



HAL
open science

Exploration de la diversité chimique des Apocynaceae par la technique des réseaux moléculaires : de la création d'une base de données vers l'annotation in silico

Alexander Fox Ramos

► To cite this version:

Alexander Fox Ramos. Exploration de la diversité chimique des Apocynaceae par la technique des réseaux moléculaires : de la création d'une base de données vers l'annotation in silico. Chimie thérapeutique. Université Paris-Saclay, 2018. Français. NNT : 2018SACLS543 . tel-02221740

HAL Id: tel-02221740

<https://theses.hal.science/tel-02221740>

Submitted on 1 Aug 2019

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

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

Exploration de la diversité chimique des Apocynaceae par la technique des réseaux moléculaires : de la création d'une base de données vers l'annotation *in silico*

Thèse de doctorat de l'Université Paris-Saclay
préparée à l'Université Paris-Sud

École doctorale n°569 Innovation thérapeutique, du fondamental à
l'appliqué
Spécialité de doctorat : Chimie des substances naturelles

Thèse présentée et soutenue à Châtenay-Malabry, le 21 Décembre 2018, par

Alexander Enrique FOX RAMOS

Composition du Jury :

M. Pascal Richomme-Peniguel Professeur, Université d'Angers (EA 921 SONAS)	Président du jury
Mme. Soizic Prado Professeur, MNHM (UMR CNRS 7245)	Rapporteuse
M. Marc Litaudon Ingénieur de recherche, ICSN (UPR CNRS 2301)	Examineur
Mme. Catherine Roullier Maître de conférences, Université de Nantes (EA 2160 MMS)	Examinatrice
M. Mehdi Beniddir Maître de conférences, Fac. Pharma. Châtenay (BioCIS - UMR 8076)	Examineur
M. Pierre Champy Professeur, Fac. Pharma. Châtenay (BioCIS - UMR 8076)	Directeur de thèse

REMERCIEMENTS

Je remercie sincèrement le Pr. Pascal Richomme-Peniguel et la Pr. Soizic Prado d'avoir accepté d'être les rapporteurs de cette thèse ainsi que le Dr. Marc Litaudon et le Dr. Catherine Roullier de participer au jury en tant qu'examineurs. Merci de me faire l'honneur de juger ce travail.

Je voudrais remercier le Prof. Pierre Champy. Son intérêt envers l'utilisation traditionnelle des plantes, ainsi que son savoir polymathique sur la taxonomie végétale est passionnante. Je lui remercie de m'avoir reçu en tant que stagiaire de M1 et d'avoir encadré cette thèse pendant ces trois ans. Je lui adresse ici toute ma reconnaissance et mon amitié.

Je souhaite ensuite adresser mes remerciements les plus amicaux au Dr. Mehdi A. Beniddir, co-encadrant de cette thèse et examinateur dans le jury. Depuis mon stage de M1, l'apprentissage de nouvelles façons de « faire de la science » ne s'est pas arrêté grâce à lui. Des discussions jusqu'à très tard au laboratoire, des précieux conseils dans le plan professionnel et personnel, ainsi que des réussites célébrées et des défaites surmontées le long de ces trois ans. Tous ces moments partagés resteront toujours dans ma mémoire.

Je voudrais également remercier le Docteur Bruno Figadère, directeur de l'UMR et du laboratoire, qui m'a permis de réaliser ces travaux au sein de l'équipe « chimie des substances naturelles » de BioCIS.

D'autre part, j'adresse mes remerciements à mes collègues du laboratoire que j'ai eu le plaisir de côtoyer, aux stagiaires que j'ai eu l'opportunité d'encadrer ainsi qu'aux enseignants de l'équipe qui m'ont beaucoup appris le long de ces trois ans et pendant le M2 également. Je pense ici particulièrement au Prof. Erwan Poupon, Prof. Michel Lebœuf, Dr. François Roblot, Dr. Alexandre Maciuk, Dr. Laurent Evanno, Dr. Christophe Fourneau, Dr. Natacha Bonneau, Pedro Vasquez, Gaëla Cauchie, Coralie Pavesi, Florian Maloberti, Chloé Vetel et Mohamed Belemlilga.

Je voudrais également remercier le Dr. Guillaume Bernadat (équipe Chimie des Substances Naturelles, BioCIS) pour les cours de modélisation moléculaire sur Python et la réalisation des simulations des spectres de dichroïsme circulaire électronique des produits isolés.

Par ailleurs, je remercie sincèrement Blandine Séon-Méniel (équipe Chimie des Substances Naturelles, BioCIS) pour la préparation des micro-tubes des composés isolés pour réaliser les analyses RMN.



Je tiens à remercier également Karine Leblanc (service chromatographie, spectrométrie de masse, BioCIS) pour toute son aide dans la détermination des conditions analytiques pour la purification des molécules ciblées et pour les analyses des échantillons sur le CLHP-Q-TOF.

Je voudrais également adresser mes remerciements à M. Jean-Christophe Jullian, responsable du service RMN de BioCIS, et Mme. Camille Dejean, d'avoir effectué les analyses RMN des composés purifiés.

J'adresse également mes remerciements à Mme. Elisabeth Mouray et au Pr. Phillippe Grellier de l'Unité de Molécules de Communication et Adaptation des Micro-organismes (MCAM) du Muséum National d'Histoire Naturelle (MNHN) pour la détermination de l'activité antipaludique de molécules isolées.

Je tiens à remercier le Dr. Phillippe M. Loiseau et le Dr. Sébastien Pomel de l'équipe de chimiothérapie antiparasitaire de BioCIS d'avoir effectué les tests d'activité antileishmanienne des composés purs.

De plus, je remercie le Dr. Christophe Duplais, du laboratoire EcoFog (CNRS, UMR 8172) en Guyane d'avoir effectué la récolte du matériel végétal pour entreprendre l'étude phytochimique des écorces de *Geissospermum laeve* (Vell.) Miers.

Je remercie également le Dr. Marc Litaudon, de l'ICSN, pour avoir mis à notre disposition les *Astonia* qu'il a récoltés en Nouvelle-Calédonie avec le Dr. Vincent Dumontet.

Je souhaite remercier M. Jean-François Gallard, responsable du service RMN à l'Institut de Chimie de Substances Naturelles (ICSN - UPR CNRS 2301), d'avoir consacré du temps pour la réalisation des analyses RMN des molécules isolées.

Un grand merci à FONDECYT (Fondo Nacional de Desarrollo Científico y Tecnológico) – CONCYTEC (Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica) pour m'avoir permis de réaliser mes études de doctorat grâce au financement de ma thèse pendant ces trois ans.

Je souhaite également remercier ma belle-famille, particulièrement Virginie et Jacques mais également Francis et Raymonde pour leur immense gentillesse et leur accueil chaleureux.

Je remercie de tout mon cœur mes parents, qui ont su m'encourager et me soutenir malgré la distance physique qui nous sépare.



Enfin, je remercie ma femme, Charlotte Fox-Alcover, pour ses encouragements, sa patience son soutien et son amour au cours de ces trois ans. Cette aventure scientifique n'aurait pas pu être possible sans toi. Ce travail lui est dédié.





TABLE DE MATIÈRES

Remerciements	3
Table de matières.....	7
Préambule.....	9
Liste des abréviations	11
Liste des images	12
Liste des tableaux.....	13
Introduction générale : L'art de l'isolement de produits naturels dans l'ère du <i>Big Data</i> ..	17
Chapitre I : Stratégies d'application du <i>molecular networking</i> pour l'isolement de produits naturels.....	21
1. Introduction	21
1.1 Approche par <i>molecular networking</i> : description générale.....	21
1.2 Traitement des données.....	22
1.3 Bases de données utilisées comme référence	22
1.4 Types de stratégie appliqués	23
1.5 Outils et stratégie visant à améliorer la pratique du <i>molecular networking</i>	23
2. Publication associée.....	27
Chapitre II : Conception d'une base de données d'alcaloïdes indolomonoterpéniques : la MIADB	53
1. Introduction	53
1.1 Description générale.....	53
1.2 Biogenèse et classification.....	54
1.3 Biogenèse	56
1.4 MIADB (<i>Monoterpene Indole Alkaloids DataBase</i>) : du concept à l'application...59	59
2. Publication associée.....	69
Chapitre III : Première application de la MIADB – Exploration de l'espace chimique de <i>Geissospermum laeve</i> (Vell.) Miers.	113
1. Introduction	113
1.1 Phylogénie et description botanique	113
1.2 Répartition géographique	114
1.3 <i>Geissospermum laeve</i> (Vell.) Miers	114
2. Publication associée.....	119
3. Résultats obtenus après publication	171



Chapitre IV : Deuxième application de la MIADB – Exploration de l’espace chimique d’*Alstonia balansae* 175

- 1. Introduction175
 - 1.1 Phylogénie et description botanique175
 - 1.2 Répartition géographique176
 - 1.3 *Alstonia balansae* Guillaumin176
 - 1.4 Travaux entrepris sur *Alstonia* spp.178
- 2. Publication associée.....183
- 3. Résultats biologiques obtenus.....250

Conclusion générale 253

Perspectives 255

Bilan des communications 259

Références bibliographiques..... 263



PRÉAMBULE

Ce manuscrit a été rédigé sous la forme d'une « thèse sur article ». Il est composé de quatre chapitres chacun présentant un article scientifique publié ou à soumettre à l'issue des travaux de recherche entrepris. Les publications originales en langue anglaise ont été incluses avec un résumé en langue française présenté au préalable.

Dans le premier chapitre, les stratégies actuelles d'application du *molecular networking* (technique des réseaux moléculaires) pour l'identification et l'isolement de nouveaux produits naturels seront abordées. Une description générale des approches de déréplication reposant sur cette innovation sera réalisée. En outre, les différentes étapes clés concernant l'utilisation de cet outil seront développées.

Le deuxième chapitre est consacré à décrire la conception d'une base de données d'alcoïdes indolomonoterpéniques (AIMs) au sein de notre laboratoire, avec les apports de plusieurs laboratoires de pharmacognosie : la MIADB (*Monoterpene Indole Alkaloids DataBase*). Une description générale des AIMs sera montrée, ainsi que leur origine biosynthétique, en introduction.

Les deux chapitres suivants correspondent à des applications de cette méthodologie, avec l'étude chimique et l'isolement de nouveaux AIMs à partir d'Apocynaceae. Les espèces étudiées seront présentées. Les articles correspondants seront ensuite proposés.

Le troisième chapitre, dédié à une première application de la MIADB, décrit l'exploration de l'espace chimique des écorces de *Geissospermum laeve* (Vell.) Miers.

Finalement, un quatrième chapitre mènera à une nouvelle approche de déréplication : l'utilisation du *molecular networking* avec une annotation *in silico* par MetWork. L'application de cette méthode pour l'exploration de la diversité chimique des feuilles d'*Alstonia balansae* Guillaumin.

Enfin, une conclusion générale est présentée, ainsi que les perspectives envisageables après ces travaux.





LISTE DES ABRÉVIATIONS

- AIM : alcaloïde indolomonoterpénique
- APG : *Angiosperm Phylogeny Group*
- Cl₅₀ : concentration inhibitrice médiane
- CLHP : chromatographie en phase liquide à haute performance
- COSY : « *correlation spectroscopy* »
- DCE (ECD) : dichroïsme circulaire électronique (« *electronic circular dichroism* »)
- DMEM : « *Dulbecco's modified Eagle's medium* » (milieu de culture d'Eagle modifié par Dulbecco)
- DMSO : diméthylsulfoxyde
- ESI : ionisation par électronébuliseur
- FCS : « *fetal calf serum* » (sérum de veau foetal)
- GNPS : « *Global Natural Products Social molecular networking* »
- HMBC : « *heteronuclear multiple bond correlation* »
- HSQC : « *heteronuclear single quantum correlation* »
- ICSN : Institut de Chimie des Substances Naturelles
- MIADB : « *Monoterpene Indole Alkaloids DataBase* » (base de données d'alcaloïdes indolomonoterpéniques)
- MN : « *molecular networking* » (technique des réseaux moléculaires)
- MNHN : Muséum national d'Histoire naturelle
- MTS : 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium
- m/z* : rapport masse sur charge d'un ion
- NOESY : « *nuclear overhauser effect spectroscopy* »
- Q : quadripôles
- Q-ToF : spectromètre de masse tandem quadripôle-temps de vol
- RMN : résonance magnétique nucléaire
- RPMI : « *Roswell Park Memorial Institute medium* » (milieu de culture du *Roswell Park Memorial*)
- RT : « *retention time* » (temps de rétention)
- SM : spectrométrie de masse
- SMHR : spectrométrie de masse haute résolution
- SM/SM : spectrométrie de masse en tandem
- VIH : virus de l'immunodéficience humaine



LISTE DES IMAGES

Figure 1. Quelques Gentianales et un exemple d'alcaloïde indolomonoterpénique	53
Figure 2. Biosynthèse des alcaloïdes indolomonoterpéniques à partir de la strictosidine	55
Figure 3. Classification des alcaloïdes indolomonoterpéniques d'après Le Men et Taylor (1965)	56
Figure 4. Schéma de filiation biogénétique des AIMs monomériques.	58
Figure 5. Bisindoles majoritaires de <i>G. laeve</i>	59
Figure 6. Évolution de la MIADB au cours du temps	65
Figure 7. Espèces du genre <i>Geissospermum</i>	113
Figure 8. Répartition géographique des espèces du genre <i>Geissospermum</i>	114
Figure 9. <i>G. laeve</i> dans l'herbier du Royal Botanical Gardens.....	114
Figure 10. Alcaloïdes indolomonoterpéniques décrits dans <i>G. laeve</i>	116
Figure 11. Position de la région D dans le réseau moléculaire total	171
Figure 12. Quelques espèces du genre <i>Alstonia</i>	175
Figure 13. Répartition géographique des espèces du genre <i>Alstonia</i>	176
Figure 14. <i>A. balansae</i> dans l'herbier du Royal Botanical Gardens.....	176
Figure 15. Alcaloïdes indolomonoterpéniques décrits dans <i>A. balansae</i>	178



LISTE DES TABLEAUX

Tableau 1. Liste de laboratoires qui ont contribué à la constitution de la MIADB	62
Tableau 2. Alcaloïdes indolomonoterpéniques décrits dans <i>G. laeve</i>	115
Tableau 3. Alcaloïdes indolomonoterpéniques décrits dans <i>A. balansae</i>	177
Tableau 4. Valeurs de CI_{50} (μM) des activités antipaludiques et antileishmaniennes et cytotoxiques des composés 1–5^a	250





INTRODUCTION GÉNÉRALE :

L'ART DE L'ISOLEMENT DES PRODUITS NATURELS DANS L'ÈRE DU *BIG DATA*





INTRODUCTION GÉNÉRALE : L'ART DE L'ISOLEMENT DE PRODUITS NATURELS DANS L'ÈRE DU *BIG DATA*

L'influence de l'ère du *big data* dans le monde a touché plusieurs aspects de nos vies.¹ Nos manières de vivre, de communiquer, de travailler et de réfléchir ont été transformées par cette nouvelle vague où l'information de notre quotidien est numérisée.

Dans ce contexte, la façon de « faire de la science » de nos jours a aussi changé. Elle est même complètement différente de celle d'il y a quelques années. Les expériences sur la paillasse sont à présent complétées par des traitements informatiques des données. De cette manière, les molécules que l'on connaissait seulement sous la forme de structures dessinées sur un papier sont devenues aujourd'hui de l'information numérisée exploitable par des méthodes modernes de traitement, interfacées avec les méthodes d'analyse.

Au cours de ces dernières années, la découverte de composés inédits dans des extraits issus de plantes ou d'autres organismes vivants est devenue plus facile, en raison de l'accès à des techniques plus sensibles d'analyse de mélanges complexes. L'image de « trouver une aiguille dans une botte de foin » évoquait de manière précise la difficulté existant dans l'art de l'isolement des produits naturels. Ainsi, l'influence du traitement de données massives a bouleversé le processus d'isolement des substances naturelles. C'est ainsi que des nouvelles techniques permettant une identification précoce des molécules à structure inédite dans des mélanges complexes ont été ajoutées au portfolio des méthodes utilisées par plusieurs équipes de recherche autour du monde.² De plus, grâce au partage des connaissances au moyen des réseaux du *big data*, des collaborations entre laboratoires pour améliorer ces techniques ont été établies.

Parallèlement, ces collaborations ont mené à la création de bases de données qui ont été mises à la disposition de toute la communauté scientifique.³ L'annotation des composés déjà décrits est donc plus efficace, permettant d'identifier précocement les molécules à cibler au cours de travaux d'exploration de la diversité chimique des mélanges complexes.

Dans ce manuscrit, ces changements de paradigme seront illustrés, avec la constitution d'une base de données issue de la collaboration entre plusieurs laboratoires ayant un historique de recherche autour des alcaloïdes indolomonoterpéniques (AIMs) et l'application d'une démarche de déréplication adaptée à l'étude phytochimique d'Apocynaceae. Le plan suivi a été présenté en préambule.

¹ S. John Walker, *Int. J. Advert.*, 2014, **33**, 181-183.

² J. Hubert, J.-M. Nuzillard, et al., *Phytochem Rev.*, 2017, **16**, 55-95.

³ M. Wang, J. J. Carver, et al., *Nat. Biotechnol.*, 2016, **34**, 828.





CHAPITRE I :

STRATÉGIES D'APPLICATION DU *MOLECULAR NETWORKING* POUR L'ISOLEMENT DE PRODUITS NATURELS





CHAPITRE I : STRATÉGIES D'APPLICATION DU *MOLECULAR NETWORKING* POUR L'ISOLEMENT DE PRODUITS NATURELS

1. INTRODUCTION

1.1 Approche par *molecular networking* : description générale

Pour les chercheurs dans le domaine des produits naturels, deux outils au minimum s'avèrent indispensables pour la caractérisation correcte de la structure de molécules : la résonance magnétique nucléaire (RMN) et la spectrométrie de masse (SM). La RMN s'est positionnée comme l'outil le plus important pour réaliser cette tâche. D'autre part, la SM a longtemps servi à principalement fournir les formules brutes des composés isolés, à côté de l'analyse élémentaire.⁴ Des approches nouvelles permettent maintenant de tenter des déterminations structurales avec la SM seule, aux stéréochimies près.⁵

Dans le cadre de l'appréhension de la complexité moléculaire des mélanges complexes, la SM s'est consolidée comme un outil de choix pour une annotation efficace de produits naturels dans un mélange. En effet, cette technique est à l'heure actuelle largement utilisée pour identifier de manière précoce des composés déjà connus, afin d'éviter l'utilisation inutile de ressources pour les isoler.² Cette façon de procéder est connue sous le nom de « déréplication ».

Depuis son implémentation en ligne, la technique du *molecular networking*⁶ (réseaux moléculaires) a largement facilité la visualisation et la façon de cibler des produits naturels dans des mélanges complexes. Cette technique fondée sur la spectrométrie de masse, permet une comparaison des spectres SM/SM des ions détectés provenant d'un échantillon suite à une séparation chromatographique. Ensuite, un score de similarité (appelé « *cosine* ») est calculé pour représenter l'analogie structurale entre chaque paire de composés du mélange. Ainsi, cette technique permet d'organiser et de visualiser des données SM/SM sous la forme d'une carte de similarité spectrale, mettant en lumière l'existence de groupes spectraux ainsi que leurs degrés de similarité. L'outil disponible pour les chercheurs – le logiciel Cytoscape et l'environnement de travail avec la base de données collaborative GNPS³, pour « *Global Natural Product Social molecular networking* » – ont été mis en place par l'équipe de P. Dorrestein, de l'*University of California* à San Diego. Un outil alternatif similaire est développé par l'ICSN-CNRS à Gif-sur-Yvette (disponible sur : <https://github.com/metgem>).⁷

⁴ N. B. Cech & K. Yu, *LC GC N. Am.*, 2013, **31**, 938-947.

⁵ D. Nikolić, *Phytochem Lett.*, 2017, **21**, 292-296.

⁶ J. Y. Yang, L. M. Sanchez, et al., *J. Nat. Prod.*, 2013, **76**, 1686-1699.

⁷ F. Olivon, N. Elie, et al., *Anal. Chem.*, 2018.



Depuis sa création, cette technique a été intégrée au « workflow » (flux de travail) quotidien de déréplication de nombreuses équipes. En effet, la comparaison des spectres d'un échantillon avec ceux de substances de référence permet l'annotation du réseau moléculaire obtenu. On peut rapidement distinguer des molécules déjà décrites et dédier les efforts d'isolement aux composés potentiellement nouveaux.

1.2 Traitement des données

Avant de soumettre les données de l'échantillon au site du GNPS, elles doivent être prétraitées. Ceci a comme but de réduire l'information qui correspond au bruit de fond, d'ôter les isotopes des massifs isotopiques, etc. Bien qu'il existe une large gamme de logiciels permettant de réaliser ce traitement, l'outil le plus utilisé à l'heure actuelle est MZmine 2.^{8, 9} Généralement, plusieurs paramètres de traitement différents doivent être testés avec ce logiciel avant d'obtenir une collection de spectres SM/SM correcte de l'échantillon analysé. Par ailleurs, d'autres outils peuvent être utilisés pour le traitement des données SM/SM, comme MassHunter® (Agilent)¹⁰ ou OpenMS¹¹.

1.3 Bases de données utilisées comme référence

Dans l'optique de réaliser une déréplication efficace, les équipes de recherche autour du monde ont créé des collections de spectres SM/SM pouvant être utilisées comme référence. Deux types de base de données peuvent être employés :

- Des bases de données de spectres SM/SM générées *in silico*. Un exemple est celle créée par Pierre-Marie Allard *et al.* en 2016,¹² constituée à partir de 170602 produits naturels issus du *Dictionnaire of Natural Products* fragmentés *in silico* avec l'outil CFM-ID (*Competitive Fragmentation Modeling for Metabolite Identification*, disponible sur : <http://cfmid.wishartlab.com/>)¹³. L'emploi de la fragmentation *in silico* pour la déréplication tend à se populariser¹⁴ et sera exploitée dans le dernier chapitre de ce manuscrit.
- Des bases de données de spectres SM/SM expérimentaux.^{15, 16} Ce sont les bases de données les plus répandues. Une grande quantité de ce genre de collections est disponible sur le site du GNPS (<https://gnps.ucsd.edu/ProteoSAFe/static/gnps-splash.jsp>). La base de données d'alcaloïdes indolomonoterpéniques constituée par notre équipe depuis 2015 fait aussi partie de ces collections (voir chapitre II).

⁸ T. Pluskal, S. Castillo, et al., *BMC Bioinform.*, 2010, **11**, 395.

⁹ F. Olivon, G. Grelier, et al., *Anal. Chem.*, 2017, **89**, 7836-7840.

¹⁰ F. Olivon, F. Roussi, et al., *Anal. Bioanal. Chem.*, 2017, **409**, 5767-5778.

¹¹ H. L. Röst, T. Sachsenberg, et al., *Nat. Methods*, 2016, **13**, 741.

¹² P.-M. Allard, T. Péresse, et al., *Anal. Chem.*, 2016, **88**, 3317-3323.

¹³ F. Allen, A. Pon, et al., *Nucleic Acids Res.*, 2014, **42**, W94-W99.

¹⁴ R. R. da Silva, M. Wang, et al., *PLoS Comput. Biol.*, 2018, **14**, e1006089.

¹⁵ D. Krug & R. Müller, *Nat. Prod. Rep.*, 2014, **31**, 768-783.

¹⁶ A. Bouslimani, L. M. Sanchez, et al., *Nat. Prod. Rep.*, 2014, **31**, 718-729.



1.4 Types de stratégie appliqués

L'information originale des réseaux moléculaires peut être complétée par d'autres types de données. Cela peut être fait de deux manières :

- Intégration d'une information complémentaire au réseau moléculaire dans le but de cibler d'une manière plus spécifique des molécules ayant une caractéristique en particulier, ou d'apporter un degré de compréhension supplémentaire à la représentation graphique de données CL-SM/SM, notamment dans le cadre de la comparaison d'échantillons.
- Construction du *network* séparément puis association de l'information relevée à celle obtenue à partir de l'utilisation d'autres techniques.

Dans le premier type de stratégie, des informations comme l'activité biologique des molécules peuvent être intégrées au réseau moléculaire. Dans ce contexte, envisageant la découverte de produits naturels bioactifs, les équipes de chimie des substances naturelles et de spectrométrie de masse de l'ICSN, avec les travaux de Florent Olivon *et al.*, ont considéré l'inclusion de l'information taxonomique et de résultats de tests biologiques dans les réseaux moléculaires. C'est ainsi qu'ils ont entrepris l'exploration raisonnée de l'espace chimique de *Bocquillonia nervosa* et *Neoguillauminia cleopatra* (Euphorbiaceae).¹⁷

En ce qui concerne le deuxième type de stratégie, l'information obtenue des réseaux peut être associée avec l'épigénétique ou le *genome mining*. Cette approche est surtout utilisée dans des études orientées vers l'identification de *clusters* de gènes à l'origine biogénétique de métabolites analogues, dans le but de faire des prédictions structurales des molécules partageant le même *cluster* dans un réseau moléculaire. C'est ainsi que l'équipe de recherche de Kleigrewe a appliqué cette méthode pour explorer la diversité structurale de cyanobactéries marines.¹⁸

1.5 Outils et stratégie visant à améliorer la pratique du *molecular networking*

En raison de l'application maintenant courante des approches de déréplication basées sur la technique du *molecular networking*, de multiples outils et algorithmes ont été développés par la communauté scientifique pour améliorer la façon d'annoter. C'est ainsi que des outils d'annotation *in silico* ont été conçus pour rendre plus informatives les réseaux moléculaires. Parmi ces outils, nous pouvons citer Sirius (disponible sur <https://bio.informatik.uni-jena.de/software/sirius/>).¹⁹ Cet algorithme analyse les *patterns* isotopiques (ou signatures isotopiques) des molécules pour déterminer des formules brutes.

¹⁷ F. Olivon, P.-M. Allard, et al., *ACS Chem. Biol.*, 2017, **12**, 2644-2651.

¹⁸ K. Kleigrewe, J. Almaliti, et al., *J. Nat. Prod.*, 2015, **78**, 1671-1682.

¹⁹ S. Böcker, M. C. Letzel, et al., *Bioinformatics*, 2008, **25**, 218-224.



De plus, les *patterns* de fragmentation des composés sont analysés en utilisant des arbres de fragmentation mis en ligne sur CSI:FingerID (disponible sur <https://www.csi-fingerid.uni-jena.de/>).²⁰

Un deuxième exemple d'outil d'amélioration de l'approche classique est la propagation de l'annotation *in silico* (NAP, pour *Network Annotation Propagation*).¹⁴ Il permet la génération *in silico* par MetFrag de structures potentielles dans le voisinage d'un nœud en particulier. Ensuite, les propositions sont classées en fonction d'un score calculé par l'outil MetFusion.

Un dernier outil récemment créé par Gregory Genta-Jouve et Yann Bauxis, appelé MetWork, se montre comme une méthode très puissante d'annotation *in silico* à partir des annotations expérimentales.²¹ Les molécules qui « *matchent* » sont métabolisées virtuellement par la plateforme en suivant des réactions sélectionnées par l'utilisateur, puis leurs spectres SM/SM sont générés *in silico* avec CFM-ID et comparés à ceux des molécules de l'échantillon.

La description et l'emploi de tous ces outils seront développés de manière plus approfondie dans la seconde partie de ce chapitre, conçue comme un article de perspectives pour le périodique *Natural Products Reports*. L'évolution de l'emploi du *molecular networking* pour la recherche de petites molécules naturelles sera présentée, avec une focalisation sur les différences de stratégies et de paramètres des équipes utilisatrices. Une lettre d'intention a été transmise au comité éditorial de *Nat. Prod. Rep.* et a obtenu une réponse favorable. Ce *highlight* est attendu pour le mois décembre prochain. La version soumise différera probablement légèrement de celle présentée ici.

²⁰ K. Dührkop, H. Shen, et al., *Proc. Natl. Acad. Sci. USA*, 2015, **112**, 12580-12585.

²¹ Y. Bauxis & G. Genta-Jouve, *Bioinformatics*, 2018, 10.1093/bioinformatics/bty864.



Stratégies de ciblage de produits naturels basées sur les réseaux moléculaires : Des approches différentes, un même objectif

Article soumis à *Natural Products Reports*

Résumé : Des progrès remarquables des outils bio-informatiques ont récemment fait progresser le domaine de la recherche sur les produits naturels (PNs), permettant aux chimistes de cette discipline d'effectuer un isolement orienté et efficace de ces molécules, en accélérant la hiérarchisation dans le flux du travail d'isolement. Parmi ces outils, la technique de réseaux moléculaire fondés sur la spectrométrie de masse tandem apparaît aujourd'hui comme une approche prometteuse pour dérégler les extraits naturels, entraînant une véritable révolution dans « l'art de l'isolement des substances naturelles » en accélérant le rythme de recherche dans ce domaine. Ce *highlight* illustre, à travers des exemples choisis et emblématiques, une nouvelle conception de l'isolement de PNs en faisant un parallèle entre les différentes philosophies sous-jacentes à l'utilisation des réseaux moléculaires pour cibler des PNs.





2. PUBLICATION ASSOCIÉE

Natural Product Reports

www.rsc.org/npr



article soumis, 2019

Natural Products Targeting Strategies Involving Molecular Networking: Different Manners, One Goal

Alexander E. Fox Ramos, Laurent Evanno, Erwan Poupon, Pierre Champy, and Mehdi A. Beniddir*

Abstract: Landmark advances in bioinformatics tools have recently enhanced the field of natural products research, putting today's natural product chemists in the enviable position of being able to perform an efficient discovery of these molecules, by expediting the prioritization of the isolation workflow. Among these advances, MS/MS molecular networking has appeared as a promising approach to dereplicate complex natural product mixtures, leading to a real revolution in the "art of natural product isolation" by accelerating the pace of research of this field. This *Highlight* illustrates through selected cornerstones studies the new thinking in NP isolation by drawing a parallel between the different underlying philosophies behind the use of MN in targeting NPs.





Natural Products Targeting Strategies Involving Molecular Networking: Different Manners, One Goal

Alexander E. Fox Ramos, Laurent Evanno, Erwan Poupon, Pierre Champy, and Mehdi A. Beniddir*

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Landmark advances in bioinformatics tools have recently enhanced the field of natural products research, putting today's natural product chemists in the enviable position of being able to perform an efficient targeting/discovery of previously undescribed molecules, by expediting the prioritization of the isolation workflow. Among these advances, MS/MS molecular networking has appeared as a promising approach to dereplicate complex natural product mixtures, leading to a real revolution in the "art of natural product isolation" by accelerating the pace of research of this field. This *review* illustrates through selected cornerstones studies the new thinking in natural product isolation by drawing a parallel between the different underlying philosophies behind the use of molecular networking in targeting natural products.

- 1 Introduction
- 2 Toward an efficient annotation of molecular networks
 - 2.1 Integration of spectrometric data
 - 2.1.1 Dereplication via in-house experimental MS/MS data
 - 2.1.2 Dereplication *via in silico* MS/MS data
 - 2.2 Harnessing mass shift differences
 - 2.3 Integration of functional annotations
 - 2.3.1 Layering of biological data
 - 2.3.2 Layering of multi-informational data for large collections
- 3 Combination of MN with other techniques
 - 3.1 Biochemometrics
 - 3.2 Mass spectrometry imaging
 - 3.3 Genome mining and metabologenomics
 - 3.4 Chemical epigenetics
 - 3.5 Stable isotope labeling
- 4 Latest improvements and tools implemented in the molecular networking practice
 - 4.1 Data pre-processing
 - 4.2 Data organization
 - 4.3 New reporting standards for MS data
 - 4.4 Advanced molecular networking annotation tools
- 5 Conclusions
- 6 Conflicts of interest
- 7 Acknowledgements
- 8 Notes and references

1 Introduction

The large amount of knowledge derived from the comprehensive study of natural products (NPs) has provided society with a wealth of fundamental insights as well as applied tangible advancements.¹

Remarkably, landmark advances in bioinformatics tools and analytical chemistry, particularly in mass spectrometry (MS), have recently enhanced the field of NP research,² putting today's practicing chemists in the enviable position of being able to efficiently speed up the NPs discovery process.^{3, 4} In this context, molecular networking (MN) has proved to be a very efficient tool to rapidly identify new NPs within complex mixtures. This emerging computer-based approach allows to visualize and organize tandem MS/MS data sets and to automate database searches for specialized metabolite identification.⁵ Since its introduction in 2012, this dereplication technique has totally revolutionized the "art of NP isolation" enabling the transition from the traditional "grind and find" model to a streamlined hypothesis-driven targeting of NPs. Although the breadth of MN applications has recently been reviewed by Dorrestein *et al.*^{6, 7} and Pevzner *et al.*⁸, we wish to disclose herein a critical assessment of the several NPs targeting strategies involving MN without paying attention to a specific class of NPs, and clearly dedicated to NP chemists from a practical point of view. Therefore, this *review* is intended to illustrate through selected cornerstones studies the new thinking in NP isolation by drawing a parallel between the different underlying philosophies behind the use of MN in targeting NPs, as developed by an increasing number of research groups. As a consequence of the overwhelming number of these studies, this *review* is not comprehensive, therefore we apologize for omitting many contributions to the advance of this exciting research field. As a general guideline to understand the organization of this *review*, the first section will cover the different strategies that have been developed in order to perform efficient annotations of molecular networks. The second section will discuss the different studies that combined MN to other techniques. The last section, will give an overview of the latest improvements that have been

Équipe « Pharmacognosie-Chimie des Substances Naturelles », BioCIS, Univ. Paris-Sud, CNRS, Université Paris-Saclay, 5 rue J.-B. Clément, 92290, Châtenay-Malabry, France. E-mail: mehdi.beniddir@u-psud.fr

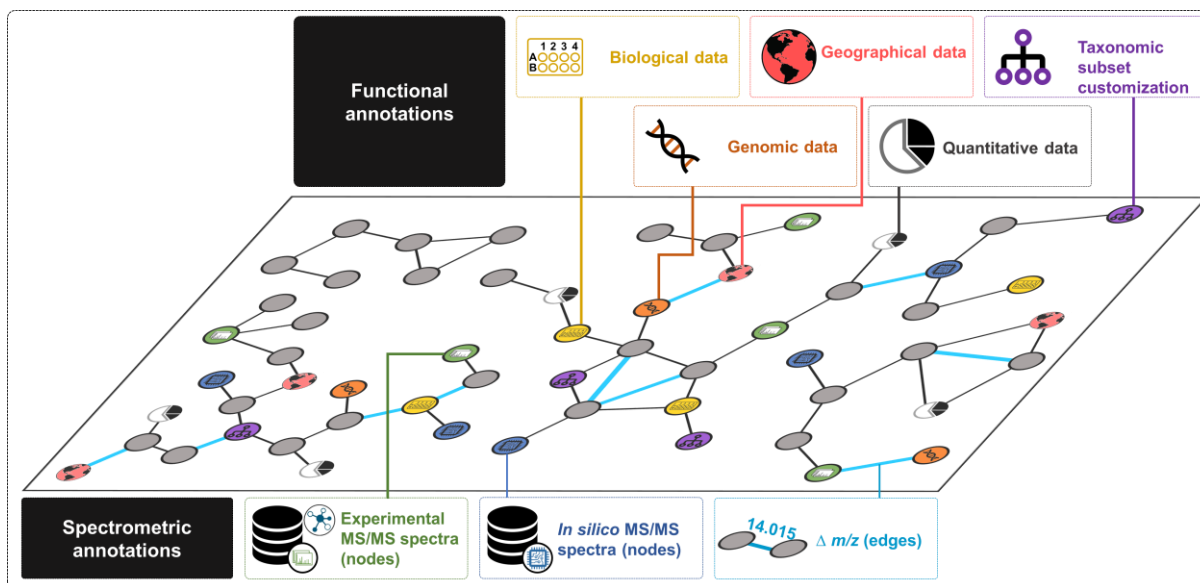


Fig. 1 Examples of multiple sources of data for the annotation of a molecular network.

recently implemented to ameliorate and sharpen the practice of MN.

2 Toward an efficient annotation of molecular networks

From an applied standpoint, a molecular network represents a road-map that can be further enriched by multiple functional annotations including different kind of data such as biological, taxonomical, or spectrometric, etc (Fig. 1). In other words, an efficient annotation of molecular networks should not only shed light on unexplored region of the chemical space but also fire the starting gun to deploy the isolation efforts toward targeted NPs. Interestingly, the quest for tackling the issue of molecular networks annotation fuelled a number of successful and creative endeavors in the NP research community. Some of them are summarized thereafter.

2.1 Integration of spectrometric data

The success of a MN-based dereplication relies greatly on the quality and the availability of MS/MS data. Even though the Global Natural Product Social Molecular Networking (GNPS,⁹ available at: <https://gnps.ucsd.edu/ProteoSAFe/static/gnps-splash.jsp>) community has already contributed more than 70000 annotated MS/MS spectra, the dereplication process annotates only a limited number of nodes. To overcome this issue, alternative solutions were sought. The latest studies related to these alternatives are reported below.

2.1.1 Dereplication via in-house experimental MS/MS data

One of our first forays in the MS/MS MN-guided isolation of NPs sought to explore the overlooked monoterpene indole alkaloids (MIAs) chemical space related to some understudied Apocynaceae plant. To address the key issue of limited node

annotations due to the low occurrence of this family of NPs in the GNPS library, we embarked on the implementation of an in-house MS/MS database for these compounds. At the outset of our endeavors, the latter was constituted of 55 Gentianales alkaloids and covered more than 50% of the 42 known MIA skeletons. Today, this database is available under the generic name of Monoterpene Indole Alkaloids DataBase (MIADB) and contains 172 MS/MS spectra that have been deposited on the GNPS library and Metabolights (study identifier: MTBLS142).¹⁰ *Geissospermum laeve*, a previously studied Apocynaceae species native to northern South America, was the subject of a study employing this strategy.¹¹ In order to prioritize the isolation workflow toward previously undescribed MIAs, an alkaloid extract of the stem bark was analyzed by HPLC-Q-TOF-MS in positive ion mode, and then the obtained MS/MS data along with the MS/MS spectra of the aforementioned in-house database were then submitted to the GNPS online platform and organized as molecular networks. As a first dereplication step, the metabolites contained in the extract were annotated by the GNPS library, affording only one hit, namely yohimbine, which did not match the standard from the MIADB. This finding supported the interest of dereplicating the extract against our in-house database. Besides, on the whole molecular network of the alkaloid extract of *G. laeve* annotated by the MIADB, two clusters appeared of particular interest, due to the presence of matches with the in-house database. The first cluster was characterized by the presence of monomers together with the already-described and structurally related bisindoles geissospermine¹² and geissosolimine.¹³ Interestingly, these two molecules were linked to an unidentified compound that resulted to be an oxidized analogue of geissospermine, assigned as 3',4',5',6'-

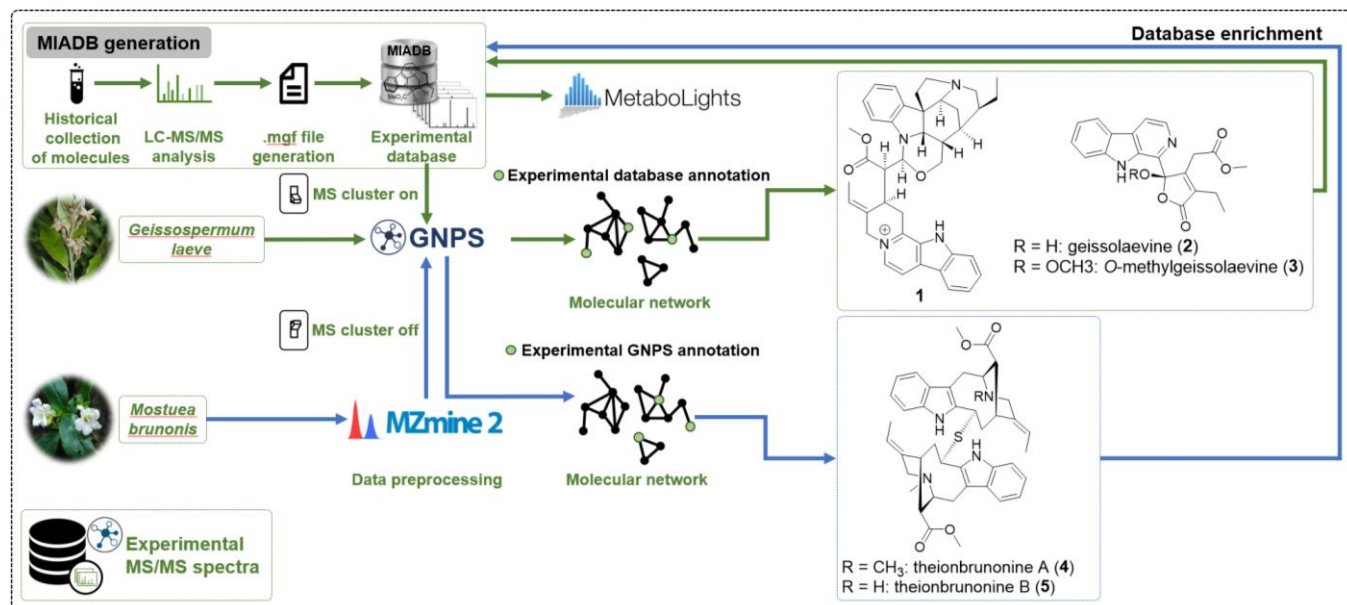


Fig. 2 MS/MS guided isolation of monoterpene indole alkaloids using MIADB.

tetrahydrogeissospermine (**1**) (Fig. 2). In the second cluster, serpentine, a MIA that has never been described in the genus *Geissospermum*, was connected to several nodes, indicating the likely presence of unexpected analogs. Satisfyingly, exploration of this cluster subsequently allowed the targeting, identification and isolation of two new NPs: geissolaevine (**2**) and *O*-methylgeissolaevine (**3**) (Fig. 2), which remarkably constituted the first examples of natural β -carboline alkaloids bridged to a butenolide ring. This discovery exemplifies how an efficient annotation of a molecular network may allow the targeting of unexpected chemistries from a previously investigated plant. Notably, in this study, the level of confidence of each match was assessed using Schymanski rules¹⁴ by comparison of HRMS data, MS² spectra, and retention times (RT). As a second example of our MIAs-discovery program using MN, we were, recently, able to target and characterize, theionbrunonines A and B (**4** and **5**) (Fig. 2), the first examples of monoterpene bisindole alkaloids linked by a thioether bridge from the stems of *Mostuea brunonis* (Gelsemiaceae).¹⁵ In this study, the MS/MS data were preprocessed by MZmine 2¹⁶ prior to their submission to the GNPS platform. This crucial step will be detailed in the last section (4.1) of this review. Furthermore, unlike the preceding study, the obtained molecular network was annotated by the MIADB hosted by the GNPS, rather than including the MS/MS spectra related to the standards into the MS/MS data of the

extract. In a similar spirit of seeking to achieve streamlined targeting of NPs using MN, our group was attracted by the chemical diversity produced by *Dactylosporgia metachromia*,^{17, 18} a Polynesian marine sponge from the Thorectidae family. This species has been extensively studied and is known to produce several compounds of the quinone sesquiterpenes and sesquiterpene benzoxazoles series.¹⁹ Among these molecules, ilimaquinone (**6**) (Fig. 3) is a well-known quinone sesquiterpene, displaying a large array of biological properties.^{20, 21} Our plan was to elaborate an efficient approach to target new analogs of this molecule. To attain this, we harnessed the clusterizing power of MN in order to match structurally related NPs using a well-tailored semisynthetic phishing probe (**7**) (Fig. 3), prepared from ilimaquinone (**6**).²² This semisynthetic compound was a typical zwitterionic quinonoid, with a *para*-benzomonoquinoneimine system, displaying a blue color. It should be noted that this motif was notably unknown in natural substances. Ethyl acetate extracts of *D. metachromia* collected from Fakarava and Rangiroa archipelagos (French Polynesia) as well as the above mentioned semisynthetic phishing probe (**7**), were analyzed by HPLC-Q-TOF-MS in positive ion mode. The obtained fragmentation data were organized by MN. A rapid exploration of the generated network allowed the identification of the reference substance connected to several nodes, suggesting the presence of potential natural analogs.

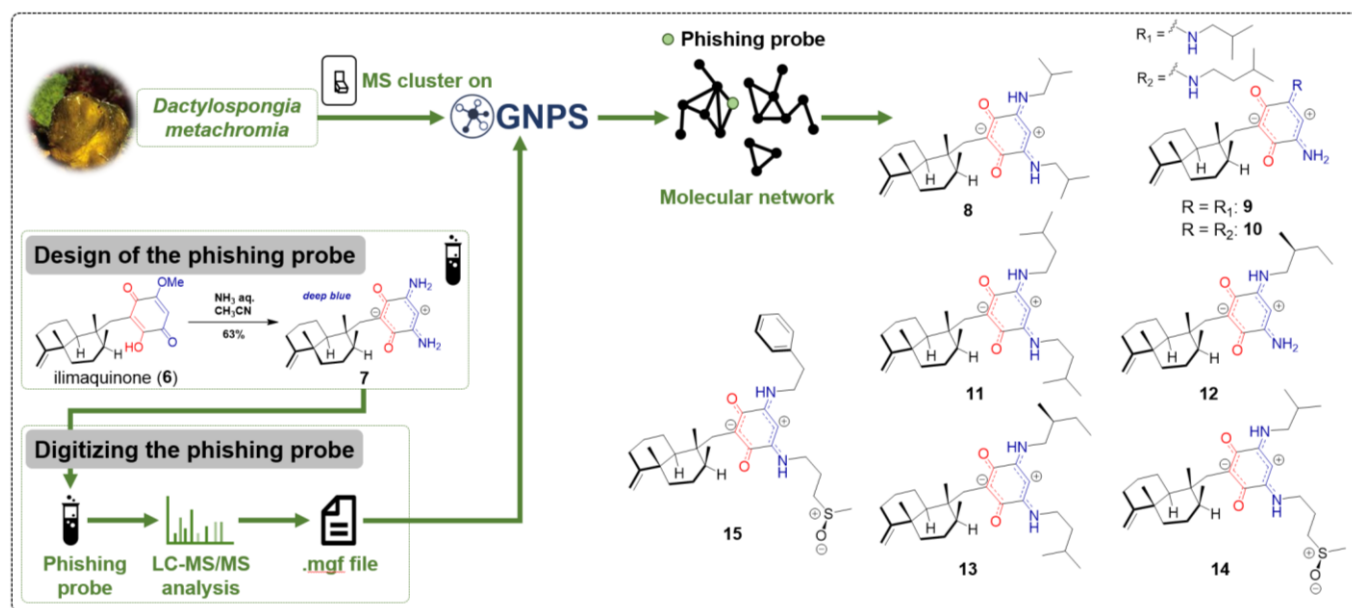


Fig. 3 Overview of the "Chemistry First" approach used for the discovery of dactylocyanines A-H (8-15).

Indeed, the MS/MS data of the incorporated phishing probe acted as "seed" spectra, providing a solid anchor in the global molecular network. Therefore, the potential natural analogs were targeted for further isolation and structural elucidation. Ultimately, this "Chemistry first" approach in combination with MN allowed the streamlined isolation of eight new natural compounds (dactylocyanines A-H) (8-15) (Fig. 3) bearing the anticipated zwitterionic diamino-*meta*-quinonoid blue scaffold.²³ This study constitutes an interesting example of anticipated NPs isolation.

In a similar line of research, Gerwick *et al.* explored the chemical space of 20 crude extracts of cultured marine cyanobacteria using quantitative MN annotated by in-house MS/MS data.²⁴ As these species are known to produce in majority low-polarity metabolites,²⁵ they underwent lipid extraction using a 2:1 mixture of CH₂Cl₂/MeOH. The extracts were then analyzed by HPLC-IT-MS with an ESI source working in positive mode. Besides, 60 pure marine NPs and NP analog samples were likewise analyzed to act as reference substances for further dereplication.

In order to map the semi-quantitative distribution of the compounds across the different samples on the MN, a script, named TORTE (for Tandem-MS Origin Tracing Engine) was applied. This tool tracks MS/MS data file behind each node of the network to generate an extracted ion chromatogram (XIC). Each ion is represented as a chromatogram peak and its corresponding area under the trace is then calculated and stored in an annotation table, with one row for each node and a column for each MS/MS data file. This allowed determining the distribution of each metabolite within the studied species, appearing in the network as pie charts with color tags for each sample. Application of this approach allowed the prompt recognition of eight matching metabolites related to dolastatines, veraguamides and barabamides series. The nodes of these well-known molecules were connected to numerous

potential analogs, among which 30 were identified. Notably, the most interesting finding in the generated network was the identification of the structurally intriguing antifungal and antitumoral lipopeptides malyngamides C (16), C acetate (17), (Fig. 4) H, I, and K as well as their producers, namely: a black *Morea* sp. and *Okeania hirsuta*. Moreover, the application of the TORTE script allowed the quantification of the molecules produced by each of them, identifying *O. hirsuta* as the most prolific producer of malyngamide C (16). As the biosynthetic pathway of 16 and 17 was enigmatic, this finding allowed to scale-up the culture of *O. hirsuta* for genome sequencing. Recently, these efforts allowed to decipher the biosynthetic pathway of type A malyngamides.²⁶

As a comment regarding the preprocessing step of the MS/MS data related to the preceding studies, it should be noted that in the work related to the phytochemical study of *G. laeve* the MS/MS data were preclusterized and converted from the .d (Agilent format) to .mgf format *via* the Auto-MS/MS (AMS) algorithm implemented in the MassHunter software prior to the GNPS upload, therefore, the multiple collision energies used in the analyses were averaged. In Gerwick's *et al.* work, the data files were converted from the .RAW format (Thermo standard data-format) to the .mzXML format using the software MS-Convert.²⁷

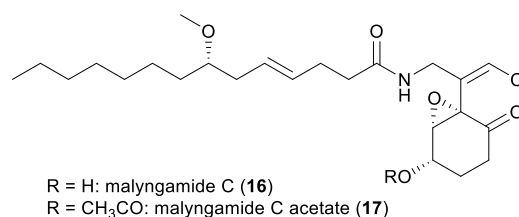


Fig. 4 Structures of malyngamide C (16) and malyngamide C acetate (17).

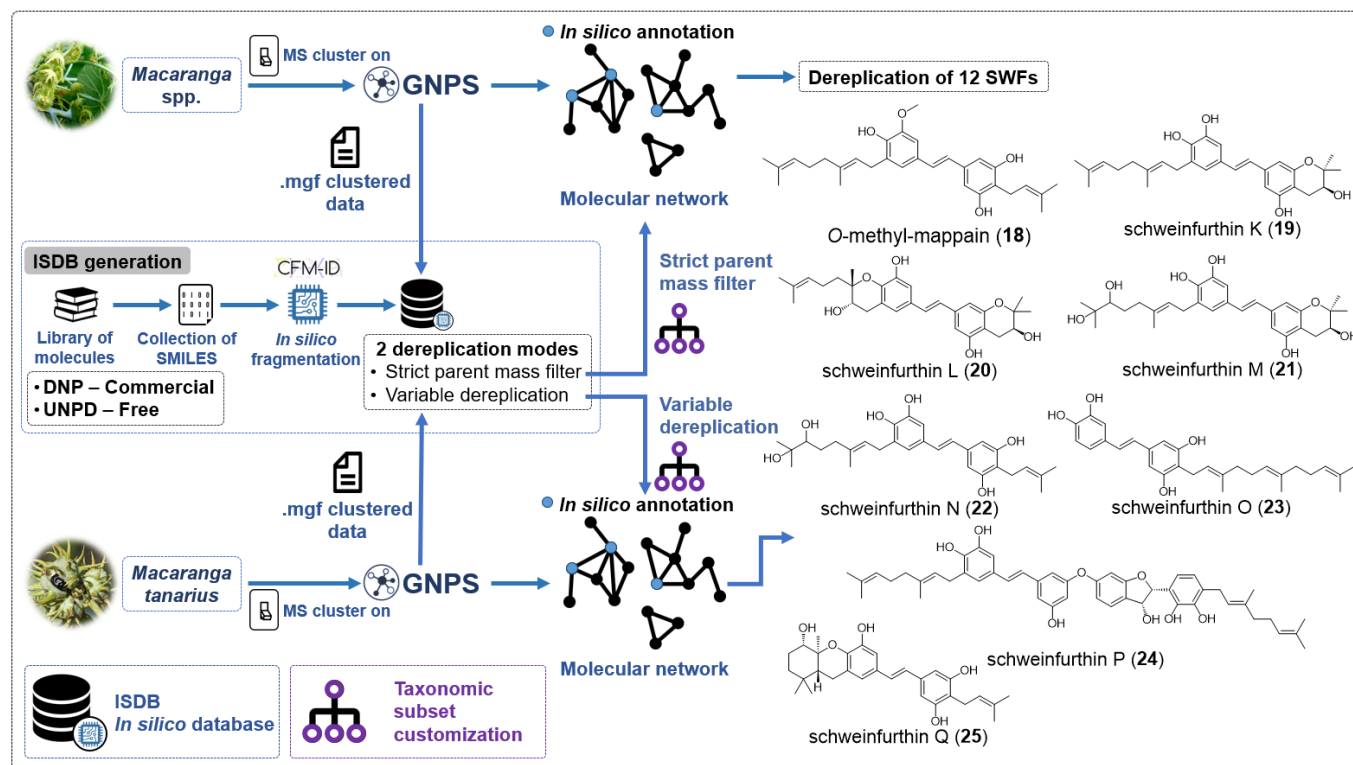


Fig. 5 ISDB-based dereplicative pipeline and structure of *O*-methyl-mappain (18) and SWFs K-Q (19-25).

As it was described in the previous studies, the incorporation of reference substances in a molecular network offers initial focal points that will match or be linked to identical or similar compounds in the mixture. However, this matching process is inevitably tied to turning ON the MS-Cluster²⁸ option, with a minimum cluster size of 2, when running the MN process. As shown in the next sections, new data-preprocessing workflows require turning OFF the MS-Cluster tool with a minimum cluster size of 1 in order to enhance MN reliability.^{29, 30} Consequently, to apply these new workflows, it is recommended to upload the experimental MS/MS data on the GNPS library in order to perform an efficient annotation of the generated molecular networks. More details on these workflows will be discussed in the last section (4.1) of this review.

2.1.2 Dereplication via *in silico* MS/MS data

Alternative methods have been developed to surpass the limitations imposed by the size of available fragmentation data hosted by existing libraries.³¹ In this context, *in silico* fragmentation approaches have been used to overcome this issue. In response to such needs, Wolfender *et al.*³² implemented an extensive *in silico* database (ISDB) from the

MS/MS data of more than 220000 NPs indexed in the *Dictionary of Natural Products* (DNP, <http://dnp.chemnetbase.com/faces/chemical/ChemicalSearch.xhtml>). The construction of this database was achieved using the SMILES (Simplified Molecular-Input Line-Entry System) input of all the non-permanently charged NPs of the DNP v. 24. The resulting collection of 221771 entries was then fragmented *in silico* using the machine learning-based tool CFM-ID v. 1.10³³ (available at: <http://sourceforge.net/projects/cfm-id/>). The generated MS/MS spectra obtained at low, medium, and high collision energies were merged and converted into .mgf files. However, as DNP is a commercial product, the generated ISDB could not be publicly shared. Therefore, the authors used the freely accessible Universal Natural Products Database (UNPD)^{34, 35} (available at: <https://github.com/clzani/DEREP-NP>) to generate another CFM-ID-based MS/MS *in silico* database with a total of 170602 compounds after filtration of duplicates and permanently charged substances.³⁶ This database was named UNPD-ISDB (available at: <http://oolonek.github.io/ISDB/>). One advantage of this approach is that the user can customize the *in silico* database in respect to a specific class of compounds.

The first step of the ISDB workflow is to generate a molecular network using the classic process. After the online treatment, the data can be dereplicated against the *in silico* reference library using the open-source tool Tremolo³⁷ (freely available at: <http://proteomics.ucsd.edu/Software/Tremolo/>) which allows comparison between the MS/MS spectra. Then, a “similarity score” threshold has to be defined to assess the comparison made between experimental and *in silico* MS/MS data. This value must not be too high, since generated *in silico* spectra only approximate experimental ones. Thus, for this library search, a similarity score threshold of 0.2 is usually used. When using the Tremolo tool, dereplication against the ISDB can be executed in two modes: (i) a “strict parent mass (PM) filter” mode, where the PM tolerance is set to 0.005 (a very stringent one, allowing to compare parent ion nodes in the network with the reference library; (ii) and a “variable dereplication mode”, where a value of 200 Da is advised to allow spectral matching with analogs having different parent ion mass but sharing MS/MS spectral similarities.

The first dereplication mode (strict PM filter mode) was applied to the exploration of the chemical space of Euphorbiaceae of the *Macaranga* genus, aiming at the identification of new schweinfurthins (SWFs), prenylated stilbenes endowed with potent antitumoral activities.³⁸ Crude extracts of 21 species of this genus were analyzed by UHPLC-HESI-Q-Orbitrap, then dereplicated against the DNP-ISDB. In this case, a subset of the ISDB restricted to only Euphorbiaceae entries was used in order to obtain more refined dereplication results. This workflow allowed to rapidly identify a cluster containing 12 already-described members of the SWF family (SWFs A-J, mappain, and vedelianin) and a potentially novel metabolite. The latter was purified from the fruits of *M. tanarius*, and identified as an *O*-methylated mappain derivative (**18**) (Fig. 5). As a continuation of these endeavors, Litaudon *et al.*³⁹ were further able to target and isolate seven new SWFs (SWF K-Q (**19-25**)) (Fig. 5) from the same plant, using the Tremolo tool in variable dereplication mode.

Since its publication, this *in silico* dereplicative pipeline (*i.e.*, ISDB) has been used by different research teams to dereplicate samples from various sources. Klein-Júnior *et al.* applied this approach to the alkaloid extract of the leaves of *Palicourea sessilis* (Rubiaceae).⁴⁰ UHPLC-ESI-Q-Orbitrap MS/MS data were acquired and organized as a network, which was annotated against the DNP-ISDB, enabling the rapid identification of a MIAs-containing cluster. The MS/MS spectra derived from the alkaloid extract were annotated against the molecules reported in the Rubiaceae and/or Loganiaceae families; using both the “strict PM filter” and the “variable dereplication” modes. This allowed the recognition of compounds bearing a strictosidine backbone within the identified MIAs cluster, making a total of 14 dereplicated molecules in this group. Within this cluster, some ions were targeted for isolation, based on the possibility that they were strictosidine-type compounds. This hypothesis was further confirmed by structural elucidation (**26-29**) (Fig. 6).

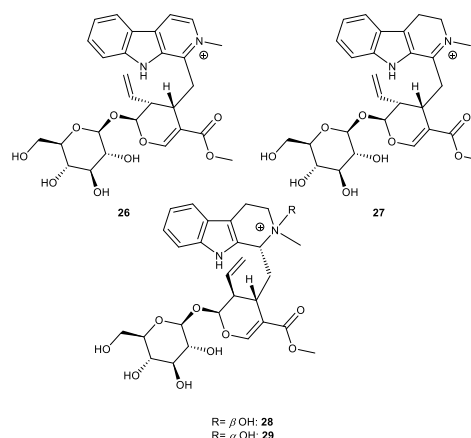


Fig. 6 Structures of targeted strictosidine like compounds (**26-29**).

Recently, the GNPS web platform introduced an online molecular network analysis tool called Network Annotation Propagation (NAP⁴¹), in which the *in silico* fragmentation algorithm MetFrag⁴² is combined with a network topological consensus and reranking of *in silico* annotations. This way of expanding the annotation of molecular networks has been recently applied by Dorrestein and Kang *et al.*⁴³ for the phytochemical investigation of the twigs of *Sageratia theezans* (Rhamnaceae). This study allowed the dereplication of several triterpenes known in the species and the streamlined isolation of 3-dicoumaroyl lignans (**30-32**) and 6-dicoumaroyl neolignans (**33-38**) (Fig. 7).

Beyond the apparent efficiency of these *in silico* fragmentation algorithms in the annotation of molecular networks, these tools still need to be improved. As such, machine learning-based methods such as CFM-ID can be trained using diverse and large training sets, leading to improve its accuracy.⁴⁴

2.2 Harnessing mass shift differences

Another way to unearth valuable information from a molecular network is to exploit *m/z* differences between related molecules. This strategy, coined meta-mass shift chemical (MeMSChem) profiling, has been recently proposed by Hartmann *et al.*⁴⁵ It identifies and annotates known chemical groups such as H₂, CH₂, COCH₂, etc. that could be linked to specific biochemical transformations. Hartmann *et al.* applied MeMSChem profiling to a dataset derived from an LC-MS/MS analysis of seven coral reef holobiont types collected in the Line Islands. After the generation of a molecular network, redundant mass differences were subsequently mined and annotated to known chemical groups when possible. Then, the MS/MS-based molecular features associated with these redundant mass shifts were quantified from the MS scan of the parent molecule using Optimus software (accessible at: <https://github.com/MolecularCartography/Optimus>). An examination of the acquired molecular features revealed that distinct mass shifts patterns could be ascribed to specific holobiont. Interestingly, by focusing on the differences in mass shifts profiles between related molecules, MeMSChem offered an efficient way to expand MN annotation beyond the

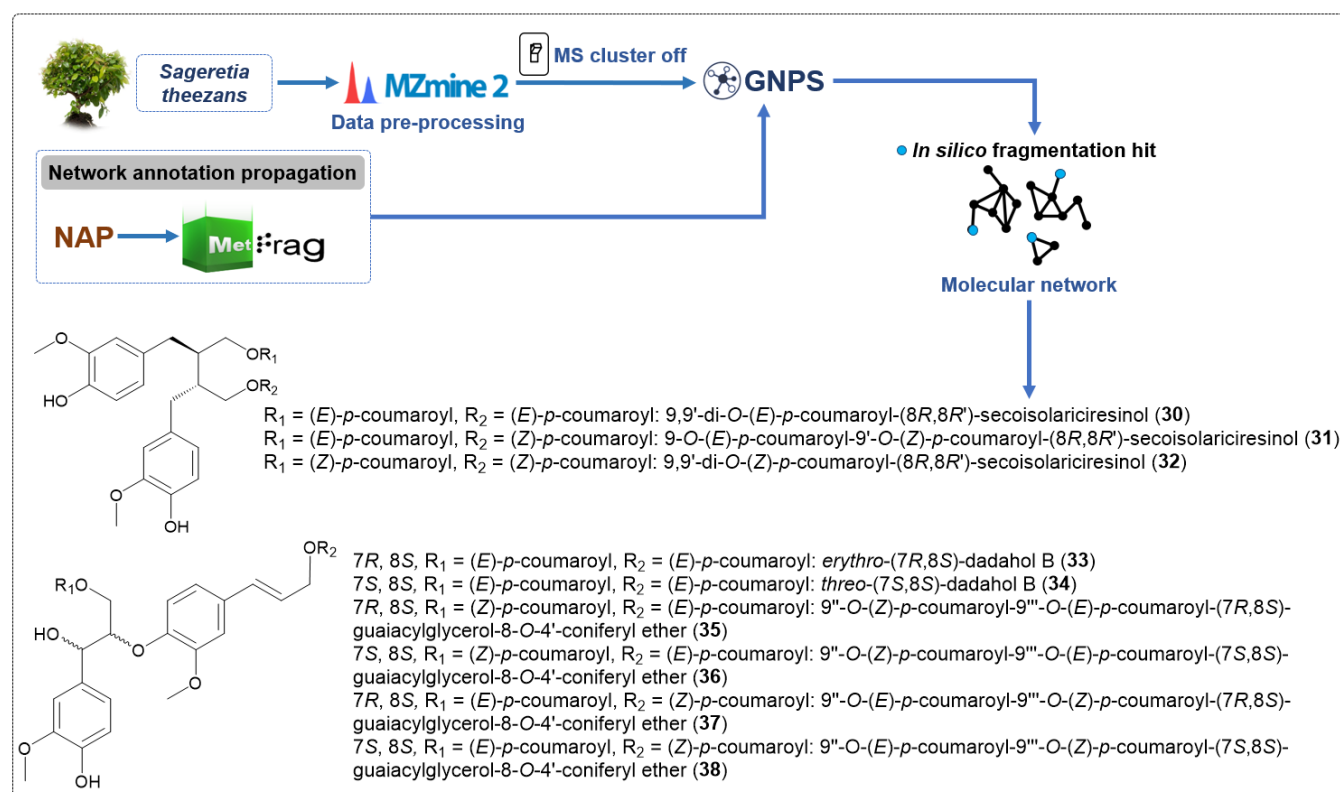


Fig. 7 NAP-based annotation of molecular network and structures of targeted compounds (30–38).

systematic spectral matching against reference libraries. Even though, this study did not result in the targeting of any NPs, one can easily imagine how MeMSChem profiling might be integrated in the NP isolation process.

2.3 Integration of functional annotations

Whereas the previous section dealt with the integration of spectrometric data to further illuminate molecular networks, this section will discuss recent articles where functional data were integrated in the networks such as biological and/or taxonomical data, culture conditions, extraction methods, and even geographical patterns, in order to target specific NPs.

2.3.1 Layering of biological data

For decades the discovery process of NPs has relied mainly on a bioactivity-based workflow, generally driven *via* iterative “extract-test-fractionate-test-purify-elucidate-test” cycles. This methodology has long been the gold standard in NPs research and has resulted in the discovery of important drugs, including camptothecin, paclitaxel, artemisinin, and vinblastine.⁴⁶ Despite these historical successes, the bio-guided process faces a number of challenges that affect the relevance of NP

research in modern biomedical science.⁴⁷ Among these, the most glaring of concession steps: rediscovery of known compounds. In this regard, new methods capable of focusing the isolation process on novel bioactive scaffolds are needed. In this context, new MN approaches have been developed in order to identify these substances efficiently in complex natural mixtures. Recently, Gerwick *et al.*, explored the chemical space of marine cyanobacterium from the *Symploca* genus (Phormidiaceae)⁴⁸ which had been extensively studied before, yielding multiple bioactive molecules, such as dolastatin 10; largazole, a cyclic depsipeptide, and santacruzamate A, a SAHA analog (both of them HDACis); symplocin A, a linear peptide (a potent cathepsin E inhibitor); symplocamide A, a cyclic depsipeptide (a chymotrypsin inhibitor). The CH₂Cl₂/MeOH extracts from 10 samples of *Symploca* spp. from different geographical origins (American Samoa, Saipan, Panama, and France) were analyzed using a HPLC-ESI-Q-TOF. The resulting data were then used to conceive a molecular network, whose annotation by the GNPS library allowed the identification of multiple families of molecules, such as chlorophyll derivatives and analogs of the bastimolides (a class of macrolides), dolastatins, and

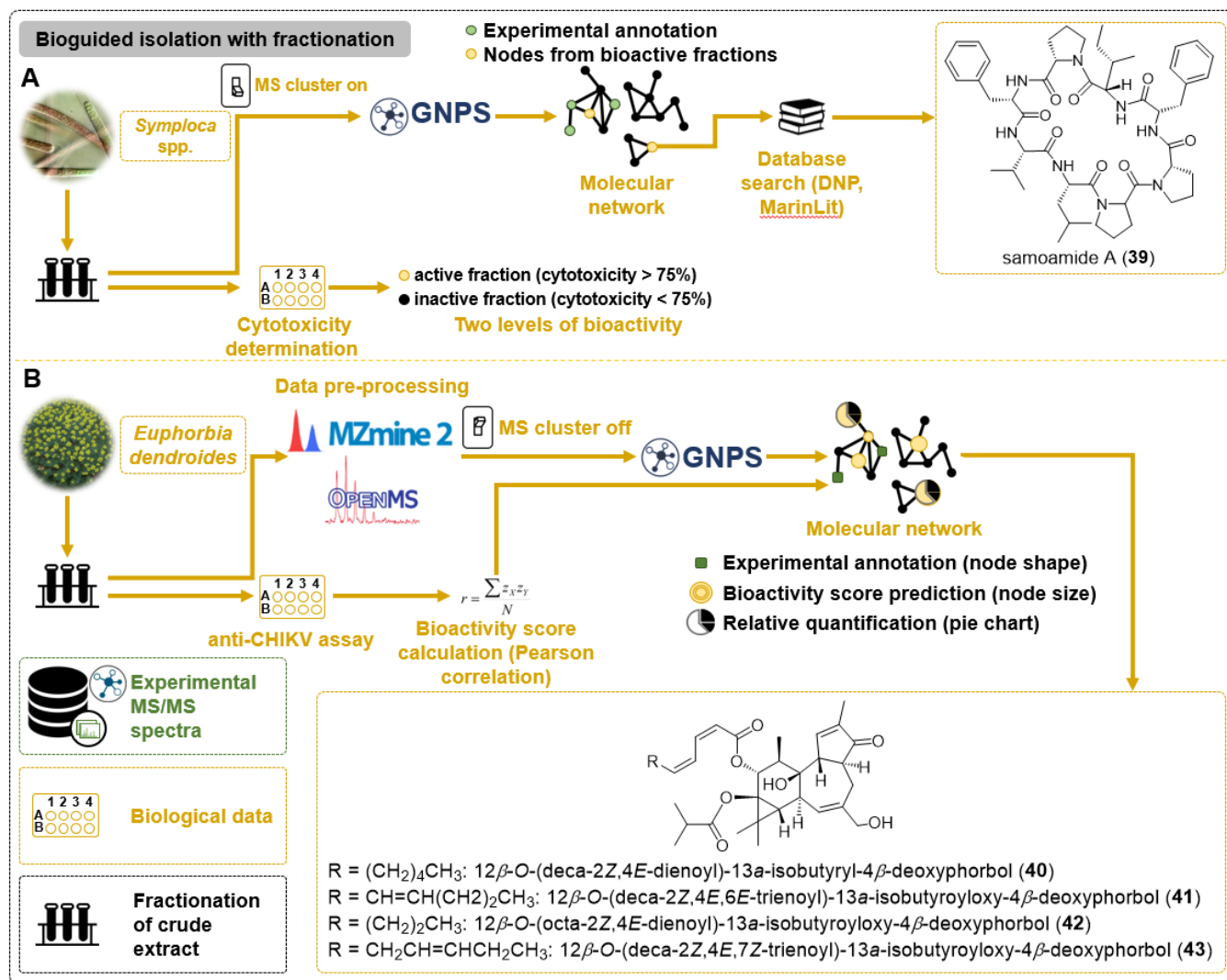


Fig. 8 Layering of biological data over molecular networks. A: targeted isolation of samoamide A (39) from *Symplocos* spp. B: targeted isolation of diterpenoids (40–43) from *Euphorbia dendroides*.

viequeamides (a family of cyclic depsipeptides). In order to complement the information drawn from the resulting network, all the *Symplocos* spp. crude extracts and their fractions were subjected to *in vitro* cytotoxicity testing in cancer cell lines. These biological data were integrated in the molecular network to target the most active fractions. These data were then combined to the results arising from the topology of the molecular network. In other words, the nodes that showed a potent cytotoxicity but were present in samples that contained dereplicated nodes were excluded from further study. Applying these criteria allowed the identification of a cluster of two nodes with a compound of interest from an American Samoan sample. Its MS/MS spectrum was unrelated to any compound present in the GNPS library. Further query of DNP and MarinLit (<http://pubs.rsc.org/marinlit>) databases for its exact mass yielded unlikely candidate molecules. Thus, this likely new metabolite corresponded to a novel cyclic octapeptide that was named samoamide A (39) (Fig. 8-A). As a comment on this study, although biological data have been integrated to the

molecular network to target bioactive compounds, this approach still required the so-called “fractionate-test” iterative cycles.

In another recent article, the concept of “Bioactivity-based molecular networking” (freely available at: https://github.com/DorresteinLaboratory/Bioactive_Molecular_Networks) was introduced by Dorrestein *et al.*⁴⁹ and consisted of three-steps:

(i) Acquisition and processing of the LC-MS/MS data using popular LC-MS feature detection software such as OpenMS,⁵⁰ or MZmine 2¹⁶ toolbox, to detect and relatively quantify the ions present in the MS/MS spectra (ii) Calculation of a bioactivity score using the Pearson correlation (a measure of the linear correlation between two variables X and Y), taking into account the ion intensity in the samples (X) and the bioactivity level of each sample (Y) (iii) Generation of a molecular network from the MS/MS data using the GNPS platform, in order to annotate detected molecules using the GNPS spectral library and taking into account the predicted bioactivity scores. The aim is the identification of clusters with

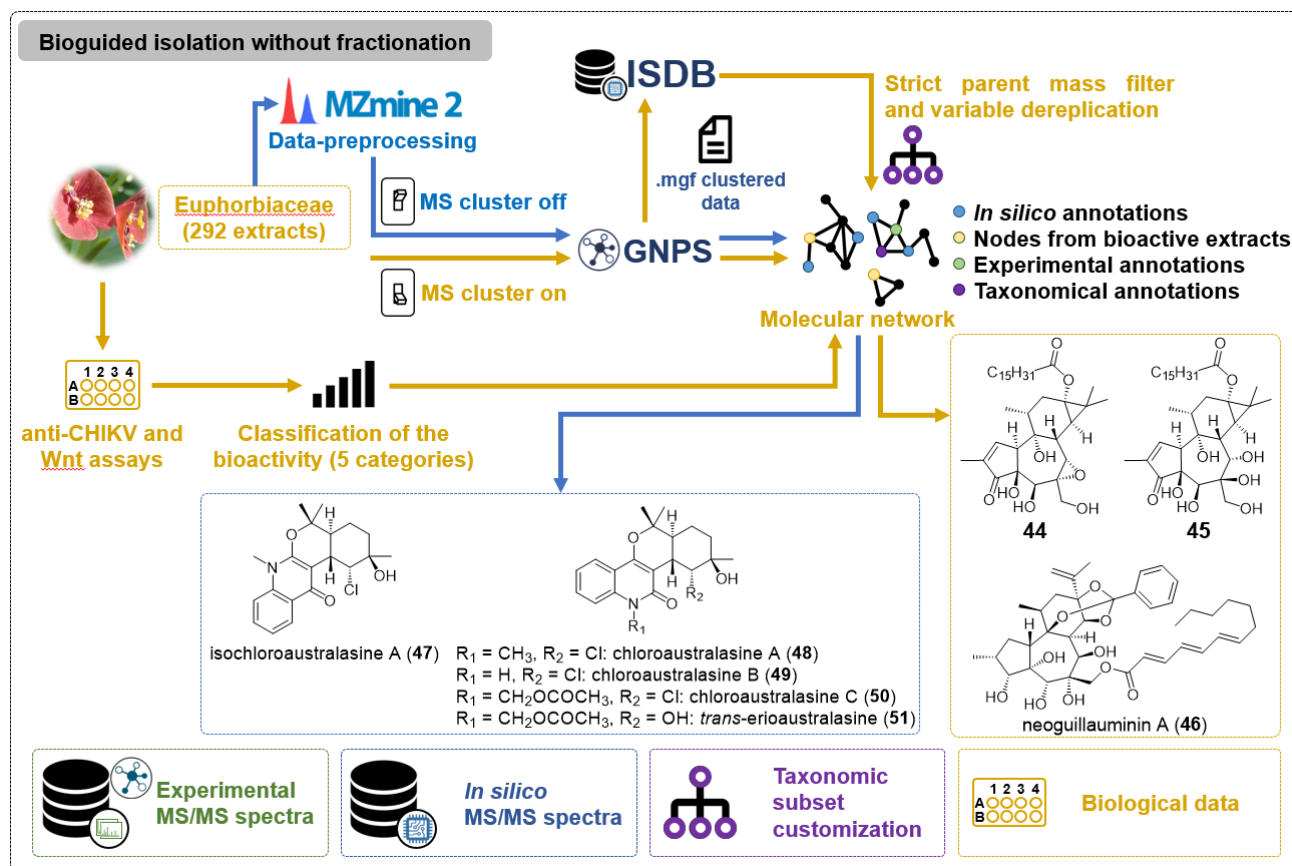


Fig. 9 Layering of multi-informational data over molecular networks and structures of targeted compounds (44–51).

a high frequency of bioactive candidates, which could evidence the presence of a common pharmacophore. This bioactivity-based dereplication pipeline was applied to reinvestigate a latex extract of *Euphorbia dendroides* (Euphorbiaceae). In a previous study, this plant demonstrated potent antiviral activity against Chikungunya virus (CHIKV) replication.⁵¹ However, none of the isolated compounds showed a selective antiviral activity against CHIKV. In this context, 18 fractions from the extract were analyzed by HPLC-IT-Orbitrap. In parallel, these samples were subjected to anti-CHIKV assays, in order to be able to include the bioactivity information in the resulting MN. These two types of data were then processed using the “Bioactivity-based molecular networking” workflow. In the resulting network, each node showed three different types of information: the bioactivity score prediction (reflected in the size of the node); the spectral annotation (symbolized by the node shape); and the relative quantification (represented by the node content, in the form of a pie chart, outlining the abundance of the molecule across the fractions. This feature

was used as an indicator of the selectivity of the antiviral activity and allowed the annotation of the clusters that contained molecules with the best bioactivity scores, constituting a total of 8.4% of the molecular network. A cursory examination of the network led to target four compounds owing to their significant bioactivity score and lack of spectral annotation matching against the GNPS library (40–43) (Fig. 8-B). The examination of their MS/MS fragmentation pattern pointed at deoxyphorbol esters. Ultimately, compounds 41 and 42 were found to be the most potent and selective CHIKV replication inhibitors with effective concentration (EC_{50}) values of 0.40 μM and 0.60 μM , respectively.

2.3.2 Layering of multi-informational data for large collections

In addition to the preceding examples of biological data integration, Litaudon *et al.*⁵² proposed a new approach that combines the layering of taxonomical and biological data to

molecular network obtained from large plants collection in order to prioritize bioactive NPs identification and isolation. The potency of this workflow was applied to explore the chemical diversity of 292 extracts from leaves, bark, twigs, whole plants and fruits of 107 New Caledonian Euphorbiaceae species. These samples were subjected to LC-MS² analyses using an UHPLC-HESI-Q-Orbitrap. In parallel the extracts were submitted to two biological evaluations: (i) CHIKV cell-based assay and (ii) oncogenic Wnt⁵³ signaling pathway assay. The obtained LC-MS² data were organized as a global molecular network constituted by 88687 nodes that were grouped in 7840 clusters. The resulting network was then annotated against a subset of the DNP-ISDB restricted to Euphorbiaceae entries,³² as described in the section 2.1.2. In this study, the ISDB workflow was executed following the two modes: "strict PM filter" mode and "variable dereplication" mode.

The integration of the bioactivity information in the resulting molecular network was based on a comparison between bioactive and inactive extracts. For this, the extracts were first classified according to their level of activity (reported as IC₅₀ or EC₅₀), applying color tags to each level of bioactivity for each biological tests. This differs from the Gerwick *et al.*⁴⁸ study, where extracts were divided only in two categories (active = cytotoxicity > 75%; inactive = cytotoxicity < 75% at 1 µg/mL), resulting in a more informative molecular network. This color mapping allowed a simple visualization of the clusters composed exclusively by ions coming from bioactive extracts, thus showing a common bioactive scaffold. As this network was very complex, a first filter allowed to explore it more efficiently. For this, a subnetwork was created for each of the determined biological activities where only nodes with at least two occurrences within the analyzed samples (thus, with at least two scans), and identified in very active extracts (IC₅₀ < 50 µg/mL in Wnt assay and EC₅₀ < 10 µg/mL in CHIKV assay). Finally, nodes within the network not having at least four neighbors at a distance of 4 were excluded. Applying these filters resulted in a molecular network that was easier to explore, as the number of nodes was reduced from 7840 to 192 and to 380, for the Wnt and CHIKV subnetworks, respectively. Envisaging a better mapping of the nodes within the generated subnetworks, taxonomical data were considered. For this, the injected samples were grouped according to their genus or species, as well as the part of the plant. This mapping highlighted 21 clusters of potential Wnt pathway inhibitors. Among them, 4 clusters related to the leaves of *Bocquillonia nervosa*, were selected. In parallel, to get more insights into the nodes contained in these four clusters, MS/MS spectra from previously isolated Euphorbiaceae diterpenoids were co-injected in the Wnt subnetwork. This allowed predicting the presence of a 12-deoxyphorbol scaffold within the selected compounds. Subsequent MS-guided purification yielded two new 12-deoxyphorbol esters (**44-45**) (Fig. 9), as well as two known ones. In order to validate the applied approach, the isolated compounds were evaluated for their capacity to inhibit the Wnt signaling pathway in HTB-19 cells. Three of the four isolated compounds proved to be highly potent inhibitors (IC₅₀

values ranging from 0.0336 to 1.18 µM). Finally, to prove that molecules from non-targeted clusters were truly inactive, two were isolated and were, indeed, found inactive (IC₅₀ > 225 µM). The subnetwork corresponding to the CHIKV cell-based assay was likewise explored after applying the taxonomical mapping, leading to the selection of less than 5% of the original molecular network. Among this selection, the aforementioned four clusters associated to *B. nervosa* extract that were highlighted in the Wnt subnetwork were also present, thus the four purified substances were also tested against CHIKV, showing extremely potent activity (EC₅₀ values ranging from 130 to 20 nM). As the four isolated molecules belonged to the deoxyphorbol series, the authors envisaged the identification of potential original CHIKV inhibitors with different skeleton. Thus, a cluster that showed a strong anti-CHIKV activity associated to the bark of *Neoguillauminia cleopatra*, a species with no previous phytochemical study, was selected for further chemical investigation. In order to obtain structural information about this cluster, it was annotated against the Euphorbiaceae subset of the DNP-ISDB, executing Tremolo in variable dereplication mode, thus allowing the identification of analogs within the database. This annotation step indicated the presence of highly oxygenated diterpenoids bearing a polyunsaturated side chain. Further analysis conducted to the isolation of a novel daphnane diterpene orthoester that was named neoguillauminin A (**46**) (Fig. 9). However, this compound showed a low EC₅₀ value (17.7 µM), probably because there were other potent molecules in the extract that could not be identified, nor isolated. In this regard, in order to increase the chances of finding the real source of the bioactivity, the authors advised to carry out the isolation of multiple molecules belonging to a presumable bioactive cluster.

In another study, Litaudon and Desrat *et al.*⁵⁴ further explored the same set of 292 New Caledonian Euphorbiaceae extracts using the same multi-informative prioritization strategy. Remarkably, this time the MS/MS data were pre-processed *via* the MZmine 2¹⁶ software affording a molecular network constituted of 17387 nodes grouped into 1231 clusters (88687 nodes grouped into 7840 clusters without preprocessing). This study led to the isolation of unprecedented chlorinated monoterpenyl quinolones (**47-51**) (Fig. 9).

As a comment on these two studies of multi-informative annotation of molecular networks of large data sets, the importance of the taxonomic mapping in highlighting unique chemistries related to spectral dissimilarity within a taxonomically homogenous set of samples can be emphasized. This concept was likewise applied by Jensen, Dorrestein, and Koyama *et al.*⁵⁵ for the exploration of a culture collections of 1000 marine microorganisms in a self-referential manner. Notably, the authors pursued the hypothesis that if a molecule was found infrequently, there would be a lesser probability that it has been previously described. To achieve this, 3000 MS/MS datasets were considered to conceive a molecular network. The exploration of this massive network was performed using a color tag in order to classify the molecules by the occurrence of their MS/MS spectra among the strains.

This allowed the identification of a molecular family of more than two dozen nodes belonging to a single strain (CNP-993). Comparison with reference databases, suggested that this was a unique molecular family in the data set, possibly representative of a novel chemical class. Within this cluster, two nodes were targeted for further structural elucidation. These new NPs were named maridric acids A and B (**52-53**) (Fig. 10). These results showed that, this approach, provides an effective tool for the untargeted prioritization of microorganisms in varied growth or extraction conditions, in order to optimize the utilization of large culture collections. In a similar study, Dorrestein, Moore *et al.* used this type of prioritization strategy to explore a collection of 146 marine *Salinispora* and *Streptomyces* strains (603 samples) by integrating culture conditions (solid *versus* liquid media, time), extraction protocols (solvent polarities), and strain locations annotations in order to quickly identify patterns in metabolite production.⁵⁶ Among 5526 nodes, 15 molecular families were identified based on MS/MS spectra annotations derived from the GNPS library.

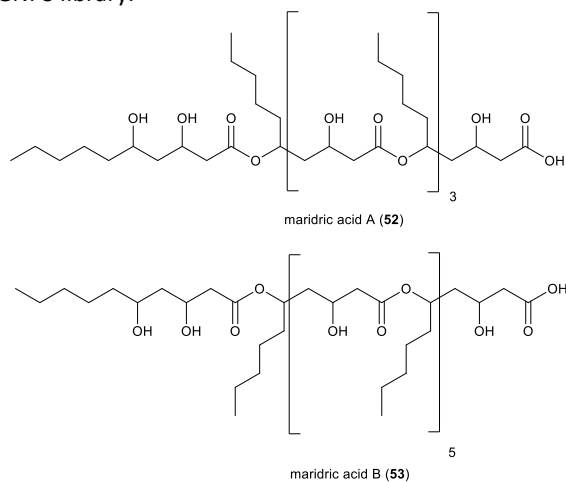


Fig. 10 Structures of maridric acids A and B (**52-53**).

Other research teams used MN to map large collection of diverse natural sources to uncover unexpected chemistries by integrating other kind of functional annotations. Taking advantage of the access to an exceptional in-house collections of marine cyanobacteria and algae, Luzzatto-Knaan Gerwick, and Dorrestein *et al.* applied an untargeted MS analysis, in order to identify novel NPs.⁵⁷ For this study, approximately 2600 fractions originating from 317 marine collections, were analysed by reversed phase ultra-performance liquid chromatography (RP-UPLC) coupled with high resolution quadrupole time-of-flight mass spectrometry (HR-qTOF-MS).

For each LC-MS/MS run, nearly 6000 spectra were collected, making a total of 15.6 million spectra that were submitted to the GNPS platform to conceive a massive MN.

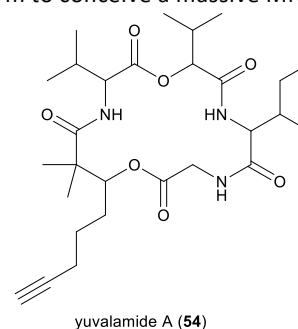


Fig. 11 Structure of yuvalamide A (**54**).

Interestingly, the percentage of compounds from this library that matched with the GNPS library was about 0.04%, indicating that the chemical space of these marine cyanobacteria and algae communities is significantly unexplored, rendering very high the probability of finding new interesting compounds. In this regard, to delve more deeply into the generated molecular network, a geographical pattern related to the provenance of each sample was considered and overlaid over the network. The metabolites contained in the samples originating from four distinct areas were differentiated. This allowed the recognition of a potentially novel molecule in a Panama-Portobelo cluster. Subsequent MS-guided isolation led to the identification of a new compound, named, yuvalamide A (**54**) (Fig. 11). The identification of **54** allowed the deduction of putative structural analogs that were present in the same cluster.

3 Combination of molecular networking with other techniques

3.1 Biochemometrics

As discussed in the previous section, the search for bioactive molecules within complex mixtures has made MN evolves towards the inclusion of bioactivity data of the studied samples, as a prediction of the biological activities of the compounds included in the network. In the first study⁵² discussed in the previous section (2.2.2), the bioactivity data were obtained through the testing of the samples against particular targets, in order to search for potential drug leads. In the second one, Dorrestein *et al.*⁴⁹ proposed to calculate a bioactivity score taking into account the ion intensity in the samples and the bioactivity level of each sample. Further creative approaches may appear in the future.

To understand the meaning of biochemometrics, we should refer, at first, to chemometrics, which is the science that applies optimal mathematical and statistical methods to process chemical data.⁵⁸ This discipline, however, has evolved since its creation, resulting in the invention of biochemometrics. This term was proposed initially by Martens *et al.* in 2006 referring to the use of chemometrics in modern biochemistry, biotechnology, and molecular biology.⁵⁹ In other words, it is the application of statistical methods to correlate chemical profile to bioactivity.⁶⁰ In this vein, Cech *et al.* developed a new workflow to prioritize the isolation of bioactive molecules within natural extracts making use of biochemometrics associated to the MN technique.⁶¹ In this study, MS, MS/MS data, and biological data were gathered. Thus, to demonstrate the capability of this approach, it was applied to the exploration of the chemical composition of the roots of *Angelica keiskei* (Apiaceae) in order to identify the components responsible for its antimicrobial activity. At first, the MeOH root extract was submitted to bioactivity screening, demonstrating a complete inhibition of methicillin-resistant *Staphylococcus aureus* (MRSA). Then, the extract was iteratively fractionated and the fractions were tested against MRSA. The most active ones were then analysed using an UPLC-LTQ-Orbitrap XL mass spectrometer in both positive and negative modes. The acquired LC-MS data were treated separately using MZmine 2¹⁶ to form the final data set for biochemometrics analysis. In order to predict the features responsible for the observed anti-MRSA activity in the fractions, biochemometrics analyses were completed using Sirius⁶² version 10.0 statistical software (Pattern Recognition Systems). This software contains statistical algorithms that were used to compute selectivity ratio plot for the bioactive molecules. With this, scores were attributed to the features present in the active samples, resulting in a list with the top contributors to the anti-MRSA activity. However, as this tool could not provide any structural insights, MN was used to have access to this information. Comparison of the resulting molecular network against the molecules selected by the biochemometrics selectivity scores allowed the annotation of already-described active molecules within the network: 4-hydroxyderricin and xanthoangelol, two chalcones that are the only known anti-MRSA agents in *A. keiskei*. This integration of biochemometrics with MN allowed also the annotation of fifteen molecules that matched accurate masses of known chalcones not endowed with antimicrobial activity. Among them, five were included within the group of potential contributors to the observed bioactivity by biochemometrics, with the one at *m/z* 491.202 being the top contributor. Bioactive fractions submitted to purification resulted in the isolation of 4-hydroxyderricin and xanthoangelol, as well as two other chalcones, one of them being inactive according to biochemometrics information. Biological evaluation of the isolated compounds confirmed this prediction. Other compounds were also part of the top contributors list but could not be isolated due to scarce amounts in the extract. This is one of the limitations inherent to the biochemometrics approach for identifying minor bioactive, as their structures

and activities cannot be confirmed without isolation. In contrast, as an advantage, biochemometric selectivity ratio analysis allows the identification of low-abundance constituents contributing to activity without being confounded by the abundance of other compounds.

3.2 Mass spectrometry imaging

In order to explore specific and particular NPs chemical spaces, MN has also been coupled to mass spectrometry imaging (MSI). This technique allows to combine molecular mass analysis and spatial information, providing the visualization of molecules on complex surfaces.⁵⁸ Within these specific chemical spaces, the interaction between two organisms can be harnessed. As this relation can be positive (commensalism, mutualism, symbiosis, etc.) or negative (predation, parasitism, antibiosis, etc.), the molecular dialogue between two organisms will be different for each kind of relation, highlighting the potential richness of these chemical spaces. In this vein, the interaction between a fungus (*Paraconiothyrium variabile* - Montagnulaceae) and a bacteria (*Bacillus subtilis* - Bacillaceae), both endophytes of *Cephalotaxus harringtonia* (Cephalotaxaceae) was explored by Prado, Brunelle *et al.* to identify the features that are present in this interspecific communication.⁶³ Starting from the observation that, when isolated, these two species showed a strong and unique antagonism that was not observed between other partners of the plant microbiota, the competition zone was explored by the MN technique in comparison against the metabolites produced by each microorganism independently. MS/MS data of the crude ethyl acetate extracts of *B. subtilis* and *C. harringtonia*, as well as the competition zone and the culture media were submitted to the GNPS in order to generate a molecular network. A first dereplication against the GNPS library allowed the annotation of a cluster containing surfactin-like molecules,⁶⁴ including, surfactins C-13, C-14, and C-15 as well as their hydrolyzed derivatives. Surprisingly, all of them were only detected in the bacterium and competition extracts. As these molecules are known to inhibit other fungus growth, it was hypothesized that *P. variabile* had developed a resistance mechanism that conducted to the hydrolysis of these features. To confirm this, an MSI of the microbial competition between these two species was carried out with MALDI-TOF and TOF-SIMS. Both of them allowed the detection of hydrolyzed surfactins in the course of the interspecific endophytic microbial competition.

3.3 Genome mining and metabologenomics

Nowadays, the search for novel NPs can be done in two different ways: "upstream," at the genome level, or "downstream," at the metabolite level.⁶⁵ MN has been coupled to genome-mining with the purpose to delve more into the biosynthetic gene clusters responsible for the production of a metabolite. The information provided by this correlation can be exploited to enhance the discovery, the isolation, as well as the structural prediction of new NPs produced by an organism. Using this association between genomics and metabolomics data, Kleigrew *et al.* performed the exploration of the

chemical diversity of marine cyanobacteria.⁶⁶ For this study, three cultured strains, *Moorea producens* 3L, *M. producens* JHB, and *M. bouillonii* PNG were chosen because they are known to produce many structurally diverse and biologically active NPs.⁶⁷ Firstly, these species were subjected to genome sequencing and analysis for their recognizable biosynthetic pathways, envisaging the identification of similar or nearly identical biosynthetic genes in the three strains. As a result, a regulatory serine histidine kinase gene was identified, in the two *M. producens* strains, as being located near a hybrid biosynthetic pathway responsible of the production of the aforementioned active compounds. Considering that this regulatory kinase was highly homologous between these two strains (96.1% of similarity), the existence of a gene encoding this regulatory enzyme within the *M. bouillonii* PNG genome sequence might identify new NP biosynthetic gene clusters. A highly homologous sequence was found in the *M. bouillonii* PNG genome and the gene neighbourhood for this kinase was explored, finding out a new and undescribed biosynthetic gene cluster with several unique features. The potential expression of metabolites by this gene cluster was evaluated by the analysis of the metabolic profile of each strain using MN. In the resulting network, clusters containing the above-mentioned molecules were rapidly identified. In addition, two molecular families produced by *M. bouillonii* PNG drew the attention of the authors because the isotopic pattern of the parent masses indicated the presence of di- and trichlorinated species. Thus, three new NPs, columbamides A, B, and C (**54-57**) (Fig. 12), were discovered.

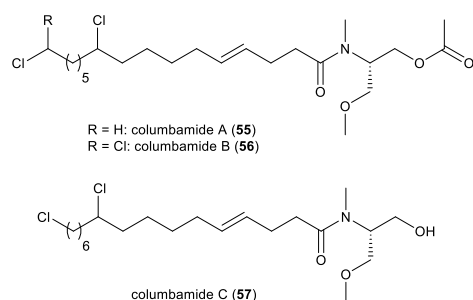


Fig. 12 Structures of columbamides A-C (**55-57**).

Another application of the association of these two techniques was reported by Maansson, Vynne, and Klitgaard *et al.* with the study of 13 genetic variants of the marine bacteria *Pseudoalteromonas luteoviolacea* for their genomic potential and ability to produce secondary metabolites.⁶⁸ Contrarily to the previous described study, the metabolomics analysis was

performed prior to the genome sequencing. At first, extracts from all the strains were analysed by an untargeted metabolomics experiment using LC-HRMS. To facilitate comparison to genomic data, all these compounds were represented as pan- and core- plots, revealing that only 2% of the molecular features were shared by all the strains, and that 30% were exclusively produced by single strains. Pan- and core-genomic analyses were then both applied to make a direct comparison with the metabolomics data of all strains. The core genome was found to constitute 65% for each strain, and 23% of the total genes were identified in a single strain. On average, 8.6% of the total genes were found to be allocated to secondary metabolism, which is a very high amount compared to other strains of *Pseudoalteromonas*. In addition, two strains were identified as hot spots for biosynthetic diversity due to the presence of singular operational biosynthetic units. As the next step, MN was used to prioritize novel chemical motifs isolation within the complete metabolome. In the resulting network, some molecular families, thus biosynthetic pathways, were rapidly identified, like the *vio* genes pathway, found in all the strains. Other already-described molecules were also found, like violacein and three of its analogs. However, the most interesting finding within the network was the presence of 313 compounds produced only by the two strains, nominated as hot spots. Among these molecules, whole series of already-known substances were identified as thiomarinol and pseudomononic acid analogs. Besides these chemical features, two novel analogs were found. Based on their molecular formulas, these molecules constituted a new type of thiomarinol.

Jensen, Moore, and Dorrestein *et al.*⁶⁹ reported the exploration of 35 strains belonging to the marine actinomycete genus *Salinispora* in order to visualize the molecular composition of their organic extracts by combination of genetics data and MN exploration. The extracts obtained from the cultures of the 35 strains were analysed by LC-HRMS/MS. This resulted in the generation of over 200000 MS¹ HRMS spectra ranging from *m/z* 304.175 to 2485.4. These data were processed by MN, resulting in a network with 1137 parent ions that was screened against reference substances, in order to identify families of already-described compounds. This step allowed the dereplication of cyclomarin and arenicolide, both connected to putative structural analogs. The information retrieved from the molecular network was coupled with the genome sequence data of these species, in a correlative analysis that was named "pattern-based genome mining". Through it, the identification of a compound within a molecular family proves the expression of the associated biosynthetic gene cluster (BGC). This correlation was observed

for molecular features like the arenicolides and cyclomarins, which were identified in the three strains that possessed the associated BGC. However, in some cases, the link was less reproducible, like for desferrioxamines, which were only detected in 1 of 21 strains that owned the corresponding BGC. In total, this relation between molecules and BGCs was observed in only 34 out of 140 cases. This result suggested that many of the BGCs were not expressed, perhaps because of the culture conditions or that the corresponding products were not extracted or went undetected in the LC-MS analyses. Application of pattern-based genome mining allowed the identification of the pathway NRPS40 as unique to strain CNT-005. In this context, the molecular network was explored in order to search for substances produced exclusively by this strain, leading to the recognition of a molecular family of quinomycin-like compounds, a group of antitumor antibiotic dimeric depsipeptides. Within this cluster, an ion was targeted for isolation, conducting to the purification of a novel non-ribosomal peptide that was named retimycin A (**58**) (Fig. 13).

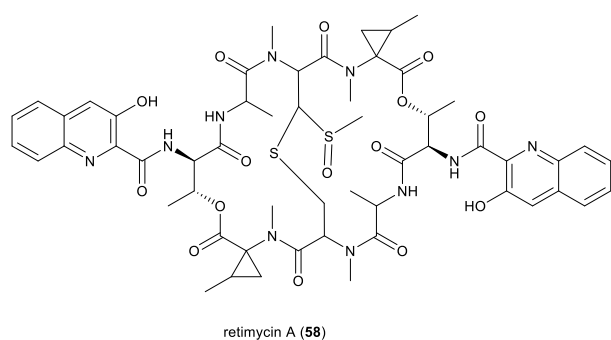


Fig. 13 Structure of retimycin A (**58**).

While the discovery of new BGC family suggests new NPs, a confirmation is still required. Without advanced knowledge of structures or bioactivities, detection of novel NPs is very tedious. To address this issue, Metcalf *et al.* introduced a new concept called metabologenomics,⁷⁰ an automated untargeted method for identifying NPs based on binary correlation between a BGC and a molecule identified by LC-HRMS. Metabologenomics has already allowed the identification of several novel NPs, including tambromycin⁷¹ and the rimosamides⁷². Recently, this strategy has been coupled with MN and resulted in the discovery of a new family of nonribosomal peptides featuring an unusual trimethylammonium tyrosine residue, that was named the tyrobetaines (**59-60**) (Fig. 14).⁷³ Briefly, LC-MS/MS analyses were conducted on a Q-Extractive mass spectrometer using high-energy collisional dissociation (HCD). Notably, the authors did not name the software used for the processing and the extraction of the resulting MS/MS data prior to their submission to the GNPS. Then, the generated molecular network was manually explored for masses of interest (*i.e.* masses that scored highly in the metabologenomics method), leading to the discovery of tyrobetaine family cluster.

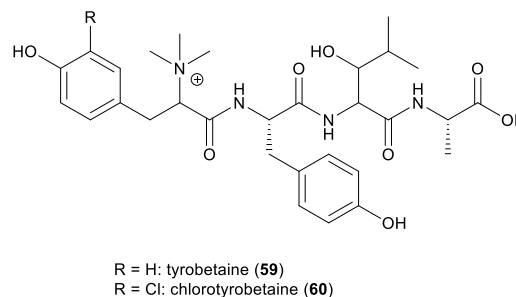


Fig. 14 Structures of tyrobetaines (**59-60**).

3.4 Chemical epigenetics

Recently, new aspercryptins were isolated from *Aspergillus nidulans* by Kelleher *et al.*, using MN-based dereplication in conjunction with chemical epigenetics.⁷⁴ The genome of this mold species is known to contain more than 50 gene clusters involved in the biosynthesis of NPs. However, only the products of 20 of them have already been described. Several strategies have been proposed to address this problem, such as the use of epigenetic regulators, including inhibitors of DNA methyltransferase (DNMT) and histone deacetylase (HDAC).⁷⁵ In this context, the metabolomes of wildtype *A. nidulans*, and of a HDAC-deficient strain were mapped by MN. Notably, the processing of the MS/MS data of this study was carried out in the same manner as the previous metabologenomics study.⁷³ In the resulting network, several molecular families were identified to be produced by only one biological state of the species. Rapidly, a family of already-described molecules, the aspercryptins (lipopeptide) was located within the metabolites produced, in majority, by the mutant strain. Aspercryptins A1 and A2 were present in this cluster and had their structure elucidated based completely on the MS information. Two other aspercryptins were also detected: aspercryptins B1 and B3. As these four molecules were linked to several nodes, their MS/MS fragmentation patterns were used as anchor points to propose the structure of 13 additional aspercryptins analogs.

3.5 Stable isotope labeling

Deciphering biosynthetic pathways using radioactive labeled substrates has historically relied on sensitive radiation detectors. Recent, advances in LC-MS instrumentation has enabled the use of stable isotope labeled, avoiding the risks inherent to the handling of radioactive material. Many studies showed that cultivation of bacteria in the presence of labeled amino acids could be used for the characterization of linear non ribosomal peptides by MS/MS analysis.⁷⁶ In this regard, Nielsen *et al.* combined stable isotope labeling and MN to detect several known and unknown compounds that were labeled, leading to the study of the biosynthesis of nidulanin A and related products produced by *Aspergillus nidulans*.⁷⁷ Accordingly, samples were taken from fungi cultivated both with and without labeled amino acids and analyzed by LC-MS/MS (positive mode). The acquired data allowed the generation of a molecular network. Using the information from

the labeling experiment led to highlight the nodes that differed in m/z according to the predicted mass shifts obtained from the incorporation of the stable isotope labeled amino acids.

4 Latest improvements and tools implemented in the molecular networking practice

4.1 Data pre-processing

Although MN allows to efficiently identifying new NPs, some limitations were addressed by several research groups with various strategies, allowing the generation of more informative and reliable molecular networks. The MS-Cluster tool, included in the GNPS architecture, has been reported to be source of some pitfalls, because it cannot distinguish between isomers, as retention times are not considered during data processing. To solve this issue, Touboul *et al.*²⁹ proposed to introduce a preprocessing workflow including the data treatment by MZmine 2¹⁶ prior to its upload on the GNPS platform. Applying this treatment to the MS/MS data enables the separation of isobaric isomers, the annotation of the molecular features with calculated chemical formulas, and precursor ion abundances. To perform this workflow, the authors built a homemade Python script (available at: <https://github.com/Florent-Olivon/MZM2-MN>). Interestingly, a GNPS export option that encompasses all the features cited above has been added as a built-in option since MZmine 2.25 (<http://mzmine.github.io/changelog.html>).

4.2 Data organization

Olivon, Touboul *et al.* developed MetGem⁷⁸, an innovative software for the generation of molecular networks without uploading the MS/MS data on the GNPS platform (available at: <https://metgem.github.io>). Furthermore, this software allows the parallel investigation of two complementary representations of the raw dataset, one based on a classic GNPS-style MN and another one based on the t-SNE (stochastic neighbour embedding) algorithm, a well-known technique used for high-dimensional data visualization.⁷⁹ Additionally, almost all parameters (cosine score value and maximum neighbour number (topK)), can be tuned in real time and new networks can be generated within few seconds for small datasets. The t-SNE graph preserves the interactions between related groups of spectra while the MN output allows an unambiguous separation of clusters. With this software, the authors wanted to address weaknesses inherent to the MN visualization architecture leading to non-clusterized compounds over the molecular network, even when they share similar scaffolds and comparable MS² spectra.

4.3 New reporting standards for MS data

Ultimately, the integration of MN in the NP discovery process brought the need for new reporting standards for LC-MS/MS data sets as well as NPs spectral properties. Consequently, one can witness today the addition of the so-called MSV and CCMSLIB codes to the traditional spectral properties related to the description of a NP. The above-mentioned codes are respectively generated upon the deposition of any full MS data sets and NP MS/MS spectrum on MassIVE, (Mass Spectrometry Interactive Virtual Environment, available at: <https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp>). With growing awareness of the NP community for the expansion of crowdsourced spectral libraries, we should likely assist to an increasing integration of the aforementioned codes in NP isolation reports.

Universal identifier. As matters transpired, the proliferation of MS spectral databases resulted ineluctably in the multiplication of accession numbers, often referring to the same mass spectrum. To address this issue, Fiehn, Schymanski, and Wohlgemuth *et al.* designed a spectral identifier, called SPLASH (SPectraL hASH), that improves the exchange and searchability of mass spectra and avoids their duplication.⁸⁰ Noteworthy, SPLASH has already been implemented in MassBank⁸¹, MoNA (<http://mona.fiehnlab.ucdavis.edu/>), GNPS⁹, HMDB⁸², MetaboLights¹⁰, and mzCloud (<https://www.mzcloud.org/>), as well as software tools such as MZmine¹⁶, MS-DIAL⁸³, RMassBank⁸⁴, BinBase⁸⁵, Bioclipse⁸⁶, and the Mass Spectrometry Development Kit (MSDK; <https://msdk.github.io/>).

4.4 Advanced molecular networking annotation tools

As outlined above, the issue of annotation of molecular networks captured the attention of many research teams in the past three years and still remains an algorithmic bottleneck.⁸⁷ To overcome this issue, several strategies based on computational processing have been proposed, including the aforementioned ISDB³² and NAP⁴¹, as well as, MS2LDA^{88, 89}, Sirius^{90, 91}, Dereplicator+⁹², Insilico Peptidic Natural Product Dereplicator⁹³, VarQuest⁹⁴, Peptidogenomics for Ribosomally Synthesized Post-translationally Modified Peptides (RiPPs) – RiPPquest,⁹⁵ and MetWork.⁹⁶ Except Sirius, MetWork, and MS2LDA, all these tools are embarked on the GNPS platform. Interested readers are referred to the documentation available at: <https://gnps.ucsd.edu/ProteoSAFe/static/gnps-theoretical.jsp>. Hence, in this subsection, we wish to disclose a brief description of the three remaining tools.

Sirius is a freely available resource (<https://bio.informatik.uni-jena.de/software/sirius/>) that allows the analysis of isotopic⁹⁷ and fragmentation⁹⁰ patterns in HRMS and MS/MS spectra,

respectively. Moreover, it uses CSI:FingerID⁹¹ (Compound Structure Identification), which combines fragmentation tree computation and machine learning to search in molecular structure databases such as PubChem.⁹⁸ Sirius requires pre-processed MS/MS peak lists as inputs. This task can be performed by popular LC-MS feature detection software such as OpenMS⁵⁰, MZmine¹⁶, or XCMS⁹⁹. Recently, a Sirius identification module has been implemented in MZmine 2.34 (<http://mzmine.github.io/changelog.html>) allowing the direct exportation of calculated molecular formulas and FingerID-identified structures in a .csv file that could annotate molecular networks.

MS2LDA is an unsupervised method, inspired by a text-mining technique called latent Dirichlet allocation (LDA),¹⁰⁰ that extracts common patterns of mass fragments and neutral losses from acquired MS/MS datasets.⁸⁹ This strategy can group molecules that share substructures (Mass2Motif) without high similarity across their entire MS/MS spectra. Interestingly, this approach enables the annotation of molecular networks at the scaffold level. MS2LDA expects information-rich MS/MS spectra (generated by ramped or stepped collision energy) as an input as well as pre-processed MS/MS peak lists. Recently, Van der Hooft and Rogers *et al.*,

developed a web application, accessible at <https://ms2lda.org>, that allows users to upload their MS/MS datasets and run MS2LDA analyses and explore the results through interactive visualizations.⁸⁸

MetWork is a recent webserver platform developed by Genta-Jouve *et al.*⁹⁶ that allows to expand the annotation of molecular networks in a different manner than the above-mentioned strategies. Indeed, this tool encompasses functionalities allowing the anticipation of new NPs. It is based on MS/MS data, organized in a molecular network, a collaborative library of (bio)chemical transformations and a MS/MS spectra prediction module based on CFM-ID.³³ Starting from one annotated node in the molecular network, the server generates putative structures. A similarity comparison with their *in silico* MS/MS spectra is then performed in order to annotate the nodes of a molecular network through the exportation of .csv file.

Hopefully, one can easily envision the upcoming integration of the above-mentioned annotations tools on a common web platform. Meanwhile, we propose to depict these multiple LC-MS/MS data processing tools interfaced with MN in a state-of-the-art dereplication pipeline in Figure 15.

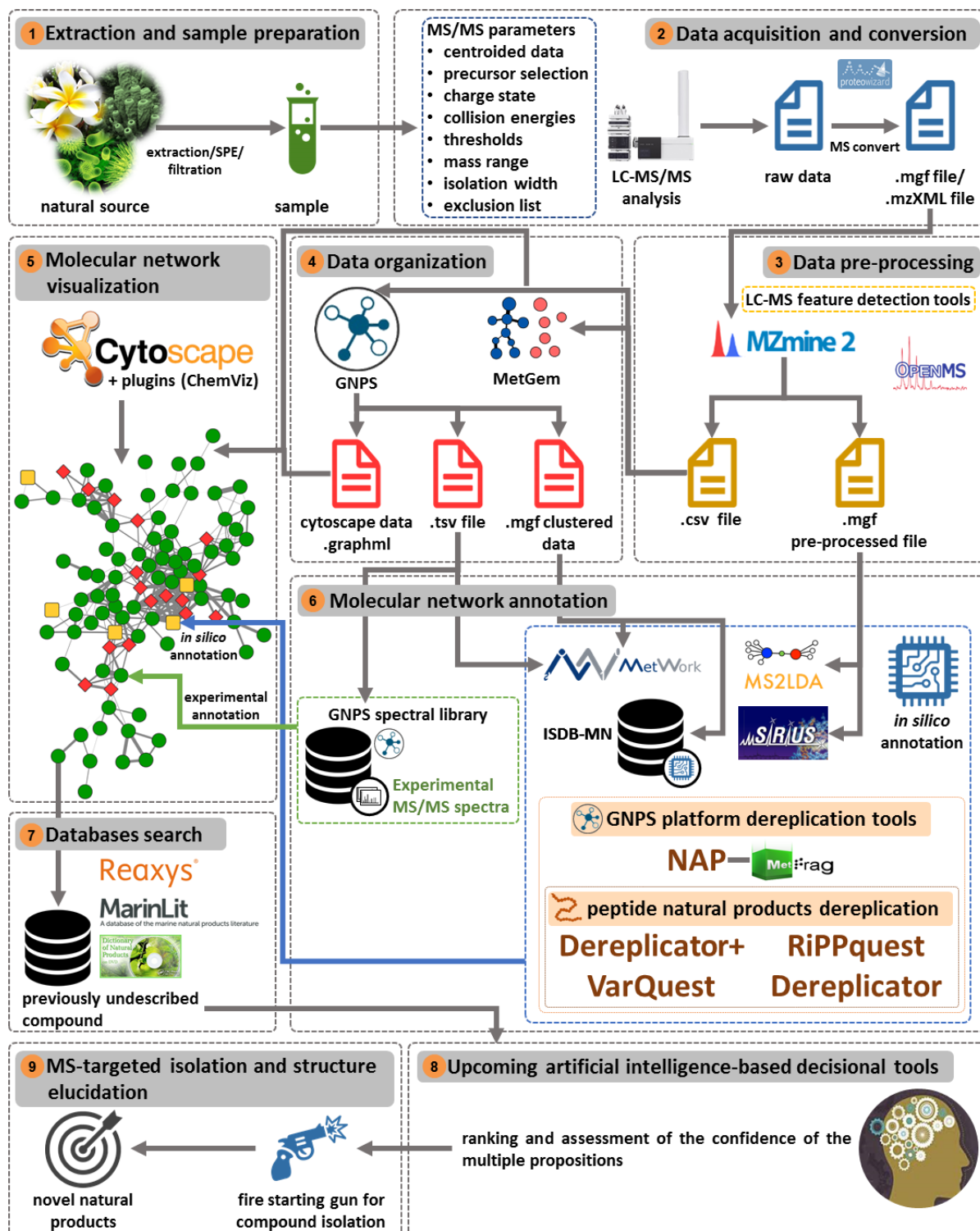


Fig. 15 A schematic workflow for a comprehensive state-of-the-art LC-MS/MS-based dereplication pipeline.

5 Conclusions

MN constitutes an accessible and adaptable means to visualize and target NPs, enabling biological research and biotechnological applications in a wide range of fields.¹⁰¹ Molecular networks are nowadays widely accepted by the NP research community as a solid and pivotal support that provides a metabolite level view of the data. Over these seven past years a number of innovative strategies along with landmark improvements arose from both the NP chemists and bio-informatics communities. In this context, it is worth noting that NPs have a long-lasting history in challenging state-of-the-art analytical techniques.¹⁰² These improvements have vastly accelerated the pace of research, from understanding biosynthetic pathways and efficient NP targeting, to the development of enhanced algorithm to sharpen the tool.

Despite the power of the above-mentioned *in silico* annotation strategies, there is still a need for a human brain to assess and rank the confidence of the large amount of the generated data. Perhaps not for long, as we will shortly assist in the emergence of artificial intelligence-assisted decisional tools that will filter those propositions (Fig. 15).

Definitely, the stage is now set for NP chemists to aim for "anticipation" in NPs isolation workflows. Indeed, the emergence of the annotation tools enable, today, to search for *in silico*-generated structures in NPs databases, which is a huge step forward compared to the traditional dereplication process based on molecular formulas and/or exact masses, or NMR chemical shifts searches. In this regard, the continued growth and enrichment of crowdsourced spectral libraries¹⁰³ will enhance machine learning-based algorithms, paving the way for more efficient structural predictions.

The ever-expanding repertoire of applications firmly positions MN at the cutting edge of NPs targeting, and holds the promise of even more exciting discoveries and inventions to come.

6 Conflicts of interest

There are no conflicts to declare.

7 Acknowledgements

The authors thank Dr. Grégory Genta-Jouve (Université Paris-Descartes) and Dr. Pierre-Marie Allard (University of Geneva) for fruitful discussions. We are grateful to FONDECYT-CONCYTEC for the fellowship 239-2015-FONDECYT (A.E.F.R.).

8 Notes and references

1. Q. Michaudel, Y. Ishihara and P. S. Baran, *Acc. Chem. Res.*, 2015, **48**, 712-721.
2. M. H. Medema and M. A. Fischbach, *Nat. Chem. Biol.*, 2015, **11**, 639.
3. M. T. Henke and N. L. Kelleher, *Nat. Prod. Rep.*, 2016, **33**, 942-950.
4. S. P. Gaudencio and F. Pereira, *Nat. Prod. Rep.*, 2015, **32**, 779-810.
5. 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. Lington, L. Zhang, H. M. Debonsi, W. H. Gerwick and P. C. Dorrestein, *J. Nat. Prod.*, 2013, **76**, 1686-1699.
6. R. A. Quinn, L.-F. Nothias, O. Vining, M. Meehan, E. Esquenazi and P. C. Dorrestein, *Trends Pharmacol. Sci.*, 2017, **38**, 143-154.
7. A. A. Aksenov, R. da Silva, R. Knight, N. P. Lopes and P. C. Dorrestein, *Nat. Rev. Chem.*, 2017, **1**, 0054.
8. H. Mohimani and P. A. Pevzner, *Nat. Prod. Rep.*, 2016, **33**, 73-86.
9. M. Wang, J. J. Carver, V. V. Phelan, L. M. Sanchez, N. Garg, Y. Peng, D. D. Nguyen, J. Watrous, C. A. Kapon, T. Luzzatto-Knaan, C. Porto, A. Bouslimani, A. V. Melnik, M. J. Meehan, W.-T. Liu, M. Crusemann, P. D. Boudreau, E. Esquenazi, M. Sandoval-Calderon, R. D. Kersten, L. A. Pace, R. A. Quinn, K. R. Duncan, C.-C. Hsu, D. J. Floros, R. G. Gavilan, K. Kleigrew, T. Northen, R. J. Dutton, D. Parrot, E. E. Carlson, B. Aigle, C. F. Michelsen, L. Jelsbak, C. Sohlenkamp, P. Pevzner, A. Edlund, J. McLean, J. Piel, B. T. Murphy, L. Gerwick, C.-C. Liaw, Y.-L. Yang, H.-U. Humpf, M. Maansson, R. A. Keyzers, A. C. Sims, A. R. Johnson, A. M. Sidebottom, B. E. Sedio, A. Klitgaard, C. B. Larson, C. A. Boya P, D. Torres-Mendoza, D. J. Gonzalez, D. B. Silva, L. M. Marques, D. P. Demarque, E. Pociute, E. C. O'Neill, E. Briand, E. J. N. Helfrich, E. A. Granatosky, E. Glukhov, F. Ryffel, H. Houson, H. Mohimani, J. J. Kharbush, Y. Zeng, J. A. Vorholt, K. L. Kurita, P. Charusanti, K. L. McPhail, K. F. Nielsen, L. Vuong, M. Elfeki, M. F. Traxler, N. Engene, N. Koyama, O. B. Vining, R. Baric, R. R. Silva, S. J. Mascuch, S. Tomasi, S. Jenkins, V. Macherla, T. Hoffman, V. Agarwal, P. G. Williams, J. Dai, R. Neupane, J. Gurr, A. M. C. Rodriguez, A. Lamsa, C. Zhang, K. Dorrestein, B. M. Duggan, J. Almaliti, P.-M. Allard, P. Phapale, L.-F. Nothias, T. Alexandrov, M. Litaudon, J.-L. Wolfender, J. E. Kyle, T. O. Metz, T. Peryea, D.-T. Nguyen, D. VanLeer, P. Shinn, A. Jadhav, R. Muller, K. M. Waters, W. Shi, X. Liu, L. Zhang, R. Knight, P. R. Jensen, B. O. Palsson, K. Pogliano, R. G. Lington, M. Gutierrez, N. P. Lopes, W. H. Gerwick, B. S. Moore, P. C. Dorrestein and N. Bandeira, *Nat. Biotechnol.*, 2016, **34**, 828-837.
10. K. Haug, R. M. Salek, P. Conesa, J. Hastings, P. de Matos, M. Rijnbeek, T. Mahendraker, M. Williams, S. Neumann, P. Rocca-Serra, E. Maguire, A. González-Beltrán, S.-A. Sansone, J. L. Griffin and C. Steinbeck, *Nucl. Acids Res.*, 2013, **41**, D781-D786.
11. A. E. Fox Ramos, C. Alcover, L. Evanno, A. Maciuk, M. Litaudon, C. Duplais, G. Bernadat, J.-F. Gallard, J.-C. Jullian, E. Mouray, P. Grellier, P. M. Loiseau, S. Pomel, E. Poupon, P. Champy and M. A. Beniddir, *J. Nat. Prod.*, 2017, **80**, 1007-1014.
12. M.-M. Janot, *Tetrahedron*, 1961, **14**, 113-125.
13. H. Rapoport and R. E. Moore, *J. Org. Chem.*, 1962, **27**, 2981-2985.
14. E. L. Schymanski, J. Jeon, R. Gulde, K. Fenner, M. Ruff, H. P. Singer and J. Hollender, *Environ. Sci. Technol.*, 2014, **48**, 2097-2098.
15. E. Otogo N'Nang, G. Bernadat, E. Mouray, B. Kumulungui, P. Grellier, E. Poupon, P. Champy and M. A. Beniddir, *Org. Lett.*, 2018, **20**, 6596-6600.

ARTICLE

16. T. Pluskal, S. Castillo, A. Villar-Briones and M. Orešič, *BMC Bioinform.*, 2010, **11**, 395.
17. A. Boufridi, S. Petek, L. Evanno, M. A. Beniddir, C. Debitus, D. Buisson and E. Poupon, *Tetrahedron Lett.*, 2016, **57**, 4922-4925.
18. A. Boufridi, D. Lachkar, D. Erpenbeck, M. A. Beniddir, L. Evanno, S. Petek, C. Debitus and E. Poupon, *Aust. J. Chem.*, 2016, **70**, 743-750.
19. G. Daletos, N. J. de Voogd, W. E. G. Müller, V. Wray, W. Lin, D. Feger, M. Kubbutat, A. H. Aly and P. Proksch, *J. Nat. Prod.*, 2014, **77**, 218-226.
20. H. S. Radeke, C. A. Digits, R. L. Casaubon and M. L. Snapper, *Chem. Biol.*, 1999, **6**, 639-647.
21. M. P. Nambiar and H. C. Wu, *Exp. Cell Res.*, 1995, **219**, 671-678.
22. L. Evanno, D. Lachkar, A. Lamali, A. Boufridi, B. Séon-Méniel, F. Tintillier, D. Saulnier, S. Denis, G. Genta-Jouve, J.-C. Jullian, K. Leblanc, M. A. Beniddir, S. Petek, C. Debitus and E. Poupon, *Eur. J. Org. Chem.*, 2018, **2018**, 2486-2497.
23. N. Bonneau, G. Chen, D. Lachkar, A. Boufridi, J.-F. Gallard, P. Retailleau, S. Petek, C. Debitus, L. Evanno, M. A. Beniddir and E. Poupon, *Chem. Eur. J.*, 2017, **23**, 14454-14461.
24. J. R. Winnikoff, E. Glukhov, J. Watrous, P. C. Dorrestein and W. H. Gerwick, *J. Antibiot.*, 2013, **67**, 105-112.
25. R. M. Van Wagoner, A. K. Drummond and J. L. C. Wright, in *Advances in Applied Microbiology*, Academic Press, 2007, vol. 61, pp. 89-217.
26. N. A. Moss, T. Leão, M. R. Rankin, T. M. McCullough, P. Qu, A. Korobeynikov, J. L. Smith, L. Gerwick and W. H. Gerwick, *ACS Chem. Biol.*, 2018, **13**, 3385-3395.
27. M. C. Chambers, B. Maclean, R. Burke, D. Amodei, D. L. Ruderman, S. Neumann, L. Gatto, B. Fischer, B. Pratt, J. Egertson, K. Hoff, D. Kessner, N. Tasman, N. Shulman, B. Frewen, T. A. Baker, M.-Y. Brusniak, C. Paulse, D. Creasy, L. Flashner, K. Kani, C. Moulding, S. L. Seymour, L. M. Nuwaysir, B. Lefebvre, F. Kuhlmann, J. Roark, P. Rainer, S. Detlev, T. Hemenway, A. Huhmer, J. Langridge, B. Connolly, T. Chadick, K. Holly, J. Eckels, E. W. Deutsch, R. L. Moritz, J. E. Katz, D. B. Agus, M. MacCoss, D. L. Tabb and P. Mallick, *Nat. Biotechnol.*, 2012, **30**, 918-920.
28. A. M. Frank, N. Bandeira, Z. Shen, S. Tanner, S. P. Briggs, R. D. Smith and P. A. Pevzner, *J. Proteome Res.*, 2008, **7**, 113-122.
29. F. Olivon, G. Grelier, F. Roussi, M. Litaudon and D. Touboul, *Anal. Chem.*, 2017, **89**, 7836-7840.
30. F. Olivon, F. Roussi, M. Litaudon and D. Touboul, *Anal. Bioanal. Chem.*, 2017, **409**, 5767-5778.
31. J.-L. Wolfender, J.-M. Nuzillard, J. J. J. van der Hooft, J.-H. Renault and S. Bertrand, *Anal. Chem.*, 2019, **91**, 704-742.
32. P.-M. Allard, T. Péresse, J. Bisson, K. Gindro, L. Marcourt, V. C. Pham, F. Roussi, M. Litaudon and J.-L. Wolfender, *Anal. Chem.*, 2016, **88**, 3317-3323.
33. F. Allen, A. Pon, M. Wilson, R. Greiner and D. Wishart, *Nuc. Acids Res.*, 2014, **42**, W94-99.
34. J. Gu, Y. Gui, L. Chen, G. Yuan, H.-Z. Lu and X. Xu, *PLoS One*, 2013, **8**, e62839.
35. C. L. Zani and A. R. Carroll, *J. Nat. Prod.*, 2017, **80**, 1758-1766.
36. At that time, the CFM-ID algorithm was not able to handle the fragmentation of permanently charged compounds. This issue has been recently fixed.
37. M. Wang and N. Bandeira, *J. Proteome Res.*, 2013, **12**, 3944-3951.
38. J. A. Beutler, J. G. Jato, G. Cragg, D. F. Wiemer, J. D. Neighbors, M. Salnikova, M. Hollingshead, D. A. Scudiero and T. G. McCloud, *Frontis*, 2006, 301-309.
39. T. Péresse, G. I. Jézéquel, P.-M. Allard, V.-C. Pham, D. T. Huong, F. Blanchard, J. r. m. Bignon, H. I. n. Lévaïque, J.-L. Wolfender and M. Litaudon, *J. Nat. Prod.*, 2017, **80**, 2684-2691.
40. L. C. Klein-Júnior, S. Cretton, P.-M. Allard, G. Genta-Jouve, C. S. Passos, J. Salton, P. Bertelli, M. Pupier, D. Jeannerat, Y. V. Heyden, A. L. Gasper, J.-L. Wolfender, P. Christen and A. T. Henriques, *J. Nat. Prod.*, 2017, **80**, 3032-3037.
41. R. R. da Silva, M. Wang, L.-F. Nothias, J. J. J. van der Hooft, A. M. Caraballo-Rodríguez, E. Fox, M. J. Balunas, J. L. Klassen, N. P. Lopes and P. C. Dorrestein, *PLOS Comput. Biol.*, 2018, **14**, e1006089.
42. C. Ruttkies, E. L. Schymanski, S. Wolf, J. Hollender and S. Neumann, *J. Cheminformatics*, 2016, **8**, 3.
43. K. B. Kang, E. J. Park, R. R. da Silva, H. W. Kim, P. C. Dorrestein and S. H. Sung, *J. Nat. Prod.*, 2018, **81**, 1819-1828.
44. I. Blaženović, T. Kind, J. Ji and O. Fiehn, *Metabolites*, 2018, **8**, 31.
45. A. C. Hartmann, D. Petras, R. A. Quinn, I. Protsyuk, F. I. Archer, E. Ransome, G. J. Williams, B. A. Bailey, M. J. A. Vermeij, T. Alexandrov, P. C. Dorrestein and F. L. Rohwer, *Proc. Natl. Acad. Sci. U.S.A.*, 2017, **114**, 11685-11690.
46. J. J. Kellogg, D. A. Todd, J. M. Egan, H. A. Raja, N. H. Oberlies, O. M. Kvalheim and N. B. Cech, *J. Nat. Prod.*, 2016, **79**, 376-386.
47. F. E. Koehn and G. T. Carter, *Nat. Rev. Drug. Discov.*, 2005, **4**, 206-220.
48. C. B. Naman, R. Rattan, S. E. Nikoulina, J. Lee, B. W. Miller, N. A. Moss, L. Armstrong, P. D. Boudreau, H. M. Debonsi, F. A. Valeriote, P. C. Dorrestein and W. H. Gerwick, *J. Nat. Prod.*, 2017, **80**, 625-633.
49. 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 and P. C. Dorrestein, *J. Nat. Prod.*, 2018, **81**, 758-767.

50. H. L. Röst, T. Sachsenberg, S. Aiche, C. Bielow, H. Weisser, F. Aicheler, S. Andreotti, H.-C. Ehrlich, P. Gutenbrunner, E. Kenar, X. Liang, S. Nahnsen, L. Nilse, J. Pfeuffer, G. Rosenberger, M. Rurik, U. Schmitt, J. Veit, M. Walzer, D. Wojnar, W. E. Wolski, O. Schilling, J. S. Choudhary, L. Malmström, R. Aebersold, K. Reinert and O. Kohlbacher, *Nat. Methods*, 2016, **13**, 741-748.
51. L.-F. Nothias-Scaglia, V. Dumontet, J. Neyts, F. Roussi, J. Costa, P. Leyssen, M. Litaudon and J. Paolini, *Fitoterapia*, 2015, **105**, 202-209.
52. 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 and M. Litaudon, *ACS Chem. Biol.*, 2017, **12**, 2644-2651.
53. J. N. Anastas and R. T. Moon, *Nat. Rev. Cancer*, 2012, **13**, 11-26.
54. F. Olivon, C. Apel, P. Retailleau, P. M. Allard, J. L. Wolfender, D. Touboul, F. Roussi, M. Litaudon and S. Desrat, *Org. Chem. Front.*, 2018, **5**, 2171-2178.
55. D. J. Floros, P. R. Jensen, P. C. Dorrestein and N. Koyama, *Metabolomics*, 2016, **12**, 145.
56. M. Crüseman, E. C. O'Neill, C. B. Larson, A. V. Melnik, D. J. Floros, R. R. da Silva, P. R. Jensen, P. C. Dorrestein and B. S. Moore, *J. Nat. Prod.*, 2017, **80**, 588-597.
57. T. Luzzatto-Knaan, N. Garg, M. Wang, E. Glukhov, Y. Peng, G. Ackermann, A. Amir, B. M. Duggan, S. Ryazanov, L. Gerwick, R. Knight, T. Alexandrov, N. Bandeira, W. H. Gerwick and P. C. Dorrestein, *eLife*, 2017, **6**, e24214.
58. S. Roussel, S. Preys, F. Chauchard and J. Lallemand, in *J. Chromatogr. A*, Springer, 2014, pp. 7-59.
59. H. Martens, S. W. Bruun, I. Adt, G. D. Sockalingum and A. Kohler, *J. Chemometr.*, 2006, **20**, 402-417.
60. E. R. Britton, J. J. Kellogg, O. M. Kvalheim and N. B. Cech, *J. Nat. Prod.*, 2018, **81**, 484-493.
61. L. K. Caesar, J. J. Kellogg, O. M. Kvalheim, R. A. Cech and N. B. Cech, *Planta Med.*, 2018, **84**, 721-728.
62. O. M. Kvalheim, H.-y. Chan, I. F. F. Benzie, Y.-t. Szeto, A. H.-c. Tzang, D. K.-w. Mok and F.-t. Chau, *Chemometr. Intell. Lab. Syst.*, 2011, **107**, 98-105.
63. M. Vallet, Q. P. Vanbellingen, T. Fu, J.-P. Le Caer, S. Della-Negra, D. Touboul, K. R. Duncan, B. Nay, A. Brunelle and S. Prado, *J. Nat. Prod.*, 2017, **80**, 2863-2873.
64. F. Peypoux, J. Bonmatin and J. Wallach, *Appl. Microbiol. Biotechnol.*, 1999, **51**, 553-563.
65. R. H. Baltz, *J. Ind. Microbiol. Biotechnol.*, 2018, <https://doi.org/10.1007/s10295-10018-12115-10294>.
66. K. Kleigrew, J. Almaliti, I. Y. Tian, R. B. Kinnel, A. Korobeynikov, E. A. Monroe, B. M. Duggan, V. Di Marzo, D. H. Sherman, P. C. Dorrestein, L. Gerwick and W. H. Gerwick, *J. Nat. Prod.*, 2015, **78**, 1671-1682.
67. M. Nagarajan, V. Maruthanayagam and M. Sundararaman, *J. Appl. Toxicol.*, 2012, **32**, 153-185.
68. M. Maansson, N. G. Vynne, A. Klitgaard, J. L. Nybo, J. Melchiorson, D. D. Nguyen, L. M. Sanchez, N. Ziemert, P. C. Dorrestein and M. R. Andersen, *mSystems*, 2016, **1**, e00028-00015.
69. K. R. Duncan, M. Crüseman, A. Lechner, A. Sarkar, J. Li, N. Ziemert, M. Wang, N. Bandeira, B. S. Moore, P. C. Dorrestein and P. R. Jensen, *Chem. Biol.*, 2015, **22**, 460-471.
70. J. R. Doroghazi, J. C. Albright, A. W. Goering, K.-S. Ju, R. R. Haines, K. A. Tchalukov, D. P. Labeda, N. L. Kelleher and W. W. Metcalf, *Nat. Chem. Biol.*, 2014, **10**, 963-968.
71. A. W. Goering, R. A. McClure, J. R. Doroghazi, J. C. Albright, N. A. Haverland, Y. Zhang, K.-S. Ju, R. J. Thomson, W. W. Metcalf and N. L. Kelleher, *ACS Cent. Sci.*, 2016, **2**, 99-108.
72. R. A. McClure, A. W. Goering, K.-S. Ju, J. A. Baccile, F. C. Schroeder, W. W. Metcalf, R. J. Thomson and N. L. Kelleher, *ACS Chem. Biol.*, 2016, **11**, 3452-3460.
73. E. I. Parkinson, J. H. Tryon, A. W. Goering, K.-S. Ju, R. A. McClure, J. D. Kembal, S. Zhukovsky, D. P. Labeda, R. J. Thomson, N. L. Kelleher and W. W. Metcalf, *ACS Chem. Biol.*, 2018, **13**, 1029-1037.
74. M. T. Henke, A. A. Soukup, A. W. Goering, R. A. McClure, R. J. Thomson, N. P. Keller and N. L. Kelleher, *ACS Chem. Biol.*, 2016, **11**, 2117-2123.
75. J. Begani, J. Lakhani and D. Harwani, *Ann. Microbiol.*, 2018, **68**, 1-14.
76. H. B. Bode, D. Reimer, S. W. Fuchs, F. Kirchner, C. Dauth, C. Kegler, W. Lorenzen, A. O. Brachmann and P. Grün, *Chem. Eur. J.*, 2012, **18**, 2342-2348.
77. A. Klitgaard, J. B. Nielsen, R. J. N. Frandsen, M. R. Andersen and K. F. Nielsen, *Anal. Chem.*, 2015, **87**, 6520-6526.
78. F. Olivon, N. Elie, G. Grelier, F. Roussi, M. Litaudon and D. Touboul, *Anal. Chem.*, 2018, **90**, 13900-13908.
79. L. van der Maaten and G. Hinton, *J. Mach. Learn. Res.*, 2008, **9**, 2579-2605.
80. G. Wohlgemuth, S. S. Mehta, R. F. Mejia, S. Neumann, D. Pedrosa, T. Pluskal, E. L. Schymanski, E. L. Willighagen, M. Wilson, D. S. Wishart, M. Arita, P. C. Dorrestein, N. Bandeira, M. Wang, T. Schulze, R. M. Salek, C. Steinbeck, V. C. Nainala, R. Mistrik, T. Nishioka and O. Fiehn, *Nat. Biotechnol.*, 2016, **34**, 1099-1101.
81. H. Horai, M. Arita, S. Kanaya, Y. Nihei, T. Ikeda, K. Suwa, Y. Ojima, K. Tanaka, S. Tanaka, K. Aoshima, Y. Oda, Y. Kakazu, M. Kusano, T. Tohge, F. Matsuda, Y. Sawada, M. Y. Hirai, H. Nakanishi, K. Ikeda, N. Akimoto, T. Maoka, H. Takahashi, T. Ara, N. Sakurai, H. Suzuki, D. Shibata, S. Neumann, T. Iida, K. Tanaka, K. Funatsu, F. Matsuura, T. Soga, R. Taguchi, K. Saito and T. Nishioka, *J. Mass Spectrom.*, 2010, **45**, 703-714.
82. D. S. Wishart, T. Jewison, A. C. Guo, M. Wilson, C. Knox, Y. Liu, Y. Djoumbou, R. Mandal, F. Aziat, E. Dong, S. Bouatra, I. Sinelnikov, D. Arndt, J. Xia, P. Liu, F. Yallou, T. Bjorn Dahl, R. Perez-Pineiro, R. Eisner, F. Allen, V. Neveu, R. Greiner and A. Scalbert, *Nucl. Acids Res.*, 2013, **41**, D801-D807.
83. H. Tsugawa, T. Cajka, T. Kind, Y. Ma, B. Higgins, K. Ikeda, M. Kanazawa, J. Van der Gheynst, O. Fiehn and M. Arita, *Nat. Methods*, 2015, **12**, 523-526.
84. M. A. Stravs, E. L. Schymanski, H. P. Singer and J. Hollender, *J. Mass Spectrom.*, 2013, **48**, 89-99.
85. K. Skogerson, G. Wohlgemuth, D. K. Barupal and O. Fiehn, *BMC Bioinform.*, 2011, **12**, 321.
86. O. Spjuth, T. Helmus, E. L. Willighagen, S. Kuhn, M. Eklund, J. Wagener, P. Murray-Rust, C. Steinbeck and J. E. Wikberg, *BMC Bioinform.*, 2007, **8**, 59.
87. X. Domingo-Almenara, J. R. Montenegro-Burke, H. P. Benton and G. Siuzdak, *Anal. Chem.*, 2018, **90**, 480-489.
88. J. Wandy, Y. Zhu, J. J. J. van der Hooff, R. Daly, M. P. Barrett and S. Rogers, *Bioinformatics*, 2018, **34**, 317-318.

Journal Name

ARTICLE

89. J. J. J. van der Hooft, J. Wandy, M. P. Barrett, K. E. V. Burgess and S. Rogers, *Proc. Natl. Acad. Sci. U.S.A.*, 2016, **113**, 13738-13743.
90. S. Böcker and K. Dührkop, *J. Cheminformatics*, 2016, **8**, 5.
91. K. Dührkop, H. Shen, M. Meusel, J. Rousu and S. Böcker, *Proc. Natl. Acad. Sci. U.S.A.*, 2015, **112**, 12580-12585.
92. H. Mohimani, A. Gurevich, A. Shlemov, A. Mikheenko, A. Korobeynikov, L. Cao, E. Shcherbin, L.-F. Nothias, P. C. Dorrestein and P. A. Pevzner, *Nat. Commun.*, 2018, **9**, 4035.
93. H. Mohimani, A. Gurevich, A. Mikheenko, N. Garg, L.-F. Nothias, A. Ninomiya, K. Takada, P. C. Dorrestein and P. A. Pevzner, *Nat. Chem. Biol.*, 2016, **13**, 30-37.
94. A. Gurevich, A. Mikheenko, A. Shlemov, A. Korobeynikov, H. Mohimani and P. A. Pevzner, *Nat. Microbiol.*, 2018, **3**, 319-327.
95. H. Mohimani, R. D. Kersten, W.-T. Liu, M. Wang, S. O. Purvine, S. Wu, H. M. Brewer, L. Pasa-Tolic, N. Bandeira, B. S. Moore, P. A. Pevzner and P. C. Dorrestein, *ACS Chem. Biol.*, 2014, **9**, 1545-1551.
96. Y. Beauxis and G. Genta-Jouve, *Bioinformatics*, 2018, bty864, <https://doi.org/810.1093/bioinformatics/bty1864>.
97. S. Böcker, M. C. Letzel, Z. Lipták and A. Pervukhin, *Bioinformatics*, 2009, **25**, 218-224.
98. E. E. Bolton, Y. Wang, P. A. Thiessen and S. H. Bryant, in *Annual Reports in Computational Chemistry*, eds. R. A. Wheeler and D. C. Spellmeyer, Elsevier, 2008, vol. 4, pp. 217-241.
99. C. A. Smith, E. J. Want, G. O'Maille, R. Abagyan and G. Siuzdak, *Anal. Chem.*, 2006, **78**, 779-787.
100. D. M. Blei, A. Y. Ng and M. I. Jordan, *J. Mach. Learn. Res.*, 2003, **3**, 993-1022.
101. S. Allard, P.-M. Allard, I. Morel and T. Gicquel, *Drug Test. Anal.*, 2018, <https://doi.org/10.1002/dta.2550>.
102. M. Köck, A. Grube, I. B. Seiple and P. S. Baran, *Angew. Chem. Int. Ed.*, 2007, **46**, 6586-6594.
103. J. B. McAlpine, S.-N. Chen, A. Kutateladze, J. B. MacMillan, G. Appendino, A. Barison, M. A. Beniddir, M. W. Biavatti, S. Bluml, A. Boufridi, M. S. Butler, R. J. Capon, Y. H. Choi, D. Coppage, P. Crews, M. T. Crimmins, M. Csete, P. Dewapriya, J. M. Egan, M. J. Garson, G. Genta-Jouve, W. H. Gerwick, H. Gross, M. K. Harper, P. Hermanto, J. M. Hook, L. Hunter, D. Jeannerat, N.-Y. Ji, T. A. Johnson, D. G. I. Kingston, H. Koshino, H.-W. Lee, G. Lewin, J. Li, R. G. Linington, M. Liu, K. L. McPhail, T. F. Molinski, B. S. Moore, J.-W. Nam, R. P. Neupane, M. Niemitz, J.-M. Nuzillard, N. H. Oberlies, F. M. M. Ocampos, G. Pan, R. J. Quinn, D. S. Reddy, J.-H. Renault, J. Rivera-Chávez, W. Robien, C. M. Saunders, T. J. Schmidt, C. Seger, B. Shen, C. Steinbeck, H. Stuppner, S. Sturm, O. Tagliabatella-Scafati, D. J. Tantillo, R. Verpoorte, B.-G. Wang, C. M. Williams, P. G. Williams, J. Wist, J.-M. Yue, C. Zhang, Z. Xu, C. Simmler, D. C. Lankin, J. Bisson and G. F. Pauli, *Nat. Prod. Rep.*, 2019, **36**, 35-107.

CHAPITRE II :

MIADB - CONCEPTION D'UNE BASE DE DONNÉES D'ALCALOÏDES INDOLOMONOTERPÉNIQUES



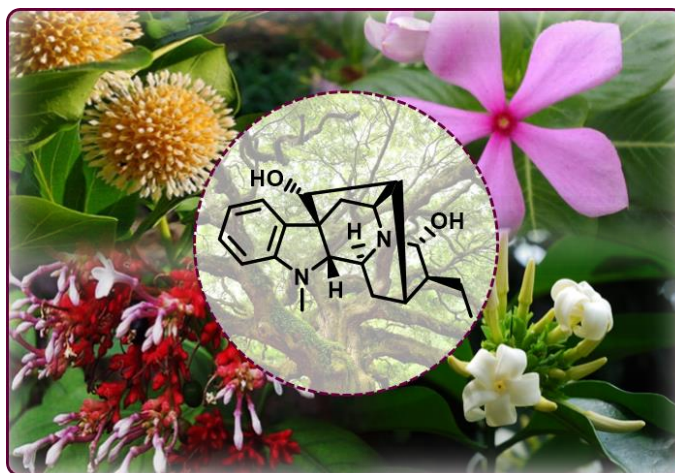


CHAPITRE II : CONCEPTION D'UNE BASE DE DONNÉES D'ALCALOÏDES INDOLOMONOTERPÉNIQUES : LA MIADB

1. INTRODUCTION

1.1 Description générale

Les alcaloïdes indolomonoterpéniques (AIMs) constituent une vaste classe de produits naturels. Avec plus de 3000 représentants décrits à ce jour, ces molécules ont été largement recherchées et étudiées au long de l'histoire de la phytochimie, en raison de leur large gamme de propriétés pharmacologiques et de leur potentiel



thérapeutique.^{22, 23} En effet, plusieurs AIMs ont été ou sont encore utilisés comme des agents anticancéreux, dans le traitement du paludisme, de l'hypertension, des arythmies cardiaques, dans certains troubles psychiatriques, etc.²⁴

Figure 1. Quelques Gentianales et un exemple d'alcaloïde indolomonoterpénique

En ce qui concerne leur distribution dans la nature, ces alcaloïdes sont quasi-exclusivement produits par des espèces de l'ordre des Gentianales (Asteridées, Lamiidées), plus précisément par celles appartenant aux familles suivantes : Apocynaceae, Loganiaceae, Rubiaceae et Gelsemiaceae (*i.e.* l'ensemble de cet ordre à l'exception des Gentianaceae ; **Figure 1**).^{24, 25, 26, 27, 28, 29, 30, 31} La production de AIMs n'est pas constante dans ces familles.

²² L. F. Szabó, *Molecules*, 2008, **13**, 1875-1896.

²³ P. M. Dewick, 3rd edn., 2009.

²⁴ M. Lahlou, *Pharmacol Pharm.*, 2013, **4**, 17-31.

²⁵ https://www.flickr.com/photos/tony_rodd/2205187291/, (accessed 22-09-2018, 2018).

²⁶

[https://keys.lucidcentral.org/keys/v3/eafrinet/weeds/key/weeds/Media/Html/Catharanthus_roseus_\(Madagascar_Periwinkle\).htm](https://keys.lucidcentral.org/keys/v3/eafrinet/weeds/key/weeds/Media/Html/Catharanthus_roseus_(Madagascar_Periwinkle).htm), (accessed 22-09-2018, 2018).

²⁷ <https://mavcure.com/sarpagandha/>, (accessed 22-09-2018, 2018).

²⁸ <https://micromedix.me/2017/06/27/microdosis-de-akuamma-nueva-medicina-alternativa-para-tratar-la-esquizofrenia-y-otros-trastornos-mentales/akuamma/>, (accessed 22-09-2018, 2018).

²⁹ J. Stöckigt, L. Barleben, et al., *Plant Physiol Biochem.*, 2008, **46**, 340-355.

³⁰ J. Le Men & W. I. Taylor, *Experientia*, 1965, **21**, 508-510.

³¹ M. W. Chase, M. Christenhusz, et al., *Bot. J. Linn. Soc.*, 2016, **181**, 1-20.

On rencontre exceptionnellement ces molécules chez des Asterids basales : Icacinaceae (ordre des Icacinales, APG IV), Nyssaceae (avec le genre *Camptotheca* ; ordre des Cornales), Cornaceae (incluant les Alangiaceae depuis l'APG II ; ordre des Cornales). Enfin, des AIMs ont également été décrits chez des Caprifoliaceae (anc.-Dipsacaceae) des genres *Pterocephalus*³² et *Triosteum*³³ (Dipsacales, chez les Asterids du groupe de Campanulids).

1.2 Biogenèse et classification

Tous les alcaloïdes indolomonoterpéniques ont comme point de départ commun la strictosidine (**3**), molécule hétérosidique qui est biosynthétisée à partir d'une molécule de tryptamine (**1** ; produit de décarboxylation du tryptophane) et du sécologanoside (un glucoside de monoterpène de la classe des séco-iridoïdes, **2**) (**Figure 2**).^{22, 29} La biosynthèse de **3** est possible grâce à la strictosidine-synthase, une enzyme qui catalyse une condensation de Pictet-Spengler stéréosélective entre les deux substrats.²⁹ C'est à partir de ce précurseur **3** que les plus de 3000 AIMs décrits dans la nature sont synthétisés (à l'exception potentielle de ceux de Dipsacales), par des réactions en cascade enzymocatalysées, généralement après action d'une glucosidase libérant une génine réactive porteuse d'une fonction aldéhyde et d'une fonction énole. De nombreuses enzymes impliquées dans la biosynthèse des AIMs ont pu être identifiées, avec notamment des réductases pouvant être impliquées dans la formation d'entités iminium comme pivots entre différents squelettes. Des peroxydases, des hydroxylases, des méthyl-transférases, des acétyl-transférases ont également été caractérisées.²³ De nombreuses enzymes impliquées dans les réarrangements, les polymérisations et les décorations des AIMs restent inconnues.

Ces produits naturels possèdent une diversité structurale importante et une variété de squelettes d'appréhension complexe, ce qui a conduit les spécialistes du domaine à proposer plusieurs modalités de classement. C'est ainsi que ces molécules ont été ordonnées selon l'organisation de l'enchaînement carboné dans la partie monoterpénique caractéristique des AIMs (types Ia, Ib, II, III - **Figure 3**)^{23, 30}, ou par leur biogenèse de manière plus fine (11 types).²² Leur squelette permet également d'établir une classification (environ 45 squelettes décrits à ce jour, présentés en annexe de l'article figurant dans ce chapitre).³⁴ Enfin, des polymérisations très diverses sont observées (homo ou hétéro-dimères, parfois trimères³⁵ et tétramères³⁶), *via* des mono- ou des di-pontages de type C-C (aryl-aryl, aryl-aliphatique, exométhylène³⁷), éther, N-C³⁸, mais aussi thio-éther³⁹.

³² D. Gülcemal, M. Masullo, et al., *Magn. Reson. Chem.*, 2010, **48**, 239-243.

³³ X. Huang, Y. Li, et al., *Phytochem Lett.*, 2014, **7**, 30-34.

³⁴ J. Buckingham, K. H. Baggaley, et al., CRC press, 2010.

³⁵ Z.-W. Liu, J. Zhang, et al., *J. Org. Chem.*, 2018, **83**, 10613-10618.

³⁶ Y. Hirasawa, S. Miyama, et al., *Org. Lett.*, 2009, **11**, 5718-5721.

³⁷ E. O. N'Nang Obiang, G. Genta-Jouve, et al., *Org. Lett.*, 2017, **19**, 6180-6183.



D'autre part, des MIAs « chimériques » comportant des carbones issus d'autres voies biosynthétiques, généralement « décoratifs », sont également décrits.

Ainsi, les différentes modalités de classement (types, squelettes) excluent quelques exemplaires de ces métabolites, notamment avec l'isolement de nouveaux AIMs à motifs chimiques originaux,⁴⁰ sans pour autant entraîner le réexamen des grands groupes usuellement proposés.

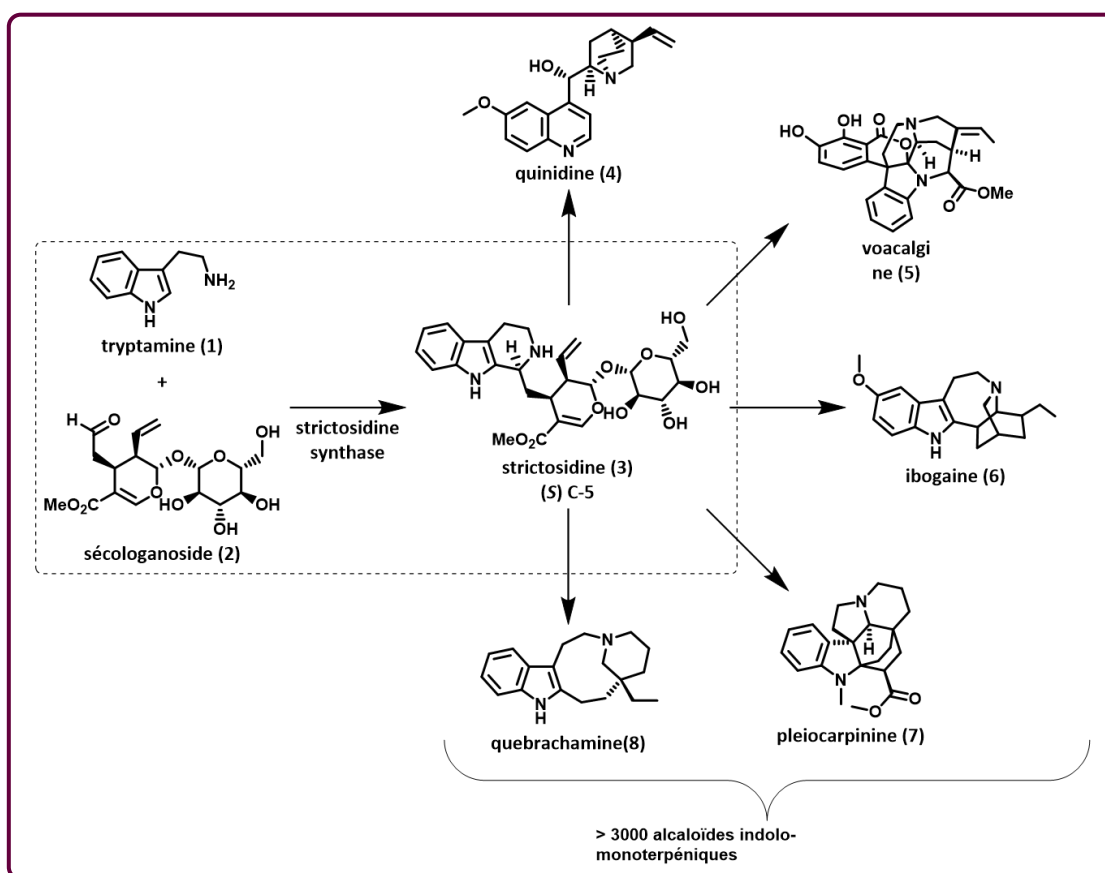


Figure 2. Biosynthèse des alcaloïdes indolomonoterpéniques à partir de la strictosidine

³⁸ A. E. Nugroho, Y. Hirasawa, et al., *J. Nat. Prod.*, 2009, **72**, 1502-1506.

³⁹ E. Otogo N'Nang, G. Bernadat, et al., *Org. Lett.*, 2018, **20**, 6596-6600.

⁴⁰ A. E. Fox Ramos, C. Alcover, et al., *J. Nat. Prod.*, 2017, **80**, 1007-1014.

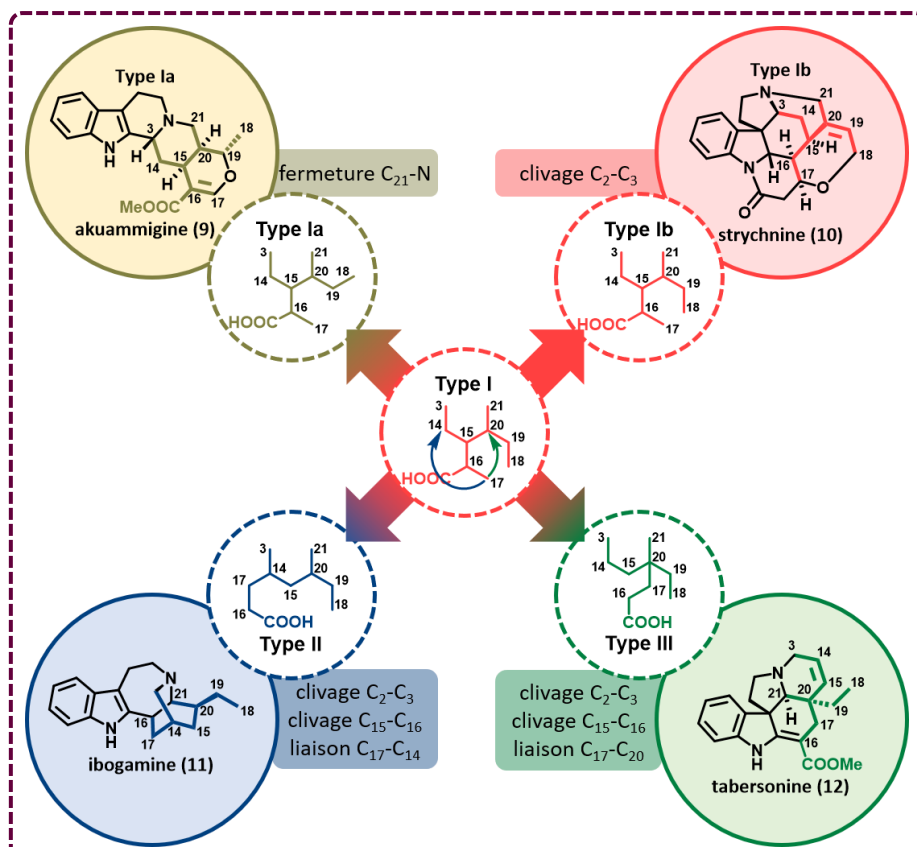


Figure 3. Classification des alcaloïdes indolomonoterpéniques d'après Le Men et Taylor (1965)

1.3 Biogenèse

Ces dernières années plusieurs travaux ont été entrepris pour mieux comprendre la biosynthèse des alcaloïdes indolomonoterpéniques. La tendance actuelle pour les réaliser exploite, en raison de son efficacité, la transposition des voies de biosynthèses générales dans des micro-organismes. Dans ce contexte, les équipes de V. Courdavault et S. E. O'Connor ont entrepris récemment la recherche des enzymes manquantes dans la biosynthèse de la vinblastine par *Catharanthus roseus* (L.) G.Don. Pour réussir cette tâche, des gènes impliqués dans la biosynthèse de la vincristine et de la vinblastine ont été transfectés dans une souche de *Escherichia coli*. Les gènes ont ainsi été exprimés, puis les hypothèses de biosynthèses testées avec plusieurs substrats de ces voies. Après de multiples expériences, deux nouvelles enzymes ont été découvertes : la PAS (précondylocarpine acétate synthase) et la DPAS (dihydroprécondylocarpine synthase). La PAS catalyse la transformation de l'acétate de stemmadénine en acétate de précondylocarpine, en tant que la DPAS, le passage de cette dernière molécule à l'acétate de dihydroprécondylocarpine. De plus, les fonctions catalytiques de deux enzymes, la CS (catharanthine synthase) et la TS (tabersonine synthase) ont été décrites pour la première fois.⁴¹

⁴¹ L. Caputi, J. Franke, et al., *Science*, 2018, **360**, 1235-1239.



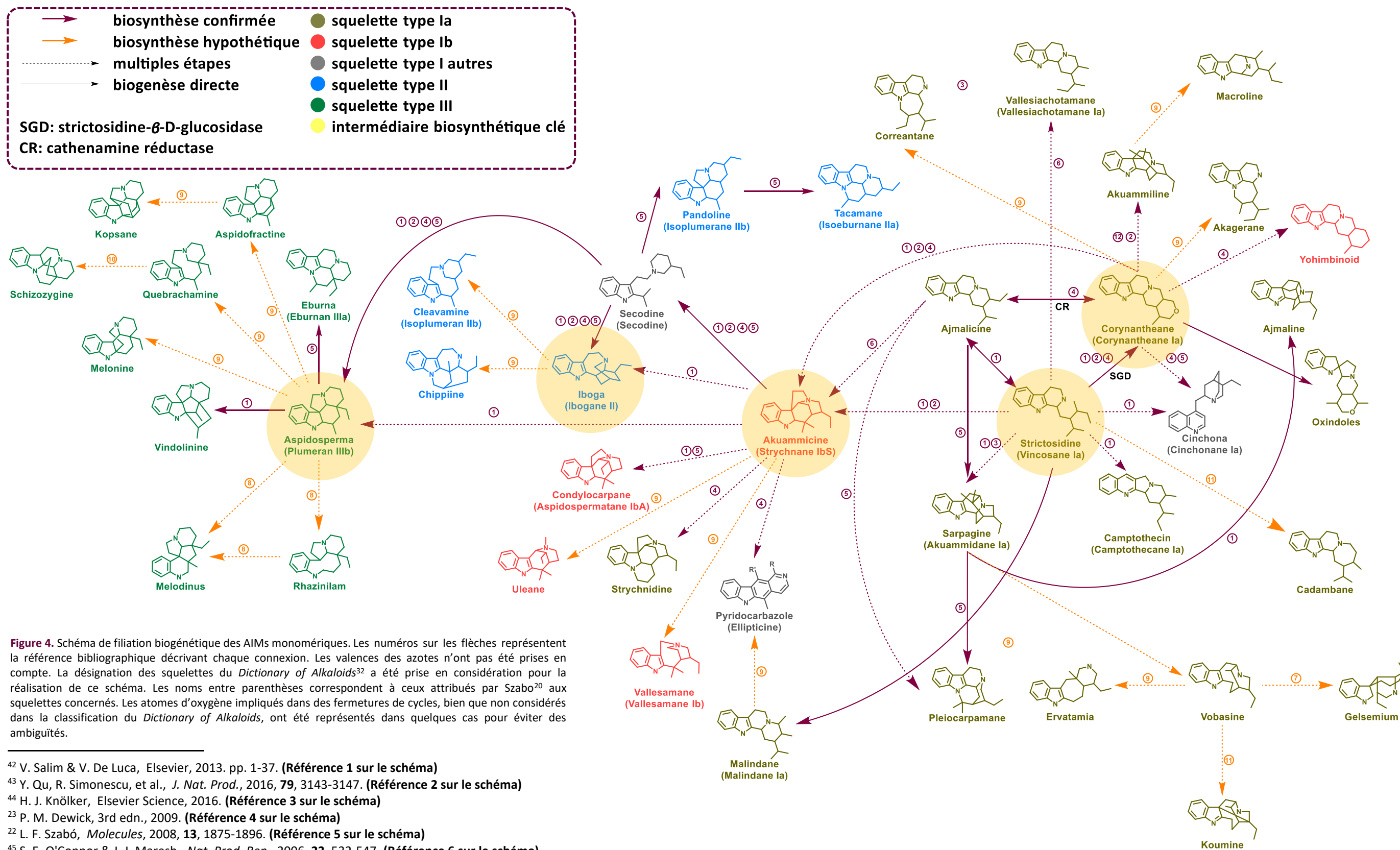
Dans ce contexte, nous proposons une mise en relation des principaux squelettes des AIMs avec les voies de filiations biogénétiques proposées dans la littérature : un schéma montrant les différents types de squelettes, reliés par rapport à leur interconnexions biosynthétiques a été conçu (**Figure 4**). La classification qui a été prise en compte est celle proposée par le *Dictionary of Alkaloids*³⁴.

Ce schéma illustre en particulier les éléments-clés suivants, qui ne demandent qu'à être éprouvés par des approches dérégulatives :

- Les squelettes du type I trouveraient principalement leur origine dans les squelettes « akuammicine », « corynanthéane » et « strictosidine ».
- Les squelettes « sécodine » et « akuammicine », tous les deux du type I, seront à l'origine de tous les squelettes du type II, sans exception.
- Les AIMs de type III ont tous comme précurseur commun le squelette « aspidospermane », également du type III. Celui-ci aurait son origine dans les squelettes « akuammicine » ou « secodine », tous les trois du type I.

Une mise en relation des dimères indolomonoterpéniques n'a pas été incluse mais mériterait d'être conçue. Un filtre taxonomique serait intéressant à appliquer.





⁴² V. Salim & V. De Luca, Elsevier, 2013, pp. 1-37. (Référence 1 sur le schéma)

⁴³ Y. Qu, R. Simonescu, et al., *J. Nat. Prod.*, 2016, **79**, 3143-3147. (Référence 2 sur le schéma)

⁴⁴ H. J. Knölker, Elsevier Science, 2016. (Référence 3 sur le schéma)

²³ P. M. Dewick, 3rd edn., 2009. (Référence 4 sur le schéma)

²² L. F. Szabó, *Molecules*, 2008, **13**, 1875-1896. (Référence 5 sur le schéma)

⁴⁵ S. E. O'Connor & J. J. Maresh, *Nat. Prod. Rep.*, 2006, **23**, 532-547. (Référence 6 sur le schéma)

⁴⁶ M. Kitajima, N. Kogure, et al., *Org. Lett.*, 2003, **5**, 2075-2078. (Référence 7 sur le schéma)

⁴⁷ Y. Yang, Y. Bai, et al., *Org. Lett.*, 2014, **16**, 6216-6219. (Référence 8 sur le schéma)

³⁴ J. Buckingham, K. H. Baggaley, et al., CRC press, 2010. (Référence 9 sur le schéma)

⁴⁸ C. H. Heathcock, M. H. Norman, et al., *J. Org. Chem.*, 1990, **55**, 798-811. (Référence 10 sur le schéma)

⁴⁹ R. T. Brown, D. M. Duckworth, et al., *Tetrahedron Lett.*, 1991, **32**, 1987-1990. (Référence 11 sur le schéma)

⁵⁰ S. Benayad, K. Ahamada, et al., *Eur. J. Org. Chem.*, 2016, **2016**, 1494-1499. (Référence 12 sur le schéma)

1.4 MIADB (*Monoterpene Indole Alkaloids DataBase*) : du concept à l'application

Au sein de notre équipe, sous l'impulsion des Dr Mehdi Beniddir et Laurent Evanno en particulier, nous nous intéressons tout particulièrement à la découverte d'alkaloïdes indolomonoterpéniques à motifs chimiques ou à squelettes originaux et à leurs relations de filiation.

Au fil des années, la tâche de l'isolement d'analogues nouveaux en série AIM est devenue de plus en plus compliquée car les métabolites majoritaires, donc les plus susceptibles d'être obtenus, ont déjà été décrits depuis longtemps dans un grand nombre de Gentianales. Pour cette raison, il est malheureusement fréquent que de longues étapes de purification aboutissent à l'isolement de molécules connues. Cependant, la littérature des années 60 et 70 recèle des mentions d'analogues identifiés, parfois isolés, mais de structures non élucidées. Dans ce contexte, nous nous sommes demandé comment cibler efficacement de nouveaux membres des AIMs. L'évolution constante et rapide des techniques analytiques déréplicatives, assistées par l'analyse comparative de données spectrales constituent un atout dans une approche de ciblage de composés nouveaux. Nous nous sommes intéressés à la technique du « *molecular networking* »⁶. Cet outil, basé sur la spectrométrie de masse tandem haute résolution, permet de cartographier les composés d'un échantillon en calculant des scores de similarité entre eux, comme cela a été décrit dans le chapitre précédent. Il a été mis à profit par le Dr Elvis Otogo N'Nang Obiang, qui a pu grâce à cette approche obtenir des MIAs aux structures insoupçonnées^{37, 39, 51} en parallèle des travaux présentés ici. Nous avons appliqué cette approche analytique à un extrait alcaloïdique d'écorces de *Geissospermum laeve*. Cependant, malgré l'obtention d'un réseau moléculaire correct, l'information que nous pouvions en tirer n'était pas suffisante pour une exploration efficace de la diversité chimique de l'échantillon.

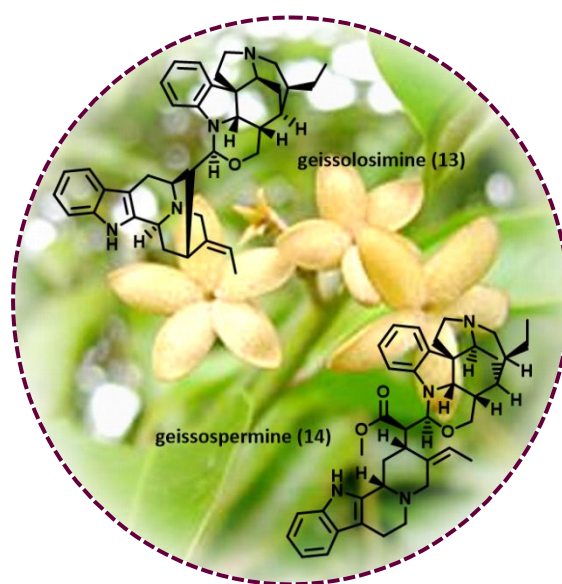


Figure 5. Bisindoles majoritaires de *G. laeve*

La recherche bibliographique des composés déjà décrits dans cette espèce a été faite en parallèle. De ce fait, nous avons pu relever leurs masses précises pour pouvoir chercher les m/z des ions $[M+H]^+$ ou $[M]^+$ correspondant dans le réseau moléculaire obtenu. Bien que nous les y retrouvions, l'identité de ces molécules ne pouvait pas être confirmée en se basant uniquement sur cette observation.

⁵¹ E. Otogo n'ngang, Paris Saclay, 2018.



C'est à ce moment que nous avons pensé à tirer profit de l'historique de recherche du laboratoire sur les AIMs. En effet, depuis les années 1970s, de nombreux composés de cette famille y ont été isolés. De plus, *G. laeve* étant une espèce déjà étudiée par notre équipe, quelques bisindoles isolés de cette plante étaient à notre disposition. Nous avons donc analysé par CLHP-SM/SM la geissolosimine (**13**) et la geissospermine (**14**) (**Figure 5**).⁵² Les données spectrales obtenues ont ainsi été soumises au GNPS (voir Chapitre I) séparément pour expérimenter la création de réseaux à partir de composés purs.

Nous avons alors obtenu un nœud individuel par composé (« *self-loop* »). Ce résultat nous a donc indiqué que l'information de ces molécules pures pouvait être incorporée au réseau complet de la plante. De ce fait, nous avons traité l'information SM/SM de l'extrait et celle des deux molécules témoins ensemble pour créer un seul réseau moléculaire. Nous avons ainsi observé pour la toute première fois un « *match* ». En effet, l'algorithme du GNPS avait bien reconnu la geissolosimine (**13**) et la geissospermine (**14**) employées comme témoins comme étant présentes dans l'extrait alcaloïdique. Nous avons donc réussi à dérépliquer nos deux premiers alcaloïdes indolomonoterpéniques dans un mélange complexe.

Malgré cette réussite, l'information tirée du réseau n'était pas encore suffisante. Nous avons alors pensé que, si deux substances de référence pouvaient être intégrées au réseau, il y avait toujours la possibilité d'en ajouter d'autres. Cette réflexion nous a conduits à réunir le reste des alcaloïdes indolomonoterpéniques de notre laboratoire (40 composés), et ceux de l'ICSN-CNRS à Gif-sur-Yvette. Nous avons donc analysé les 53 composés récupérés et leurs données spectrales SM/SM ont été obtenues. Avec une telle quantité d'informations, nous avons rapidement perçu leur utilité pour la déréplication d'autres extraits issus de taxons d'intérêt. C'est ainsi que l'idée de créer une base de données d'AIMs est née.

Bien que l'appellation « base de données d'alcaloïdes indolomonoterpéniques » ait eu du succès, nous avons opté pour la version anglophone pour pouvoir donner un acronyme au nom. Le choix était donc fait, « *Monoterpene Indole Alkaloids DataBase* », abrégé en MIADB serait le nom de cette base de données spectrales.

De cette manière, nous avons utilisé cette première version de la MIADB pour explorer l'espace chimique de *G. laeve*. La totalité de l'information concernant ce projet peut être retrouvée dans le troisième chapitre de ce manuscrit. La réussite de ces travaux nous a encouragés à continuer de nourrir cette base de données avec de nouvelles molécules. C'est ainsi que nous avons commencé à y rajouter les composés que nous avons isolés, à l'issue de sa propre utilisation.

⁵² <http://www.beljanskiblog.com/pao-pereira-commerce-equitable-et-durable/>, (accessed 15-09-2018, 2018).



Nous nous sommes ensuite demandé s'il était possible d'élargir encore plus cette base de données. Cela avait comme but d'englober plus de squelettes d'alcaloïdes indolomonoterpéniques, menant à une dérégulation encore plus efficace. Nous avons donc pensé à établir des collaborations avec des laboratoires travaillant avec ces composés. De cette manière, des laboratoires en Espagne, en Belgique et en Malaisie, ainsi que d'autres équipes en France ont collaboré au projet afin d'enrichir cette base de données (**Tableau 1**). Grâce à leur contribution, les spectres SM/SM de 117 alcaloïdes indolomonoterpéniques ont y été rajoutés, portant ainsi à 172 le nombre de molécules incluses dans la MIADB.

Du fait que cette base de données était le résultat d'une collaboration scientifique, nous avons décidé de la rendre publique aux chercheurs de la planète entière. C'est ainsi que tous les spectres composant la MIADB ont été mis en ligne sur le site du GNPS (<https://gnps.ucsd.edu/ProteoSAFe/gnpslibrary.jsp?library=MIADB>). De cette façon, nous avons réussi à mettre en place un outil au service de la communauté scientifique internationale.

L'accessibilité de la MIADB sur le site du GNPS a changé aussi la façon de déréguler les extraits de plantes. Au départ, la collection de spectres SM/SM de notre échantillon était traitée sur le GNPS en même temps que les données des substances de référence. Les témoins de la MIADB avaient donc une influence sur la topologie des réseaux. A présent, l'information SM/SM de la MIADB étant disponible sur le GNPS, seules les données de l'échantillon sont à envoyer sur la plateforme. Les données de la MIADB n'ont plus qu'un rôle d'annotation et permettent d'obtenir des réseaux plus « naturels ».

On notera que les analyses CL-SM/SM des extraits d'intérêt et des AIMs témoins sont réalisées dans des conditions standardisées (voir ci-dessous). Nous disposons ainsi d'une information complémentaire : les temps de rétention des produits. Ceux-ci constituent également un élément important pour l'attribution d'un degré de confiance pouvant être apporté à un *match*.



Tableau 1. Liste de laboratoires qui ont contribué à la constitution de la MIADB

Laboratoire contributeur	Directeur	Pays	Origine des produits
Équipe "Pharmacognosie-Chimie des Substances Naturelles" BioCIS, Univ. Paris-Sud, CNRS, Université Paris-Saclay	Dr. Bruno Figadère	France	Commercial, extraction et hémisynthèse
Centre for Natural Products and Drug Discovery, University of Malaya	Prof. Dr. Mohd Rais Bin Mustafa	Malaisie	Extraction
EA921, SONAS, SFR QUASAV, UBL, Université d'Angers	Prof. Pascal Richomme-Peniguel / Dr. Marie-Agnès Jacques	France	Extraction
Institute of Chemical Research of Catalonia (ICIQ), Barcelona Institute of Science and Technology	Prof. Miquel A. Pericàs	Espagne	Synthèse totale
Departament de Química Orgànica i Analítica, Universitat Rovira i Virgili, C	Prof. Michel Frédéric	Espagne	Synthèse totale
Laboratory of Pharmacognosy, CIRM, University of Liège	Prof. Michel Frederich	Belgique	Extraction
Laboratoire de Pharmacognosie, UMR/CNRS 8638 COMETE, Faculté de Pharmacie de Paris, Université Paris Descartes	Prof. Sylvie Michel	France	Extraction
Laboratoire Ecologie et Biologie des Interactions, Équipe Microbiologie de l'Eau, UMR CNRS 7267, Université de Poitiers	Prof. Jean Marc Berjeaud	France	Extraction
Institut de Chimie des Substances Naturelles, Équipe de Substances Naturelles, CNRS-ICSN, UPR 2301, Université Paris-Saclay	Dr. Angela Marinetti / Dr. Fanny Roussi / Dr. Marc Litaudon	France	Extraction
EA1069 Laboratoire de Chimie des Substances Naturelles, Faculté de Pharmacie, Université de Limoges	Prof. Vincent Sol	France	Extraction



Nos conditions de travail sont les suivantes :

- **Concentration des extraits et des témoins** : 1,0 mg/mL dans le MeOH ;
- **Volume d'injection** : 5 μ L ;
- **Phase stationnaire** : colonnes analytiques octadécylsilylées (C₁₈) Sunfire® C₁₈ Waters (150 \times 2,1 mm ; 3,5 μ m) ou XBridge® C₁₈ Waters (150 \times 2,1 mm ; 3,5 μ m) et précolonnes correspondantes (10 mm) ;
- **Phase mobile** : gradient d'élution MeOH/H₂O + 0,1% HCOOH (5/95 \rightarrow 100/0 en 30 min) ;
- **Débit** : 0,25 mL/min ;
- Colonne non thermostatée.

Le matériel employé et les conditions d'analyse spectrométrique (dont les énergies de collision employées) sont décrits dans les parties expérimentales des chapitres suivants.

La mise au point des conditions analytiques et le traitement des données pour les témoins AIMs ont également impliqué d'autres membres du laboratoire :

- Charlotte Alcover,
- Elvis Otogo N'Nang Obiang,
- Gaëla Cauchie,
- Hazrina Hazni,
- Mehdi Beniddir.

L'évolution des techniques de déréplication et d'annotation des composés dans des mélanges complexes reste importante. Nous avons été témoins du développement de plusieurs outils pour améliorer la façon de cibler des molécules. Parmi ces techniques, nous avons décidé de tester la version 0.2.2 de MetWork (développé par le Dr Grégory Genta-Jouve, disponible sur <https://metwork.pharmacie.parisdescartes.fr/>), un algorithme qui permet une propagation de l'annotation *in silico*.²¹ Pour ce faire, il faut comme point de départ un « match » avec une base de données de référence. MetWork assure l'anticipation des structures de dérivés de composés connus, *via* des réactions métaboliques sélectionnées, puis génère des spectres SM/SM simulés avec l'outil CFM-ID® (*Competitive fragmentation modeling for metabolite identification*; <http://cfmid.wishartlab.com/predict>)¹³, pour ensuite les confronter avec les données SM/SM expérimentales, assurant une annotation virtuelle des nœuds. C'est ainsi que le projet d'exploration de l'espace chimique d'*Alstonia balansae* en utilisant cette technique avec la MIADB comme référence a été pensé. Les détails concernant ce sujet peuvent être retrouvés dans le quatrième chapitre de ce manuscrit.



Le chemin parcouru par notre équipe pour constituer la MIADB comme base de données internationale a été long, et a comporté quelques obstacles que nous avons réussi à surmonter. C'est ainsi qu'à l'heure actuelle, l'information SM/SM de 172 alcaloïdes indolomonoterpéniques peut être utilisée par n'importe quelle équipe de recherche sur la planète. Nous espérons que la MIADB pourra encore être nourrie avec de nouvelles molécules ciblées et isolées grâce à son utilisation (**Figure 6**).

Cette base de données a fait l'objet d'un dépôt sur la plateforme *Metabolights* (<https://www.ebi.ac.uk/metabolights/MTBLS142>) ainsi que sur le GNPS (<https://gnps.ucsd.edu/ProteoSAFe/gnpslibrary.jsp?library=MIADB>), et d'un article, soumis en octobre 2018 au journal *Scientific Data*, qui figure ci-après et qui constitue le corps de ce chapitre.



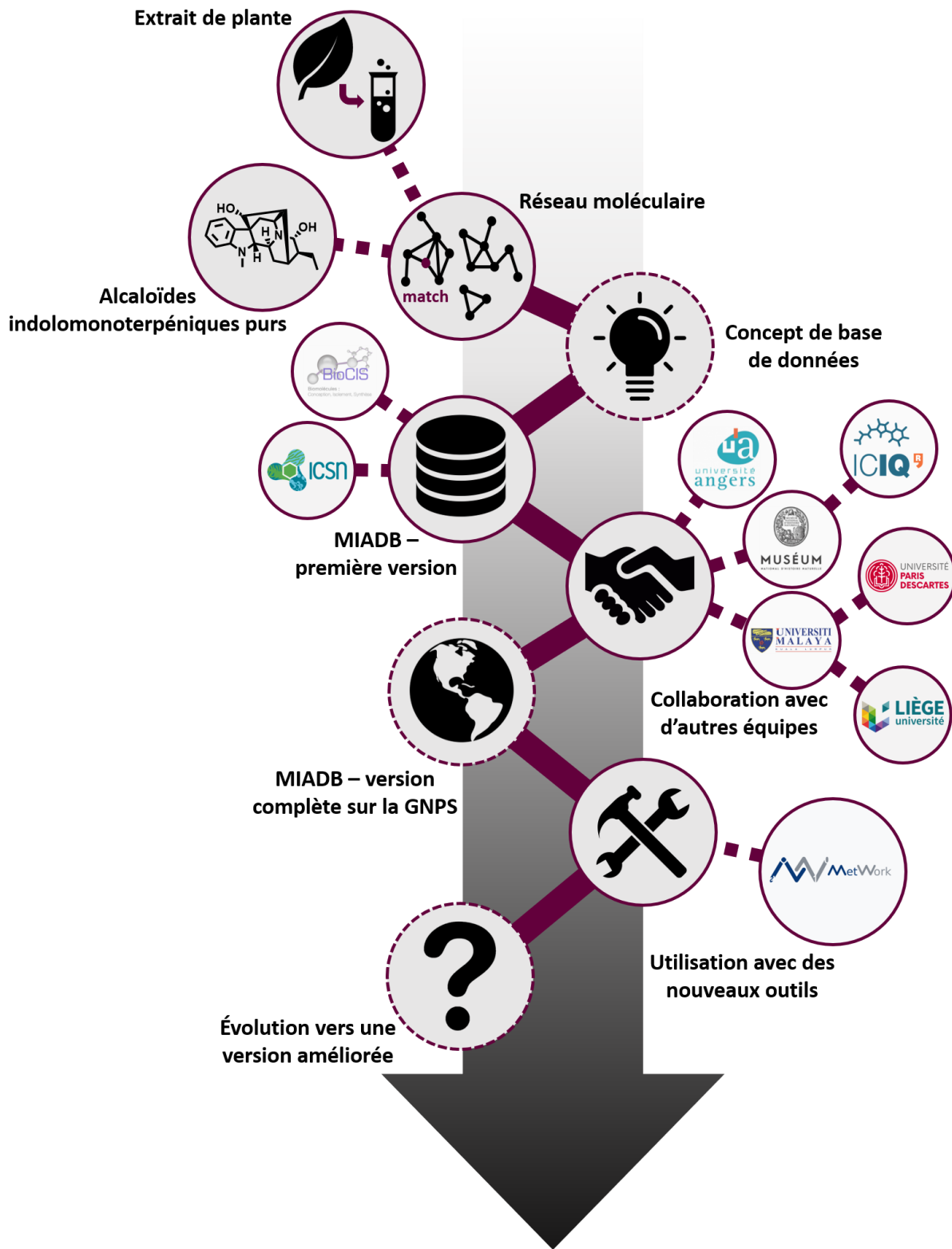


Figure 6. Évolution de la MIADB au cours du temps





MIADB : Base de données spectrale SM/SM d'alcaloïdes indolomonoterpéniques

Article publié dans *Scientific Data*, 2019

Résumé : Cet article décrit la constitution et la construction d'une base de données d'alcaloïdes indolomonoterpéniques (MIADB), constituée d'une collection cumulée de 172 spectres de spectrométrie de masse en tandem (SM/SM) issus de plusieurs projets de recherche menés dans huit laboratoires de chimie de produits naturels depuis les années 1960. Toutes les données ont été annotées et organisées pour promouvoir leur réutilisation par la communauté scientifique. Étant une collection unique de ces produits naturels complexes, ces données peuvent être utilisées pour guider la déréplication et le ciblage de nouveaux alcaloïdes indolomonoterpéniques dans des mélanges complexes lors de l'application d'approches informatiques, telles que la technique du « *molecular networking* ». Chaque spectre a son propre numéro d'accès, du CCMSLIB00004679916 au CCMSLIB00004680087 sur le GNPS. La MIADB est disponible en téléchargement direct à partir de MetaboLights sous l'identifiant : MTBLS142 (<https://www.ebi.ac.uk/metabolights/MTBLS142>).

Cet article décrit également une validation, constituée de la confrontation de la MIADB avec :

- un extrait de feuilles de pervenche de Madagascar (*Catharanthus roseus*, Apocynaceae), réalisée par le Dr. Pierre Le Pogam. La plante a été choisie en raison de l'excellente caractérisation de ses AIMs.
- les AIMs déjà présents sur la plateforme GNPS.





2. PUBLICATION ASSOCIÉE

SCIENTIFIC DATA 

2019

Collected mass spectrometry data on monoterpene indole alkaloids from natural product chemistry research

Alexander E. Fox Ramos, Pierre Le Pogam, Charlotte Alcover, Elvis Otogo N'Nang, Gaëla Cauchie, Hazrina Hazni, Khalijah Awang, Dimitri Bréard, Antonio M. Echavarren, Michel Frederich, Thomas Gaslonde, Marion Girardot, Raphaël Grougnet, Mariia S. Kirillova, Marina Kritsanida, Christelle Lémus, Anne-Marie Le Ray, Guy Lewin, Marc Litaudon, Lengo Mambu, Sylvie Michel, Fedor M. Miloserdov, Michael E. Muratore, Pascal Richomme-Peniguel, Fanny Roussi, Laurent Evanno, Erwan Poupon, Pierre Champy*, and Mehdi A. Beniddir*

Abstract: This Data Descriptor announces the submission to public repositories of the monoterpene indole alkaloid database (MIADB), a cumulative collection of 172 tandem mass spectrometry (MS/MS) spectra from multiple research projects conducted in eight natural product chemistry laboratories since the 1960s. All data have been annotated and organized to promote reuse by the community. Being a unique collection of these complex natural products, these data can be used to guide the dereplication and targeting of new related monoterpene indole alkaloids within complex mixtures when applying computer-based approaches, such as molecular networking. Each spectrum has its own accession number from CCMSLIB00004679916 to CCMSLIB00004680087 on the GNPS. The MIADB is available for download from MetaboLights under the identifier: MTBLS142 (<https://www.ebi.ac.uk/metabolights/MTBLS142>).





Collected mass spectrometry data on monoterpene indole alkaloids from natural product chemistry research

Authors

Alexander E. Fox Ramos¹, Pierre Le Pogam¹, Charlotte Fox Alcover¹, Elvis Otogo N'Nang¹, Gaëla Cauchie¹, Hazrina Hazni^{1,2}, Khalijah Awang², Dimitri Bréard³, Antonio M. Echavarren^{4,5}, Michel Frédérick⁶, Thomas Gaslonde⁷, Marion Girardot⁸, Raphaël Grougnet⁷, Mariia S. Kirillova⁴, Marina Kritsanida⁷, Christelle Lémus⁷, Anne-Marie Le Ray³, Guy Lewin¹, Marc Litaudon⁹, Lengo Mambu¹⁰, Sylvie Michel⁷, Fedor M. Miloserdov⁴, Michael E. Muratore⁴, Pascal Richomme-Peniguel³, Fanny Roussi⁹, Laurent Evanno¹, Erwan Poupon¹, Pierre Champy¹ & Mehdi A. Beniddir¹

Affiliations

1. Équipe "Pharmacognosie-Chimie des Substances Naturelles" BioCIS, Univ. Paris-Sud, CNRS, Université Paris-Saclay, 5 Rue J.-B. Clément, 92290 Châtenay-Malabry, France.
2. Centre for Natural Products and Drug Discovery, University of Malaya, Jalan Universiti, 50603 Kuala Lumpur, Wilayah Persekutuan Kuala Lumpur, Malaysia.
3. EA921, SONAS, SFR QUASAV, UBL/Angers University, Campus du végétal, 42 rue Georges Morel, 49070 Beaucouzé, France.
4. Institute of Chemical Research of Catalonia (ICIQ), Barcelona Institute of Science and Technology, Avenue Països Catalans 16, 43007 Tarragona, Spain.
5. Departament de Química Organica i Analítica, Universitat Rovira i Virgili, C/Marcel·lí Domingo s/n, 43007 Tarragona, Spain.
6. Laboratory of Pharmacognosy, CIRM, University of Liège, Quartier Hopital, 15 Avenue Hippocrate, Sart Tilman, 4000 Liège, Belgium.
7. Laboratoire de Pharmacognosie, UMR/CNRS 8638 COMETE, Faculté de Pharmacie de Paris, Université Paris Descartes, Sorbonne Paris Cité, 4 Avenue de l'Observatoire, 75006 Paris, France.
8. Laboratoire Écologie et Biologie des Interactions, Équipe Microbiologie de l'Eau, UMR CNRS 7267, Université de Poitiers, 5 rue Albert Turpain, 86073 Poitiers CEDEX 09, France.
9. Institut de Chimie des Substances Naturelles, CNRS-ICSN, UPR 2301, Université Paris-Saclay, 1 Avenue de la Terrasse, 91198 Gif-sur-Yvette, France.
10. Département de Pharmacognosie, Laboratoire PEIRENE-EA 7500, Faculté de Pharmacie, Université de Limoges, 2 rue du Dr Marcland, 87025 Limoges CEDEX, France.

Corresponding author: Mehdi A. Beniddir (mehdi.beniddir@u-psud.fr)

Abstract

This Data Descriptor announces the submission to public repositories of the monoterpene indole alkaloid database (MIADB), a cumulative collection of 172 tandem mass spectrometry (MS/MS) spectra from multiple research projects conducted in eight natural product chemistry laboratories since the 1960s. All data have been annotated and organized to promote reuse by the community. Being a unique collection of these complex natural products, these data can be used to guide the dereplication and targeting of new related monoterpene indole alkaloids within complex mixtures when applying computer-based approaches, such as molecular networking. Each spectrum has its own accession number from CCMSLIB00004679916 to CCMSLIB00004680087 on the GNPS. The MIADB is available for download from MetaboLights under the identifier: MTBLS142 (<https://www.ebi.ac.uk/metabolights/MTBLS142>).

Background & Summary

Monoterpene indole alkaloids (MIAs) constitute a broad class of nitrogen-containing plant-derived natural products composed of more than 3000 members.¹ This natural product class is found in hundreds of plant species from the Apocynaceae, Loganiaceae, Rubiaceae, Icacinaceae, Nyssaceae, and Gelsemiaceae plant families. Throughout the six past decades, the structural intricacies and biological activities of these molecules have captured the interest of many researchers all over the world.² Examples of MIAs are the antimalarial drug of choice till the mid of the last century, quinine; the antihypertensive reserpine, and vincristine and vinblastine, which are used directly or as derivatives for the treatment of several cancer types. Recently, much effort was directed toward understanding and manipulating the underlying biosynthetic pathways of MIAs in order to engineer them in microorganisms to allow industrial production of medically relevant compounds.³⁻⁵ Although a large amount of knowledge has been accumulated concerning the early steps⁶⁻⁸ and the assembly of key intermediates, many questions are still unanswered, and the discovery of new members of this family may illuminate unexpected enzymes involved in the biosynthesis of this intriguing group of natural products.

As part of our continuing interest in MIA chemistry,⁹⁻¹² we developed a streamlined molecular networking¹³ dereplication pipeline based on the implementation of an in-house MS/MS database, constituted of a cumulative collection of MIAs.¹⁴ In order to enrich this database, seven prominent practitioners from the global natural products research community shared their historical collections, leading to the construction of the largest MS/MS dataset of MIAs to date, that we named: Monoterpene Indole Alkaloids DataBase (MIADB) (Fig. 1). The MIADB contains MS/MS data of 172 standard compounds, comprising 128 monoindoles and 44 bisindoles (these compounds are presented in Supplementary Table 1) and covers more than 70% of the known (30/42) MIA skeletons. The information that can be drawn from this dataset is valuable for the scientific community that envisages the isolation of new MIAs.

The purpose of this Data Descriptor is to announce the deposition of the MIADB on the Global Natural Product Social Molecular Networking (GNPS¹⁵) and MetaboLights.¹⁶ Each spectrum of the MIADB has its own accession number from CCMSLIB00004679916 to CCMSLIB00004680087 on GNPS (accessed via: <https://gnps.ucsd.edu/ProteoSAFe/static/gnps-splash.jsp>). The spectral collection is also available for download from MetaboLights under the identifier: MTBLS142.¹⁷

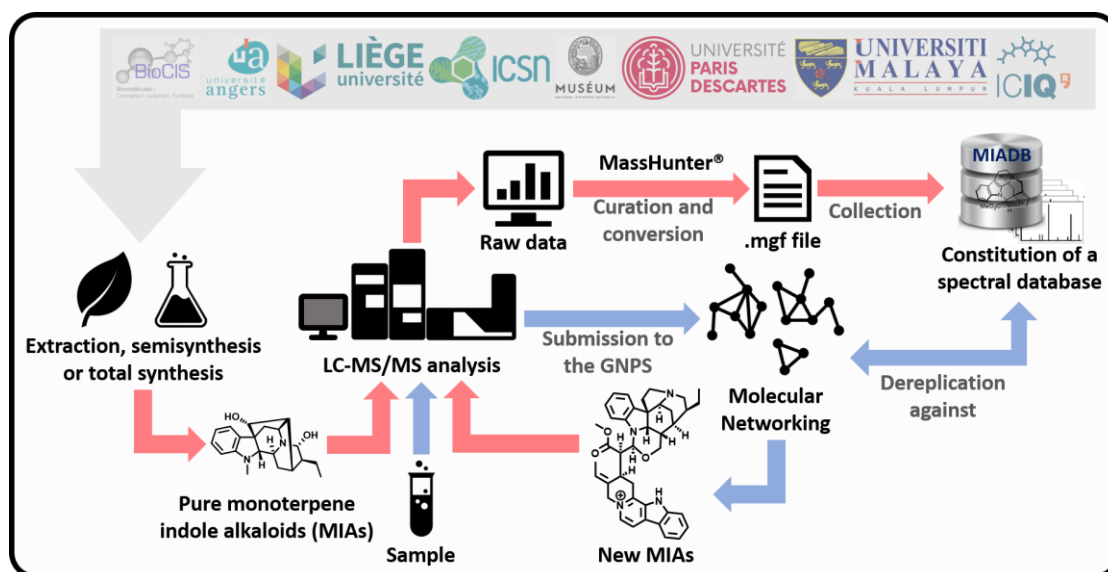


Figure 1. Construction of the MIADB (red arrows) and application in a molecular networking-based dereplication workflow (blue arrows)

Methods

Sample preparation

Each of the collected MIA was diluted to a concentration of 1 mg/mL using HPLC-grade (High Performance Liquid Chromatography) with MeOH (Methanol) as solvent. The solution was then transferred in 1.5 mL HPLC vials and analyzed by LC-MS/MS (Liquid Chromatography-tandem Mass Spectrometry). Chemicals and solvents were purchased from Sigma-Aldrich.

Data acquisition

Samples were analyzed using an Agilent LC-MS (Liquid Chromatography Mass Spectrometry) system composed of an Agilent 1260 Infinity HPLC coupled to an Agilent 6530 ESI-Q-TOF-MS (ElectroSpray Ionization Quadrupole Time of Flight Mass Spectrometry) operating in positive mode. A Sunfire[®] analytical C₁₈ column (150 × 2.1 mm; i.d. 3.5 μm, Waters) was used, with a flow rate of 250 μL/min and a linear gradient from 5% B (A: H₂O + 0.1% formic acid, B: MeOH) to 100% B over 30 min. The column temperature was maintained at 25 °C. ESI conditions were set with the capillary temperature at 320 °C, source voltage at 3.5 kV, and a sheath gas flow rate of 10 L/min. Injection volume was set at 5 μL. The mass spectrometer was operated in Extended Dynamic Range mode (2 GHz). The divert valve was set to waste for the first 3 min. There were four scan events: positive MS, window from *m/z* 100–1200, then three data-dependent MS/MS scans of the first, second, and third most intense ions from the first scan event.

MS/MS settings were: three fixed collision energies (30, 50, and 70 eV), default charge of 1, minimum intensity of 5000 counts, and isolation width of *m/z* 1.3. Purine C₅H₄N₄ [M + H]⁺ ion (*m/z* 121.050873) and hexakis(1*H*,1*H*,3*H*-tetrafluoropropoxy)-phosphazene C₁₈H₁₈F₂₄N₃O₆P₃ [M + H]⁺ ion (*m/z* 922.009798) were used as internal lock masses. Full scans were acquired at a resolution of 11 000 (at *m/z* 922) and 4000 at (*m/z* 121). A permanent MS/MS exclusion list criterion was set to prevent oversampling of the internal calibrant.

Database constitution

The analysis of each of these substances resulted in 172 files with the standard .d format (Agilent standard data-format). A list of individual compounds for each sample was generated from an Auto MS/MS data mining process implemented in MassHunter[®] software on every single file. Averaged as well as monocollisional energy MS/MS spectra were generated from the three retained collision energies (30, 50, and 70 eV). Within this list, the molecular formula (as well as the exact mass) of the expected compound (in its charged state) was identified. Then, depuration of the other features was carried out. Finally, each spectrum was converted into the .mgf (Mascot Generic Format) using the export tool of the MassHunter[®] software.

Code availability

The LC-MS feature detection software (MassHunter[®]) used in this work is commercially available from Agilent[®].

Data Records

All data described in this article have been uploaded to GNPS and MetaboLights. Each spectrum of the 172 compounds of the MIADB has its own accession number from CCMSLIB00004679916 to CCMSLIB00004680087 on the Global Natural Product Social Molecular Networking (GNPS) (accessed via: <https://gnps.ucsd.edu/ProteoSAFe/static/gnps-splash.jsp>). The spectral collection in its two versions (*i.e.* averaged and separate collision energy MS/MS spectra at 30, 50, and 70 eV) is available for download from MetaboLights under the identifier: MTBLS142.¹⁷

Metadata

The MS/MS spectra of the MIADB library are recorded with a variety of details including: LC-MS/MS acquisition parameters, instrument details, organism, organism part, smiles and Inchi codes, CAS numbers, CHEBI IDs, retention times, and chemical formula. These metadata are available on the GNPS and MetaboLights websites.

Technical Validation

Spectroscopic validation of MIADB compounds

The structural identity of the alkaloids being implemented in the MIADB reference metabolite index was established through extensive spectroscopic analyses, including, NMR (Nuclear Magnetic Resonance) and HRMS (High-Resolution Mass Spectrometry). The analyses were carried out by the various collaborators having contributed to the establishment of the database. The obtained mass spectra were individually inspected to verify the occurrence of either the protonated molecular or molecular ion as the precursor mass.

Selected strategies for the validation of the MIADB

The validation of the MIADB was achieved following two strategies: (i) dereplication of the profiled compounds from a methanol extract of the leaves of *Catharanthus roseus* (L.) G. Don. (Apocynaceae) (see supplementary Tables 2 and 3), and (ii) the dereplication of the MIADB against the MIAs previously available on the GNPS library before the upload of the MIADB.

Molecular networking-based dereplication of *Catharanthus roseus* methanol extract

Molecular networking-based dereplication using MIADB-uploaded GNPS libraries was attempted on the methanol extract of *Catharanthus roseus*, the MIAs content of which was thoroughly studied. Accordingly, more than 130 different compounds were reported from the different tissues of the plant.¹⁸ In the displayed network, the experimental data of *C. roseus* methanol extract are depicted as green rectangles and nodes representing a consensus of experimental data and database records (*i.e.*, MIADB-uploaded in the GNPS libraries) are displayed as red rectangles (Fig. 2). As expected, molecular networking of the *C. roseus* leaves methanol extract allowed dereplication of previously known metabolites within this plant including: tabersonine, catharanthine, vindolinine, perivine, geissoschizine, pericyclivine, serpentine, raubasine, and akuammigine (Table 1). All the dereplicated compounds were assigned a level of confidence 1 according to Schymanski et al.¹⁹ based on HMRS, MS/MS and retention time matching, except for geissoschizine, serpentine; and alloyohimbine. The latter were attributed a level of confidence of 2, due to a delta of retention time (RT) superior to 1.5 min. The molecular networking-based dereplication provided a comprehensive coverage of *C. roseus* alkaloids by regards to the available standards, despite the noticeable lack of a vinblastine hit. This missing observation is likely due to the vinblastine concentration that is known to be very low in the plant (ranging from 0.0003% to 0.001% w/w dry weight).²⁰ Conversely, some unexpected matches could also be evidenced throughout the obtained dereplication: burnamine and vobasine. Although none of these were previously described in *C. roseus*, both these structural assignments can be deemed reasonable based on biosynthetic considerations. Being an akuammiline-derived MIA, such as akuammine²¹ and the monomer precursors of the bisindoles vingramine and methylvingramine²² that have been reported to occur in *C. roseus*, the detection of burnamine is not unexpected. Likewise, the co-dereplication in the depicted molecular network of the formerly described vobasane-type perivine supports the identification of vobasine within this plant. Such examples emphasize the dereplicative interest of MIADB especially on such a deeply dug plant model. Prior to its GNPS upload, *i.e.*, as an in-house database, the ability of the MIADB to pinpoint tentatively new MIAs was demonstrated through the streamlined isolation of geissolaevine along with its *O*-methylether derivative and 3',4',5',6'-tetrahydrogeissospermine from the formerly vastly studied *Geissospermum leave* (Vell.) Miers (Apocynaceae).¹⁴ Altogether, the currently garnered results support the valuable contribution of MIADB either for the straightforward identification of monoterpene indole alkaloids or to highlight putative structural novelty among this privileged structural class. The topology of the obtained network also reveals that a further extent of information could yet be accessed from *C. roseus* extracts. Indeed, most dereplicated MIAs are tightly associated within cluster A. Since clusterization depends on structural similarity, a single match to the MIADB-implemented GNPS allows for the propagation of the structure throughout an entire molecular family, indicating that most if not all the nodes of this cluster refer to MIAs. The seminal contribution of the MIADB to the tandem mass spectrometric databanks of MIA is expected to pave the way for the upload of such data by the numerous teams involved in MIA research all over the world, thereby contributing to making this tool more and more efficient to reach a quick and sharp insight into the MIAs content of any producing organism.

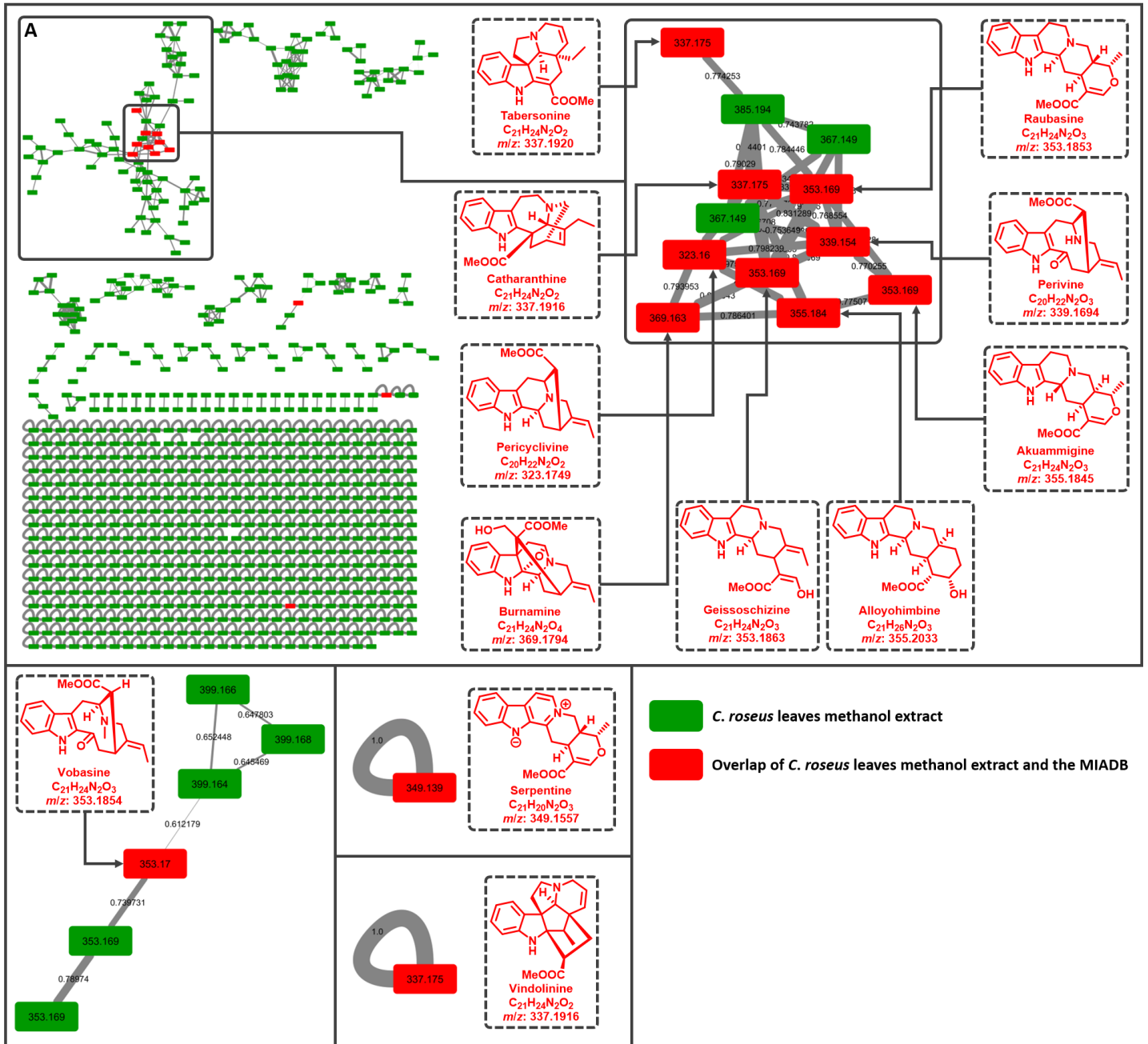


Figure 2. Full molecular network of the profiled compounds from a methanol extract of *C. roseus* leaves annotated by the MIADB. The cosine similarity score cutoff for the molecular network was set at 0.6, the parent ion mass tolerance at 0.02, the fragment ion mass tolerance at 0.02, the score library threshold at 0.6 and the minimum matched peaks at 6. The cosine similarity score are depicted on the edges.

Table 1. Matches between the profiled compounds from a methanol extract of *C. roseus* and MIADB

Compound	Match score	Comment	Δ RT (min)	Confidence level
akuammigine	0.69	Described in <i>C. roseus</i>	0.63	1
alloyohimbine	0.76	Not described in <i>C. roseus</i>	1.95	2
burnamine	0.62	Not described in <i>C. roseus</i>	1.07	1
catharanthine	0.67	Described in <i>C. roseus</i>	0.67	1
geissoschizine	0.71	Described in <i>C. roseus</i>	1.52	2
pericyclivine	0.79	Described in <i>C. roseus</i>	0.00	1
perivine	0.73	Described in <i>C. roseus</i>	0.01	1
raubasine	0.64	Described in <i>C. roseus</i>	0.26	1
serpentine	0.66	Described in <i>C. roseus</i>	6.65	2
tabersonine	0.80	Described in <i>C. roseus</i>	1.48	1
vindolinine	0.77	Described in <i>C. roseus</i>	1.44	1
vobasine	0.65	Not described in <i>C. roseus</i>	0.16	1

Dereplication of the MIADB against the MIAs previously available on the GNPS library

As a second validation assay, the MIADB was dereplicated against the GNPS library. For this purpose, the 172 .mgf files were submitted to the GNPS online platform and all the hits between the MIADB and the GNPS were annotated. 19 of the total MIAs were identified as hits by the GNPS platform (Table 2).

These results indicate that the compounds from the 19 matches were correctly identified within the GNPS library, except in the case of epimers or isomers. Indeed, it should be noted that the matching process does not take into account the stereochemistry of the compounds (Table 2).

Table 2. MIADB matches with the GNPS library

Compounds (GNPS)	Match score	Compounds (MIADB)	Comments
brucine	0.78	brucine	
reserpiline	0.86	reserpiline	
tabernaemontanine	0.74	tabernaemontanine	
voachalotine	0.93	voachalotine	
ajmaline	0.75	ajmaline	
vincamine	0.78	vincamine	
methyl reserpate	0.83	methyl reserpate	
camptothecin	0.73	camptothecin	
reserpine	0.86	reserpine	
strychnine	0.80	strychnine	
akuammigine	0.86	raubasine	epimer
raubasine	0.88	akuammigine	epimer
corynanthine	0.90	yohimbine	epimer
yohimbine	0.92	corynanthine	epimer
vincosamide	0.93	strictosamide	epimer
strictosamide	0.93	vincosamide	epimer
yohimbine	0.90	pseudoyohimbine	epimer
elegantissine	0.73	carapanaubine	isomer
yohimbine	0.89	alloyohimbine	epimer

Acknowledgements

We gratefully acknowledge the productive collaborations and the fruitful contributions that allowed the generation of this natural products MS/MS database. This research was funded by FONDECYT-CONCYTEC (grant contract number 239-2015-FONDECYT), and by the French ANR grant (ANR-15-CE29-0001). We express our thanks to Séverine Amand (MNHN) for her assistance in the collection of the MIAs from the MNHN chemical library.

Author contributions

M.A.B. conceived the project. A.E.F.R., C.A., E.O.N., G.C., and H.H. performed the acquisition of the LC-MS/MS data. P.L.P. performed the technical validation of the MIADB. M.S.K., F.M.M., M.E.M., and A.M.E., performed the total synthesis of grandilodines B and C and lundurine. M.G. and L.M. isolated and collected MNHN compounds. M.K., R.G., T.G., C.L. and S.M. isolated and collected compounds from Université Paris-Descartes. K.A. advised on the extraction aspects of the work conducted by H.H. D.B., A.-M.L., and P.R. collected compounds from Université d'Angers. M.F. isolated and collected compounds from Université de Liège. F.R. and M.L. collected and supervised the isolation of compounds from ICSN. G.L. performed isolation and structure elucidation of lanciferine and vincamajine. L.E. and E.P. conceived and performed the synthetic experimental works related to the obtention of bipleiophylline, voacalgine A, and leucoridine A. A.E.F.R., P.L.P., P.C. and M.A.B. supervised isolation work at Université Paris-Sud and wrote the manuscript. All the authors read and commented the manuscript.

Competing interests

The authors declare that they have no conflict of interest.

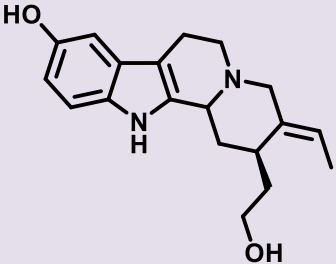
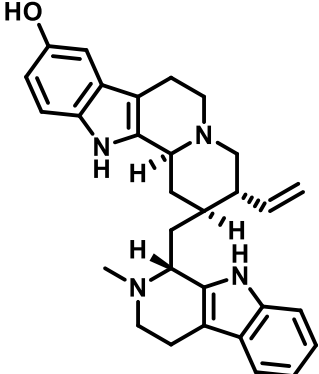
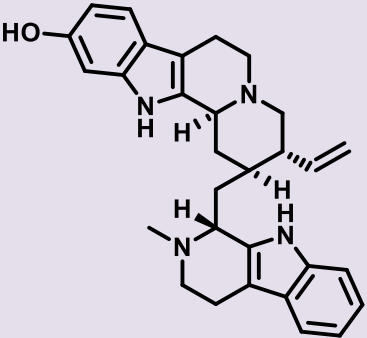
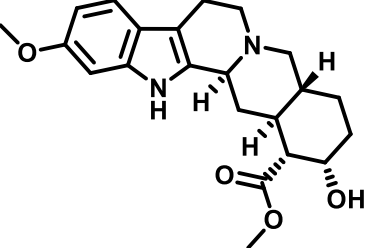
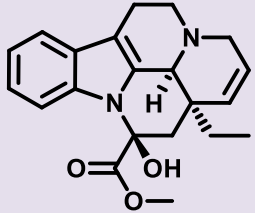
References

- 1 Pan, Q., Mustafa, N. R., Tang, K., Choi, Y. H. & Verpoorte, R. Monoterpenoid indole alkaloids biosynthesis and its regulation in *Catharanthus roseus*: A literature review from genes to metabolites. *Phytochemistry Rev.* **15**, 221-250 (2016).
- 2 Pritchett, B. P. & Stoltz, B. M. Enantioselective palladium-catalyzed allylic alkylation reactions in the synthesis of *Aspidosperma* and structurally related monoterpene indole alkaloids. *Nat. Prod. Rep.* **35**, 559-574 (2018).
- 3 Qu, Y. *et al.* Completion of the seven-step pathway from tabersonine to the anticancer drug precursor vindoline and its assembly in yeast. *Proc. Natl. Acad. Sci. USA* **112**, 6224-6229 (2015).
- 4 Caputi, L. *et al.* Missing enzymes in the biosynthesis of the anticancer drug vinblastine in Madagascar periwinkle. *Science* **360**, 1235-1239 (2018).
- 5 Dang, T.-T. T. *et al.* Sarpagan bridge enzyme has substrate-controlled cyclization and aromatization modes. *Nat. Chem. Biol.* **14**, 760-763 (2018).
- 6 Miettinen, K. *et al.* The seco-iridoid pathway from *Catharanthus roseus*. *Nat. Commun.* **5**, 3606 (2014).
- 7 Salim, V., Yu, F., Altarejos, J. & Luca, V. Virus-induced gene silencing identifies *Catharanthus roseus* 7-deoxyloganic acid-7-hydroxylase, a step in iridoid and monoterpene indole alkaloid biosynthesis. *Plant J.* **76**, 754-765 (2013).
- 8 Asada, K. *et al.* A 7-deoxyloganic acid glucosyltransferase contributes a key step in secologanin biosynthesis in Madagascar periwinkle. *Plant Cell* **25**, 4123-4134 (2013).
- 9 Lachkar, D. *et al.* Unified biomimetic assembly of voacalgine A and bipleiophylline via divergent oxidative couplings. *Nat. Chem.* **9**, 793 (2017).
- 10 Ootogo N'Nang Obiang, E. *et al.* Pleiokomenines A and B: Dimeric aspidofractinine alkaloids tethered with a methylene group. *Org. Lett.* **19**, 6180-6183 (2017).

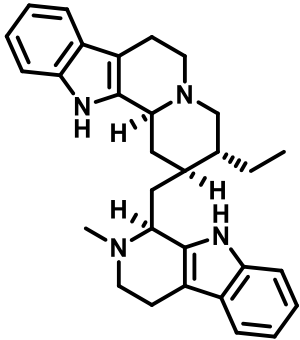
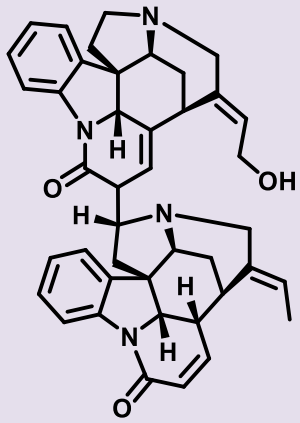
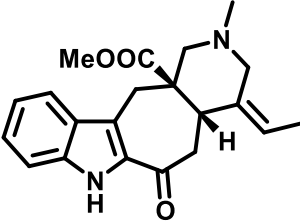
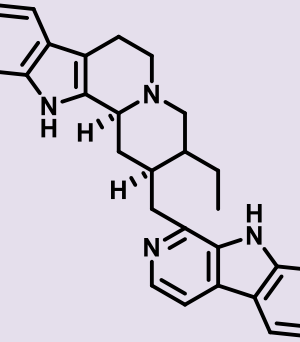
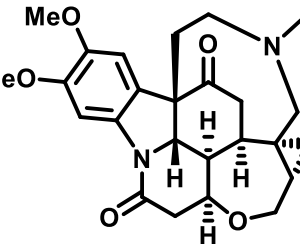
- 11 Beniddir, M. A., Genta-Jouve, G. & Lewin, G. Resolving the (19*R*) absolute configuration of lanciferine, a monoterpene indole alkaloid from *Alstonia boulindaensis*. *J. Nat. Prod.* **81**, 1075-1078 (2018).
- 12 Ootago N'Nang, E. *et al.* Theionbrunonines A and B: Dimeric vobasine alkaloids tethered by a thioether bridge from *Mostuea brunonis*. *Org. Lett.* **20**, 6596-6600 (2018).
- 13 Yang, J. Y. *et al.* Molecular networking as a dereplication strategy. *J. Nat. Prod.* **76**, 1686-1699 (2013).
- 14 Fox Ramos, A. E. *et al.* Revisiting previously investigated plants: A molecular networking-based study of *Geissospermum laeve*. *J. Nat. Prod.* **80**, 1007-1014 (2017).
- 15 Wang, M. *et al.* Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **34**, 828-837 (2016).
- 16 Haug, K. *et al.* MetaboLights—an open-access general-purpose repository for metabolomics studies and associated meta-data. *Nucl. Acids Res.* **41**, D781-D786 (2013).
- 17 Fox Ramos, A. E. *et al.* Collected mass spectrometry data on monoterpene indole alkaloids from natural product chemistry research. *MetaboLights* <https://www.ebi.ac.uk/metabolights/MTBLS142> (2018).
- 18 Heijden, R. V. D., Jacobs, D., I., Snoeijer, W., Hallard, D. & Verpoorte, R. The *Catharanthus* alkaloids: Pharmacognosy and biotechnology. *Curr. Med. Chem.* **11**, 607-628 (2004).
- 19 Schymanski, E. L. *et al.* Identifying small molecules via high resolution mass spectrometry: Communicating confidence. *Environment. Sci. Technol.* **48**, 2097-2098 (2014).
- 20 Uniyal, G. C., Bala, S., Mathur, A. K. & Kulkarni, R. N. Symmetry C18 column: a better choice for the analysis of indole alkaloids of *Catharanthus roseus*. *Phytochem. Anal.* **12**, 206-210 (2001).
- 21 Ramirez, A. & Garcia-Rubio, S. Current progress in the chemistry and pharmacology of akuammiline alkaloids. *Curr. Med. Chem.* **10**, 1891-1915 (2003).
- 22 Jossang, A., Fodor, P. & Bodo, B. A new structural class of bisindole alkaloids from the seeds of *Catharanthus roseus*: Vingramine and methylvingramine. *J. Org. Chem.* **63**, 7162-7167 (1998).

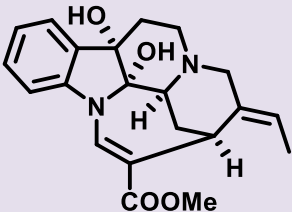
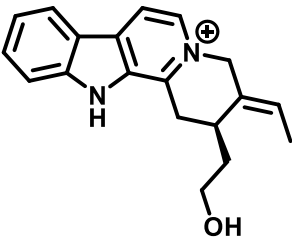
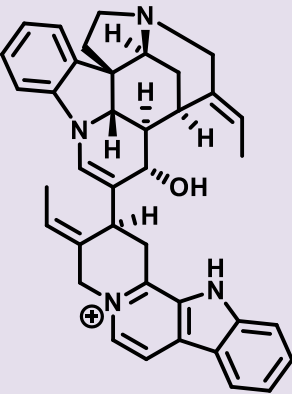
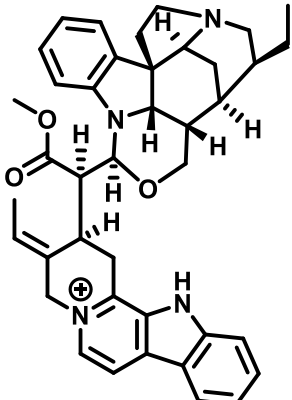
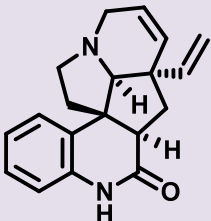
Supplementary information:

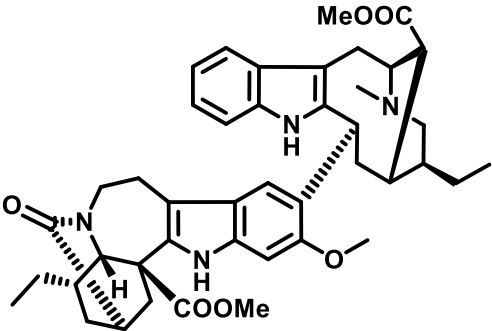
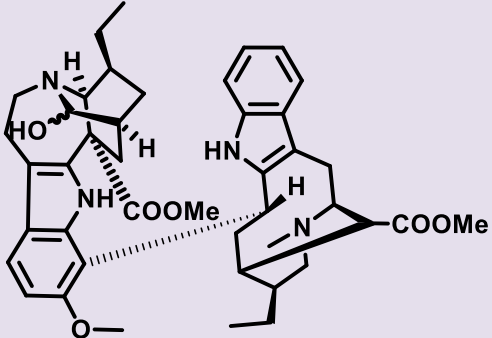
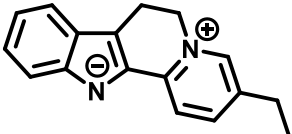
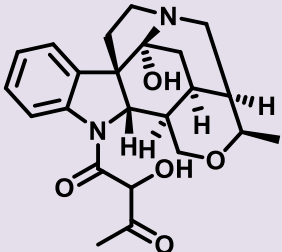
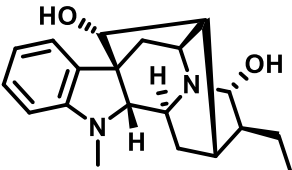
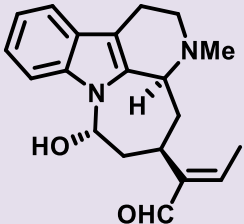
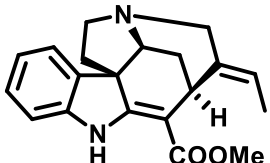
Table S1. Monoterpene indole alkaloids included in the MIADB

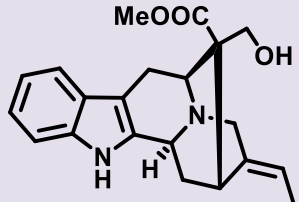
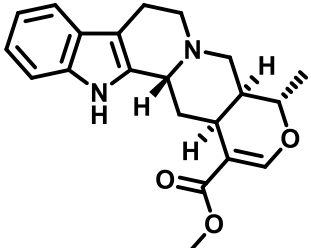
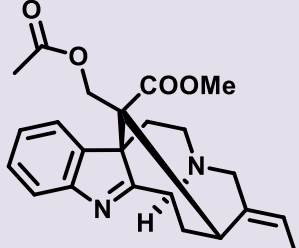
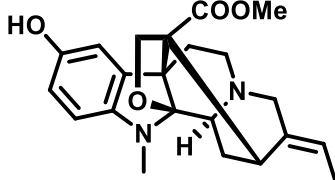
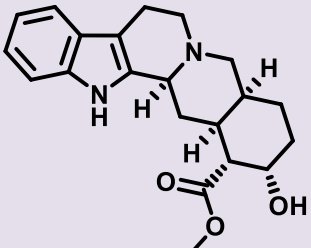
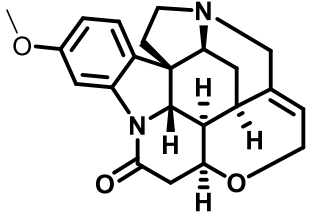
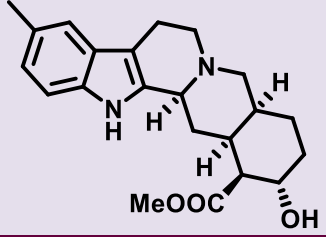
Substance	Measured m/z	Source	Retention time (min)	Structure
10-Hydroxygeissoschizol	313.1903	Natural	11.65	
10-Hydroxyusambarine	467.2822	Natural	13.60	
11-Hydroxyusambarine	467.2794	Natural	13.33	
11-Methoxyyohimbine	385.2125	Natural	14.08	
14,15-Dehydrovincamine	353.1859	Natural	12.98	

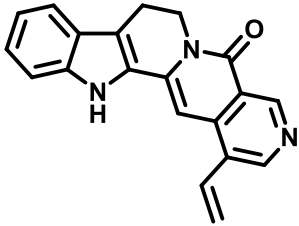
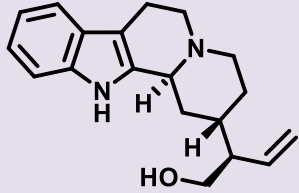
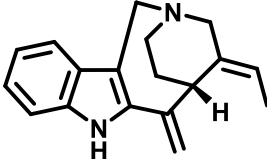
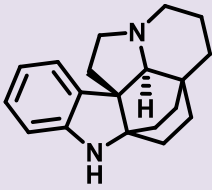
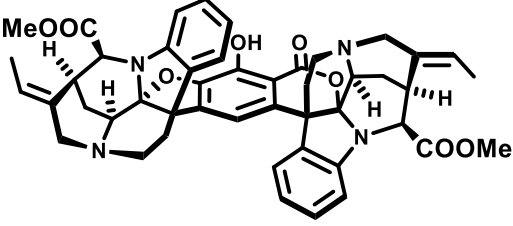
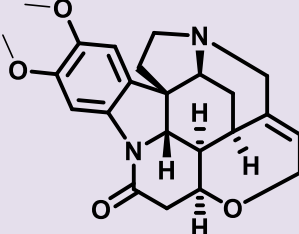
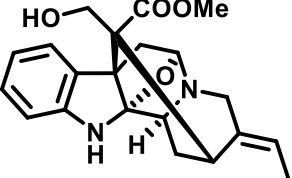


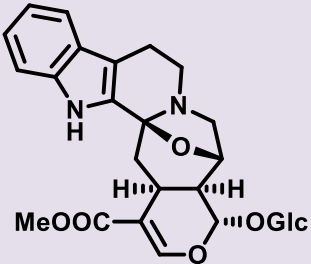
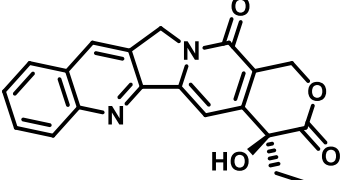
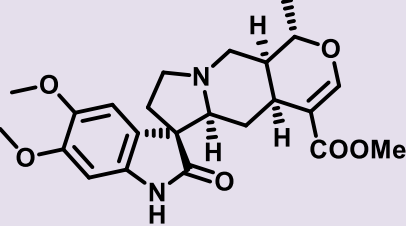
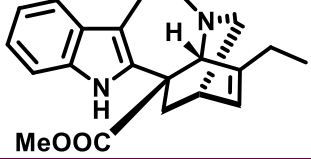
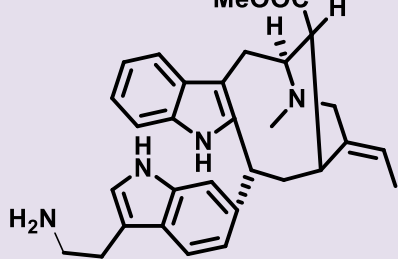
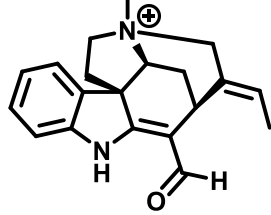
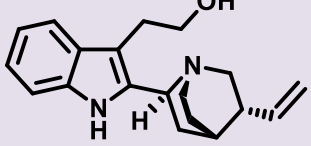
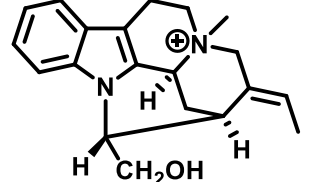
18,19-Dihydrousambarine	453.3021	Natural	14.74	
18-Hydroxyisosungucine	651.3325	Natural	15.35	
19,20-Didehydroervatamine	353.1860	Natural	14.26	
19,20-Dihydrousambarensine	435.2554	Natural	14.53	
19,20-Epoxyovacine	441.2012	Natural	10.05	

2,7-Dihydroxyapogeissoschizine	369.1818	Natural	16.49	 <p>The structure shows a complex polycyclic alkaloid with two indole-like rings. It features two hydroxyl groups (HO) and a methyl ester group (COOMe) attached to the rings. Stereochemistry is indicated with wedges and dashes.</p>
3,4,5,6-Tetrahydrogeissoschizol	293.1645	Natural	16.74	 <p>The structure shows a complex polycyclic alkaloid with two indole-like rings. It features a hydroxyl group (OH) and a vinyl group attached to the rings. The nitrogen atom in one of the rings is positively charged (N⁺).</p>
3',4',5',6'-Tetrahydroolongicaudatine Y	567.3127	Natural	15.20	 <p>The structure shows a complex polycyclic alkaloid with two indole-like rings. It features a hydroxyl group (OH) and a vinyl group attached to the rings. The nitrogen atom in one of the rings is positively charged (N⁺).</p>
3',4',5',6'-Tetrahydrogeissospermine	629.3499	Natural	15.81	 <p>The structure shows a complex polycyclic alkaloid with two indole-like rings. It features a methoxy group (OCH₃) and a vinyl group attached to the rings. The nitrogen atom in one of the rings is positively charged (N⁺).</p>
3-Epimeloscine	293.1655	Natural	11.70	 <p>The structure shows a complex polycyclic alkaloid with two indole-like rings. It features a vinyl group and a carbonyl group (C=O) attached to the rings. The nitrogen atom in one of the rings is positively charged (N⁺).</p>

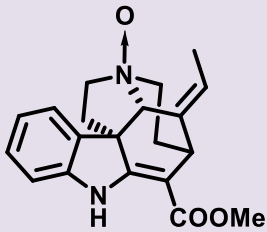
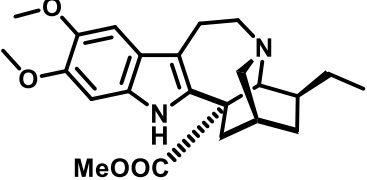
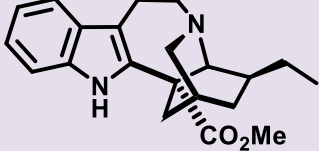
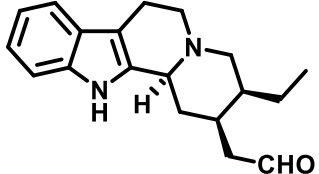
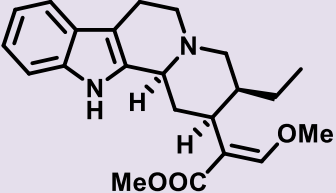
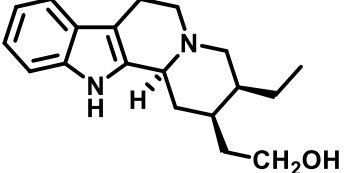
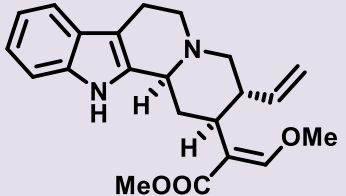
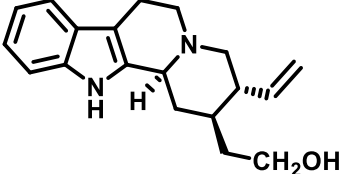
3'-Oxotabernaegantine B	721.3944	Natural	22.61	
3-R,S-Hydroxytabernaegantine A	705.4009 [M+H-H ₂ O] ⁺	Natural	15.07	
6,7-Dihydroflavopereirine	249.1378	Natural	16.28	
Acetyl-splendoline	413.2077	Natural	14.36	
Ajmaline	327.2096	Natural	12.55	
Akagerine	325.1930	Natural	15.89	
Akuammicine	323.1770	Natural	16.07	

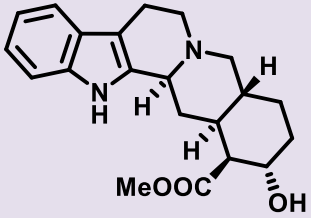
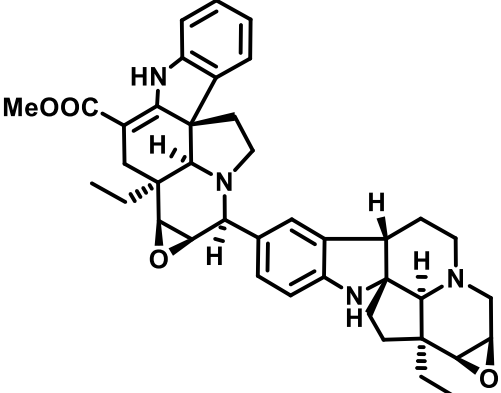
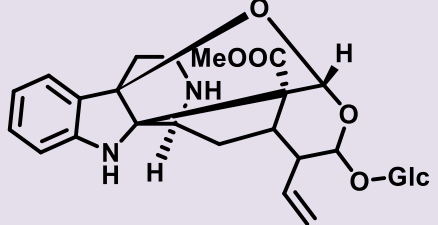
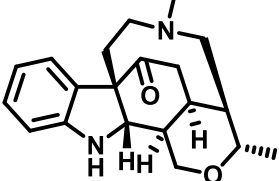
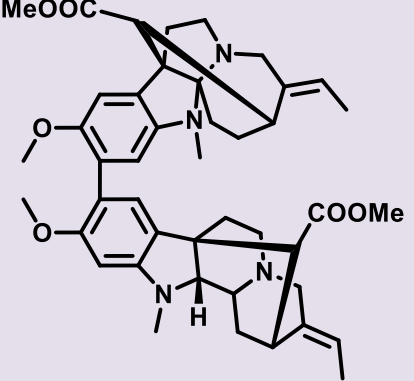
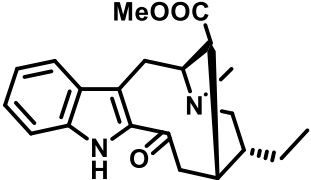
Akuamidine	353.1864	Natural	11.11	
Akuammigine	353.1843	Natural	18.14	
Akuamiline	395.1969	Natural	13.52	
Akuamine	383.1962	Natural	12.57	
Alloyohimbine	355.2034	Natural	14.30	
Alpha-colubrine	365.1861	Natural	13.80	
Alpha-methylohimbine	369.2181	Natural	16.76	

Angustine	314.1281	Natural	28.84	
Antirhine	297.1958	Natural	18.29	
Apparicine	265.1687	Natural	18.50	
Aspidofractinine	281.2012	Natural	9.26	
Bipleiophylline	795.3385	Semisynthesis	19.49	
Brucine	395.1967	Natural	12.95	
Burnamine	369.1791	Natural	14.62	

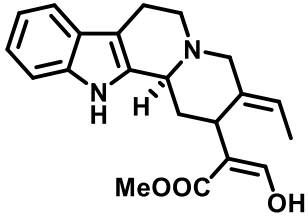
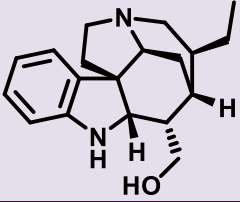
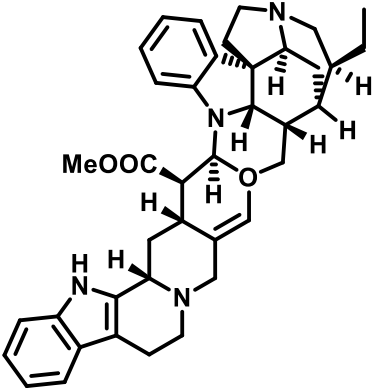

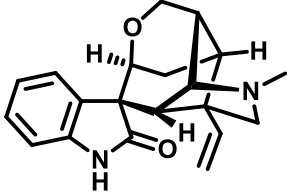
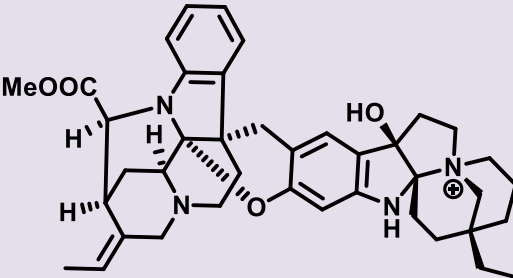
Cadambine	545.2161	Natural	16.74	
Camptothecin	349.1174	Commercial (Alfa Aesar, 2016)	27.21	
Carpanaubine	429.2067	Natural	15.75	
Catharanthine	337.1914	Natural	16.11	
Ceridimine	497.2898	Natural	12.62	
C-fluorocurarine	307.1808	Natural	11.29	
Cinchonamine	297.1951	Natural	16.68	
C-mavacurine	309.1965	Natural	14.82	



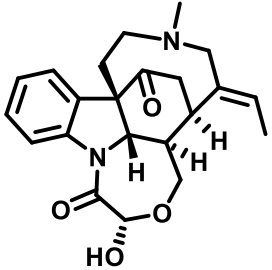

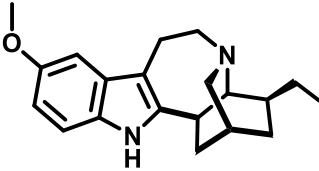
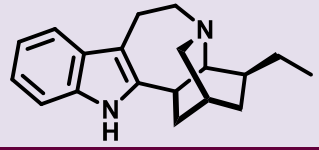
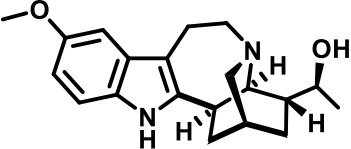

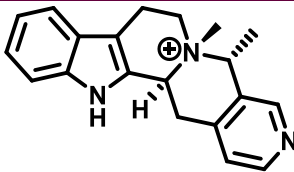
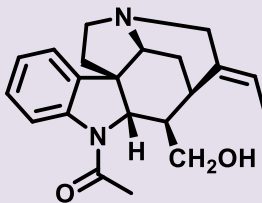
Condyllocarpine <i>N</i> -oxide	339.1719	Natural	15.35	
Conopharyngine	399.2292	Natural	16.30	
Coronaridine	339.2056	Natural	16.85	
Corynantheidal	297.1950	Natural	14.43	
Corynantheidine	369.2183	Natural	16.78	
Corynantheidol	299.2120	Natural	14.05	
Corynantheine	367.2006	Natural	16.70	
Corynantheol	297.1963	Natural	13.66	

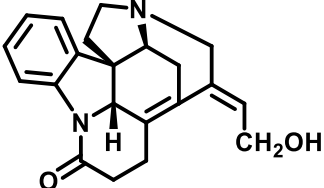
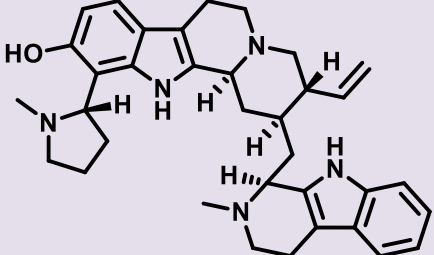
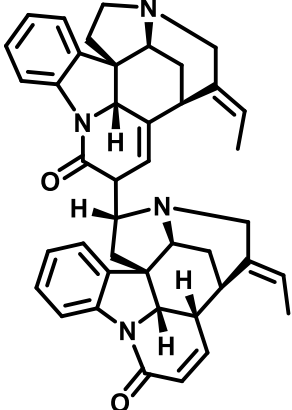
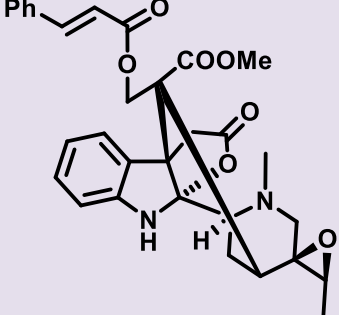
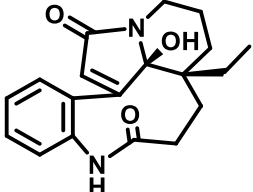
Corynanthine	355.2011	Natural	14.25	
Criophylline	647.3569	Natural	15.99	
Cymoside	547.2286	Natural	10.35	
Desacetyl-isosplendine	327.2073	Natural	15.95	
Desoxycabufiline	735.4127	Natural	16.90	
Dregamine	355.2012	Natural	11.90	

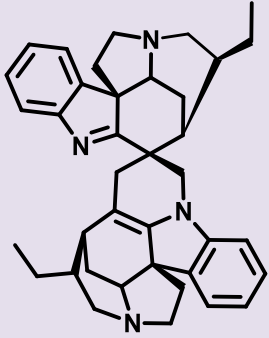
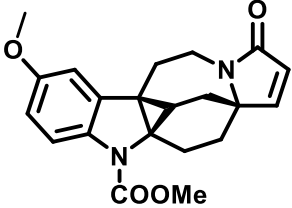
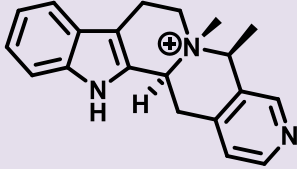
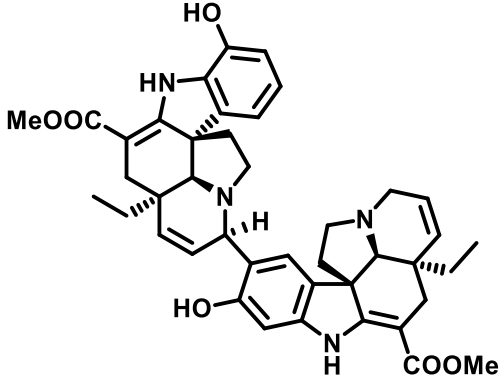
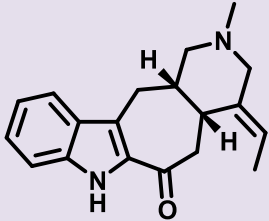
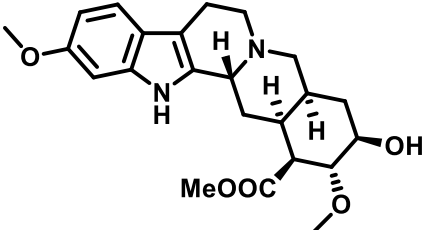
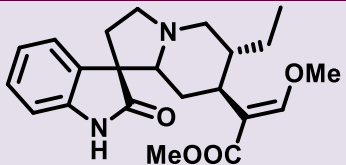
Echitamidine	341.1868	Natural	10.85	
Echitamine	385.2123	Natural	13.40	
Ervafoline	645.3433	Natural	25.80	
Ervatamine	355.2023	Natural	14.99	
Ervitsine	293.1648	Natural	14.24	
Geissolaevine	367.1298	Natural	23.22	
Geissolosimine	573.3609	Natural	13.83	

Geissoschizine	353.1869	Natural	18.31	
Geissoschizoline	299.2114	Natural	10.36	
Geissospermine	633.3801	Natural	14.42	
Gelsemicine	359.1964	Natural	16.28	
Gelsemine	323.1769	Natural	9.75	
Goniomedine A	649.3741	Natural	12.49	

Goniomedine B	633.3790	Natural	14.66	
Goniomedinone	663.3532	Natural	13.83	
Goniomedine A N-oxide	665.3705	Natural	13.40	
Goniomitine	299.2104	Natural	17.49	
Grandilodine B	455.1821	Total synthesis	25.06	
Grandilodine C	381.1448	Total synthesis	24.09	

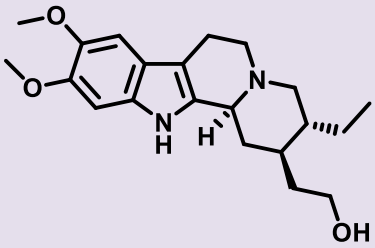
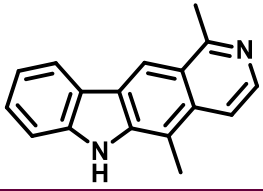
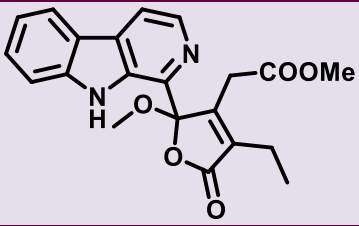
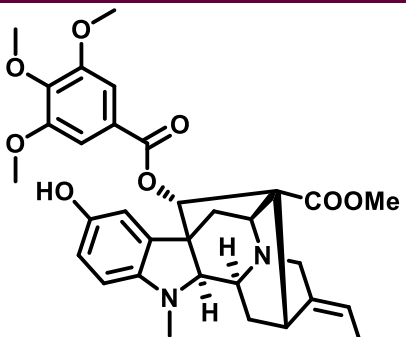
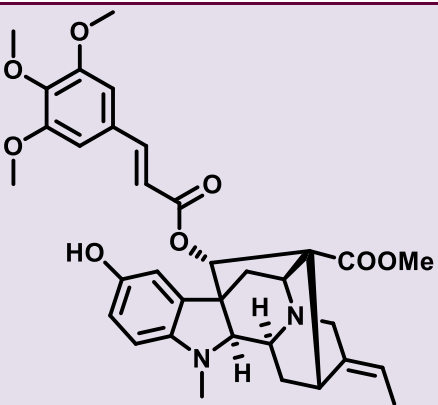
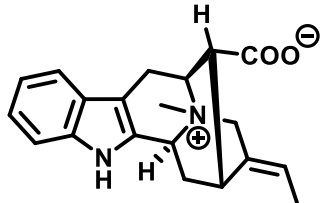
Holstiine	383.1976	Natural	12.06	
Holstiline	397.2141	Natural	16.00	
Ibogaine	311.2118	Natural	13.57	
Ibogamine	281.1998	Natural	15.34	
Iboxygaine	327.2077	Natural	14.90	
Icajine	365.1866	Natural	13.64	
Isomalindine	304.1802	Natural	10.88	
Isoretuline	339.2057	Natural	12.55	

Isostrychnine	335.1741	Natural	12.14	
Isostrychnopentamine A	550.3544	Natural	13.16	
Isosungucine	635.3363	Natural	16.03	
Lanciferine	545.2268	Natural	29.88	
Leuconolam	327.1702	Natural	18.12	

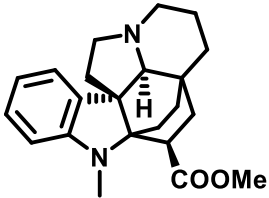
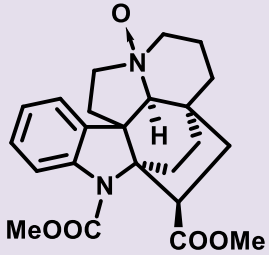
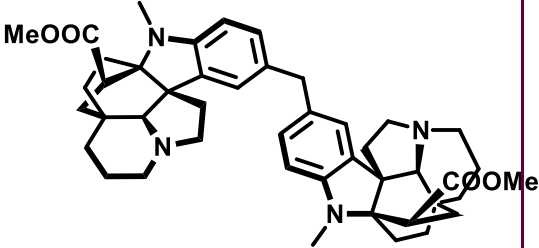
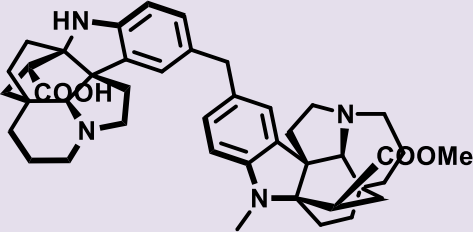
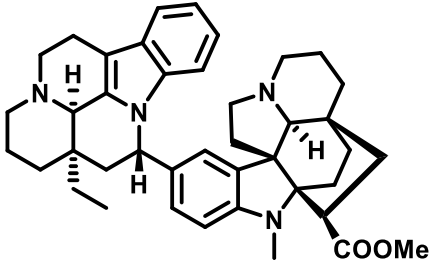
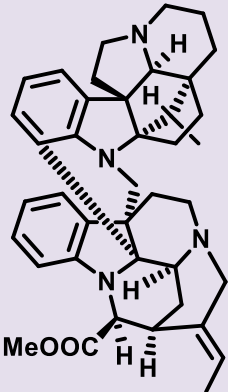
Leucoridine A	557.3640	Semisynthesis	16.27	
Lundurine A	367.1652	Total synthesis	28.08	
Malindine	304.1819	Natural	10.20	
Melosuavine E	703.3472	Natural	16.70	
Methuenine	295.1804	Natural	14.35	
Methyle reserpate	415.2234	Natural	16.12	
Mitrinermine	385.2141	Natural	16.43	



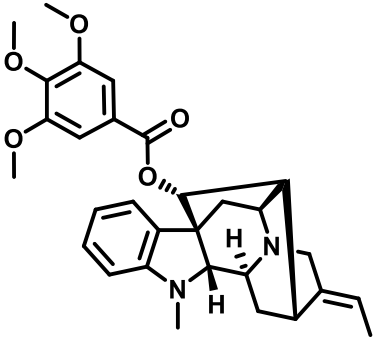
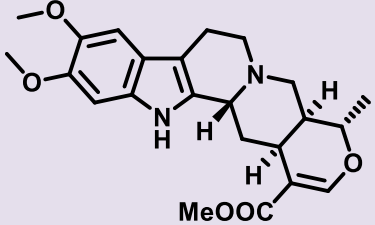
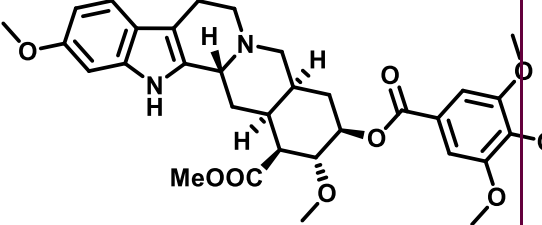
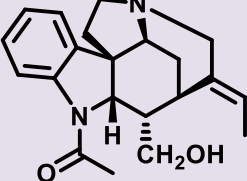
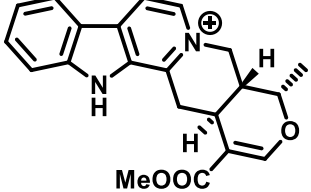
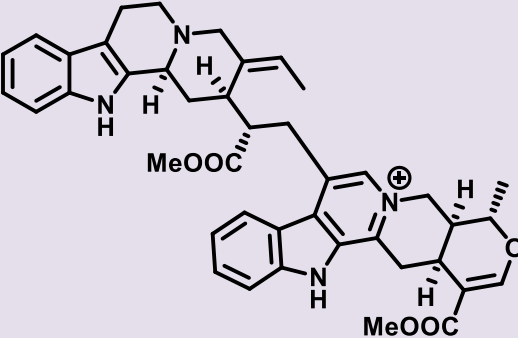
N⁴-methylantirhine	311.2123	Natural	14.23	
Naucleidinal	337.1543	Natural	27.33	
Naulafine	312.1124	Natural	22.14	
N⁶-methylusambarensine	447.2546	Natural	15.48	
Novacine	425.2065	Natural	13.05	
Ochrolifuanine A	439.2850	Natural	16.62	
Ochropamine	367.2021	Natural	16.30	

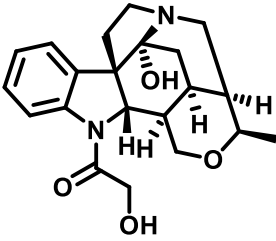
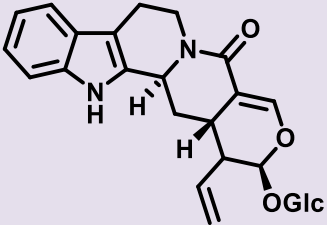
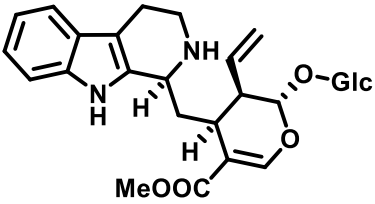
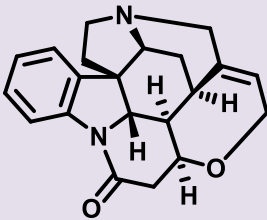
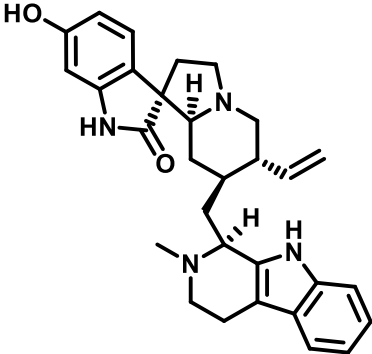
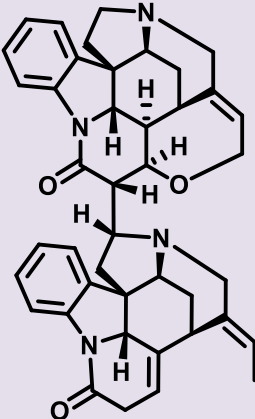
Ochroposinine	359.2310	Natural	16.60	
Olivacine	247.1228	Natural	21.48	
O-methylgeissolaevine	381.1446	Natural	26.97	
O-trimethoxy-3,4,5-benzoyl-OH-vincamajine	577.255	Natural	15.76	
O-3,4,5-trimethoxycinnamate-OH-vincamajine	603.2701	Natural	21.00	
Panarine	323.1754	Natural	15.64	

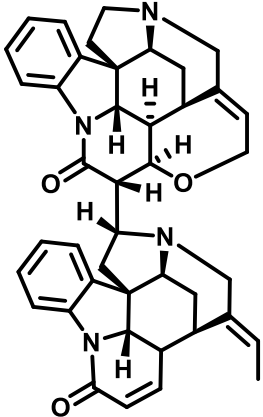
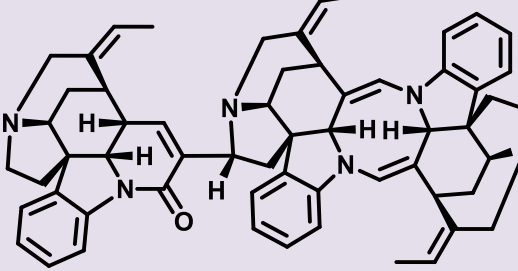
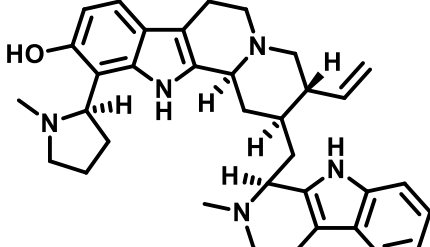
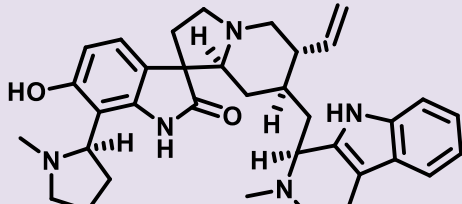
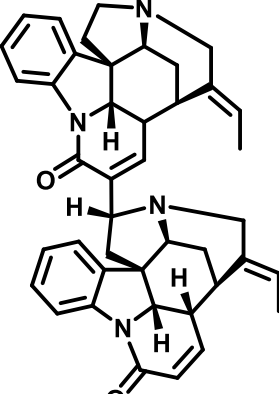
Pandine	353.1843	Natural	14.82	
Pericyclivine	323.1749	Natural	16.17	
Perivine	339.1694	Natural	13.40	
Picaline	411.1914	Natural	17.32	
Pleiocarpamine	323.1742	Natural	17.46	
Pleiocarpine	397.2122	Natural	26.94	
Pleiocarpinilam	367.2032	Natural	15.06	

Pleioarpinine	353.2217	Natural	17.23	
Pleioarpoline	413.2088	Natural	17.16	
Pleiokomenine A	717.4375	Natural	19.52	
Pleiokomenine B	689.4051	Natural	17.96	
Pleiomutine	631.4026	Natural	18.73	
Pleiomutinine	615.3667	Natural	15.01	

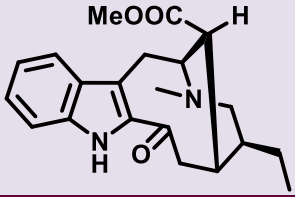
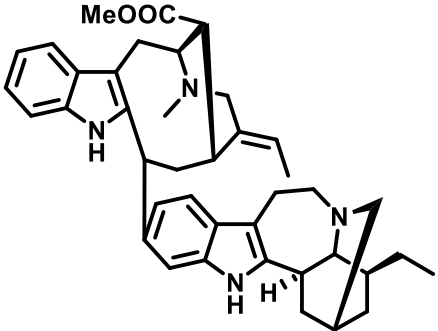
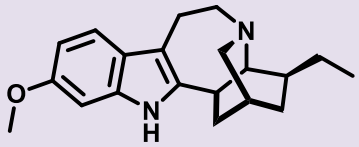
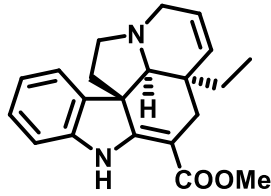
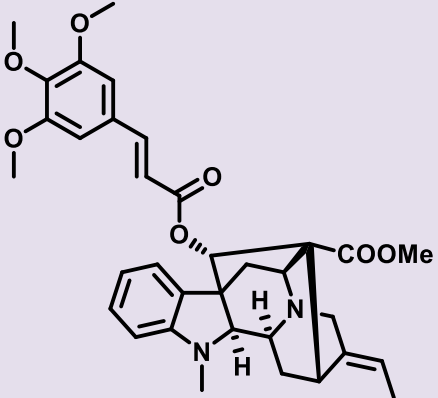
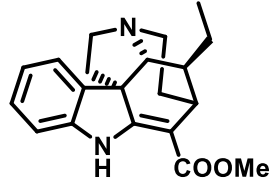
Polyneuridine	353.1874	Natural	12.95	
Protostrychnine	353.1865	Natural	10.05	
Pseudostrychnine	351.1686	Natural	13.27	
Pseudoyohimbine	355.2030	Natural	13.63	
Quebrachamine	283.2185	Natural	15.05	
Quebrachidine	353.1861	Natural	14.91	
Quinidine	325.1907	Natural	13.69	
Raubasine	353.1849	Natural	17.54	

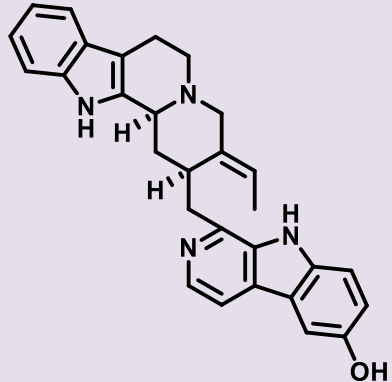
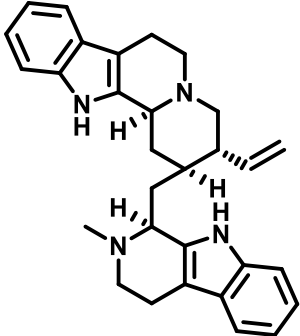
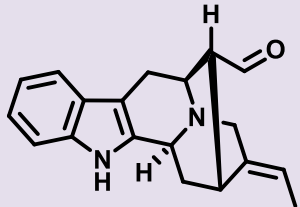
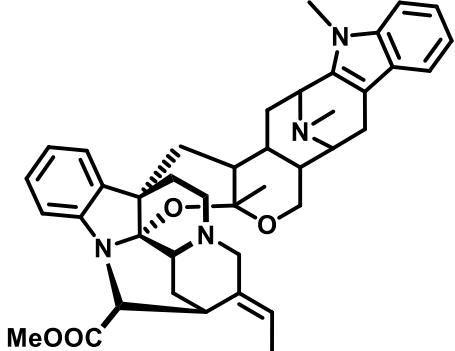
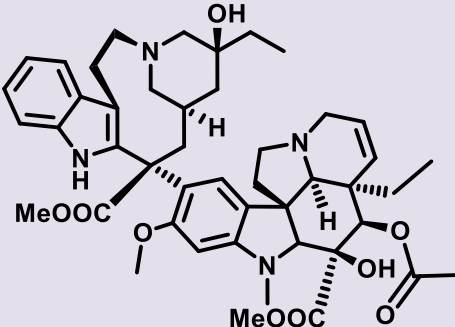
Rauvomitine	503.2531	Natural	22.61	
Reserpiline	413.2077	Natural	17.02	
Reserpine	609.2817	Natural	21.85	
Retuline	339.2063	Natural	9.59	
Serpentine	349.1557	Natural	13.15	
20'-episerpentinine	685.3381	Natural	18.20	

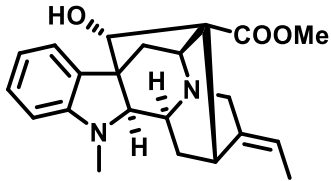
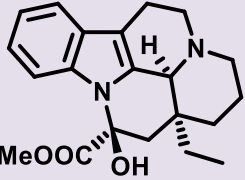
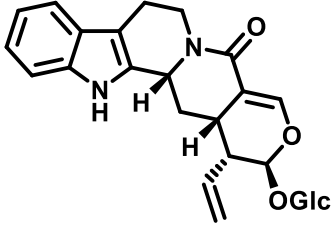
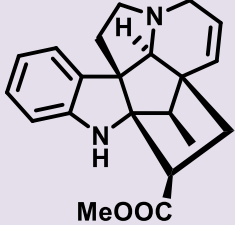
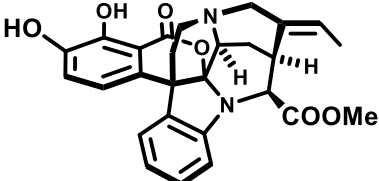
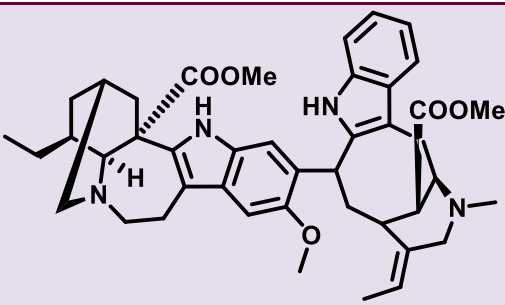
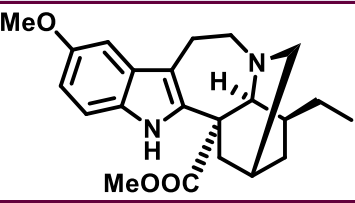
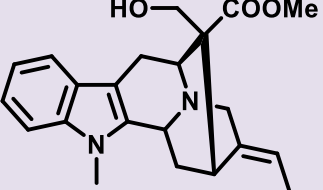
Splendoline	371.1973	Natural	11.85	
Strictosamide	499.2071	Natural	26.66	
Strictosidine	531.2330	Natural	16.96	
Strychnine	335.1760	Natural	11.00	
Strychnofoline	483.2757	Natural	11.94	
Strychnogucine A	651.3347	Natural	15.55	

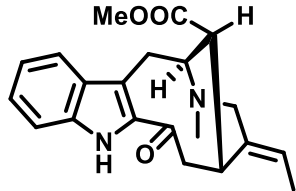
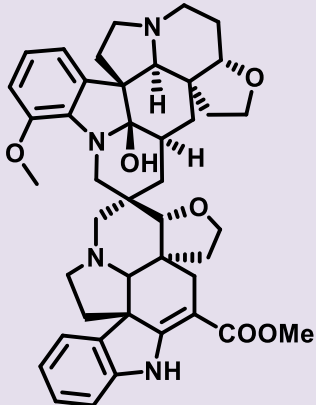
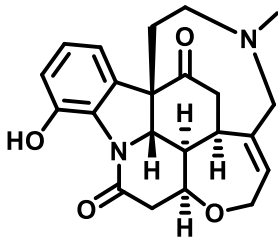
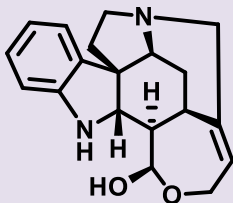
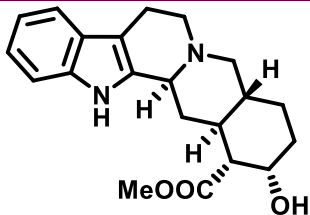
Strychnogucine C	651.3361	Natural	16.33	
Strychnohexamine	869.4934	Natural	17.81	
Strychnopentamine	550.3547	Natural	12.58	
Strychnophylline	566.3494	Natural	8.71	
Sungucine	635.3390	Natural	15.32	



Tabernaemontanine	355.2008	Natural	14.73	
Tabernamine	617.3835	Natural	15.28	
Tabernanthine	311.2109	Natural	18.63	
Tabersonine	337.1928	Natural	18.23	
Trimethoxy-3,4,5-cinnamate-vincamajine	587.2748	Natural	17.77	
Tubotaiwine	325.1897	Natural	13.70	

10-hydroxyusambarensine	449.2337	Natural	16.64	
Usambarine	451.2857	Natural	14.80	
Velloimine	293.1651	Natural	14.71	
Villalstonine	661.3737	Natural	17.38	
Vinblastine	811.4268	Natural	18.62	

Vincamajine	367.2013	Natural	12.01	
Vincamine	355.2019	Natural	19.99	
Vincosamide	499.2101	Natural	26.86	
Vindoline	337.1916	Natural	13.46	
Voacalgine A	475.1853	Natural	19.64	
Voacamine	705.3984	Natural	17.52	
Voacangine	369.2179	Natural	16.76	
Voachalotine	367.2033	Natural	14.25	

Vobasine	353.1854	Natural	15.32	
Vobtusine	719.3819	Natural	23.26	
Vomicine	381.1809	Natural	13.39	
Wieland-Gumlich aldehyde	311.1758	Natural	9.62	
Yohimbine	355.2014	Natural	15.20	

First validation assay

The plant material was a commercial batch of the leaves of *Catharanthus roseus* (L.) G.Don. used for teaching purposes (Univ. Paris-Sud), bearing no voucher number. It was extracted as follows:

For LC-MS/MS analysis, 1 g of dry and powdered leaves of *C. roseus* were extracted with MeOH (20 mL). The MeOH crude extract was filtered under vacuum and concentrated by rotatory evaporation to yield 10 mg of residue. This residue was analyzed in identical conditions to that used for data acquisition of individual compounds.



Table S2. Molecular Networking parameters for the first validation essay

Parameter	Value
Minimum pairs cosine	0.6
Parent mass ion tolerance	0.02
Fragment ion mass tolerance	0.02
Minimum matched peaks	6
Top K	10
Minimum cluster size	2
Maximum connected component size	100
Run MScluster	on
Library search score threshold	0.6
Library search minimum matched peaks	6

A molecular network was created using the online workflow at GNPS. The data was then clustered with MS-Cluster with a parent mass tolerance of 0.02 Da and a MS/MS fragment ion tolerance of 0.02 Da to create consensus spectra. Further, consensus spectra that contained less than 2 spectra were discarded. A network was then created where edges were filtered to have a cosine score above 0.6 and more than 6 matched peaks. Further edges between two nodes were kept in the network if and only if each of the nodes appeared in each other's respective top 10 most similar nodes. The spectra in the network were then searched against GNPS' spectral libraries. All matches kept between network spectra and library spectra were required to have a score above 0.6 and at least 6 matched peaks.

Second validation assay

Table S3. Molecular Networking parameters for the second validation essay

Parameter	Value
Minimum pairs cosine	0.6
Parent mass ion tolerance	0.02
Fragment ion mass tolerance	0.02
Minimum matched peaks	6
Top K	10
Minimum cluster size	1
Maximum connected component size	100
Run MScluster	off
Library search score threshold	0.7
Library search minimum matched peaks	6



A molecular network was created using the online workflow at GNPS. The data was filtered by removing all MS/MS peaks within ± 17 Da of the precursor m/z . MS/MS spectra were window filtered by choosing only the top 6 peaks in the ± 50 Da window throughout the spectrum. A network was then created where edges were filtered to have a cosine score above 0.6 and more than 6 matched peaks. Further edges between two nodes were kept in the network if and only if each of the nodes appeared in each other's respective top 10 most similar nodes. The spectra in the network were then searched against GNPS' spectral libraries. The library spectra were filtered in the same manner as the input data. All matches kept between network spectra and library spectra were required to have a score above 0.7 and at least 6 matched peaks.





CHAPITRE III :

PREMIÈRE APPLICATION DE LA MIADB - ÉTUDE PHYTOCHIMIQUE DE *GEISSOSPERMUM LAEVE* MIERS.





CHAPITRE III : PREMIÈRE APPLICATION DE LA MIADB – EXPLORATION DE L’ESPACE CHIMIQUE DE *GEISSOSPERMUM LAEVE* (VELL.) MIERS.

1. INTRODUCTION

1.1 Phylogénie et description botanique

Le genre *Geissospermum* Allemão appartient à la famille des Apocynaceae⁵³, sous-famille des *Rauvolfioideae* Kostel., tribu des *Aspidospermatæ* Miers.⁵³ Ce genre regroupe six espèces : *G. argenteum* Woodson, *G. fuscum* Markgr., *G. laeve* (Vell.) Miers, *G. reticulatum* A.H.Gentry, *G. sericeum* Miers et *G. urceolatum* A.H.Gentry. On retrouve des synonymies avec les genres *Aspidosperma* et *Tabernaemontana* pour certaines.

De manière générale, les plantes du genre *Geissospermum* se présentent toutes sous forme de grands arbres, à écorce très épaisse. Elles se caractérisent par la présence des feuilles alternes et d’inflorescences extra-axillaires. Les corolles, tubulaires, ont une forme de plateau avec des segments à légère convolution dextre, caractéristiques des Apocynaceae, et des anthères cordées. Leur formule florale est également caractéristique des Apocynaceae (5 sépales, 5 pétales, 5 étamines, 2 carpelles, fusionnés ici). Les fruits sont des baies en forme de poire (**Figure 7**).^{54, 55, 56, 57}

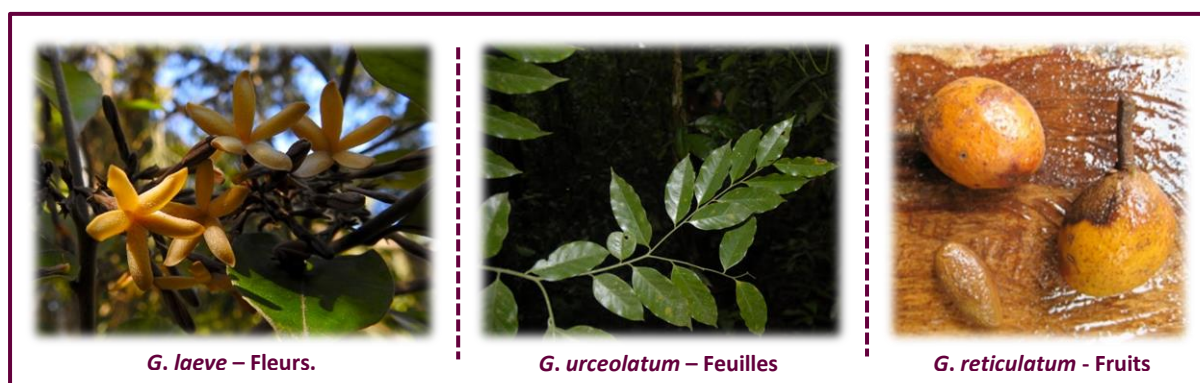


Figure 7. Espèces du genre *Geissospermum*

⁵³ M. E. Endress, S. Liede-Schumann, et al., *Phytotaxa*, 2014, **159**, 175-194.

⁵⁴ J. Miers, Taylor and Francis, 1878. pp. 1-5.

⁵⁵ <https://www.amigosjb.org.br/galeria/agostosembro-2015/#jp-carousel-7431>, (accessed 08-09-2018, 2018).

⁵⁶ http://static1.kew.org/science/tropamerica/imagetdatabase/large73/cat_single73-2.htm, (accessed 08-09-2018, 2018).

⁵⁷

http://andesamazonfieldschool.com/Andes_and_Amazon_Field_School/Geissospermum_reticulatum_Apocynaceae.html, (accessed 08-09-2018, 2018).



1.2 Répartition géographique

Les espèces du genre *Geissospermum* sont originaires de l'Amérique du Sud, et plus exactement des pays suivants : Brésil, Pérou, Bolivie, Guyana, Surinam et Guyane française, où il y a un climat subtropical (Figure



8),^{58, 59}

Elles poussent dans la *Terra firme*

de la forêt amazonienne (ou zone qui n'est pas inondée par la crue en période de hautes eaux), dans la forêt semi-décidue (forêt caractérisée par la présence d'une majorité d'arbres dont les feuilles caduques, tombent au rythme des saisons) et dans la forêt tropicale.⁶⁰

Figure 8. Répartition géographique des espèces du genre *Geissospermum*

1.3 *Geissospermum laeve* (Vell.) Miers



Figure 9. *G. laeve* dans l'herbier du Royal Botanical Gardens

Geissospermum laeve (Vell.) Miers. (Figure 9,⁶¹ connu aussi sous les noms de *G. vellosii* Allemão, *G. martianum* Miers et *Tabernaemontana laevis* Vell.) est l'espèce la plus étudiée du genre. Elle est utilisée en médecine traditionnelle comme antipaludique, sous la forme de décocté des écorces.⁶² L'emploi sous forme de macérations alcooliques est également notable.⁶³ *G. laeve* est utilisé aussi pour traiter la constipation, les troubles hépatiques et comme stimulant sexuel.⁶⁴

En raison de ces propriétés médicinales et de l'importance de ses emplois traditionnels, les écorces ont fait l'objet d'une demande d'inscription à la pharmacopée française, par l'association pour l'étude et le développement des plantes aromatiques et médicinales en Guyane (GADEPAM).

⁵⁸ J. C. P. Steele, N. C. Veitch, et al., *J. Nat. Prod.*, 2002, **65**, 85-88.

⁵⁹ <https://doi.org/10.15468/39omei>, (accessed 18-09-2018).

⁶⁰ I. R. Koch, A.; Simões, A.O.; Kinoshita, L.S.; Spina, A.P.; Castello, A.C.D., <http://reflora.jbrj.gov.br/jabot/floradobrasil/FB4605>, (accessed 18-09-2018).

⁶¹ <http://specimens.kew.org/herbarium/K000965595>, (accessed 18-09-2018).

⁶² S. Bertani, G. Bourdy, et al., *J. Ethnopharmacol.*, 2005, **98**, 45-54.

⁶³ P. Grenand, C. Moretti, et al., ORSTOM, 1987. p. 126.

⁶⁴ F. Mbeunkui, M. H. Grace, et al., *J. Chromatogr. B*, 2012, **885**, 83-89.

Après examen de la bibliographie par un comité *ad hoc* de l'ANSM (Agence Nationale de Sécurité des Médicaments) en 2014, la « présence d'alcaloïdes toxiques et les risques de confusions avec d'autres plantes »⁶⁵, ont mené à son inscription sur la liste B des plantes médicinales de la Pharmacopée française (plantes utilisées traditionnellement dont le rapport bénéfice / risques est défavorable). On notera que la plante est également célèbre pour avoir été proposée, de manière très controversée, comme anti-VIH par le Pr. Beljansky.⁶³

Plusieurs études phytochimiques ont été entreprises sur *G. laeve* en utilisant des procédés traditionnels d'identification et d'isolement des produits naturels. Ces travaux ont abouti à la description de 13 alcaloïdes indolomonoterpéniques (**Tableau 2** et **Figure 10**).^{66, 67, 68, 69, 70, 71, 72, 73, 74}

Tableau 2. Alcaloïdes indolomonoterpéniques décrits dans *G. laeve*

Nom de la molécule	Décrite par	Type de squelette	Source
5,6-dihydroflavopereirine (15)	V. Muñoz et al. ⁷¹	Camptothecine	Écorces
apogeissoschizine (16)	H. Rapoport et al. ⁶⁸	Strychnidine	Écorces
flavopereirine (17)	N.A. Hughes et al. ⁶⁶	Camptothecine	Écorces
geissolosimine (13)	H. Rapoport et al. ⁶⁹	Corynanthe- strychnos	Écorces
geissoschizine (18)	H. Rapoport et al. ⁶⁷	Corynanthean	Écorces
geissoschizone (19)	F. Mbeunkui et al. ⁷⁴	Strychnidine	Écorces
geissoschizoline (20)	H. Rapoport et al. ⁶⁷	Akuammicine	Écorces
geissospermine (14)	H. Rapoport et al. ⁶⁷	Corynanthe- strychnos	Écorces
geissovelline (21)	R.E. Moore et al. ⁷⁰	Condylocarpan	Écorces
pausperadine A (22)	H. Ishiyama et al. ⁷²	Ulean	Écorces
vellosimine (23)	H. Rapoport et al. ⁶⁹	Sarpagine	Écorces
vellosiminol (24)	H. Rapoport et al. ⁶⁹	Sarpagine	Écorces
vellosine (25)	H. Rapoport et al. ⁶⁷	Aspidofractine	Écorces

⁶⁵

https://ansm.sante.fr/var/ansm_site/storage/original/application/c0085a65bb75c11555a6aa91ccc4c750.pdf, (accessed 10-09-2018, 2018).

⁶⁶ N. A. Hughes & H. Rapoport, *J. Am. Chem. Soc.*, 1958, **80**, 1604-1609.

⁶⁷ H. Rapoport, T. P. Onak, et al., *J. Am. Chem. Soc.*, 1958, **80**, 1601-1604.

⁶⁸ H. Rapoport, R. J. Windgassen Jr, et al., *J. Am. Chem. Soc.*, 1960, **82**, 4404-4414.

⁶⁹ H. Rapoport & R. E. Moore, *J. Org. Chem.*, 1962, **27**, 2981-2985.

⁷⁰ R. E. Moore & H. Rapoport, *J. Org. Chem.*, 1973, **38**, 215-230.

⁷¹ V. Muñoz, M. Sauvain, et al., *J. Ethnopharmacol.*, 2000, **69**, 127-137.

⁷² H. Ishiyama, M. Matsumoto, et al., *Heterocycles*, 2005, **66**, 651-658.

⁷³ J. A. Lima, R. S. Costa, et al., *Pharmacol. Biochem. Behav.*, 2009, **92**, 508-513.

⁷⁴ F. Mbeunkui, M. H. Grace, et al., *J. Ethnopharmacol.*, 2012, **139**, 471-477.

Université Paris-Saclay

Espace Technologique / Immeuble Discovery

Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France



Tirant parti de l'évolution des techniques d'identification et d'isolement des produits naturels, notre équipe a décidé d'entreprendre l'exploration de l'espace chimique de cette espèce déjà étudiée en utilisant ces nouvelles techniques, en aspirant à découvrir des molécules à motifs chimiques inédits.



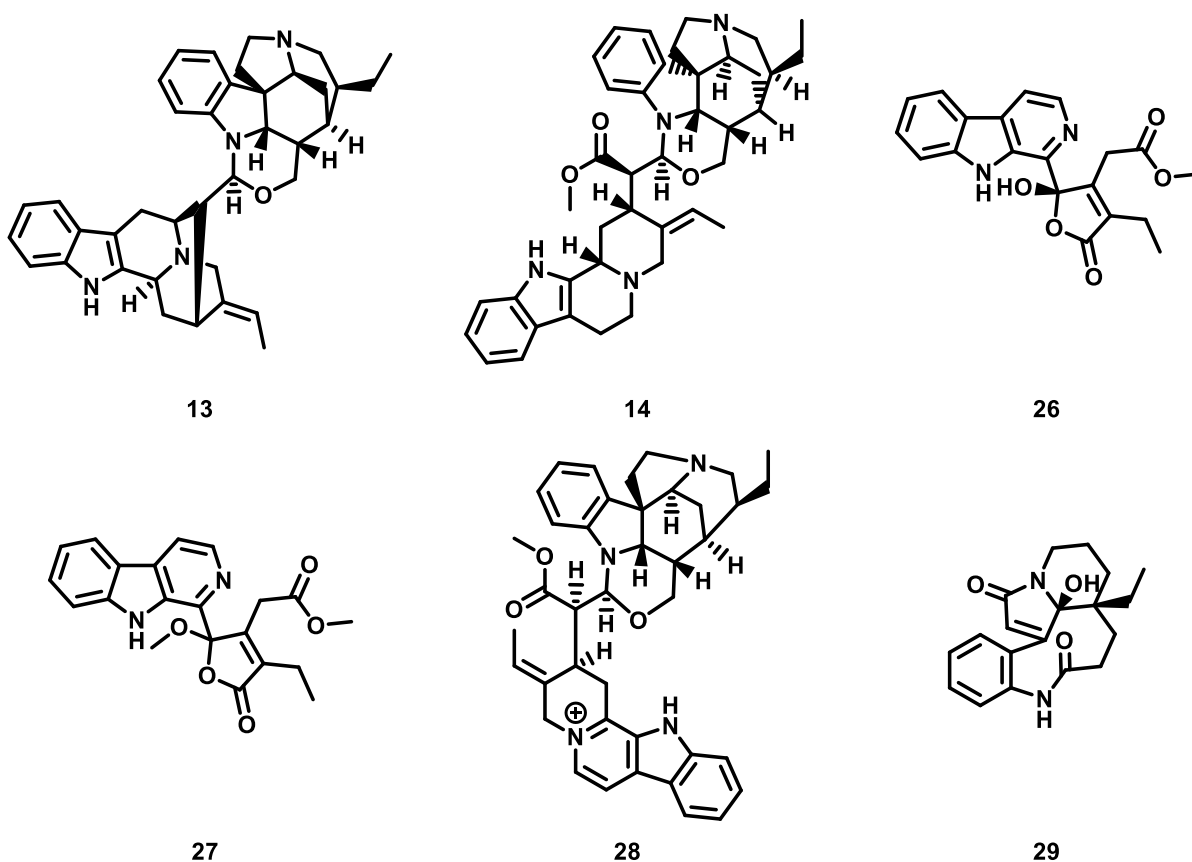
Figure 10. Alcaloïdes indolomonoterpéniques décrits dans *G. laeve*

Dans ce cadre, nous avons tiré profit de la première version de la MIADB, constituée de 55 molécules à ce moment-là, pour servir de référence à cette étude. C'est ainsi qu'un extrait alcaloïdique des écorces de *G. laeve* a été préparé et, par la suite, dérépliqué contre cette base de données en appliquant la technique du « *molecular networking* ». L'importance de ce travail réside dans le fait que c'était le premier pas de notre laboratoire vers une phytochimie utilisant des outils informatiques d'exploitation de données de CL-SM/SM, au service d'une exploration plus efficace de l'espace chimique des plantes déjà étudiées au cours de l'histoire. Ce travail a fait l'objet d'une publication en 2017 dans la revue *Journal of Natural Products*.

Réinvestigation des plantes déjà explorées : une étude de *Geissospermum laeve* basée sur les réseaux moléculaires

Article publié dans *Journal of Natural Products*, 2017

Résumé : Trois nouveaux AIMs (**26–28**) ont été isolés de l'écorce de *Geissospermum laeve*, ainsi que des alcaloïdes connus (–)-leuconolam (**29**), géissolosimine (**13**) et géissospermine (**14**). Les structures de **26–28** ont été élucidées par l'analyse de leurs données spectroscopiques SMHR et RMN. Parmi ces trois composés, **26** et **27** ont un squelette jamais décrit auparavant, constitué d'une partie γ -lactone α,β -insaturé. La configuration absolue de la geissolaevine (**26**) a été déduite de la comparaison des spectres DCE expérimental et calculé. La méthode d'isolement était guidée par une stratégie de déréplication basée sur la technique des réseaux moléculaires en utilisant une base de données interne d'AIMs. En outre, cinq composés connus, non décrits précédemment dans le genre *Geissospermum*, ont été déréplicés du réseau de l'extrait alcaloïdique de *G. laeve* ; différents niveaux d'identification ont pu être attribués pour ces molécules. Les activités antiparasitaires contre *Plasmodium falciparum* et *Leishmania donovani* ainsi que l'activité cytotoxique contre la souche cellulaire MRC-5 ont été déterminées pour les composés **13** et **26–29**.





2. PUBLICATION ASSOCIÉE

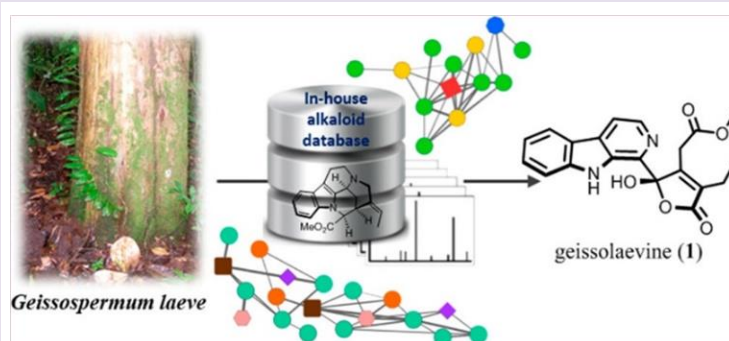
JOURNAL OF NATURAL PRODUCTS

2017

Revisiting Previously Investigated Plants: A Molecular Networking-Based Study of *Geissospermum laeve*

Alexander E. Fox Ramos, Charlotte Alcover, Laurent Evanno, Alexandre Maciuk, Marc Litaudon, Christophe Duplais, Guillaume Bernadat, Jean-François Gallard, Jean-Christophe Jullian, Elisabeth Mouray, Philippe Grellier, Philippe M. Loiseau, Sébastien Pomel, Erwan Poupon, Pierre Champy*, and Mehdi A. Beniddir*

Abstract: Three new monoterpene indole alkaloids (**1–3**) have been isolated from the bark of *Geissospermum laeve*, together with the known alkaloids (–)-leuconolam (**4**), geissolosimine (**5**), and geissospermine (**6**). The structures of **1–3** were



elucidated by analysis of their HRMS and NMR spectroscopic data. The absolute configuration of geissolaevine (**1**) was deduced from the comparison of experimental and theoretically calculated ECD spectra. The isolation workflow was guided by a molecular networking-based dereplication strategy using an in-house database of monoterpene indole alkaloids. In addition, five known compounds previously undescribed in the *Geissospermum* genus were dereplicated from the *G. laeve* alkaloid extract network and were assigned with various levels of identification confidence. The antiparasitic activities against *Plasmodium falciparum* and *Leishmania donovani* as well as the cytotoxic activity against the MRC-5 cell line were determined for compounds **1–5**.





Revisiting Previously Investigated Plants: A Molecular Networking-Based Study of *Geissospermum laeve*

Alexander E. Fox Ramos,[†] Charlotte Alcover,[†] Laurent Evanno,[†] Alexandre Maciuk,[†] Marc Litaudon,^{‡,§} Christophe Duplais,[§] Guillaume Bernadat,[#] Jean-François Gallard,[‡] Jean-Christophe Jullian,[†] Elisabeth Mouray,[⊥] Philippe Grellier,[⊥] Philippe M. Loiseau,^{||} Sébastien Pomel,^{||} Erwan Poupon,[†] Pierre Champy,^{*,†} and Mehdi A. Beniddir^{*,†,§}

[†]Équipe "Pharmacognosie-Chimie des Substances Naturelles" BioCIS, Univ. Paris-Sud, CNRS, Université Paris-Saclay, 5 Rue J.-B. Clément, 92290 Châtenay-Malabry, France

[‡]Institut de Chimie des Substances Naturelles, CNRS, ICSN UPR 2301, Université Paris-Saclay, 21 Avenue de la Terrasse, 91198 Gif-sur-Yvette, France

[§]CNRS, UMR8172 EcoFoG, AgroParisTech, Cirad, INRA, Université des Antilles, Université de Guyane, 23 Avenue Pasteur, 97300 Cayenne, France

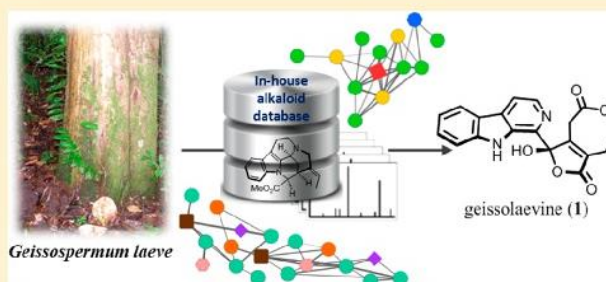
[#]Équipe "Molécules Fluorées et Chimie Médicinale" BioCIS, Univ. Paris-Sud, CNRS, Université Paris-Saclay, 5 rue J.-B. Clément, 92290 Châtenay-Malabry, France

[⊥]Unité Molécules de Communication et Adaptation des Microorganismes (MCAM, UMR 7245), Muséum National d'Histoire Naturelle, CNRS, Sorbonne Universités, CP52, 57 Rue Cuvier, 75005 Paris, France

^{||}Équipe "Chimiothérapie Antiparasitaire" BioCIS, Univ. Paris-Sud, CNRS, Université Paris-Saclay, 5 Rue J.-B. Clément, 92290 Châtenay-Malabry, France

Supporting Information

ABSTRACT: Three new monoterpene indole alkaloids (1–3) have been isolated from the bark of *Geissospermum laeve*, together with the known alkaloids (–)-leuconolam (4), geissolosimine (5), and geissospermine (6). The structures of 1–3 were elucidated by analysis of their HRMS and NMR spectroscopic data. The absolute configuration of geissolaevine (1) was deduced from the comparison of experimental and theoretically calculated ECD spectra. The isolation workflow was guided by a molecular networking-based dereplication strategy using an in-house database of monoterpene indole alkaloids. In addition, five known compounds previously undescribed in the *Geissospermum* genus were dereplicated from the *G. laeve* alkaloid extract network and were assigned with various levels of identification confidence. The antiparasitic activities against *Plasmodium falciparum* and *Leishmania donovani* as well as the cytotoxic activity against the MRC-5 cell line were determined for compounds 1–5.



Throughout the six past decades, the plant family Apocynaceae has drawn the interest of generations of chemists due to the presence of monoterpene indole alkaloids, a therapeutically valuable class of natural products.^{1,2} Remarkably, a thorough literature survey has revealed that most of these compounds have been very poorly evaluated biologically and that many plants containing them remain untapped from both a biological and a chemical perspective. For many years, the resolution of the complex structures of these natural compounds was hindered by the availability of limited spectroscopic instrumentation and relied mainly on the degradation, derivatization, and synthesis of structural subunits for molecular resolution.³ Therefore, understudied Apocynaceae species may represent a potential overlooked source of unprecedented secondary metabolite structures that could

afford new pharmacophoric scaffolds. Hence, as part of a continuing interest in indole alkaloid chemistry, as well as in the identification of antiprotozoal compounds from tropical plants,^{4,5} a reinvestigation of *Geissospermum laeve* (Vell.) Miers, a previously studied Apocynaceae species, was conducted using molecular networking as a dereplication strategy.⁶

Geissospermum (Apocynaceae) is a small genus of Amazonian trees native to northern South America, distributed from Guyana to Brazil.⁷ Several *Geissospermum* species including *G. laeve* (syn. *G. vellosii*) and *G. sericeum* are known to possess antimalarial properties.⁸ They are usually ingested as bark

Received: November 5, 2016

Published: March 10, 2017

Chart 1

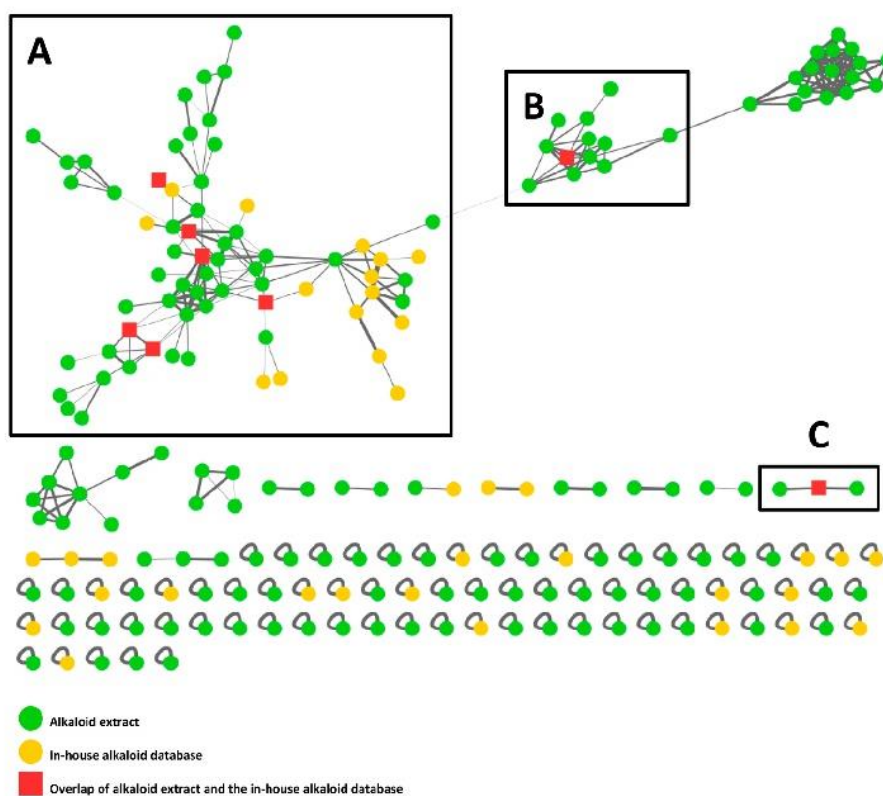
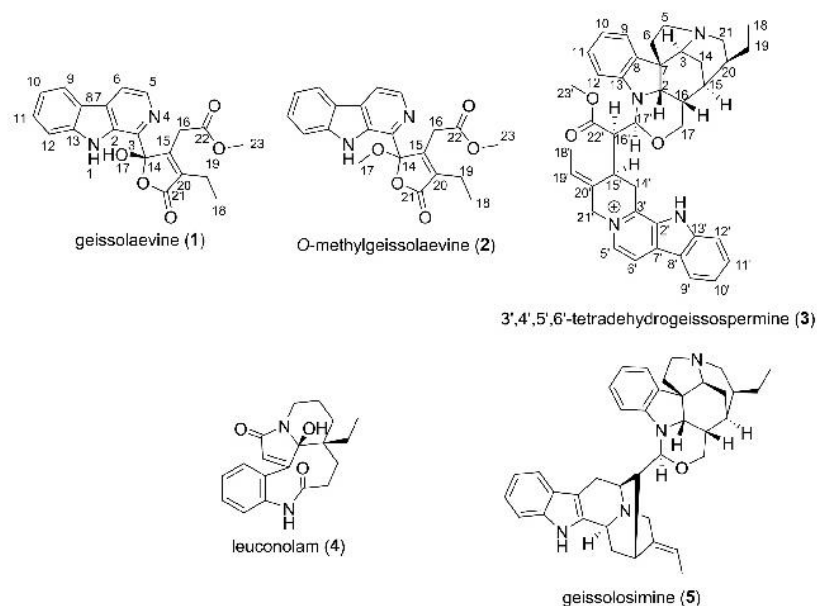


Figure 1. Full molecular network realized using MS/MS data from the alkaloid extract from the bark of *Geissospermum laeve* and the reference compounds from the in-house alkaloid database. The cosine similarity score cutoff for the molecular network was set at 0.6. Details for clusters A–C are presented in Figures S3 and S4, [Supporting Information](#).

decoctions prepared in water (Brazil, Guyana, and Surinam) or as alcoholic macerates (French Guiana).⁹ Within this genus, *G. laeve* is the most well-studied species, with 22 isolated indole alkaloids of the strychnan (akuammicine subtype), corynanthean, aspido-spermatan, and flavopereirine β -carboline subtypes having been obtained.^{10–19}

Recent advances in bioinformatics and analytical chemistry, particularly in mass spectrometry, have enhanced the field of natural product discovery.²⁰ Recently, molecular networking is emerging as a promising computer-based approach to visualize and organize tandem MS/MS data sets and to automate database searches for secondary metabolite identification within

complex mixtures.^{21,22} One major advantage of a molecular networking-based dereplication strategy is the possibility of identifying known compounds and potential analogues, thereby allowing for the prioritization of the isolation workflow and guiding the user toward unknown molecules.

Despite these advantages, it must be pointed out that the success of a molecular networking-based dereplication study to achieve the rapid characterization of known compounds greatly relies on the availability and quality of MS/MS databases.²³ Existing libraries, including MassBank,²⁴ Metlin,²⁵ ReSpec,²⁶ NIST,²⁷ and GNPS,²⁸ host MS/MS data of 22 644 individual compounds. They are far from being comparable with the 281 897 natural products described to date (CRC Press, www.dnp.chemnetbase.com; checked in November 2016). To address this issue, an elegant approach developed by Allard et al. has allowed for a search of experimental MS/MS data in a batch mode against an extensive in silico database (UNPD²⁹-ISDB) of 220 000 natural products.³⁰ Unfortunately, this MS/MS database (UNPD-ISDB) does not contain permanently charged compounds, a feature shared by a large number of monoterpene indole alkaloid derivatives. Therefore, an in-house MS/MS database for these compounds was implemented by taking advantage of a cumulative collection of plant alkaloids. Today, this database contains MS/MS data of 55 standard compounds (Figure S1, Supporting Information) and covers more than 50% of the 42 known monoterpene indole alkaloid skeletons (Figure S2, Supporting Information).³¹ The incorporation of these reference compounds in the *G. laeve* alkaloid extract molecular network provided an efficient and streamlined way to target new analogues as well as unknown compounds.

The present report describes the construction of the in-house database, the dereplication of known alkaloids from the dried stem bark of *G. laeve*, and the MS/MS guided isolation, structure elucidation, and evaluation of antiparasitic and cytotoxic activities of three new alkaloids (1–3), along with the known (–)-leuconolam (4), geissolosimine (5), and geissospermine (6).

RESULTS AND DISCUSSION

Molecular Networking-Based Prioritization of the Isolation Workflow. In order to prioritize the isolation workflow toward previously undescribed monoterpene indole alkaloids, an alkaloid extract of the stem bark of *G. laeve* was analyzed by HPLC-HRMS/MS (positive-ion mode), and then the fragmentation data obtained were organized by molecular networking and subsequently dereplicated against spectra contained in the in-house database. In the molecular network generated, 1598 individual MS/MS spectra were organized into 207 nodes forming 14 clusters of connected nodes, among which three (clusters A–C) appeared of particular interest, due to the presence of matches with the in-house database.

In the network, the experimental data of *G. laeve* alkaloid extract nodes are represented by green circles; the in-house database nodes that do not overlap with experimental data are shown as yellow circles; and nodes that are a consensus of experimental data and database information are shown as red squares (Figure 1).

As expected, molecular networking of the *G. laeve* alkaloid extract allowed dereplication of the previously known compounds, geissolosimine (5), geissospermine (6), and geissoschizoline (18) (Figure S3, Supporting Information). In addition, five known compounds previously undescribed in the

genus *Geissospermum* were also dereplicated, namely, raubasine (10), ibogamine (48), quebrachamine (54) (Figure S3, Supporting Information), serpentine (14) (Figure S4, Supporting Information), and leuconolam (4). Moreover, the dereplication against the GNPS libraries proposed only one hit, namely, yohimbine (44), without matching our own standard, thus supporting the interest of building an in-house database. In the case of geissolosimine (5) and geissospermine (6, corynanthe-strychnos-type), geissoschizoline (18, akuammicine-type), quebrachamine (54, quebrachamine-type), and ibogamine (48, iboga-type), unambiguous identification could be performed (level 1 of confidence according to Schymanski et al.;³² see Figure S5, Supporting Information for MS, retention time, MS/MS, and reference compound comparisons). Despite MS, MS/MS, and retention time matching, raubasine (10, ajmalicine-type), leuconolam (4, rhazinilam-type), and serpentine (14, ajmalicine-type) could be assigned only the identification level 2a³² due to the occurrence of several known epimers reported to date. Interestingly, this approach allowed the identification of unexpected skeletons, which, however, have been described in the closely related *Tabernaemontana* and *Alstonia* genera.

Further analysis of the molecular network revealed the following: (i) In cluster A, the structurally related bisindoles geissospermine (6) and geissolosimine (5) were directly connected to two additional nodes (Figure S3, Supporting Information), suggesting the presence of two new putative geissospermine-like bisindoles at m/z 629.348 [$C_{40}H_{45}N_4O_3^+$] and 587.374 [$C_{39}H_{46}N_4O$]. In addition, clear connections between raubasine (10), geissoschizoline (18), quebrachamine (54), and ibogamine (48) with putative analogues could be observed. (ii) In cluster B, serpentine (14) or one of its epimers was connected to several nodes (Figure S4, Supporting Information), indicating the unexpected presence of analogues in the alkaloid extract. (iii) In cluster C, leuconolam (4) was connected to two additional nodes at m/z 341.186 and 353.229 (Figure S4, Supporting Information).

As a result of this molecular networking-based dereplication analysis, several compounds were potentially of interest for isolation regarding their putative novelty. However, the LC-DAD-MS analysis of *G. laeve* chromatographic fractions indicated that only six compounds could be isolated in amounts suitable for full structure elucidation in order to confirm their identification. The selected compounds were 1 and 2 (m/z 367.1281 and 381.1438, cluster B) and 3 (m/z 629.3476, cluster A). The known leuconolam (4), geissolosimine (5), and geissospermine (6) were also targeted, for biological evaluation.

HRMS-Based Dereplication. Molecular formulas of the aforementioned selected metabolites were generated using the MassHunter software. For every compound, only one or two molecular formulas matched HRMS data generated within the tolerance of the instrument (5 ppm m/z tolerance and lowest *i*-fit below 2).³³ The results of their search against the *Dictionary of Natural Products* (DNP)³⁴ and Reaxys³⁵ databases are summarized in Table S6, Supporting Information. According to this database search, (i) 12 and 11 isomers exist for compounds 1 and 2, respectively. None have been described in the genus *Geissospermum*. (ii) One hit (macrospetrine, a sarpagane-macrolane bisindole alkaloid isolated from the roots of *Rauwolfia verticillata*)³⁶ could be proposed for compound 3. This assignment is dubious, since 3 was located in the geissospermine-like cluster (A) as outlined above. (iii) Twenty-three hits were proposed for compound 4. Since one of these

Table 1. NMR Spectroscopic Data for Geissolaevine (1) and O-Methylgeissolaevine (2) in DMSO-*d*₆

position	1		2	
	δ_{H} , mult. (J in Hz) ^a	δ_{C} ^b	δ_{H} , mult. (J in Hz) ^a	δ_{C} ^b
2		133.1		135.4
3		138.9		137.9
5	8.20, d (5.1)	136.4	8.22, d (5.2)	136.8
6	8.12, d (5.1)	115.4	8.16, d (5.2)	115.8
7		119.9		120.1
8		129.9		130.1
9	8.23, d (7.9)	121.5	8.25, d (7.8)	121.6
10	7.24, t (7.5)	119.4	7.26, t (7.5)	119.5
11	7.55, t (7.6)	128.4	7.59, t (7.7)	128.6
12	7.78, d (8.3)	112.8	7.75, d (8.2)	112.5
13		140.9		140.9
14		106.7		109.5
15		153.4		151.1
16	3.40, d (15.1); 3.46, d (15.1)	31.7	3.54, d (16.8); 3.48, d (16.8)	31.7
17			3.43, brs	50.8
18	1.14, t (7.4)	12.1	1.15, t (7.5)	12.1
19	2.36, q (7.4)	16.9	2.36, dq (13.7, 7.5); 2.43, dq (13.7, 7.5, 7.3)	17.3
20		132.2		133.1
21		171.1		170.5
22		168.3		168.2
23	3.30, s	51.7	3.40, s	51.9
NH-1	11.16, brs		11.24, brs	
OH-17	8.55, brs			

^aData recorded at 400 MHz. ^bData recorded at 100 MHz.

isomers (leuconolam) was part of the in-house alkaloid database, the structural information related to its fragmentation pattern allowed compound **4** to be assigned as leuconolam or its C-21 epimer (*epi*-leuconolam). Interestingly, the structure of *epi*-leuconolam as entered in databases was recently revised by Kam et al. to 6,7-dehydroleuconoxine.³⁷ Here, the isolation and structure elucidation of **4** confirmed the assignment of the node to leuconolam.³⁸ (iv) Although four and seven isomers were proposed for compounds **5** and **6**, respectively, they were assigned unambiguously as geissolosimine (**5**) and geissospermine (**6**), based on chemotaxonomic considerations and on their spectral data, as suggested by their position within the network. This assignment was further confirmed by isolation and structure elucidation.

Structure Elucidation. Compound **1** was obtained as a dark brown, amorphous solid, $[\alpha]_{\text{D}}^{25} +80$ (c 0.025, MeOH). The IR spectrum showed typical absorption bands at ν_{max} 3385 cm^{-1} (OH and NH), 1739, and 1684 cm^{-1} (α,β -unsaturated γ -lactone moiety). The UV spectrum showed absorption maxima at 235, 292, and 350 nm, characteristic of a β -carboline chromophore.³⁹ HRESIMS indicated a $[M + H]^+$ ion peak at m/z 367.1281, consistent with the molecular formula, $\text{C}_{20}\text{H}_{18}\text{N}_2\text{O}_5$ (corresponding to a double-bond equivalent (DBE) value of 13). The ^{13}C NMR spectrum (Table 1) accounted for 20 carbon resonances, comprising two methyl groups, two sp^3 methylenes, six sp^2 methines, two sp^2 oxygenated quaternary carbons, seven sp^2 quaternary carbons, and one sp^3 quaternary carbon. The ^1H NMR spectrum (Table 1) showed four aromatic protons at δ_{H} 8.23 (1H, d, $J = 7.9$ Hz), 7.24 (1H, t, $J = 7.5$ Hz), 7.55 (1H, t, $J = 7.6$ Hz), and 7.78 (1H, d, $J = 8.3$ Hz), and two *ortho*-coupled doublets at δ_{H} 8.20 and 8.12 with $J = 5.1$ Hz, assignable to pyridine ring protons. These characteristic signals suggested the presence of a monosub-

stituted β -carboline moiety in **1**. The ^1H NMR spectrum also displayed a broad downfield one-proton singlet at δ_{H} 11.16 assignable to an indole NH proton, a one-proton broad singlet at δ_{H} 8.55 corresponding to an OH proton, two signals for an ethyl side chain at δ_{H} 1.14 (3H, t, $J = 7.4$ Hz); 2.36 (2H, q, $J = 7.4$ Hz), a pair of AB doublets at δ_{H} 3.40 (1H, d, $J = 15.1$ Hz); 3.46 (1H, d, $J = 15.1$ Hz) indicative of diastereotopic protons of a methylene (CH_2 -16); and a signal at δ_{H} 3.30 corresponding to a methoxy group. The β -carboline moiety together with the above-mentioned two sp^2 -oxygenated carbons accounted for 11 DBE, and the remaining two degrees of unsaturation required the presence of an additional monounsaturated ring in compound **1**. The construction of the latter was deduced by the observed HMBC correlations arising from both the ethyl side chain and the methylene AB system group (CH_2 -16). The ethyl side chain was connected at C-20 by HMBC correlations from H₃-18 to C-20 and from H₂-19 to C-15, C-20, and C-21. Similarly, HMBC cross-peaks from H-16a and H-16b to C-14, C-15, C-20, and C-22 were used to confirm the location of the methylene AB system group between the carbomethoxy group and the unsaturated ring (Figure 2). Afterward, the position of the hydroxy group (OH-17) in the structure was established from the HMBC cross-peak between H-17 and C-14. The

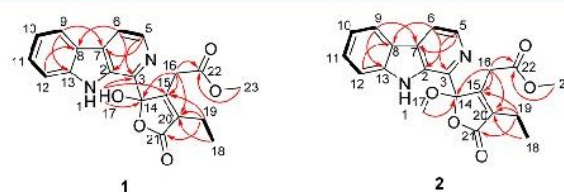


Figure 2. Selected COSY (bold) and HMBC (red arrows) correlations for compounds **1** and **2**.

downfield shift of the hemiacetal C-14 carbon indicated that it was linked to two oxygen atoms, consistent with a γ -hydroxybutenolide unit in the molecule. The junction between the β -carboline unit and the γ -hydroxybutenolide ring was deduced from HMBC correlations observed between H-17 and C-3, C-14, and C-15.

In order to determine the absolute configuration of **1**, its experimental electronic circular dichroism (ECD) spectrum was compared with the theoretically calculated⁴⁰ ECD curves of both the 14R and 14S enantiomers. The absolute configuration of **1** was assigned as 14S, since the experimental and calculated ECD spectra were similar (Figure 3). The compound was given the trivial name geissolaevine (**1**).

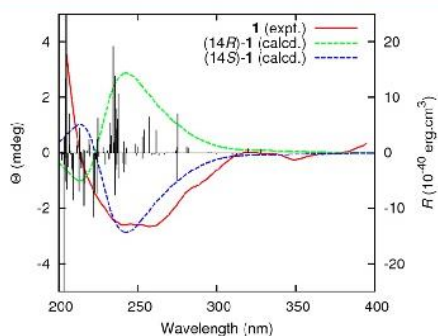


Figure 3. Comparison between the experimental (red) and calculated (green and blue) ECD spectra of geissolaevine (**1**).

Compound **2** was isolated as a dark brown, amorphous solid. The HRESIMS of the protonated molecular positive-ion peak $[M + H]^+$ at m/z 381.1438 established the molecular formula as $C_{21}H_{20}N_2O_5$ (13 DBE), indicating a molecular weight of 14 mass units higher than **1**. In its 1H NMR spectrum (Table 1), the signals observed closely matched those of geissolaevine (**1**). The main differences between **1** and **2** were the presence of an additional signal for a methoxy group at δ_H 3.43 in **2** and the absence of the broad singlet signal corresponding to the hydroxy group. The position of the additional methoxy group (OMe-17) in the structure was established using the HMBC cross-peak observed between H-17 and C-14 (Figure 2). These observations, in combination with a thorough analysis of COSY, HSQC, and HMBC data, supported the assignment of compound **2** as *O*-methylgeissolaevine (**2**).

Compound **2** was isolated as an optically inactive compound (with no optical rotation or Cotton effects in the ECD spectrum when recorded in MeOH). Even if this puzzling feature could call into question the natural origin of *O*-methylgeissolaevine (**2**), this compound was detected unambiguously in the alkaloid extract before any purification process, eliminating an artifactual origin. Regarding its racemic nature, it is likely that **2** derives naturally from the hydrolysis and ring opening of the hydroxy-butenolide of **1** with concomitant ring closure and methoxylation, although the artifactual nature of racemization remains unresolved. Compound **2** was also obtained, in a reasonable yield, when treating **1** with MeOH, overnight at room temperature (see Figure S7, Supporting Information). To the best of our knowledge, compounds **1** and **2** are the first examples of natural β -carboline alkaloids bearing a butenolide moiety. It is worth noting that in a recent medicinal chemistry study the preparation of new β -

carboline- and γ -lactone-based molecular hybrids was reported.⁴¹

Compound **3** was obtained as a yellow, amorphous solid, with $[\alpha]_D^{146} -21$ (*c* 1.5, MeOH). The UV spectrum showed absorption maxima at 253, 308, and 367 nm, characteristic of a β -carbolinium chromophore.⁴² The HRESIMS of the positive-ion peak $[M]^+$ at m/z 629.3476 established the molecular formula as $C_{40}H_{45}N_4O_3^+$ (21 DBE), indicating a molecular weight three mass units lower than geissospermine (**6**). In its 1H NMR spectrum (Table 2), the signals observed matched

Table 2. NMR Spectroscopic Data for Compound **3** in DMSO- d_6

position	δ_H , mult. (J in Hz) ^a	δ_C ^b
2	3.92, m	64.4
3	3.40, d (11.2)	61.6
5	2.87, m; 3.18, m	52.7
6	1.81, m; 2.06, m	43.7
7		50.5
8		133.1
9	7.14, d (6.9)	122.6
10	7.09, t (7.7)	128.4
11	6.69, t (7.4)	119.7
12	6.63, d (8.0)	106.7
13		149.6
14	1.22, m; 1.52, d (14.2)	23.6
15	0.93, s	29.3
16	1.57, m	34.0
17	3.02, d (10.6); 3.06, d (10.4)	63.9
18	0.30, t (7.0)	10.7
19	0.40, m; 0.65, m	23.1
20	1.30, m	40.1
21	1.94, m; 2.79, m	52.6
2'		134.3
3'		139.8
5'	8.55, d (6.5)	132.3
6'	8.66, d (6.5)	115.6
7'		133.1
8'		119.7
9'	8.43, d (8.0)	123.1
10'	7.73, t (7.7)	131.2
11'	7.41, t (7.4)	121.3
12'	7.79, d (8.1)	113.1
13'		143.8
14'	3.58, dd (17.5; 4.5); 4.47, dd (17.3; 3.0)	28.5
15'	3.82, m	32.0
16'	3.30, t (9.2)	51.8
17'	5.79, d (10.1)	81.2
18'	1.79, d (7.0)	12.9
19'	5.87, q (7.0)	124.3
20'		129.9
21'	5.35, d (16.2); 5.45, d (16.2)	58.1
22'		170.9
23'	3.50, s	51.5

^aData recorded at 400 MHz. ^bData recorded at 100 MHz.

closely those of geissospermine (**6**)¹⁷ (see Figures S26 and S32, Supporting Information). The main differences between **3** and **6** were the presence of two additional deshielded pyridine protons at δ_H 8.55 and 8.66 in **3** and the absence of the broad singlet signal at δ_H 3.95, corresponding to H-3' in **6**. These observations, in combination with an analysis of COSY, HSQC,

Table 3. IC₅₀ Values (μM) of the Antiplasmodial, Antileishmanial, and Cytotoxic Activities of Compounds 1–5^a

compound	<i>P. falciparum</i> strain FcB1	<i>L. donovani</i> LV9 axenic amastigotes	<i>L. donovani</i> LV9 intramacrophage amastigotes	MRC-5
1	74.2 ± 1.3	>100	25.7 ± 1.7	>50
2	56.0 ± 9.6	>100	59.6 ± 5.2	>50
3	2.9 ± 1.0	51.8 ± 10.8	57.8 ± 14.4	3.0 ± 0.1
4	>100	>100	25.0 ± 0.7	>50
5	1.9 ± 1.4	9.4 ± 1.4	4.5 ± 0.6	0.022 ± 0.001
standard ^b	0.05 ± 0.02	1.4 ± 0.20	5.2 ± 0.4	0.00052 ± 0.00005

^aValues are mean ± standard errors of three independent experiments. ^bChloroquine (FcB1); miltefosine B (LV9); paclitaxel (MRC-5).

and HMBC data as well as the high structure relatedness evidenced in the molecular network (Figure S3, Supporting Information), strongly suggested that compound 3 is an oxidized form of 6, namely, 3',4',5',6'-tetrahydrogeissospermine.

The NOESY correlations from H-17' to H-17α, H-16', H-15', and H₃-18' and from H-17α to H-15 (Figure S8, Supporting Information, blue arrows) indicated the cofacial orientation of these protons. On the other hand, NOESY correlations from H-2 to H-6β and H-16 were used to place them on the opposite face (Figure S8, Supporting Information, red arrows). The NOESY cross-peaks of compound 3 were comparable to those observed in the NOESY spectrum of geissospermine (6) (see Figures S31 and S33, Supporting Information), suggesting the same stereo-orientation of the protons for the two compounds. Furthermore, despite the presence of an additional chromophore in 3, i.e., a β-carbolinium, a comparison of the experimental ECD spectra of both geissospermine (6) and 3 revealed the same overall Cotton effects (Figure S9, Supporting Information). On the basis of these observations, the absolute configuration of 3 was assigned as shown.⁴³

Biological Testing of Alkaloids 1–5. The purified compounds were evaluated in vitro for their antiparasitic activity against the chloroquine-resistant strain of *Plasmodium falciparum* FcB1 and the axenic and intramacrophage (RAW 264.7) amastigote forms of *Leishmania donovani* (LV9) as well as for their cytotoxic activity against the MRC-5 (human fetal lung fibroblast) cell line (Table 3). Compounds 1 and 2 showed weak or no antiparasitic activity and no cytotoxic activity. Compound 3, in contrast, exhibited moderate antiplasmodial and cytotoxic activities. In addition, no morphological alteration was observed microscopically on infected RAW 264.7 cells at 100 μM for compounds 1, 2, and 4 after a 72 h incubation time, whereas they appeared altered at 50 μM for compound 3, at 12.5 μM for miltefosine, and at 6.25 μM for compound 5. The known compounds (–)-leuconolam (4) and geissosolimine (5) were also evaluated for their antiparasitic and cytotoxic activities. Although geissospermine (6) was not tested, its antiplasmodial and cytotoxic activities have been described previously.¹⁷ In the present work, (–)-leuconolam (4) showed moderate antileishmanial activity against the intramacrophage amastigote form. In addition to its submicromolar antiplasmodial activity, geissosolimine (5) exhibited potent antileishmanial activity for both the axenic and intramacrophage amastigote forms. However, the submicromolar level of cytotoxicity against the MRC-5 human cell line also shown for 5 indicated a low selectivity.

EXPERIMENTAL SECTION

General Experimental Procedures. Optical rotations were measured at 25 °C on a Polar 32 polarimeter. UV spectra were recorded on a Lightwave II+ WPA 7126 V. 1.6.1 spectrophotometer. ECD spectra were measured at 25 °C on a JASCO J-810 spectropolarimeter. IR spectra were recorded with a PerkinElmer type 257 spectrometer. The NMR spectra were recorded on Bruker AM-300 (300 MHz), AM-400 (400 MHz), and AM-600 (600 MHz) NMR spectrometers using CD₃OD, CDCl₃, and DMSO-*d*₆ as solvents. The solvent signals were used as references. Sunfire preparative C₁₈ columns (150 × 19 mm, i.d. 5 μm and 150 × 30 mm, i.d. 5 μm; Waters) were used for preparative HPLC separations using a Waters Delta Prep equipped with a binary pump (Waters 2525) and a UV-visible diode array detector (190–600 nm, Waters 2996). Silica 330 g, 120 g, and 24 g Grace cartridges were used for flash chromatography using an Armen Instrument spot liquid chromatography flash apparatus. Chemicals and solvents were purchased from Sigma-Aldrich.

Plant Material. The bark of *Geissospermum laeve* was collected near Roura (French Guiana) in July 2013. The botanical identification was confirmed by Pierre Silland. A voucher specimen (128470) has been deposited at the “Herbier IRD de Guyane” (Cayenne, French Guiana).

Extraction and Isolation. The air-dried and powdered bark of *G. laeve* (1 kg) was alkalized with 6 M NH₄OH and extracted with CH₂Cl₂ (3 × 1.5 L, 1 h each, 20 °C, atmospheric pressure). The CH₂Cl₂ crude extract was concentrated under vacuum at 38 °C and dissolved in CH₂Cl₂ (50 mL). This solution was subjected to extraction using a solution of hydrochloric acid 0.1 M (1 L).

The mixture was alkalized with the NH₄OH solution until pH 10, and the CH₂Cl₂ crude extract was concentrated under vacuum at 38 °C to yield 27.7 g of an alkaloid extract. A part of this residue (F1, 8.4 g) was subjected to flash chromatography using a silica 120 g Grace cartridge with a gradient of CH₂Cl₂–MeOH (100:0 to 0:100) at 80 mL/min to afford eight fractions, F1.1–F1.8, according to their TLC profiles. Fraction F1.2 (1.2 g) underwent a preparative HPLC separation using a gradient of CH₃CN–H₂O with 0.2% trifluoroacetic acid (TFA) (20:80 to 60:40) to afford geissolaevine (1, 24.3 mg, yield 0.008%). Fraction F1.4 (2.4 g) was fractionated by flash chromatography using a silica 24 g Grace cartridge with a gradient of CH₂Cl₂–MeOH (100:0 to 0:100) at 36 mL/min to afford seven fractions, F1.4.1–F1.4.7, according to their TLC profiles. Fraction F1.4.1 (1.1 g) was subjected to fractionation by silica gel column chromatography using a gradient of CH₂Cl₂–MeOH (100:0 to 80:20) of increasing polarity, leading to eight fractions (F1.4.1.1–F1.4.1.8), on the basis of TLC. Fraction F1.4.1.3 (69.8 mg) was subjected to a preparative HPLC separation using a gradient of MeOH–H₂O with 0.2% TFA (50:50 to 70:30) to give (–)-leuconolam (4, 4.8 mg, yield 0.0005%). Another part of the residue (F2, 3.3 g) was fractionated by flash chromatography using a silica 40 g Grace cartridge with a gradient of CH₂Cl₂–MeOH (100:0 to 0:100) at 15 mL/min to afford eight fractions, F2.1–F2.8, according to their TLC profiles. Fraction F2.3 (105.8 mg) underwent fractionation by flash chromatography using a silica 40 g Grace cartridge with a gradient of CH₂Cl₂–MeOH (100:0 to 0:100) at 10 mL/min to afford 10 fractions, F2.3.1–F2.3.10, on the basis of TLC. Fraction F2.3.3 (28.6 mg) was subjected to preparative HPLC separation using a gradient of MeOH–H₂O with 0.2% formic acid (60:40 to 85:15) to yield *O*-methylgeissolaevine (2, 1.6 mg, yield

0.0013%). The rest of the residue (F3, 16 g) was subjected to flash chromatography using a 330 g silica cartridge with a gradient of CH_2Cl_2 –MeOH (100:0 to 0:100) at 80 mL/min to afford 15 fractions, F3.1–F3.15, according to their TLC profiles. Fraction F3.11 (1.1 g) was further purified by preparative HPLC using a gradient of MeOH– H_2O with 0.2% formic acid (20:80 to 50:50) to afford 3',4',5',6'-tetrahydrogeissospermine (**3**, 2.9 mg, yield 0.0005%).

Fraction F3.13 (1 g) was subjected to preparative HPLC using a gradient of MeOH– H_2O with 0.2% formic acid (20:80–80:20) to give geissosolimine (**5**, 68.1 mg, yield 0.0118%) and geissospermine (**6**, 1 mg, yield 0.0002%).

Geissolaevine (1): dark brown, amorphous solid; $[\alpha]_{\text{D}}^{25} +80$ (c 0.025, MeOH); UV (MeOH) λ_{max} (log ϵ) 235 (4.2), 292 (3.9), 350 (3.3) nm; IR ν_{max} 3385, 2927, 1739, 1684, 1433, 1205 cm^{-1} ; ^1H and ^{13}C NMR data, see Table 1; HRESIMS m/z $[\text{M} + \text{H}]^+$ 367.1281 (calcd for $\text{C}_{20}\text{H}_{19}\text{N}_2\text{O}_5$, 367.1288).

(±)-O-Methylgeissolaevine (2): dark brown, amorphous solid; $[\alpha]_{\text{D}}^{20} 0$ (c 0.017, MeOH); UV (MeOH) λ_{max} (log ϵ) 234 (4.2), 291 (3.9), 347 (3.4) nm; IR ν_{max} 3385, 2927, 1739, 1684, 1433, 1205 cm^{-1} ; ^1H and ^{13}C NMR data, see Table 1; HRESIMS m/z $[\text{M} + \text{H}]^+$ 381.1438 (calcd for $\text{C}_{21}\text{H}_{21}\text{N}_2\text{O}_5$, 381.1445).

3',4',5',6'-Tetrahydrogeissospermine (3): yellow, amorphous powder; $[\alpha]_{\text{D}}^{14.6} -21$ (c 1.5, MeOH); UV (MeOH) λ_{max} (log ϵ) 253 (4.7), 308 (4.5), 367 (3.8) nm; IR ν_{max} 3600–3200, 2963–2877, 1730, 1636, 1488, 1461 cm^{-1} ; ^1H and ^{13}C NMR data, see Table 2; HRESIMS m/z M^+ 629.3476 (calcd for $\text{C}_{40}\text{H}_{45}\text{N}_4\text{O}_3$, 629.3486).

Mass Spectrometry Analysis. Samples were analyzed using an Agilent LC-MS system comprising an Agilent 1260 Infinity HPLC coupled to an Agilent 6530 Q-TOF-MS equipped with an ESI source operating with a positive polarity. A Sunfire analytical C_{18} column (150 \times 2.1 mm; i.d. 3.5 μm , Waters) was used, with a flow rate of 250 $\mu\text{L}/\text{min}$ and a linear gradient from 5% B (A: H_2O + 0.1% formic acid, B: MeOH) to 100% B over 30 min. ESI conditions were set with the capillary temperature at 320 $^\circ\text{C}$, source voltage at 3.5 kV, and a sheath gas flow rate of 10 L/min. The divert valve was set to waste for the first 3 min. There were four scan events: positive MS, window from m/z 100–1200, then three data-dependent MS/MS scans of the first, second, and third most intense ions from the first scan event. MS/MS settings were three fixed collision energies (30, 50, and 70 eV), default charge of 1, minimum intensity of 5000 counts, and isolation width of m/z 2. In the positive-ion mode, purine $\text{C}_5\text{H}_4\text{N}_4$ $[\text{M} + \text{H}]^+$ ion (m/z 121.050873) and the hexakis(1*H*,1*H*,3*H*-tetrafluoropropoxy)-phosphazene $\text{C}_{18}\text{H}_{18}\text{F}_{24}\text{N}_3\text{O}_6\text{P}_3$ $[\text{M} + \text{H}]^+$ ion (m/z 922.009 798) were used as internal lock masses. Full scans were acquired at a resolution of 11 000 (at m/z 922). A permanent MS/MS exclusion list criterion was set to prevent oversampling of the internal calibrant.

Molecular Networking. The network was created using the HPLC-HRMS/MS data from the alkaloid extract of the bark of *G. laeve* as well as the 55 reference compounds of the in-house alkaloid database. Data were converted from the standard .d format (Agilent standard data-format) to the .mgf format (Mascot Generic Format) using the software MassHunter Workstation from Agilent. The converted data files were processed using the molecular networking method developed by Dorrestein and co-workers.⁶ This generation of the molecular network was carried out using the online workflow at GNPS (<http://gnps.ucsd.edu>). The resulting analysis and parameters for the network can be accessed via the link <http://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=0824b6f510df454f90d6a3ca528cf4ec> (crude extract) and <http://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=f367ad47423c405697c94a71290a0126> (crude extract + alkaloid database substances). The following settings were used for generation of the network: minimum pairs cos 0.6; parent mass tolerance, 1.0 Da; ion tolerance, 0.3; network topK, 10; minimum matched peaks, 6; minimum cluster size, 2. The molecular networking data were analyzed and visualized using Cytoscape (ver. 2.8.2).⁴⁴

Computational Methods. Conformations of compound **1** were fully optimized in vacuo and without constraint using the DFT method,^{45,46} with the hybrid Becke3LYP functional^{47,48} and the 6-31G* basis,⁴⁹ as implemented in the Gaussian 09 software package.⁵⁰ Vibrational analysis within the harmonic approximation was performed

at the same level of theory upon geometrical optimization convergence, and local minima were characterized by the absence of imaginary frequency. The UCSF Chimera v1.11 software package was used for the depiction of the most stable conformer of compound **1** (Figure S11, Supporting Information).⁵¹ Excitation energies and corresponding rotational strengths for the first 30 electronic transitions of each of the five most populated conformations (accounting for a cumulated population of >90%) were calculated using the TDDFT method⁵² at the B3LYP/6-31G* level (Table S10, Supporting Information). Individual ECD spectra were simulated by the summation of Gaussian functions⁵³ using SpecDis^{54,55} v1.64 software with a sigma/gamma value of 0.4 eV. The final predicted ECD spectrum of compound **1** was a Boltzmann population-weighted average of these individual spectra and was plotted with Gnuplot v4.6.⁵⁶ The experimental ECD spectrum of **1** was smoothed by moving-average.

Biological Assays. See Experimental Method S12, Supporting Information.

■ ASSOCIATED CONTENT

● Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jnatprod.6b01013.

NMR spectra of compounds **1–3** as well as ^1H NMR and NOESY spectra of geissospermine (**6**), description of the in-house alkaloid database, comparison of EICs and MS/MS spectra of the matches and the model coordinates of the most stable conformation of **1** (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: pierre.champy@u-psud.fr (P. Champy).

*Tel: +33 1 46 83 55 87. Fax: +33 1 46 83 53 99. E-mail: mehdi.beniddir@u-psud.fr (M. Beniddir).

ORCID

Marc Litaudon: 0000-0002-0877-8234

Mehdi A. Beniddir: 0000-0003-2153-4290

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work has been developed with the funding from Cienciaactiva, an initiative of the National Council for Science, Technology and Technological Innovation (CONCYTEC) contract 239-2015-FONDECYT. In addition, this work was also supported by the French ANR grant ANR-15-CE29-0001. We express our thanks to Dr. J. Bignon (ICSN) for performing the cytotoxic assays. We also thank K. Leblanc (BioCIS) for her assistance in the preparative HPLC purifications. We are grateful to Dr. M. Wang, Prof. P. Dorrestein, and Prof. N. Bandeira (University of California, San Diego) for developing and making public the GNPS platform.

■ REFERENCES

- O'Connor, S. E.; Maresh, J. J. *Nat. Prod. Rep.* **2006**, *23*, 532–547.
- Szabó, L. *Molecules* **2008**, *13*, 1875–1896.
- Kurita, K. L.; Linington, R. G. *J. Nat. Prod.* **2015**, *78*, 587–596.
- Beniddir, M. A.; Grellier, P.; Rasoanaivo, P.; Loiseau, P. M.; Bories, C.; Dumontet, V.; Gueritte, F.; Litaudon, M. *Eur. J. Org. Chem.* **2012**, *2012*, 1039–1046.
- Beniddir, M. A.; Martin, M.-T.; Tran Huu Dau, M.-E.; Grellier, P.; Rasoanaivo, P.; Guéritte, F.; Litaudon, M. *Org. Lett.* **2012**, *14*, 4162–4165.

- (6) Yang, J. Y.; Sanchez, L. M.; Rath, C. M.; Liu, X.; Boudreau, P. D.; Bruns, N.; Glukhov, E.; Wodtke, A.; de Felicio, R.; Fenner, A.; Wong, W. R.; Linington, R. G.; Zhang, L.; Debonsi, H. M.; Gerwick, W. H.; Dorrestein, P. C. *J. Nat. Prod.* **2013**, *76*, 1686–1699.
- (7) Morton, J. F. *Atlas of Medicinal Plants of Middle America*; C. C., Thomas: Springfield, IL, 1981.
- (8) Steele, J. C. P.; Veitch, N. C.; Kite, G. C.; Simmonds, M. S. J.; Warhurst, D. C. *J. Nat. Prod.* **2002**, *65*, 85–88.
- (9) Milliken, W.; Albert, B. *Econ. Bot.* **1997**, *51*, 264–278.
- (10) Hughes, N. A.; Rapoport, H. *J. Am. Chem. Soc.* **1958**, *80*, 1604–1609.
- (11) Rapoport, H.; Onak, T. P.; Hughes, N. A.; Reinecke, M. G. *J. Am. Chem. Soc.* **1958**, *80*, 1601–1604.
- (12) Rapoport, H.; Windgassen, R. J.; Hughes, N. A.; Onak, T. P. *J. Am. Chem. Soc.* **1960**, *82*, 4404–4414.
- (13) Rapoport, H.; Moore, R. E. *J. Org. Chem.* **1962**, *27*, 2981–2985.
- (14) Moore, R. E.; Rapoport, H. *J. Org. Chem.* **1973**, *38*, 215–230.
- (15) Ishiyama, H.; Matsumoto, M.; Sekiguchi, M.; Shigemori, H.; Ohsaki, A.; Kobayashi, J. *Heterocycles* **2005**, *66*, 651–658.
- (16) Lima, J. A.; Costa, R. S.; Epifanio, R. A.; Castro, N. G.; Rocha, M. S.; Pinto, A. C. *Pharmacol., Biochem. Behav.* **2009**, *92*, S08–S13.
- (17) Mbeunkui, F.; Grace, M. H.; Lategan, C.; Smith, P. J.; Raskin, I.; Lila, M. A. *J. Ethnopharmacol.* **2012**, *139*, 471–477.
- (18) Mbeunkui, F.; Grace, M. H.; Lila, M. A. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2012**, *885–886*, 83–89.
- (19) Lima, J. A.; Costa, T. W. R.; Silva, L. L.; Miranda, A. L. P.; Pinto, A. C. *An. Acad. Bras. Cienc.* **2016**, *88*, 237–248.
- (20) Medema, M. H.; Fischbach, M. A. *Nat. Chem. Biol.* **2015**, *11*, 639–648.
- (21) Watrous, J.; Roach, P.; Alexandrov, T.; Heath, B. S.; Yang, J. Y.; Kersten, R. D.; van der Voort, M.; Pogliano, K.; Gross, H.; Raaijmakers, J. M.; Moore, B. S.; Laskin, J.; Bandeira, N.; Dorrestein, P. C. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, E1743–E1752.
- (22) Bouslimani, A.; Sanchez, L. M.; Garg, N.; Dorrestein, P. C. *Nat. Prod. Rep.* **2014**, *31*, 718–729.
- (23) Hubert, J.; Nuzillard, J.-M.; Renault, J.-H. *Phytochem. Rev.* **2015**, *14*, 1–41.
- (24) MassBank <http://www.massbank.jp/> (accessed Feb 10, 2017).
- (25) Metlin <https://metlin.scripps.edu/> (accessed Feb 17, 2017).
- (26) ReSpec <http://spectra.psc.riken.jp/menta.cgi/respect/index> (accessed Feb 17, 2017).
- (27) The National Institute of Standards and Technology. NIST <http://www.nist.gov/srd/nist1a.cfm> (accessed Feb 17, 2017).
- (28) Wang, M.; Carver, J. J.; Phelan, V. V.; Sanchez, L. M.; Garg, N.; Peng, Y.; Nguyen, D. D.; Watrous, J.; Kaponov, C. A.; Luzzatto-Knaan, T.; Porto, C.; Bouslimani, A.; Melnik, A. V.; Meehan, M. J.; Liu, W.-T.; Crusemann, M.; Boudreau, P. D.; Esquenazi, E.; Sandoval-Calderon, M.; Kersten, R. D.; Pace, L. A.; Quinn, R. A.; Duncan, K. R.; Hsu, C.-C.; Floros, D. J.; Gavilan, R. G.; Kleigrewe, K.; Northen, T.; Dutton, R. J.; Parrot, D.; Carlson, E. E.; Aigle, B.; Michelsen, C. F.; Jelsbak, L.; Sohlenkamp, C.; Pevzner, P.; Edlund, A.; McLean, J.; Piel, J.; Murphy, B. T.; Gerwick, L.; Liaw, C.-C.; Yang, Y.-L.; Humpf, H.-U.; Maansson, M.; Keyzers, R. A.; Sims, A. C.; Johnson, A. R.; Sidebottom, A. M.; Sedio, B. E.; Klitgaard, A.; Larson, C. B.; Boya, P. C. A.; Torres-Mendoza, D.; Gonzalez, D. J.; Silva, D. B.; Marques, L. M.; Demarque, D. P.; Pociute, E.; O'Neill, E. C.; Briand, E.; Helfrich, E. J. N.; Granatosky, E. A.; Glukhov, E.; Ryffel, F.; Houson, H.; Mohimani, H.; Kharbush, J. J.; Zeng, Y.; Vorholt, J. A.; Kurita, K. L.; Charusanti, P.; McPhail, K. L.; Nielsen, K. F.; Vuong, L.; Elfeki, M.; Traxler, M. F.; Engene, N.; Koyama, N.; Vining, O. B.; Baric, R.; Silva, R. R.; Mascuch, S. J.; Tomasi, S.; Jenkins, S.; Macherla, V.; Hoffman, T.; Agarwal, V.; Williams, P. G.; Dai, J.; Neupane, R.; Gurr, J.; Rodriguez, A. M. C.; Lamsa, A.; Zhang, C.; Dorrestein, K.; Duggan, B. M.; Almaliti, J.; Allard, P.-M.; Phapale, P.; Nothias, L.-F.; Alexandrov, T.; Litaudon, M.; Wolfender, J.-L.; Kyle, J. E.; Metz, T. O.; Peryea, T.; Nguyen, D.-T.; VanLeer, D.; Shinn, P.; Jadhav, A.; Muller, R.; Waters, K. M.; Shi, W.; Liu, X.; Zhang, L.; Knight, R.; Jensen, P. R.; Palsson, B. O.; Pogliano, K.; Linington, R. G.; Gutierrez, M.; Lopes, N. P.; Gerwick, W. H.; Moore, B. S.; Dorrestein, P. C.; Bandeira, N. *Nat. Biotechnol.* **2016**, *34*, 828–837.
- (29) Gu, J.; Gui, Y.; Chen, L.; Yuan, G.; Lu, H.-Z.; Xu, X. *PLoS One* **2013**, *8*, e62839.
- (30) Allard, P.-M.; Péresse, T.; Bisson, J.; Gindro, K.; Marcourt, L.; Pham, V. C.; Roussi, F.; Litaudon, M.; Wolfender, J.-L. *Anal. Chem.* **2016**, *88*, 3317–3323.
- (31) Buckingham, J.; Baggaley, K. H.; Roberts, A. D.; Szabó, L. F. *Dictionary of Alkaloids, with CD-ROM*; CRC Press/ Taylor & Francis Group: New York, 2010; p 59.
- (32) Schymanski, E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H. P.; Hollender, J. *Environ. Sci. Technol.* **2014**, *48*, 2097–2098.
- (33) Kind, T.; Fiehn, O. *BMC Bioinf.* **2006**, *7*, 234.
- (34) DNP <http://dnp.chemnetbase.com/> (accessed Feb 17, 2017).
- (35) Reaxys <https://www.reaxys.com> (accessed Feb 17, 2017).
- (36) Lin, M.; Yang, B. Q.; Yu, D. Q.; Lin, X. Y.; Zhang, Y. J. *Acta Pharm. Sin.* **1987**, *22*, 833–836.
- (37) Low, Y.-Y.; Hong, F.-J.; Lim, K.-H.; Thomas, N. F.; Kam, T.-S. *J. Nat. Prod.* **2014**, *77*, 327–338.
- (38) Goh, S. H.; Wei, C.; M. Ali, A. R. *Tetrahedron Lett.* **1984**, *25*, 3483–3484.
- (39) Ohmoto, T.; Koike, K.; Higuchi, T.; Ikeda, K. *Chem. Pharm. Bull.* **1985**, *33*, 3356–3360.
- (40) Pescitelli, G.; Bruhn, T. *Chirality* **2016**, *28*, 466–474.
- (41) Singh, D.; Devi, N.; Kumar, V.; Malakar, C. C.; Mehra, S.; Rattan, S.; Rawal, R. K.; Singh, V. *Org. Biomol. Chem.* **2016**, *14*, 8154–8166.
- (42) Frédéricich, M.; Quetin-Leclercq, J.; Biala, R. G.; Brandt, V.; Penelle, J.; Tits, M.; Angenot, L. *Phytochemistry* **1998**, *48*, 1263–1266.
- (43) Chiaroni, A.; Riche, C.; Païs, M.; Goutarel, R. *Tetrahedron Lett.* **1976**, *17*, 4729–4730.
- (44) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. *Genome Res.* **2003**, *13*, 2498–2504.
- (45) Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B864–B871.
- (46) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133–A1138.
- (47) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785–789.
- (48) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- (49) Hehre, W. J. R. L.; Schleyer, P. V. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986.
- (50) Frisch, M. J.; Trucks, H. B.; Schlegel, G. W.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09 Revision D.01*; Gaussian, Inc.: Wallingford, CT, 2013.
- (51) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25*, 1605–1612.
- (52) Bauernschmitt, R.; Häser, M.; Treutler, O.; Ahlrichs, R. *Chem. Phys. Lett.* **1997**, *264*, 573–578.
- (53) Stephens, P. J.; Harada, N. *Chirality* **2010**, *22*, 229–233.
- (54) Bruhn, T.; Schaumlöffel, A.; Hemberger, Y. *SpecDis*, Version 1.64; University of Würzburg: Germany, 2015.
- (55) Bruhn, T.; Schaumlöffel, A.; Hemberger, Y.; Bringmann, G. *Chirality* **2013**, *25*, 243–249.
- (56) Williams, T.; Kelley, C.; et al. *Gnuplot 4.6: an Interactive Plotting Program*; 2014. <http://gnuplot.sourceforge.net/>.

Supplementary information:

Figure S1. Generation and composition of the in-house database

Generation of the in-house monoterpene indole alkaloid database. This database was generated from the acquisition of MS/MS spectra of 55 monoterpene indole alkaloids [(47 monomers (**4**, **9-54**), 8 bisindoles (**5**, **6**, **55-60**)], which were previously isolated in the laboratory from various plants in the order Gentianales, or obtained as commercial standards. These reference compounds were characterized by analysis of their NMR, UV-vis and HRMS data. For each compound, an average MS/MS spectrum of three collision energies (30, 50 and 70 eV) was obtained, in the positive mode, using mono-charged molecular or pseudomolecular ions. These three values were chosen after several tests. Indeed, we noticed that bisindoles and monoindoles gave different MS/MS fragmentation pattern depending on the value of the collision energy. Thus, to address this issue we decided to use three collision energies to afford averaged MS/MS spectra. The acquisition of the spectra was done using an Agilent 1260 Infinity Q-TOF 6530 accurate mass spectrometer with a SunFire® analytical C18 column (150 × 2.1 mm; i.d. 3.5 µm, Waters) using a gradient of MeOH/H₂O with 0.2% formic acid (5:95-100:0 in 30 min) at 0.25 mL/min. All the reference substances were dissolved in MeOH to obtain 2.53 mM solutions. The injection volume of the samples was set at 5 µL.



Monoindoles

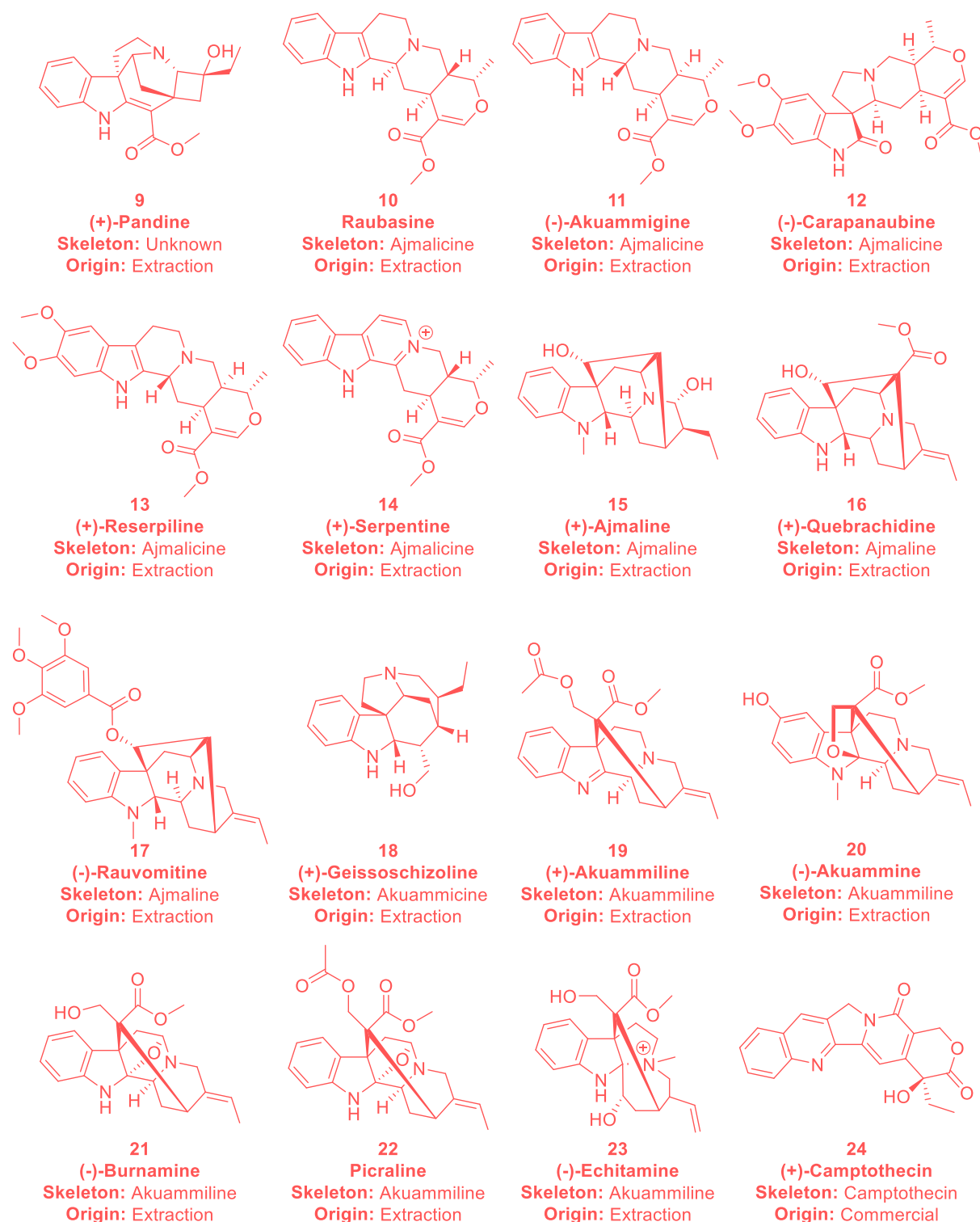


Figure S1. continued



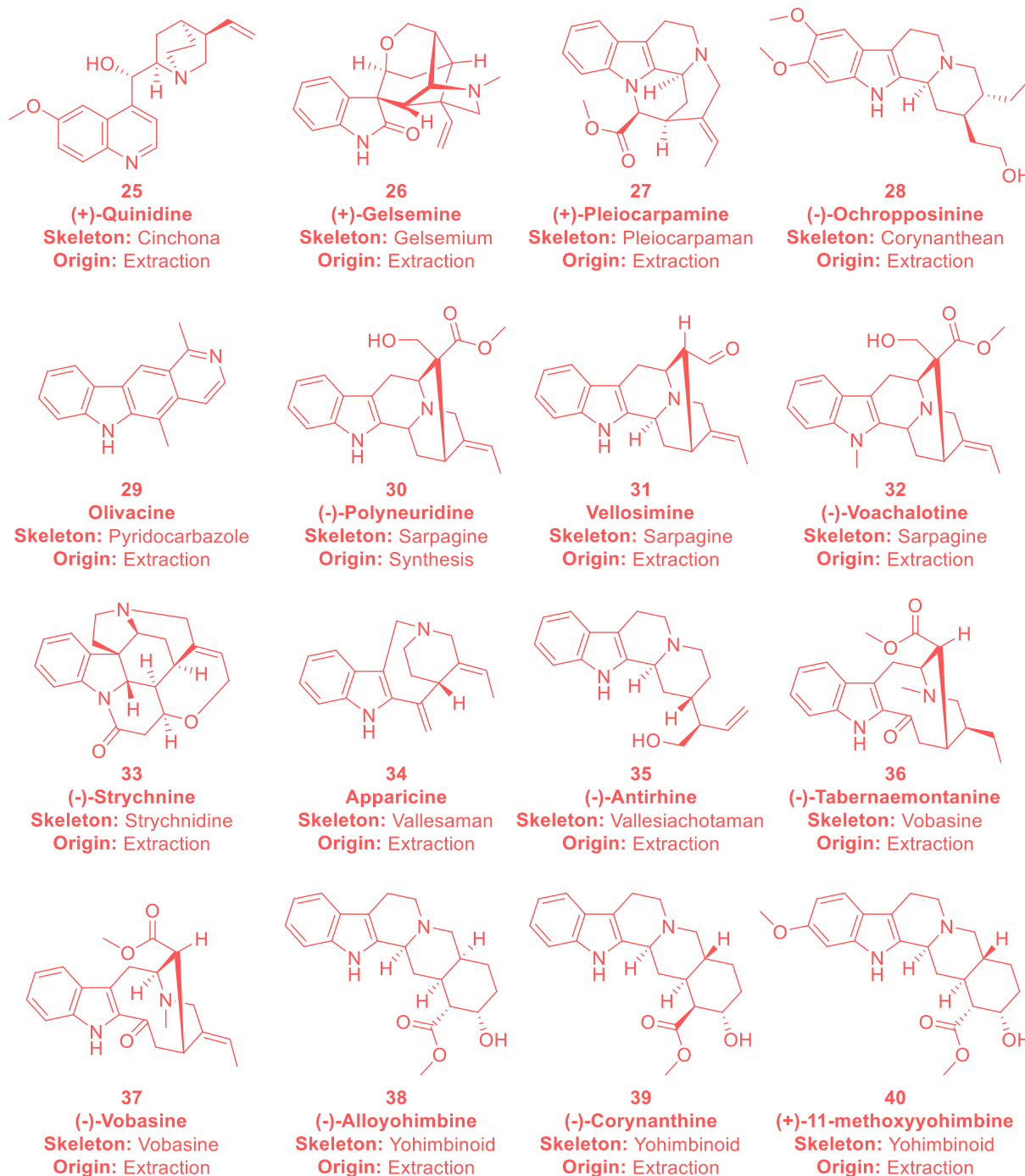


Figure S1. continued



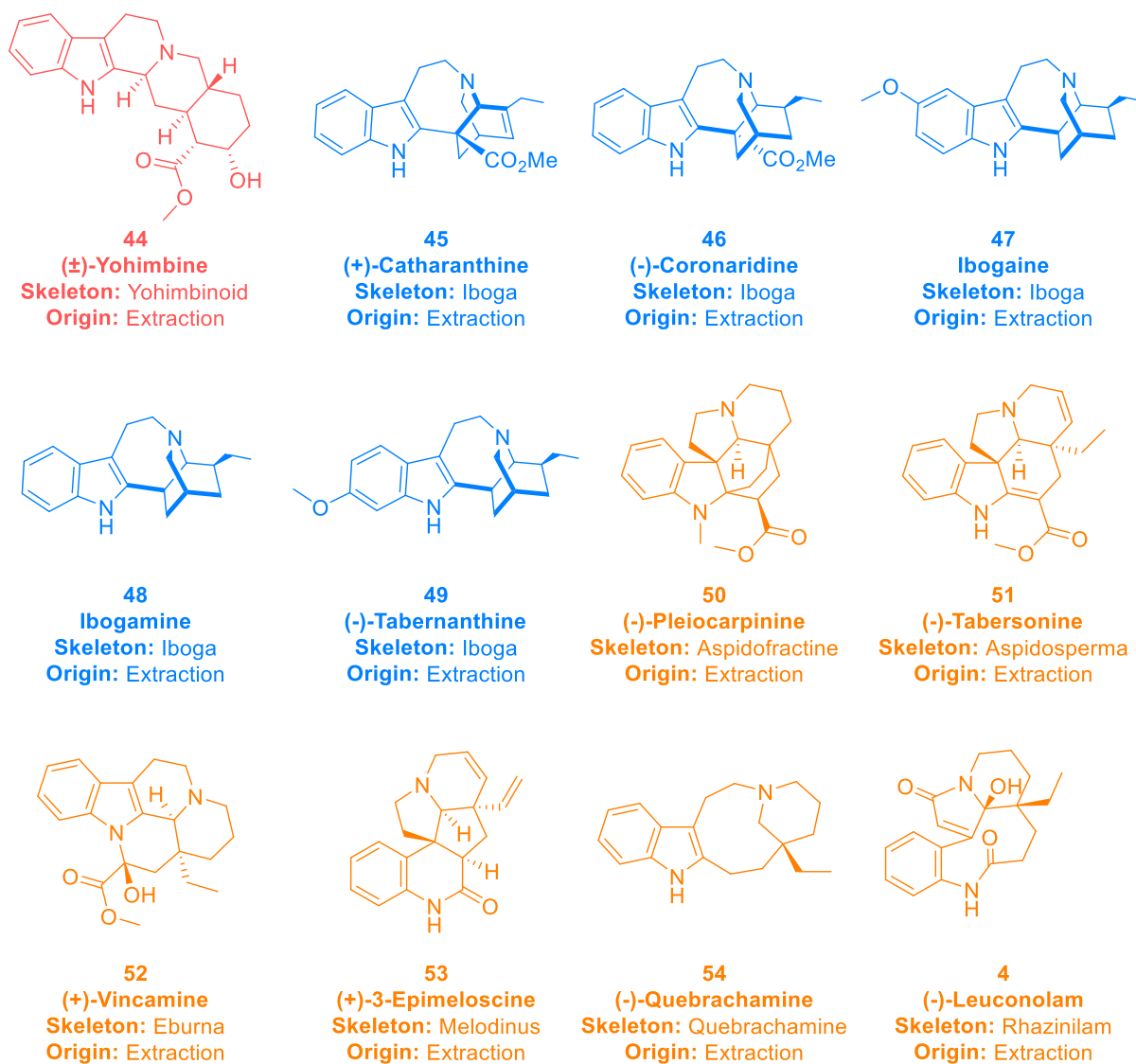


Figure S1. Chemical structures of the monoterpene indole (mono-indoles) alkaloids of the in-house database. Red: Type I; Blue: Type II; Orange: Type III.

Bisindoles

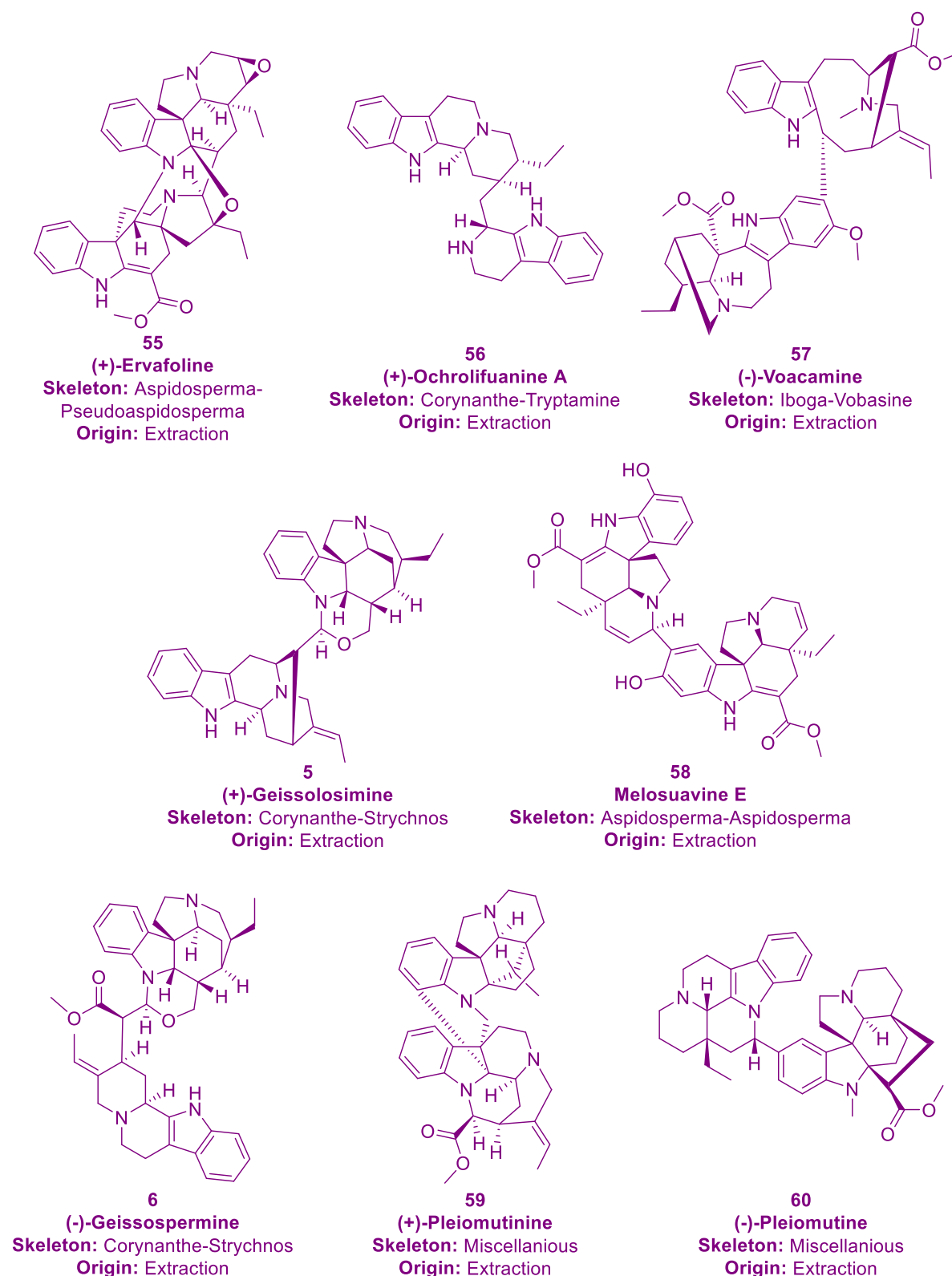
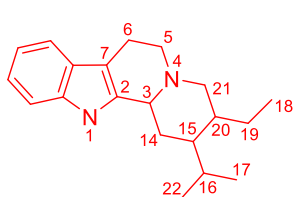


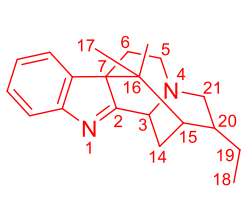
Figure S1. Chemical structures of the monoterpene indole alkaloids (bisindoles) of the in-house alkaloid database.



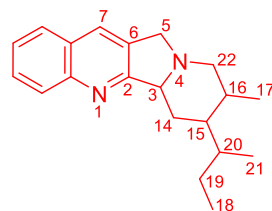
Figure S2. Monoterpene indole alkaloids skeletons according to the dictionary of alkaloids¹



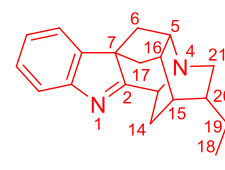
Ajmalicine



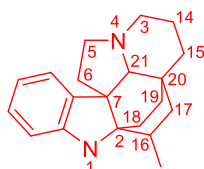
Akuammiline



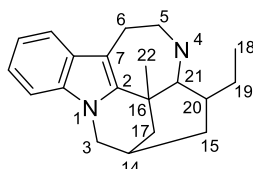
Camptothecin



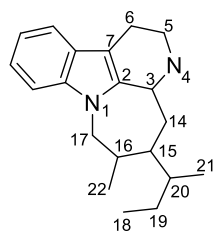
Ajmaline



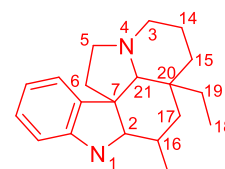
Aspidofractine



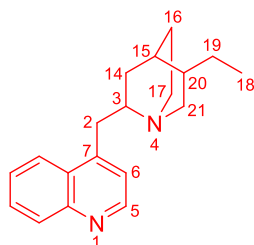
Chippiine



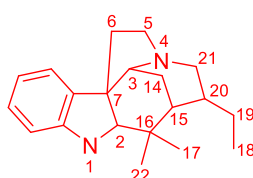
Akageran



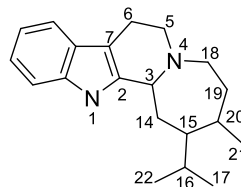
Aspidosperma



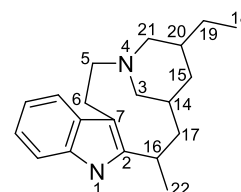
Cinchona



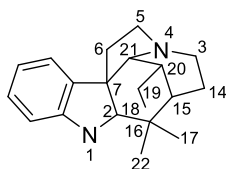
Akuammicine



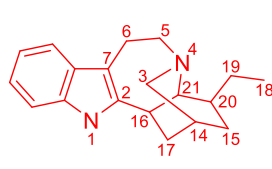
Cadamban



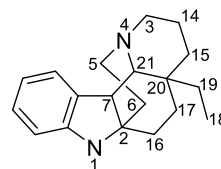
Cleavamine



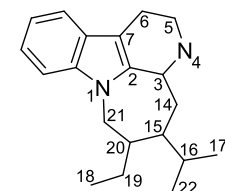
Condylocarpan



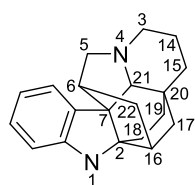
Iboga



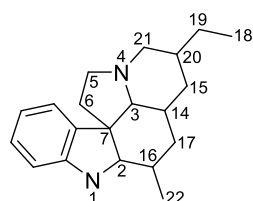
Melonine



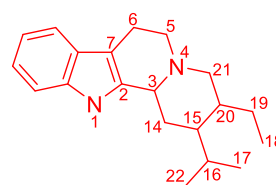
Correantane



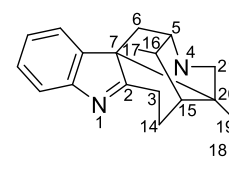
Kopsane



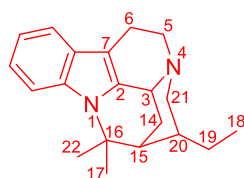
Pandoline



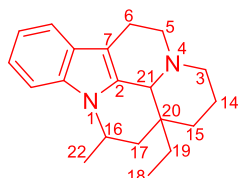
Corynanthean



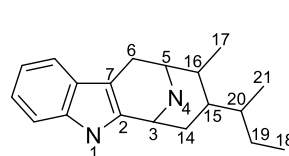
Koumine



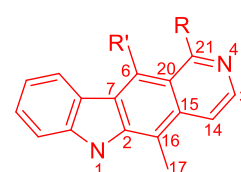
Pleiocarpaman



Eburna



Macroline



Pyridocarbazole

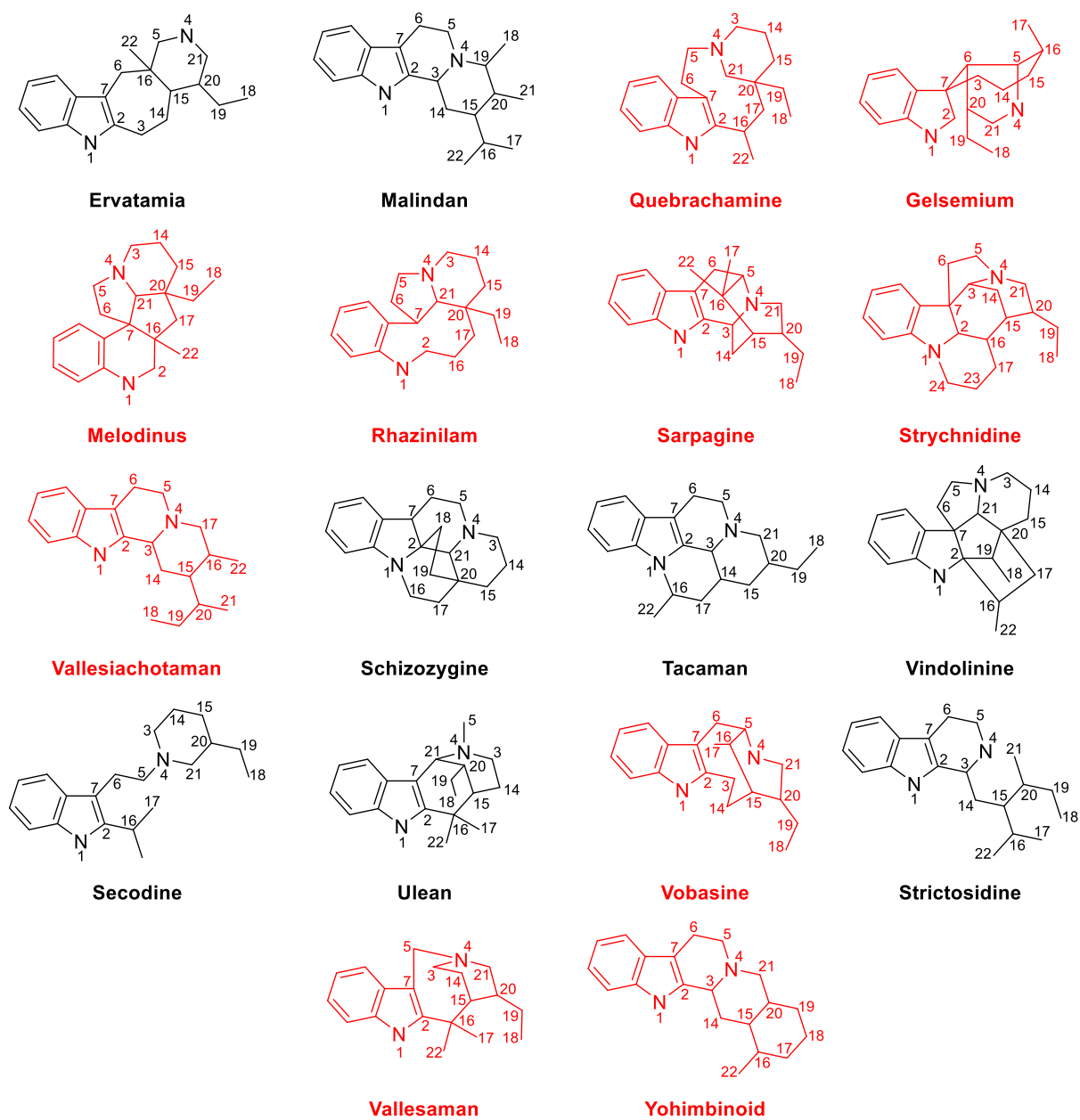


Figure S2. Monoterpene indole alkaloids skeletons. In red, the skeletons included in the in-house database



Figure S3. Molecular network: Cluster A

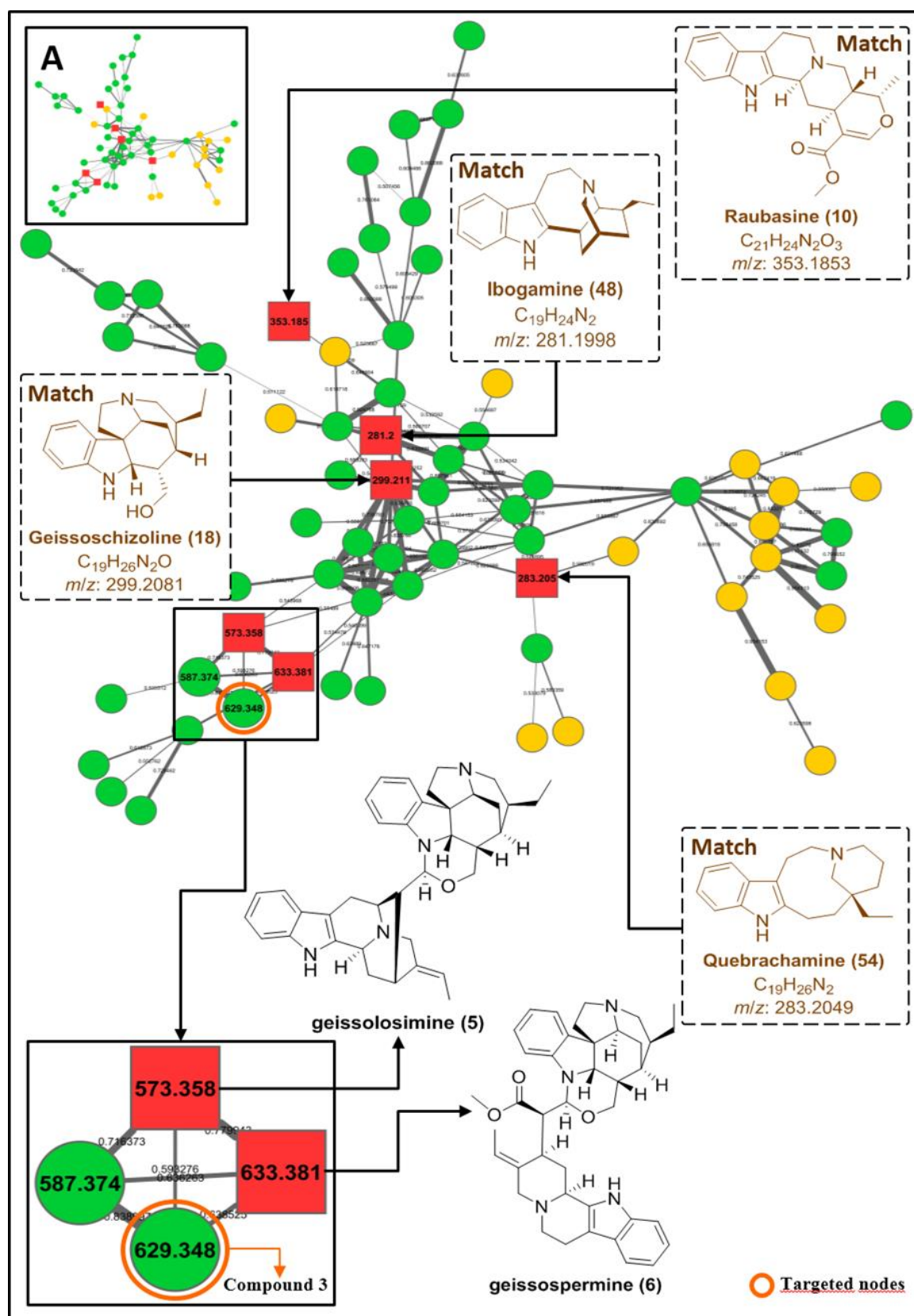


Figure S3. Cluster A of the full molecular network of the alkaloid extract from the bark of *Geissospermum laeve*, and the reference compounds from the in-house alkaloid database.



Figure S4. Molecular network: Clusters B and C

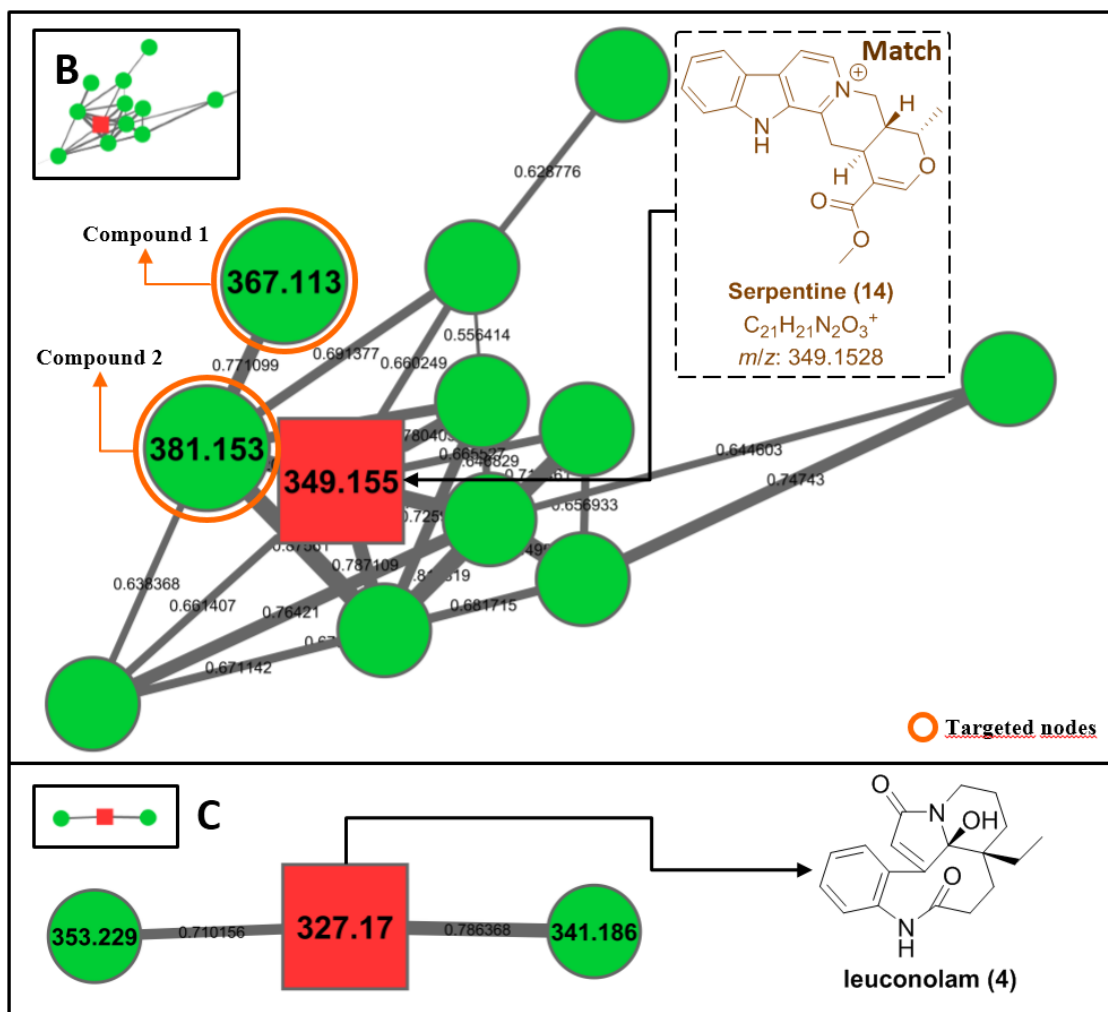
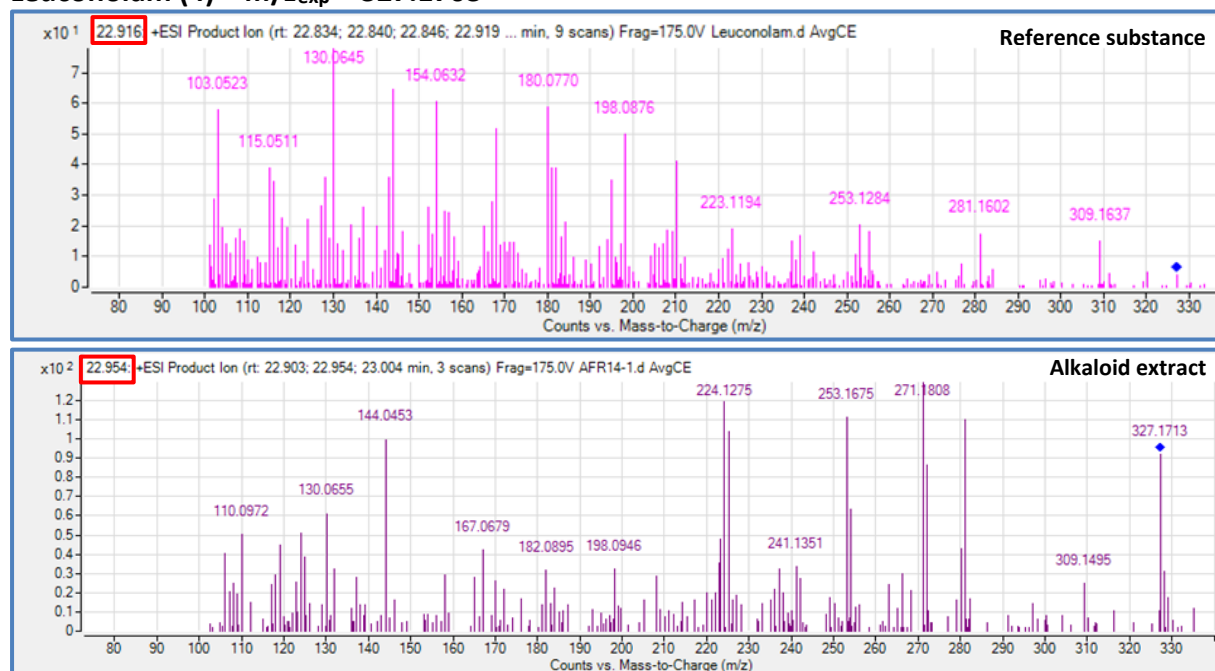


Figure S4. Cluster B, C, and D of the full molecular network of the alkaloid extract from the bark of *Geissospermum laeve*, and the reference compounds from the in-house alkaloid database.

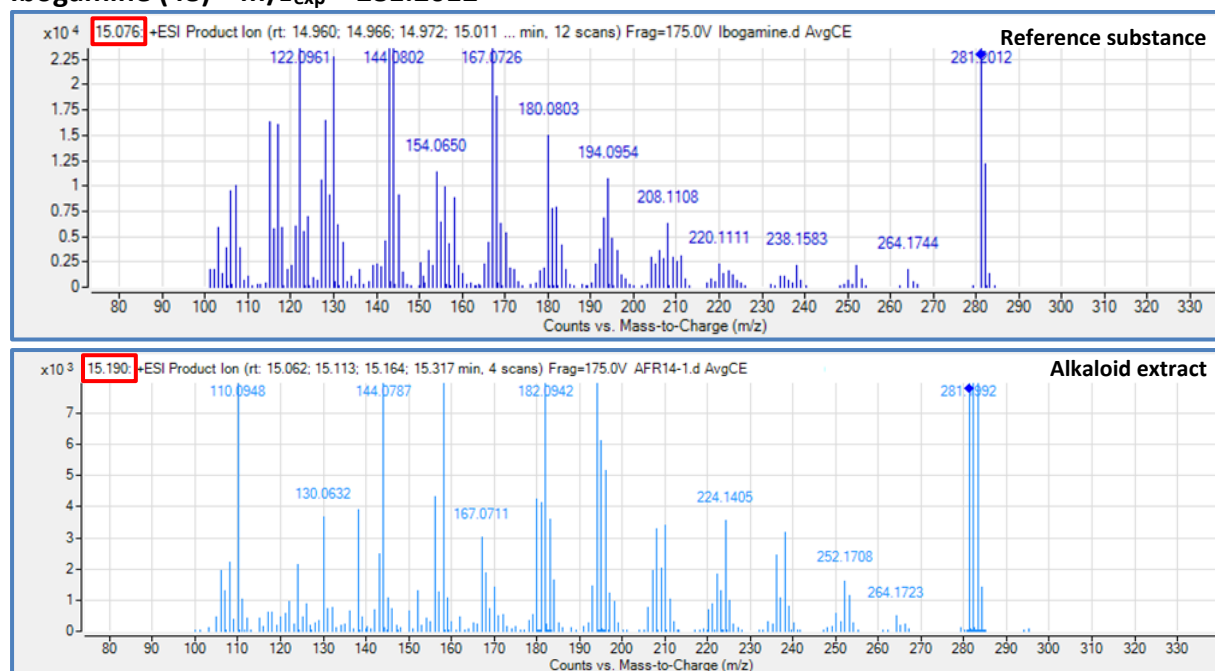
Figure S5. Comparison of MS/MS spectra from the references substances of the in-house database and the matches in the extract

It must be pointed out that all matches between network spectra and library spectra were required to have a score above 0.6 and at least 6 matched peaks.

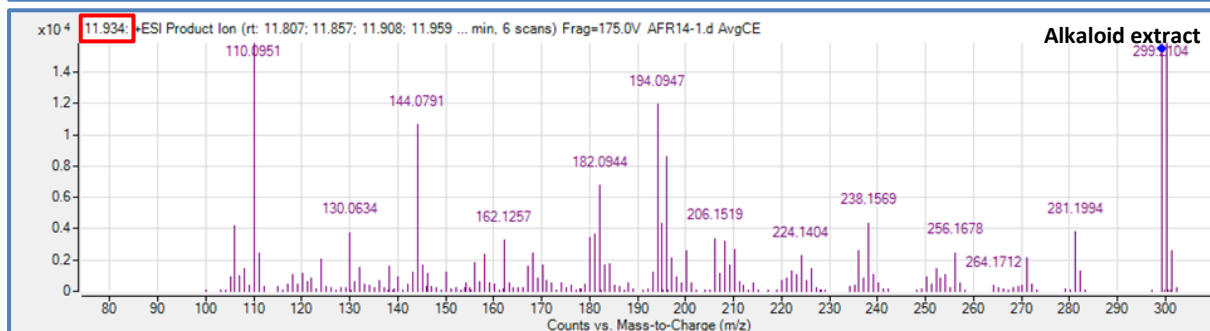
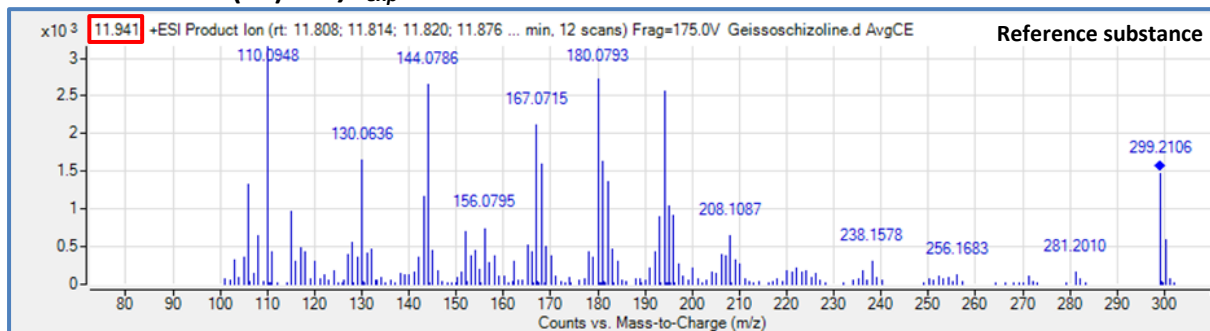
Leuconolam (4) – $m/z_{\text{exp}} = 327.1703$



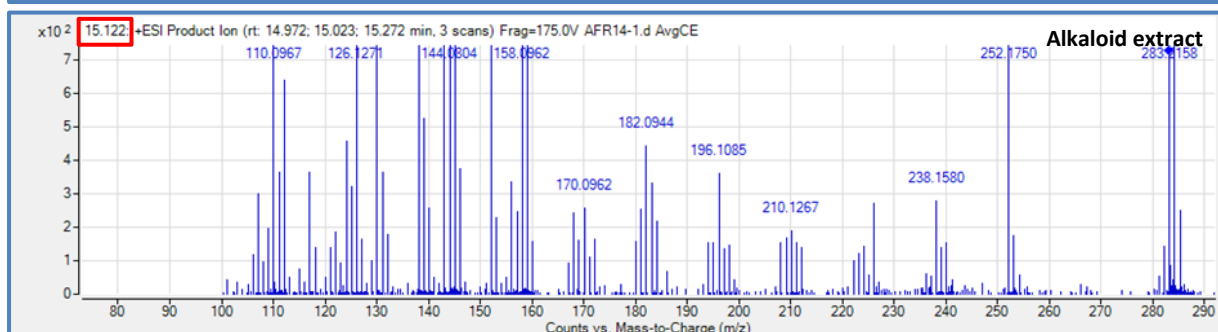
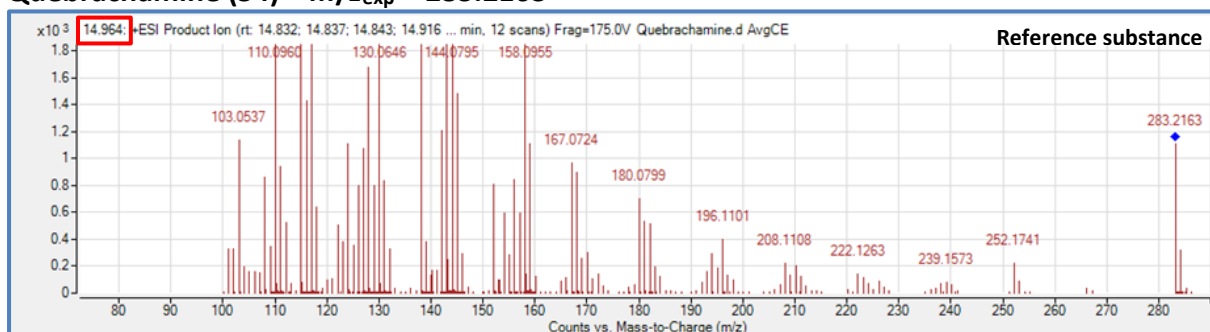
Ibogamine (48) – $m/z_{\text{exp}} = 281.2012$



Geissoschizoline (18) – $m/z_{exp} = 299.2106$

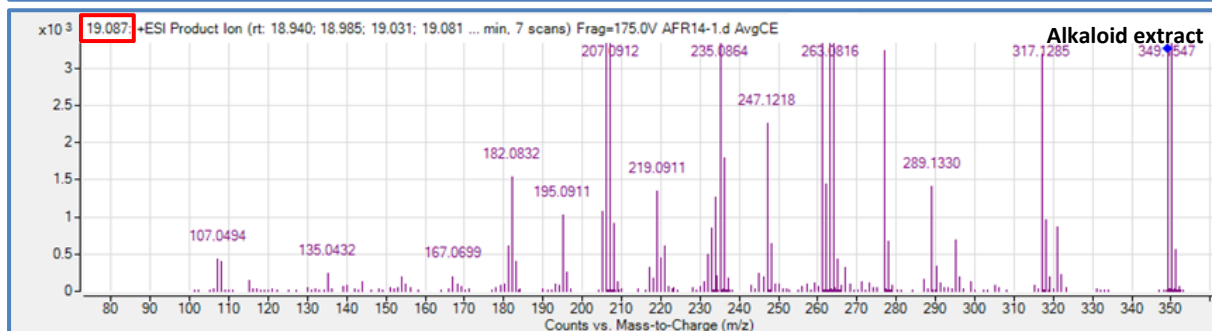
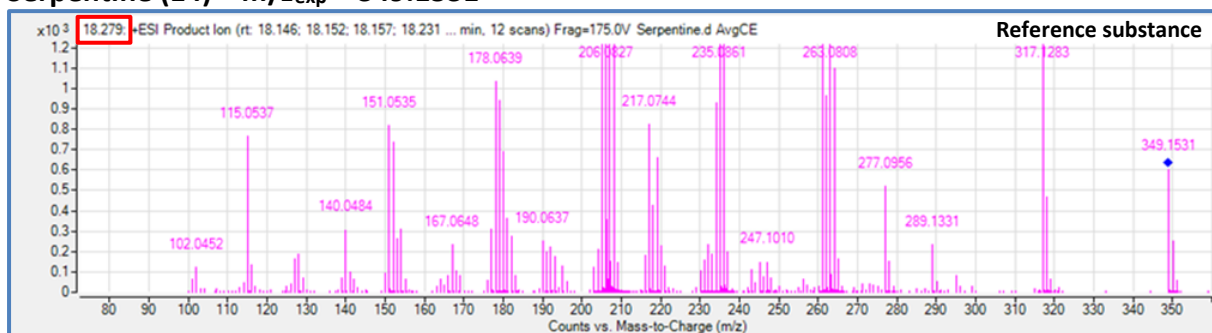


Quebrachamine (54) – $m/z_{exp} = 283.2163$

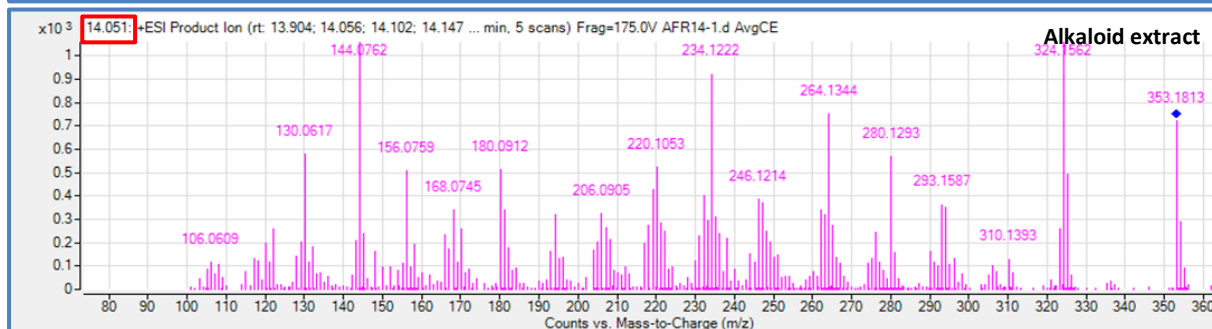
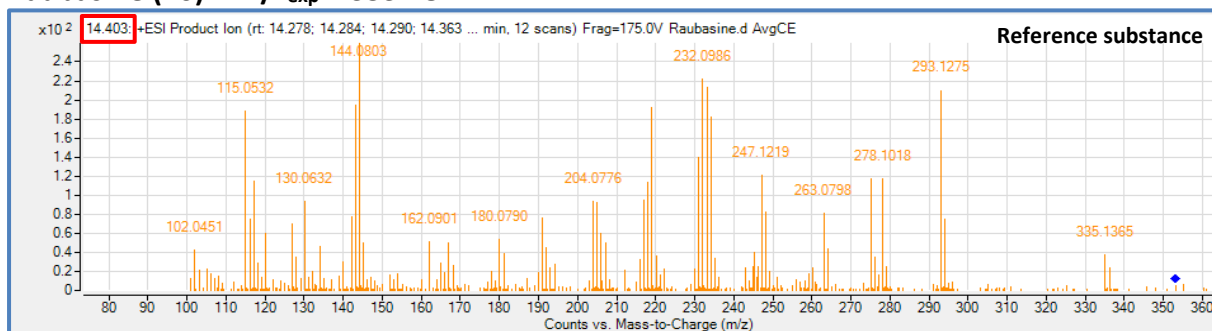




Serpentine (14) – $m/z_{exp} = 349.1531$



Raubasine (10) – $m/z_{exp} = 353.1842$



Akuammigine (11) – $m/z_{\text{exp}} = 353.1835$

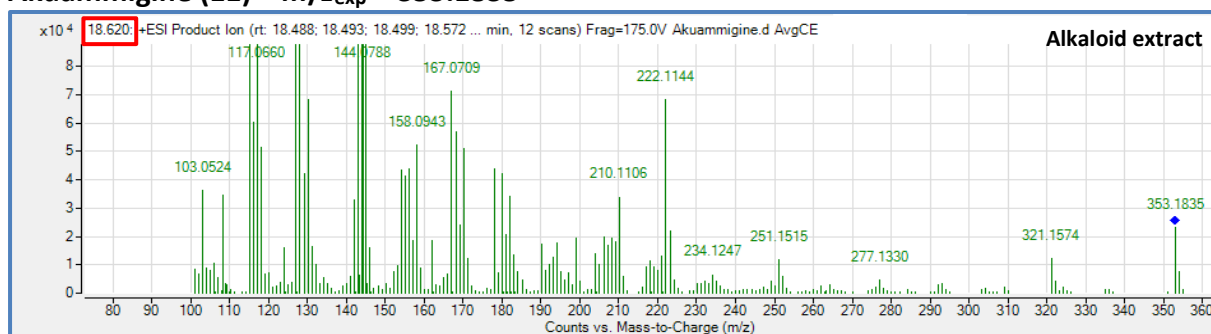


Figure S5. The MS/MS spectra of the reference substances that matched in the Molecular Network of the alkaloid extract and the in-house database were compared. All of the matches had similar MS/MS spectra and retention time (RT) in both the reference sample and the extract. To ensure that the compound at RT = 14.051 and $m/z = 353.1842$ was truly raubasine (**10**) at not one of its epimers, such as akuammigine (**11**), we compared the RT of each reference substance with the one present in the extract. The RT of raubasine (**10**) is closer than that of akuammigine (**11**) in regard to the peak of interest.

The other substances that matched also have epimers that have been described in nature, a list of which can be found below.

- Leuconolam (**4**): no epimers.
- Serpentine (**14**): Alstonine (20-epimer); epi-alstonine (19,20-diepimer); 19-episerpentine (19-epimer).
- Raubasine (**10**): Isoajmalicine (3-epimer); mayumbine (19-epimer); tetrahydroalstonine (20-epimer); 19-epi-3-isoajmalicine (3,19-diepimer); rauniticine (19,20-diepimer); 3-isorauniticine (3,19,20-triepimer).
- Geissoschizoline (**18**): no epimers.
- Ibogamine (**48**): no epimers.
- Quebrachamine (**54**): two enantiomers.



Table S6. HRMS-based dereplication – Selected compounds

Table S6: Dereplication of the selected compounds against DNP and Reaxys® databases

compounds	species	experimental <i>m/z</i>	calculated <i>m/z</i>	hits	molecular formula
1	[M+H] ⁺	367.1281	367.1288	12 hits - Not described in <i>G. laeve</i>	C ₂₀ H ₁₈ N ₂ O ₅
2	[M+H] ⁺	381.1438	381.1445	11 hits - Not described in <i>G. laeve</i>	C ₂₁ H ₂₀ N ₂ O ₅
3	[M] ⁺	629.3476	629.3486	1 hit = macrosepatrine - Not described in <i>G. laeve</i>	C ₄₀ H ₄₅ N ₄ O ₃ ⁺
4	[M+H] ⁺	327.1699	327.1703	23 hits	C ₁₉ H ₂₂ N ₂ O ₃
5	[M+H] ⁺	573.3588	573.3588	4 hits	C ₃₈ H ₄₄ N ₄ O
6	[M+H] ⁺	633.3795	633.3799	7 hits	C ₄₀ H ₄₈ N ₄ O ₃

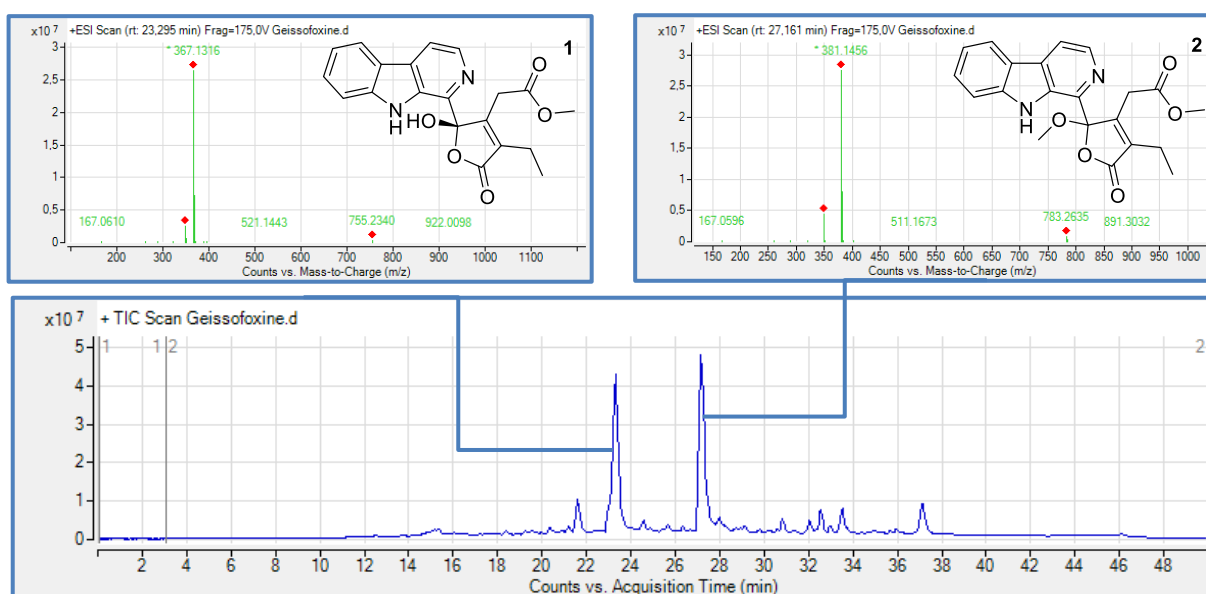
Figure S7. Generation of *O*-methylgeissolaevine (**2**) from geissolaevine (**1**)Figure S7. TIC-ESI(+) chromatogram of compound **1** after one night in MeOH and extracted MS spectra of compounds **1** and **2**.

Figure S8. Comparison between the principal NOESY correlations of 3',4',5',6'-tetrahydrogeissospermine (**3**) and geissospermine (**6**)

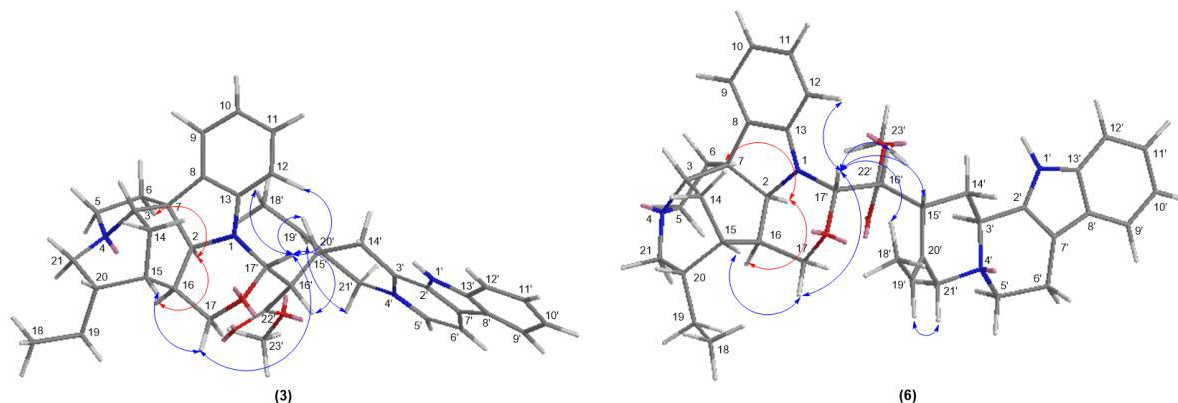


Figure S8. Principal NOESY correlations of 3',4',5',6'-tetrahydrogeissospermine (**3**) and geissospermine (**6**).

Figure S9. Comparison between the experimental ECD spectra of 3',4',5',6'-tetrahydrogeissospermine (**3**) and geissospermine (**6**)

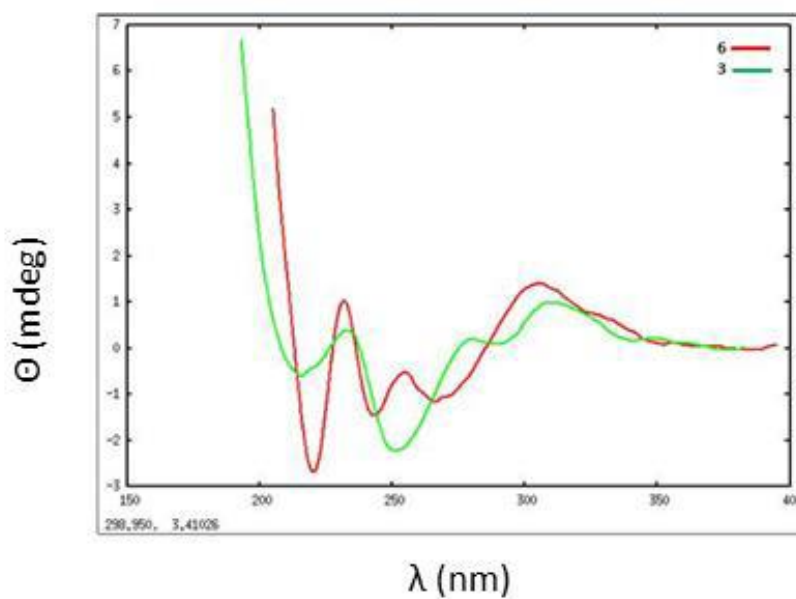


Figure S9. Comparison between the experimental ECD spectra of compounds **3** (green) and **6** (red).

Table S10 and Figure S11. Computational methods

Table S10. Conformations and corresponding relative free energies, populations and cumulated populations used for the simulation of circular dichroism spectrum for compound **1**.

Conformation	G^0_{rel} (kcal/mol)	P (%)	P_c (%)
1	0.00	40.0	40.0
2	0.11	33.1	73.1
3	0.98	7.7	80.8
4	1.13	6.0	86.7
5	1.14	5.9	92.6

Model coordinates

Compound **1** (E(B3LYP): -1258.63423932 Ha, 17.2363 cm^{-1})

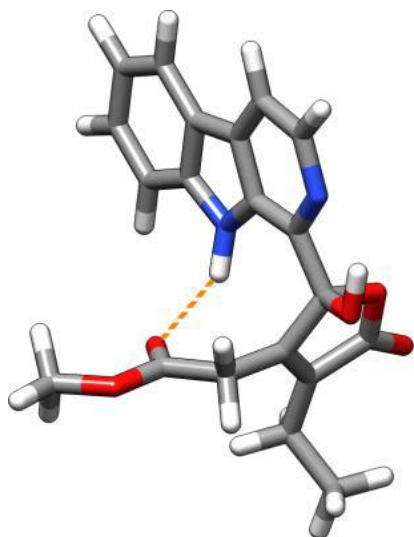


Figure S11. Geometry of the most stable conformation of compound **1** found at the B3LYP/6-31G* level (The orange dashed line represents an intramolecular hydrogen bond).



Coordinates, energies and lowest frequency of the most stable conformer found for compound 1:

Sum of electronic and thermal Free Energies: -1258.334998 Ha

1 O 2.06157 2.29927 1.46314
2 C 1.37090 1.60503 0.48968
3 C 1.99504 0.22598 0.33225
4 C 2.05217 -0.72512 1.49013
5 C 1.31890 -2.03952 1.25528
6 O 1.41865 -2.83196 2.33238
7 C 0.75817 -4.10964 2.23763
8 O 0.72610 -2.35760 0.24146
9 C 2.52041 0.10391 -0.89447
10 C 3.31489 -0.99573 -1.53083
11 C 4.80114 -0.62439 -1.69677
12 C 2.23911 1.36672 -1.64165
13 O 1.56173 2.22081 -0.82202
14 O 2.53082 1.63527 -2.78195
15 C -0.13915 1.68704 0.76989
16 N -0.47001 2.69002 1.58569
17 C -1.75970 2.97511 1.83258
18 C -2.81561 2.26665 1.27030
19 C -2.50404 1.19348 0.43202
20 C -1.13363 0.89664 0.18072
21 N -1.05849 -0.19372 -0.66730
22 C -2.34453 -0.64169 -0.93213
23 C -2.75558 -1.73677 -1.69842
24 C -4.12337 -1.95623 -1.82339
25 C -5.06611 -1.11056 -1.20594
26 C -4.65301 -0.02700 -0.44031
27 C -3.27967 0.21240 -0.29258
28 H 1.39680 2.92639 1.83683
29 H 1.64619 -0.25668 2.39335
30 H 3.09189 -0.97594 1.73731
31 H -0.31200 -3.97263 2.06517
32 H 1.18246 -4.69759 1.42005
33 H 0.93552 -4.59765 3.19569
34 H 2.88979 -1.19500 -2.52260
35 H 3.21344 -1.91785 -0.94991
36 H 4.90429 0.27204 -2.31458
37 H 5.34496 -1.44135 -2.18318
38 H 5.27297 -0.42999 -0.72707
39 H -1.94242 3.80987 2.50410
40 H -3.84304 2.53917 1.49185
41 H -0.23765 -0.78460 -0.75751
42 H -2.03227 -2.38904 -2.17840
43 H -4.47237 -2.79862 -2.41444
44 H -6.12613 -1.31064 -1.33165
45 H -5.38191 0.62179 0.03808



Experimental method S12. Biological assays

Antiplasmodial evaluation. The chloroquine-resistant strain FcB1/Colombia of *Plasmodium falciparum* was obtained from the National Museum Natural History collection, Paris, France (n°MNHN-CEU-224-PfFCB1). Parasites were maintained in vitro in human erythrocytes in RPMI 1640 medium supplemented by 8% (v/v) heat-inactivated human serum at 37 °C under an atmosphere of 3% CO₂, 6% O₂, and 91% N₂. Human red blood cells and serum were provided by the Etablissement Français du Sang Ile de France under the C-CPSL-UNT approval n°13/EFS/126. In vitro drug susceptibility was measured by [³H]-hypoxanthine incorporation as described previously.² Stock solutions of drugs were prepared in DMSO. Compounds were serially diluted two-fold with 100 µL culture medium in 96-well plates. Asynchronous parasite cultures (100 µL, 1% parasitemia and 1% final hematocrit) were then added to each well and incubated for 24 h at 37 °C prior to the addition of 0.5 µCi of [³H]-hypoxanthine (GE Healthcare; 1 to 5 Ci.mmol/mL) per well. After a further incubation of 24 h, plates were frozen and thawed. Cell lysates were then collected onto glass-fiber filters and counted in a liquid scintillation spectrometer. The growth inhibition for each drug concentration was determined by comparison of the radioactivity incorporated in the treated culture with that in the control culture maintained on the same plate. The concentration causing 50% growth inhibition (IC₅₀) was obtained from the drug concentration-response curve and the results were expressed as mean values ± standard deviations as determined from three independent experiments. Chloroquine diphosphate, used as a positive control, was purchased from Sigma (purity>99%).

Antileishmanial evaluation. The in vitro evaluation was performed on the *L. donovani* (MHOM/ET/67/HU3) line, called LV9, kindly supplied by Prof. S. Croft, London School of Hygiene and Tropical Medicine, UK. The evaluation on axenic amastigotes was adapted from Cheikh-Ali et al.³ Axenic amastigotes were induced by incubating a five-day late log-phase promastigote culture in M199 supplemented with 100 µM adenosine, 0.5 mg/L hemin, 40 mM HEPES, and 10% heat inactivated fetal bovine serum at pH 5.5 at 37 °C with 5% CO₂ for 24 h. Two-fold serial dilutions of the test extract and compounds were then performed in 100 µL of the same medium in 96-well microplates. Axenic amastigotes were then added to each well at 106/ml in 200 µL final volume. After 72 h of incubation at 37 °C with 5% CO₂, 20 µL of 450 µM resazurin was added to each well and further incubated in the dark for 24 h at 37°C with 5% CO₂. In living cells, resazurin is reduced to resofurin. This conversion was monitored by measuring the absorbance at specific wavelengths of resofurin (570 nm) and resazurin (600 nm) using a microplate reader (Lab systems Multiskan MS). Final activities were expressed as IC₅₀ values. Miltefosine was used as the reference compound.

Evaluation using intramacrophage amastigotes was adapted from the method of Cheikh-Ali et al.³ The cell line was RAW 264.7 macrophage (from ATCC), a mouse monocyte macrophage. Briefly, RAW 264.7 macrophages were plated in 96-well microplates at 2 × 10⁴ cells per well and incubated for 24 h at 37°C with 5% CO₂. Axenic amastigotes were induced as described above, centrifuged at 2000 g for 10 min, resuspended in DMEM complete medium, and added to each well at 3.2 × 10⁵ parasites per well to reach a parasite to macrophage ratio 16 :1.



After 24 h of infection at 37 °C with 5% CO₂, extracellular parasites were removed and 100 µL of DMEM complete medium containing two-fold serial dilutions of each test material was added to each well. After 48 h of treatment, the medium was removed and replaced by 100 µL of DirectPCR Lysis Reagent (Euromedex, Souffelweyersheim, France) before three freeze-thaw cycles at room temperature, addition of 50 µg/mL proteinase K, and a final incubation at 55 °C for 4 h to allow cell lysis. Then 10 µL of each cell extract were then added to 40 µL of DirectPCR Lysis reagent containing 0.05% Sybr Green I (Invitrogen, Saint-Aubin, France). DNA fluorescence was monitored using Mastercycler® realplex (Eppendorf, Montesson, France). The activity of the test materials was expressed as IC₅₀ values. Miltefosine was used as the reference compound.

Cell culture and proliferation assay. The MRC-5 cell line was obtained from the American Type Culture Collection (Rockville, MD, USA) and cultured according to the supplier's instructions. Human MRC-5 cells were grown in DMEM supplemented with 10% fetal calf serum (FCS) and 1% glutamine. The cell line was maintained at 37 °C in a humidified atmosphere containing 5% CO₂. Cell growth inhibition was determined by an MTS assay according to the manufacturer's instructions (Promega, Madison, WI, USA). Briefly, the cells were seeded in 96-well plates (2.5 × 10³ cells/well) containing 200 µL of growth medium. After 24 h of culture, the cells were treated with the test compounds at different final concentrations. After 72 h of incubation, 40 µL of resazurin was added for 2 h before recording absorbance at 490 nm with a spectrophotometric plate reader. The IC₅₀ value corresponded to the concentration of compound inducing a decrease of 50% in absorbance of drug-treated cells compared with untreated cells. Experiments were performed in triplicate. Paclitaxel was used as the reference compound.



Figure S13. ¹H NMR spectrum (400 MHz, DMSO-*d*₆) of geissolaevine (1)

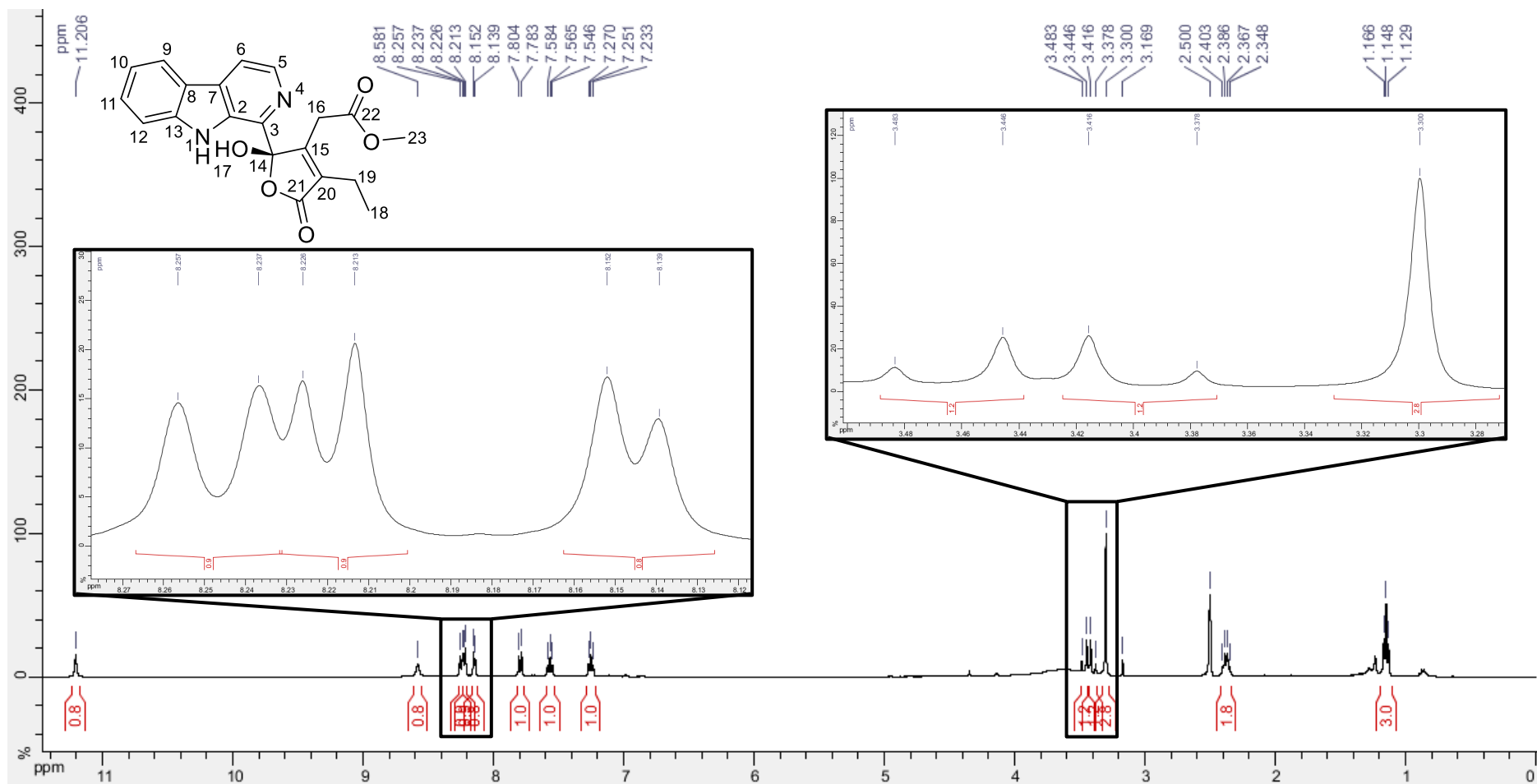


Figure S14. ¹³C NMR spectrum (100 MHz, DMSO-*d*₆) of geissolaevine (1)

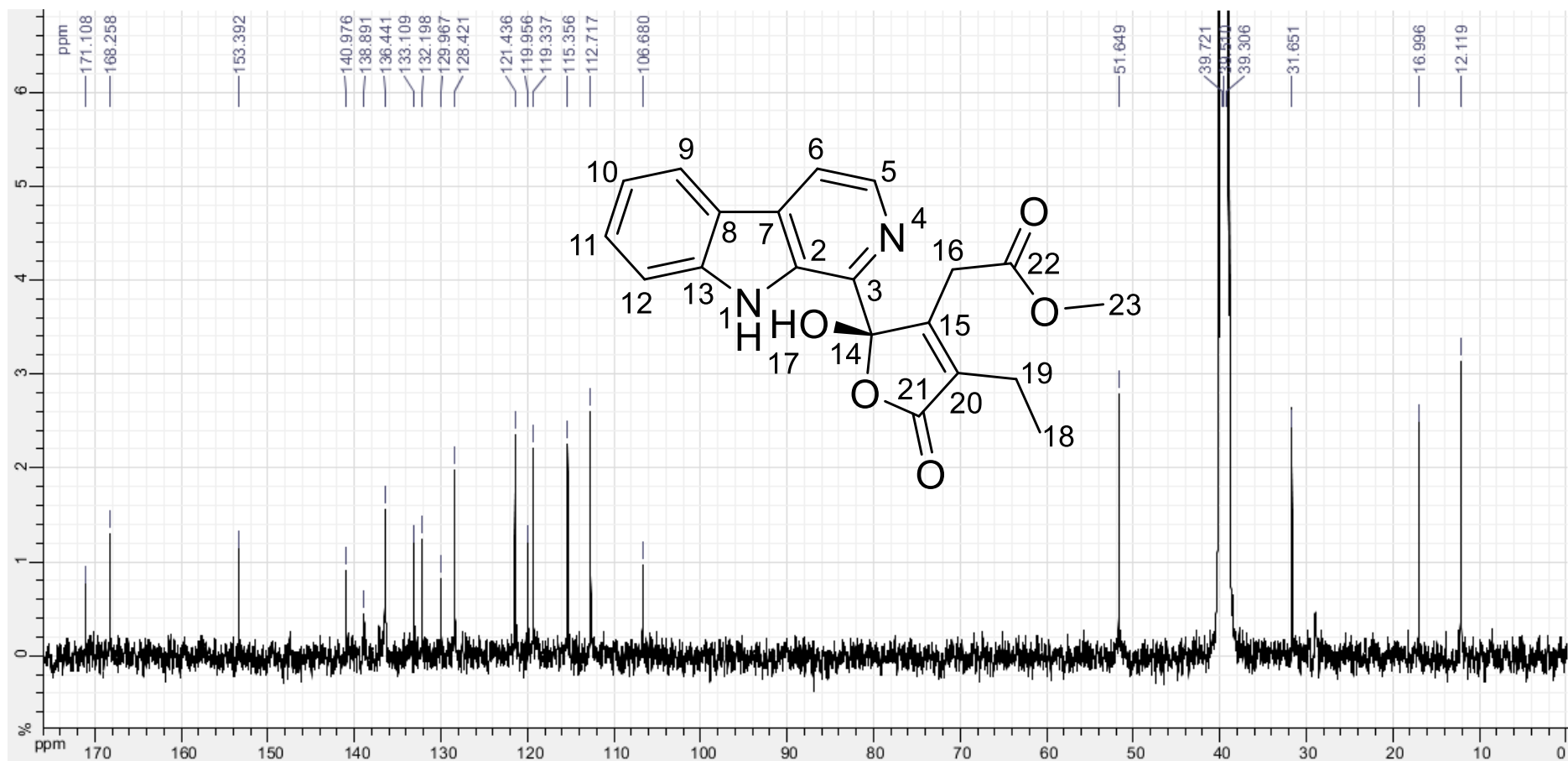


Figure S15. ¹³C J-MOD NMR spectrum (100 MHz, DMSO-d₆) of geissolaevine (1)

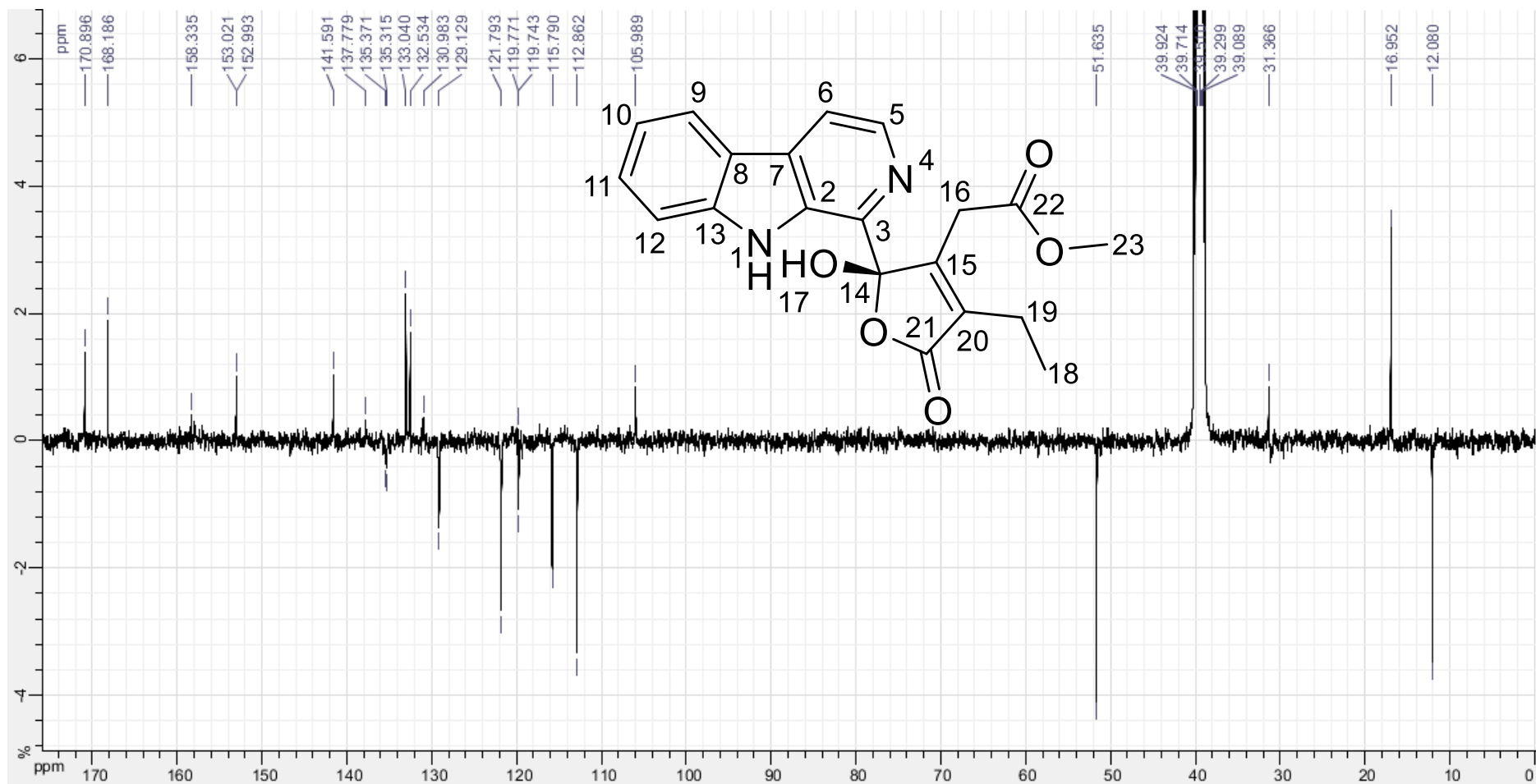


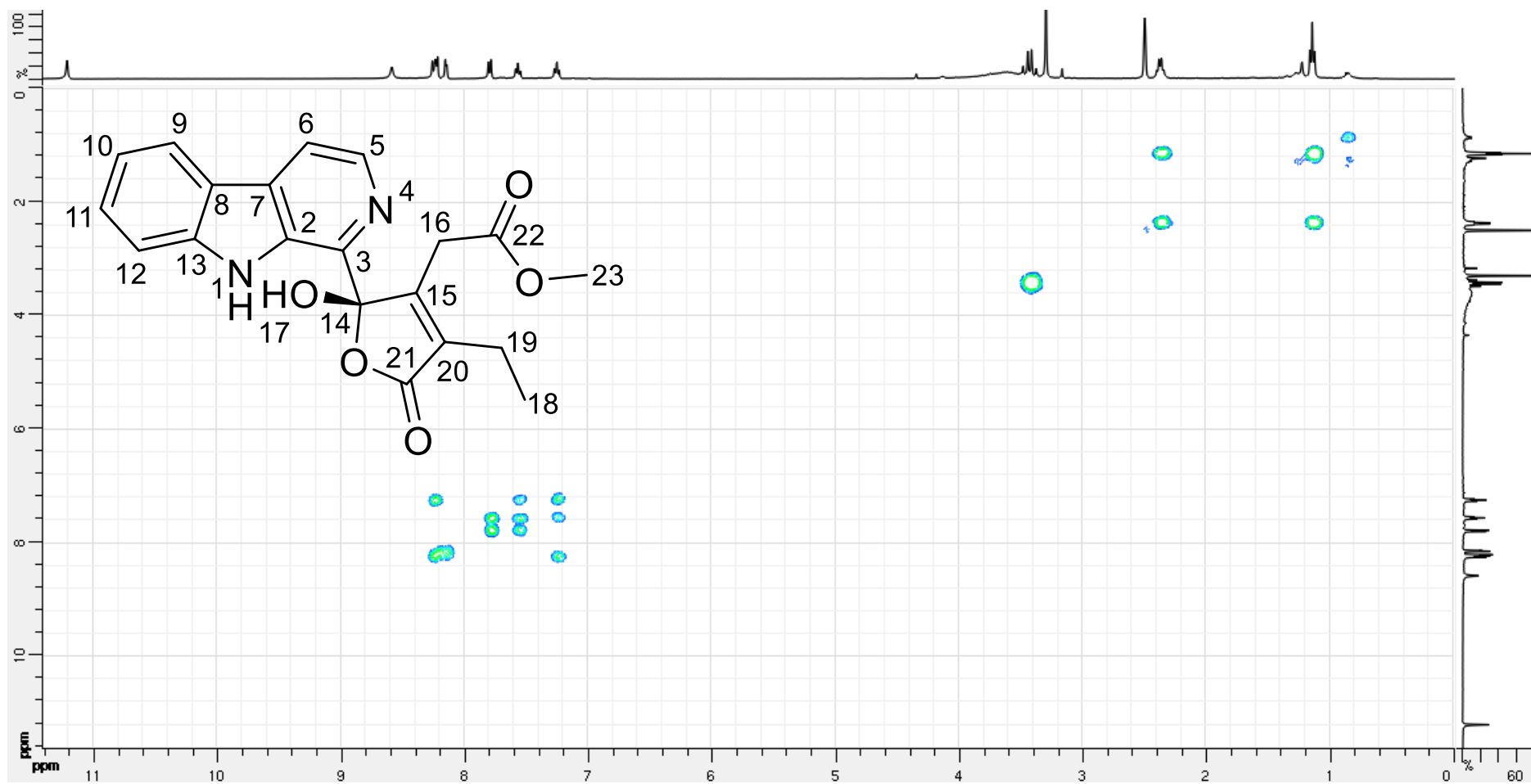
Figure S16. COSY spectrum (400 MHz, DMSO-*d*₆) of geissolaevine (1)

Figure S17. HSQC spectrum (400 MHz, DMSO-*d*₆) of geissolaevine (1)

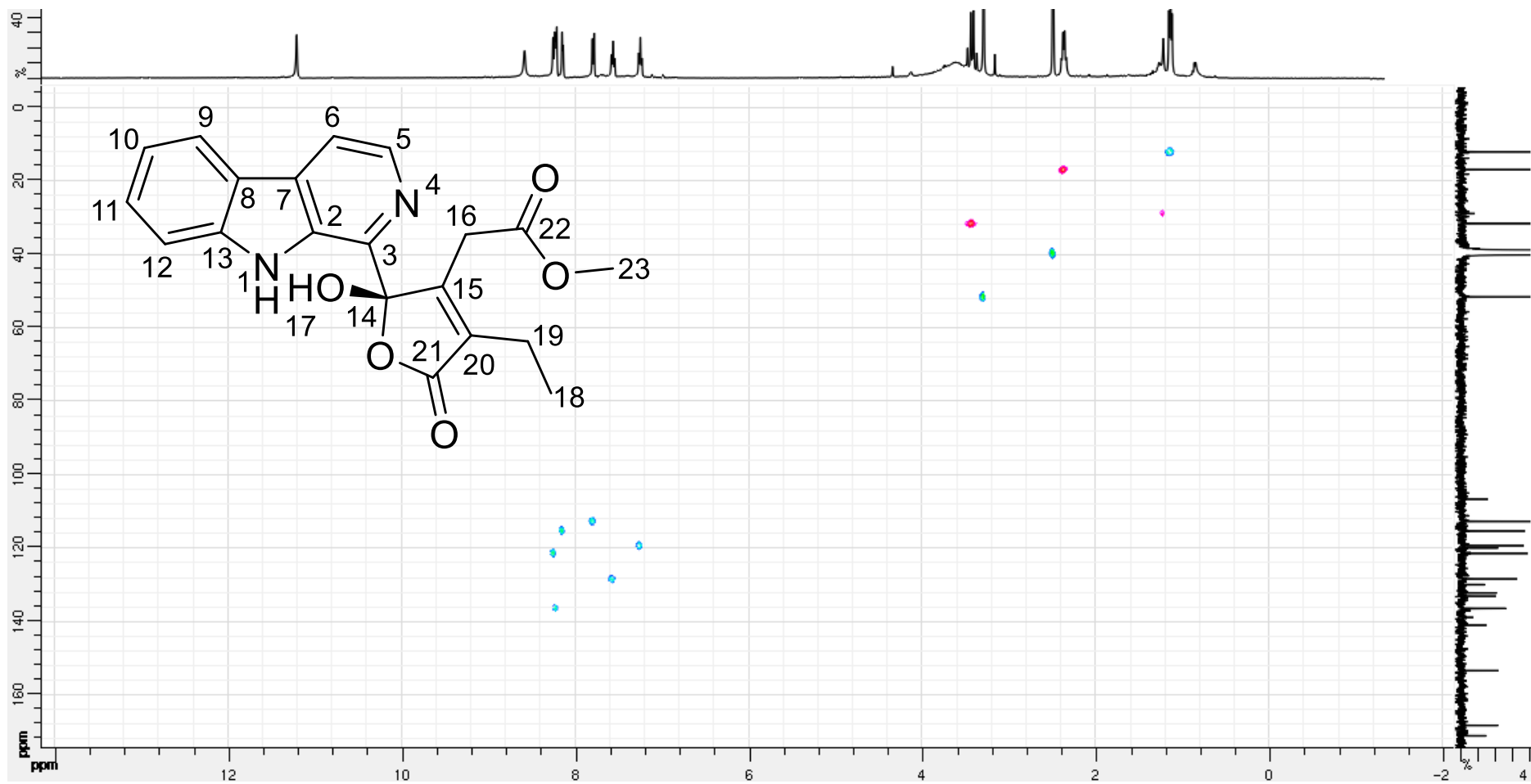


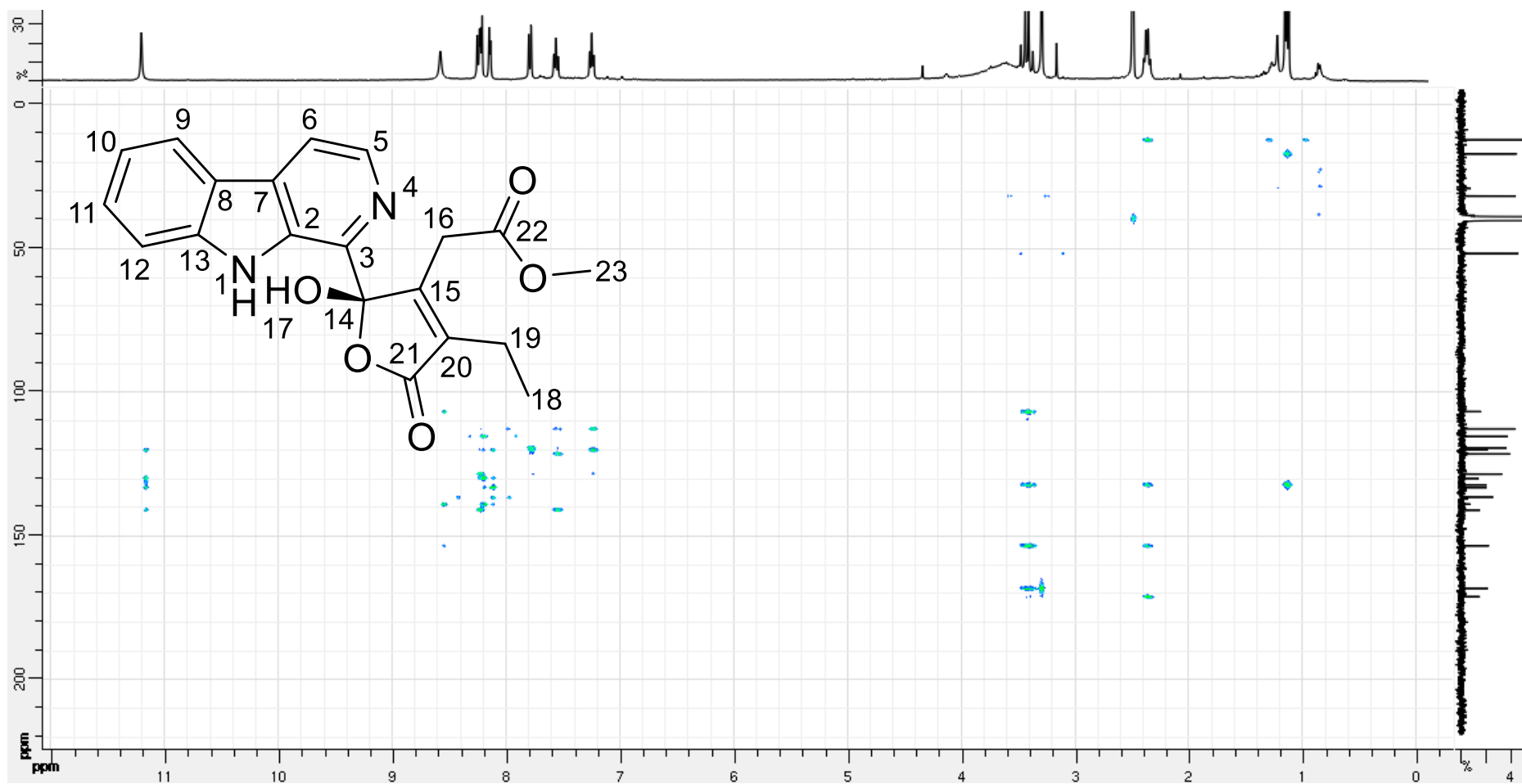
Figure S18. HMBC spectrum (400 MHz, DMSO- d_6) of geissolaevine (1)

Figure S19. NOESY spectrum (400 MHz, DMSO-*d*₆) of geissolaevine (1)

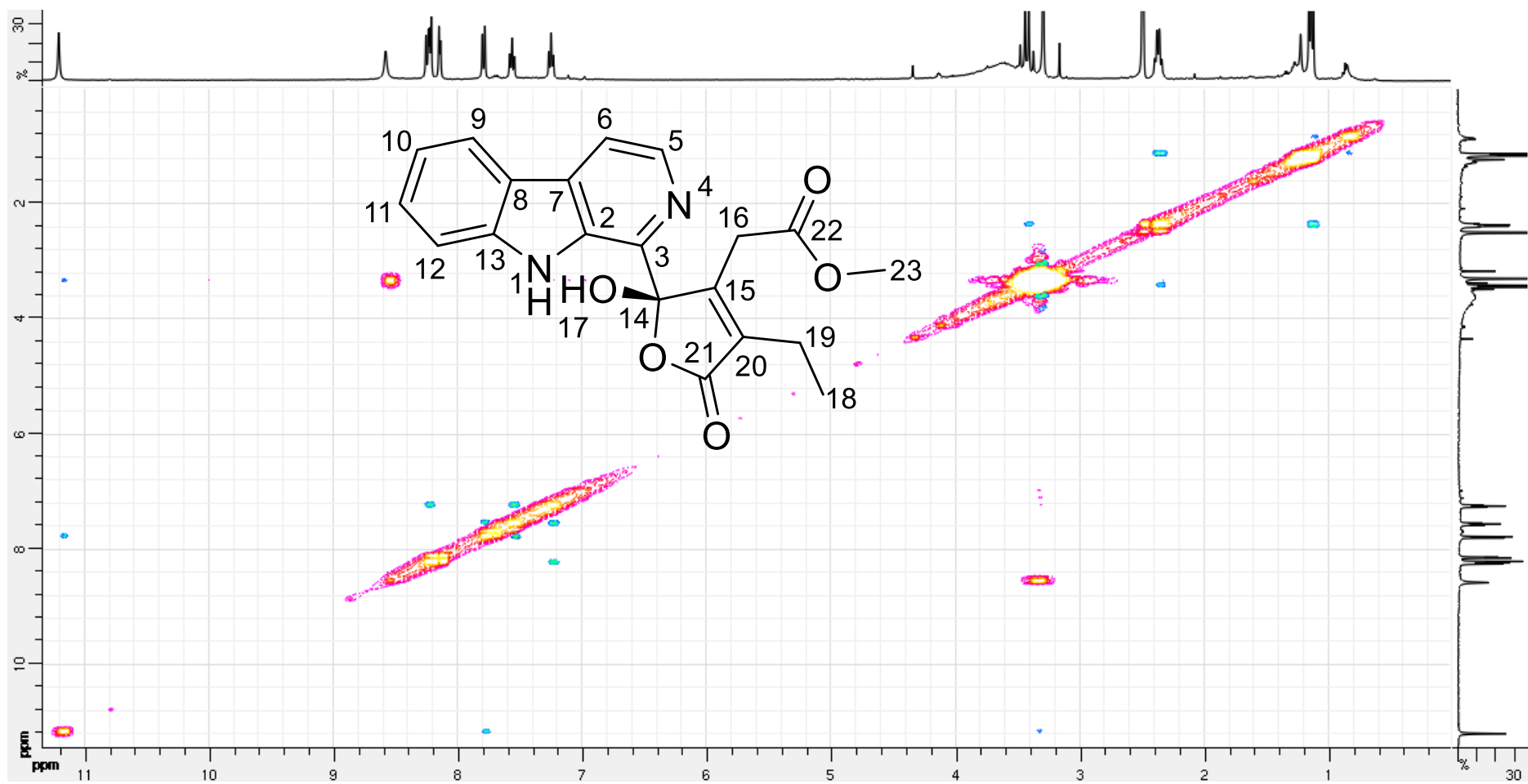


Figure S20. ¹H NMR spectrum (600 MHz, DMSO-d₆) of *O*-methylgeissolaevine (2)

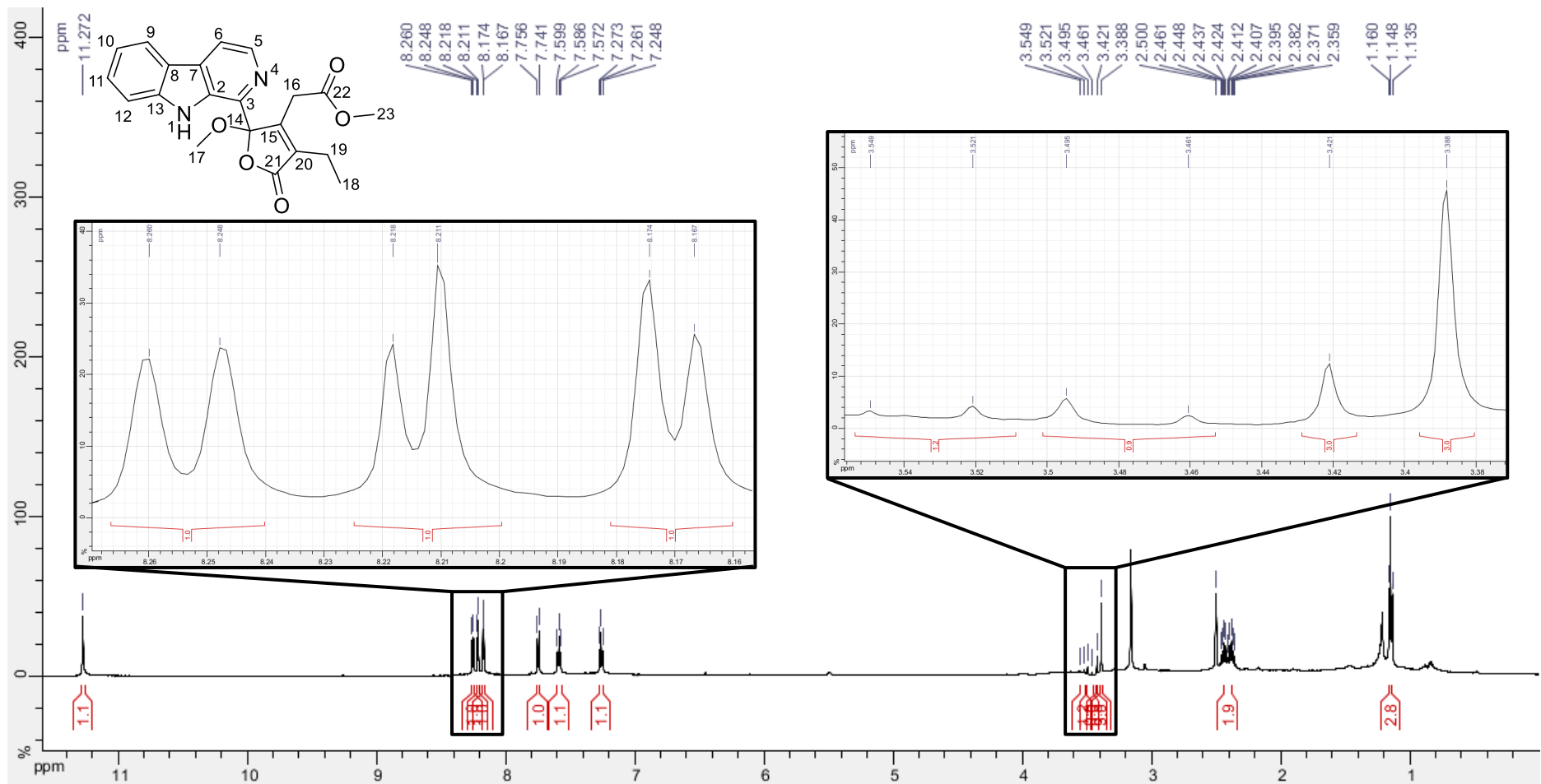


Figure S21. ¹³C NMR spectrum (100 MHz, DMSO-d₆) of *O*-methylgeissolaevine (2)

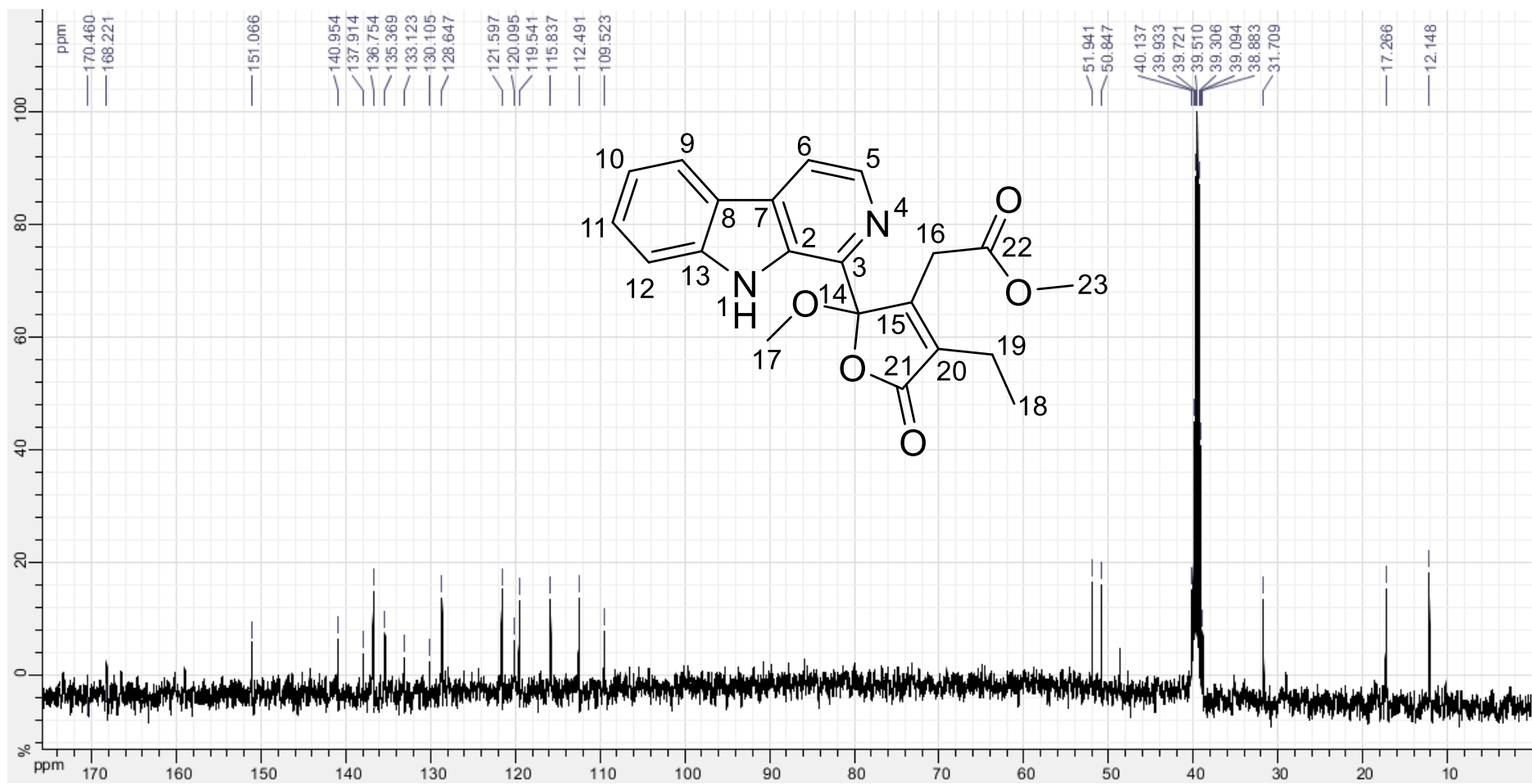


Figure S22. COSY spectrum (400 MHz, DMSO-*d*₆) of *O*-methylgeissolaevine (2)

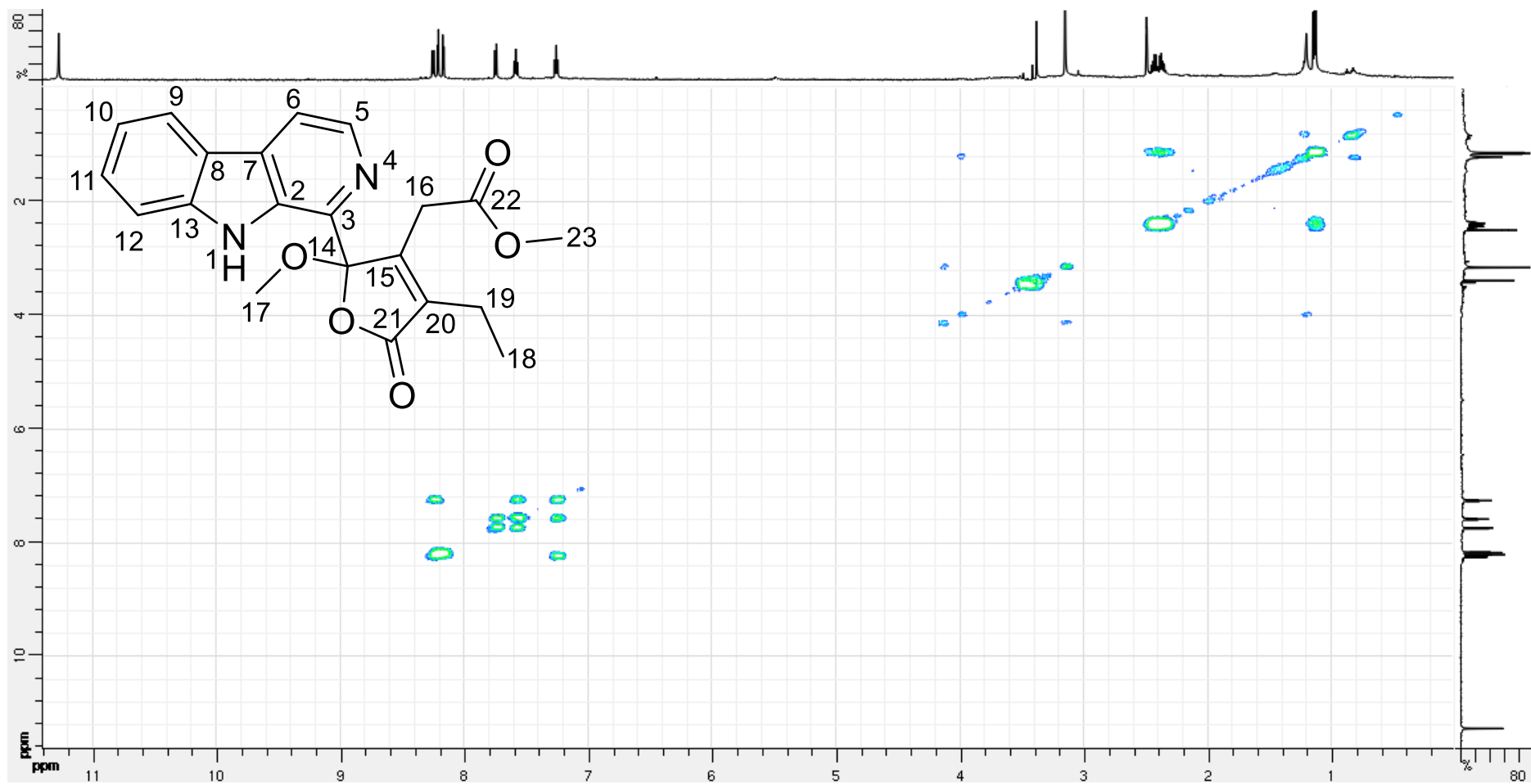


Figure S23. HSQC spectrum (400 MHz, DMSO-*d*₆) of *O*-methylgeissolaevine (2)

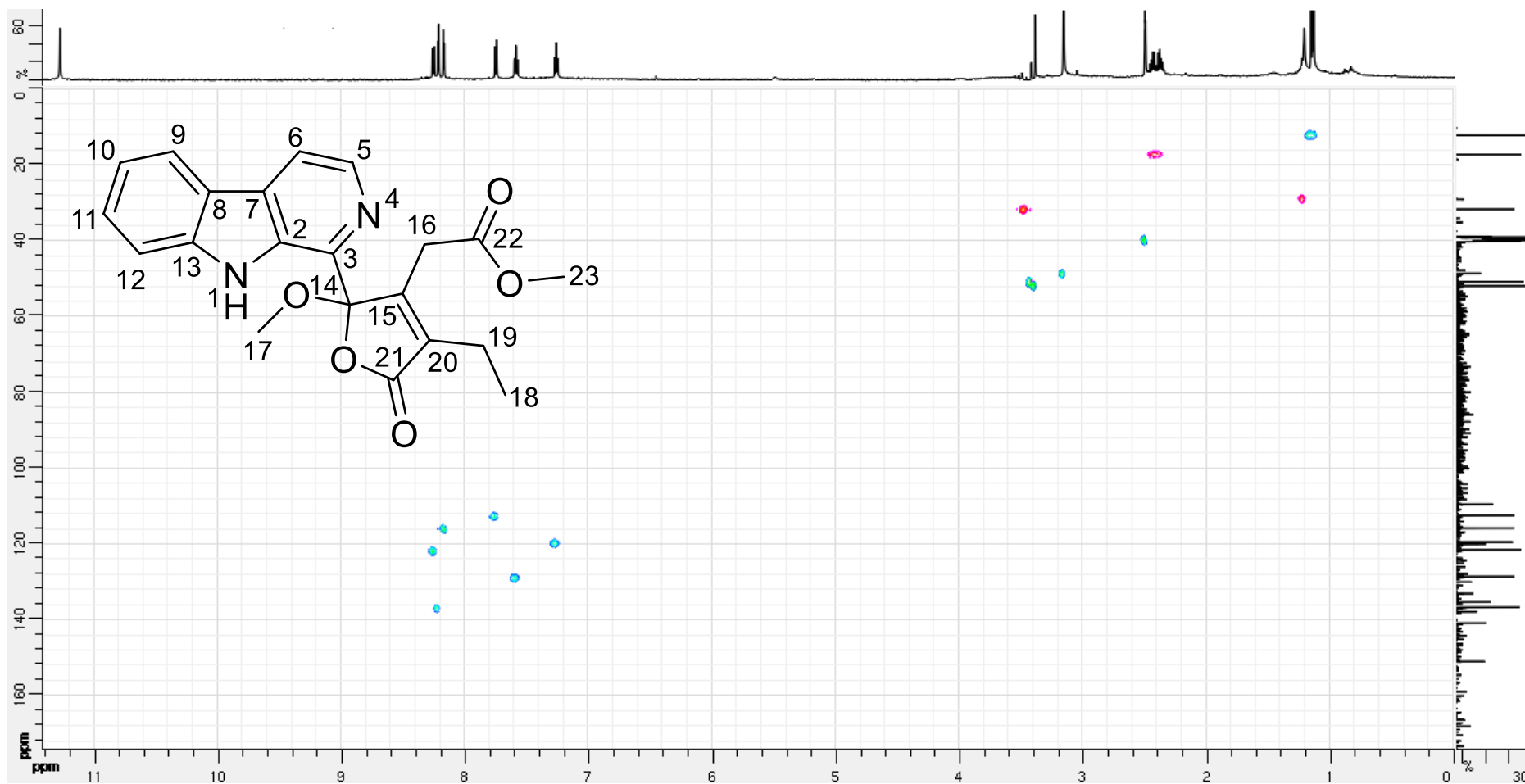


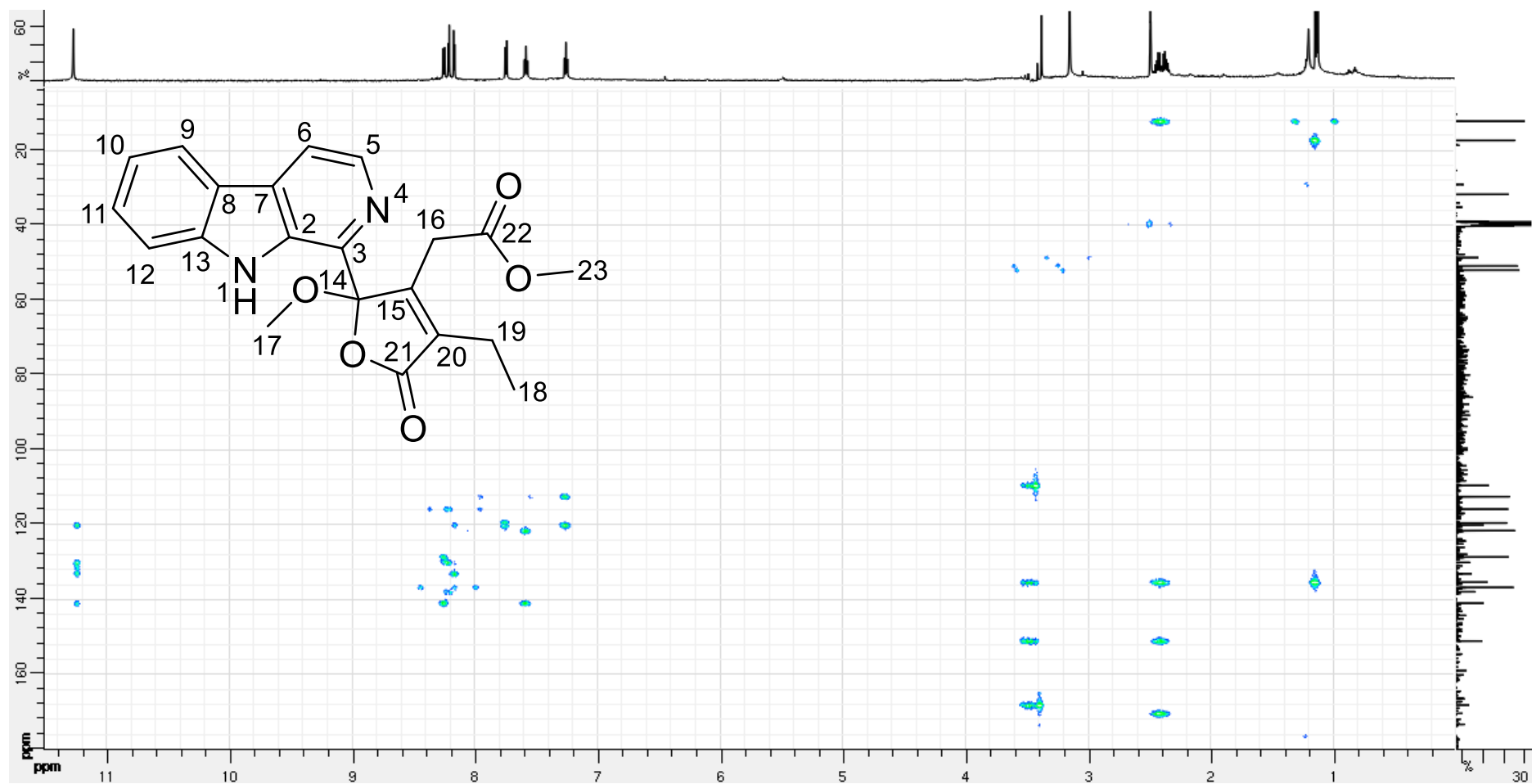
Figure S24. HMBC spectrum (400 MHz, DMSO- d_6) of *O*-methylgeissolaevine (2)

Figure S25. NOESY spectrum (400 MHz, DMSO-*d*₆) of *O*-methylgeissolaevine (2)

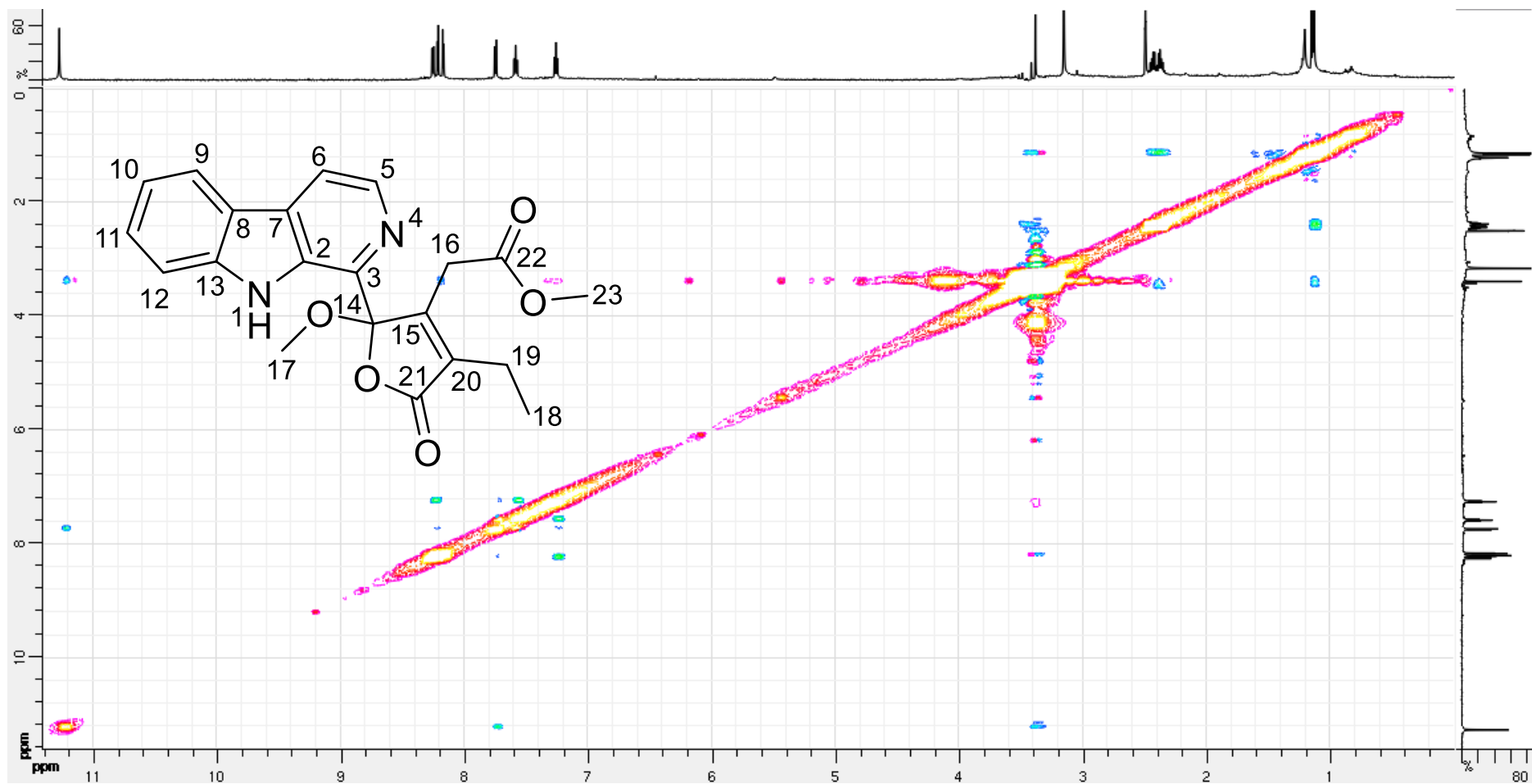


Figure S26. ¹H NMR spectrum (400 MHz, DMSO-d₆) of 3',4',5',6'-tetrahydrogeissospermine (3)

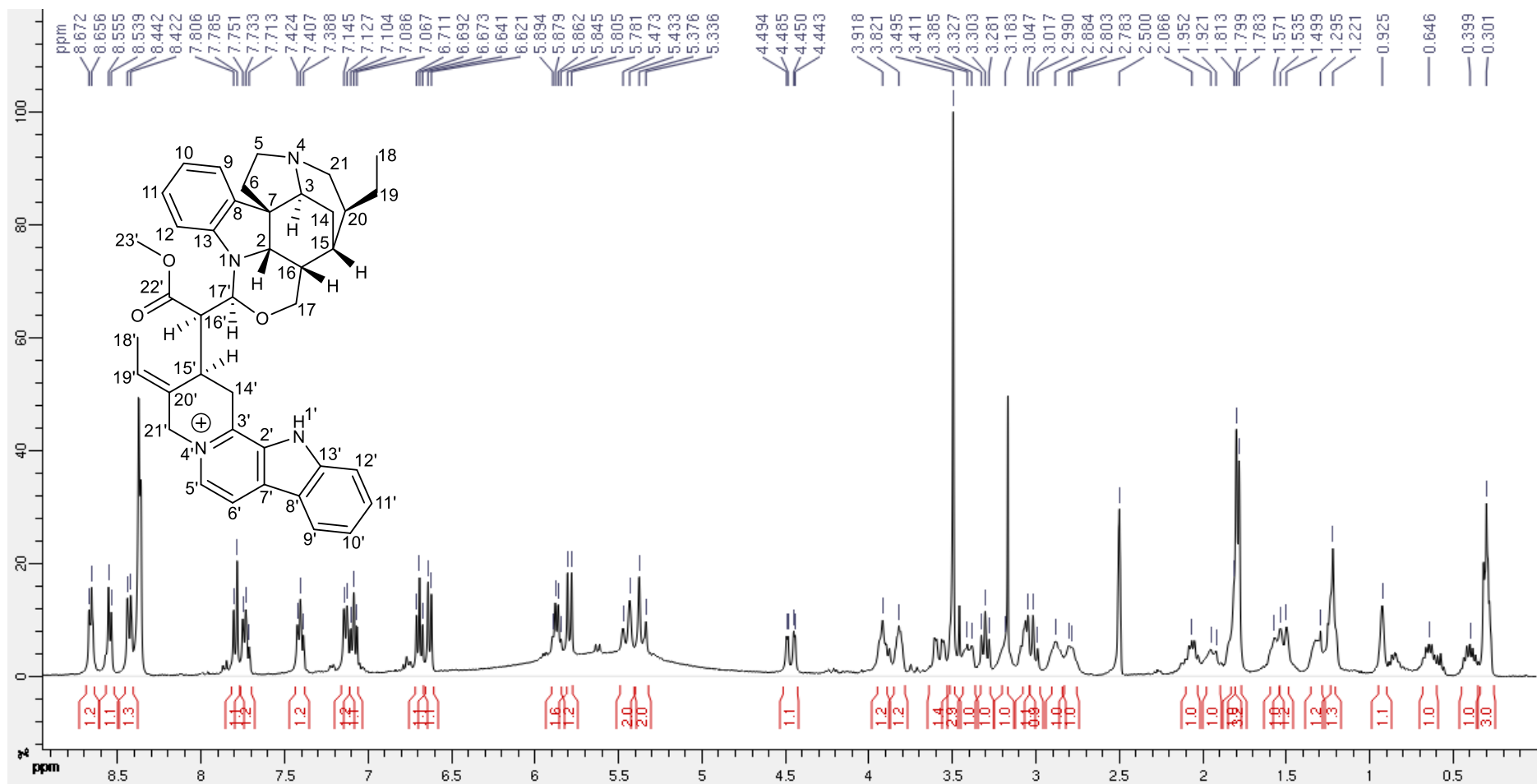


Figure S27. ¹³C NMR spectrum (100 MHz, DMSO-d₆) of 3',4',5',6'-tetrahydrogeissospermine (3)

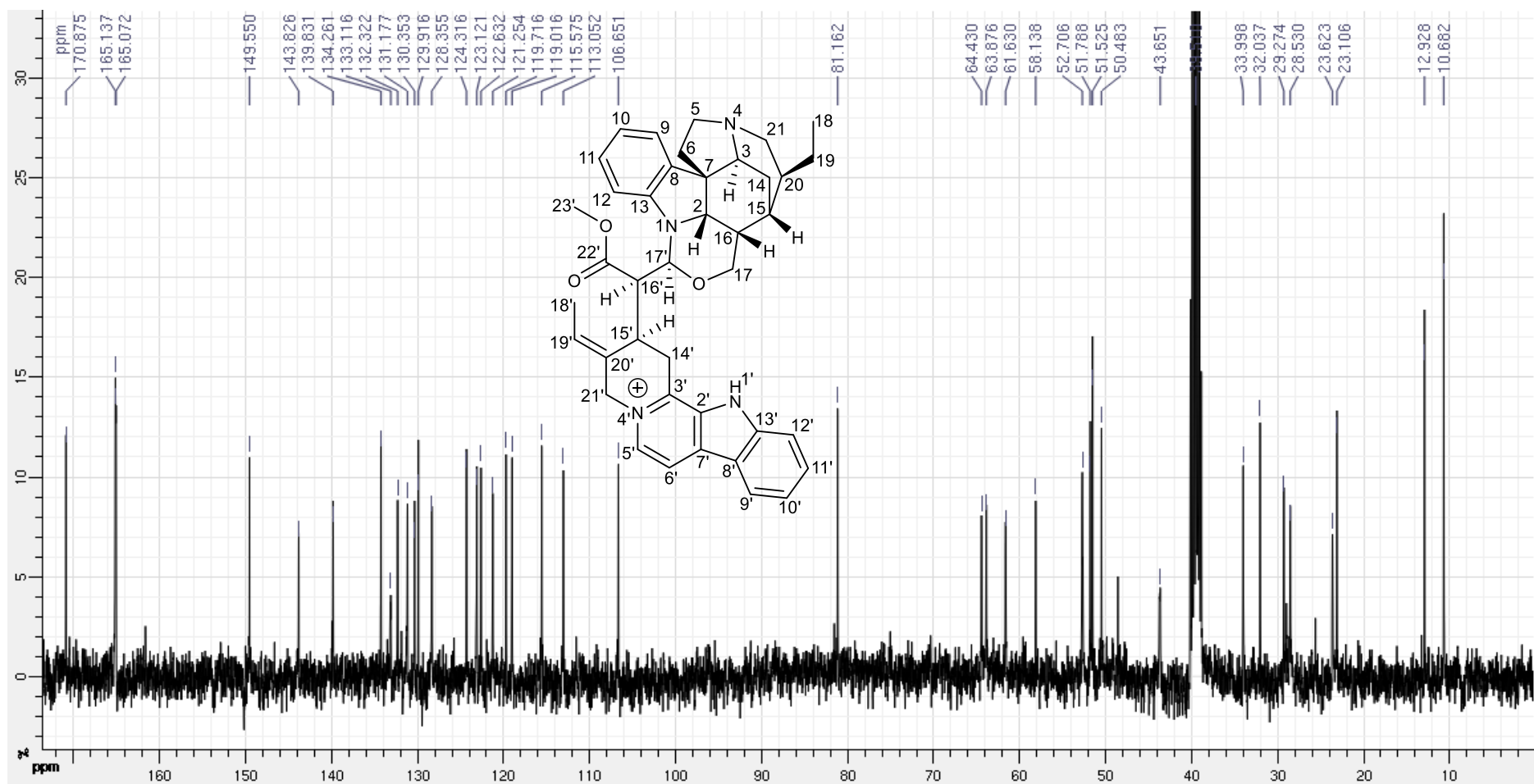


Figure S28. COSY spectrum (400 MHz, DMSO-*d*₆) of 3',4',5',6'-tetrahydrogeissospermine (3)

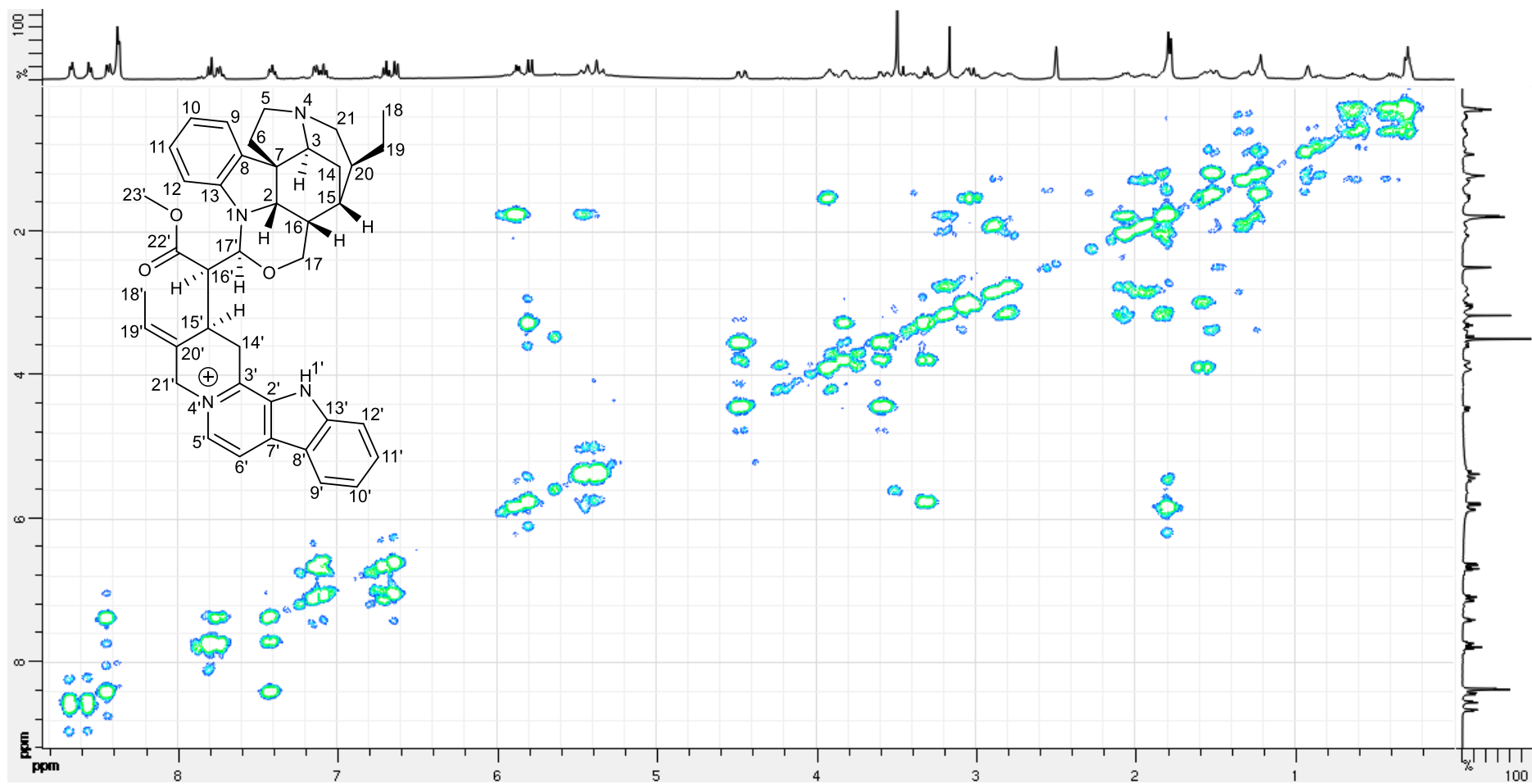


Figure S29. HSQC spectrum (400 MHz, DMSO-*d*₆) of 3',4',5',6'-tetrahydrogeissospermine (3)

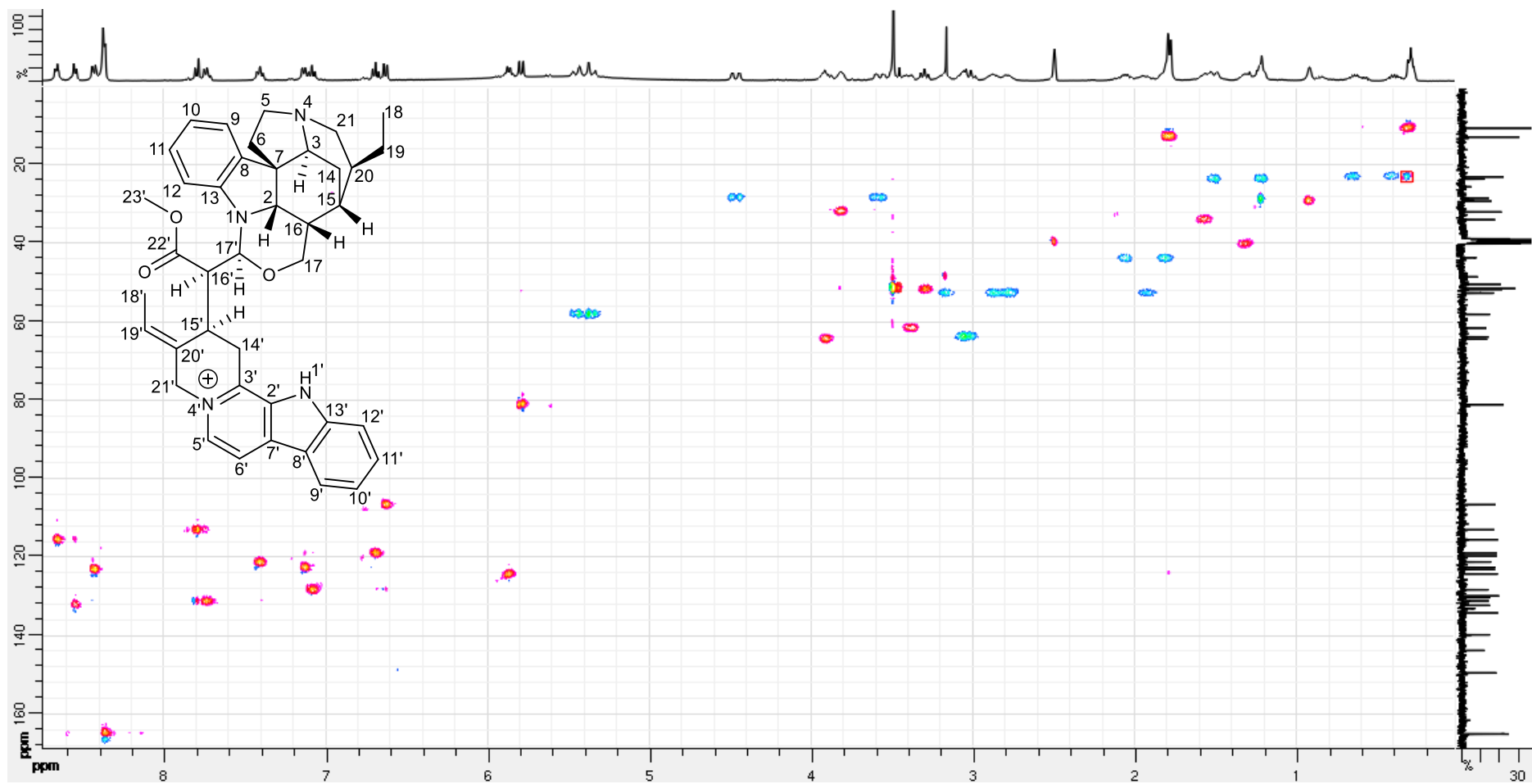


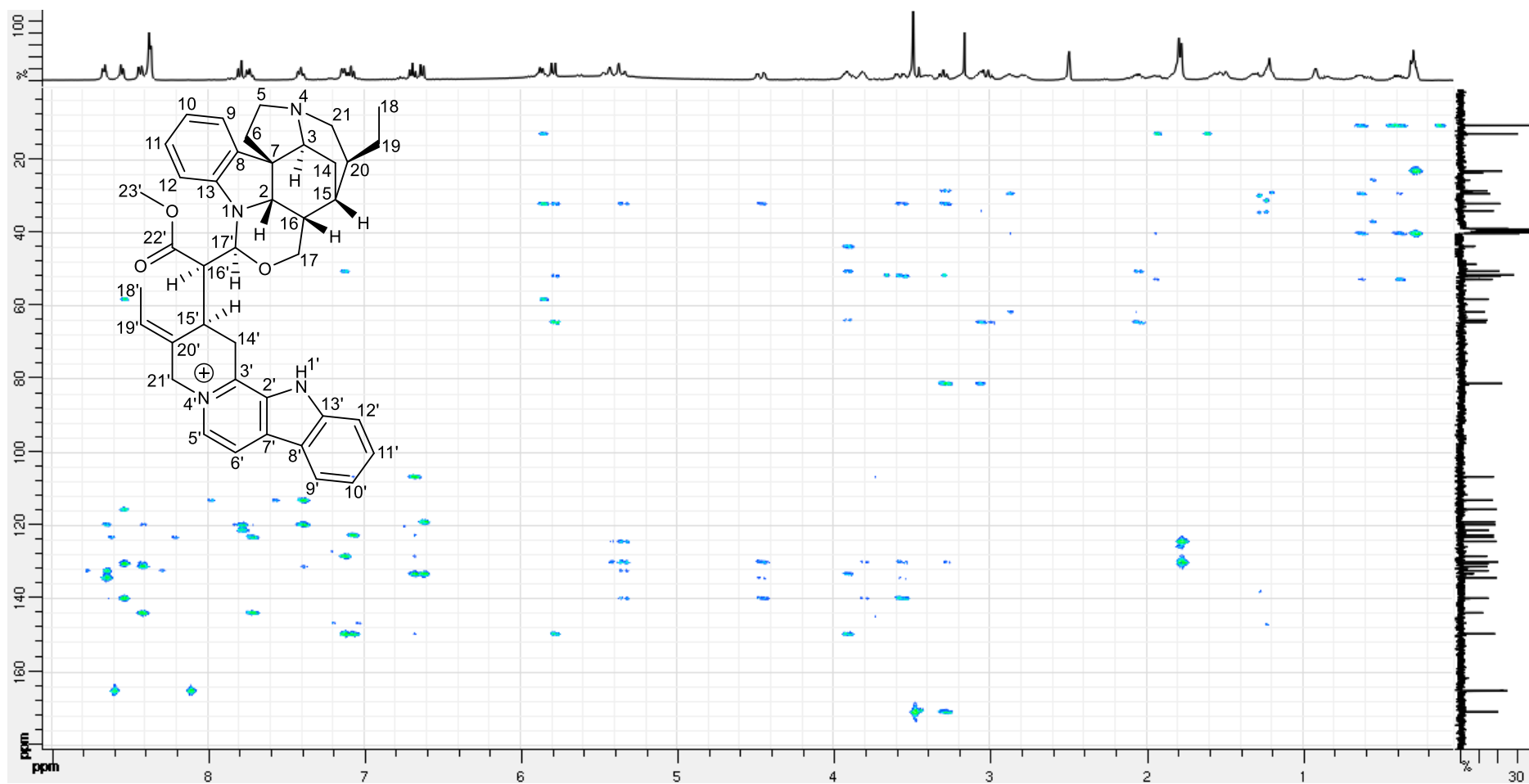
Figure S30. HMBC spectrum (400 MHz, DMSO- d_6) of 3',4',5',6'-tetrahydrogeissospermine (3)

Figure S31. NOESY spectrum (400 MHz, DMSO-*d*₆) of 3',4',5',6'-tetrahydrogeissospermine (3)

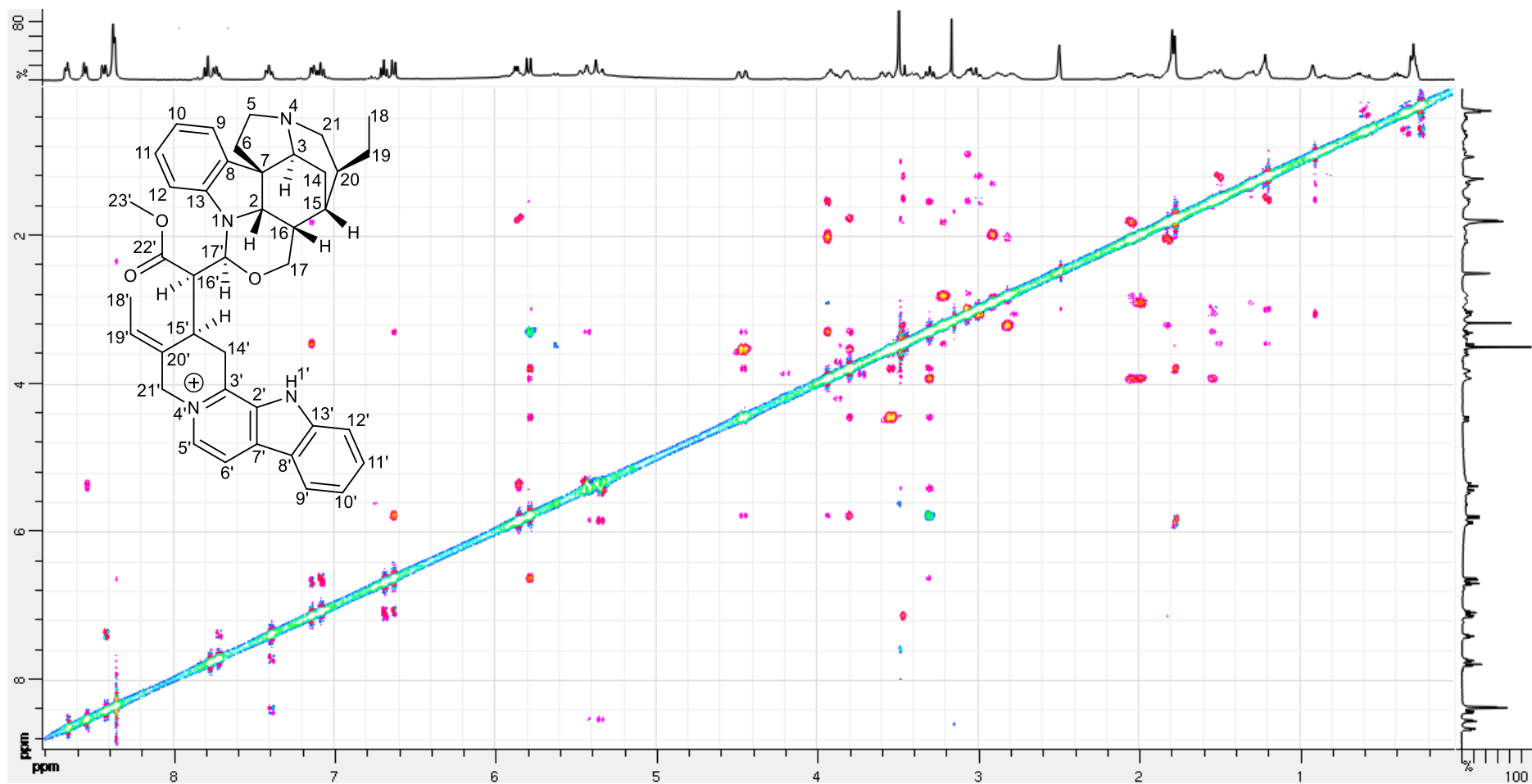


Figure S32. ¹H NMR spectrum (400 MHz, DMSO-d₆) of geissospermine (6)

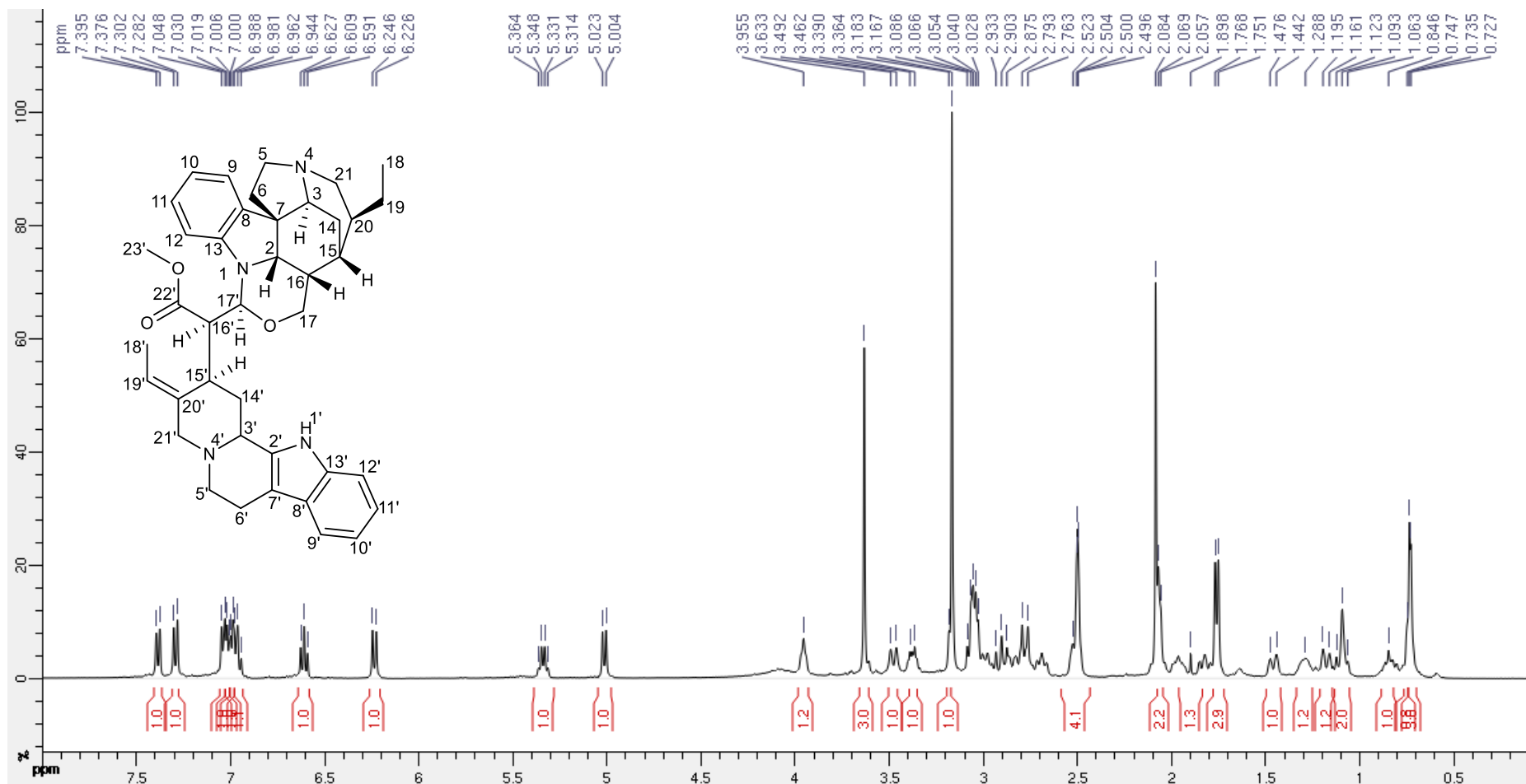
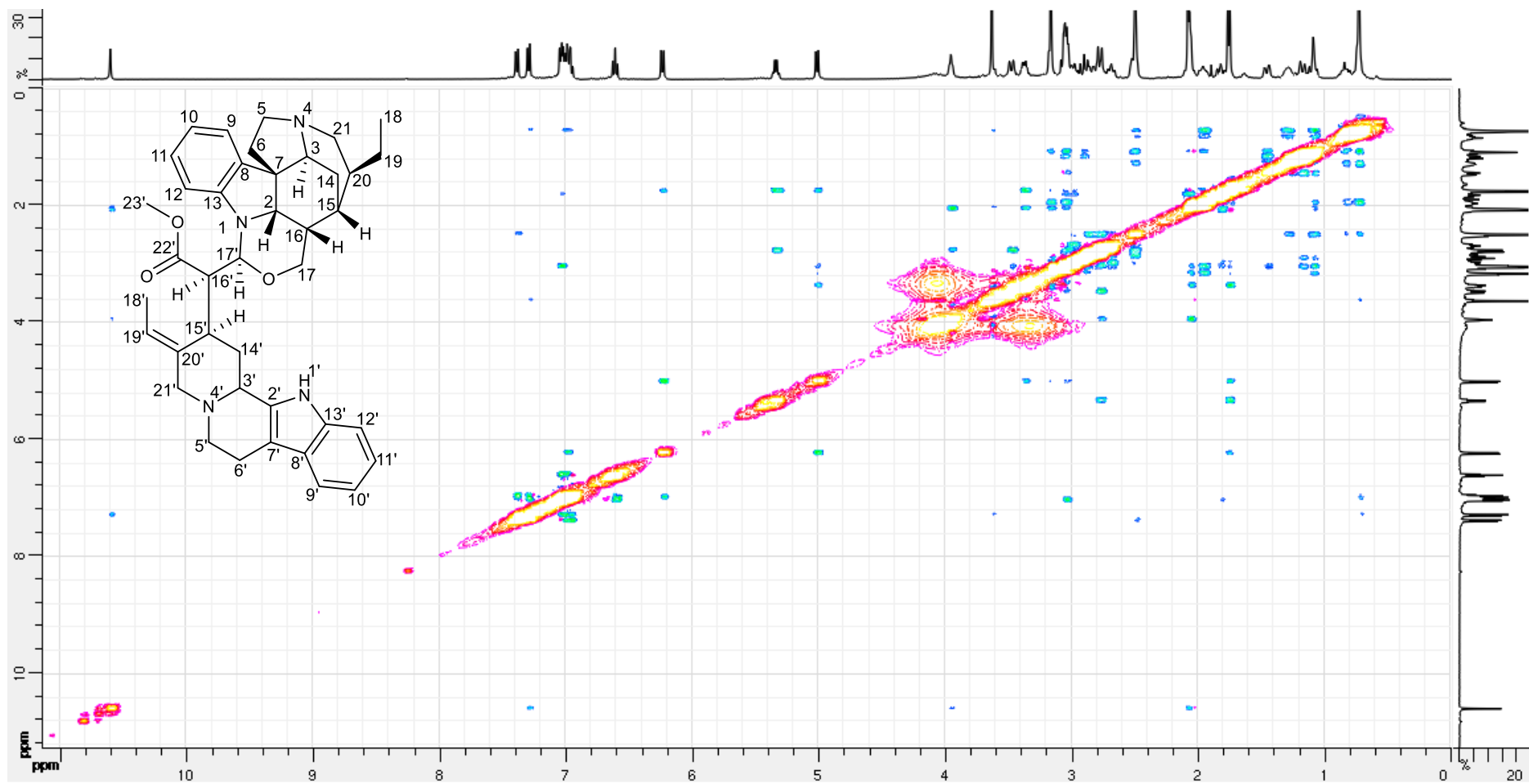


Figure S33. NOESY spectrum (400 MHz, DMSO-*d*₆) of geissospermine (6)



References:

(1) Buckingham, J.; Baggaley, K. H.; Roberts, A. D.; Szabó, L. F. Dictionary of Alkaloids, with CD-ROM; CRC Press/ Taylor & Francis Group: New York, **2010**; p 59.

(2) Guillon, J.; Grellier, P.; Labaied, M.; Sonnet, P.; Léger, J.-M.; Déprez-Poulain, R.; Forfar-Bares, I.; Dallemagne, P.; Lemaître, N.; Péhourcq, F.; Rochette, J.; Sergheraert, C.; Jarry, C. J. *Med. Chem.* **2004**, *47*, 1997-2009.

(3) Cheikh-Ali, Z.; Caron, J.; Cojean, S.; Bories, C.; Couvreur, P.; Loiseau, P. M.; Desmaële, D.; Poupon, E.; Champy, P. *ChemMedChem* **2015**, *10*, 411-418.



3. RÉSULTATS OBTENUS APRÈS PUBLICATION

L'exploration d'une dernière région du réseau moléculaire total a été entreprise par notre équipe, du fait de l'absence de match avec la MIADB (Région D - **Figure 11**).

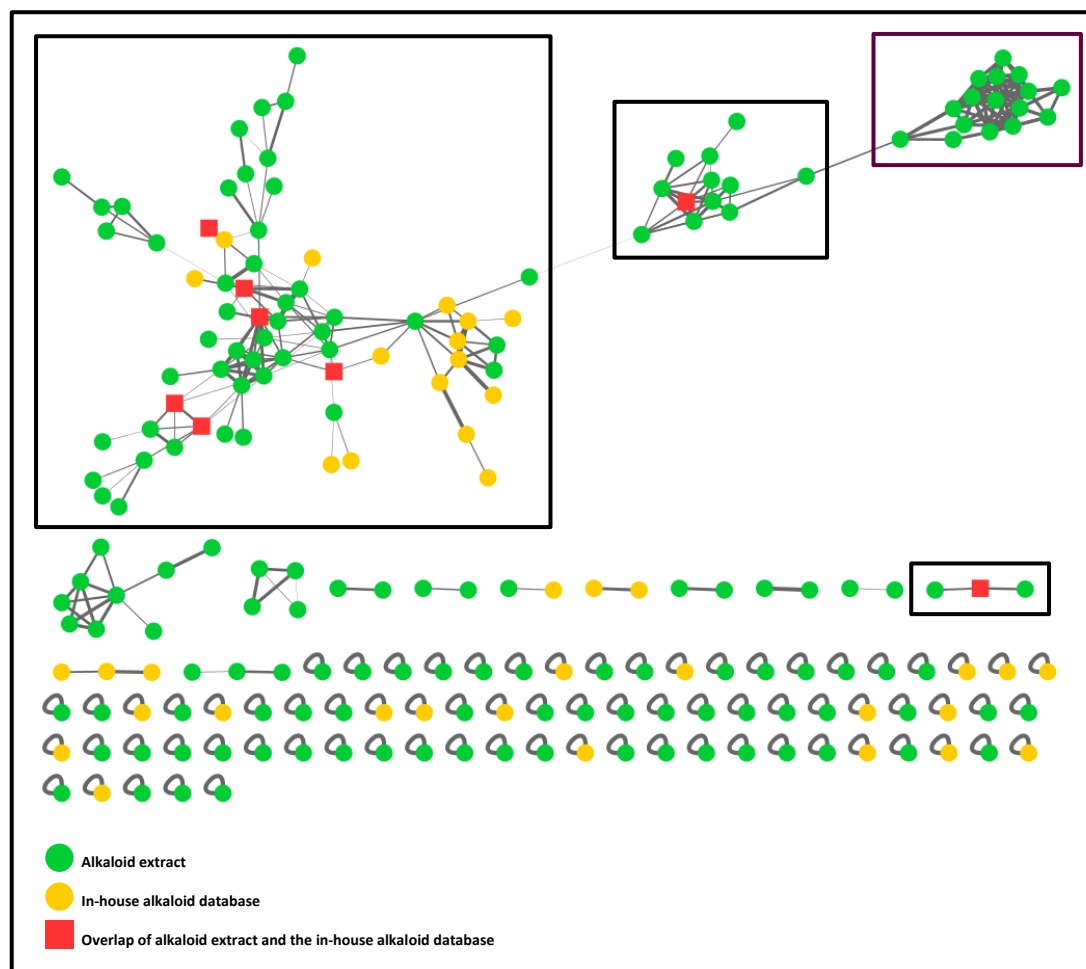


Figure 11. Position de la région D dans le réseau moléculaire total

Parmi les nœuds appartenant à cette région, seulement deux composés étaient présents en quantité suffisante pour être isolés et caractérisés structurellement. Les formules brutes des nœuds correspondant, à m/z mesurés 491,2444 et 595,2914 ont été générées en utilisant le logiciel MassHunter® puis recherchées sur des bases de données de produits naturels (DNP et Reaxys®). Nous avons donc obtenu les résultats suivants :

- m/z 491.2444 ($m/z_{\text{théorique}}$: 491.2442, erreur : 0.41 ppm) – $[M+H]^+$ pour $C_{31}H_{30}N_4O_2$: Cette formule brute ne correspondait à aucun composé déjà décrit dans la littérature.

- m/z 595.2914 ($m/z_{\text{théorique}}$: 595.2915, erreur : 0.17 ppm) – $[M+H]^+$ pour $C_{35}H_{38}N_4O_5$: Une seule proposition a été retrouvée dans le DNP : le bactériophaerophorbide-e, un produit du métabolisme de la chlorophylle de Chlorobiaceae (bactéries sulfureuses vertes).⁷⁵ Pour des raisons chimiotaxonomiques, cette proposition ne semblait pas correcte.

Par conséquent, ces composés ont été ciblés pour suivre le processus d'isolement. Malheureusement, ces molécules se sont avérées très instables à l'issu de chaque essai de purification. C'est ainsi que leur élucidation structurale n'a pas pu être réalisée.

⁷⁵ R. Baltenweck-Guyot & R. Ocampo, *Org. Geochem.*, 2007, **38**, 1580-1584.



CHAPITRE IV :

DEUXIÈME APPLICATION DE LA MIADB - ÉTUDE PHYTOCHIMIQUE D'*ALSTONIA* *BALANSAE* GUILLAUMIN





CHAPITRE IV : DEUXIÈME APPLICATION DE LA MIADB – EXPLORATION DE L’ESPACE CHIMIQUE D’*ALSTONIA BALANSAE*

1. INTRODUCTION

1.1 Phylogénie et description botanique

Le genre *Alstonia* R.Br. appartient aussi à la famille des Apocynaceae³¹, sous-famille des *Rauvolfioideae* Kostel., tribu des *Alstonieae* G.Don.⁵³ Ce genre a été décrit pour la première fois par Robert Brown dans les années 1810, et a reçu ce nom en hommage au professeur Charles Alston de l’Université d’Écosse.⁷⁶ À l’époque seulement quatre espèces étaient décrites dans le genre. À l’heure actuelle, il regroupe un total de 43 espèces acceptées, avec 104 synonymes décrits dans la littérature.⁷⁷ Les espèces du genre *Alstonia* sont, pour la plupart, des arbres avec une taille pouvant atteindre 60 m et un diamètre maximal de 200 cm. D’autre part, ils existent aussi sous la forme d’arbustes. Le genre produit du latex, comme la majorité des représentants de la famille. Les feuilles des *Alstonia* spp. ont des dispositions, des formes et des tailles très variées. Les fleurs du genre sont petites et fines pouvant atteindre une taille maximale de 40 mm dans le bourgeon mature, et un diamètre maximal de 2,5 mm. Le tube formant la corolle est court, pubescent sous les étamines, glabres au-dessus. Les fleurs sont réparties en cymes thyrsoides. Elles sont parfumées et ouvertes pendant la journée (Figure 12).^{78, 79, 80, 81} L’ovaire est généralement formé de deux loges séparées. Les fruits sont des follicules appariés, sauf chez *A. rostrata* C.E.C.Fisch. où ils sont unis et apparaissent donc comme un seul follicule.⁸¹ Les graines sont oblongues. Une clé de détermination a notamment été proposée par Monachino en 1949.⁷⁶

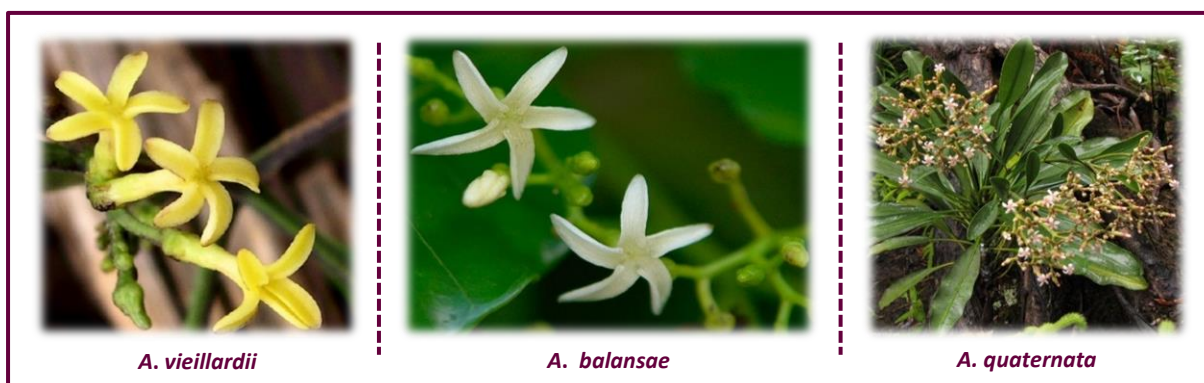


Figure 12. Quelques espèces du genre *Alstonia*

⁷⁶ J. Monachino, *Pac. Sci.*, 1949, **3**, 133-182.

⁷⁷ <http://www.theplantlist.org/tpl1.1/search?q=alstonia>, (accessed 19-09-2018, 2018).

⁷⁸ <http://endemia.nc/flore/fiche1665>, (accessed 18-09-2018).

⁷⁹ <http://endemia.nc/flore/fiche3284>, (accessed 20-09-2018).

⁸⁰ <http://endemia.nc/flore/fiche696>, (accessed 20-09-2018).

⁸¹ K. Sidiyasa, *Blumea, Suppl.*, 1998, **11**, 1-230.

1.2 Répartition géographique

Les représentants du genre *Alstonia* sont répartis dans les régions tropicales et subtropicales de l'Afrique, de l'Asie, de l'Amérique du Sud et Centrale mais surtout de la Polynésie et de l'Australie (**Figure 13**).

82



Figure 13. Répartition géographique des espèces du genre *Alstonia*

Parmi les espèces du genre, quelques-unes sont endémiques de certains pays. En effet, des plantes comme *A. quaternata* Van Heurck & Müll.Arg., *A. lanceolifera* S.Moore ou *A. coriacea* Pancher ex S.Moore ne peuvent être retrouvées qu'en Nouvelle-Calédonie ; *A. venenata* R. Br. n'a été décrite qu'en Inde, par exemple.⁸³ Les espèces de ce genre sont utilisées en médecine traditionnelle, plus spécifiquement dans le traitement du paludisme et des troubles gastro-intestinaux. Elles sont aussi utilisées en tant que toniques généraux, aphrodisiaques, anticholériques et antipyrétiques.⁸⁴ Cela a conduit à plusieurs études phytochimiques de ces plantes.^{85, 86}

1.3 *Alstonia balansae* Guillaumin

Alstonia balansae Guillaumin (synonyme *A. undulata* Guillaumin) est une plante endémique de Nouvelle-Calédonie (**Figure 14**),⁸⁷ synonyme *A. undulata* Guillaumin) caractérisée par la présence d'alcaloïdes indolomonoterpéniques (AIMs) d'après quelques auteurs (

Tableau 3 et Figure 15).^{79, 88, 89}



Figure 14. *A. balansae* dans l'herbier du Royal Botanical Gardens

⁸² <https://doi.org/10.15468/39omei>, (accessed 19-09-2018).

⁸³ <http://endemia.nc/flore/fiche256>, (accessed 20-09-2018).

⁸⁴ M. S. Khyade, D. M. Kasote, et al., *Journal of ethnopharmacology*, 2014, **153**, 1-18.

⁸⁵ A. P. G. Macabeo, K. Krohn, et al., *Phytochemistry*, 2005, **66**, 1158-1162.

⁸⁶ N. Keawpradub, E. Eno-Amooquaye, et al., *Planta Med.*, 1999, **65**, 311-315.

⁸⁷ <http://specimens.kew.org/herbarium/K000857281>, (accessed 18-09-2018).

⁸⁸ T.-M. Pinchon, J.-M. Nuzillard, et al., *Phytochemistry*, 1990, **29**, 3341-3344.

⁸⁹ D. Guillaume, A. Morfaux, et al., *Phytochemistry*, 1984, **23**, 2407-2408.

Tableau 3. Alcaloïdes indolomonoterpéniques décrits dans *A. balansae*

Nom de la molécule	Décrite par	Type de squelette	Source
10-hydroxypéricyclivine (30)	T.-M. Pinchon et al. ⁸⁸	Sarpagine	Feuilles
10-méthoxypéricyclivine (31)	T.-M. Pinchon et al. ⁸⁸	Sarpagine	Feuilles
11-méthoxyakuammicine (32)	D. Guillaume et al. ⁸⁹	Akuammicine	Écorces
11-méthoxyakuammicine N-oxyde (33)	D. Guillaume et al. ⁸⁹	Akuammicine	Écorces
alstonidine (34)	D. Guillaume et al. ⁸⁹	Strictosidine	Écorces
cabucraline (35)	D. Guillaume et al. ⁸⁹	Akuammiline	Écorces, feuilles
cabucraline N-oxyde (36)	D. Guillaume et al. ⁸⁹	Ajmaline	Écorces
cathafoline (37)	D. Guillaume et al. ⁸⁹	Akuammiline	Feuilles
deplancheine (38)	D. Guillaume et al. ⁸⁹	Corynanthe	Écorces
désoxycabufiline (39)	D. Guillaume et al. ⁸⁹	Akuammiline-akuammiline	Écorces
fluorocarpamine (40)	D. Guillaume et al. ⁸⁹	Pleiocarpamane	Écorces
gentiacraline (41)	D. Guillaume et al. ⁸⁹	Akuammiline	Écorces
N-méthyl-10-hydroxypéricyclivine (42)	T.-M. Pinchon et al. ⁸⁸	Sarpagine	Feuilles
N-méthyl-10-méthoxypéricyclivine (43)	T.-M. Pinchon et al. ⁸⁸	Sarpagine	Feuilles
N-méthyl-16-épépéricyclivine (44)	T.-M. Pinchon et al. ⁸⁸	Sarpagine	Feuilles
péricyclivine (45)	D. Guillaume et al. ⁸⁹	Sarpagine	Feuilles
pleiocarpamine (46)	D. Guillaume et al. ⁸⁹	Pleiocarpamane	Écorces
plumocraline (47)	D. Guillaume et al. ⁸⁹	Akuammiline	Écorces
tétrahydroalstonine (48)	D. Guillaume et al. ⁸⁹	Ajmalicine	Écorces, feuilles
vincamedine (49)	D. Guillaume et al. ⁸⁹	Ajmaline	Écorces
vincorine (50)	D. Guillaume et al. ⁸⁹	Akuammiline	Feuilles
voachalotinal (51)	T.-M. Pinchon et al. ⁸⁸	Sarpagine	Feuilles



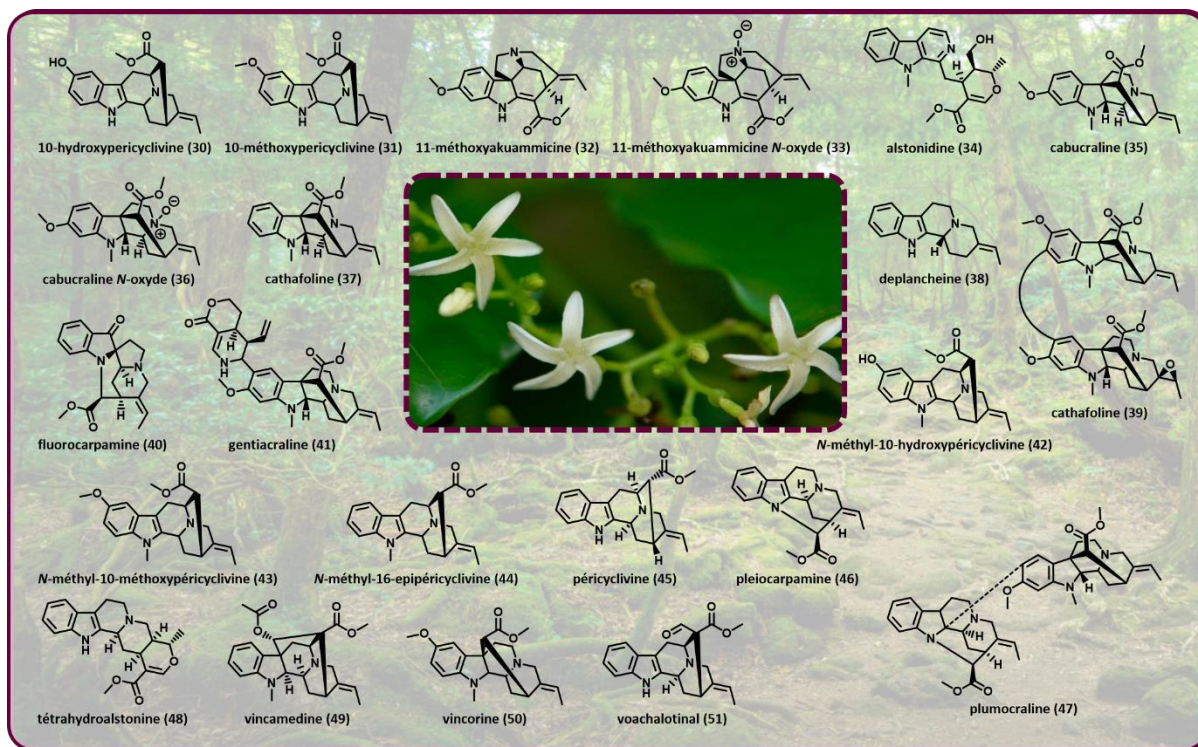


Figure 15. Alcaloïdes indolomonoterpéniques décrits dans *A. balsanae*

1.4 Travaux entrepris sur *Alstonia* spp.

Suite à nos travaux de déréplication sur *G. laeve*, le Dr Marc Litaudon (ICSN-CNRS) nous a fourni les feuilles et les écorces de trois espèces du genre *Alstonia* endémiques de Nouvelle Calédonie :

- *Alstonia balsanae* Guillaumin,
- *Alstonia vielliardii* Van Heurck & Müll.Arg.,
- *Alstonia quaternata* Van Heurck & Müll.Arg.

Comme point de départ, les extraits alcaloïdiques des feuilles et des écorces de trois espèces ont suivi une déréplication par *molecular networking* en utilisant comme référence la MIADB (résultats non présentés dans ce manuscrit). Une première exploration du réseau moléculaire obtenu nous a permis de différencier deux types de *clusters* :

- *Clusters* formés par des molécules provenant de plusieurs espèces (donc, des molécules communes aux espèces du genre).
- *Clusters* formés exclusivement par des molécules provenant d'une seule espèce (donc, de molécules potentiellement plus intéressantes).



Compte tenu de notre intérêt pour la recherche de molécules à motif chimique innovant, le deuxième type de *cluster* a attiré notre attention. Certains de ces *clusters* étaient formés exclusivement par des substances produites par une partie spécifique d'une des trois espèces. C'est ainsi que nous avons décidé d'explorer l'espace chimique de l'extrait alcaloïdique des feuilles d'*Alstonia balansae* en utilisant la technique du *molecular networking* dans une stratégie de déréplication, comme cela a été fait pour réaliser l'étude de *G. laeve*.⁴⁰

Cependant, cette fois-ci nous avons décidé d'associer cette stratégie à MetWork.²¹ À partir des « *matches* » obtenus avec la MIADB, nous avons employé cet outil pour prédire les structures des composés dérivés. C'est ainsi que nous avons identifié, pour des nœuds dont les masses n'étaient pas décrites dans la littérature, cinq structures prédites par Metwork. Nous avons donc ciblé ces composés pour confirmer ces structures anticipées. Les résultats obtenus ont été mis en forme pour soumission d'une publication à *Angewandte Chemie*.





CANPA : Anticipation de produits naturels assistée par ordinateur

Article soumis à *Angewandte Chemie*

Résumé : Dans notre quête visant à découvrir des nouveaux alcaloïdes indolomonoterpéniques (AIMs), cinq alcaloïdes *N*-oxydés de type sarpagine, non décrits précédemment, ont été isolés des feuilles d'*Alstonia balansae*. L'élucidation structurale a été réalisée en analysant leurs données spectroscopiques SMHR et RMN. La configuration absolue de tous les composés isolés a été déduite de la comparaison entre les spectres ECD expérimentaux mesurés et ceux décrits dans la littérature pour les composés de type sarpagine. Le *workflow* d'isolement était guidé par une stratégie de déréplication basée sur la technique des réseaux moléculaires assistée par l'outil MetWork. En outre, six composés déjà connus, non décrits dans le genre *Alstonia*, ont été dérèpliqués dans le réseau moléculaire de l'extrait alcaloïdique des feuilles d'*A. balansae* et ont pu se voir attribuer différents niveaux de confiance d'identification. Cette étude constitue le premier exemple de *workflow* d'isolement de produits naturels où les structures ciblées ont été initialement générées et anticipées par ordinateur avant leur isolement.





2. PUBLICATION ASSOCIÉE

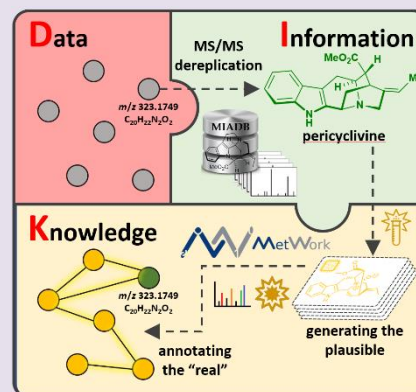


article soumis, 2019

CANPA: Computer-Assisted Natural Products Anticipation

Alexander E. Fox Ramos, Coralie Pavesi, Marc Litaudon, Vincent Dumontet, Erwan Poupon, Pierre Champy, Grégory Genta-Jouve*, and Mehdi A. Beniddir*

Abstract: In our quest to discover new monoterpene indole alkaloids, five previously undescribed sarpagine-like N-oxide alkaloids have been targeted and isolated from the leaves of *Alstonia balansae*, using a molecular networking-based dereplication strategy fueled by computer-generated annotations. This study constitutes the first example of a natural product discovery workflow in which the targeted structures were initially generated and therefore anticipated by a computer prior to their isolation.





CANPA: Computer-Assisted Natural Products Anticipation

Alexander E. Fox Ramos,^[a] Coralie Pavesi,^[a] Marc Litaudon,^[b] Vincent Dumontet,^[b] Erwan Poupon,^[a] Pierre Champy,^[a] Grégory Genta-Jouve,^{[c, d],*} and Mehdi A. Beniddir,^{[a],*}

Dedicated to Michel Rohmer on the occasion of his 70th birthday

Abstract: In our quest to discover new monoterpene indole alkaloids, five previously undescribed sarpagine-like *N*-oxide alkaloids have been targeted and isolated from the leaves of *Alstonia balansae*, using a molecular networking-based dereplication strategy fueled by computer-generated annotations. This study constitutes the first example of a natural product discovery workflow in which the targeted structures were initially generated and therefore anticipated by a computer prior to their isolation.

Extending knowledge and principles of structural chemical diversity is one fundamental outcome derived from the natural product (NP) discovery process. From an applied standpoint, this process often consists of an extraction step followed by a spate of finely tuned scenarios, each making use of different targeting strategies (*i.e.*, bioactivity-based^[1] and/or structure-based^[2]), leading to the identification of NP structures. Remarkably, discovery-based workflows systematically need a crucial NPs database search that identifies and filters out any known compounds that are present in the extract.^[3] Usually, this dereplication step relies mainly on spectrometric data queries such as exact masses (chemical formulas), tandem mass spectra (MS/MS), and NMR chemical shifts. Interestingly however, despite significant efforts made in the field, searching for structures early in the discovery pipeline still remain a long-standing challenge. Exploiting MS/MS data to generate reasonable molecular structures of fully unknown metabolites has emerged as a particularly appealing strategy.^[4] The use of CSI:FingerID^[5] and MS2LDA^[6], just to name a few, have allowed the generation of molecular fingerprints and substructures, respectively.

Since Laskin's, Bandeira's, and Dorrestein's seminal article, published seven years ago,^[7] the field of MS/MS-based targeting

of NPs has experienced a dramatic increase, resulting in the development of a number of successful and creative endeavors by several groups,^[8] including ours.^[2a, 9] The success of a molecular networking (MN)-based dereplication relies greatly on the quality and the availability of MS/MS data. Although the Global Natural Products Social Molecular Networking (GNPS) library^[10] provides more than 70000 MS/MS spectra to accomplish the high-throughput dereplication of the submitted data, only a limited number of nodes can be annotated. As such, the field recently culminated by the development of several strategies to annotate molecular networks.^[5, 11] From the data-information-knowledge-wisdom (DIKW) perspective,^[12] these approaches are able to convert DATA into INFORMATION, however, only few of them encompass all the functionalities required for the generation of KNOWLEDGE, *i.e.* the identification of the relationship between two or more compounds and the creation of meaning. Thereby, accepting the idea that similar compounds share similar properties, the ability to foresee appears as a necessity to harness new natural products in complex matrices. Recently, we have introduced a web server named MetWork^[13] based on MS/MS data (organized as molecular network), a collaborative library of (bio)chemical transformations and a MS/MS spectra prediction module based on CFM-ID.^[14] It represents a full implementation of the metabolome consistency concept,^[15] which was then reworded as the "virtuous cycle of metabolites identification" two years ago.^[16] This approach (detailed on figure 1) consists in three linked components: structural database expansion, spectral prediction, and data annotation. Indeed, having one node identified in the molecular network, the server generates structures (plausible and known structures of natural products) and, a similarity comparison between theoretical and experimental MS/MS spectra is then performed in order to annotate the nodes.

Pursuing our interest in exploring the chemistry of monoterpene indole alkaloids (MIAs) derived from previously investigated plants,^[2a, 9a, 17] the phytochemical investigation of the leaves of *Alstonia balansae* Guillaumin (Apocynaceae) was carried out, using a MN-based dereplication strategy fueled by computer-generated metabolites using the MetWork web server. An alkaloid extract of the leaves of *A. balansae* was analyzed by LC-HRMS/MS (positive-ion mode). The fragmentation data were preprocessed *via* the MZmine 2 software^[18] and then organized by MN and annotated by the MIADB^[19] (hosted by the GNPS), allowing the identification of some already-described compounds in the genus *Alstonia* like, akuammicine (1), antirhine (2), pericyclivine (3), pleiocarpamine (4), tabersonine (5) and tubotaiwine (6).

- [a] Dr. A. Fox Ramos, C. Pavesi, Prof. Dr. E. Poupon, Prof. Dr. P. Champy, Dr. M. A. Beniddir
Équipe "Pharmacognosie-chimie des substances naturelles" BioCIS, Univ. Paris-Sud, CNRS, Université Paris-Saclay, 5 rue J.-B. Clément, 92290 Châtenay-Malabry, France
E-mail: mehdi.beniddir@u-psud.fr
- [b] Dr. M. Litaudon, Dr. V. Dumontet
Institut de Chimie des Substances Naturelles, CNRS, ICSN UPR 2301, Université Paris-Saclay, 21 avenue de la Terrasse, 91198, Gif-sur-Yvette, France
- [c] Dr. G. Genta-Jouve
C-TAC UMR CNRS 8038 CiTCoM, Faculté de Pharmacie de Paris, Université Paris Descartes, 4 avenue de l'Observatoire, 75006 Paris, France.
E-mail: gregory.genta-jouve@parisdescartes.fr
- [d] Unité Molécules de Communication et Adaptation des Micro-organismes (UMR 7245), Sorbonne Universités, Muséum National d'Histoire Naturelle, CNRS, Paris, France

Supporting information for this article is given via a link at the end of the document.

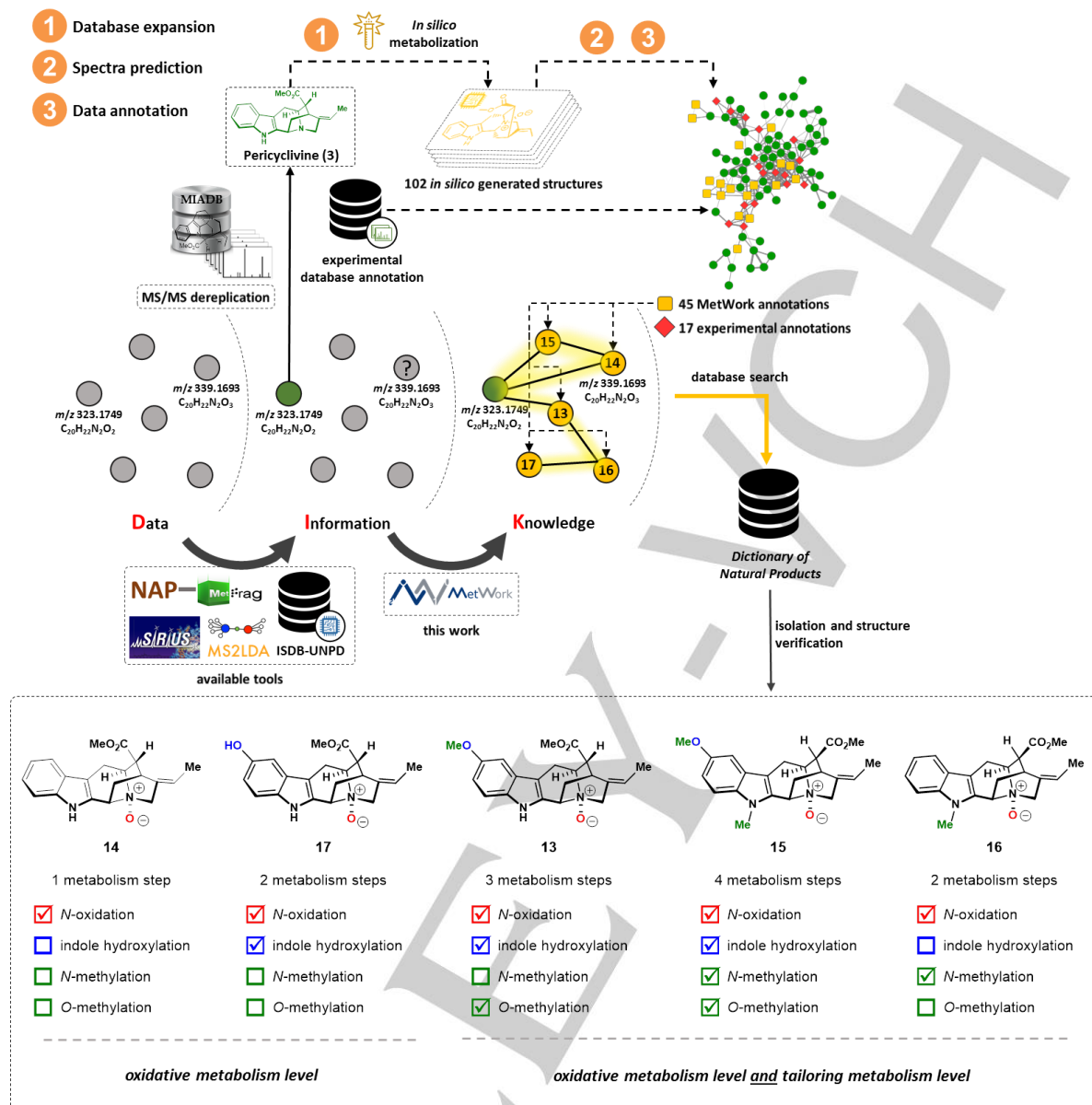


Figure 1. Overview of the CANPA workflow and structures of the targeted compounds (13–17) (See Figures S1, S2, and S3 in SI for full annotated molecular and metabolization networks).

Additionally, six known compounds that have never been described in *Alstonia* spp. were also dereplicated: condylocarpine *N*-oxide (7), corynantheidal (8), corynantheidol (9), geissoschizine (10), splendoline (11), and voachalotine (12). The confidence levels of compound annotations, attributed to the dereplicated MIAs according to Fiehn *et al.*^[20] are summarized in **Table S1 – S1**.

In addition to these experimental annotations, the molecular network was further enriched by computer-generated metabolites obtained with MetWork. These *in silico*-generated compounds were obtained through four reaction steps, consistent with the MIA biochemistry, including, *N*-, and *O*-methylation, *N*-oxidation, and *para*-indole hydroxylation. In the resulting molecular network, metabolites from the alkaloid extract from the leaves of *A. balansae* are represented by green circles; and nodes that have been annotated by the GNPS and

MetWork are shown as red rhombus and as yellow squares, respectively (**Figure 1**, **Figure S1** and **S2 – S1**).

Starting from the GNPS annotations, putative structures were generated through the MetWork server and compared to experimental data, leading to the annotation of additional nodes (yellow squares) in the neighboring of the aforementioned identified ones (**Figure S1** and **S2 – S1**). The result of the database search of the 99 computer-generated structures (45 unique structures) against the *Dictionary of Natural Products* (DNP),^[21] is summarized in **Table S2 – S1**. Interestingly, this search revealed that 39 were unknown, whereas six of them were described in other species of the Apocynaceae family (**Table S2 – S1**). Remarkably, some masses were related to a certain number of previously undescribed structures (for *e.g.*, m/z $[M+H]^+$ 339.1703 – 5 unknown structures; m/z $[M+H]^+$ 353.1860 – 3 unknown structures; m/z $[M+H]^+$ 355.1658 – 4

unknown structures; m/z $[M+H]^+$ 369.1809 – 10 unknown structures; m/z $[M+H]^+$ 383.1965 – 7 unknown structures, **Table S3 - S1**). Furthermore, most of those masses belong to the same metabolization cluster (A, **Figure S3 - S1**), and their corresponding structures derived from the *in silico* metabolization of pericyclivine (**3**) (Figure 1). At this stage, these masses were of interest for isolation in regard to their putative novelty. In addition, LC-DAD-ELSD-MS analysis of *A. balansae* chromatographic fractions indicated that some compounds related to the above-mentioned MetWork annotated nodes could be isolated in amounts suitable for full structure elucidation. Accordingly, the isolation workflow was directed toward compounds **13** (m/z 369.1809), **14** (m/z 339.1703), **15** (m/z 383.1965), **16** (m/z 353.1860), and **17** (m/z 355.1658). The extract was investigated and as anticipated, the MN-guided isolation of the targeted molecules led to the identification of five previously undescribed sarpagine *N*-oxide alkaloids (**13-17**). The structures of compounds **13-17** were elucidated by NMR spectroscopic data analysis and their absolute configurations were assigned by comparison of experimental ECDs spectra (see SI for full structure elucidation and absolute configuration assignment).

MetWork: What we thought, what we have got

As it has been outlined above, the utilization of MetWork allowed the early detection of potentially novel molecules in the extract. After the structural elucidation of the isolated compounds, a fair evaluation of this strategy prompted us to evoke the following observations:

- All the isolated compounds were annotated and anticipated by MetWork. However, only **14**, **16**, and **17** had their correct structures among the first-ranked ones (**Table S2 - S1**).
- Despite the correct computer-generation of the structures of **13** and **15**, their misannotations highlight the limitation of the MS/MS spectra prediction module embarked by MetWork. In this context, further efforts are needed to sharpen the accuracy of MS/MS spectra prediction algorithms.
- Submitting the structures generated by MetWork to a NP database search provided an efficient way to shortlist the proposed structures, allowing to guide the isolation workflow towards new compounds.
- The epimerization at C-16 of compounds **15** and **16** was not expected and pointed out the limitation of MS/MS-based prediction approaches, such as MetWork, to correctly identify tridimensional structures.

In summary, we applied an unprecedented natural product discovery pipeline and demonstrated its utility by isolating five new sarpagine *N*-oxide alkaloids. To the best of our knowledge, this study constitutes the first example of a natural product chemistry workflow in which the targeted structures were initially generated by a computer prior to their isolation. As we celebrate the 150th anniversary of the periodic table of elements proposed by Dmitri Mendeleev, it is important to take a step back on the way natural products chemistry has been realized in the last decades. Convinced that the essence is a function of the relationship^[22], the anticipation of new natural products has to be

considered as a paradigm shift in the NP field and modern analysts have to chase for well-defined structures.

Acknowledgements

This work has been developed with the funding from Cienciactiva, an initiative of the National Council for Science, Technology and Technological Innovation (CONCYTEC) contract 239-2015-FONDECYT. This was also supported by the French ANR grant (ANR-15-CE29-0001). We express our thanks to Jean-Christophe Jullian (BioCIS) for performing the NMR analysis of the isolated compounds. We also thank Karine Leblanc (BioCIS) for her assistance in the preparative HPLC purifications. The authors are very grateful to North Province of New Caledonia and Falconbridge New Caledonia that have facilitated our field investigation.

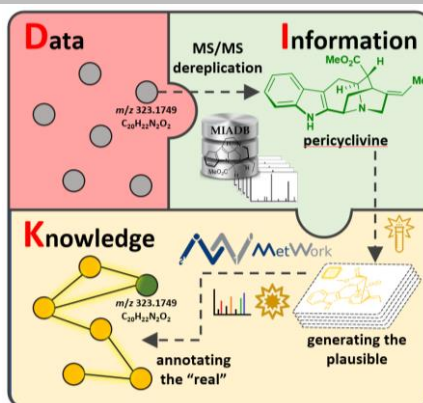
Keywords: Anticipation, *in silico* metabolization, monoterpene indole alkaloids, *N*-oxide, dereplication, MetWork,

- [1] a) 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 and P. C. Dorrestein, *J. Nat. Prod.* **2018**, *81*, 758-767; b) 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 and M. Litaudon, *ACS Chem. Biol.* **2017**, *12*, 2644-2651; c) C. B. Naman, R. Rattan, S. E. Nikoulina, J. Lee, B. W. Miller, N. A. Moss, L. Armstrong, P. D. Boudreau, H. M. Debonsi, F. A. Valeriote, P. C. Dorrestein and W. H. Gerwick, *J. Nat. Prod.* **2017**, *80*, 625-633; d) K. L. Kurita, E. Glassey and R. G. Lington, *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 11999-12004.
- [2] a) E. Otogo N'Nang, G. Bernadat, E. Mouray, B. Kumulungui, P. Grellier, E. Poupon, P. Champy and M. A. Beniddir, *Org. Lett.* **2018**, *20*, 6596-6600; b) N. Wada, R. D. Kersten, T. Iwai, S. Lee, F. Sakurai, T. Kikuchi, D. Fujita, M. Fujita and J.-K. Weng, *Angew. Chem. Int. Ed.* **2018**, *57*, 3671-3675; c) T. Grkovic, R. H. Pouwer, M.-L. Vial, L. Gambini, A. Noël, J. N. A. Hooper, S. A. Wood, G. D. Mellick and R. J. Quinn, *Angew. Chem. Int. Ed.* **2014**, *53*, 6070-6074; d) F. Qiu, J. B. McAlpine, D. C. Lankin, I. Burton, T. Karakach, S.-N. Chen and G. F. Pauli, *Anal. Chem.* **2014**, *86*, 3964-3972; e) S. T. Deyrup, L. E. Eckman, P. H. McCarthy, S. R. Smedley, J. Meinwald and F. C. Schroeder, *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 9753-9758.
- [3] M. T. Henke and N. L. Kelleher, *Nat. Prod. Rep.* **2016**, *33*, 942-950.
- [4] S. Böcker, *Curr. Opin. Chem. Biol.* **2017**, *36*, 1-6.
- [5] K. Dührkop, H. Shen, M. Meusel, J. Rousu and S. Böcker, *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 12580-12585.
- [6] J. J. J. van der Hooft, J. Wandy, M. P. Barrett, K. E. V. Burgess and S. Rogers, *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, 13738-13743.
- [7] J. Watrous, P. Roach, T. Alexandrov, B. S. Heath, J. Y. Yang, R. D. Kersten, M. van der Voort, K. Pogliano, H. Gross, J. M. Raaijmakers, B. S. Moore, J. Laskin, N. Bandeira and P. C. Dorrestein, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, E1743-E1752.
- [8] R. A. Quinn, L.-F. Nothias, O. Vining, M. Meehan, E. Esquenazi and P. C. Dorrestein, *Trends Pharmacol. Sci.* **2017**, *38*, 143-154.
- [9] a) A. E. Fox Ramos, C. Alcover, L. Evanno, A. Maciuk, M. Litaudon, C. Duplais, G. Bernadat, J.-F. Gallard, J.-C. Jullian, E. Mouray, P. Grellier, P. M. Loiseau, S. Pomel, E. Poupon, P. Champy and M. A. Beniddir, *J. Nat. Prod.* **2017**, *80*, 1007-1014; b) N. Bonneau, G. Chen, D. Lachkar, A. Boufridi, J.-F. Gallard, P. Retailleau, S. Petek, C. Debitus, L. Evanno, M. A. Beniddir and E. Poupon, *Chem. Eur. J.* **2017**, *23*, 14454-14461; c) C. Audoin, A. Zampalégré, N. Blanchet, A. Giuliani, E. Roulland, O. Laprévotte and G. Genta-Jouve, *Molecules* **2018**, *23*, 1237.
- [10] M. Wang, J. J. Carver, V. V. Phelan, L. M. Sanchez, N. Garg, Y. Peng, D. D. Nguyen, J. Watrous, C. A. Kapon, T. Luzzatto-Knaan, C. Porto, A. Bouslimani, A. V. Melnik, M. J. Meehan, W.-T. Liu, M. Crusemann, P. D. Boudreau, E. Esquenazi, M. Sandoval-Calderon, R. D. Kersten, L. A. Pace, R. A. Quinn, K. R. Duncan, C.-C. Hsu, D. J. Floros, R. G. Gavilan, K. Kleigrew, T. Northen, R. J. Dutton, D. Parrot, E. E. Carlson, B. Aigle, C. F. Michelsen, L. Jelsbak, C. Sohlenkamp, P. Pevzner, A. Edlund, J. McLean, J. Piel, B. T. Murphy, L. Gerwick, C.-C. Liaw, Y.-L. Yang, H.-U. Humpf, M. Maansson, R. A. Keyzers, A. C. Sims, A. R. Johnson, A. M. Sidebottom, B. E. Sedo, A.

- Klitgaard, C. B. Larson, C. A. Boya P, D. Torres-Mendoza, D. J. Gonzalez, D. B. Silva, L. M. Marques, D. P. Demarque, E. Pociute, E. C. O'Neill, E. Briand, E. J. N. Helfrich, E. A. Granatosky, E. Glukhov, F. Ryffel, H. Houson, H. Mohimani, J. J. Kharbush, Y. Zeng, J. A. Vorholt, K. L. Kurita, P. Charusanti, K. L. McPhail, K. F. Nielsen, L. Vuong, M. Elfeki, M. F. Traxler, N. Engene, N. Koyama, O. B. Vining, R. Baric, R. R. Silva, S. J. Mascuch, S. Tomasi, S. Jenkins, V. Macherla, T. Hoffman, V. Agarwal, P. G. Williams, J. Dai, R. Neupane, J. Gurr, A. M. C. Rodriguez, A. Lamsa, C. Zhang, K. Dorrestein, B. M. Duggan, J. Almaliti, P.-M. Allard, P. Phapale, L.-F. Nothias, T. Alexandrov, M. Litaudon, J.-L. Wolfender, J. E. Kyle, T. O. Metz, T. Peryea, D.-T. Nguyen, D. VanLeer, P. Shinn, A. Jadhav, R. Muller, K. M. Waters, W. Shi, X. Liu, L. Zhang, R. Knight, P. R. Jensen, B. O. Palsson, K. Pogliano, R. G. Linington, M. Gutierrez, N. P. Lopes, W. H. Gerwick, B. S. Moore, P. C. Dorrestein and N. Bandeira, *Nat. Biotechnol.* **2016**, *34*, 828-837.
- [11] a) P.-M. Allard, T. Péresse, J. Bisson, K. Gindro, L. Marcourt, V. C. Pham, F. Roussi, M. Litaudon and J.-L. Wolfender, *Anal. Chem.* **2016**, *88*, 3317-3323; b) R. R. da Silva, M. Wang, L.-F. Nothias, J. J. J. van der Hooft, A. M. Caraballo-Rodríguez, E. Fox, M. J. Balunas, J. L. Klassen, N. P. Lopes and P. C. Dorrestein, *PLOS Comput. Biol.* **2018**, *14*, e1006089; c) S. Böcker and K. Dührkop, *J. Cheminformatics* **2016**, *8*, 5; d) J.-L. Wolfender, J.-M. Nuzillard, J. J. J. van der Hooft, J.-H. Renault and S. Bertrand, *Anal. Chem.* **2019**, *91*, 704-742; e) J. Wandy, Y. Zhu, J. J. J. van der Hooft, R. Daly, M. P. Barrett and S. Rogers, *Bioinformatics* **2018**, *34*, 317-318.
- [12] J. Rowley, *J. Inf. Sci.* **2007**, *33*, 163-180.
- [13] Y. Beauxis and G. Genta-Jouve, *Bioinformatics* **2018**, bty864, <https://doi.org/10.1093/bioinformatics/bty1864>.
- [14] F. Allen, A. Pon, M. Wilson, R. Greiner and D. Wishart, *Nucleic Acids Res.* **2014**, *42*, W94-99.
- [15] C. Audoin, V. Cocandeau, O. P. Thomas, A. Bruschini, S. Holderith and G. Genta-Jouve, *Metabolites* **2014**, *4*, 421-432.
- [16] P.-M. Allard, G. Genta-Jouve and J.-L. Wolfender, *Curr. Opin. Chem. Biol.* **2017**, *36*, 40-49.
- [17] a) E. Otogo N'Nang Obiang, G. Genta-Jouve, J.-F. Gallard, B. Kumulungui, E. Mouray, P. Grellier, L. Evanno, E. Poupon, P. Champy and M. A. Beniddir, *Org. Lett.* **2017**, *19*, 6180-6183; b) M. A. Beniddir, G. Genta-Jouve and G. Lewin, *J. Nat. Prod.* **2018**, *81*, 1075-1078.
- [18] T. Pluskal, S. Castillo, A. Villar-Briones and M. Orešič, *BMC Bioinform.* **2010**, *11*, 395.
- [19] MIADB is a MS/MS spectral database of 172 monoterpene indole alkaloids that can be accessed online on the GNPS, see: <https://gnps.ucsd.edu/ProteoSAFe/libraries.jsp> (accessed Jan 27, 2019).
- [20] I. Blaženović, T. Kind, J. Ji and O. Fiehn, *Metabolites* **2018**, *8*, 31.
- [21] DNP, accessed: Jan 26, 2019.
- [22] G. Bachelard, *La valeur inductive de la relativité*, Paris, Vrin, **1929**, p. 208.

Communication

Guided by molecular networking-based dereplication strategy fueled by MetWork annotations, five new sarpagine *N*-oxide monoterpene indole alkaloids were anticipated and isolated from the leaves of the New Caledonian plant *Alstonia balansae*. This strategy constitutes the first computer-assisted anticipation applied to the isolation of natural products.



Alexander E. Fox Ramos, Coralie Pavesi, Marc Litaudon, Vincent Dumontet, Pierre Champy, Grégory Genta-Jouve* and Mehdi A. Beniddir*

Page No. – Page No.

CANPA: Computer-Assisted Natural Products Anticipation

WILEY-VCH

Supplementary information:

Figure S1. Full molecular network of the alkaloid extract of the leaves of *Alstonia balansae*

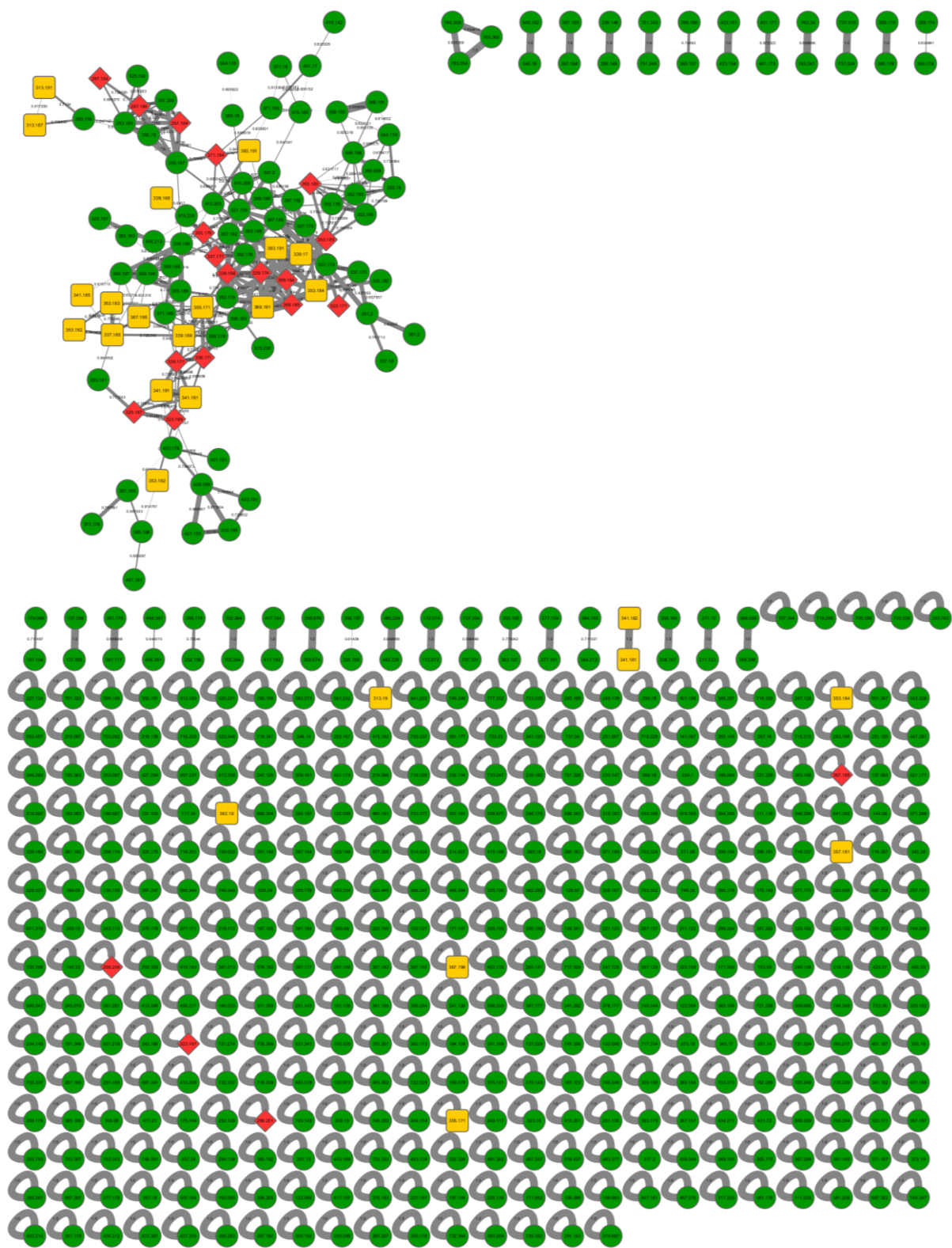


Figure S1. Full molecular network of the alkaloid extract of the leaves of *Alstonia balansae*



Figure S2. Molecular network: targeted nodes

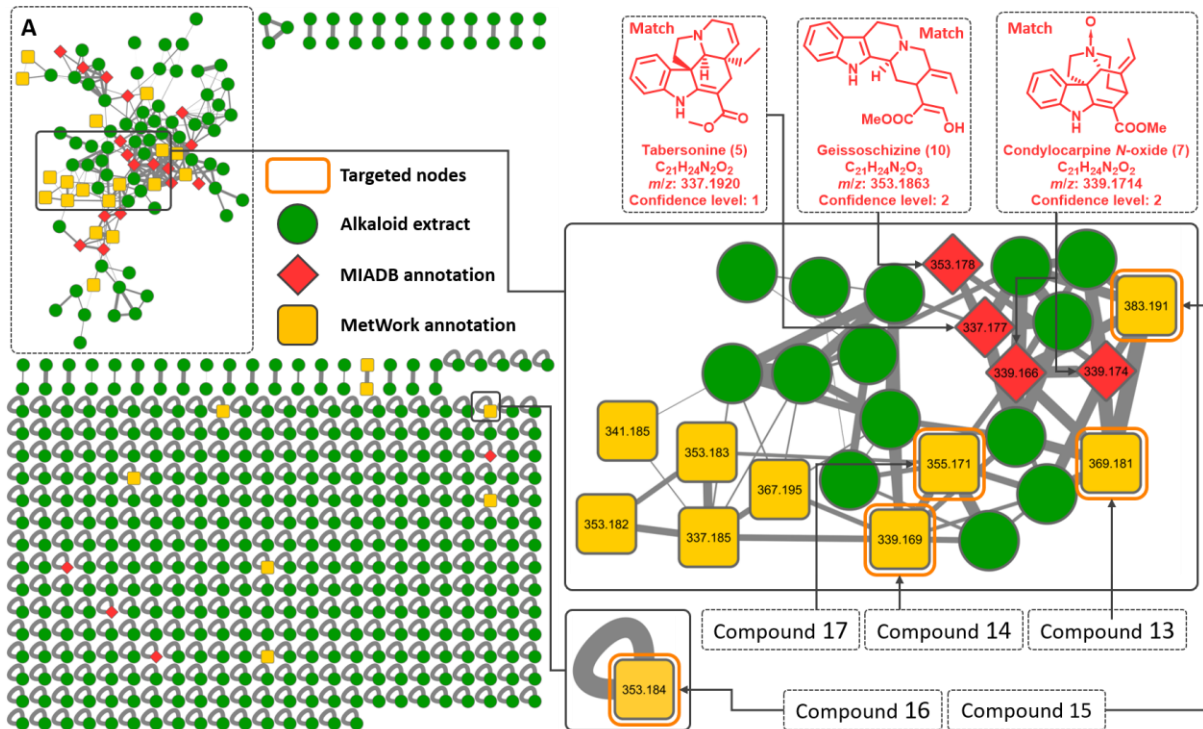


Figure S2. Targeted nodes for isolation in the obtained molecular network from the alkaloid extract of the leaves of *Alstonia balansae*.

Table S1. Confidence levels attributed to the dereplicated MIAs against the MIADB.

Table S1. Confidence levels attributed to the annotated substances in the alkaloid extract by the MIADB, according to Blaženović et al.¹ Δ RT: retention time difference between the compound in the extract and the analogue from the MIADB.

Compound	Confidence level	Δ RT (min)	Cosine	Matching data
Akuammicine (1)	1	0.58	0.79	MS/MS spectra and RT
Antirhine (2)	2	2.49	0.85	MS/MS spectra and RT
Pericyclivine (3)	2	2.66	0.75	MS/MS spectra
Pleiocarpamine (4)	1	1.19	0.75	MS/MS spectra
Tabersonine (5)	1	0.23	0.61	MS/MS spectra and RT
Tubotaiwine (6)	2	3.80	0.80	MS/MS spectra
Condylocarpine <i>N</i> -oxide (7)	2	2.16	0.71	MS/MS spectra
	2	3.50	0.69	MS/MS spectra
	2	4.75	0.64	MS/MS spectra
Corynantheidal (8)	1	0.61	0.80	MS/MS spectra and RT
	2	3.40	0.70	MS/MS spectra
Corynantheidol (9)	1	0.02	0.76	MS/MS spectra and RT
	2	2.91	0.63	MS/MS spectra
Geissoschizine (10)	1	1.49	0.65	MS/MS spectra and RT
	2	2.79	0.65	MS/MS spectra
Splendoline (11)	2	2.89	0.69	MS/MS spectra
Voachalotine (12)	2	1.58	0.64	MS/MS spectra



Figure S3. MetWork metabolization network

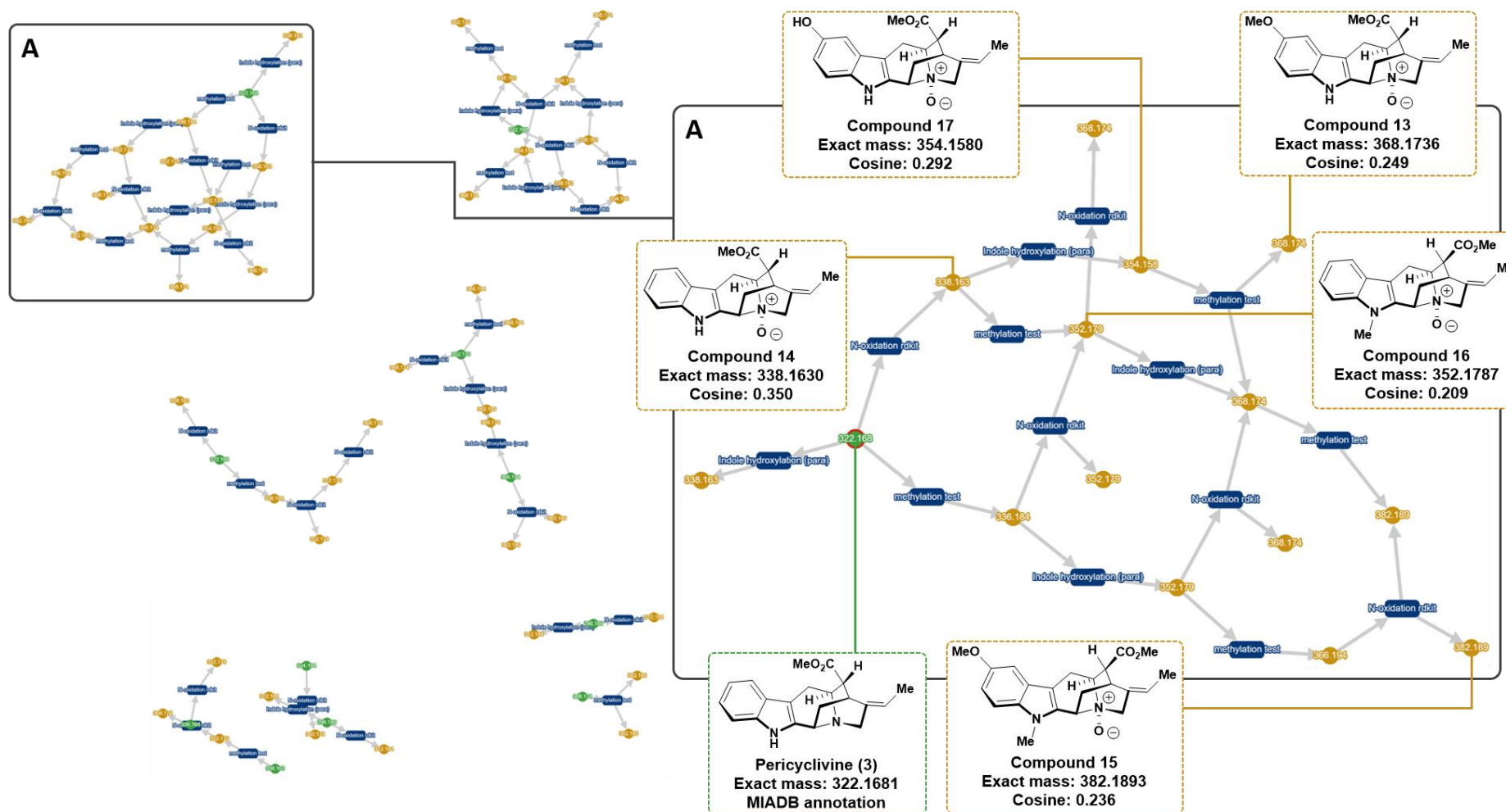
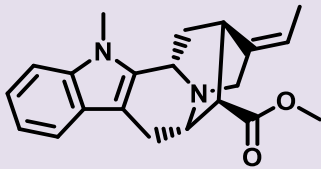
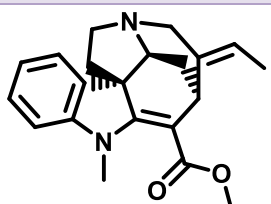
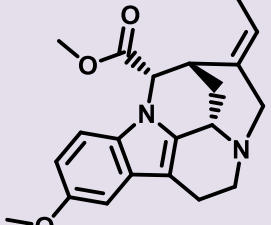
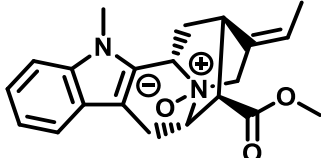
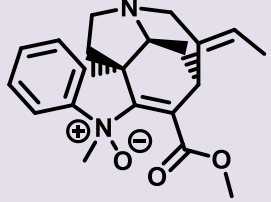
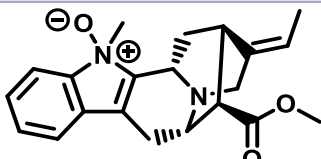


Figure S3. Overview In-silico annotations by MetWork of the isolated compounds

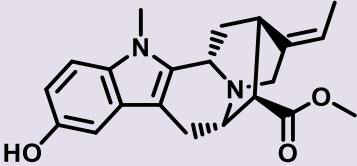
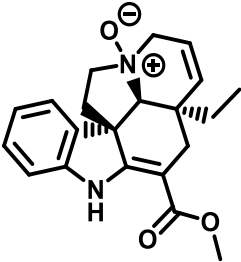
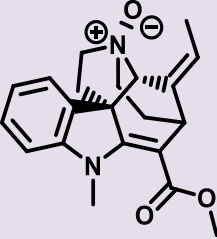
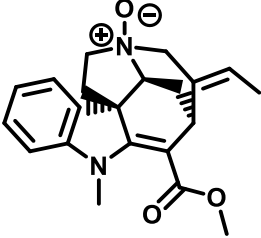
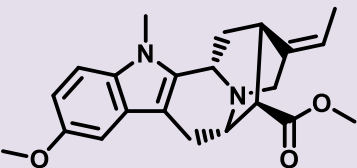
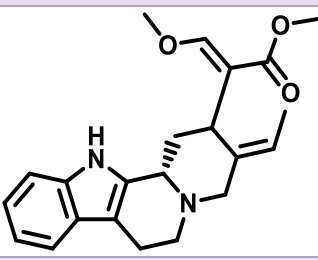
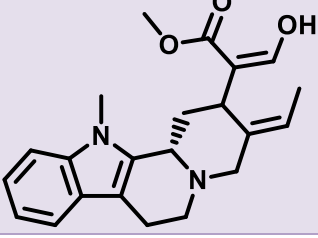


Table S2. Results of the search of the structures proposed by MetWork against the DNP

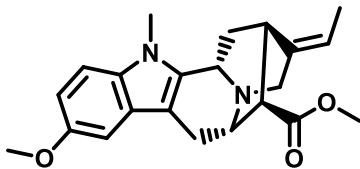
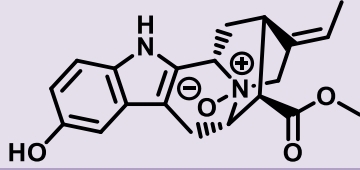
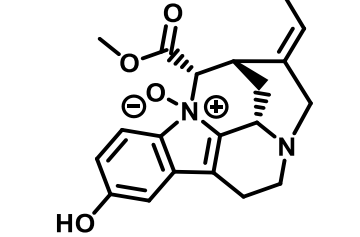
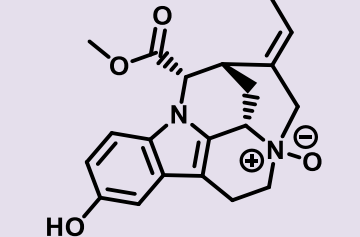
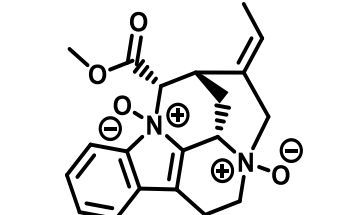
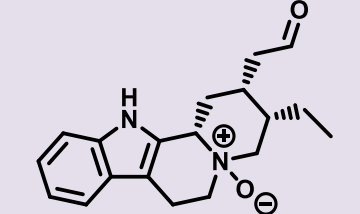
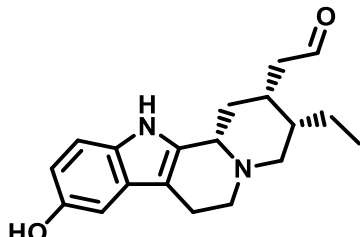
Table S2. Annotated molecules by MetWork in the full molecular network. It should be noted that, for the most part of them, multiple annotations were proposed. 99 structures were generated by MetWork corresponding to 45 different in-silico generated structures. Noteworthy, the stereochemistry of the MetWork-generated metabolites arises from that of the GNPS hits (i.e., as they were entered in the GNPS library).

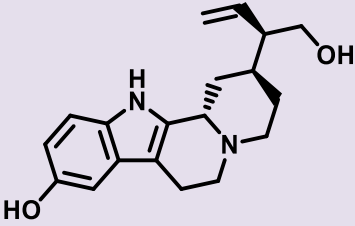
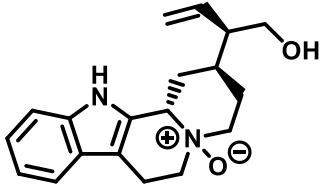
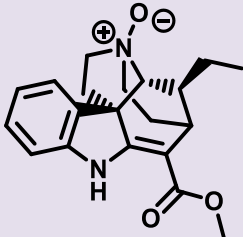
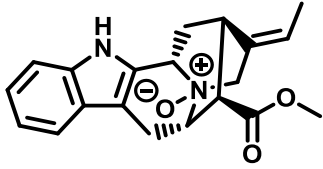
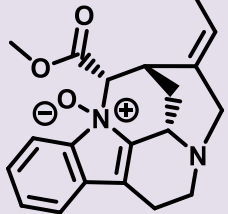
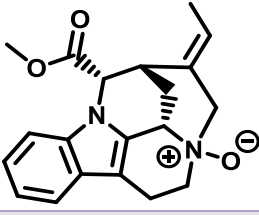
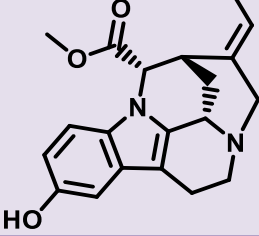
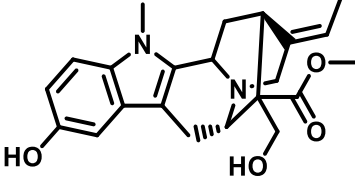
Exact mass	Cosine	Structure	Number of hits (DNP)	Comment
336.1838	0.432		2	17-carboxylic acid, <i>N</i> ₁ -methyl, methylester tombozine; 16-epimer, 17-carboxylic acid, <i>N</i> ₁ -methyl, methylester tombozine – Not described in the genus
	0.426		0	-
352.1787	0.432		0	-
	0.430		0	Structure of compound 16
	0.426		0	-
	0.425		0	-

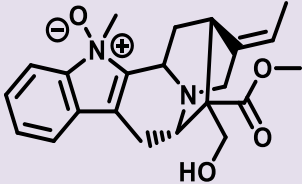
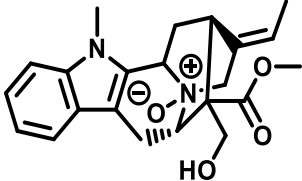

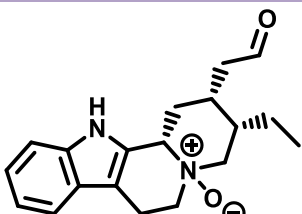
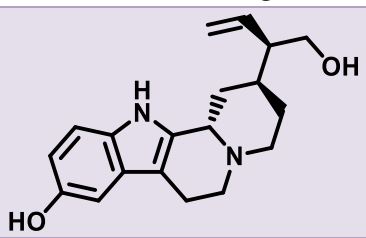
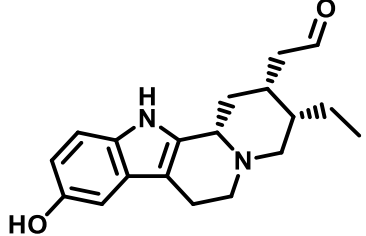
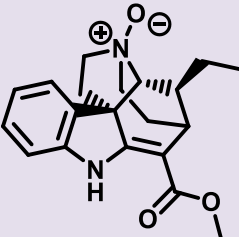


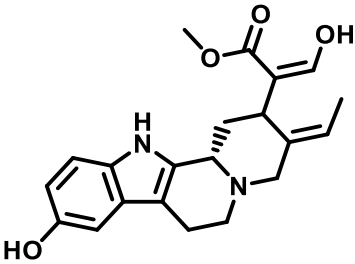
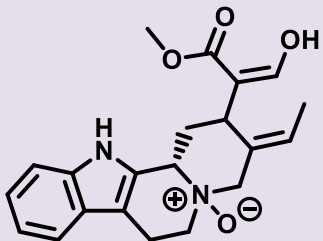
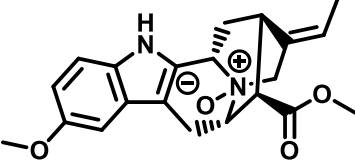
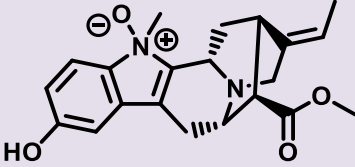
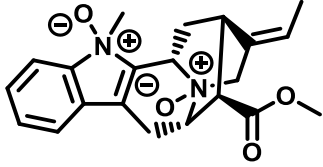
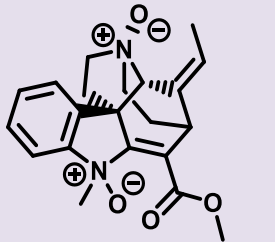
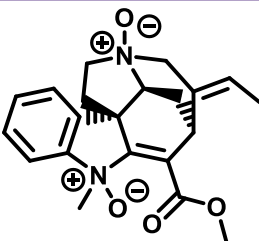
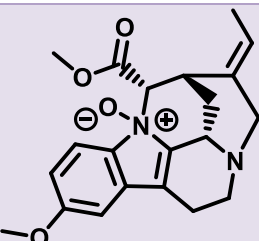
	0.424		1	16-epimer, 17-carboxylic acid, N^1 -methyl, methylester sarpagine – Not described in the genus
	0.418		0	-
	0.412		0	-
	0.376		0	-
366.1943	0.360		1	16-epimer, 17-carboxylic acid, methylether, N^1 -methyl, methylester sarpagine – Not described in the genus
	0.351		2	3-epimer, methylether geissoschizine; methylether geissoschizine – Not described in the genus
	0.317		0	-

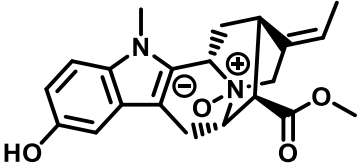
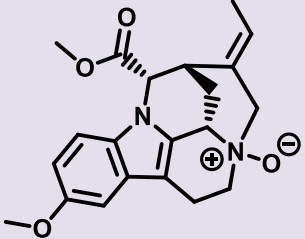
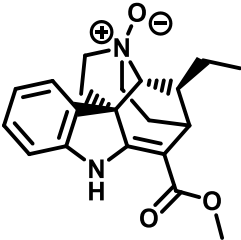
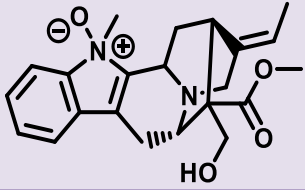
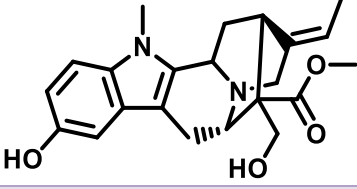
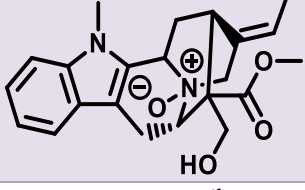
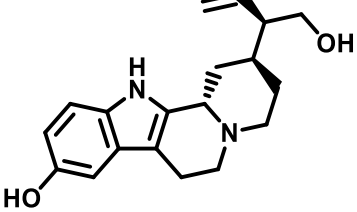
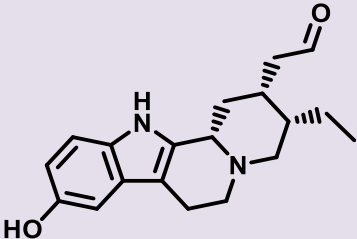
352.1787	0.447		0	Structure of compound 16
	0.445		0	-
	0.444		0	-
	0.444		0	-
	0.442		0	-
	0.382		1	16-epimer, 17-carboxylic acid, N ¹ -methyl, methylester sarpagine – Not described in the genus
	0.380		0	-
	0.351		0	-

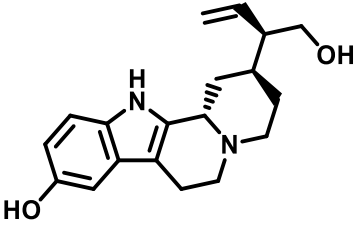
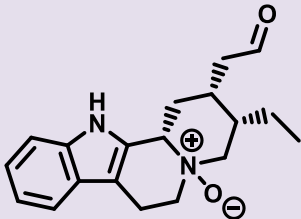
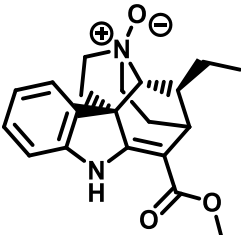
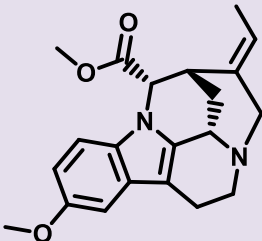
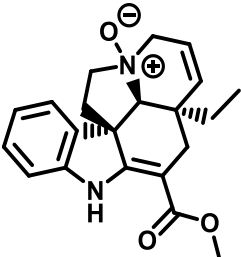
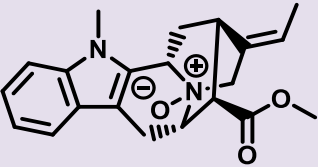
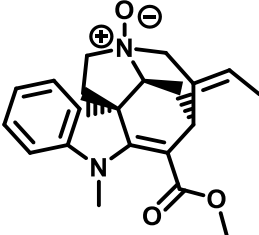
366.1943	0.227		1	16-epimer, 17-carboxylic acid, methylether, N ¹ -methyl, methylester sarpagine – Not described in the genus
354.1580	0.292		0	Structure of compound 17
	0.283		0	-
	0.281		0	-
	0.257		0	-
312.1838	0.287		0	-
	0.287		0	-

	0.274		1	10-hydroxyantirrhine – Not described in the genus
	0.258		0	-
340.1787	0.292		0	-
	0.235		0	Structure of compound 14
	0.210		0	-
338.1630	0.208		0	-
	0.206		0	-
382.1893	0.238		0	-

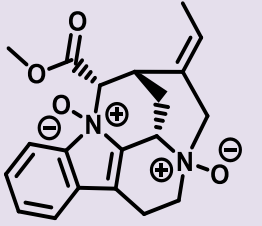
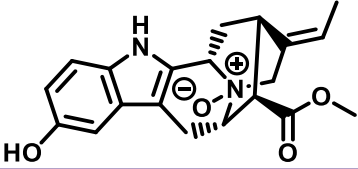
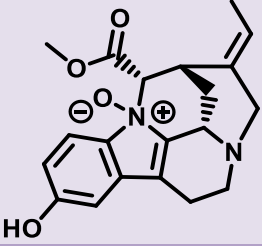
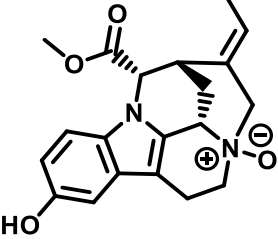
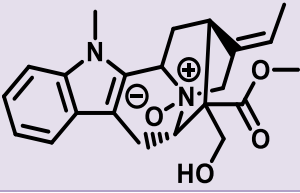
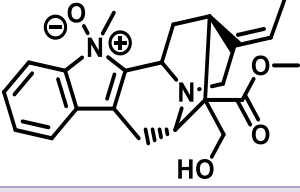
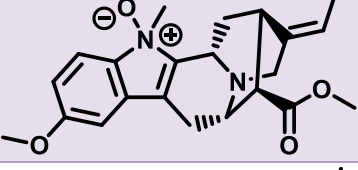
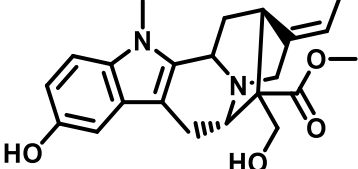
	0.237		0	-
	0.230		0	-
340.1787	0.361		0	-
	0.231		0	-
312.1838	0.206		1	10-hydroxyantirrhine – Not described in the genus
	0.201		0	-
340.1787	0.292		0	-

368.1736	0.291		0	-
	0.255		0	-
	0.249		0	Structure of compound 13
	0.245		0	-
	0.240		0	-
	0.226		0	-
	0.225		0	-
	0.223		0	-

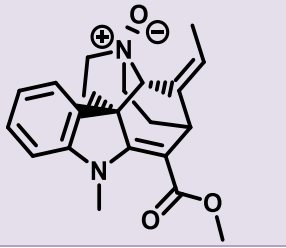
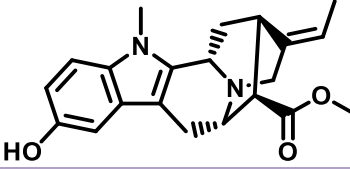
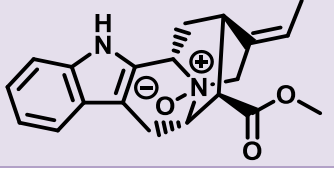
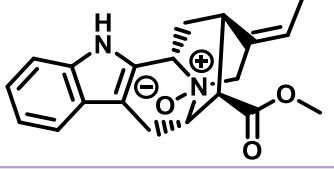
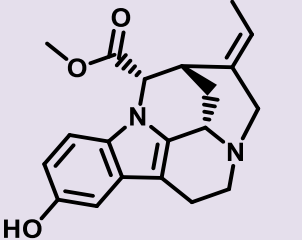
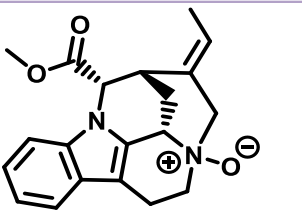
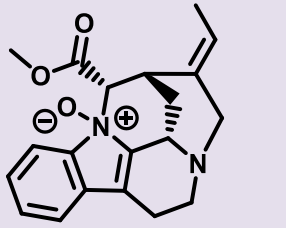
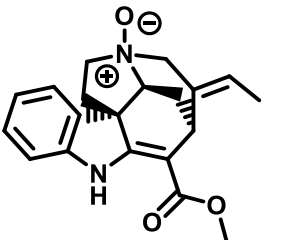
	0.218		0	
	0.214		0	-
340.1787	0.354		0	-
	0.293		0	-
382.1893	0.258		0	-
	0.249		0	-
312.1838	0.298		1	10-hydroxyantirhine – Not described in the genus
	0.275		0	-

	0.260		1	10-hydroxyantirrhine – Not described in the genus
	0.226		0	-
340.1787	0.354		0	-
	0.363		0	-
352.1787	0.353		0	-
	0.342		0	Structure of compound 16
	0.342		0	-

	0.342		0	-
	0.341		1	16-epimer, 17-carboxylic acid, N ¹ -methyl, methylester sarpagine – Not described in the genus
	0.327		0	-
	0.307		0	-
352.1787	0.253		0	-
	0.223		0	-
	0.223		1	16-epimer, 17-carboxylic acid, N ¹ -methyl, methylester sarpagine – Not described in the genus
	0.209		0	Structure of compound 16

354.1580	0.571		0	-
	0.568		0	Structure of compound 17
	0.559		0	-
	0.549		0	-
382.1893	0.327		0	-
	0.317		0	-
	0.286		0	-
	0.282		0	-

	0.236		0	Structure of compound 15
	0.220		0	-
	0.211		0	-
352.1787	0.262		0	-
	0.240		0	Structure of compound 16
	0.238		0	-
	0.235		0	-
	0.226		1	16-epimer, 17-carboxylic acid, <i>N</i> ¹ -methyl, methylester sarpagine – Not described in the genus

	0.217		0	-
	0.210		1	16-epimer, 17-carboxylic acid, N ¹ -methyl, methylester sarpagine – Not described in the genus
338.1630	0.235		0	Structure of compound 14
	0.350		0	Structure of compound 14
	0.349		0	-
338.1630	0.341		0	-
	0.338		0	-
	0.331		0	-

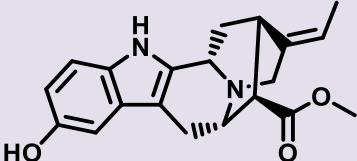
	0.329		1	16-epimer, 17-carboxylic acid, methylester sarpagine – Not described in the genus
--	-------	---	---	---

Table S3. Results of the search of the structures proposed by MetWork against the DNP – Number of already-described and unknown structures

Table S3. Number of already-described and unknown structures proposed by MetWork in the full molecular network.

Exact mass	m/z [M+H] ⁺	Number of already-described structures in DNP	Number of unknown structures in DNP	Comments
368.1736	369.1809	0	10	Compound 13
352.1787	353.1820	1	7	
352.1787	353.1833	1	7	
352.1787	353.1820	1	7	
382.1893	383.1965	0	7	Compound 15
352.1787	353.1842	2	5	
338.1630	339.1703	1	5	Compound 14
354.1580	355.1658	0	4	Compound 17
338.1630	339.1702	0	4	
354.1580	355.1709	0	4	
312.1838	313.1909	1	3	
382.1893	383.1905	0	3	
382.1893	383.1910	0	3	
352.1787	353.1860	1	3	Compound 16
312.1838	313.1901	1	2	
312.1838	313.1869	2	2	
336.1838	337.1847	2	1	
366.1943	367.1951	3	1	
340.1787	341.1824	0	1	
340.1787	341.1846	0	1	
340.1787	341.1813	0	1	
340.1787	341.1814	0	1	
340.1787	341.1814	0	1	
338.1630	339.1675	0	1	
366.1943	367.1956	1	0	

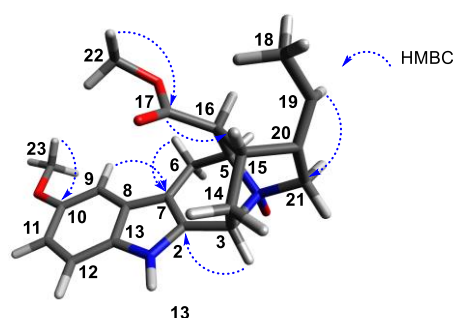


Structural elucidation

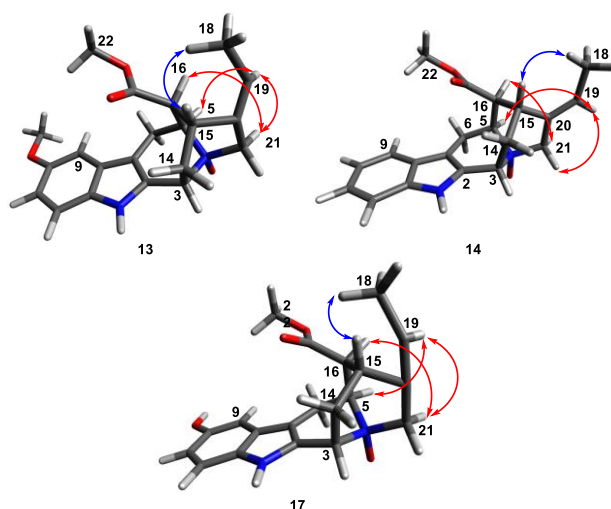
Compounds **13–17** are structurally closely related, as can be evidenced by several shared spectral features. The UV spectra show absorption bands at 221, 275 and 294 nm, consistent with a tetrahydro- β -carboline chromophore², and the produced IR spectra exhibit absorption bands arising from amine groups (3400 cm⁻¹) and ester groups (1720 cm⁻¹).

Compound **13** was obtained as a brown, amorphous solid, $[\alpha]_D^{29} -80$ (c 0.05, MeOH). The HRESIMS data indicated a $[M+H]^+$ ion peak at $m/z = 369.1809$, which corresponds to the molecular formula C₂₁H₂₄N₂O₄, requiring eleven degrees of unsaturation. The ¹³C NMR data of **13** (Table S4) indicated the presence of 21 carbon signals, comprising three methyl groups, three sp³ methylenes, four sp³ methines, four sp² methines, two sp² oxygenated carbons and five sp² quaternary carbons. Analysis of the ¹H NMR data (Table S4) showed three aromatic protons at δ_H 6.94 (1H, d, $J = 2.2$ Hz, H-9), 6.77 (1H, dd, $J = 8.8, 2.2$ Hz, H-11) and 7.22 (1H, d, $J = 8.8$ Hz, H-12), corresponding to a monosubstituted indole moiety, two methoxy groups δ_H 3.11 (3H, s, H₃-22) and 3.81 (3H, s, H₃-23), two aminomethines δ_H 4.67 (1H, d, $J = 10.3$ Hz, H-3) and 4.08 (1H, d, $J = 15.6$ Hz, H-5), an aminomethylene δ_H 4.49 (1H, d, $J = 15.2$ Hz, H-21a) and 4.13 (1H, d, $J = 2.3$ Hz, H-21b), one double-bonded methine δ_H 5.41 (1H, q, $J = 6.6$ Hz, H-19) and one methyl from an ethylidene side chain δ_H 1.67 (3H, d, $J = 6.6$ Hz, H-18). As both aminomethines and the aminomethylene had deshielded proton chemical shifts when compared to those linked to a non-oxidized nitrogen atom², *N*-oxidation was suspected. The position of each of the two methoxy groups in the molecule was deduced by the observed HMBC correlations arising from both H₃-22 and C-17, and H₃-23 and C-10. The *O*-CH₃ (C-22) at δ_C 52.1 was deduced to be part of a carbomethoxy group, while the one at δ_C 56.3 (C-23) was identified to be linked to the indole. A monosubstituted tryptoline moiety, counting for seven double-bonds equivalents (DBE), was constructed from this indole with the HMBC correlations between H-9 and C-7, H₂-6 and C-7, and H-3 and C-2. COSY correlations between H-3 and H-14b, and H-14a and H-15 allowed to anchor this chain to the tetrahydro- β -carboline moiety. Thereafter, the aminomethylene at δ_C 71.7 (C-21) was connected to the ethylidene at δ_C 132.7 (C-20) based on the HMBC correlation between H-19 and C-21. This part of the molecule was defined as an additional heterocycle linked to the already-defined tryptoline moiety, forming a four-cycle structure. HMBC correlation between H-15 and C-17 allowed the anchor of the carbomethoxy group to the C-16 (Figure S4). Finally, as the DBE total for the constructed molecule was 10, the presence of an additional cycle was suspected. Indeed, COSY and HMBC correlations indicated that carbons C-5, C-15, and C-16 were linked together to form a supplementary cycle. The structure was further confirmed by comparison against the NMR data of the non *N*-oxide version of this molecule² and therefore named 10-methoxypericyclivine *N*-oxide (**13**).



Figure S4. Key HMBC correlations for compound **13**

The determination of the relative configuration of **13** was carried out using the NOESY data. The NOESY correlations from H-5 to H-6a, H-16, and H-21b indicated the cofacial orientation of these protons (Figure S5, red arrows). On the other hand, the NOESY correlation from H-18 to H-15 (Figure S5, blue arrow) was used to place them on the opposite face, suggesting that the geometry of the C-19 – C-20 double bond was *E*. Finally, the configuration of C-16 was assigned as *S** by the characteristic shielded chemical shift ($\delta_{\text{H}} = 3.11$) of the methoxy group, due to the anisotropic effect of the indole ring. The relative configuration of **13** was assigned as 3*S**, 4*S**, 5*S**, 15*R**, 16*S**.

Figure S5. Selected NOESY correlations for compounds **13**, **14**, and **17**.

Compound **14** was obtained as a brown, amorphous solid, $[\alpha]_{\text{D}}^{28} -120$ (c 0.05, MeOH). Its molecular formula was determined to be $\text{C}_{20}\text{H}_{22}\text{N}_2\text{O}_3$, based on its HRESIMS data ($[\text{M}+\text{H}]^+$ m/z 339.1703), corresponding to eleven degrees of unsaturation, like compound **13**. The ^1H NMR spectrum showed aromatic signals for an unsubstituted indole moiety, and a signal for a methoxy group. HMBC correlations were observed from H-19 to C-21 and from H-15 to C-17 (Table S4). As ^1H NMR data of **14** were similar to those of **13**, and taking into account the information from the molecular network, compound **14** was suspected to be an analogue of **13**. This was confirmed by the 2D NMR spectroscopic data (COSY and HMBC). Indeed, the noticeable difference between both molecules was the replacement of the methoxy signal by an additional aromatic proton at C-10 in **14**. All of these data led to assign the structure of pericyclivine *N*-oxide to compound **14**.



An examination of the NOESY data of **14** permitted the observation of similar correlations to those of compound **13** leading to conclude that both compounds have the same relative configuration.

Compound **15** was obtained as a brown, amorphous solid, $[\alpha]^{29}_D -100$ (c 0.05, MeOH). The HRESIMS data exhibits a $[M+H]^+$ ion peak at $m/z = 383.1965$, which corresponded to the molecular formula $C_{22}H_{26}N_2O_4$, one methylene less in regard to **13**. Analysis of the 1H NMR data of **15** showed signals for a monosubstituted indole moiety, and signals for two methoxy groups and one *N*-Me group (Table S5). The HMBC correlation observed between the protons *N*-Me group and C-2 confirmed its location. Consequently, **15** was identified as *N*-methyl-10-methoxy-16-epi-pericyclivine *N*-oxide.

Compound **16** was obtained as a brown, amorphous solid, $[\alpha]^{29}_D -100$ (c 0.05, MeOH). Its molecular formula, $C_{21}H_{24}N_2O_3$, was established from its HRESIMS data ($[M+H]^+$ m/z 353.1860). The 1H NMR data (Table S5) exhibited signals for four aromatic protons, suggesting the presence of an unsubstituted indole moiety, a signal for only one methoxy group (instead of two in compound **15**) and another one for the *N*-Me group. Analysis of COSY and HMBC spectra strongly suggested that **16** corresponds to a demethoxy version of **15**. Thus, **16** was assigned as *N*-methyl-16-epi-pericyclivine *N*-oxide.

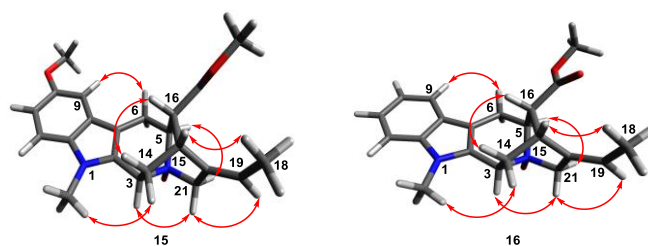
The relative configurations of **15** and **16** were determined together, as they showed common spectral traits. Indeed, as indicated in Figure S6, both compounds shared the following NOESY correlations: 1-*N*-Me and H-14a, H-3 and H-21a, H-5 and H-21b, H-6b and H-9, H-14b and H-16, H-15 and H-18, and H-19 and H-21a. Consequently, the relative configuration of **15** and **16** was assigned as 3*S**, 4*S**, 5*S**, 15*R**, 16*R**.

Compound **17** was obtained as a brown, amorphous solid, $[\alpha]^{28}_D -280$ (c 0.025, MeOH). The HRESIMS data ($[M+H]^+$ m/z 355.1658) of **17** established the molecular formula as $C_{20}H_{22}N_2O_4$, 14 mass units less than **13**. The 1H NMR data (Table S4) showed three aromatic signals, characteristic of a monosubstituted indole moiety and a signal of one methoxy group. The only difference between **13** and **17** was the absence of a second methoxy group in **17**, suggesting the presence of another substituent linked to C-10. These data established the nature of the substituent as an hydroxy group, leading to assign the structure of 10-hydroxypericyclivine *N*-oxide to **17**.

The relative configuration of **17** was determined to be identical with that of **1** on the basis of their similar NOESY correlations (Figure S5).



Figure S6. Selected NOESY correlations for 15 and 16.



Absolute configurations of compounds 13–17 determination

The absolute configurations of compounds 13–17 were determined using electronic circular dichroism spectroscopy (ECD) (**Figure S37**). Indeed, previous studies correlated a typical positive Cotton effect observed at 250–280 nm region to the (3*S*) configuration in sarpagine structures.³ Thus, compounds 13, 14, and 17 were assigned the absolute configuration 3*S*, 4*S*, 5*S*, 15*R*, 16*S*. Whereas compounds 15 and 16 were assigned the absolute configuration 3*S*, 4*S*, 5*S*, 15*R*, 16*R*.

Table S4. ¹H and ¹³C NMR data for 10-methoxypericyclivine *N*-oxide (13), pericyclivine *N*-oxide (14) and 10-hydroxypericyclivine *N*-oxide (17).

No.	1 ^a		2 ^a		5 ^a	
	δ _C	δ _H , (J in Hz)	δ _C	δ _H , (J in Hz)	δ _C	δ _H , (J in Hz)
2	134.2		133.9		134.4	
3	67.6	4.67, d (10.3)	67.8	4.67, d (8.3)	67.7	4.61, d (10.3)
5	69.1	4.08, d (15.6)	69.1	4.10, m	69.1	4.07, m
6a	20.9	3.48, dd (16.2, 3.7)	20.9	3.52, dd (16.2, 2.9)	20.9	3.46, dd (15.7, 3.3)
6b		3.39, d (16.0)		3.42, d (16.0)		3.32, m
7	103.6		103.8		103.2	
8	127.7		127.4		128.1	
9	101.2	6.94, d (2.2)	118.9	7.44, d (7.4)	103.3	6.81, d (1.1)
10	155.5		120.3	7.02, t (7.4)	151.8	
11	112.8	6.77, dd (8.8, 2.2)	122.8	7.10, t (7.4)	112.8	6.66, dd (8.3, 1.1)
12	113.1	7.22, d (8.8)	112.3	7.33, d (7.4)	112.7	7.15, d (8.3)
13	133.9		138.8		133.3	
14a	30.3	2.89, dd (13.1, 4.3)	30.3	2.90, dd (13.1, 4.1)	30.3	2.88, dd (13.1, 3.9)
14b		2.34, t (11.2)		2.37, t (10.8)		2.35, td (11.5, 0.8)
15	27.5	3.10, m	27.6	3.11, m	27.6	3.10, m
16	45.3	3.23, d (10.7)	45.3	3.25, d (11.3)	45.3	3.23, dd (11.1, 2.8)
17	171.4		171.5		171.5	
18	12.6	1.67, d (6.6)	12.6	1.68, d (6.4)	12.6	1.68, d (6.2)
19	119.4	5.41, q (6.6)	119.1	5.44, q (6.4)	119.2	5.43, q (6.2)
20	132.7		133.2		133.0	
21a	71.7	4.49, d (15.2)	72.1	4.51, d (15.0)	71.9	4.49, d (15.0)
21b		4.13, d (2.3)		4.07, m		4.07, m
22-OCH₃	52.1	3.11, s	52.0	3.07, s	52.1	3.12, s
23-OCH₃	56.3	3.81, s				

^a Measured in CD₃OD at 400 and 100 MHz for ¹H and ¹³C NMR, respectively



Table S5. ^1H and ^{13}C NMR data for 10-methoxypericyclivine *N*-oxide (15), pericyclivine *N*-oxide (2) and 10-hydroxypericyclivine *N*-oxide (16).

No.	3^a		4^a	
	δ_{C}	δ_{H} , (J in Hz)	δ_{C}	δ_{H} , (J in Hz)
2	135.8	-	134.9	-
3	66.0	4.74, d (10.3)	65.6	4.82, d (10.6)
5	67.7	4.11, dd (5.4, 6.6)	67.6	4.16, dd (7.1, 5.2)
6a	24.8	3.54, dd (16.0, 4.9)	24.7	3.54, dd (16.1, 4.9)
6b		2.78, d (16.0)		2.82, dd (16.0, 0.7)
7	101.8		102.1	
8	128.1		127.7	
9	101.5	6.99, d (2.1)	119.3	7.48, d (7.4)
10	155.8	-	120.7	7.08, t (7.4)
11	112.9	6.84, dd (8.9, 2.1)	123.1	7.19, t (7.4)
12	111.1	7.28, d (8.9)	110.4	7.36, d (7.4)
13	134.7		139.4	
14a	33.3	2.54, t (11.8)	33.1	2.55, ddd (11.3, 12.2, 1.0)
14b		2.06, dd (12.8, 3.1)		2.05, ddd (13.1, 4.1, 1.4)
15	28.7	3.38, s	28.6	3.37, t (2.0)
16	48.5	2.94, d (7.6)	48.4	2.93, dd (7.7, 1.3)
17	172.5		172.3	
18	12.9	1.62, d (6.5)	12.9	1.62, d (6.7)
19	121.1	5.41, brs	121.4	5.40, q (6.1)
20	130.9		130.5	
21a	71.7	4.39, d (15.3)	71.4	4.41, dt (15.5, 2.6)
21b		4.02, d (15.4)		4.05, dt (15.4, 0.9)
22-OCH₃	53.0	3.72, s	53.0	3.71, s
23-OCH₃	56.3	3.81, s		
1-N-Me	30.0	3.63, s	29.8	3.63, s

^a Measured in CD₃OD at 400 and 100 MHz for ^1H and ^{13}C NMR, respectively



Experimental Section

General experimental procedures

Optical rotations were measured at 25 °C on a Polar 32 polarimeter. UV spectra were recorded on a Lightwave II+ WPA 7126 V. 1.6.1 spectrophotometer. ECD spectra were measured at 25 °C on a JASCO J-810 spectropolarimeter. IR spectra were recorded with a PerkinElmer type 257 spectrometer. The NMR spectra were recorded on a Bruker AM-400 (400 MHz) NMR spectrometer using CD₃OD as solvent. The solvent signals were used as references. A Sunfire[®] analytical C18 column (150 × 4. mm, i.d. 5 μm; Waters) was used to determine the HPLC conditions for the purification of the isolated compounds. Sunfire[®] preparative C18 columns (150 × 19 mm, i.d. 5 μm and 150 × 30 mm, i.d. 5 μm; Waters) were used for preparative HPLC separations using a Waters Delta Prep equipped with a binary pump (Waters 2525) and a UV-visible diode array detector (190–600 nm, Waters 2996). A silica 120 g Grace[™] cartridge was used for flash chromatography using an Armen Instrument spot liquid chromatography flash apparatus. Chemicals and solvents were purchased from Sigma-Aldrich.

Plant material

The stem bark and leaves of *A. balansae* were collected in June 2003 from the Koniambo massif, New Caledonia, and were identified by Vincent Dumontet. A voucher specimen (No. DUM-0311) has been deposited at the Herbarium of IRD Nouméa (NOU), New Caledonia.

Extraction and isolation

An alkaloid extract of the air-dried and powdered leaves of *A. balansae* (757 g) was prepared adding a 6 M NH₄OH solution to the plant sample. The mixture was extracted three times with CH₂Cl₂ (3 × 1.5 L, stirring, 3 h each, 20 °C, atmospheric pressure) and then the CH₂Cl₂ crude extract was concentrated under vacuum at 38 °C and redissolved in CH₂Cl₂ (50 mL). This solution was subjected to extraction using a solution of hydrochloric acid 0.1 M (1 L).

The mixture was alkalinized to pH 10 with NH₄OH, counter-extracted with CH₂Cl₂ and the extract was concentrated under vacuum at 38 °C to yield 12.85 g of an alkaloid extract (1.65% m/m). A part of this residue (F1, 7 g) was subjected to flash chromatography using a silica 120 g Grace cartridge with a gradient of CH₂Cl₂–MeOH (100:0 to 0:100) at 80 mL/min to afford ten fractions, F1.1–F1.10, according to their TLC profiles. Fraction F1.5 (137.7 mg) underwent a preparative HPLC separation using a gradient of MeOH–H₂O with 0.2% formic acid (FA) (30:70 to 60:40 in 20 min) to afford 10-methoxypericyclivine *N*-oxide (**13**, 2.6 mg, yield 0.0003%), pericyclivine *N*-oxide (**14**, 3.0 mg, yield 0.0004%), *N*-methyl-10-methoxy-16-epi-pericyclivine *N*-oxide (**14**, 2.7 mg, yield 0.0004%) and *N*-methyl-16-epi-pericyclivine *N*-oxide (**16**, 6.6 mg, yield 0.0009%). Fraction F1.6 (89.3 mg) was subjected to a preparative HPLC preparative separation using a gradient of MeOH–H₂O with 0.2% FA (15:85 to 50:50 in 20 min) to give 10-hydroxypericyclivine *N*-oxide (**17**, 1.1 mg, yield 0.0001%).



Mass Spectrometry Analysis

Samples were analyzed using an Agilent LC-MS system composed of an Agilent 1260 Infinity HPLC coupled to an Agilent 6530 ESI-Q-TOF-MS operating in positive mode. A Sunfire[®] analytical C₁₈ column (150 × 2.1 mm; i.d. 3.5 μm, Waters) was used, with a flow rate of 250 μL/min and a linear gradient from 5% B (A: H₂O + 0.1% formic acid, B: MeOH) to 100% B over 30 min. ESI conditions were set with the capillary temperature at 320 °C, source voltage at 3.5 kV, and a sheath gas flow rate of 10 L/min. The divert valve was set to waste for the first 3 min. There were four scan events: positive MS, window from *m/z* 100–1200, then three data-dependent MS/MS scans of the first, second, and third most intense ions from the first scan event. MS/MS settings were: three fixed collision energies (30, 50, and 70 eV), default charge of 1, minimum intensity of 5000 counts, and isolation width of *m/z* 2. Purine C₅H₄N₄ [M + H]⁺ ion (*m/z* 121.050873) and hexakis(1H,1H,3H-tetrafluoropropoxy)-phosphazene C₁₈H₁₈F₂₄N₃O₆P₃ [M + H]⁺ ion (*m/z* 922.009 798) were used as internal lock masses. Full scans were acquired at a resolution of 11 000 (at *m/z* 922). A permanent MS/MS exclusion list criterion was set to prevent oversampling of the internal calibrant.

MZmine 2 Data-Preprocessing Parameters

The MS² data files were converted from the .d (Agilent) standard data-format to .mzXML format using the MSConvert software, part of the ProteoWizard package⁴. All .mzXML were then processed using MZmine 2 v32⁵. The mass detection was realized keeping the noise level at 1.0 E4. The ADAP chromatogram builder was used using a minimum group size of scans of 4, a group intensity threshold of 10000, a minimum highest intensity of 10000 and *m/z* tolerance of 0.02 Da (or 15 ppm)⁶. The ADAP wavelets deconvolution algorithm was used with the following standard settings: S/N threshold = 70, minimum feature height = 3000, coefficient/area threshold = 5, peak duration range 0.02 - 2 min, RT wavelet range 0.02 – 0.2. MS² scans were paired using a *m/z* tolerance range of 0.03 Da and RT tolerance range of 1.0 min. Isotopologues were grouped using the isotopic peaks grouper algorithm with a *m/z* tolerance of 0.02 Da (or 15 ppm) and a RT tolerance of 0.3 min. Peak alignment was performed using the join aligner module (*m/z* tolerance = 0.02 Da (or 15 ppm), weight for *m/z* = 2, weight for RT = 2.0, absolute RT tolerance 0.3 min). The peak list was gap-filled with the same RT and *m/z* range gap filler module (*m/z* tolerance of 0.02 Da (or 15 ppm)). Eventually, the .mgf preclustered spectral data file and its corresponding .csv metadata file (for RT, areas and formulas integration) were exported using the dedicated “Export for GNPS” and “Export to CSV file” built-in options. See further documentation at <https://bixlab.ucsd.edu/display/Public/GNPS+data+analysis+workflow+2.0>.



Molecular Networking

A molecular network was created using the online workflow at GNPS (<http://gnps.ucsd.edu>) with a parent mass tolerance of 0.02 Da and an MS/MS fragment ion tolerance of 0.02 Da, minimum cluster size of 1, run MScluster and filter precursor window tools were turned off. A network was then created where edges were filtered to have a cosine score above 0.6 and more than 6 matched peaks. Further edges between two nodes were kept in the network if and only if each of the nodes appeared in each other's respective top 10 most similar nodes. The spectra in the network were then searched against GNPS spectral libraries. All matches kept between network spectra and library spectra were required to have a score above 0.6 and at least 6 matched peaks. The resulting analysis and parameters for the network can be accessed via the link <https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=43c725a0274442b58058a04fc042524b>. The molecular networking data were analyzed and visualized using Cytoscape (ver. 3.5.1)⁷.

MetWork Annotation Parameters

MetWork annotation was achieved using the online tool available at <https://metwork.pharmacie.parisdescartes.fr>. *In silico* generated compounds were metabolized 4 times at most in a row (depth limit of 4) following selected reactions schemes according to the metabolites family (i.e., MIAs), namely “indole hydroxylation (para)”: Reaction_ID=71, “methylation”: Reaction_ID=55, and “N-oxidation”: Reaction_ID=68. A metabolization network was then created using a *m/z* tolerance of 0.02, where edges were filtered to have a cosine score above 0.2 and more than 4 matched peaks.

Spectroscopic data

10-methoxypericyclivine N-oxide (13): Brown, amorphous solid; $[\alpha]_D^{29}$ -80 (c 0.05, MeOH); UV (MeOH) λ_{\max} (log ϵ) 221 (3.87) nm, 275 (3.34) nm, 294 (3.32) nm; IR (dry film) ν_{\max} 3580 (amine), 3350, 1715 (ester), 1710, 1615, 1580 cm^{-1} ; ^1H and ^{13}C NMR data, see **Table S4**; HRESIMS *m/z* 369.1809 $[\text{M} + \text{H}]^+$ (calcd for $\text{C}_{21}\text{H}_{24}\text{N}_2\text{O}_4 + \text{H}$, 369.1804). MS/MS spectrum was deposited in the GNPS spectral library.

Pericyclivine N-oxide (14): Brown, amorphous solid; $[\alpha]_D^{28}$ -120 (c 0.05, MeOH); UV (MeOH) λ_{\max} (log ϵ) 221 (3.84) nm, 281 (3.27) nm, 289 (3.22) nm; IR (dry film) ν_{\max} 3400 (amine), 1720 (ester carbonyl) cm^{-1} ; ^1H and ^{13}C NMR data, see **Table S4**; HRESIMS *m/z* 339.1703 $[\text{M} + \text{H}]^+$ (calcd for $\text{C}_{20}\text{H}_{22}\text{N}_2\text{O}_3 + \text{H}$, 339.1693). MS/MS spectrum was deposited in the GNPS spectral library.

N-methyl-10-methoxy-16-epi-pericyclivine N-oxide (15): Brown, amorphous solid; $[\alpha]_D^{29}$ -100 (c 0.05, MeOH); UV (MeOH) λ_{\max} (log ϵ) 225 (4.02) nm, 280 (3.49) nm, 297 (3.37) nm; IR (dry film) ν_{\max} 3500 (amine), 1715 (ester carbonyl) cm^{-1} ; ^1H and ^{13}C NMR data, see **Table S5**; HRESIMS *m/z* 383.1965 $[\text{M} + \text{H}]^+$ (calcd for $\text{C}_{22}\text{H}_{26}\text{N}_2\text{O}_4 + \text{H}$, 383.1956). MS/MS spectrum was deposited in the GNPS spectral library.

N-methyl-16-epi-pericyclivine N-oxide (16): Brown, amorphous solid; $[\alpha]_D^{29}$ -100 (c 0.05, MeOH); UV (MeOH) λ_{\max} (log ϵ) 225 (4.06) nm, 283 (3.38) nm, 292 (3.06) nm; IR (dry film) ν_{\max} 3460 (amine), 1740 (ester carbonyl) cm^{-1} ; ^1H and ^{13}C NMR data, see **Table S5**; HRESIMS *m/z* 353.1860 $[\text{M} + \text{H}]^+$ (calcd for $\text{C}_{21}\text{H}_{24}\text{N}_2\text{O}_3 + \text{H}$, 353.1848). MS/MS spectrum was deposited in the GNPS spectral library.



10-hydroxypericyclivine N-oxide (**17**): Brown, amorphous solid; $[\alpha]^{28}_{\text{D}}$ -280 (c 0.025, MeOH); UV (MeOH) λ_{max} (log ϵ) 225 (4.02) nm, 283 (3.23) nm, 292 (3.06) nm; IR (dry film) ν_{max} 3520 (amine), 1740 (ester carbonyl) cm^{-1} ; ^1H and ^{13}C NMR data, see **Table S4**; HRESIMS m/z 355.1658 $[\text{M} + \text{H}]^+$ (calcd for $\text{C}_{20}\text{H}_{22}\text{N}_2\text{O}_4 + \text{H}$, 355.1666). MS/MS spectrum was deposited in the GNPS spectral library.



Figure S7. ¹H NMR Spectrum (400 MHz, CD₃OD) of 10-methoxypericyclivine N-oxide (13)

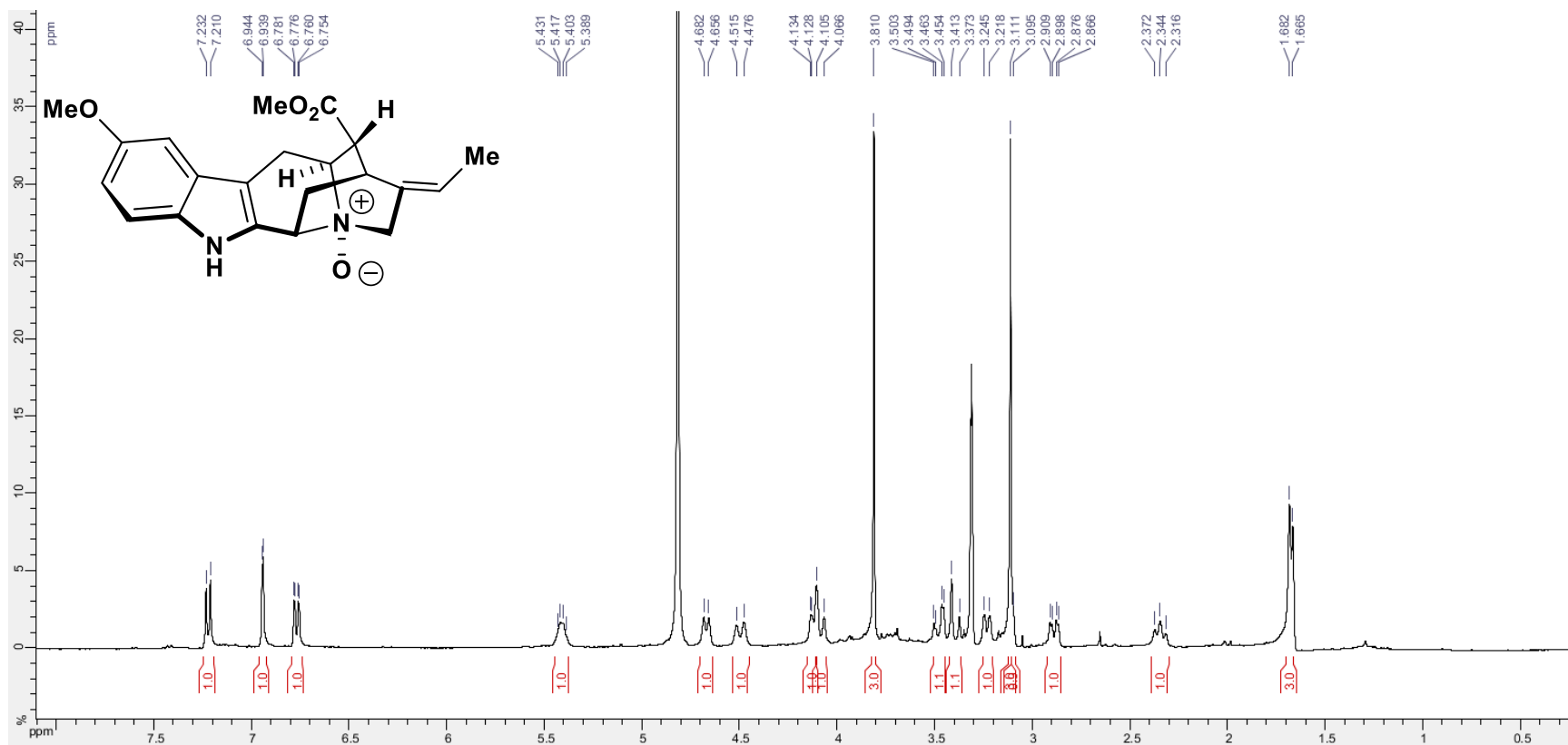


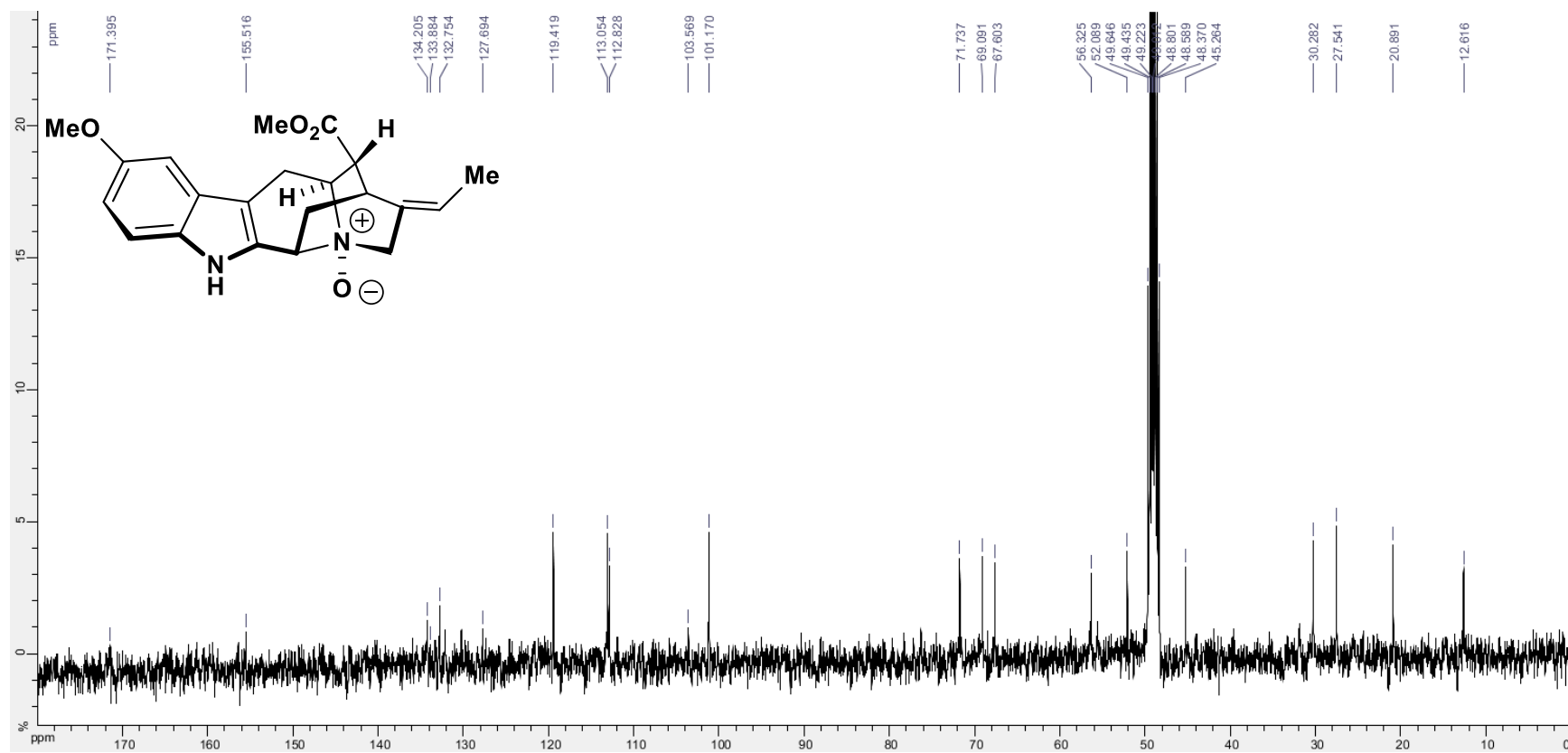
Figure S8. ^{13}C NMR Spectrum (100 MHz, CD_3OD) of 10-methoxypericyclivine *N*-oxide (13)

Figure S9. COSY spectrum (400 MHz, CD₃OD) of 10-methoxypericyclivine *N*-oxide (13)

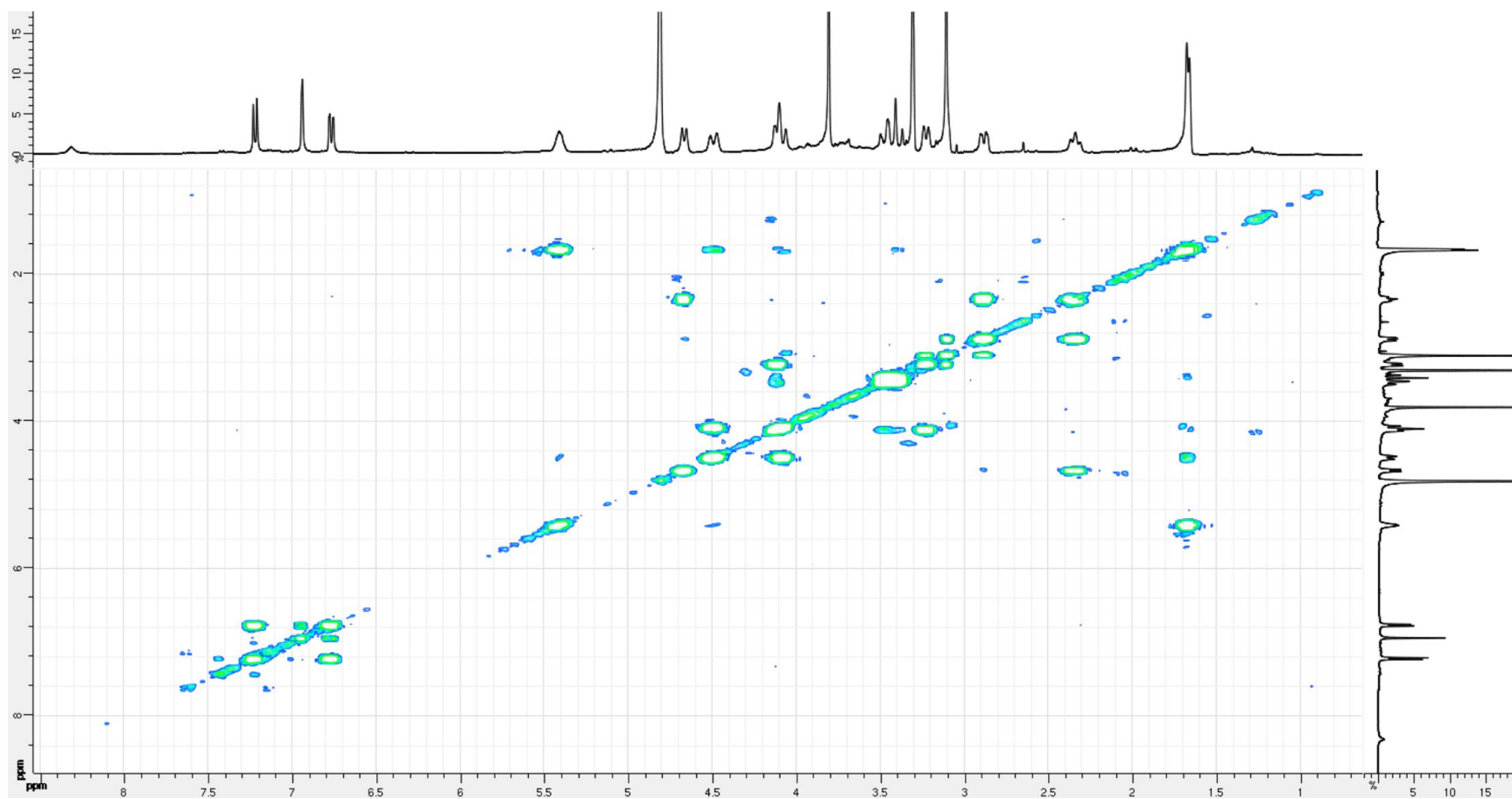


Figure S10. HSQC spectrum (400 MHz, CD₃OD) of 10-methoxypericyclivine *N*-oxide (13)

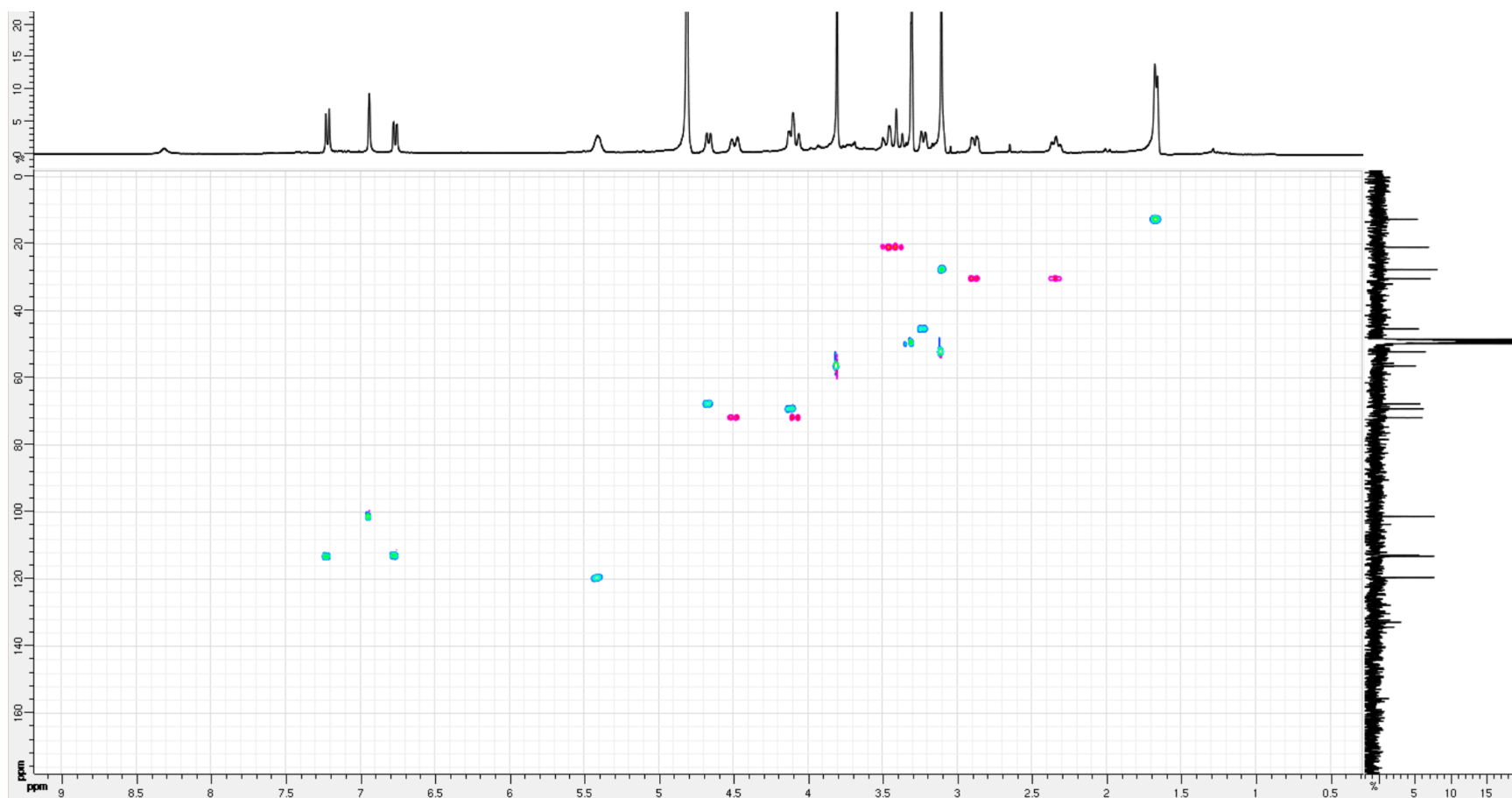


Figure S11. HMBC spectrum (400 MHz, CD₃OD) of 10-methoxypericyclivine *N*-oxide (13)

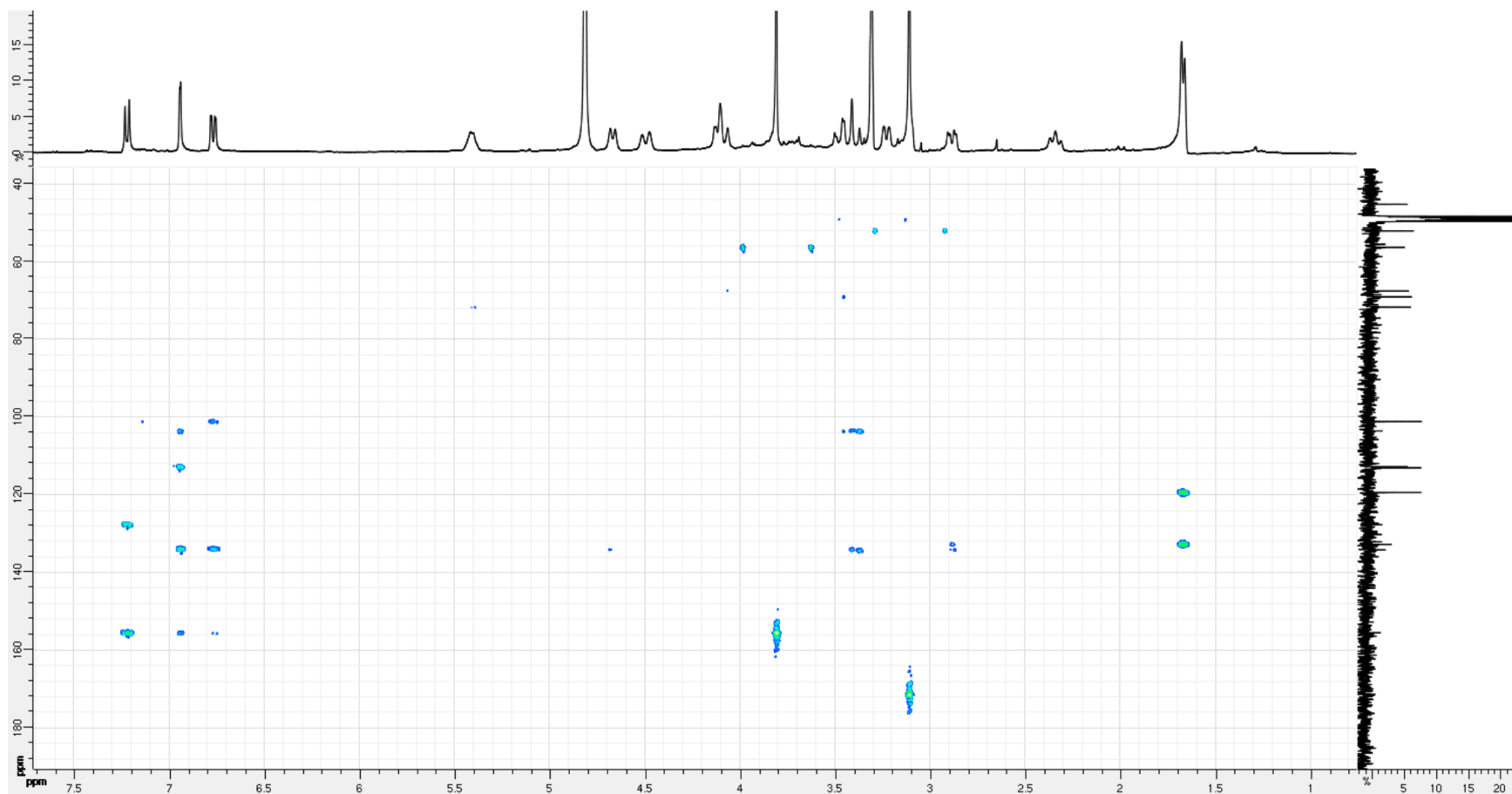


Figure S12. NOESY spectrum (400 MHz, CD₃OD) of 10-methoxypericyclivine *N*-oxide (13)

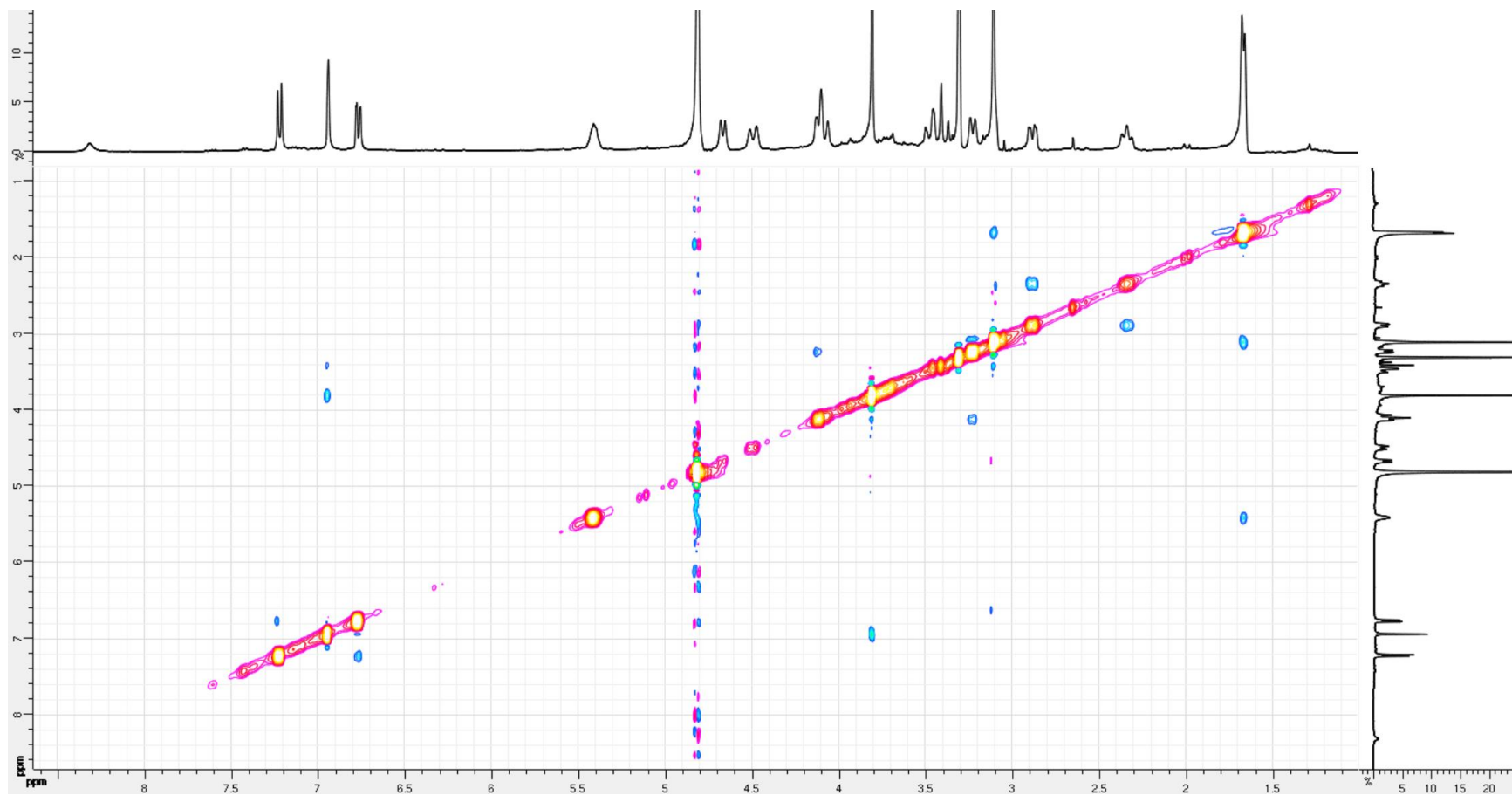


Figure S13. ¹H NMR Spectrum (400 MHz, CD₃OD) of pericyclivine N-oxide (14)

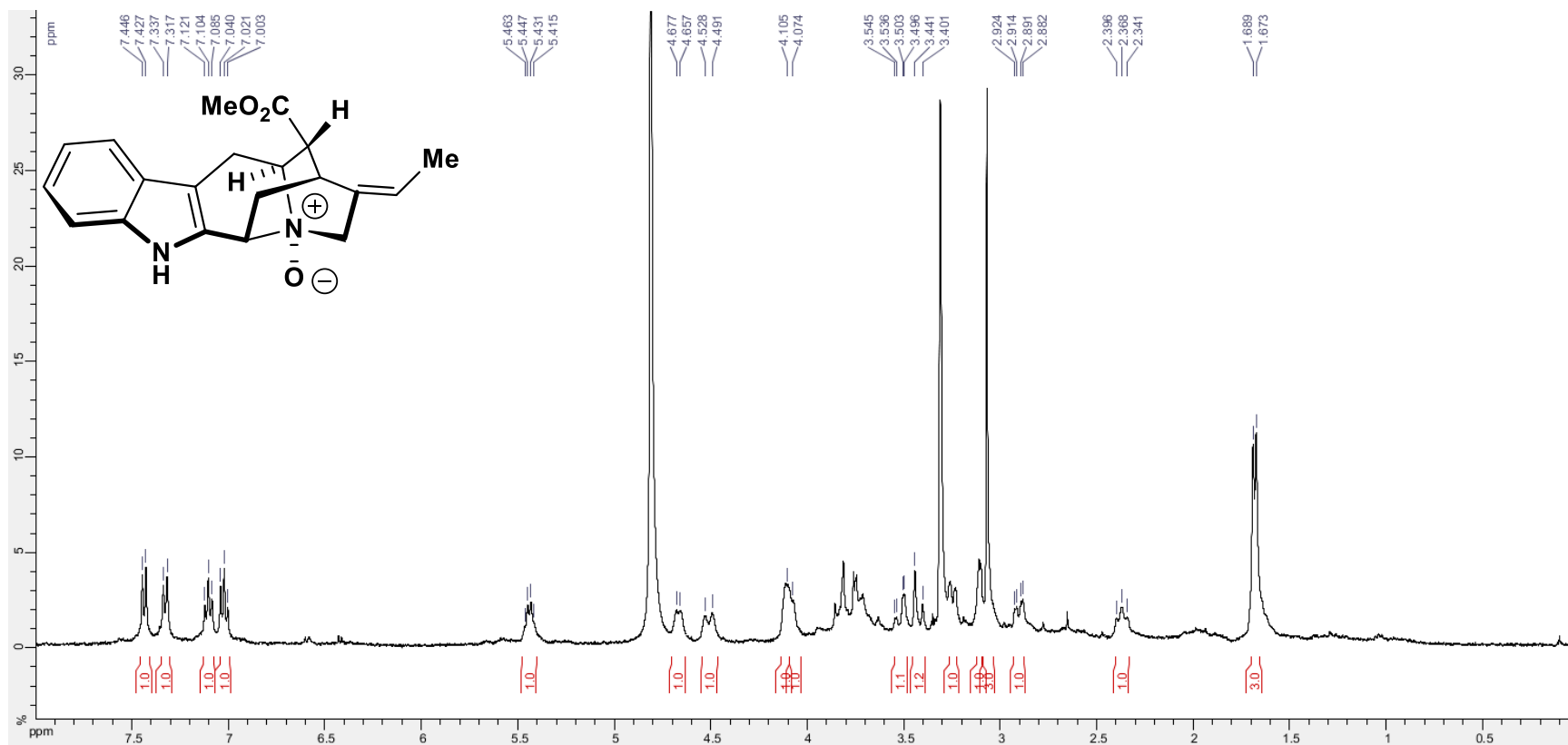


Figure S14. ¹³C NMR Spectrum (100 MHz, CD₃OD) of pericyclivine *N*-oxide (14)

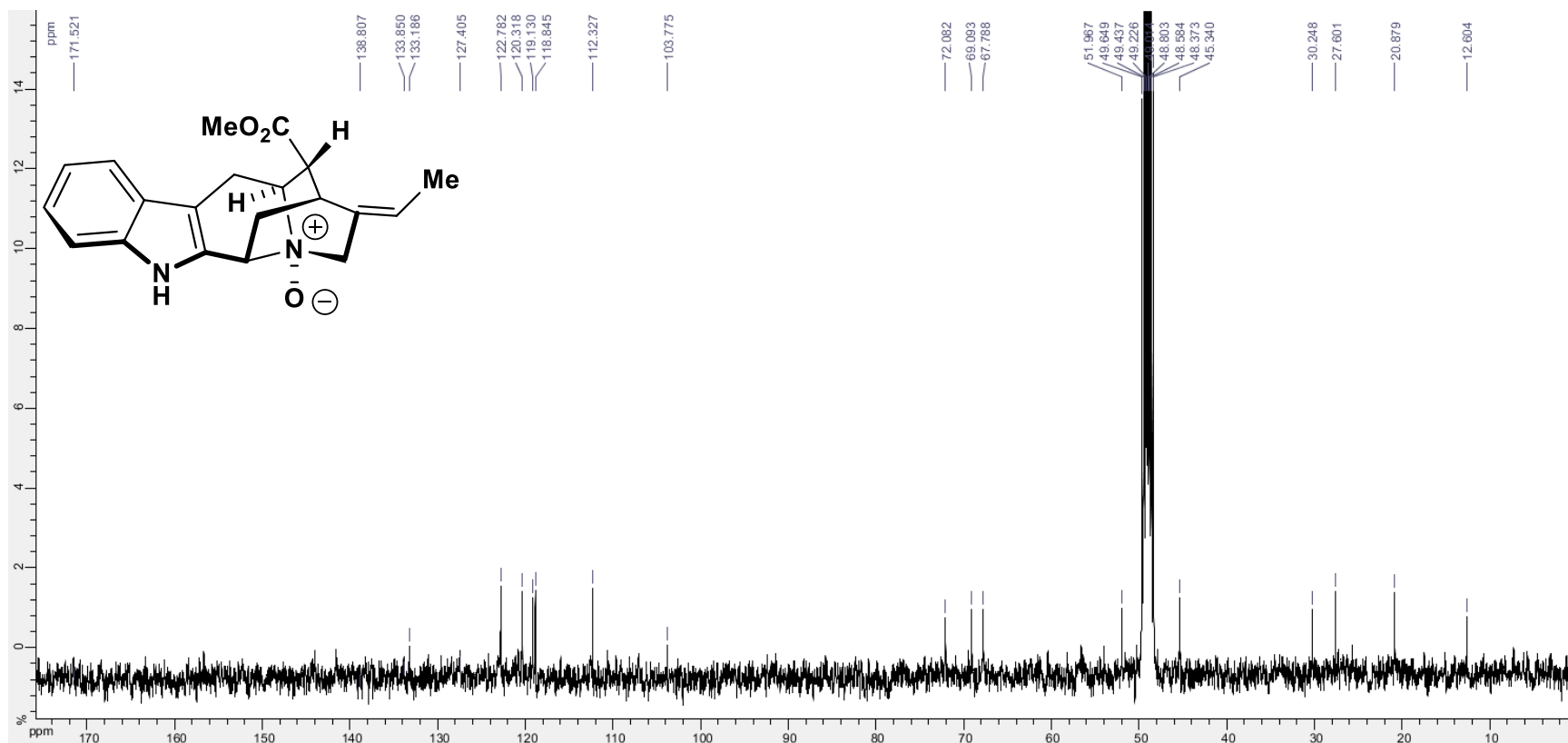


Figure S15. COSY spectrum (400 MHz, CD₃OD) of pericyclivine *N*-oxide (14)

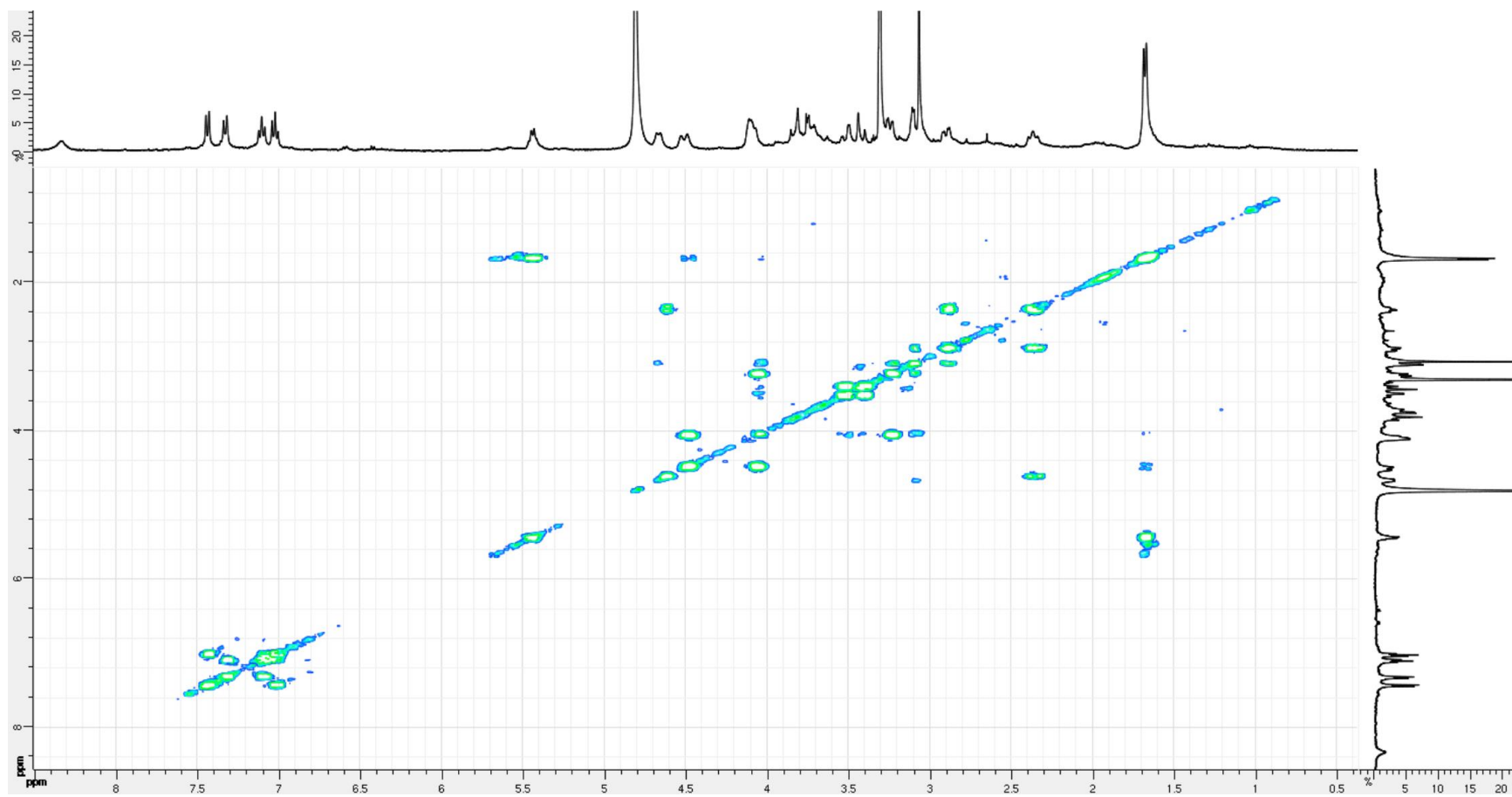


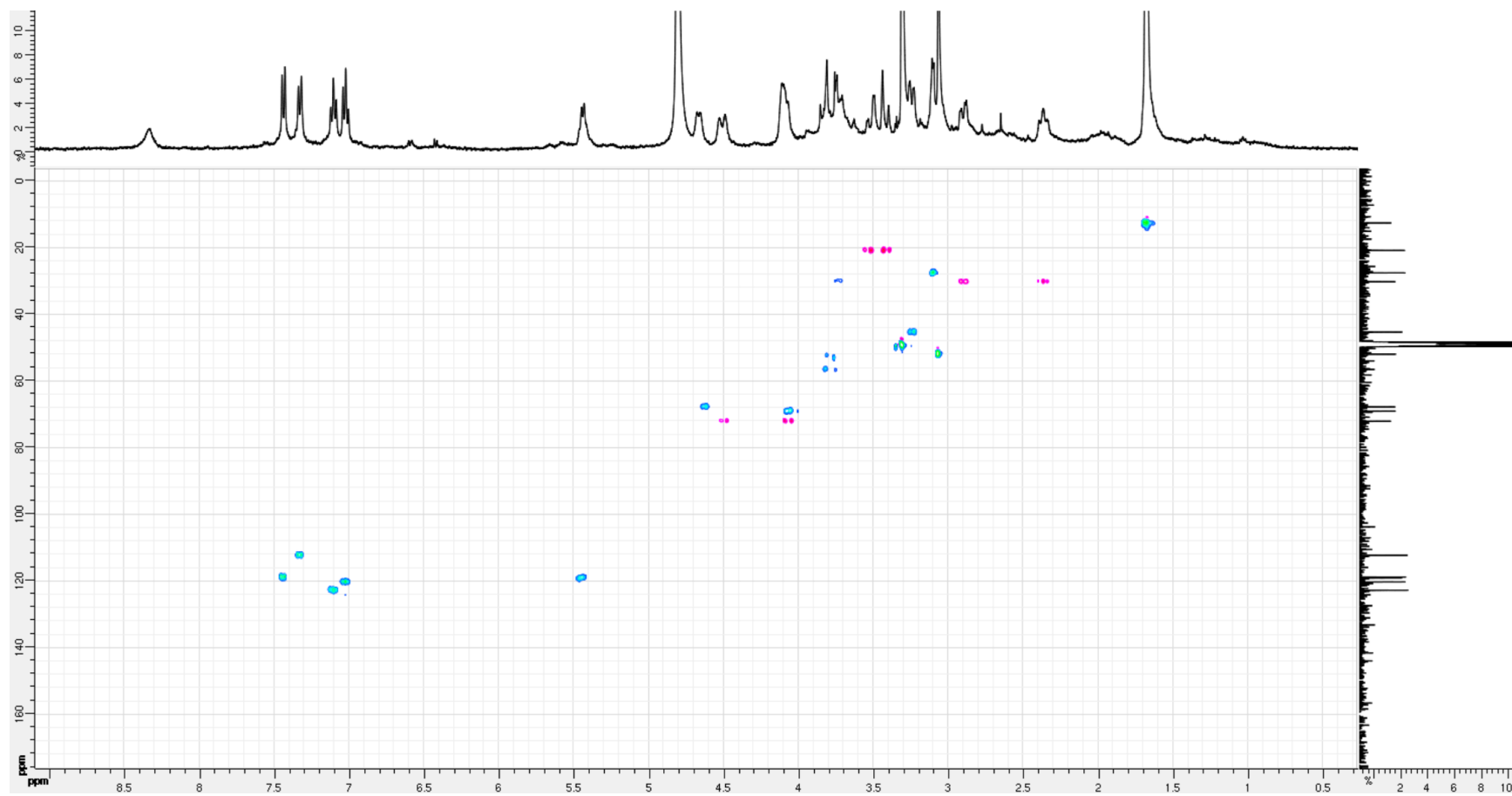
Figure S16. HSQC spectrum (400 MHz, CD₃OD) of pericyclivine *N*-oxide (14)

Figure S17. HMBC spectrum (400 MHz, CD₃OD) of pericyclivine *N*-oxide (14)

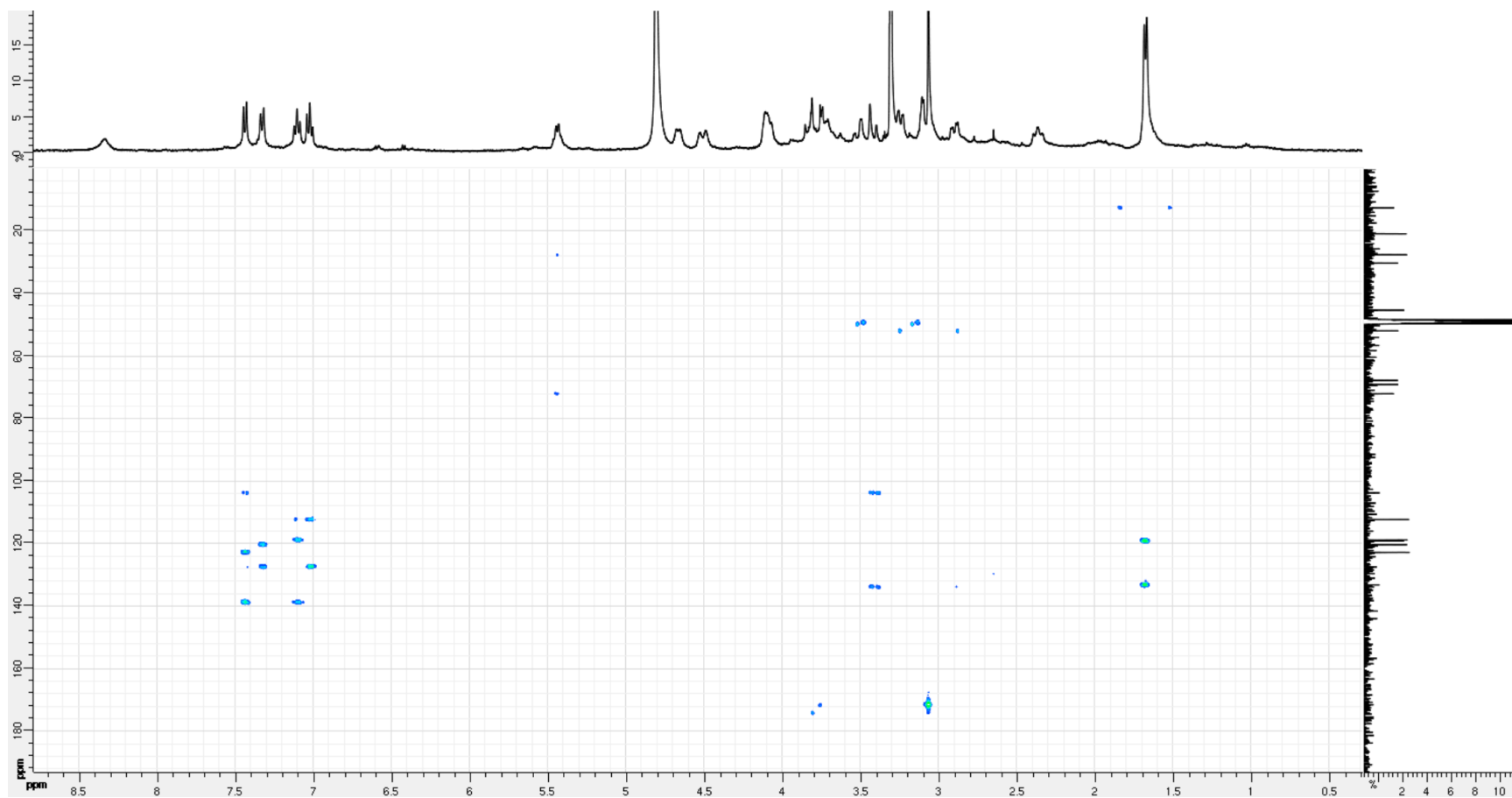


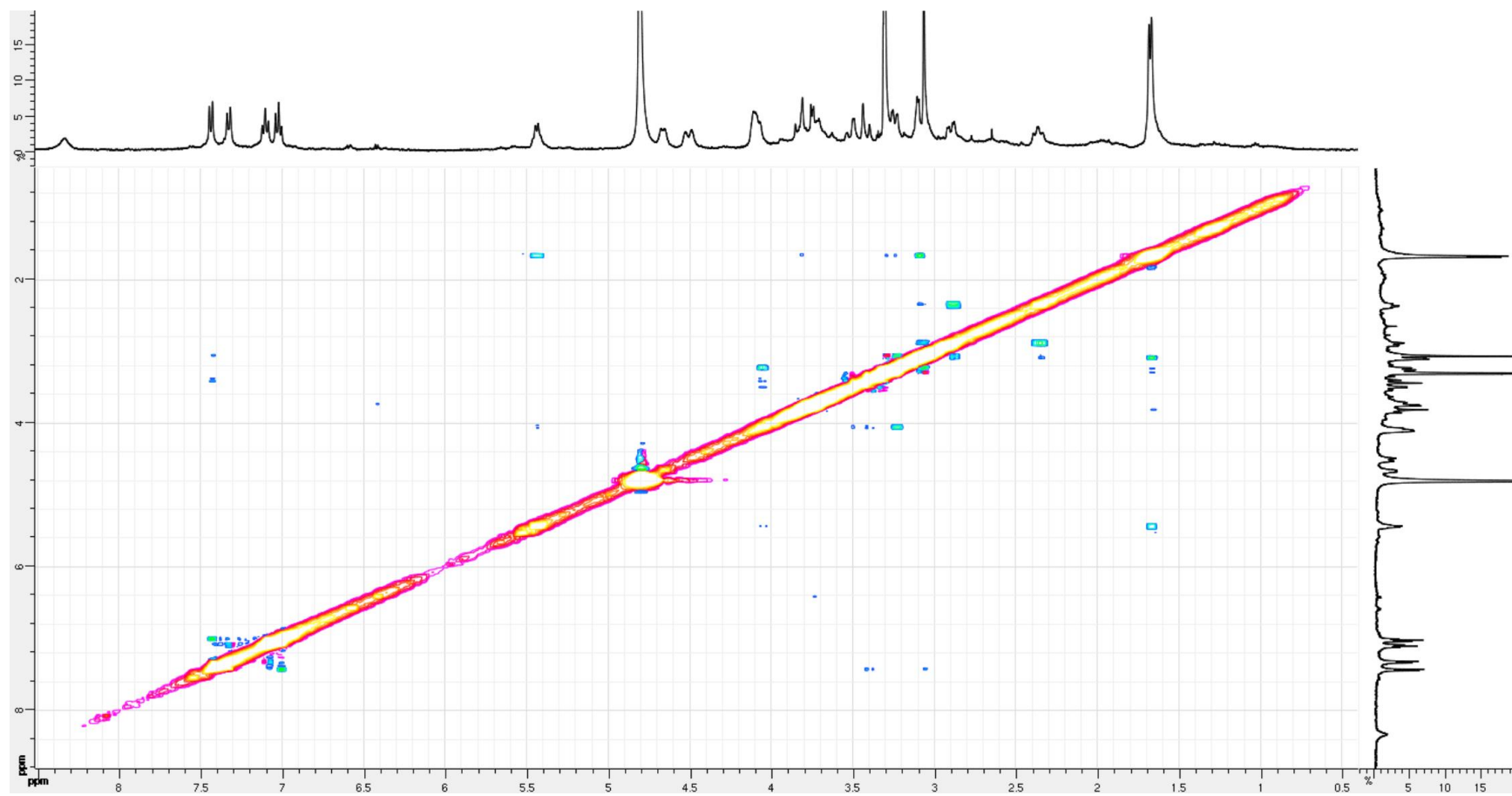
Figure S18. NOESY spectrum (400 MHz, CD₃OD) of pericyclivine *N*-oxide (14)

Figure S19. ¹H NMR Spectrum (400 MHz, CD₃OD) of *N*-methyl-10-methoxy-16-epi-pericyclivine *N*-oxide (15)

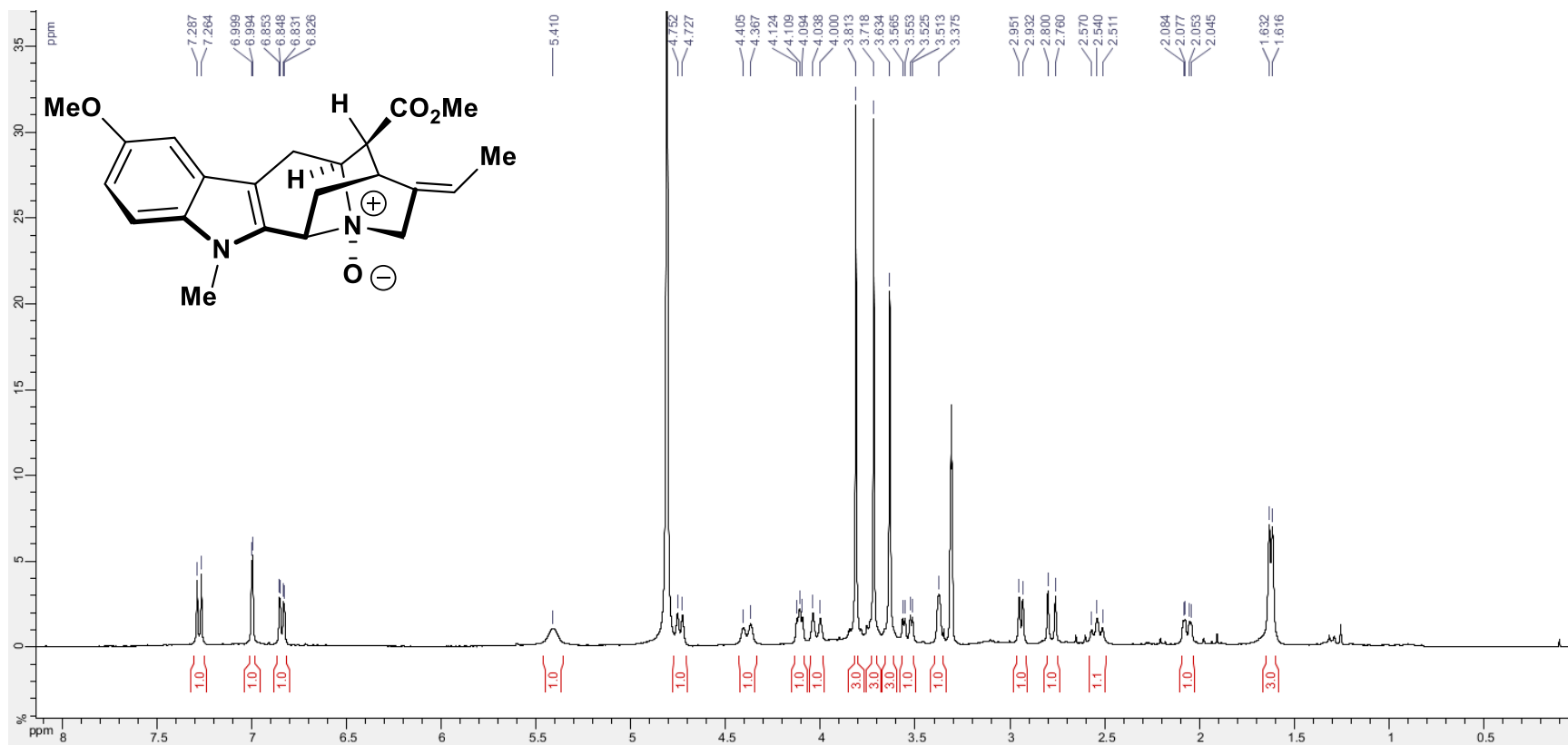


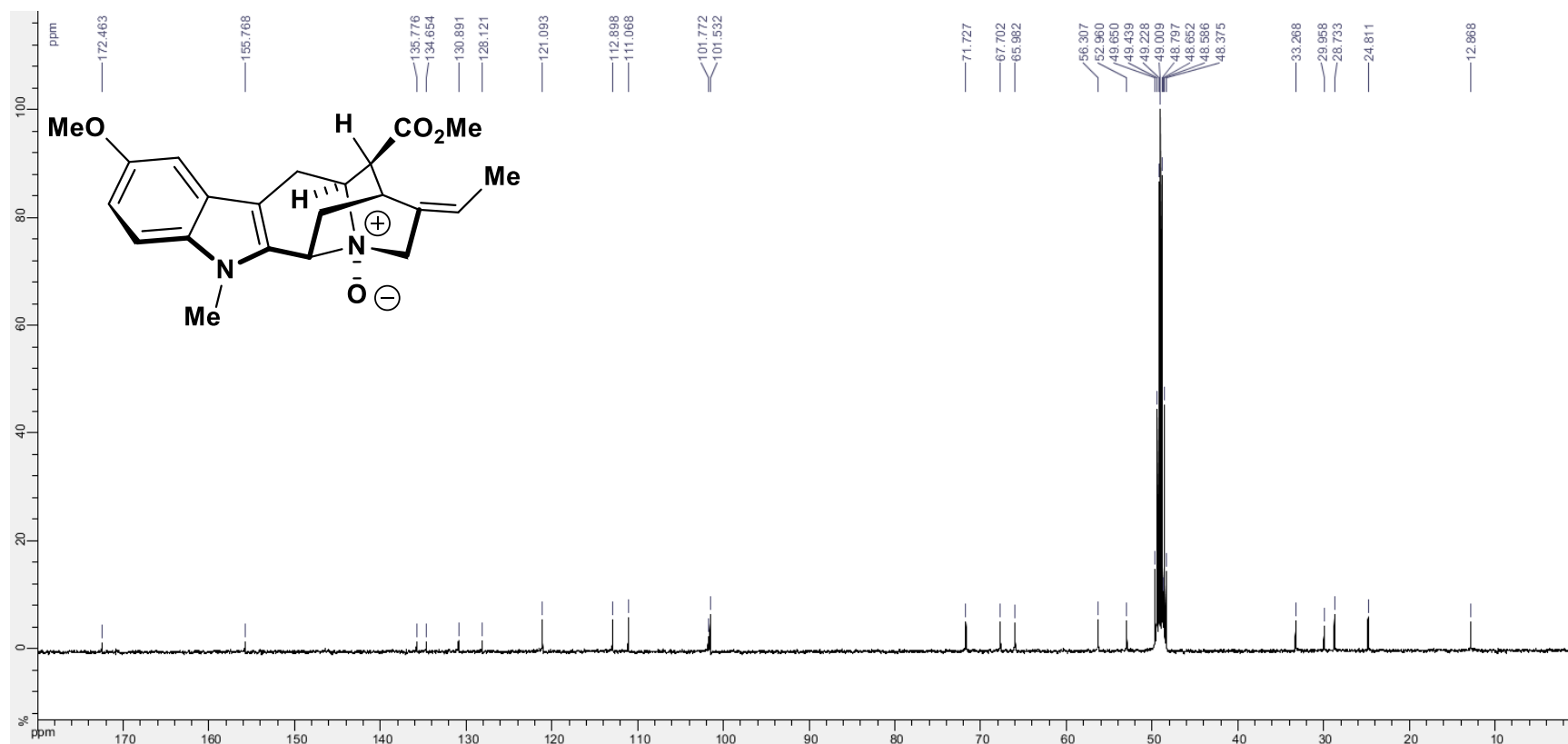
Figure S20. ^{13}C NMR Spectrum (100 MHz, CD_3OD) of *N*-methyl-10-methoxy-16-epi-pericyclivine *N*-oxide (15)

Figure S21. COSY spectrum (400 MHz, CD₃OD) of *N*-methyl-10-methoxy-16-*epi*-pericyclivine *N*-oxide (15)

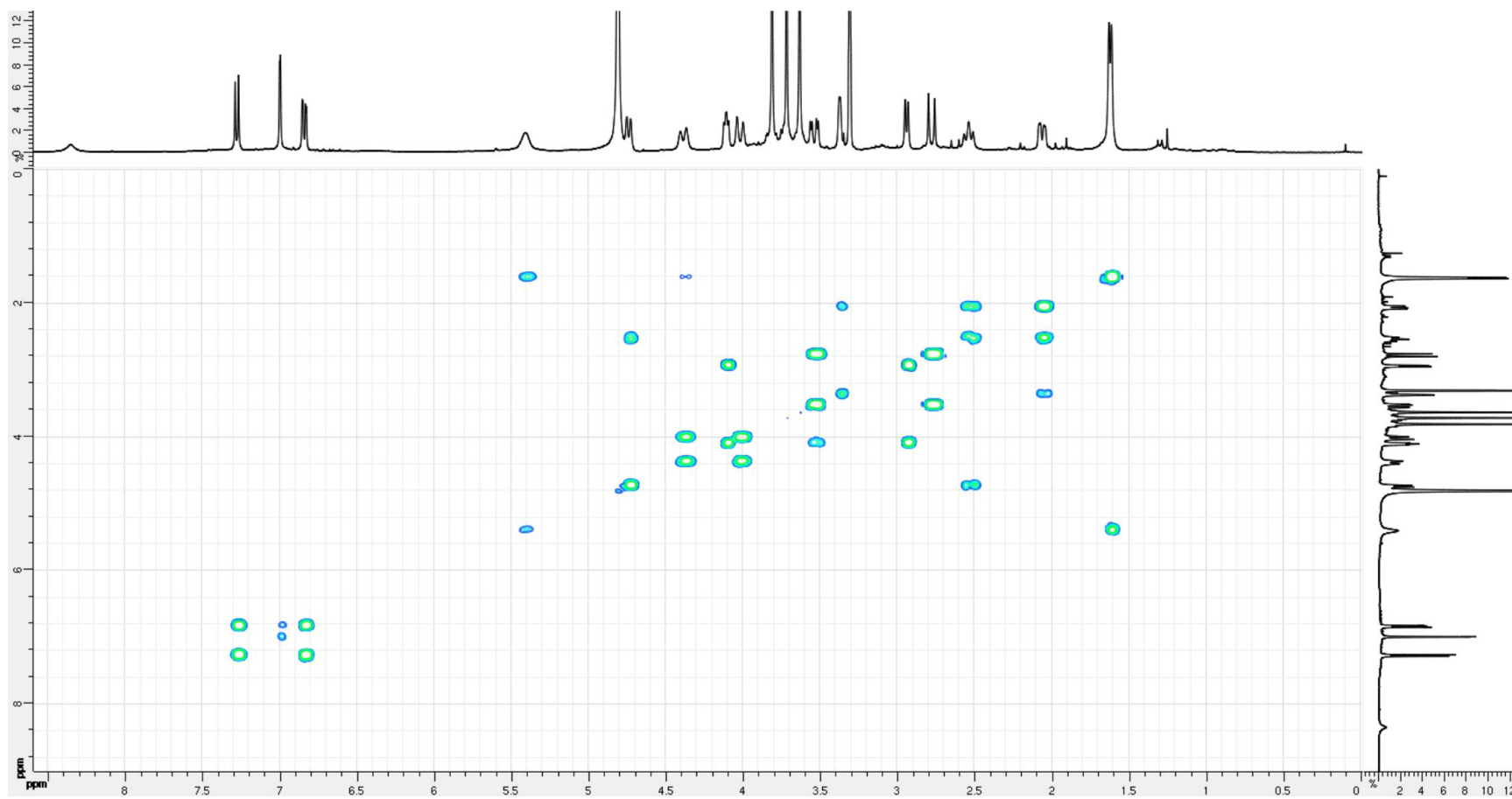


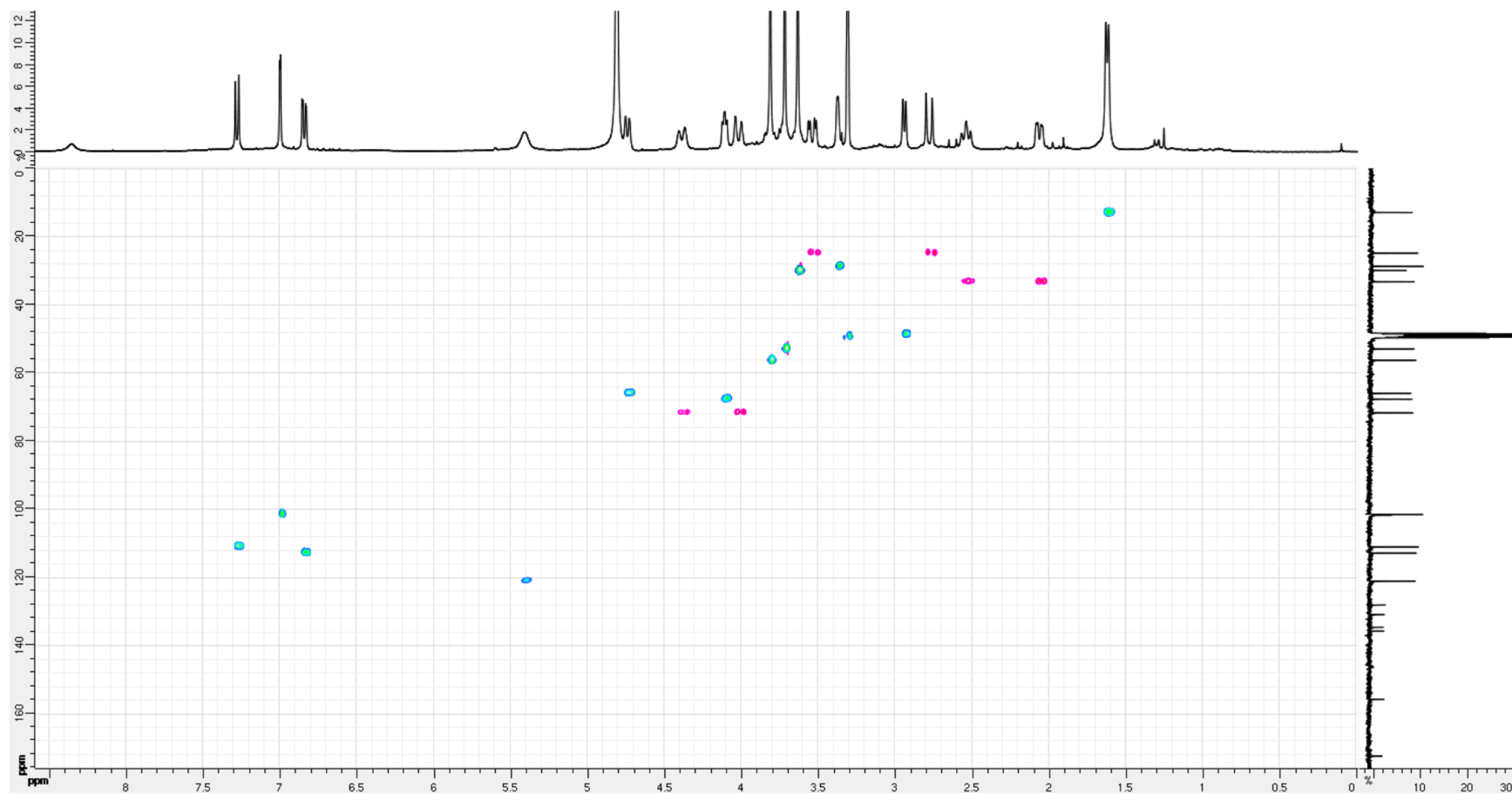
Figure S22. HSQC spectrum (400 MHz, CD₃OD) of *N*-methyl-10-methoxy-16-*epi*-pericyclivine *N*-oxide (15)

Figure S23. HMBC spectrum (400 MHz, CD₃OD) of *N*-methyl-10-methoxy-16-*epi*-pericyclivine *N*-oxide (15)

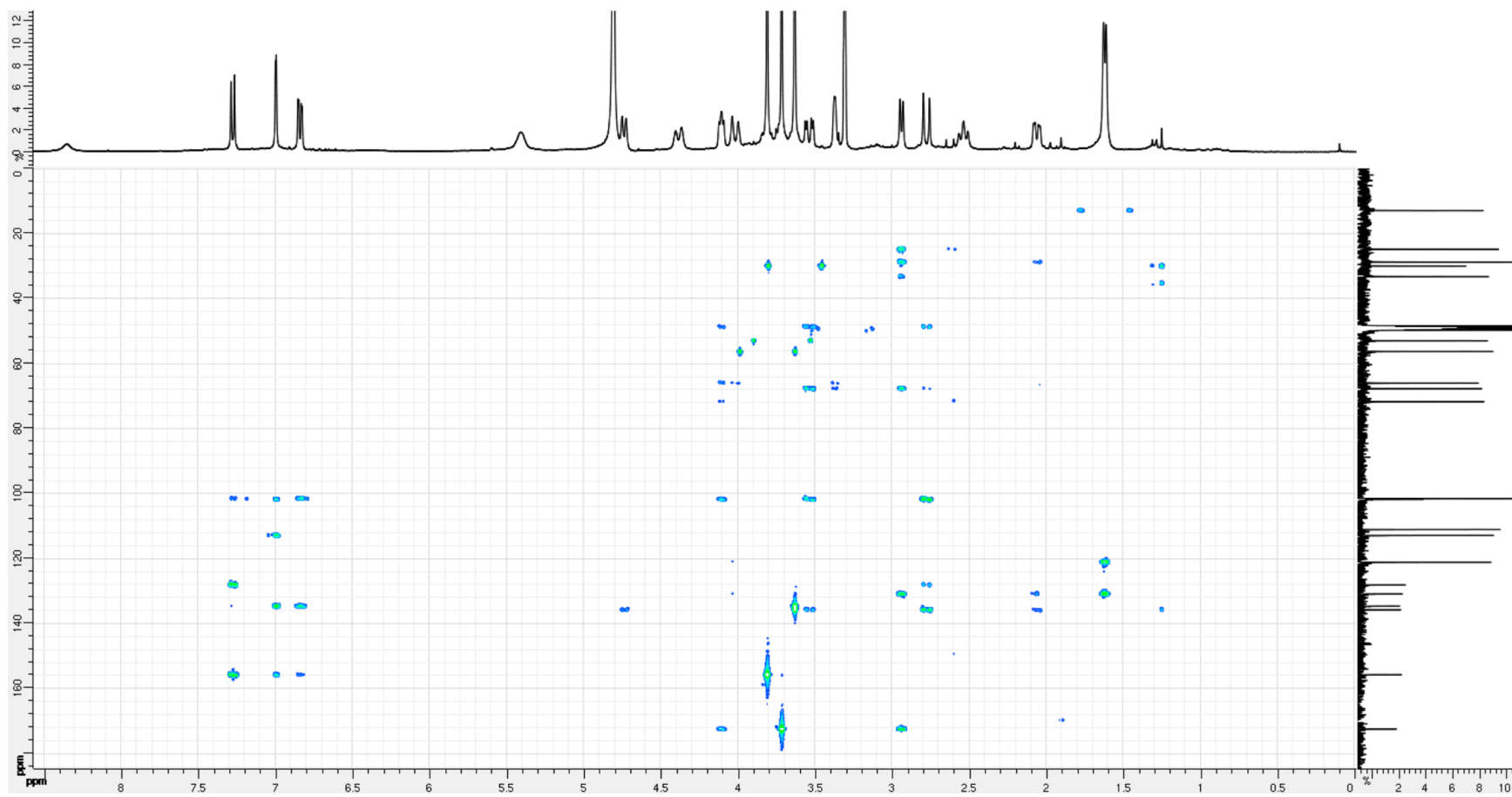


Figure S24. NOESY spectrum (400 MHz, CD₃OD) of *N*-methyl-10-methoxy-16-*epi*-pericyclivine *N*-oxide (15)

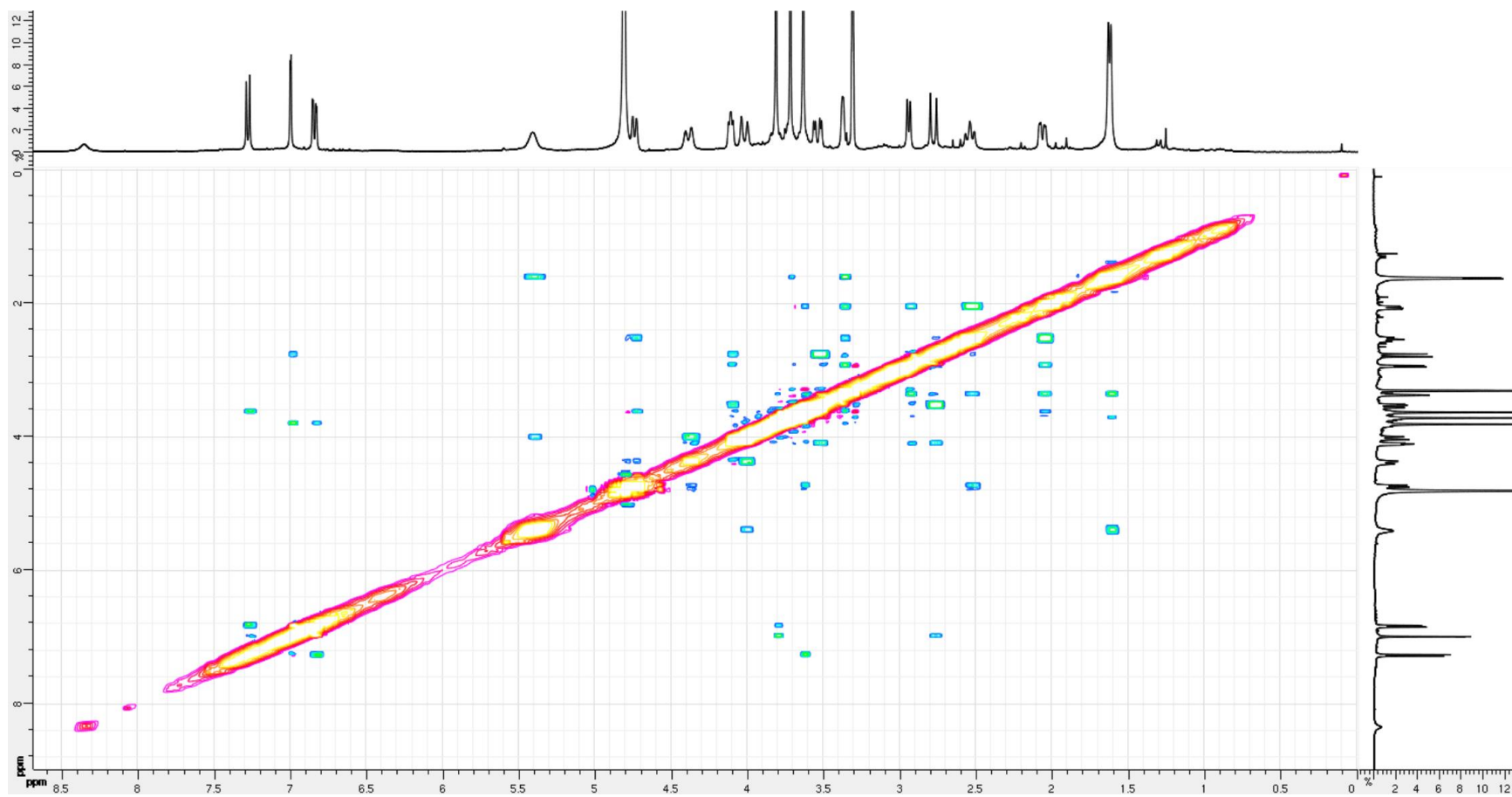


Figure S25. ¹H NMR Spectrum (400 MHz, CD₃OD) of *N*-methyl-16-*epi*-pericyclivine *N*-oxide (16)

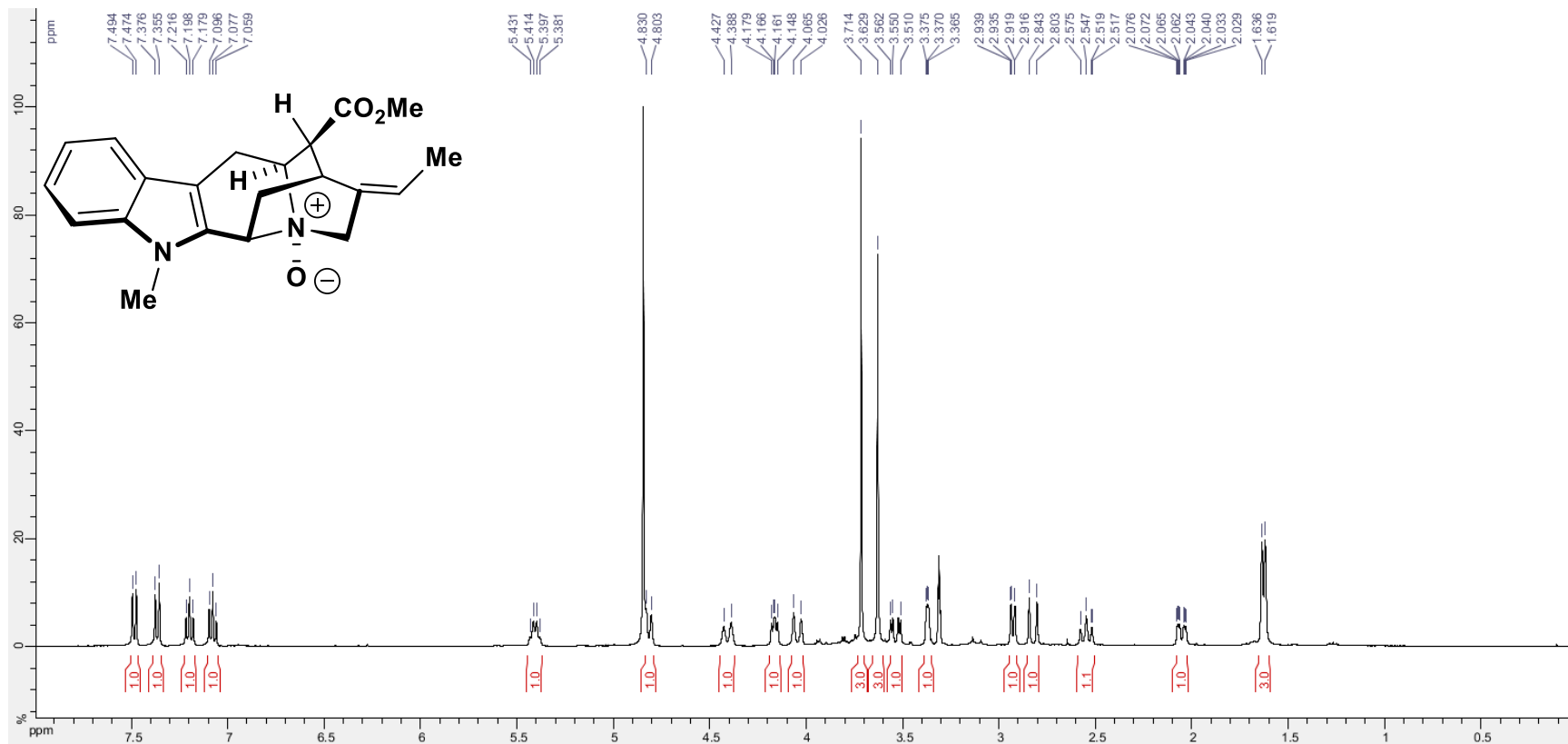


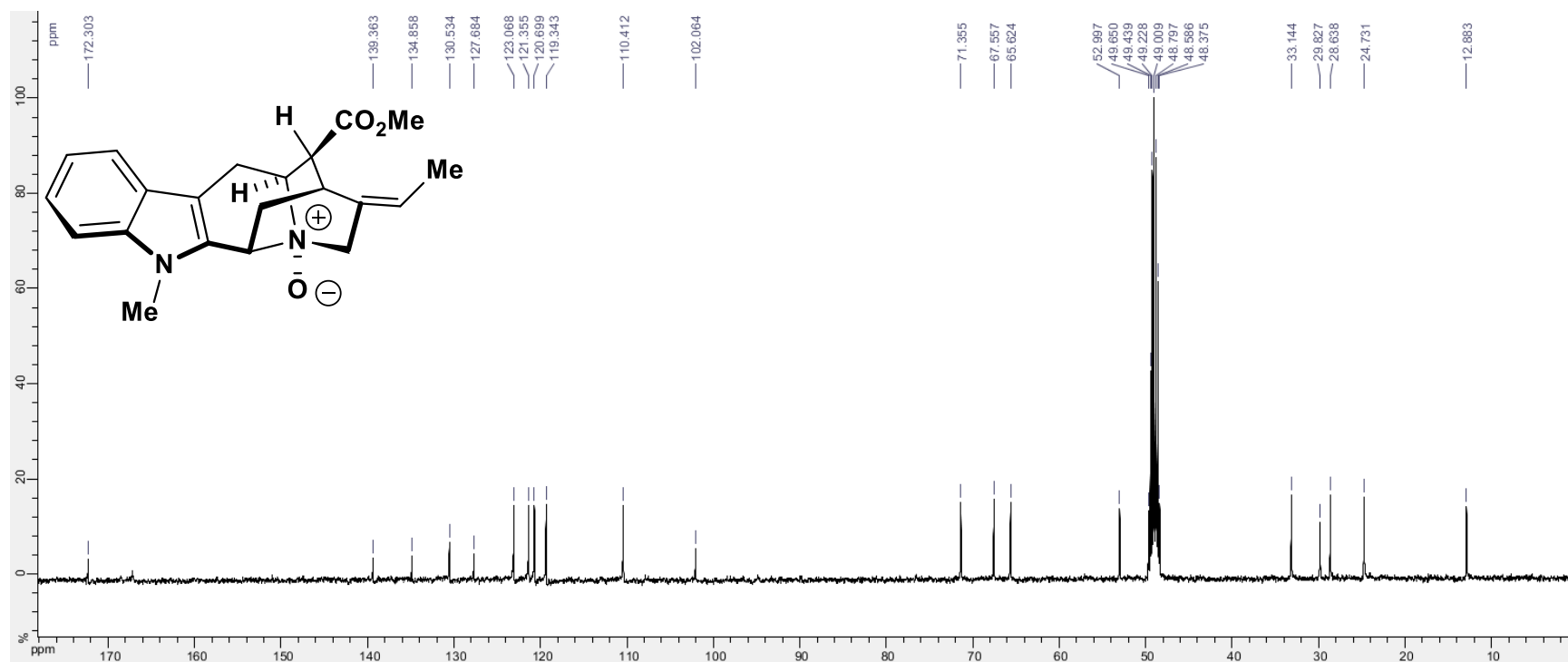
Figure S26. ^{13}C NMR Spectrum (100 MHz, CD_3OD) of *N*-methyl-16-*epi*-pericyclivine *N*-oxide (16)

Figure S27. COSY spectrum (400 MHz, CD₃OD) of *N*-methyl-16-*epi*-pericyclivine *N*-oxide (16)

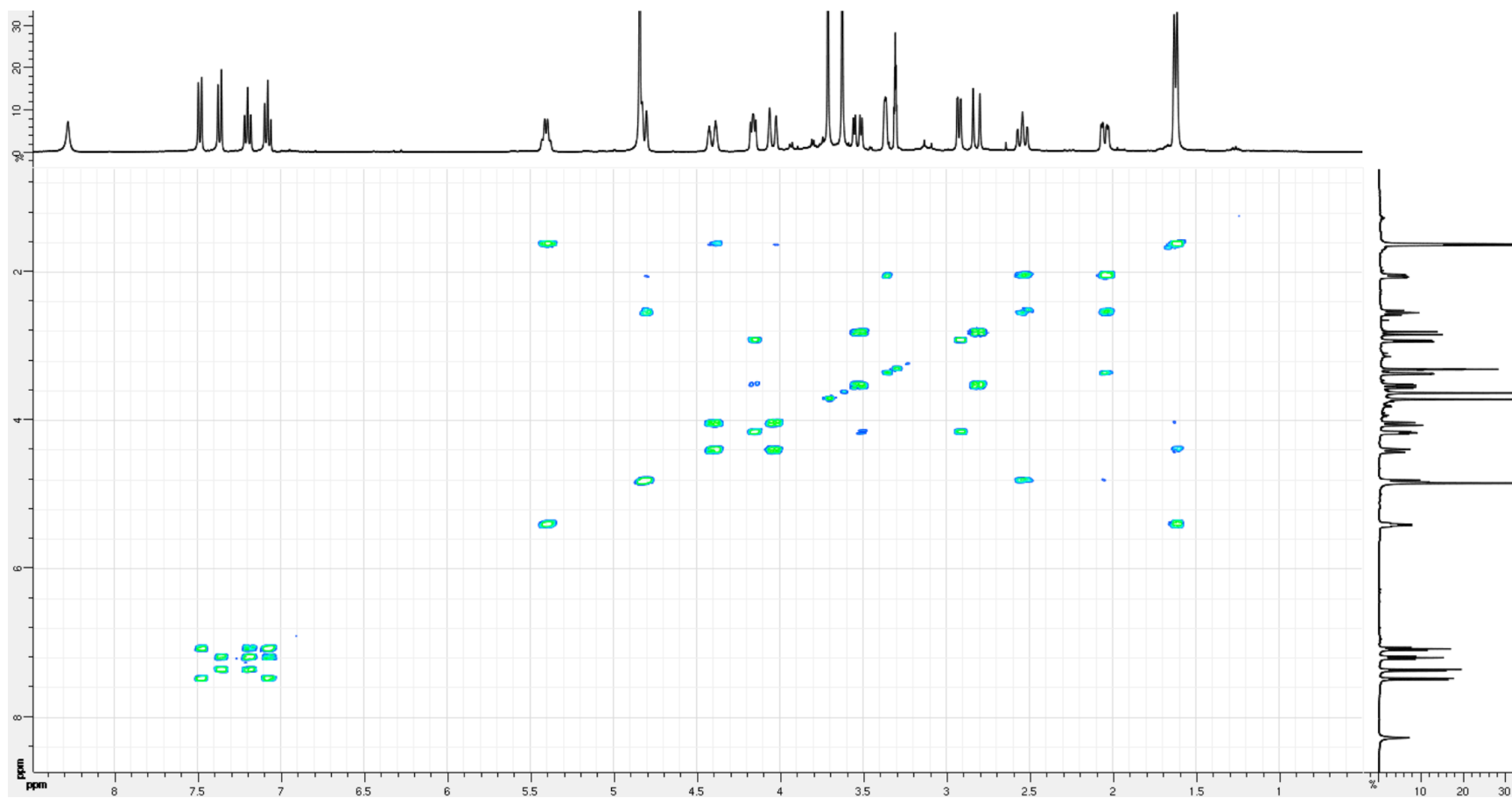


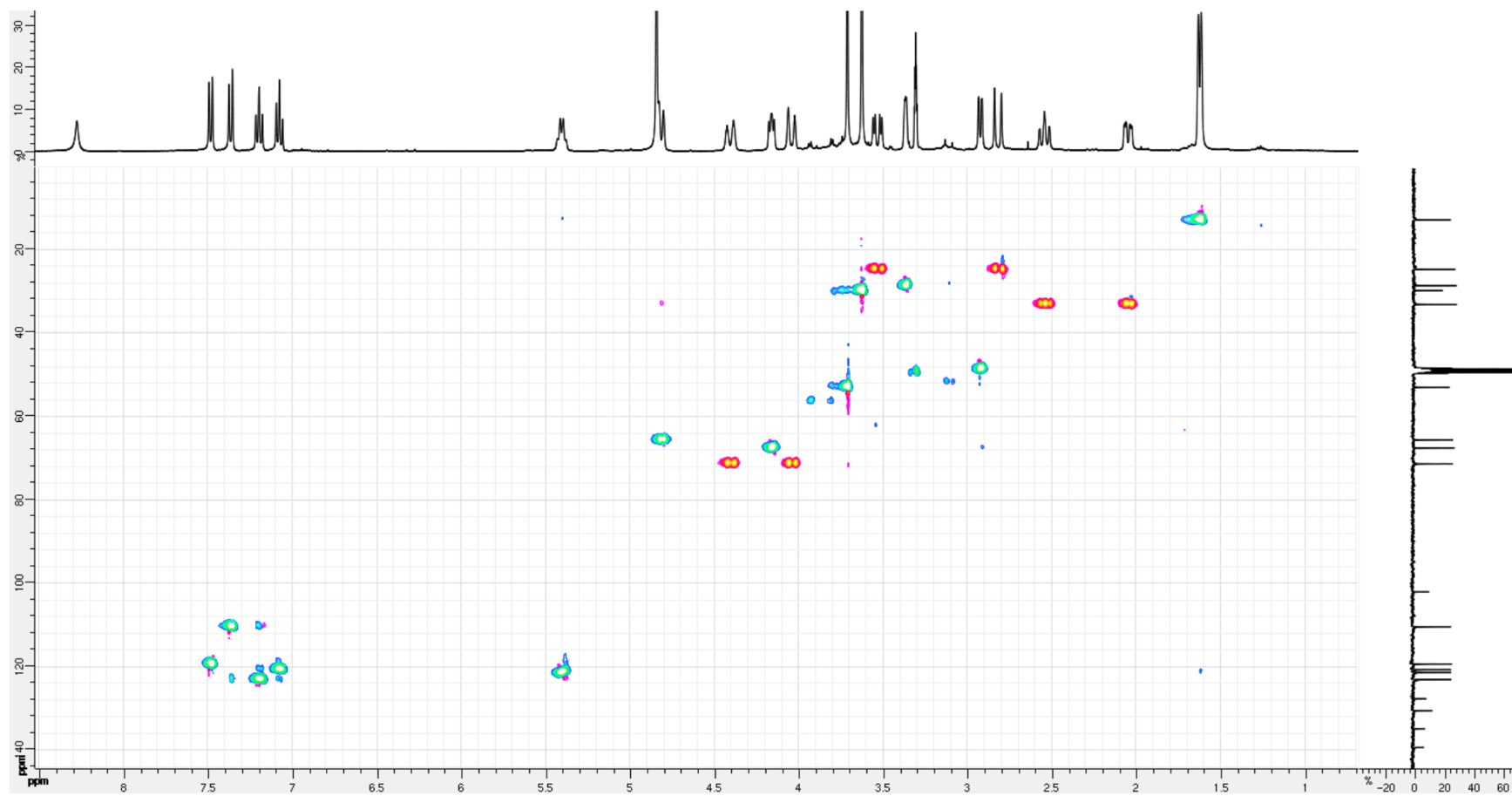
Figure S28. HSQC spectrum (400 MHz, CD₃OD) of *N*-methyl-16-*epi*-pericyclivine *N*-oxide (16)

Figure S29. HMBC spectrum (400 MHz, CD₃OD) of *N*-methyl-16-*epi*-pericyclivine *N*-oxide (16)

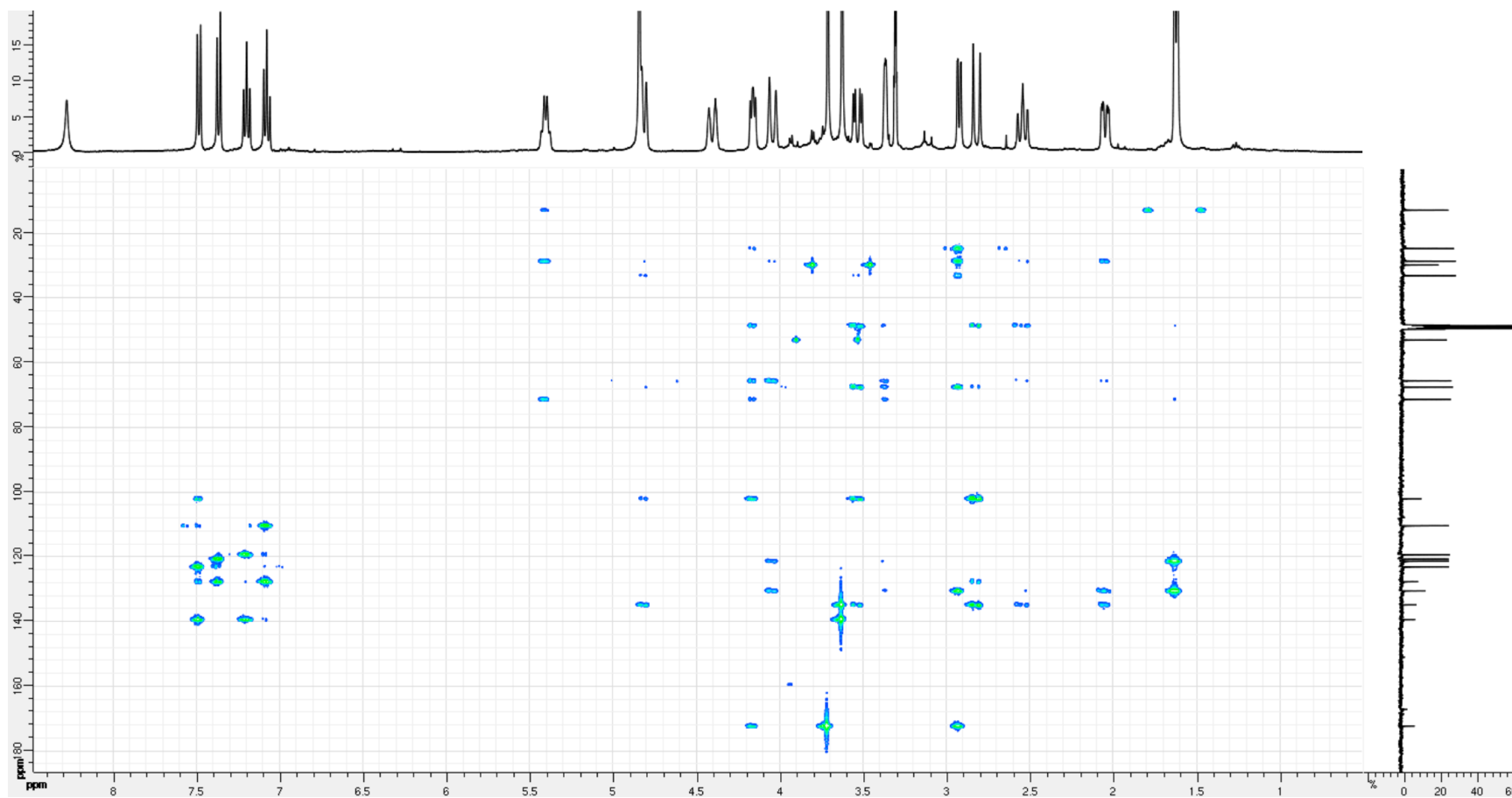


Figure S30. NOESY spectrum (400 MHz, CD₃OD) of *N*-methyl-16-*epi*-pericyclivine *N*-oxide (16)

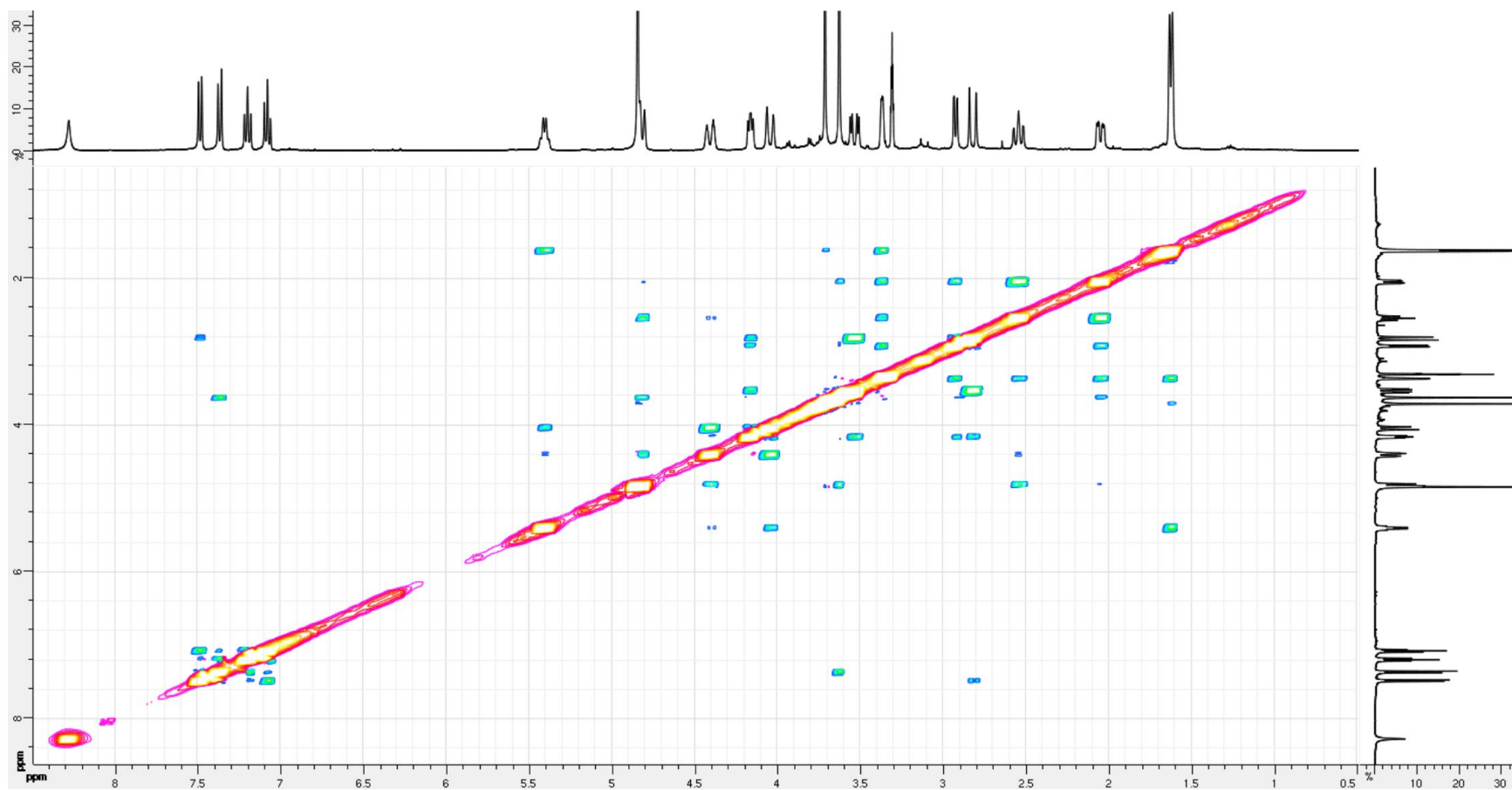


Figure S31. ¹H NMR Spectrum (400 MHz, CD₃OD) of 10-hydroxypericyclivine N-oxide (17)

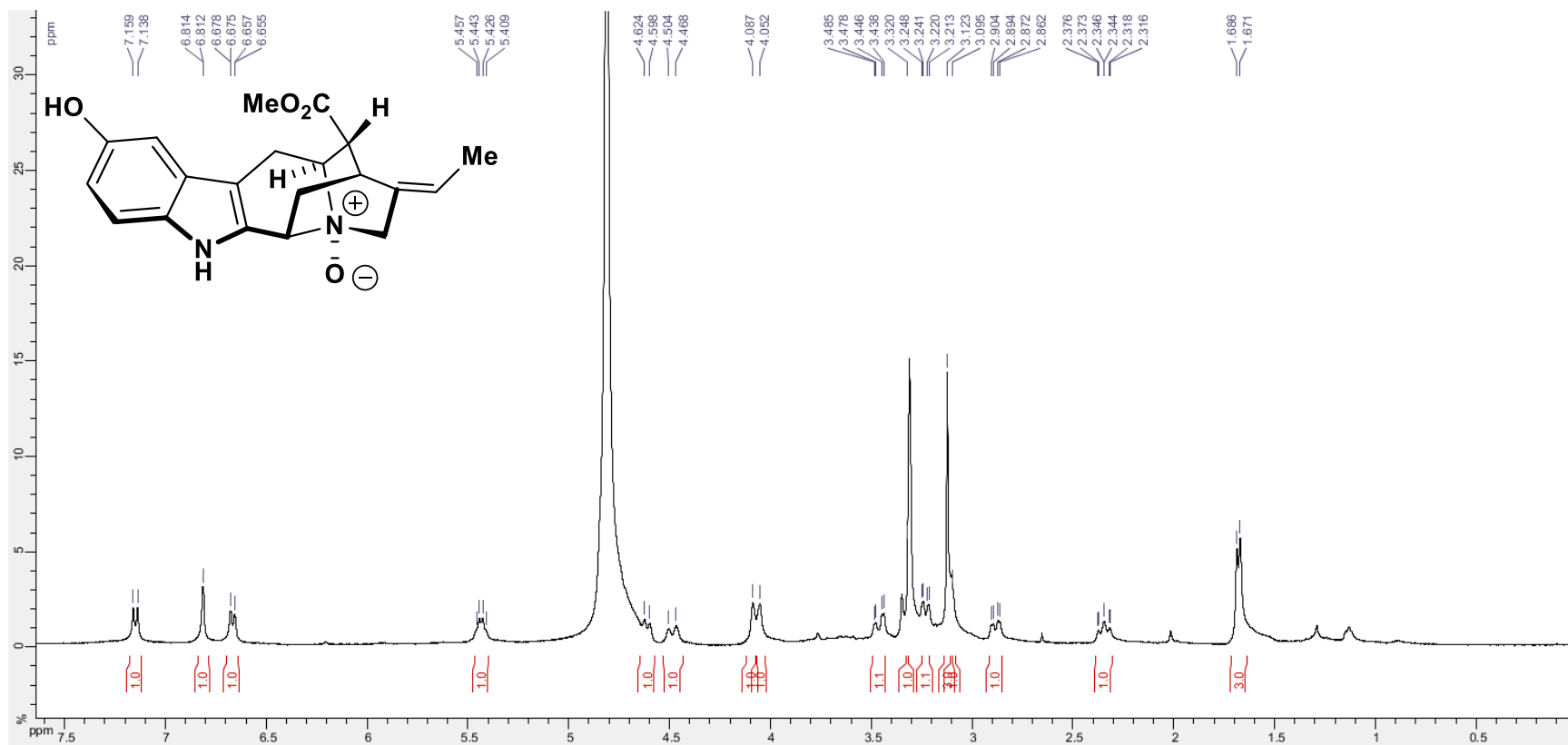


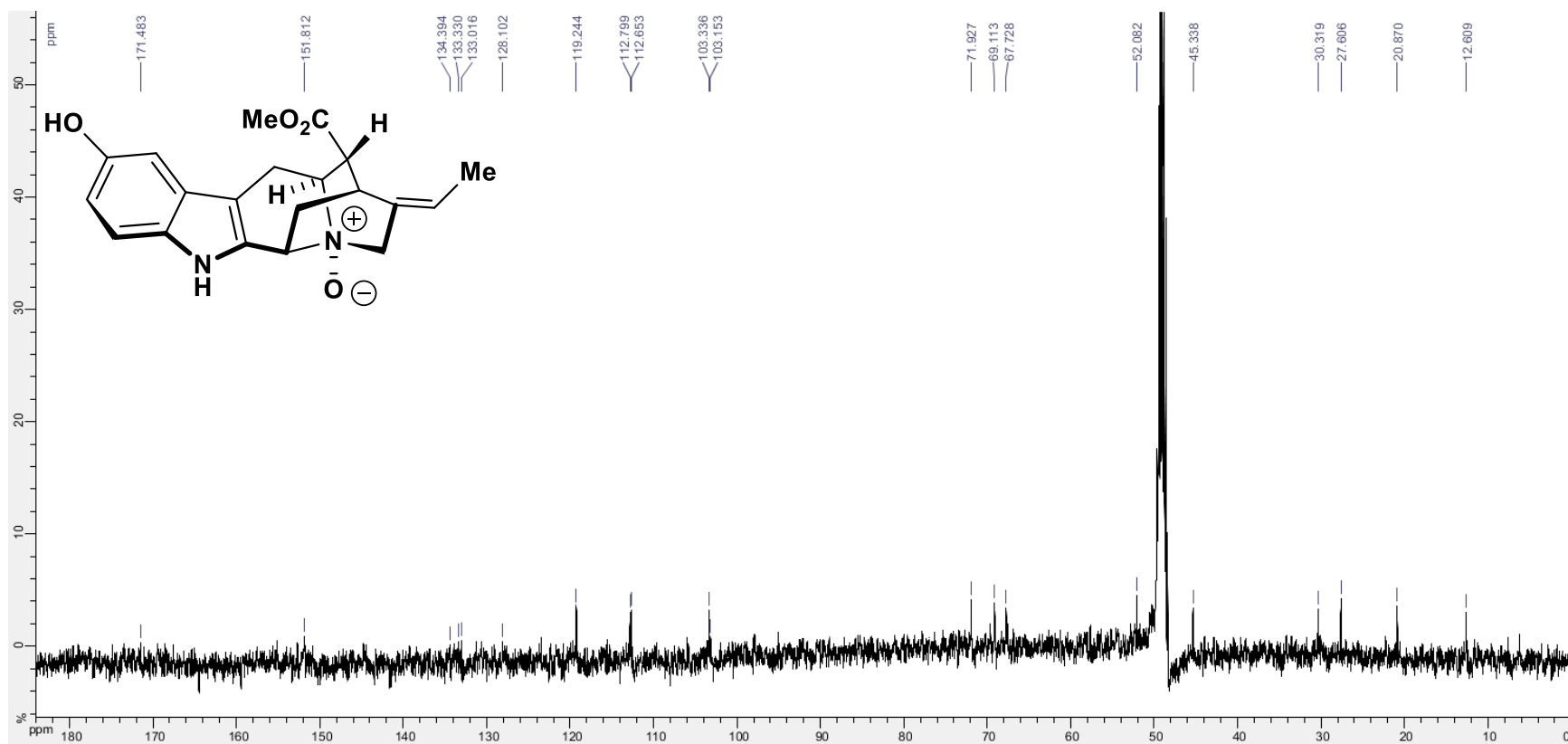
Figure S32. ^{13}C NMR Spectrum (100 MHz, CD_3OD) of 10-hydropericyclivine *N*-oxide (17)

Figure S33. COSY spectrum (400 MHz, CD₃OD) of 10-hydroxypericyclivine *N*-oxide (17)

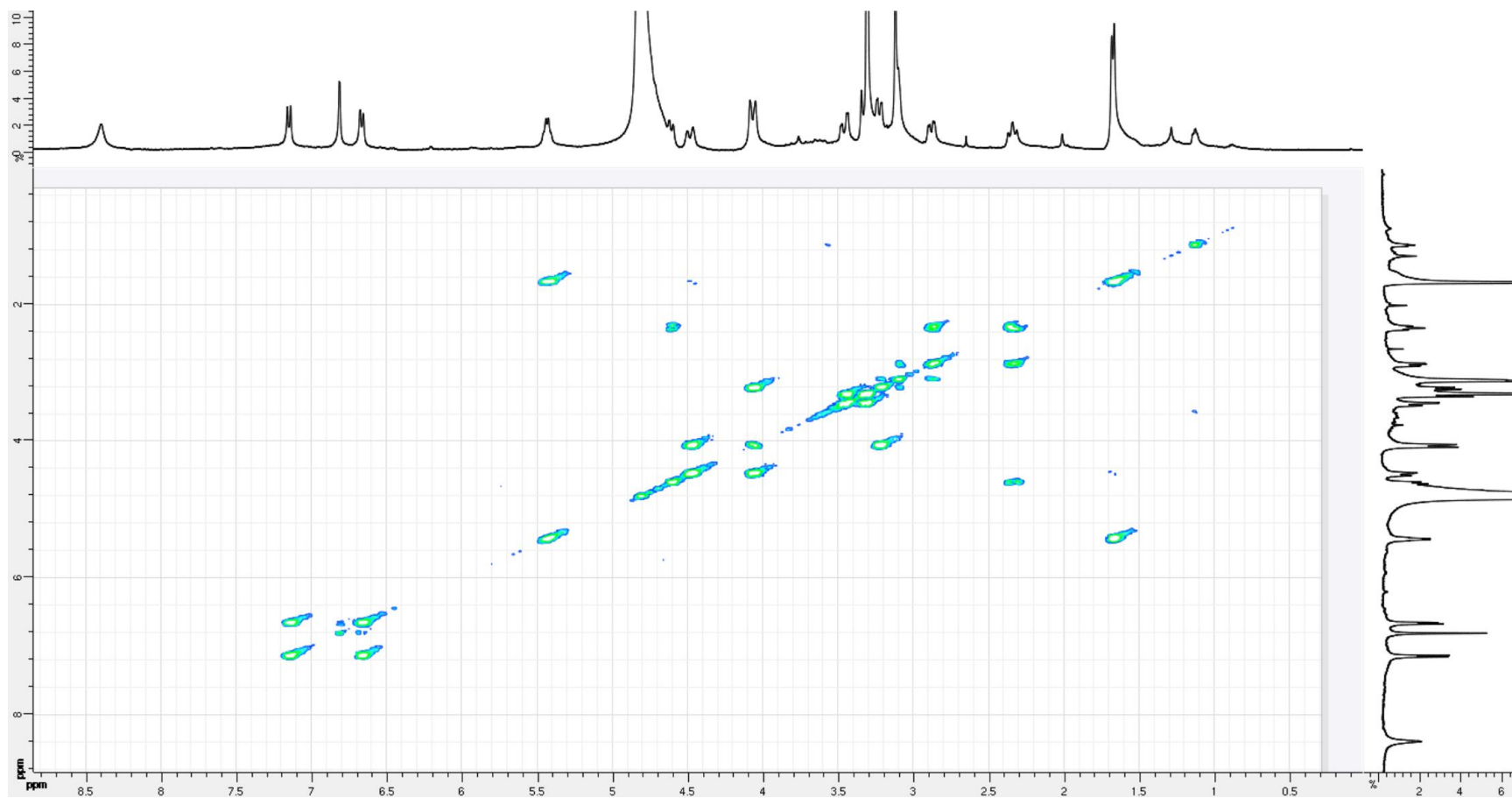


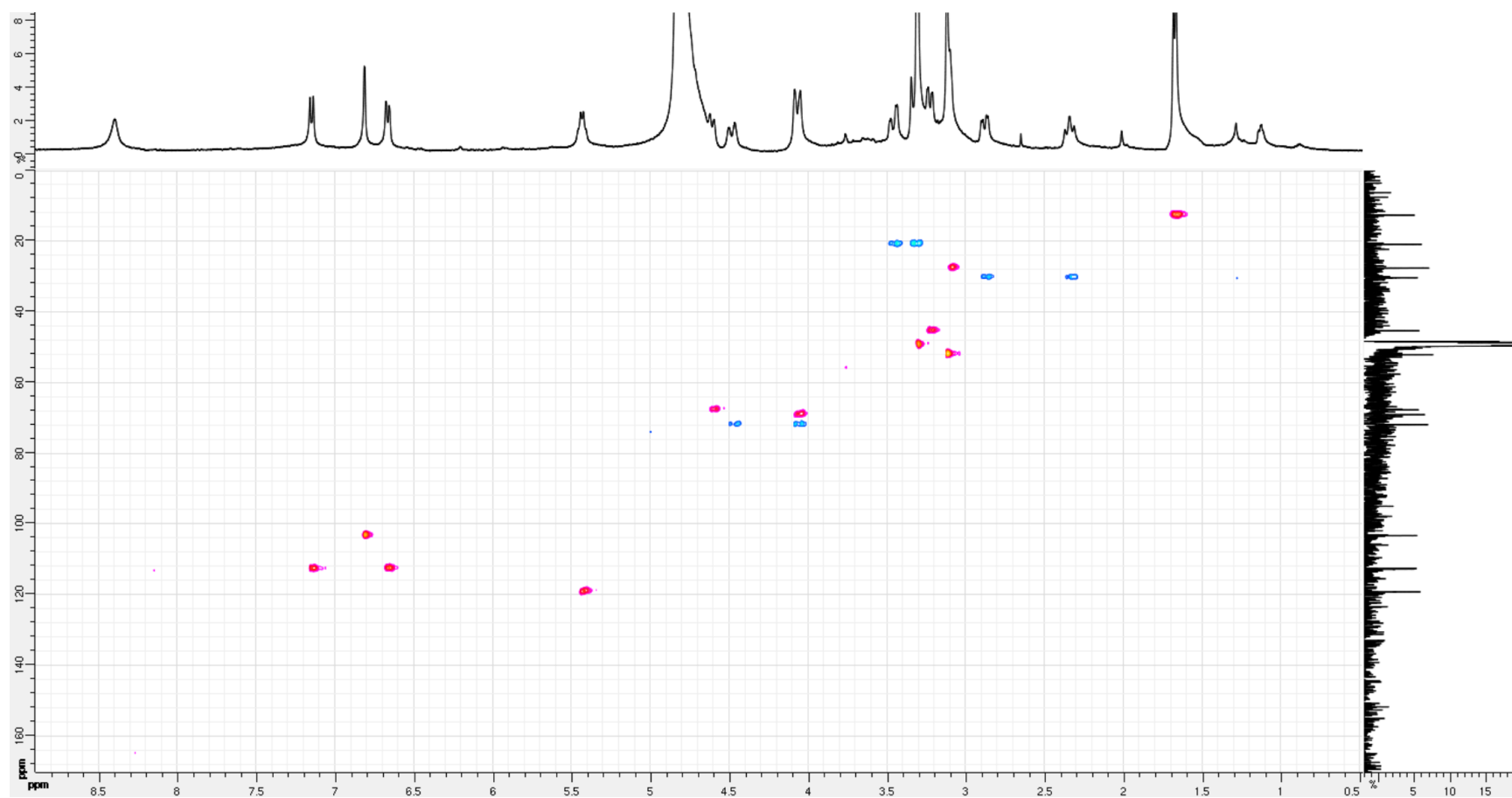
Figure S34. HSQC spectrum (400 MHz, CD₃OD) of 10-hydroxypericyclivine *N*-oxide (17)

Figure S35. HMBC spectrum (400 MHz, CD₃OD) of 10-hydroxypericyclivine *N*-oxide (17)

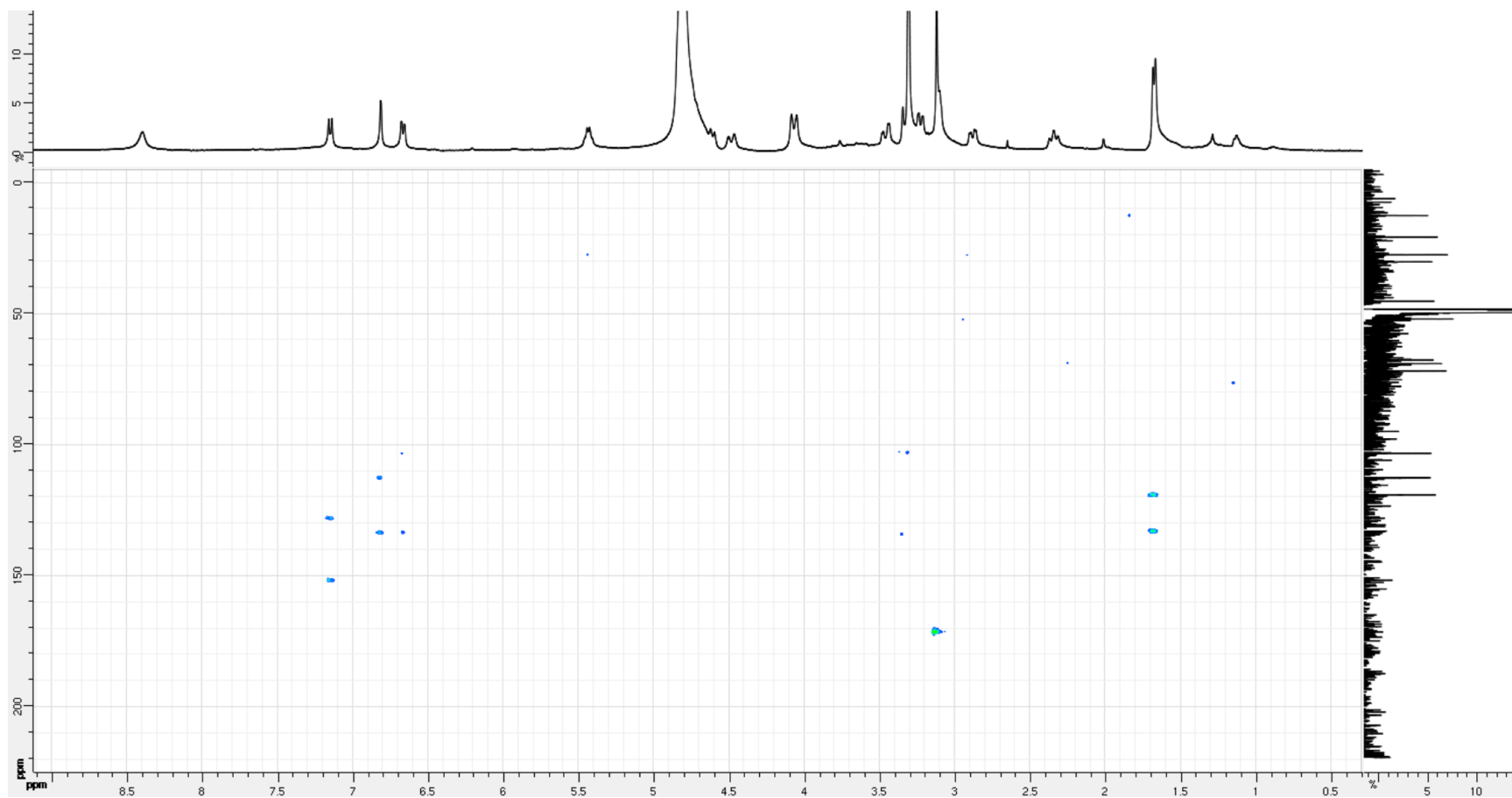


Figure S36. NOESY spectrum (400 MHz, CD₃OD) of 10-hydroxypericyclivine *N*-oxide (17)

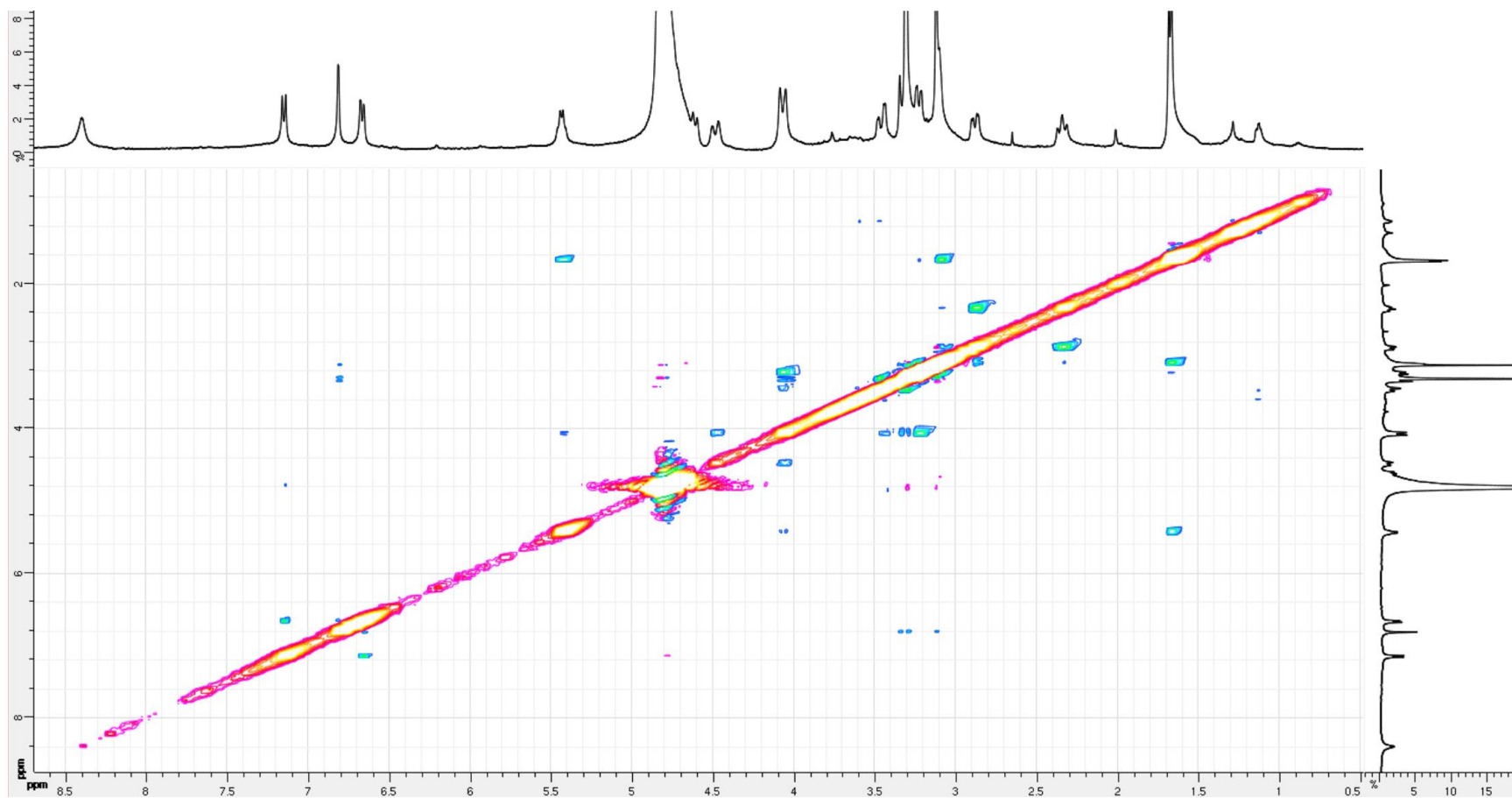
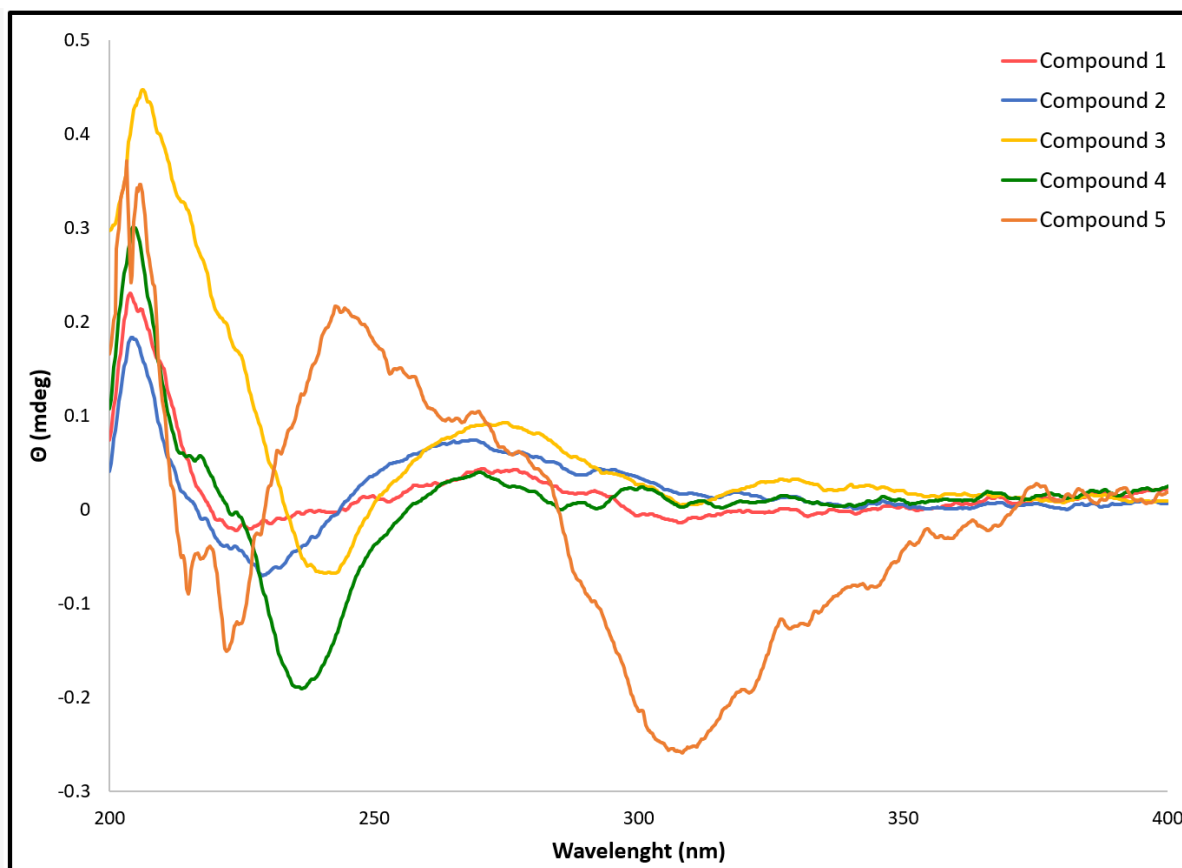


Figure S37. ECD spectra from compounds 13–17 in MeOH



References:

- (1) I. Blaženović, T. Kind, J. Ji and O. Fiehn, *Metabolites* **2018**, *8*, 31.
- (2) T.-M. Pinchon, J.-M. Nuzillard, B. Richard, G. Massiot, L. Le Men-Olivier and T. Sevenet, *Phytochemistry* **1990**, *29*, 3341-3344.
- (3) a) K. Bláha, Z. Kobicová and J. Trojanek, *Collect. Czech. Chem. Commun.* **1974**, *39*, 3168-3176; b) A. E. Nugroho and H. Morita, *Heterocycles* **2012**, *84*, 101-113.
- (4) M. C. Chambers, B. Maclean, R. Burke, D. Amodei, D. L. Ruderman, S. Neumann, L. Gatto, B. Fischer, B. Pratt and J. Egerton, *Nat. Biotechnol.* **2012**, *30*, 918.
- (5) T. Pluskal, S. Castillo, A. Villar-Briones and M. Orešič, *BMC Bioinform.* **2010**, *11*, 395.
- (6) O. D. Myers, S. J. Sumner, S. Li, S. Barnes and X. Du, *Anal. Chem.* **2017**, *89*, 8696-8703.
- (7) P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski and T. Ideker, *Genome Res.* **2003**, *13*, 2498-2504.



3. RÉSULTATS BIOLOGIQUES OBTENUS

Les cinq nouveaux composés isolés ont été testés pour déterminer leurs activités antiparasitaires contre une souche de *Plasmodium falciparum* FcB1 résistante à la chloroquine, et contre les formes axénique et intramacrophagique (RAW 264.7) des amastigotes de *Leishmania donovani* (LV9). De plus, leurs cytotoxicités contre la lignée cellulaire MRC-5 (fibroblastes embryonnaires humains) ont été évaluées (**Tableau 4**). Les composés 1–5 n’ont pas présenté d’activité antiparasitaire ni de cytotoxicité.

Tableau 4. Valeurs de CI_{50} (μ M) des activités antipaludiques et antileishmaniennes et cytotoxiques des composés 1–5^a

sample	<i>P. falciparum</i> strain FCB1	<i>L. donovani</i> LV9 axenic amastigotes	<i>L. donovani</i> LV9 intramacrophagic amastigotes	MRC-5
1	>50	> 100	> 100	> 20
2	>50	> 100	> 100	> 20
3	>50	> 100	> 100	> 20
4	>50	> 100	> 100	> 20
5	>50	> 100	> 100	> 20
référence ^b	0.06 ± 0.02	0.42 ± 0.07	2.04 ± 0.54	$0.55 \times 10^{-3} \pm 0.03 \times 10^{-3}$

^aLes valeurs montrées dans le tableau sont les moyennes \pm écart type des trois expériences indépendantes. ^bchloroquine (FcB1); miltefosine (LV9); paclitaxel (MRC-5).

L'évaluation biologique des composés a été réalisée en suivant les méthodologies décrites par la suite :

- **Détermination de l'activité antipaludique**

La détermination de l'activité antipaludique a été faite par Elizabeth Mouray sous la direction de Philippe Grellier au Muséum Nationale d'Histoire Naturelle (61 rue Buffon (CP52), 75005 Paris, France).

La souche résistante à la chloroquine FcB1/Colombie de *Plasmodium falciparum* a été obtenue de la collection du Muséum Nationale d'Histoire Naturelle à Paris, France (n°MNHN-CEU-224-PfFCB1). Les parasites ont été maintenus in vitro dans des érythrocytes humaines, dans un milieu de culture RPMI 1640 additionné de sérum humain 8% (v/v) inactivé par la chaleur à 37 °C sous une atmosphère de 3% de CO₂, 6% de O₂ et 91% de N₂.



Les globules rouges humaines et le sérum ont été fournis par l'Établissement Français du Sang d'Île de France sous l'agrément n°13/EFS/126 de la C-CPSLUNT. La sensibilité au médicament *in vitro* a été mesurée par l'incorporation de [³H]-hypoxanthine comme a été décrit précédemment.⁹⁰ Des solutions mères de médicaments ont été préparées dans du DMSO. Les composés ont été dilués en série deux fois avec 100 µL de milieu de culture dans des plaques à 96 puits. Des cultures parasitaires asynchrones (100 µL, 1% de parasitémie et 1% d'hématocrite final) ont ensuite été ajoutées à chaque puit puis incubées pendant 24 h à 37 °C avant l'ajout de 0,5 µCi de [³H]-hypoxanthine (GE Healthcare; 1 à 5 Ci.mmol/mL) par puit. Après une incubation supplémentaire de 24 h, les plaques ont été congelées et décongelées. Les lysats cellulaires ont été ensuite recueillis sur des filtres en fibre de verre et comptés dans un spectromètre à scintillation liquide. L'inhibition de la croissance pour chaque concentration de produit a été déterminée par la comparaison de la radioactivité incorporée dans le milieu de culture traitée à celle du milieu de culture témoin maintenue sur la même plaque. La concentration entraînant une inhibition de 50% de la croissance (CI₅₀) a été obtenue à partir de la courbe dose-réponse et les résultats ont été exprimés en tant que valeurs moyennes ± écarts-types déterminés à partir de trois expériences indépendantes. Le diphosphate de chloroquine, utilisé comme contrôle positif, a été acheté chez Sigma (pureté > 99%).

- **Détermination de l'activité antileishmanienne**

L'activité antileishmanienne des composés isolés a été déterminée par Sébastien Pomel sous la direction de Philippe M. Loiseau à la faculté de pharmacie de l'université Paris-Sud – Paris-Saclay (5 Rue Jean-Baptiste Clément, 92290 Châtenay-Malabry, France).

L'évaluation *in vitro* a été réalisée sur la lignée cellulaire de *L. donovani* (MHOM/ET/67/HU3), souche LV9, fournie par le professeur S. Croft, de la *London School of Hygiene and Tropical Medicine*, au Royaume-Uni. L'évaluation sur les amastigotes axéniques a été adaptée de la méthode utilisée par Cheikh-Ali *et al.*⁹¹ Dans les cellules vivantes, la résazurine est réduite en résosfurine. Cette conversion a été contrôlée en mesurant l'absorbance à des longueurs d'ondes spécifiques de la résosfurine (570 nm) et de la résazurine (600 nm), en utilisant un lecteur de microplaques (Lab systems Multiskan MS). Les activités finales ont été exprimées sous forme de valeurs de CI₅₀. La miltefosine a été utilisée comme composé de référence.

⁹⁰ J. Guillon, P. Grellier, et al., *J. Med. Chem.*, 2004, **47**, 1997-2009.

⁹¹ Z. Cheikh-Ali, J. Caron, et al., *ChemMedChem*, 2015, **10**, 411-418.



L'évaluation en utilisant d'amastigotes intramacrophagiques a également été adaptée à partir de la méthode de Cheikh-Ali.⁹¹ La fluorescence de l'ADN de ces cellules a été contrôlée par un realplex Mastercycler® (Eppendorf, Montesson, France). L'activité des matériaux testés a été exprimée en valeurs de CI_{50} . La miltefosine a été utilisée comme composé de référence.

- **Determination de la cytotoxicité**

La détermination de la cytotoxicité a été réalisée par Jerome Bignon à ICSN - CNRS (1, avenue de la Terrasse, 91198 Gif-sur-Yvette, France).

La lignée cellulaire MRC-5 a été obtenue auprès de l'*American Type Culture Collection* (Rockville, Maryland, États-Unis) et a été cultivée conformément aux instructions du fournisseur. Les cellules MRC-5 ont été cultivées dans du DMEM additionné de sérum de veau fœtal (SVF) (10%) et de glutamine (1%). La lignée cellulaire a été maintenue à 37 °C dans une atmosphère humidifiée contenant 5 % de CO₂. L'inhibition de la croissance cellulaire a été déterminée par un test MTS ((3-(4,5-diméthylthiazol-2-yl)-5-(3-carboxyméthoxyphényl)-2-(4-sulfophényl)-2H-tétrazolium) conformément aux instructions du fabricant (Promega, Madison, WI, USA). La valeur de la CI_{50} correspond à la concentration de composé induisant une diminution de 50% de l'absorbance des cellules traitées par le médicament par rapport aux cellules non traitées. Les expériences ont été effectuées en triplicats. Le paclitaxel a été utilisé comme composé de référence.



CONCLUSION GÉNÉRALE

Au cours de ce travail, l'approche de déréplication basée sur la technique du *molecular networking* a été présentée, ainsi que l'évolution de cette approche (**Chapitre I**). Les différents modes pour le ciblage des produits naturels originaux qui ont été appliqués ces dernières années ont été développés. C'est ainsi que la tendance à une amélioration de cette technique par la communauté scientifique internationale est observée. Avec le développement des nouveaux outils d'annotation expérimentale et *in silico*, la déréplication des substances naturelles devient de plus en plus efficace.

Dans le but d'une annotation orientée vers la nouveauté structurale chez les alcaloïdes indolomonoterpéniques, une base de données spectrale, la MIADB a été créée (**Chapitre II**). Les spectres SM/SM de 172 alcaloïdes qui la composent sont maintenant disponibles sur le site du GNPS pour l'annotation des mélanges complexes. La MIADB a été ensuite testée pour déréplicer un extrait alcaloïdique des feuilles de *C. roseus*, à titre de validation de cette base de données.

Ensuite, l'exploration de l'espace chimique d'un extrait alcaloïdique des écorces de *Geissospermum laeve* (Apocynaceae) a été décrite (**Chapitre III**). L'application d'une approche de déréplication basée sur *molecular networking* en utilisant comme référence une première version de la MIADB, incluse dans le réseau et influençant sa topologie, a permis l'isolement de trois nouveaux alcaloïdes indolomonoterpéniques. Parmi ces molécules, deux ont un motif chimique nouveau dans la série avec une γ -lactone α,β -insaturée. Cette première exploration de l'espace chimique d'une Apocynaceae déjà étudiée par *molecular networking* démontre l'intérêt de la réinvestigation d'espèces. Bien que trois nouvelles molécules aient pu être isolées, d'autres composés d'intérêt ont également été ciblés mais leur purification s'est avérée impossible.

Avec comme but l'amélioration de l'approche de déréplication décrite auparavant, l'étude phytochimique de l'extrait alcaloïdique des feuilles de *Alstonia balansae* (Apocynaceae) a été entreprise (**Chapitre IV**). L'annotation expérimentale (via MIADB incluse dans GNPS), combinée avec l'annotation *in silico* grâce à MetWork, un outil récemment créé de métabolisation et de modélisation de spectres SM/SM *in silico*, a permis l'identification de cinq nouveaux alcaloïdes indolomonoterpéniques du type *N*-oxyde sarpagine. L'évolution de l'approche déréplicative entre ces deux derniers chapitres montre l'intérêt d'une annotation *in silico* avec comme point de départ des données expérimentales. Il s'agit ici du premier exemple d'isolement de substances naturelles dont la structure a été initialement prédite par ordinateur.



Dans ce manuscrit, la description des travaux entrepris nous montre une évolution vers une déréduplication plus efficace qui s'appuie sur des outils de génération *in silico* des données des molécules, elles à leurs tours numérisées précédemment. L'évolution vers ce genre d'approches est donc la tendance actuelle dans l'art de l'isolement des produits naturels.



PERSPECTIVES

L'exploration de l'extrait alcaloïdique des feuilles de *G. laeve* (**Chapitre III**) et d'*A. balansae*, ainsi que des autres *Alstonia* disponibles (**Chapitre IV**) montre qu'un ciblage efficace des molécules potentiellement nouvelles n'assure pas leur isolement. En effet, des tentatives de purification non décrites ici se sont avérées infructueuses : soit les composés étaient instables, soit les quantités présentes étaient trop faibles. L'emploi d'autres techniques pour purifier certains de ces composés sont potentiellement envisageables. L'utilisation de la chromatographie de partage centrifuge semblerait une solution potentielle dans certains cas, évitant notamment le passage sur une phase stationnaire solide acide comme la silice. Cette approche a d'ailleurs déjà été appliquée au laboratoire pour des séries alcaloïdiques comparables.⁹² La quantité du produit contenu dans l'échantillon est, quant à elle, un paramètre à prendre en compte précocement pour savoir si la purification est envisageable pour une caractérisation structurale ultérieure. Si elle peut être appréhendée « manuellement » par l'examen des chromatogrammes UV et SM comparé aux masses disponibles d'échantillons, l'inclusion au traitement des données de la notion de quantification relative peut être intéressante. Cette notion a déjà été abordée par Olivon *et al.*¹⁰ en proposant une approche semi-quantitative basée sur l'aire des pics chromatographiques et sa représentation dans les réseaux. On conclura cette partie « isolement / élucidation structurale » en soulignant que de nombreuses molécules restent encore à obtenir dans les plantes que nous avons étudiées. Beaucoup n'ont en effet pas pu être dérépliquées ou s'inscrivent de manière intéressante dans les réseaux

Comme indiqué précédemment (**Chapitre II**) les alcaloïdes indolomonoterpéniques sont liés par de multiples voies des biogenèses organisées autour d'intermédiaires-clés. Nous avons pu constater qu'il n'existait pas de relation entre la biogenèse de ces substances et leurs liaisons par *molecular networking*, ce qui était anticipable. En comparant les relations biosynthétiques ou l'appartenance à une même classe de MIA avec les associations en réseau, nous avons aussi pu relever que quelques molécules ayant des squelettes sans rapport biogénétique direct se montraient fortement liées.

Une rationalisation de la « *clusterisation* » (néologisme employé au laboratoire pour « appariement ») de ces composés par *molecular networking* a été envisagée. C'est ainsi que les 172 composés de la MIADB ont été soumis à une analyse de ressemblance par le GNPS, pour la genèse de réseaux moléculaires avec une variation des paramètres de traitement. Les multiples réseaux obtenus ont été comparés, puis le plus informatif a été choisi. Nous avons pu constater que quelques molécules ayant des structures assez différentes se montraient liées dans le réseau. Ces composés ont été choisis comme « molécules diagnostiques ».

⁹² M. Girardot, A. Gadea, et al., *Eur. J. Org. Chem.*, 2012, **2012**, 2816-2823.



Les spectres SM/SM de paires biogénétiquement aberrantes de molécules seront donc étudiés pour déterminer quels ions fragments sont à l'origine de ces liaisons dans le réseau. Des « règles de *clusterisation* » pourraient donc être établies en utilisant cette information comme base. Si des petits ions diagnostiques sont déjà identifiables d'après la littérature, la rationalisation des réseaux reste à établir au laboratoire.

Comme évoqué auparavant, il existe une tendance vers l'annotation *in silico* dans les approches déréplicatives utilisés à l'heure actuelle. L'utilisation de MetWork (**Chapitre IV**) l'illustre. Par ailleurs, malgré la pertinence de l'information obtenu à partir d'un réseau annoté avec de l'information *in silico*, il existe quelques limites à cette approche. En effet, les spectres SM/SM générés de cette manière représentent seulement une approximation des spectres SM/SM expérimentaux. Dans quelques cas, cette approximation engendre une qualité des résultats encore insuffisante, compte tenu de la complexité des voies de fragmentation suivies par les molécules. C'est ainsi que, comme observé au cours de l'exploration de l'espace chimique de *A. balansae* (**Chapitre IV**), les spectres *in silico* générés n'arrivent pas à être correctement « matchés » à un spectre expérimental dans quelques cas. Par conséquent, MetWork propose une liste de molécules probables par *match* entre les composés générés (et fragmentés) *in silico* et les composés de l'échantillon, avec des scores de similarité différents. Bien que la structure proposée avec le score de similarité le plus élevée puisse s'avérer correcte, ce n'est pas le cas pour toutes les molécules. Ceci relève une des limites de la génération de spectres SM/SM *in silico*. Pour surmonter cet obstacle, le « *machine learning* » se montre comme une solution envisageable. L'apprentissage automatique avec de spectres SM/SM expérimentaux pourrait être donc une manière d'améliorer la qualité de l'approximation *in silico* faite par ces outils, notre base de données constituant un beau jeu de données pour cet apprentissage. Ces spectres expérimentaux pourraient être obtenus en profitant, également, des bases de données spectrales en libre accès, avec d'autres AIMs déposés sur le GNPS, notamment (voir la validation analytique n°2 du **Chapitre II**).

Par ailleurs, le module de MetWork dédié à la métabolisation des *matches* reste à améliorer également. En effet, plusieurs réactions impliquées dans les voies de biosynthèses classiques des produits naturels n'ont pas encore été incluses parmi les 60 déjà proposées à ce jour par cet outil. L'intégration de ces réactions pourrait donc mener à une meilleure annotation *in silico* du réseau. Il s'agit d'ailleurs d'un intéressant défi en termes de rationalisation des associations de monomères pour former des bis-indoles, voire des analogues de plus haute masse. Le module des réactions de métabolisation de Metwork peut être nourri directement sur le site internet de l'outil avec des nouvelles réactions, qui peuvent être importées depuis un fichier du type .mrv. Ceci sera envisagé en particulier avec le début de travaux sur des *Strychnos* (Loganiaceae) qui doit débiter en 2019 au laboratoire, en collaboration avec le laboratoire de pharmacognosie de Liège (Pr. M. Frédéric), contributeur de la MIADB.



Les perspectives offertes par l'approche des réseaux moléculaires sont nombreuses et offrent un regard neuf sur la diversité chimique et la répartition botanique des AIMs. Outre la poursuite de la recherche d'analogues à forte originalité structurale, notamment à partir de nos collections de plantes, des pistes relatives aux filiations biosynthétiques ou à des relations chimiotaxonomiques seront poursuivies dans un avenir proche.





BILAN DES COMMUNICATIONS

Communication orale, congrès international : 1

SINAPSIS 2017 – Berlin, Allemagne, 24 - 26 octobre 2017

- Fox Ramos, A. E., Gallard J.-F., Dumontet V., Litaudon M., Champy P., Beniddir M. A. **Prioritizing natural product diversity in three *Alstonia* species endemic to New Caledonia through massive multi-informative molecular networks.**

Communications orales, congrès nationaux : 3

PSE Young Scientists' Meeting Lille 2017 – Villeneuve d'Ascq, France, 28 juin - 1 juillet 2017

- Fox Ramos, A. E., Gallard J.-F., Dumontet V., Litaudon M., Champy P., Beniddir M. A. **Prioritizing natural product diversity in three *Alstonia* species endemic to New Caledonia through massive multi-informative molecular networks.** – Prix de la meilleure communication orale.

Journée des doctorants de 2^e année (séminaire annuel de l'UMR) – Châtenay-Malabry, France, 14 juin 2017

- Fox Ramos, A. E., Champy P., Beniddir M. A. **Exploration of the chemical diversity of South American and New Caledonian Gentianales with antiparasitic potential by the molecular networking technique**

SINAPSIS 2016 – Berlin, Allemagne, 24 - 26 octobre 2017

- Fox Ramos, A. E., Alcover, C., Maciuk, A., Duplais, C., Bernadat, G., Dejean, C., Jullian, J.-C., Grellier, P., Poupon, E., Evanno, L., Champy, P., Beniddir, M. A. **Unveiling the secrets of forgotten plants: molecular networking-based phytochemical study of *Geissospermum laeve* (Vell.) Miers.**

Communications par affiche, congrès international : 2

9th Joint Natural Products Conference 2016 – Copenhague, Danemark, 24 - 27 juillet 2016 ; des actes des congrès ont été publiés

- Fox Ramos, A. E., Alcover, C., Maciuk, A., Duplais, C., Bernadat, G., Dejean, C., Jullian, J.-C., Grellier, P., Poupon, E., Evanno, L., Champy, P., Beniddir, M. A. **Unveiling the secrets of forgotten plants: molecular networking-based phytochemical study of *Geissospermum laeve* (Vell.) Miers.**



- Beniddir, M. A., Fox Ramos, A. E., Otogo N'ngang, E., Alcover, C., Maciuk, A., Evanno, L., Poupon, E., Champy, P. **Out of fashion plants at the “big data” Era: Illuminating the overlooked Apocynaceae alkaloids chemical space by molecular networking.**

Communications par affiche, congrès nationaux : 4

XVII^{ème} Journée de l'École doctorale 569 « Innovation thérapeutique : du fondamental à l'appliqué » – Châtenay-Malabry, France, 23 juin 2017

- Fox Ramos, A. E., Gallard J.-F., Dumontet V., Litaudon M., Champy P., Beniddir M. A. **Prioritizing natural product diversity in three *Alstonia* species endemic to New Caledonia through massive multi-informative molecular networks.**

24th Young Research Fellow Meeting – Châtenay-Malabry, France, 8 - 10 février 2017

- Fox Ramos, A. E., Gallard J.-F., Dumontet V., Litaudon M., Champy P., Beniddir M. A. **Prioritizing natural product diversity in three *Alstonia* species endemic to New Caledonia through massive multi-informative molecular networks.**

XVI^{ème} Journée de l'École doctorale 569 « Innovation thérapeutique : du fondamental à l'appliqué » – Châtenay-Malabry, France, 17 juin 2016

- Fox Ramos, A. E., Alcover C., Evanno L., Maciuk A., Litaudon M., Duplais C., Bernadat G., Gallard J.-F., Jullian J.-C., Mouray E., Grellier P., Loiseau P. M., Pomel S., Poupon E., Champy P., Beniddir M. A. **Unveiling the secrets of forgotten plants: molecular networking-based phytochemical study of *Geissospermum laeve* (Vell.) Miers.**

Journée de la Recherche de la Faculté de Pharmacie – Châtenay-Malabry, France, 20 novembre 2015

- Fox Ramos, A. E., Alcover C., Maciuk A., Duplais C., Jullian J.-C., Dejean C., Poupon E., Evanno L., Champy P., Beniddir M. A. **Étude phytochimique de *Geissospermum laeve* (Vell.) M.**

Actes de congrès : 2

9th Joint Natural Products Conference 2016 – Copenhague, Danemark, 24 - 27 juillet 2016 ; des actes des congrès ont été publiés

- Fox Ramos, A. E., Alcover, C., Maciuk, A., Duplais, C., Bernadat, G., Dejean, C., Jullian, J.-C., Grellier, P., Poupon, E., Evanno, L., Champy, P., Beniddir, M. A. **Unveiling the secrets of forgotten plants: molecular networking-based phytochemical study of *Geissospermum laeve* (Vell.) Miers.** *Planta Med.* **2016**, 81(S 01), 31.
- Beniddir, M. A., Fox Ramos, A. E., Otogo N'ngang, E., Alcover, C., Maciuk, A., Evanno, L., Poupon, E., Champy, P. **Out of fashion plants at the “big data” Era: Illuminating the overlooked Apocynaceae alkaloids chemical space by molecular networking.** *Planta Med.* **2016**, 81(S 01), 46.



Articles publiés : 2

- **Fox Ramos A. E.**, Alcover C., Evanno L., Maciuk A., Litaudon M., Duplais C., Bernadat G., Gallard J.-F., Jullian J.-C., Mouray E., Grellier P., Loiseau P. M., Pomel S., Poupon E., Champy P., Beniddir M. A. **2017. Revisiting previously investigated plants: a molecular networking-based study of *Geissospermum laeve*. *J. Nat. Prod.*, **80**(4), 1107-1014.**
- **Fox Ramos A. E.**, Le Pogam, P., Fox Alcover C., Otogo N’Nang, E., Cauchie G., Hazni H., Awang K., Bréard D., Echavarren A. M., Frédérick M., Gaslonde T., Girardot M., Grougnet R., Kirillova M. S., Kritsanida M., Lémus C., Le Ray A.-M., Lewin G., Litaudon M., Mambu L., Michel S., Miloserdov F. M., Muratore M. E., Richomme-Peniguel P., Roussi F., Evanno L., Poupon E., Champy P., Beniddir M. A. **2019. Collected mass spectrometry data on monoterpene indole alkaloids from natural product chemistry research. *Sci. Data.*, in press.**

Articles soumis en évaluation : 2

- **Fox Ramos A. E.**, Evanno L., Poupon E., Champy P., Beniddir M. A., **2019. Natural products targeting strategies involving molecular networking: Different manners, one goal.** Soumis à *Nat. Prod. Rep.*
- **Fox Ramos A. E.**, Pavesi C., Litaudon M., Dumontet V., Poupon E., Champy P., Genta-Jouve G., Beniddir M. A. **2019. CANPA: Computer-Assisted Natural Products Anticipation.** Soumis à *Angew. Chem.*





RÉFÉRENCES BIBLIOGRAPHIQUES

Ces références sont celles citées en notes de bas de page dans le corps de texte. Certaines des références citées dans les articles acceptés, soumis ou en préparation n'y figurent pas. Les sources iconographiques apparaissent également dans cette liste.

1. S. John Walker, Big Data: a revolution that will transform how we live, work, and think, *Int. J. Advert.*, 2014, **33**, 181-183.
2. J. Hubert, J.-M. Nuzillard et J.-H. Renault, Dereplication strategies in natural product research: How many tools and methodologies behind the same concept?, *Phytochem Rev.*, 2017, **16**, 55-95.
3. M. Wang, J. J. Carver, V. V. Phelan, L. M. Sanchez, N. Garg, Y. Peng, D. D. Nguyen, J. Watrous, C. A. Kapon et T. Luzzatto-Knaan, Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking, *Nat. Biotechnol.*, 2016, **34**, 828.
4. N. B. Cech et K. Yu, Mass spectrometry for natural products research, *LC GC N. Am.*, 2013, **31**, 938-947.
5. D. Nikolić, CASMI 2016: A manual approach for dereplication of natural products using tandem mass spectrometry, *Phytochem Lett.*, 2017, **21**, 292-296.
6. 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. Debonsi, W. H. Gerwick et P. C. Dorrestein, Molecular Networking as a dereplication strategy, *J. Nat. Prod.*, 2013, **76**, 1686-1699.
7. F. Olivon, N. Elie, G. Grelier, F. Roussi, M. Litaudon et D. Touboul, MetGem software for the generation of molecular networks based on t-SNE algorithm, *Anal. Chem.*, 2018.
8. T. Pluskal, S. Castillo, A. Villar-Briones et M. Orešič, MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data, *BMC Bioinform.*, 2010, **11**, 395.
9. F. Olivon, G. Grelier, F. Roussi, M. Litaudon et D. Touboul, MZmine 2 data-preprocessing to enhance molecular networking reliability, *Anal. Chem.*, 2017, **89**, 7836-7840.
10. F. Olivon, F. Roussi, M. Litaudon et D. Touboul, Optimized experimental workflow for tandem mass spectrometry molecular networking in metabolomics, *Anal. Bioanal. Chem.*, 2017, **409**, 5767-5778.
11. H. L. Röst, T. Sachsenberg, S. Aiche, C. Bielow, H. Weisser, F. Aicheler, S. Andreotti, H.-C. Ehrlich, P. Gutenbrunner et E. Kenar, OpenMS: a flexible open-source software platform for mass spectrometry data analysis, *Nat. Methods*, 2016, **13**, 741.
12. P.-M. Allard, T. Péresse, J. Bisson, K. Gindro, L. Marcourt, V. C. Pham, F. Roussi, M. Litaudon et J.-L. Wolfender, Integration of molecular networking and in-silico MS/MS fragmentation for natural products dereplication, *Anal. Chem.*, 2016, **88**, 3317-3323.
13. F. Allen, A. Pon, M. Wilson, R. Greiner et D. Wishart, CFM-ID: a web server for annotation, spectrum prediction and metabolite identification from tandem mass spectra, *Nucleic Acids Res.*, 2014, **42**, W94-W99.
14. R. R. da Silva, M. Wang, L.-F. Nothias, J. J. van der Hooft, A. M. Caraballo-Rodríguez, E. Fox, M. J. Balunas, J. L. Klassen, N. P. Lopes et P. C.



- Dorrestein, Propagating annotations of molecular networks using in silico fragmentation, *PLOS Comput. Biol.*, 2018, **14**, e1006089.
15. D. Krug et R. Müller, Secondary metabolomics: the impact of mass spectrometry-based approaches on the discovery and characterization of microbial natural products, *Nat. Prod. Rep.*, 2014, **31**, 768-783.
 16. A. Bouslimani, L. M. Sanchez, N. Garg et P. C. Dorrestein, Mass spectrometry of natural products: current, emerging and future technologies, *Nat. Prod. Rep.*, 2014, **31**, 718-729.
 17. 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 et M. Litaudon, Bioactive natural products prioritization using massive multi-informational molecular networks, *ACS Chem. Biol.*, 2017, **12**, 2644-2651.
 18. K. Kleigrew, J. Almaliti, I. Y. Tian, R. B. Kinnel, A. Korobeynikov, E. A. Monroe, B. M. Duggan, V. Di Marzo, D. H. Sherman, P. C. Dorrestein, L. Gerwick et W. H. Gerwick, Combining mass spectrometric metabolic profiling with genomic analysis: a powerful approach for discovering natural products from cyanobacteria, *J. Nat. Prod.*, 2015, **78**, 1671-1682.
 19. S. Böcker, M. C. Letzel, Z. Lipták et A. Pervukhin, SIRIUS: decomposing isotope patterns for metabolite identification, *Bioinformatics*, 2008, **25**, 218-224.
 20. K. Dührkop, H. Shen, M. Meusel, J. Rousu et S. Böcker, Searching molecular structure databases with tandem mass spectra using CSI: FingerID, *Proc. Natl. Acad. Sci. USA*, 2015, **112**, 12580-12585.
 21. Y. Beauxis et G. Genta-Jouve, Metwork: a web server for natural products anticipation, *Bioinformatics*, 2018, DOI: 10.1093/bioinformatics/bty864.
 22. L. F. Szabó, Rigorous biogenetic network for a group of indole alkaloids derived from strictosidine, *Molecules*, 2008, **13**, 1875-1896.
 23. P. M. Dewick, *Medicinal Natural Products: A Biosynthetic Approach*, 3rd edn., 2009.
 24. M. Lahlou, The success of natural products in drug discovery, *Pharmacol Pharm.*, 2013, **4**, 17-31.
 25. Nauclea *orientalis* 031208-3067, https://www.flickr.com/photos/tony_rodd/2205187291/, (accessed 22-09-2018, 2018).
 26. *Catharanthus roseus* (Madagascar Periwinkle), [https://keys.lucidcentral.org/keys/v3/eafrinet/weeds/key/weeds/Media/Html/Catharanthus_roseus_\(Madagascar_Periwinkle\).htm](https://keys.lucidcentral.org/keys/v3/eafrinet/weeds/key/weeds/Media/Html/Catharanthus_roseus_(Madagascar_Periwinkle).htm), (accessed 22-09-2018, 2018).
 27. Sarpagandha: health benefits, dosage, side effects, <https://mavcure.com/sarpagandha/>, (accessed 22-09-2018, 2018).
 28. Akuamma, <https://micromedix.me/2017/06/27/microdosis-de-akuamma-nueva-medicina-alternativa-para-tratar-la-esquizofrenia-y-otros-trastornos-mentales/akuamma/>, (accessed 22-09-2018, 2018).
 29. J. Stöckigt, L. Barleben, S. Panjikar et E. A. Loris, 3D-Structure and function of strictosidine synthase—the key enzyme of monoterpenoid indole alkaloid biosynthesis, *Plant Physiol Biochem.*, 2008, **46**, 340-355.
 30. J. Le Men et W. I. Taylor, A uniform numbering system for indole alkaloids, *Experientia*, 1965, **21**, 508-510.



31. M. W. Chase, M. Christenhusz, M. Fay, J. Byng, W. Judd, D. Soltis, D. Mabberley, A. Sennikov, P. Soltis et P. Stevens, An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV, *Bot. J. Linn. Soc.*, 2016, **181**, 1-20.
32. D. Gülcemal, M. Masullo, O. Alankuş-Çalışkan, T. Karayıldırım, S. G. Şenol, S. Piacente et E. Bedir, Monoterpenoid glucoindole alkaloids and iridoids from *Pterocephalus pinardii*, *Magn. Reson. Chem.*, 2010, **48**, 239-243.
33. X. Huang, Y. Li, Y. Su, X. Chai et S. Yan, Monoterpene indole alkaloids and monoterpene diglycosides from the roots of *Triosteum pinnatifidum*, *Phytochem Lett.*, 2014, **7**, 30-34.
34. J. Buckingham, K. H. Baggaley, A. D. Roberts et L. F. Szabo, *Dictionary of alkaloids, with CD-ROM*, CRC press, 2010.
35. Z.-W. Liu, J. Zhang, S.-T. Li, M.-Q. Liu, X.-J. Huang, Y.-L. Ao, C.-L. Fan, D.-M. Zhang, Q.-W. Zhang et W.-C. Ye, Ervadivamines A and B, Two Unusual Trimeric Monoterpenoid Indole Alkaloids from *Ervatamia divaricata*, *J. Org. Chem.*, 2018, **83**, 10613-10618.
36. Y. Hirasawa, S. Miyama, T. Hosoya, K. Koyama, A. Rahman, I. Kusumawati, N. C. Zaini et H. Morita, Alasmontamine A, a first tetrakis monoterpene indole alkaloid from *Tabernaemontana elegans*, *Org. Lett.*, 2009, **11**, 5718-5721.
37. E. O. N'Nang Obiang, G. Genta-Jouve, J.-F. Gallard, B. Kumulungui, E. Mouray, P. Grellier, L. Evanno, E. Poupon, P. Champy et M. A. Beniddir, Pleiokomenines A and B: dimeric aspidofractinine alkaloids tethered with a methylene group, *Org. Lett.*, 2017, **19**, 6180-6183.
38. A. E. Nugroho, Y. Hirasawa, N. Kawahara, Y. Goda, K. Awang, A. H. A. Hadi et H. Morita, Bisnicalaterine A, a vobasine– vobasine bisindole alkaloid from *Hunteria zeylanica*, *J. Nat. Prod.*, 2009, **72**, 1502-1506.
39. E. Otogo N'Nang, G. Bernadat, E. Mouray, B. Kumulungui, P. Grellier, E. Poupon, P. Champy et M. A. Beniddir, Theionbrunonines A and B: dimeric vobasine alkaloids tethered by a thioether bridge from *Mostuea brunonis*, *Org. Lett.*, 2018, **20**, 6596-6600.
40. A. E. Fox Ramos, C. Alcover, L. Evanno, A. Maciuk, M. Litaudon, C. Duplais, G. Bernadat, J.-F. Gallard, J.-C. Jullian, E. Mouray, P. Grellier, P. M. Loiseau, S. Pomel, E. Poupon, P. Champy et M. A. Beniddir, Revisiting previously investigated plants: a molecular networking-based study of *Geissospermum laeve*, *J. Nat. Prod.*, 2017, **80**, 1007-1014.
41. L. Caputi, J. Franke, S. C. Farrow, K. Chung, R. M. Payne, T.-D. Nguyen, T.-T. T. Dang, I. S. T. Carqueijeiro, K. Koudounas, T. D. de Bernonville, B. Ameyaw, D. M. Jones, I. J. Curcino Vieira, V. Courdavault et S. E. O'Connor, Missing enzymes in the biosynthesis of the anticancer drug vinblastine in Madagascar periwinkle, *Science*, 2018, **360**, 1235-1239.
42. V. Salim et V. De Luca, in *Advances in Botanical Research*, Elsevier, 2013, vol. 68, pp. 1-37.
43. Y. Qu, R. Simonescu et V. De Luca, Monoterpene indole alkaloids from the fruit of *Tabernaemontana littoralis* and differential alkaloid composition in various fruit components, *J. Nat. Prod.*, 2016, **79**, 3143-3147.
44. H. J. Knölker, *The alkaloids*, Elsevier Science, 2016.
45. S. E. O'Connor et J. J. Maresh, Chemistry and biology of monoterpene indole alkaloid biosynthesis, *Nat. Prod. Rep.*, 2006, **23**, 532-547.



46. M. Kitajima, N. Kogure, K. Yamaguchi, H. Takayama et N. Aimi, Structure reinvestigation of gelsemoxonine, a constituent of *Gelsemium elegans*, reveals a novel, azetidone-containing indole alkaloid, *Org. Lett.*, 2003, **5**, 2075-2078.
47. Y. Yang, Y. Bai, S. Sun et M. Dai, Biosynthetically inspired divergent approach to monoterpene indole alkaloids: total synthesis of mersicarpine, leuconodines B and D, leuconoxine, melodinine E, leuconolam, and rhazinilam, *Org. Lett.*, 2014, **16**, 6216-6219.
48. C. H. Heathcock, M. H. Norman et D. A. Dickman, Total synthesis of (.+.-)-vallesamidine, *J. Org. Chem.*, 1990, **55**, 798-811.
49. R. T. Brown, D. M. Duckworth et C. A. Santos, Biogenetically patterned synthesis of cadambine, *Tetrahedron Lett.*, 1991, **32**, 1987-1990.
50. S. Benayad, K. Ahamada, G. Lewin, L. Evanno et E. Poupon, Prekuammicine: A Long-Awaited Missing Link in the Biosynthesis of Monoterpene Indole Alkaloids, *Eur. J. Org. Chem.*, 2016, **2016**, 1494-1499.
51. E. Otogo n'ngang, Paris Saclay, 2018.
52. Pao-pereira : Commerce équitable et durable, <http://www.beljanskiblog.com/pao-pereira-commerce-equitable-et-durable/>, (accessed 15-09-2018, 2018).
53. M. E. Endress, S. Liede-Schumann et U. Meve, An updated classification for Apocynaceae, *Phytotaxa*, 2014, **159**, 175-194.
54. J. Miers, *On the Apocynaceae of South America*, Taylor and Francis, 1878. pp. 1-5.
55. *Geissospermum laevis*, <https://www.amigosib.org.br/galeria/agostosembro-2015/#jp-carousel-7431>, (accessed 08-09-2018, 2018).
56. *Geissospermum* urceolatum,
http://static1.kew.org/science/tropamerica/imagedatabase/large73/cat_single73-2.htm, (accessed 08-09-2018, 2018).
57. *Geissospermum* reticulatum,
<http://andesamazonfieldschool.com/Andes and Amazon Field School/Geissospermum reticulatum Apocynaceae.html>, (accessed 08-09-2018, 2018).
58. J. C. P. Steele, N. C. Veitch, G. C. Kite, M. S. J. Simmonds et D. C. Warhurst, Indole and β -Carboline alkaloids from *Geissospermum sericeum*, *J. Nat. Prod.*, 2002, **65**, 85-88.
59. *Geissospermum* Allem. GBIF Secretariat, <https://doi.org/10.15468/39omei>, (accessed 18-09-2018).
60. I. R. Koch, A.; Simões, A.O.; Kinoshita, L.S.; Spina, A.P.; Castello, A.C.D., Apocynaceae, <http://reflora.ibri.gov.br/jabot/floradobrasil/FB4605>, (accessed 18-09-2018).
61. *Geissospermum* *laeve* (Vell.) Miers
<http://specimens.kew.org/herbarium/K000965595>, (accessed 18-09-2018).
62. S. Bertani, G. Bourdy, I. Landau, J. Robinson, P. Esterre et E. Deharo, Evaluation of French Guiana traditional antimalarial remedies, *J. Ethnopharmacol.*, 2005, **98**, 45-54.
63. P. Grenand, C. Moretti, H. Jacquemin et M.-F. Prévost, *Pharmacopées traditionnelles en Guyane: créoles, palikur, wayãpi*, ORSTOM, 1987. p. 126.
64. F. Mbeunkui, M. H. Grace et M. A. Lila, Isolation and structural elucidation of indole alkaloids from *Geissospermum vellosii* by mass spectrometry, *J. Chromatogr. B*, 2012, **885**, 83-89.
65. Comité français de la pharmacopée « Plantes médicinales et huiles essentielles » — CP022014043,



- https://ansm.sante.fr/var/ansm_site/storage/original/application/c0085a65bb75c11555a6aa91ccc4c750.pdf, (accessed 10-09-2018, 2018).
66. N. A. Hughes et H. Rapoport, Flavopereirine, an alkaloid from *Geissospermum vellosii*, *J. Am. Chem. Soc.*, 1958, **80**, 1604-1609.
 67. H. Rapoport, T. P. Onak, N. A. Hughes et M. G. Reinecke, Alkaloids of *Geissospermum vellosii*, *J. Am. Chem. Soc.*, 1958, **80**, 1601-1604.
 68. H. Rapoport, R. J. Windgassen Jr, N. A. Hughes et T. P. Onak, Alkaloids of *Geissospermum vellosii*. Further Studies on geissospermine and the structures of the indolic cleavage products, geissoschizine and apogeissoschizine, *J. Am. Chem. Soc.*, 1960, **82**, 4404-4414.
 69. H. Rapoport et R. E. Moore, Alkaloids of *Geissospermum vellosii*. Isolation and structure determinations of vellosimine, vellosiminol, and geissolosimine, *J. Org. Chem.*, 1962, **27**, 2981-2985.
 70. R. E. Moore et H. Rapoport, Geissovelline, a new alkaloid from *Geissospermum vellosii*, *J. Org. Chem.*, 1973, **38**, 215-230.
 71. V. Muñoz, M. Sauvain, G. Bourdy, J. Callapa, S. Bergeron, I. Rojas, J. Bravo, L. Balderrama, B. Ortiz et A. Gimenez, A search for natural bioactive compounds in Bolivia through a multidisciplinary approach: Part I. Evaluation of the antimalarial activity of plants used by the Chacobo Indians, *J. Ethnopharmacol.*, 2000, **69**, 127-137.
 72. H. Ishiyama, M. Matsumoto, M. Sekiguchi, H. Shigemori, A. Ohsaki et J. i. Kobayashi, Two new indole alkaloids from *Aspidosperma subincanum* and *Geissospermum vellosii*, *Heterocycles*, 2005, **66**, 651-658.
 73. J. A. Lima, R. S. Costa, R. A. Epifânio, N. G. Castro, M. S. Rocha et A. C. Pinto, *Geissospermum vellosii* stem bark: anticholinesterase activity and improvement of scopolamine-induced memory deficits, *Pharmacol. Biochem. Behav.*, 2009, **92**, 508-513.
 74. F. Mbeunkui, M. H. Grace, C. Lategan, P. J. Smith, I. Raskin et M. A. Lila, In vitro antiplasmodial activity of indole alkaloids from the stem bark of *Geissospermum vellosii*, *J. Ethnopharmacol.*, 2012, **139**, 471-477.
 75. R. Baltenweck-Guyot et R. Ocampo, Identification of a novel 7-desmethyl-7-acetonyl bacteriophageophorbide c series in a recent sediment, *Org. Geochem.*, 2007, **38**, 1580-1584.
 76. J. Monachino, A revision of the genus *Alstonia* (Apocynaceae), *Pac. Sci.*, 1949, **3**, 133-182.
 77. The plant list, <http://www.theplantlist.org/tpl1.1/search?q=alstonia>, (accessed 19-09-2018, 2018).
 78. *Alstonia vieillardii* Van Heurck, <http://endemia.nc/flore/fiche1665>, (accessed 18-09-2018).
 79. *Alstonia balansae* Guillaumin, <http://endemia.nc/flore/fiche3284>, (accessed 20-09-2018).
 80. *Alstonia quaternata* Van Heurck, <http://endemia.nc/flore/fiche696>, (accessed 20-09-2018).
 81. K. Sidiyasa, Taxonomy, phylogeny, and wood anatomy of *Alstonia* (Apocynaceae), *Blumea, Suppl.*, 1998, **11**, 1-230.
 82. *Alstonia* R.Br. GBIF Secretariat, <https://doi.org/10.15468/39omei>, (accessed 19-09-2018).



83. Endemia.nc - Faune et flore de Nouvelle-Calédonie, <http://endemia.nc/flore/fiche256>, (accessed 20-09-2018).
84. M. S. Khyade, D. M. Kasote et N. P. Vaikos, *Alstonia scholaris* (L.) R. Br. and *Alstonia macrophylla* Wall. ex G. Don: A comparative review on traditional uses, phytochemistry and pharmacology, *Journal of ethnopharmacology*, 2014, **153**, 1-18.
85. A. P. G. Macabeo, K. Krohn, D. Gehle, R. W. Read, J. J. Brophy, G. A. Cordell, S. G. Franzblau et A. M. Aguinaldo, Indole alkaloids from the leaves of Philippine *Alstonia scholaris*, *Phytochemistry*, 2005, **66**, 1158-1162.
86. N. Keawpradub, E. Eno-Amooquaye, P. Burke et P. Houghton, Cytotoxic activity of indole alkaloids from *Alstonia macrophylla*, *Planta Med.*, 1999, **65**, 311-315.
87. *Alstonia balansae* Guillaumin, <http://specimens.kew.org/herbarium/K000857281>, (accessed 18-09-2018).
88. T.-M. Pinchon, J.-M. Nuzillard, B. Richard, G. Massiot, L. Le Men-Olivier et T. Sevenet, Alkaloids from *Alstonia undulata*, *Phytochemistry*, 1990, **29**, 3341-3344.
89. D. Guillaume, A. Morfaux, B. Richard, G. Massiot, L. Le Men-Olivier, J. Pusset et T. Sevenet, Some alkaloids of *Alstonia undulata*, *Phytochemistry*, 1984, **23**, 2407-2408.
90. J. Guillon, P. Grellier, M. Labaied, P. Sonnet, J.-M. Léger, R. Déprez-Poulain, I. Forfar-Bares, P. Dallemagne, N. Lemaître, F. Péhourcq, J. Rochette, C. Sergheraert et C. Jarry, Synthesis, antimalarial activity, and molecular modeling of new pyrrolo[1,2-a]quinoxalines, bispyrrolo[1,2-a]quinoxalines, bispyrido[3,2-e]pyrrolo[1,2-a]pyrazines, and bispyrrolo[1,2-a]thieno[3,2-e]pyrazines, *J. Med. Chem.*, 2004, **47**, 1997-2009.
91. Z. Cheikh-Ali, J. Caron, S. Cojean, C. Bories, P. Couvreur, P. M. Loiseau, D. Desmaële, E. Poupon et P. Champy, "Squalenoylcurcumin" nanoassemblies as water-dispersible drug candidates with antileishmanial activity, *ChemMedChem*, 2015, **10**, 411-418.
92. M. Girardot, A. Gadea, C. Deregnacourt, A. Deville, L. Dubost, B. Nay, A. Maciuk, P. Rasoanaivo et L. Mambu, Tabernaelegantinals: unprecedented cytotoxic bisindole alkaloids from *Muntafara sessilifolia*, *Eur. J. Org. Chem.*, 2012, **2012**, 2816-2823.



Titre : Exploration de la diversité chimique des Apocynaceae par la technique des réseaux moléculaires : de la création d'une base de données vers l'annotation *in silico*

Mots clés : alcaloïdes indolomonoterpéniques, Apocynaceae, base de données, déréplication, réseau moléculaire, spectrométrie de masse

Résumé : Les alcaloïdes indolo-monoterpéniques (AIMs) constituent une classe de molécules naturelles très étudiée en raison d'un fort potentiel pharmacologique et thérapeutique et d'une grande diversité structurale. D'autre part, les techniques de déréplication par chromatographie liquide couplée à la spectrométrie de masse tandem ont évolué récemment, avec l'émergence de l'approche par réseaux moléculaires (*molecular networking*). Dans ce contexte, nous avons créé une base de données de spectres de masse tandem moyennés de 172 AIMs témoins, en collaboration avec plusieurs équipes de pharmacognosie dans le monde. Cette base de données, nommée MIADB (*Monoterpene Indole Alkaloids DataBase*), rendue publique, peut être utilisée comme référence dans des stratégies de déréplication fondées sur l'utilisation des réseaux moléculaires. Nous avons ensuite exploité la MIADB pour l'étude phytochimique de deux Apocynaceae : *Geissospermum laeve* (Vell.) Miers et *Alstonia balansae* Guillaumin.

Dans un premier temps, l'annotation par la MIADB d'un extrait alcaloïdique des écorces de *G. laeve* a permis l'isolement et l'élucidation structurale de 3 nouveaux AIMs, dont deux à motif butyrolactone. Par la suite, notre approche a été améliorée par l'emploi du nouvel outil d'annotation *in silico* MetWork, fondé sur une prédiction métabolique et la modélisation de spectres de masse tandem. C'est ainsi que l'exploration de l'espace chimique d'un extrait alcaloïdique des feuilles d'*A. balansae* a permis l'identification, puis l'isolement, de 5 nouveaux AIMs du type *N*-oxyde-sarpagane. La stéréochimie des nouveaux composés a pu être déterminée par l'exploitation de spectres prédits et expérimentaux de dichroïsme circulaire électronique.

Ce manuscrit décrit, après une introduction consacrée notamment aux emplois du *molecular networking* pour le ciblage et la découverte de petites molécules naturelles puis aux interconnexions biosynthétiques en série AIM, la création puis l'évolution de la MIADB, puis son utilisation dans un *workflow* de déréplication efficace et de ciblage de nouveaux composés dans des mélanges complexes issus d'Apocynaceae.

Title: Exploration of the chemical diversity of Apocynaceae plants using molecular networking: From the creation of a spectral database to *in silico* annotations

Keywords: Apocynaceae, database, dereplication, mass spectrometry, Molecular Networking, monoterpene indole alkaloids

Abstract: Monoterpene indole alkaloids (MIAs) constitute a class of natural products that has been extensively studied due to its important pharmacological and therapeutic potentials, and to its large structural diversity. Dereplication techniques based on liquid chromatography coupled to tandem mass spectrometry have recently evolved, with the implementation of molecular networking-based approaches.

In this context, we have created a spectral database that encompasses the averaged tandem mass spectra of 172 reference MIAs, in collaboration with several pharmacognosy research teams around the world. This database, named MIADB (standing for Monoterpene Indole Alkaloids DataBase), was made publicly available and can be used as a reference in the application of molecular networking as a dereplication strategy. Thereafter, we used the MIADB to carry out the phytochemical investigation of two Apocynaceae species: *Geissospermum laeve* (Vell.) Miers and *Alstonia balansae* Guillaumin.

As a first application, the MIADB-based annotation of an alkaloid extract of the barks of *G. laeve* led to the isolation and the structural elucidation of three new MIAs, two having a butyrolactone moiety. Afterwards, this approach was improved by the application of a new tool for *in silico* annotation called MetWork, which is based on metabolic prediction and on the generation of predicted tandem mass spectra. Following this approach, the exploration of the chemical space of an alkaloid extract of the leaves of *A. balansae* allowed the anticipation and further isolation of five novel MIAs of the *N*-oxide-sarpagine type. The stereochemistry of all the new molecules could be determined on the basis of experimental and predicted electronic circular dichroism spectra.

In the introduction of this manuscript the multiple uses of molecular networking for the identification of small natural molecules are described, as well as the biosynthetic interconnections in the MIAs group. The creation and evolution of the MIADB are then presented, followed by its utilization in efficient dereplication workflows for the targeting of new natural products within complex mixtures from Apocynaceae species.

