coefficients (RC)
 - Relative concentrations of the corresponding metabolite in each group (C)
 - An information of Importance Levels according to VIP and C
- VIP ≥ 1 & C(x) > C(x-1) → "High", VIP < 1 or C(x) < C(x-1) → "Low", Others → "Medium"
- Plots (.pdf)
 - PLS-DA_summary
 - PLS-DA classification and feature selection
 - XY Loadings
 - XY Score
 - VIP Score plot
 - Coef Score plot

```

# Create "PLS-DA" working directory
dir.create(file.path(WD, "PLS-DA"))
setwd(file.path(WD, "PLS-DA"))

# Apply PLS-DA function
mSet$dataSet$norm <- mSet$dataSet$norm[, 2:ncol(mSet$dataSet$norm)]
mSet <- PLSR.Anal(mSet, reg=TRUE)

detach("package:MetaboAnalystR", unload=TRUE)
detach("package:caret", unload=TRUE)

require(MetaboAnalystR)
require(caret)

# PLS-DA cross validation
mSet <- PLSDA.CV(mSet, methodName=methodName, compNumber, choice=choice
)

# Recover and export fit.info (R2 and Q2)

```

```

fit.info<-data.frame(mSet$analSet$plsda$fit.info)
write.table(cbind(parameters=row.names(fit.info),fit.info),file="PLS-DA_CV.csv",sep=";",dec=".",row.names=F)

# Plot and export PLS-DA_summary and PLS-DA classification
mSet<-PlotPLS.Classification(mSet, "PLS-DA_CV", "pdf", width=NA)
mSet<-PlotPLSPairSummary(mSet, "PLS-DA_Pairsummary","pdf",72, width=NA, 5)

# Recover Important Statistical results and merge results in a table:
# - Very Important Projection score: "PLS-DA_VIP"
# - Weighted sum of absolute Regression Coefficients: "PLS-DA_RC"
# - Relative concentrations of the corresponding metabolite in each group: "C"
data=mSet$analSet$plsda$vip.mat[,feat.nm]
imp.vec<-data
imp.vec <- rev(sort(imp.vec))[1:(feat.num.tot+1)]
imp.vec <- sort(imp.vec)
mns <- by(mSet$dataSet$norm[,names(imp.vec)], mSet$dataSet$cls,
function(x) {
  apply(x, 2, mean, trim = 0.1)
})
mns <- t(matrix(unlist(mns), ncol = feat.num.tot, byrow = TRUE))
row.names(mns)<-names(imp.vec)
mns<-mns[order(row.names(mns)),]
data<-data.frame(mSet$analSet$plsda$vip.mat[1:feat.num.tot, feat.nm])

# Very Important Projection score (VIP)
PLS_DA_VIP<-merge(data,mns,by="row.names")
row.names(PLS_DA_VIP)<-PLS_DA_VIP[,1]
PLS_DA_VIP<-PLS_DA_VIP[, -1]

# Weighted sum of absolute Regression Coefficients (RC) and Relative concentrations of the corresponding features in each group: "C"
if (mSet$dataSet$cls.num==2){
  PLS_DA_RC<-merge(mSet$analSet$plsda$coef.mat[, "Overall"],mns,by="row.names")
}else {PLS_DA_RC<-merge(mSet$analSet$plsda$coef.mat[, "coef.mean"],mns,by="row.names")
}
row.names(PLS_DA_RC)<-PLS_DA_RC[,1]
PLS_DA_RC<-PLS_DA_RC[, -1]

# Merge VIP and RC
if (length(mSet$analSet$plsda$coef.mat[,1])<1){
  PLS_DA<-PLS_DA_VIP
}else{
  PLS_DA<-cbind(PLS_DA_VIP,PLS_DA_RC[, "x"])
}

```

```

}

# Define Importance Levels of each features according to the "VIP" and "C"
if (mSet$dataSet$cls.num == 2 & length(mSet$analSet$plsda$coef.mat) < 1) {
  colnames(PLS_DA) <- c("VIP", "C1", "C2")
  Level <- ifelse(PLS_DA[, "VIP"] >= 1 & PLS_DA[, "C2"] > PLS_DA[, "C1"], "High",
    ifelse(PLS_DA[, "VIP"] < 0.5 & PLS_DA[, "VIP"] < 1
      & PLS_DA[, "C2"] > PLS_DA[, "C1"], "Medium", "Low"))
} else if (mSet$dataSet$cls.num == 3 & length(mSet$analSet$plsda$coef.mat) < 1) {
  colnames(PLS_DA) <- c("VIP", "C1", "C2", "C3")
  Level <- ifelse(PLS_DA[, "VIP"] >= 1 & PLS_DA[, "C3"] > PLS_DA[, "C2"] & PLS_DA[, "C3"] > PLS_DA[, "C1"], "High",
    ifelse(PLS_DA[, "VIP"] < 0.5 & PLS_DA[, "VIP"] < 1
      & PLS_DA[, "C3"] > PLS_DA[, "C2"] & PLS_DA[, "C3"] > PLS_DA[, "C1"], "Medium", "Low"))
} else if (mSet$dataSet$cls.num == 4 & length(mSet$analSet$plsda$coef.mat) < 1) {
  colnames(PLS_DA) <- c("VIP", "C1", "C2", "C3", "C4")
  Level <- ifelse(PLS_DA[, "VIP"] >= 1 & PLS_DA[, "C4"] > PLS_DA[, "C3"] & PLS_DA[, "C4"] > PLS_DA[, "C2"] & PLS_DA[, "C4"] > PLS_DA[, "C1"], "High",
    ifelse(PLS_DA[, "VIP"] < 0.5 & PLS_DA[, "VIP"] < 1
      & PLS_DA[, "C4"] > PLS_DA[, "C3"] & PLS_DA[, "C4"] > PLS_DA[, "C2"] & PLS_DA[, "C4"] > PLS_DA[, "C1"], "Medium", "Low"))
} else if (mSet$dataSet$cls.num == 5 & length(mSet$analSet$plsda$coef.mat) < 1) {
  colnames(PLS_DA) <- c("VIP", "C1", "C2", "C3", "C4", "C5")
  Level <- ifelse(PLS_DA[, "VIP"] >= 1 & PLS_DA[, "C5"] > PLS_DA[, "C4"] & PLS_DA[, "C5"] > PLS_DA[, "C3"] & PLS_DA[, "C5"] > PLS_DA[, "C2"] & PLS_DA[, "C5"] > PLS_DA[, "C1"], "High",
    ifelse(PLS_DA[, "VIP"] < 0.5 & PLS_DA[, "VIP"] < 1
      & PLS_DA[, "C5"] > PLS_DA[, "C4"] & PLS_DA[, "C5"] > PLS_DA[, "C3"] & PLS_DA[, "C5"] > PLS_DA[, "C2"] & PLS_DA[, "C5"] > PLS_DA[, "C1"], "Medium", "Low"))
} else if (mSet$dataSet$cls.num == 6 & length(mSet$analSet$plsda$coef.mat) < 1) {
  colnames(PLS_DA) <- c("VIP", "C1", "C2", "C3", "C4", "C5", "C6")
  Level <- ifelse(PLS_DA[, "VIP"] >= 1 & PLS_DA[, "C6"] > PLS_DA[, "C5"] & PLS_DA[, "C6"] > PLS_DA[, "C4"] & PLS_DA[, "C6"] > PLS_DA[, "C3"] & PLS_DA[, "C6"] > PLS_DA[, "C2"] & PLS_DA[, "C6"] > PLS_DA[, "C1"], "High",
    ifelse(PLS_DA[, "VIP"] >= 0.5 & PLS_DA[, "VIP"] < 1
      & PLS_DA[, "C6"] > PLS_DA[, "C5"] & PLS_DA[, "C6"] > PLS_DA[, "C4"] & PLS_DA[, "C6"] > PLS_DA[, "C3"] & PLS_DA[, "C6"] > PLS_DA[, "C2"] & PLS_DA[, "C6"] > PLS_DA[, "C1"], "Medium", "Low"))
} else if (mSet$dataSet$cls.num == 2 & length(mSet$analSet$plsda$

```



```

coef.mat)>1) {
  colnames(PLS_DA)<-c("VIP", "C1", "C2", "Coef.mean")
  Level<-ifelse(PLS_DA[, "VIP"] >=1 & PLS_DA[, "C2"]>PLS_DA[, "C1"]
, "High",
               ifelse(PLS_DA[, "VIP"]<0.5 & PLS_DA[, "VIP"]<1
& PLS_DA[, "C2"]>PLS_DA[, "C1"], "Medium", "L
ow"))
}else if (mSet$dataSet$cls.num == 3 & length(mSet$analSet$plsda$
coef.mat)>1) {
  colnames(PLS_DA)<-c("VIP", "C1", "C2", "C3", "Coef.mean")
  Level<-ifelse(PLS_DA[, "VIP"] >=1 & PLS_DA[, "C3"]>PLS_DA[, "C2"]
&PLS_DA[, "C3"]>PLS_DA[, "C1"], "High",
               ifelse(PLS_DA[, "VIP"]<0.5 &PLS_DA[, "VIP"]<1
&PLS_DA[, "C3"]>PLS_DA[, "C2"]&PLS_DA[, "C3"]
]>PLS_DA[, "C1"], "Medium", "Low"))
}else if (mSet$dataSet$cls.num == 4 & length(mSet$analSet$plsda$
coef.mat)>1) {
  colnames(PLS_DA)<-c("VIP", "C1", "C2", "C3", "C4", "Coef.mean")
  Level<-ifelse(PLS_DA[, "VIP"] >=1 & PLS_DA[, "C4"]>PLS_DA[, "C3"]
&PLS_DA[, "C4"]>PLS_DA[, "C2"]&PLS_DA[, "C4"]>PLS_DA[, "C1"], "High",
               ifelse(PLS_DA[, "VIP"]<0.5 &PLS_DA[, "VIP"]<1
& PLS_DA[, "C4"]>PLS_DA[, "C3"]&PLS_DA[, "C4"]
]>PLS_DA[, "C2"]&PLS_DA[, "C4"]>PLS_DA[, "C1"], "Medium", "Low"))
}else if (mSet$dataSet$cls.num == 5 & length(mSet$analSet$plsda$
coef.mat)>1) {
  colnames(PLS_DA)<-c("VIP", "C1", "C2", "C3", "C4", "C5", "Coef.mean"
)
  Level<-ifelse(PLS_DA[, "VIP"] >=1 &PLS_DA[, "C5"]>PLS_DA[, "C4"]&
PLS_DA[, "C5"]>PLS_DA[, "C3"]&PLS_DA[, "C5"]>PLS_DA[, "C2"]&PLS_DA[, "C5"]
]>PLS_DA[, "C1"], "High",
               ifelse(PLS_DA[, "VIP"]<0.5 &PLS_DA[, "VIP"]<1
&PLS_DA[, "C5"]>PLS_DA[, "C4"]& PLS_DA[, "C5"]
]>PLS_DA[, "C3"]&PLS_DA[, "C5"]>PLS_DA[, "C2"]&PLS_DA[, "C5"]>PLS_DA[, "
C1"], "Medium", "Low"))
}else if (mSet$dataSet$cls.num == 6 & length(mSet$analSet$plsda$
coef.mat)>1) {
  colnames(PLS_DA)<-c("VIP", "C1", "C2", "C3", "C4", "C5", "C6", "Coef.
mean")
  Level<-ifelse(PLS_DA[, "VIP"] >=1 &PLS_DA[, "C6"]>PLS_DA[, "C5"]&
PLS_DA[, "C6"]>PLS_DA[, "C4"]& PLS_DA[, "C6"]>PLS_DA[, "C3"]&PLS_DA[, "C6"]
]>PLS_DA[, "C2"]&PLS_DA[, "C6"]>PLS_DA[, "C1"], "High",
               ifelse(PLS_DA[, "VIP"]>=0.5 &PLS_DA[, "VIP"]<1
&PLS_DA[, "C6"]>PLS_DA[, "C5"]&PLS_DA[, "C6"]
]>PLS_DA[, "C4"]& PLS_DA[, "C6"]>PLS_DA[, "C3"]&PLS_DA[, "C6"]>PLS_DA[, "
C2"]&PLS_DA[, "C6"]>PLS_DA[, "C1"], "Medium", "Low"))

}else {
  print("no group")}

# Merge

```

```

PLS_DA<-cbind(PLS_DA,Level)
PLS_DA[,-ncol(PLS_DA)]<-round(PLS_DA[,-ncol(PLS_DA)],2)
write.table(cbind(features=row.names(PLS_DA),PLS_DA),file="PLS-D
A.csv",sep=";",dec=".",row.names = F)

# View and export Graphs
# PLS-DA_Score plot
mSet<-PlotPLS2DScore(mSet, "PLS-DA_score", "pdf",72, width=NA, C
ompx,Compy,0.95,1,0,use.sparse = FALSE)

# PLS-DA_Loadings plot
mSet<-PlotPLSLoading(mSet, "PLS-DA_loading", "pdf",72, width=NA,
Compx,Compy,"scatter", 1)

# PLS-DA_Score plot
mSet<-PlotPLS2DScore(mSet, "PLS-DA_score", "pdf",72, width=NA, C
ompx,Compy,0.95,1,0,use.sparse = FALSE)

# PLS-DA_Loadings plot
mSet<-PlotPLSLoading(mSet, "PLS-DA_loading", "pdf",72, width=NA,
Compx,Compy,"scatter", 1)

# PLS-DA_VIP plot
{
  if (OpSys == "mac"){
    quartz()
  }
  else {
    X11()
  }
  # Define the number of features
  feat.num = 11
  # Recover VIP score of the feature
  data=mSet$analSet$plsda$vip.mat[, feat.nm]
  imp.vec<-data
  # Choose colored VIP plot
  color.BW = FALSE
  # Legend of x axis
  xlbl = "VIPS"
  cls.len <- length(levels(mSet$dataSet$cls))
  if (cls.len == 2) {
    rt.mrg <- 5
  }else if (cls.len == 3) {
    rt.mrg <- 6
  }else if (cls.len == 4) {
    rt.mrg <- 7
  }else if (cls.len == 5) {
    rt.mrg <- 8
  }else if (cls.len == 6) {

```

```

    rt.mrg <- 9
  }else {
    rt.mrg <- 11
  }
  op <- par(mar = c(5, 7, 3, rt.mrg))
  if (feat.num <= 0) {
    feat.num = 15
  }
  if (feat.num > length(imp.vec)) {
    feat.num <- length(imp.vec)
  }
  imp.vec <- rev(sort(imp.vec))[1:feat.num]
  imp.vec <- sort(imp.vec)
  mns <- by(mSet$dataSet$norm[, names(imp.vec)], mSet$dataSet$cl
s,
          function(x) {
            apply(x, 2, mean, trim = 0.1)
          })
  mns <- t(matrix(unlist(mns), ncol = feat.num, byrow = TRUE))
  vip.nms <- substr(names(imp.vec), 1, 14)
  row.names(mns) <- rev(vip.nms)
  names(imp.vec) <- NULL
  dotcolour <- ifelse(color.BW, "darkgrey", "black")
  dotchart(imp.vec, bg = dotcolour, xlab = xlbl, cex = 1.3)

  # Define colours of features labels according to their importa
nce Level
  # Create three groups
  vip.nms1 <- row.names(PLS_DA)[which(PLS_DA$Level=="High")]
  vip.nms2 <- row.names(PLS_DA)[which(PLS_DA$Level=="Medium")]
  vip.nms3 <- row.names(PLS_DA)[which(PLS_DA$Level=="Low")]
  # Define colours
  colour = rep(NA, length=length(vip.nms))
  colour[match(vip.nms1, row.names(mns))] = "blue"
  colour[match(vip.nms2, row.names(mns))] = "#FF9966"
  colour[match(vip.nms3, row.names(mns))] = "grey"
  # Labels
  mtext(side = 2, at = 1:feat.num, vip.nms, las = 2, line = 1, co
l = colour)
}
axis.lims <- par("usr")
shift <- 2 * par("cxy")[1]
lgd.x <- axis.lims[2] + shift
x <- rep(lgd.x, feat.num)
y <- 1:feat.num
par(xpd = T)
suppressMessages(library(RColorBrewer))
nc <- ncol(mns)

# Colour scale

```

```

whiteblack<-colorRampPalette(c("black","grey","white"))(20)
col <- colorRampPalette(whiteblack)(nc)
if (color.BW)
  col <-col
bg <- matrix("", nrow(mns), nc)
for (m in 1:nrow(mns)) {
  bg[m, ] <- (col[nc:1])[rank(mns[m, ])]
}
cls.lbl<-levels(mSet$dataSet$cls)

for (n in 1:ncol(mns)) {
  points(x, y, bty = "n", pch = 22, bg = bg[, n], cex = 3)
  text(x[1], axis.lims[4], cls.lbl[n], srt = 45, adj = c(0.2,
                                                       0.5))

  x <- x + shift/1.25
}
col <- colorRampPalette(whiteblack)(50)
if (color.BW)
  col <- rev(col)
nc <- length(col)
x <- rep(x[1] + shift, nc)
shifty <- (axis.lims[4] - axis.lims[3])/3
starty <- axis.lims[3] + shifty
endy <- axis.lims[3] + 2 * shifty
y <- seq(from = starty, to = endy, length = nc)
points(x, y, bty = "n", pch = 15, col = rev(col), cex = 2)
text(x[1], endy + shifty/8, "+")
text(x[1], starty - shifty/8, "-")
legend("topright",inset =-0.9, title, c("black","grey","white"
), cex=0.5)

# Save the plot
dev.copy2pdf(file="PLS-DA_VIP.pdf",device=dev.cur())

# Close the graphical window
dev.off()

# PLS-DA_Coef plot
{
  if (OpSys == "mac"){
    quartz()
  }
  else {
    X11()
  }

  if (length(mSet$analSet$plsda$coef.mat)<1) {
    dev.off()
  } else {

```

```

feat.num = 11
if (mSet$dataSet$cls.num==2){
  data=mSet$analSet$plsda$coef.mat[, "Overall"]}
else {data=mSet$analSet$plsda$coef.mat[, "coef.mean"]}
imp.vec<-data
color.BW = FALSE
xlbl = "Coef"
cls.len <- length(levels(mSet$dataSet$cls))
if (cls.len == 2) {
  rt.mrg <- 5
}else if (cls.len == 3) {
  rt.mrg <- 6
}else if (cls.len == 4) {
  rt.mrg <- 7
}else if (cls.len == 5) {
  rt.mrg <- 8
}else if (cls.len == 6) {
  rt.mrg <- 9
}else {
  rt.mrg <- 11
}
op <- par(mar = c(5, 7, 3, rt.mrg))
if (feat.num <= 0) {
  feat.num = 15
}
if (feat.num > length(imp.vec)) {
  feat.num <- length(imp.vec)
}
imp.vec <- rev(sort(imp.vec))[1:feat.num]
imp.vec <- sort(imp.vec)
mns <- by(mSet$dataSet$norm[, names(imp.vec)], mSet$dataSet$
cls,
      function(x) {
        apply(x, 2, mean, trim = 0.1)
      })
mns <- t(matrix(unlist(mns), ncol = feat.num, byrow = TRUE))
vip.nms <- substr(names(imp.vec), 1, 14)
names(imp.vec) <- NULL
dotcolour <- ifelse(color.BW, "darkgrey", "black")
dotchart(imp.vec, bg = dotcolour, xlab = xlbl, cex = 1.3)
mtext(side = 2, at = 1:feat.num, vip.nms, las = 2, line = 1)
axis.lims <- par("usr")
shift <- 2 * par("cxy")[1]
lgd.x <- axis.lims[2] + shift
x <- rep(lgd.x, feat.num)
y <- 1:feat.num
par(xpd = T)
suppressMessages(library(RColorBrewer))
nc <- ncol(mns)

```

```

whiteblack<-colorRampPalette(c("black","grey","white"))(50)
col <- colorRampPalette(whiteblack)(nc)
if (color.BW)
  col <-col
bg <- matrix("", nrow(mns), nc)
for (m in 1:nrow(mns)) {
  bg[m, ] <- (col[nc:1])[rank(mns[m, ])]
}
cls.lbl <- levels(mSet$dataSet$cls)
for (n in 1:ncol(mns)) {
  points(x, y, bty = "n", pch = 22, bg = bg[, n], cex = 3)
  text(x[1], axis.lims[4], cls.lbl[n], srt = 45, adj = c(0.2
,
0.5
))
  x <- x + shift/1.25
}
col <- colorRampPalette(whiteblack)(50)
nc <- length(col)
x <- rep(x[1] + shift, nc)
shifty <- (axis.lims[4] - axis.lims[3])/3
starty <- axis.lims[3] + shifty
endy <- axis.lims[3] + 2 * shifty
y <- seq(from = starty, to = endy, length = nc)
points(x, y, bty = "n", pch = 15, col = rev(col), cex = 2)
text(x[1], endy + shifty/8, "+")
text(x[1], starty - shifty/8, "-")

# Save plot
dev.copy2pdf(file="PLS-DA_Coef.pdf",device=dev.cur())

# Close the graphical window
dev.off()
}
}

```

4.5 Recover all the results

- Graphical representation with the four biochemometrics models

```

# Create and define the working directory "Summary-results"
dir.create(file.path(WD, "Summary-Results"))
setwd(file.path(WD, "Summary-Results"))

# Add the last plot (PLS-DA plot) in the graphical window to obtain
a graphical representation of the four biochemometrics models used
in the script

if (OpSys == "mac"){
  dev.set(4)
} else {

```

```

dev.set(4)
op <- par(mar = c(5, 7, 3, rt.mrg))
}
# Define the number of features for the graphical representation
feat.num = 11
# Recover the data for the features selected
imp.vec<-mSet$analSet$plsda$vip.mat[, feat.nm]
# Choose to obtain a plot with colour
color.BW = FALSE
# Define "x" axis
xlbl = "VIPS"
# Define the scale of "x" axis
cls.len <- length(levels(mSet$dataSet$cls))
if (cls.len == 2) {
  rt.mrg <- 1
}else if (cls.len == 3) {
  rt.mrg <- 2
}else if (cls.len == 4) {
  rt.mrg <- 3
}else if (cls.len == 5) {
  rt.mrg <- 4
}else if (cls.len == 6) {
  rt.mrg <- 5
}else {
  rt.mrg <- 11
}
op <- par(mar = c(5, 5, 3, rt.mrg),pty="s")
if (feat.num <= 0) {
  feat.num = 15
}
if (feat.num > length(imp.vec)) {
  feat.num <- length(imp.vec)
}
# Ordered data to obtain the most important features at the top of
the graph
imp.vec <- rev(sort(imp.vec))[1:feat.num]
imp.vec <- sort(imp.vec)
mns <- by(mSet$dataSet$norm[, names(imp.vec)], mSet$dataSet$cls,
function(x) {
  apply(x, 2, mean, trim = 0.1)
})
mns <- t(matrix(unlist(mns), ncol = feat.num, byrow = TRUE))
vip.nms <- substr(names(imp.vec), 1, 14)
names(imp.vec) <- NULL
row.names(mns)<-rev(vip.nms)
# Define dot colour
dotcolour <- ifelse(color.BW, "darkgrey", "black")
dotchart(imp.vec, bg = dotcolour, xlab = xlbl, cex = 0.8, main="PL
S-DA",cex.main=1.2,cex.lab=0.8,cex.axis=0.8)

```

```

# Recover the names of features in the PLS-DA table according to their importance levels in three groups
vip.nms1<-row.names(PLS_DA)[which(PLS_DA$Level=="High")]
vip.nms2<-row.names(PLS_DA)[which(PLS_DA$Level=="Medium")]
vip.nms3<-row.names(PLS_DA)[which(PLS_DA$Level=="Low")]

# Define colours according to the importance levels of the three groups
colour = rep(NA, length=length(vip.nms))
colour[match(vip.nms1,row.names(mns))] = "blue"
colour[match(vip.nms2,row.names(mns))] = "#FF9966"
colour[match(vip.nms3,row.names(mns))] = "grey"

# Parameters for the Labels
mtext(side = 2, at = 1:feat.num, vip.nms, las = 2, line = 1,col = colour,cex=0.8)

axis.lims <- par("usr")
shift <- 1 * par("cxy")[1]
lgd.x <- axis.lims[2] + shift
x <- rep(lgd.x, feat.num)
y <- 1:feat.num
par(xpd = T)
suppressMessages(library(RColorBrewer))
nc <- ncol(mns)

# Define the colour for the graph
whiteblack<-colorRampPalette(c("black","grey","white"))(50)
col <- colorRampPalette(whiteblack)(nc)

# A matrix is created to defined a colour for each peak area according to
bg <- matrix("", nrow(mns), nc)
for (m in 1:nrow(mns)) {
  bg[m, ] <- (col[nc:1])[rank(mns[m, ])]
}

# The number of groups according to biological activity
cls.lbl <- levels(mSet$dataSet$cls)
for (n in 1:ncol(mns)) {
  points(x, y, bty = "n", pch = 22, bg = bg[, n], cex = 1.5)
  text(x[1], 11.5, cls.lbl[n], srt = 45, adj = c(0.2, 0.5),cex=0.6)

  x <- x + shift/1
}
# Legend
col <- colorRampPalette(whiteblack)(50)
if (color.BW)
  col <- rev(col)

```



```

nc <- length(col)
x <- rep(x[1] + shift, nc)
shifty <- (axis.lims[4] - axis.lims[3])/3
starty <- axis.lims[3] + shifty
endy <- axis.lims[3] + 2 * shifty
y <- seq(from = starty, to = endy, length = nc)

points(x, y, bty = "n", pch = 15, col = rev(col), cex = 2)
text(x[1], endy + shifty/8, "+", cex=0.6)
text(x[1], starty - shifty/8, "-", cex=0.6)

# Save the plot
dev.copy2pdf(file="Four_models.pdf", device=dev.cur())

```

- Merge all statistical results in a table (Data_result.csv)

```

# Order all data tables by row names to merge results
r <- r[order(row.names(r)),]
fpca <- fpca[order(row.names(fpca)),]
PLS <- PLS[order(row.names(PLS)),]
PLS_DA <- PLS_DA[order(row.names(PLS_DA)),]

# Summary of all statistical models
Data_result <- cbind(r[, -ncol(r)],
                    fpca[, 1],
                    PLS[, -ncol(PLS)], PLS_DA[, -ncol(PLS_DA)])

if (length(mSet$analSet$plsda$coef.mat) > 1 & mSet$dataSet$cls.num
== 2) {
  colnames(Data_result) <- c("Spearman_r",
                            "F-PCA_score",
                            "PLS_dist1", "PLS_dist2", "PLS_VIP", "PL
S_RC", "PLS_SR",
                            "PLS-DA_VIP", "PLS-DA_C1", "PLS-DA_C2",
                            "PLS-DA_Coef.mean")
} else if (length(mSet$analSet$plsda$coef.mat) > 1 & mSet$dataSet$cls
ls.num == 3) {
  colnames(Data_result) <- c("Spearman_r",
                            "F-PCA_score",
                            "PLS_dist1", "PLS_dist2", "PLS_VIP", "PL
S_RC", "PLS_SR",
                            "PLS-DA_VIP", "PLS-DA_C1", "PLS-DA_C2",
                            "PLS-DA_C3", "PLS-DA_Coef.mean")
} else if (length(mSet$analSet$plsda$coef.mat) > 1 & mSet$dataSet$cls
ls.num == 4) {
  colnames(Data_result) <- c("Spearman_r",
                            "F-PCA_score",
                            "PLS_dist1", "PLS_dist2", "PLS_VIP", "PL

```

```

S_RC", "PLS_SR",
                                "PLS-DA_VIP", "PLS-DA_C1", "PLS-DA_C2",
"PLS-DA_C3", "PLS-DA_C4", "PLS-DA_Coef.mean")

    }else if (length(mSet$analSet$plsda$coef.mat)>1 & mSet$dataSet$cls.num == 5) {
        colnames(Data_result)<-c("Spearman_r",
                                "F-PCA_score",
                                "PLS_dist1", "PLS_dist2", "PLS_VIP", "PL
S_RC", "PLS_SR",
                                "PLS-DA_VIP", "PLS-DA_C1", "PLS-DA_C2",
"PLS-DA_C3", "PLS-DA_C4", "PLS-DA_C5", "PLS-DA_Coef.mean")
    }else if (length(mSet$analSet$plsda$coef.mat)>1 & mSet$dataSet$cls.num == 6) {
        colnames(Data_result)<-c("Spearman_r",
                                "F-PCA_score",
                                "PLS_dist1", "PLS_dist2", "PLS_VIP", "PL
S_RC", "PLS_SR",
                                "PLS-DA_VIP", "PLS-DA_C1", "PLS-DA_C2",
"PLS-DA_C3", "PLS-DA_C4", "PLS-DA_C5", "PLS-DA_C6", "PLS-DA_Coef.mean")
    }else if (mSet$dataSet$cls.num == 2 & length(mSet$analSet$plsda$coef.mat)<1) {
        colnames(Data_result)<-c("Spearman_r",
                                "F-PCA_score",
                                "PLS_dist1", "PLS_dist2", "PLS_VIP", "PL
S_RC", "PLS_SR",
                                "PLS-DA_VIP", "PLS-DA_C1", "PLS-DA_C2",
"PLS-DA_Coef.mean")

    }else if (mSet$dataSet$cls.num == 3 & length(mSet$analSet$plsda$coef.mat)<1) {
        colnames(Data_result)<-c("Spearman_r",
                                "F-PCA_score",
                                "PLS_dist1", "PLS_dist2", "PLS_VIP", "PL
S_RC", "PLS_SR",
                                "PLS-DA_VIP", "PLS-DA_C1", "PLS-DA_C2",
"PLS-DA_C3")

    }else if (mSet$dataSet$cls.num == 4 & length(mSet$analSet$plsda$coef.mat)<1) {
        colnames(Data_result)<-c("Spearman_r",
                                "F-PCA_score",
                                "PLS_dist1", "PLS_dist2", "PLS_VIP", "PL
S_RC", "PLS_SR",
                                "PLS-DA_VIP", "PLS-DA_C1", "PLS-DA_C2",
"PLS-DA_C3", "PLS-DA_C4")

    }else if (mSet$dataSet$cls.num == 5 & length(mSet$analSet$plsda$coef.mat)<1) {
        colnames(Data_result)<-c("Spearman_r",

```

```

        "F-PCA_score",
        "PLS_dist1", "PLS_dist2", "PLS_VIP", "PL
S_RC", "PLS_SR",
        "PLS-DA_VIP", "PLS-DA_C1", "PLS-DA_C2",
"PLS-DA_C3", "PLS-DA_C4", "PLS-DA_C5")
    }else if (mSet$dataSet$cls.num == 6 & length(mSet$analSet$plsda$
coef.mat)<1) {
        colnames(Data_result)<-c("Spearman_r",
        "F-PCA_score",
        "PLS_dist1", "PLS_dist2", "PLS_VIP", "PL
S_RC", "PLS_SR",
        "PLS-DA_VIP", "PLS-DA_C1", "PLS-DA_C2",
"PLS-DA_C3", "PLS-DA_C4", "PLS-DA_C5", "PLS-DA_C6")

    }else {
        print("no group")
    }

# Export table
Data_result1<-as.matrix(Data_result)
Data_result1[,3:7]<-format(as.numeric(Data_result1[,3:7]),scientific
ic = TRUE,digits = 3)
write.table(cbind(features=row.names(Data_result1),Data_result1),f
ile="Data_result.csv",sep=";",dec=".",row.names=F)

```

5. Combination of the result obtained for the four models

5.1 Recover the most important features

Create a table with the most important features according to:

- The statistical parameters :
 - Spearman: $r \geq 0.5$
 - fpca: f_PCA_score with the most important significant positive correlation
 - PLS: selectivity ratio (SR) ≥ 1 and regression coefficient (RC) > 0
 - PLS-DA: VIP ≥ 1 and coefficient for bioactive groups higher than coefficient for the other groups
- The maximal number of interested features selected in each model

```

# A. Define the columns names
if (mSet$dataSet$cls.num == 2 & length(mSet$analSet$plsda$coef.m
at)<1) {
    colnamesdr<-c("Spearman_r", "F-PCA_score",
        "PLS_dist1", "PLS_dist2", "PLS_VIP", "PLS_RC", "PLS_
SR",
        "PLS-DA_VIP", "PLS-DA_C1", "PLS-DA_C2")
    }else if (mSet$dataSet$cls.num == 3 & length(mSet$analSet$plsda$
coef.mat)<1) {
    colnamesdr<-c("Spearman_r", "F-PCA_score",
        "PLS_dist1", "PLS_dist2", "PLS_VIP", "PLS_RC", "PLS_

```

```

SR",
                                "PLS-DA_VIP", "PLS-DA_C1", "PLS-DA_C2", "PLS-DA_C3"
)
}else if (mSet$dataSet$cls.num == 4 & length(mSet$analSet$plsda$
coef.mat)<1) {
    colnamesdr<-c("Spearman_r", "F-PCA_score",
                  "PLS_dist1", "PLS_dist2", "PLS_VIP", "PLS_RC", "PLS_
SR",
                                "PLS-DA_VIP", "PLS-DA_C1", "PLS-DA_C2", "PLS-DA_C3"
, "PLS-DA_C4")
}else if (mSet$dataSet$cls.num == 5 & length(mSet$analSet$plsda$
coef.mat)<1) {
    colnamesdr<-c("Spearman_r", "F-PCA_score",
                  "PLS_dist1", "PLS_dist2", "PLS_VIP", "PLS_RC", "PLS_
SR",
                                "PLS-DA_VIP", "PLS-DA_C1", "PLS-DA_C2", "PLS-DA_C3"
, "PLS-DA_C4", "PLS-DA_C5")
}else if (mSet$dataSet$cls.num == 6 & length(mSet$analSet$plsda$
coef.mat)<1) {
    colnamesdr<-c("Spearman_r", "F-PCA_score",
                  "PLS_dist1", "PLS_dist2", "PLS_VIP", "PLS_RC", "PLS_
SR",
                                "PLS-DA_VIP", "PLS-DA_C1", "PLS-DA_C2", "PLS-DA_C3"
, "PLS-DA_C4", "PLS-DA_C5", "PLS-DA_C6")

}else if (mSet$dataSet$cls.num == 2 & length(mSet$analSet$plsda$
coef.mat)>1) {
    colnamesdr<-c("Spearman_r", "F-PCA_score",
                  "PLS_dist1", "PLS_dist2", "PLS_VIP", "PLS_RC", "PLS_
SR",
                                "PLS-DA_VIP", "PLS-DA_C1", "PLS-DA_C2", "PLS-DA_Coe
f.mean")
}else if (mSet$dataSet$cls.num == 3 & length(mSet$analSet$plsda$
coef.mat)>1) {
    colnamesdr<-c("Spearman_r", "F-PCA_score",
                  "PLS_dist1", "PLS_dist2", "PLS_VIP", "PLS_RC", "PLS_
SR",
                                "PLS-DA_VIP", "PLS-DA_C1", "PLS-DA_C2", "PLS-DA_C3"
, "PLS-DA_Coef.mean")
}else if (mSet$dataSet$cls.num == 4 & length(mSet$analSet$plsda$
coef.mat)>1) {
    colnamesdr<-c("Spearman_r", "F-PCA_score",
                  "PLS_dist1", "PLS_dist2", "PLS_VIP", "PLS_RC", "PLS_
SR",
                                "PLS-DA_VIP", "PLS-DA_C1", "PLS-DA_C2", "PLS-DA_C3"
, "PLS-DA_C4", "PLS-DA_Coef.mean")
}else if (mSet$dataSet$cls.num == 5 & length(mSet$analSet$plsda$
coef.mat)>1) {
    colnamesdr<-c("Spearman_r", "F-PCA_score",

```

```

        "PLS_dist1", "PLS_dist2", "PLS_VIP", "PLS_RC", "PLS_
SR",
        "PLS-DA_VIP", "PLS-DA_C1", "PLS-DA_C2", "PLS-DA_C3"
, "PLS-DA_C4", "PLS-DA_C5", "PLS-DA_Coef.mean")
    } else if (mSet$dataSet$cls.num == 6 & length(mSet$analSet$plsda$
coef.mat) > 1) {
        colnamesdr <- c("Spearman_r", "F-PCA_score",
        "PLS_dist1", "PLS_dist2", "PLS_VIP", "PLS_RC", "PLS_
SR",
        "PLS-DA_VIP", "PLS-DA_C1", "PLS-DA_C2", "PLS-DA_C3"
, "PLS-DA_C4", "PLS-DA_C5", "PLS-DA_C6", "PLS-DA_Coef.mean")

    } else {
        print("no group")
    }

    # Recover the most important features according to each biochemome
tric model
    # Spearman
    Spearman_Hit <- which(Data_result[, "Spearman_r"] >= sort(Data_resul
t[, "Spearman_r"], decreasing=T) [NBofInterest] & Data_result[, "Spearm
an_r"] >= 0.5 )
    if (length(Spearman_Hit) < 1) {
        Spearman_Hit <- which(Data_result[, "Spearman_r"] >= 0.5) [1:NBofInt
erest]
    } else {
        print("ok")
    }

    # F-PCA
    F_PCA_Hit <- which(Data_result[, "F-PCA_score"] >= sort(Data_result[
, "F-PCA_score"], decreasing=T) [NBofInterest] & Data_result[, "F-PCA_
score"] >= rayonsign)
    if (length(F_PCA_Hit) < 1) {
        F_PCA_Hit <- which(Data_result[, "F-PCA_score"] >= rayonsign) [1:NB
ofInterest]
    } else {
        print("ok")
    }

    # PLS
    PLS_Hit <- which(Data_result[, "PLS_SR"] >= sort(Data_result[, "PLS_S
R"] [which(Data_result[, "PLS_SR"] >= 1 & Data_result[, "PLS_RC"] >= 0)], de
creasing = T) [NBofInterest] &
        Data_result[, "PLS_RC"] >= 0 & Data_result[, "PLS_
SR"] >= 1)
    if (length(PLS_Hit) < 1) {
        PLS_Hit <- which(Data_result[, "PLS_SR"] >= 1 & Data_result[, "PLS_
RC"] >= 0) [1:NBofInterest]
    } else {

```

```

    print("ok")
  }

  # PLS-DA
  if (mSet$dataSet$cls.num == 3) {
    PLS_DA_Hit<-which(Data_result[, "PLS-DA_VIP"] >= sort(Data_result[, "PLS-DA_VIP"][which(Data_result[, "PLS-DA_VIP"]>=1 &
Data_result[, "PLS-DA_C3"]>=Data_result[, "PLS-DA_C2"] &
Data_result[, "PLS-DA_C3"]>=Data_result[, "PLS-DA_C1"])), decreasing =
T) [NBofInterest] &
    Data_result[, "PLS-DA_VIP"]>=1 &
    Data_result[, "PLS-DA_C3"]>=Data_result[, "P
LS-DA_C2"] &
    Data_result[, "PLS-DA_C3"]>=Data_result[, "P
LS-DA_C1"])

  } else if (mSet$dataSet$cls.num == 3 & length(PLS_DA_Hit)<1) {
    PLS_DA_Hit<- which(Data_result[, "PLS-DA_VIP"]>=1 &
    Data_result[, "PLS-DA_C3"]>=Data_result[, "
PLS-DA_C2"] &
    Data_result[, "PLS-DA_C3"]>=Data_result[, "
PLS-DA_C1"])[1:NBofInterest]

  } else if (mSet$dataSet$cls.num == 4) {
    PLS_DA_Hit<- which(Data_result[, "PLS-DA_VIP"]>=sort(Data_result[, "PLS-DA_VIP"][which(Data_result[, "PLS-DA_VIP"]>=1 &
Data_result[, "PLS-DA_C4"]>=Data_result[, "PLS-DA_C3"]&
Data_result[, "PLS-DA_C4"]>=Data_result[, "PLS-DA_C2"] &
Data_result[, "PLS-DA_C4"]>=Data_result[, "PLS-DA_C1"])), decreasing =
T) [NBofInterest] &
    Data_result[, "PLS-DA_VIP"]>=1 &
    Data_result[, "PLS-DA_C4"]>=Data_result[, "
PLS-DA_C3"] &
    Data_result[, "PLS-DA_C4"]>=Data_result[, "
PLS-DA_C2"] &
    Data_result[, "PLS-DA_C4"]>=Data_result[, "
PLS-DA_C1"])

  } else if (mSet$dataSet$cls.num == 4 & length(PLS_DA_Hit)<1) {
    PLS_DA_Hit<- which(Data_result[, "PLS-DA_VIP"]>=1 &
    Data_result[, "PLS-DA_C4"]>=Data_result[, "
PLS-DA_C3"] &
    Data_result[, "PLS-DA_C4"]>=Data_result[, "
PLS-DA_C2"] &
    Data_result[, "PLS-DA_C4"]>=Data_result[, "

```

```

PLS-DA_C1"))[1:NBofInterest]

  } else if (mSet$dataSet$cls.num == 5) {
    PLS_DA_Hit<- which(Data_result[, "PLS-DA_VIP"]>=sort(Data_resul
t[, "PLS-DA_VIP"][which(Data_result[, "PLS-DA_VIP"]>=1 &
Data_result[, "PLS-DA_C5"]>=Data_result[, "PLS-DA_C4"]&
Data_result[, "PLS-DA_C5"]>=Data_result[, "PLS-DA_C3"]&
Data_result[, "PLS-DA_C5"]>=Data_result[, "PLS-DA_C2"] &
Data_result[, "PLS-DA_C5"]>=Data_result[, "PLS-DA_C1"])),decreasing =
T) [NBofInterest] &
                                Data_result[, "PLS-DA_VIP"]>=1 &
                                Data_result[, "PLS-DA_C5"]>=Data_result[, "
PLS-DA_C4"] &
                                Data_result[, "PLS-DA_C5"]>=Data_result[, "
PLS-DA_C3"] &
                                Data_result[, "PLS-DA_C5"]>=Data_result[, "
PLS-DA_C2"] &
                                Data_result[, "PLS-DA_C5"]>=Data_result[, "
PLS-DA_C1"])

  } else if (mSet$dataSet$cls.num == 5 & length(PLS_DA_Hit)<1) {
    PLS_DA_Hit<- which(Data_result[, "PLS-DA_VIP"]>=1 &
                                Data_result[, "PLS-DA_VIP"]>=1 &
                                Data_result[, "PLS-DA_C5"]>=Data_result[, "
PLS-DA_C4"] &
                                Data_result[, "PLS-DA_C5"]>=Data_result[, "
PLS-DA_C3"] &
                                Data_result[, "PLS-DA_C5"]>=Data_result[, "
PLS-DA_C2"] &
                                Data_result[, "PLS-DA_C5"]>=Data_result[, "
PLS-DA_C1"]) [1:NBofInterest]

  } else if (mSet$dataSet$cls.num == 6) {
    PLS_DA_Hit<- which(Data_result[, "PLS-DA_VIP"]>=sort(Data_resul
t[, "PLS-DA_VIP"][which(Data_result[, "PLS-DA_VIP"]>=1 &
Data_result[, "PLS-DA_C6"]>=Data_result[, "PLS-DA_C5"]&
Data_result[, "PLS-DA_C6"]>=Data_result[, "PLS-DA_C4"]&
Data_result[, "PLS-DA_C6"]>=Data_result[, "PLS-DA_C3"]&
Data_result[, "PLS-DA_C6"]>=Data_result[, "PLS-DA_C2"] &
Data_result[, "PLS-DA_C6"]>=Data_result[, "PLS-DA_C1"])),decreasing =
T) [NBofInterest] &

```

```

Data_result[, "PLS-DA_VIP"] >= 1 &
Data_result[, "PLS-DA_C6"] >= Data_result[, "
PLS-DA_C5"] &
Data_result[, "PLS-DA_C6"] >= Data_result[, "
PLS-DA_C4"] &
Data_result[, "PLS-DA_C6"] >= Data_result[, "
PLS-DA_C3"] &
Data_result[, "PLS-DA_C6"] >= Data_result[, "
PLS-DA_C2"] &
Data_result[, "PLS-DA_C6"] >= Data_result[, "
PLS-DA_C1"])

} else if (mSet$dataSet$cls.num == 6 & length(PLS_DA_Hit) < 1) {
  PLS_DA_Hit <- which(Data_result[, "PLS-DA_VIP"] >= 1 &
    Data_result[, "PLS-DA_VIP"] >= 1 &
    Data_result[, "PLS-DA_C6"] >= Data_result[, "
PLS-DA_C5"] &
    Data_result[, "PLS-DA_C6"] >= Data_result[, "
PLS-DA_C4"] &
    Data_result[, "PLS-DA_C6"] >= Data_result[, "
PLS-DA_C3"] &
    Data_result[, "PLS-DA_C6"] >= Data_result[, "
PLS-DA_C2"] &
    Data_result[, "PLS-DA_C6"] >= Data_result[, "
PLS-DA_C1"]) [1:NBofInterest]

} else {
  print("ok")
}

# Recover data for a step
all <- unique(c(Spearman_Hit, F_PCA_Hit, PLS_Hit, PLS_DA_Hit))
Super_list <- Data_result[all, colnames(dr)]

```

5.2 Add annotation of the model used to select highlighted features

Add "Spearman", "fPCA", "PLS" and/or "PLS-DA" annotation for each feature according to the same conditions defined previously:

- Spearman: $r \geq 0.5$
- fPCA: f_PCA_score with the most important significant positive correlation
- PLS: selectivity ratio (SR) ≥ 1 and regression coefficient (RC) > 0
- PLS-DA: VIP ≥ 1 and coefficient for bioactive groups higher than coefficient for the other groups

```

# Create four columns
Super_list <- cbind(Super_list, modp = '', modf = '', modpls = '', modplsda =
'', Models = "")
Super_list[, "modp"] <- as.character(Super_list[, "modp"])
Super_list[, "modf"] <- as.character(Super_list[, "modf"])
Super_list[, "modpls"] <- as.character(Super_list[, "modpls"])

```



```

Super_list[, "modplsda"] <- as.character(Super_list[, "modplsda"])

# If the number of features with "high" importance level is higher than the maximal number defined at the beginning of the script

if (mSet$dataSet$cls.num == 2) {
  Super_list[, "modp"] [which(Super_list[, "Spearman_r"] >= sort(Data_result[, "Spearman_r"], decreasing=T)[NBofInterest] &
    (Super_list[, "Spearman_r"] >= 0.5) )
] <- paste('Spearman')

  Super_list[, "modf"] [which(Super_list[, "F-PCA_score"] >= sort(Data_result[, "F-PCA_score"], decreasing=T) [NBofInterest] &
    (Super_list[, "F-PCA_score"] >= rayon
sign))] <- paste('F-PCA')

  Super_list[, "modpls"] [which(Super_list[, "PLS_SR"] >= sort(Data_result[, "PLS_SR"] [which(Data_result[, "PLS_SR"] >= 1 & Data_result[, "PLS_RC"] >= 0)], decreasing = T)[NBofInterest] &
    (Super_list[, "PLS_RC"] >= 0) & Super_list[, "PLS_SR"] >= 1 )] <- paste('PLS')

  Super_list[, "modplsda"] [which(Super_list[, "PLS-DA_VIP"] >= sort(Data_result[, "PLS-DA_VIP"] [which(Data_result[, "PLS-DA_VIP"] >= 1 &
Data_result[, "PLS-DA_C2"] >= Data_result[, "PLS-DA_C1"])], decreasing = T) [NBofInterest] &
    (Super_list[, "PLS-DA_C2"] >= Super_list[, "PLS-DA_C1"] &
    Super_list[, "PLS-DA_VIP"] >= 1))] <- paste('PLS-DA')

} else if (mSet$dataSet$cls.num == 3) {

  Super_list[, "modp"] [which(Super_list[, "Spearman_r"] >= sort(Data_result[, "Spearman_r"], decreasing=T)[NBofInterest] &
    (Super_list[, "Spearman_r"] >= 0.5) )
] <- paste('Spearman')

  Super_list[, "modf"] [which(Super_list[, "F-PCA_score"] >= sort(Data_result[, "F-PCA_score"], decreasing=T) [NBofInterest] &
    (Super_list[, "F-PCA_score"] >= rayon
sign) )] <- paste('F-PCA')

  Super_list[, "modpls"] [which(Super_list[, "PLS_SR"] >= sort(Data_result[, "PLS_SR"] [which(Data_result[, "PLS_SR"] >= 1 & Data_result[, "PLS_RC"] >= 0)], decreasing = T)[NBofInterest] &
    (Super_list[, "PLS_RC"] >= 0 & Super_list[, "PLS_SR"] >= 1))] <- paste('PLS')

```

```

    Super_list[, "modplsda"][which(Super_list[, "PLS-DA_VIP"] >= sort(Data_result[, "PLS-DA_VIP"][which(Data_result[, "PLS-DA_VIP"] >= 1 & Data_result[, "PLS-DA_C3"] >= Data_result[, "PLS-DA_C2"] & Data_result[, "PLS-DA_C3"] >= Data_result[, "PLS-DA_C1"])], decreasing = T) [NBofInterest] & (Super_list[, "PLS-DA_C3"] >= Super_list[, "PLS-DA_C2"] & Super_list[, "PLS-DA_C3"] >= Super_list[, "PLS-DA_C1"] & Super_list[, "PLS-DA_VIP"] >= 1))] <- paste('PLS-DA')

  } else if (mSet$dataSet$cls.num == 4) {

    Super_list[, "modp"] [which(Super_list[, "Spearman_r"] >= sort(Data_result[, "Spearman_r"], decreasing=T) [NBofInterest] & (Super_list[, "Spearman_r"] >= 0.5) ) ] <- paste('Spearman')
    Super_list[, "modf"] [which(Super_list[, "F-PCA_score"] >= sort(Data_result[, "F-PCA_score"], decreasing=T) [NBofInterest] & (Super_list[, "F-PCA_score"] >= rayon sign) )] <- paste('F-PCA')

    Super_list[, "modpls"] [which(Super_list[, "PLS_SR"] >= sort(Data_result[, "PLS_SR"] [which(Data_result[, "PLS_SR"] >= 1 & Data_result[, "PLS_RC"] >= 0)], decreasing = T) [NBofInterest] & (Super_list[, "PLS_RC"] >= 0 & Super_list[, "PLS_SR"] >= 1) )] <- paste('PLS')

    Super_list[, "modplsda"] [which(Super_list[, "PLS-DA_VIP"] >= sort(Data_result[, "PLS-DA_VIP"] [which(Data_result[, "PLS-DA_VIP"] >= 1 & Data_result[, "PLS-DA_C4"] >= Data_result[, "PLS-DA_C3"] & Data_result[, "PLS-DA_C4"] >= Data_result[, "PLS-DA_C2"] & Data_result[, "PLS-DA_C4"] >= Data_result[, "PLS-DA_C1"])], decreasing = T) [NBofInterest] & (Super_list[, "PLS-DA_C4"] >= Super_list[, "PLS-DA_C3"] & Super_list[, "PLS-DA_C4"] >= Super_list[, "PLS-DA_C2"] & Super_list[, "PLS-DA_C4"] >= Super_list[, "PLS-DA_C1"] & Super_list[, "PLS-DA_VIP"] >= 1))] <- paste('PLS-DA')

  } else if (mSet$dataSet$cls.num == 5) {

    Super_list[, "modp"] [which(Super_list[, "Spearman_r"] >= sort(Data

```

```

_result[,"Spearman_r"],decreasing=T)[NBofInterest] &
      (Super_list[,"Spearman_r"]>=0.5) )
] <-paste('Spearman')
      Super_list[,"modf"][which(Super_list[,"F-PCA_score"]>=sort(Data
a_result[,"F-PCA_score"],decreasing=T) [NBofInterest] &
      (Super_list[,"F-PCA_score"]>=rayon
sign ))] <-paste('F-PCA')

      Super_list[,"modpls"][which(Super_list[,"PLS_SR"] >= sort(Data
_result[,"PLS_SR"][which(Data_result[,"PLS_SR"]>=1 & Data_result[,"P
LS_RC"]>=0)],decreasing = T)[NBofInterest] &
      (Super_list[,"PLS_RC"]>=0 & Supe
r_list[,"PLS_SR"]>=1) )]<-paste('PLS')

      Super_list[,"modplsda"][which(Super_list[,"PLS-DA_VIP"] >= sor
t(Data_result[,"PLS-DA_VIP"][which(Data_result[,"PLS-DA_VIP"]>=1 &
Data_result[,"PLS-DA_C5"]>=Data_result[,"PLS-DA_C4"]&Data_result[,"P
LS-DA_C5"]>=Data_result[,"PLS-DA_C3"] &Data_result[,"PLS-DA_C5"]>=Da
ta_result[,"PLS-DA_C2"]
&Data_result[,"PLS-DA_C5"]>=Data_result[,"PLS-DA_C1"])],decreasing =
T) [NBofInterest] &
      (Super_list[,"PLS-DA_C5"]>=Sup
er_list[,"PLS-DA_C4"]&Super_list[,"PLS-DA_C5"]>=Super_list[,"PLS-DA_
C3"]&Super_list[,"PLS-DA_C5"]>=Super_list[,"PLS-DA_C2"]&
      Super_list[,"PLS-DA_C5"]>=S
uper_list[,"PLS-DA_C1"] & Super_list[,"PLS-DA_VIP"]>=1))<-paste('PL
S-DA')

} else if (mSet$dataSet$cls.num == 6) {

      Super_list[,"modp"][which(Super_list[,"Spearman_r"]>=sort(Data
_result[,"Spearman_r"],decreasing=T)[NBofInterest] &
      (Super_list[,"Spearman_r"]>=0.5) )
] <-paste('Spearman')
      Super_list[,"modf"][which(Super_list[,"F-PCA_score"]>=sort(Data
a_result[,"F-PCA_score"],decreasing=T) [NBofInterest] &
      (Super_list[,"F-PCA_score"]>=rayon
sign))] <-paste('F-PCA')

      Super_list[,"modpls"][which(Super_list[,"PLS_SR"] >= sort(Data
_result[,"PLS_SR"][which(Data_result[,"PLS_SR"]>=1 & Data_result[,"P
LS_RC"]>=0)],decreasing = T)[NBofInterest] &
      (Super_list[,"PLS_SR"]>=1 & Supe
r_list[,"PLS_RC"]>=0) )]<-paste('PLS')

      Super_list[,"modplsda"][which(Super_list[,"PLS-DA_VIP"] >= sor
t(Data_result[,"PLS-DA_VIP"][which(Data_result[,"PLS-DA_VIP"]>=1 &

```

```

Data_result[,"PLS-DA_C6"]>=Data_result[,"PLS-DA_C5"]&Data_result[,"PLS-DA_C6"]>=Data_result[,"PLS-DA_C4"] & Data_result[,"PLS-DA_C6"]>=Data_result[,"PLS-DA_C3"]

&Data_result[,"PLS-DA_C6"]>=Data_result[,"PLS-DA_C2"]&Data_result[,"PLS-DA_C6"]>=Data_result[,"PLS-DA_C1"])),decreasing = T) [NBofInterest] &
(Super_list[,"PLS-DA_C6"]>=Super_list[,"PLS-DA_C5"]&Super_list[,"PLS-DA_C6"]>=Super_list[,"PLS-DA_C4"]&Super_list[,"PLS-DA_C6"]>=Super_list[,"PLS-DA_C3"]&Super_list[,"PLS-DA_C6"]>=Super_list[,"PLS-DA_C2"]&
Super_list[,"PLS-DA_C6"]>=Super_list[,"PLS-DA_C1"] & Super_list[,"PLS-DA_VIP"]>=1))<-paste('PLS-DA')

} else {
  print("ok")
}

# If the number of interested features (features with "high" importance levels ) is less than the maximal number defined at the beginning of the script

if (length(which(Super_list[,"modp"]=="Spearman"))<1) {
  Super_list[,"modp"][which(Super_list[,"Spearman_r"]>=0.5)][1:NBofInterest] <-paste('Spearman')
} else if (length(which(Super_list[,"modf"]=="F-PCA"))<1) {
  Super_list[,"modf"][which(Super_list[,"F-PCA_score"]>=rayonsign)][1:NBofInterest] <-paste('F-PCA')
} else if (length(which(Super_list[,"modpls"]=="PLS"))<1) {
  Super_list[,"modpls"][which(Super_list[,"PLS_RC"]>=0 & Super_list[,"PLS_SR"]>=1)][1:NBofInterest]<-paste('PLS')
} else if (length(which(Super_list[,"modplsda"]=="PLS-DA"))<1 & mSet$dataSet$cls.num == 2 ) {
  Super_list[,"modplsda"][which(Super_list[,"PLS-DA_C2"]>=Super_list[,"PLS-DA_C1"] &
Super_list[,"PLS-DA_VIP"]>=1)]
[1:NBofInterest]<-paste('PLS-DA')
} else if (length(which(Super_list[,"modplsda"]=="PLS-DA"))<1 & mSet$dataSet$cls.num == 3) {
  Super_list[,"modplsda"][which(Super_list[,"PLS-DA_C3"]>=Super_list[,"PLS-DA_C2"]&
Super_list[,"PLS-DA_C3"]>=Super_list[,"PLS-DA_C1"] &
Super_list[,"PLS-DA_VIP"]>=1)]
[1:NBofInterest]<-paste('PLS-DA')
} else if (length(which(Super_list[,"modplsda"]=="PLS-DA"))<1 & mSet$dataSet$cls.num == 4) {
  Super_list[,"modplsda"][which(Super_list[,"PLS-DA_C4"]>=Super_list[,"PLS-DA_C3"]&Super_list[,"PLS-DA_C4"]>=Super_list[,"PLS-DA_C2"]

```

```

]&
                                Super_list[,"PLS-DA_C4"]>=Super
r_list[,"PLS-DA_C1"] & Super_list[,"PLS-DA_VIP"]>=1)][1:NBofInterest
]<-paste('PLS-DA')
    } else if (length(which(Super_list[,"modplsda"]=="PLS-DA"))<1 &
mSet$dataSet$cls.num == 5) {
        Super_list[,"modplsda"][which (Super_list[,"PLS-DA_C5"]>=Super
_list[,"PLS-DA_C4"]&Super_list[,"PLS-DA_C5"]>=Super_list[,"PLS-DA_C3
"]&Super_list[,"PLS-DA_C5"]>=Super_list[,"PLS-DA_C2"]&
                                Super_list[,"PLS-DA_C5"]>=Sup
er_list[,"PLS-DA_C1"] & Super_list[,"PLS-DA_VIP"]>=1)][1:NBofInteres
t]<-paste('PLS-DA')
    } else if (length(which(Super_list[,"modplsda"]=="PLS-DA"))<1 &
mSet$dataSet$cls.num == 6) {
        Super_list[,"modplsda"][which (Super_list[,"PLS-DA_C6"]>=Super
_list[,"PLS-DA_C5"]&Super_list[,"PLS-DA_C6"]>=Super_list[,"PLS-DA_C4
"]&Super_list[,"PLS-DA_C6"]>=Super_list[,"PLS-DA_C3"]&Super_list[,"P
LS-DA_C6"]>=Super_list[,"PLS-DA_C2"]&
                                Super_list[,"PLS-DA_C6"]>=Sup
er_list[,"PLS-DA_C1"] & Super_list[,"PLS-DA_VIP"]>=1)][1:NBofInteres
t]<-paste('PLS-DA')

    } else {
        print("ok")
    }

# Remove NA introduced during the selection
Super_list<-Super_list[(which(row.names(Super_list)!="NA")),]

# Recover the Super_List for Venn diagram
Super_list_venn<-Super_list

# Merge the annotation of the model used to select each features i
n the same column
Super_list[,"Models"]<-paste(Super_list[,"modp"],Super_list[,"modf
"],Super_list[,"modpls"],Super_list[,"modplsda"],sep=" ")
Super_list<-subset(Super_list,select=c("Spearman_r", "F-PCA_score",
"PLS_SR", "PLS-DA_VIP", "Models"))

```

5.3 Normalization of each biochemometrics score

```

# Normalization (x-xmin)/(xmax-xmin)
Super_list_norm<-Super_list
Super_list_norm[,"Spearman_r"]<-(as.numeric(Super_list[,"Spearman_
r"])-min(as.numeric(Super_list[,"Spearman_r"])))/(max(as.numeric(Sup
er_list[,"Spearman_r"])-min(as.numeric(Super_list[,"Spearman_r"])))
Super_list_norm[,"F-PCA_score"]<-(as.numeric(Super_list[,"F-PCA_sc
ore"])-min(as.numeric(Super_list[,"F-PCA_score"])))/(max(as.numeric(
Super_list[,"F-PCA_score"])-min(as.numeric(Super_list[,"F-PCA_score
"])))
Super_list_norm[,"PLS_SR"]<-(as.numeric(Super_list[,"PLS_SR"])-min

```

```

(as.numeric(Super_list[, "PLS_SR"]))) / (max(as.numeric(Super_list[, "PLS_SR"])) - min(as.numeric(Super_list[, "PLS_SR"])))
Super_list_norm[, "PLS-DA_VIP"] <- (as.numeric(Super_list[, "PLS-DA_VIP"]) - min(as.numeric(Super_list[, "PLS-DA_VIP"]))) / (max(as.numeric(Super_list[, "PLS-DA_VIP"])) - min(as.numeric(Super_list[, "PLS-DA_VIP"])))

Super_list_norm <- cbind(Super_list_norm, SCORE = "")
Super_list_norm$SCORE <- (Super_list_norm$SCORE)
# Rename the columns
colnames(Super_list_norm) <- c("Spearman_r", "F-PCA_score", "PLS_SR", "PLS-DA_VIP", "Models", "SCORE")

```

5.4 Determination of the Global score

The global score for all models is obtained according to the computational mode defined at the beginning of the script

```

# Sum
if(SCORE == "SUM") {
  Super_list_norm[, "SCORE"] <- (Super_list_norm[, "F-PCA_score"] + Super_list_norm[, "Spearman_r"] + Super_list_norm[, "PLS-DA_VIP"] + Super_list_norm[, "PLS_SR"])
# Euclidean distance
}else if(SCORE == "DIST") {
  Super_list_norm[, "SCORE"] <- sqrt(Super_list_norm[, "F-PCA_score"]^2 + Super_list_norm[, "Spearman_r"]^2 + Super_list_norm[, "PLS-DA_VIP"]^2 + as.numeric(Super_list_norm[, "PLS_SR"])^2)
# Mean
}else if(SCORE == "MEAN") {
  Super_list_norm[, "SCORE"] <- mean(Super_list_norm[, "F-PCA_score"], Super_list_norm[, "Spearman_r"], Super_list_norm[, "PLS-DA_VIP"], Super_list_norm[, "PLS_SR"])
}
}else {
  print("no score")
}

# Add the "global score" to the Super-list
Super_list <- data.frame(Super_list, Super_list_norm[, "SCORE"])
# Rename the columns to keep only parameters defined for classification of importance levels
colnames(Super_list) <- c("Spearman_r", "F-PCA_score", "PLS_SR", "PLS-DA_VIP", "Models", "SCORE")
# Super_list is ordered according to the global score obtained for each feature
Super_list <- Super_list[order(-Super_list[, "SCORE"]),]
# Round the values
Super_list <- as.matrix(Super_list)
Super_list[, -5] <- round(as.numeric(Super_list[, -5]), 2)
# Save the table in the working directory

```

```
write.table(cbind(features=row.names(Super_list),Super_list),file=
"Super_list.csv",sep=";",dec=".",row.names = F)
```

5.5 Graphical representation: radar chart and venn diagram

```
# Features are ordered according to their "global score" obtained
from all models
Super_list_norm<-Super_list_norm[order(-Super_list_norm[, "SCORE"])
,]
# remove "models" column in the table
Imp_ions2<-subset(Super_list_norm,select = -c(Models))

# a. Radar chart

# References scores
ref<-c(1,1,1,1)

# Merge references scores for the four models with the normalized
scores obtained for highlited features
ref<-rbind(Imp_ions2,ref)
colnames(ref)<-c("Spearman", "F-PCA", "PLS", "PLS-DA")

# Number of graphical windows required for radars charts
NbWindows<-ceiling((dim(Imp_ions2)[1]+1)/16)
for(i in 1:NbWindows){ #i<-1
  if(i == 1){
    if (OpSys == "mac"){
      quartz("radar_chart_1")
    } else {
      X11(15,15,title="radar_chart_1")
    }
    op<-par(mar=c(1,1,1,1),mfrow=c(4,4),ylbias=0.2,yaxs="r",yaxt="
s")

    # Fifteen radar chart appears on the graphical window
    y=1:15
    x=1:(ncol(Imp_ions2)-1)

    # Apply function spider for the referencess scores: Spearman=1
,F-PCA=1,PLS=1 and PLS-DA=1
    spider(nrow(ref),1:(ncol(ref)-1),ref,main=,cxy=NULL,labels=col
names(ref))
    spider(y,x,Imp_ions2,center=FALSE,connect=TRUE,add=FALSE,cex=1
,main=row.names(Imp_ions2),labels = "")
    dev.copy2pdf(file="radar_chart_1.pdf",device=dev.cur())

  }else{
    if (OpSys == "mac"){
      quartz(paste("radar_chart_",i,sep=""))
    }
  }
}
```

```

    } else {
      X11(15,15,paste("radar_chart_",i,sep=""))
    }
  op<-par(mar=c(1,1, 1,1),mfrow=c(4,4),ylbias=0.2,yaxs="r",yaxt=
"s")

  # Define the number of radar chart on the next graphical window
  according to xy coordinates
  y=(16+(i-2)*17):min((16+(i-2)*17+16),dim(Imp_ions2)[1])
  x=1:(ncol(Imp_ions2)-1)

  # Apply function spider
  spider(y,x,Imp_ions2,center=FALSE,connect=TRUE,add=FALSE,cex=1
,main=rownames(Imp_ions2),labels = "")
  # Save plot(s)
  dev.copy2pdf(file=paste("radar_chart_",i,".pdf",sep=""),device
=dev.cur())
}

}

# b. VennDiagram plot

if (OpSys == "mac"){
  quartz(10,10, title="Venn")
  op<-(mar=c(1,1,1,1))
}else {
  X11(16,10,title="Venn")
  op<-(mar=c(1,1,1,1))
}

# Define data

area1<-length(Super_list_venn[,"modp"][which(Super_list_venn[,"m
odp"]=="Spearman")])
area3<-length(Super_list_venn[,"modf"][which(Super_list_venn[,"m
odf"]=="F-PCA")])
area4<-length(Super_list_venn[,"modpls"][which(Super_list_venn[,
"modpls"]=="PLS")])
area2<-length(Super_list_venn[,"modplsda"][which(Super_list_venn
[, "modplsda"]=="PLS-DA")])
n12<-length(Super_list_venn[,"modp"][which(Super_list_venn[,"mod
p"]=="Spearman" & Super_list_venn[,"modplsda"]=="PLS-DA")])
n13<-length(Super_list_venn[,"modp"][which(Super_list_venn[,"mod
p"]=="Spearman" & Super_list_venn[,"modf"]=="F-PCA")])
n14<-length(Super_list_venn[,"modp"][which(Super_list_venn[,"mod
p"]=="Spearman" & Super_list_venn[,"modpls"]=="PLS")])
n23<-length(Super_list_venn[,"modplsda"][which(Super_list_venn[,
"modplsda"]=="PLS-DA" & Super_list_venn[,"modf"]=="F-PCA")])
n24<-length(Super_list_venn[,"modplsda"][which(Super_list_venn[,

```



```

"modplsda"]=="PLS-DA" & Super_list_venn[, "modpls"]=="PLS"))
  n34<-length(Super_list_venn[, "modf"][which(Super_list_venn[, "mod
f"]=="F-PCA" & Super_list_venn[, "modpls"]=="PLS"))
  n123<-length(Super_list_venn[, "modp"][which(Super_list_venn[, "mo
dp"]=="Spearman" & Super_list_venn[, "modplsda"]=="PLS-DA" & Super_li
st_venn[, "modf"]=="F-PCA"))
  n124<-length(Super_list_venn[, "modp"][which(Super_list_venn[, "mo
dp"]=="Spearman" & Super_list_venn[, "modplsda"]=="PLS-DA" & Super_li
st_venn[, "modpls"]=="PLS"))
  n134<-length(Super_list_venn[, "modp"][which(Super_list_venn[, "mo
dp"]=="Spearman" & Super_list_venn[, "modf"]=="F-PCA" & Super_list_ve
nn[, "modpls"]=="PLS"))
  n234<-length(Super_list_venn[, "modplsda"][which(Super_list_venn[
, "modplsda"]=="PLS-DA" & Super_list_venn[, "modf"]=="F-PCA" & Super_l
ist_venn[, "modpls"]=="PLS"))
  n1234<-length(Super_list_venn[, "modp"][which(Super_list_venn[, "m
odp"]=="Spearman" & Super_list_venn[, "modf"]=="F-PCA" & Super_list_v
enn[, "modpls"]=="PLS" & Super_list_venn[, "modplsda"]=="PLS-DA"))

# Build the venn.plot
venn.plot <- draw.quad.venn(area1=area1,
                           area2=area2,
                           area3=area3,
                           area4=area4,
                           n12=n12,
                           n13=n13,
                           n14=n14,
                           n23=n23,
                           n24=n24,
                           n34=n34,
                           n123=n123,
                           n124=n124,
                           n134=n134,
                           n234=n234,
                           n1234=n1234,
                           category = c("Spearman", "PLS-DA", "F-
PCA", "PLS"),
                           fill = c("white", "white", "white",
"white"),
                           lty = "dashed",
                           cex = 2,
                           cat.cex = 2,
                           filename = NULL)

# Function to obtain the Venn diagram
grid.draw(venn.plot)

# Save the plot
dev.copy2pdf(file="VennDiagram.pdf", device=dev.cur())
dev.off()

```

```
if (length(Super_list)>1){
  nblast = nrow(Super_list)
  message("FINISHED -",nblast," bioactive features HIGHLIGHTED =
D")
}else {
  message("FINISHED - Sorry, no bioactive features HIGHLIGHTED =
(")
}
```


Titre : Du bioguidage à la métabolomique pour l'identification de métabolites fongiques actifs sur des modèles cellulaires du cancer du sein- Etude de l'activité de l'ergostérol.

Mots clés : Champignons marins, cancer du sein, bioguidage, métabolomique, biochimométrie, ergosterol

Résumé : Les champignons marins produisent des métabolites d'une grande diversité chimique et potentiellement valorisables en thérapeutique. Les travaux réalisés lors de cette thèse ont porté sur la mise en évidence de métabolites ayant une action antiproliférative sur des lignées cancéreuses mammaires. Ainsi, différentes approches de fractionnement bioguidé, de profilage métabolomique (LC-UV-HRMS) et de biochimométrie, réalisées sur une souche de *Penicillium chrysogenum* ont permis d'isoler et d'identifier une seule molécule responsable de l'activité observée : l'ergostérol, ainsi que d'autres molécules. Afin d'identifier précocement les métabolites bioactifs au sein d'un mélange complexe, il a été proposé et développé un « workflow » comprenant microfractionnement, acquisition des données chimiques et biologiques, et traitement bioinformatique.

Dans ce cadre, un programme combinant 4 modèles statistiques (Spearman, ACP, PLS et PLSDA) a été développé sous R. Appliqué à la souche *P. chrysogenum*, ce workflow a permis de confirmer l'importance de l'ergostérol dans l'activité de l'extrait. L'étude de ce stéroïde sur nos modèles de cellules cancéreuses du sein a montré qu'il réduit le cholestérol en favorisant son efflux (via les transporteurs ABC), et en diminuant probablement sa captation. La réduction de l'expression de la flotilline-2, de l'EGFR et de la cavéoline-1 dans les radeaux lipidiques témoigne de son action sur les voies de signalisation actives dans les cellules cancéreuses. Par conséquent, l'ergostérol exerce une activité intéressante sur des cellules cancéreuses du sein : il diminue la prolifération cellulaire en ciblant le métabolisme du cholestérol.

Title : From Bioguided fractionation to metabolomics for identification of actives fungal metabolites on breast cancer cell models- Study of biological activity of ergosterol.

Keywords: Marine fungi, breast cancer, bioguided assay, metabolomics, biochimométrie, ergosterol

Abstract: Marine fungi produce metabolites of wide chemical diversity that can potentially be used for therapeutic purposes. The work carried out during this thesis focused on the detection of bioactive metabolites with antiproliferative action on mammary cancer cell lines. Thus, different approaches of bioguided fractionation, metabolomic profiling (LC-UV-HRMS) and biochemometrics performed on a *Penicillium chrysogenum* strain, have allowed the isolation and identification of a single molecule responsible for the observed activity: ergosterol, and other molecules. In order to identify bioactive metabolites in a complex mixture at an early stage, a workflow has been proposed and developed, including microfractionation chemical and biological data acquisition and bioinformatics treatment.

In this context, a combining 4 statistical models (Spearman, PCA, PLS and PLSDA) was developed with R. Application of this workflow to the *P. chrysogenum* strain confirmed the importance of ergosterol in the extract activity. The study of this steroid on our breast cancer cell models has shown that it reduces cholesterol by promoting its efflux (via ABC carriers) and decreasing probably its uptake. The reduction in the expression of flotillin-2, EGFR and caveolin-1 in lipid rafts is an evidence of its action on active signalling pathways in cancer cells. These results indicate that ergosterol has an interesting activity on breast cancer cells: it reduces cell proliferation by targeting cholesterol metabolism.