



Identification rapide par RMN de substances naturelles en mélange (dereplication)

Mariacaterina LIANZA et Jean-Marc NUZILLARD

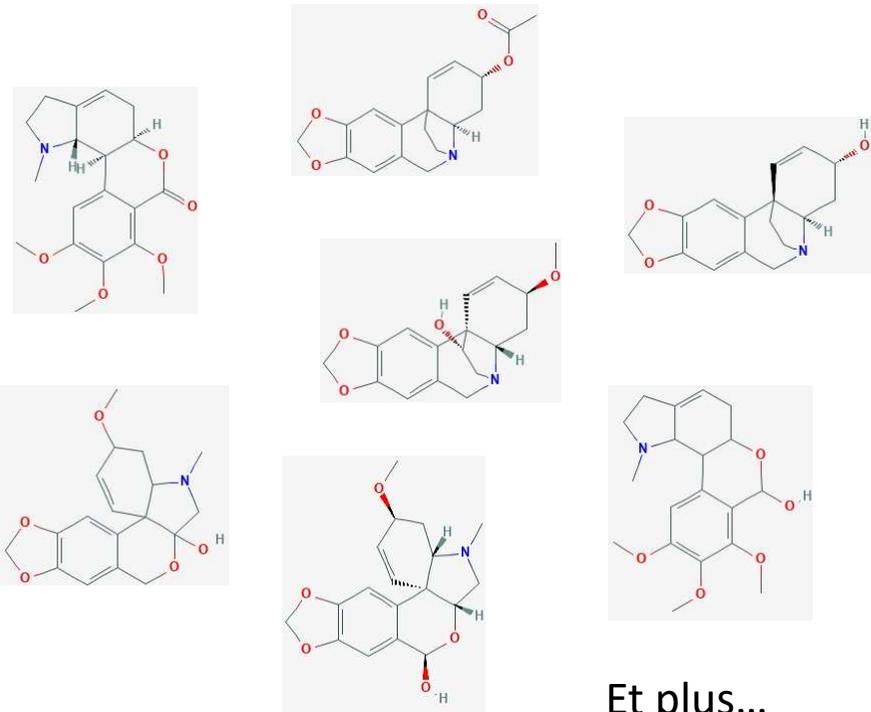
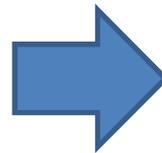
*Institut de Chimie Moléculaire de Reims (ICMR), UMR CNRS 7312
Université de Reims Champagne Ardenne, France
Université de Bologne, Italie*

jm.nuzillard@univ-reims.fr

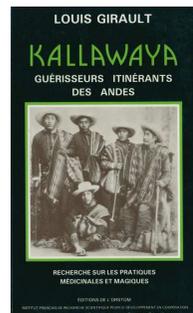


Position du problème

Bulbes de *Urceolina peruviana* (Amaryllidaceae)



Louis Girault (ORSTOM/IRD), dans son livre "Kallawayá, guérisseurs itinérants des Andes: recherches sur les pratiques médicinales et magiques" mentionne que les bulbes de *U. peruviana* sont mélangés à la de la graisse de porc ou de lama pour réaliser un onguent destiné à **traiter les tumeurs et les abcès**.



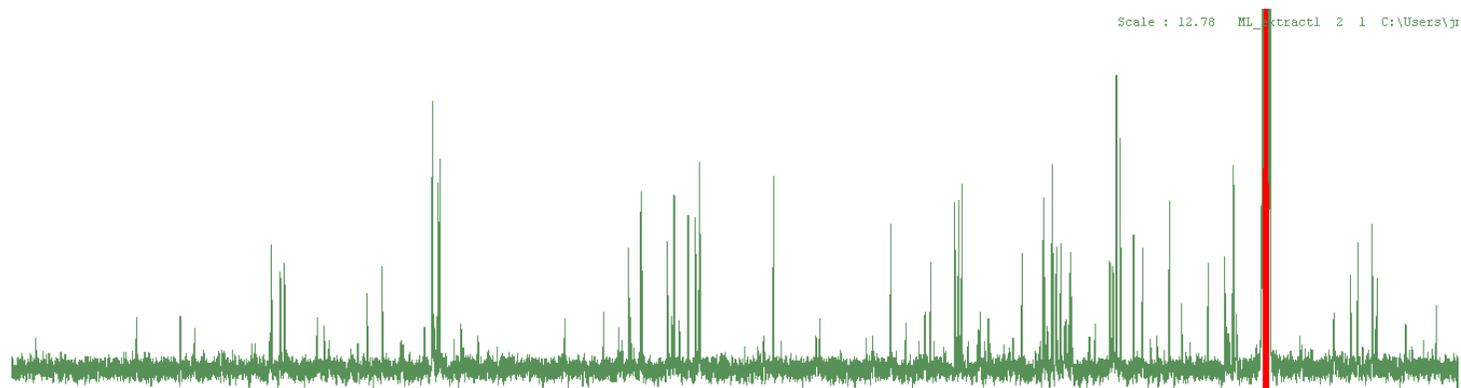


Extraction des bulbes

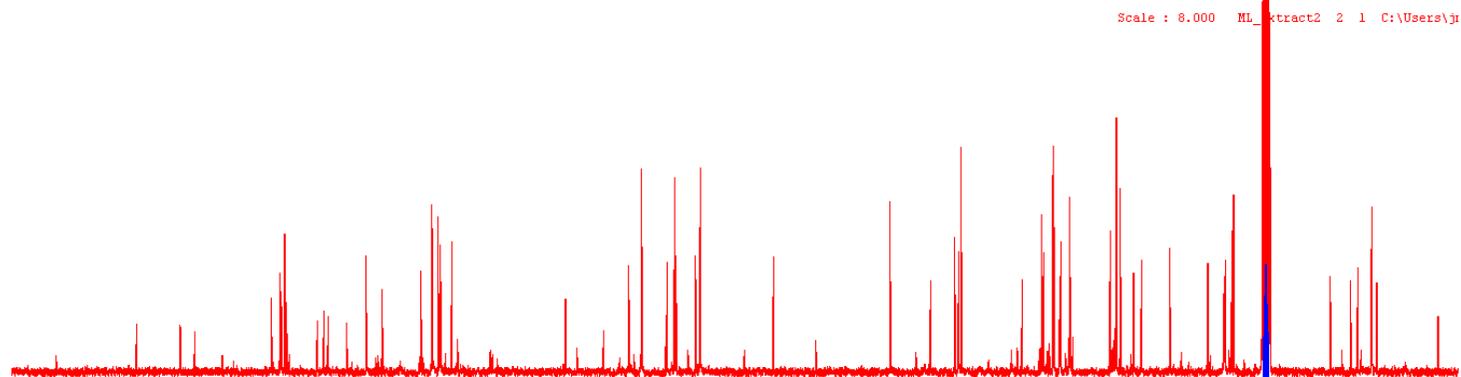
- Matériel végétal : bulbes lyophilisés et broyés
- Deux protocoles d'extraction
 - Méthode I, peu sélective
 - *Natural product research* **2014**, 28(10), 704-710
 - Sur un bulbe (1,3 g) -> **Extrait « 1 » (61 mg)**
 - Méthode II, adaptée aux **alcaloïdes**
 - Patent WO **2006/064105 A1**, préparation de la galanthamine
 - Sur un bulbe (1,3 g) -> **Extrait « 2 » (20 mg)**
 - Sur 270 g de bulbes -> **Extrait « 3 » (2,74 g)**



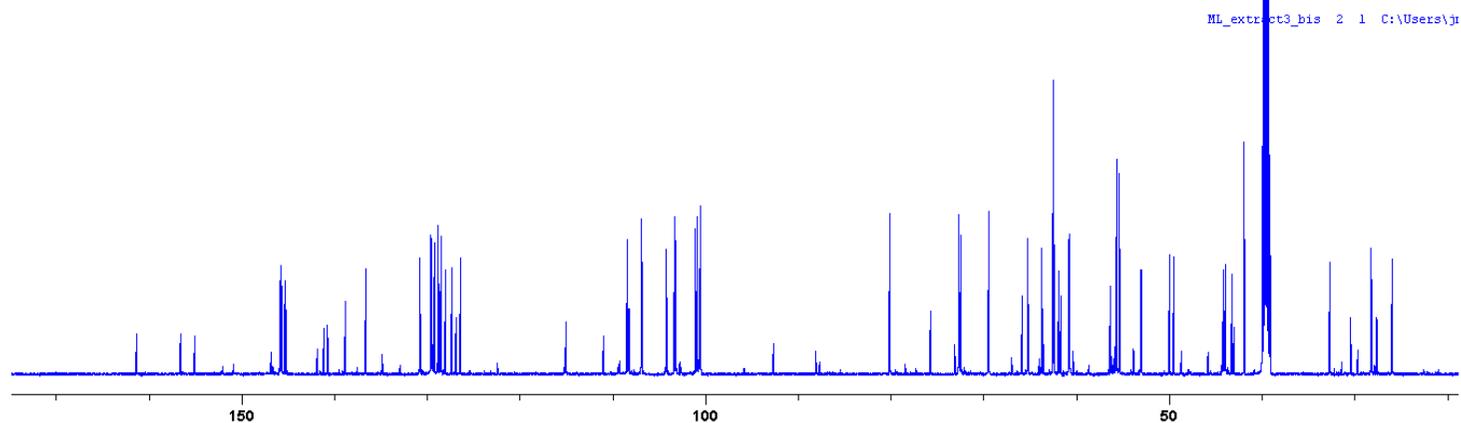
Extraits, RMN du ^{13}C , DMSO- d_6



Extrait 1



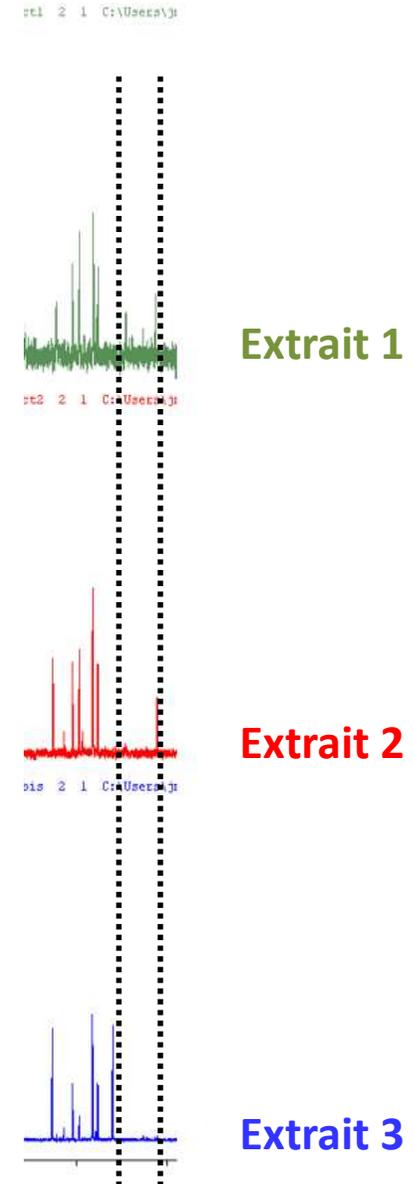
Extrait 2



Extrait 3

Extraits, RMN du ^{13}C , DMSO- d_6

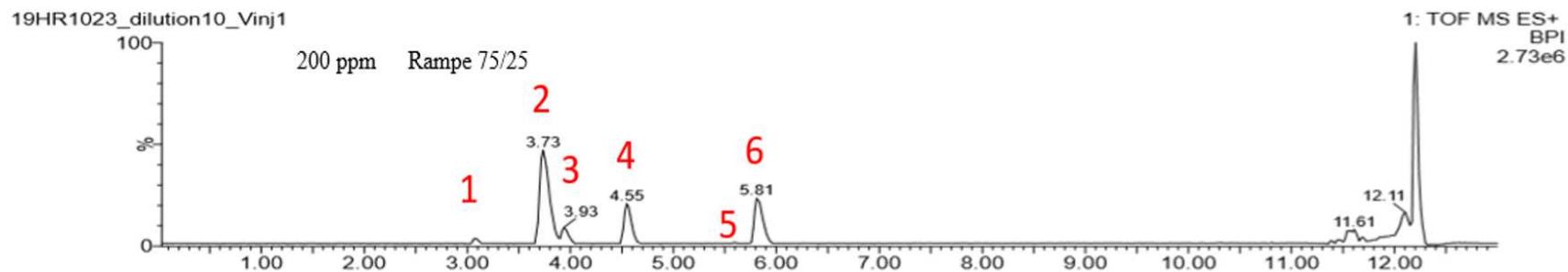
- Extraits 1 et 2 :
 - Produits avec les mêmes masses de bulbe mais des protocoles différents
- Extraits 2 et 3 :
 - Produits avec le même protocole mais avec des masses différentes
- => Importance de l'extraction sur la nature des composés à identifier...



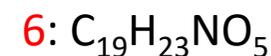
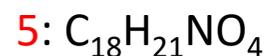
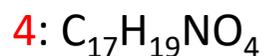
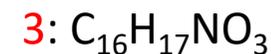
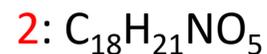
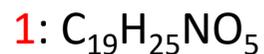


Etude préliminaire de l'extrait 2

- En attendant la préparation de l'extrait 3
- Etude par UPLC, puis par UPLC-HRMS



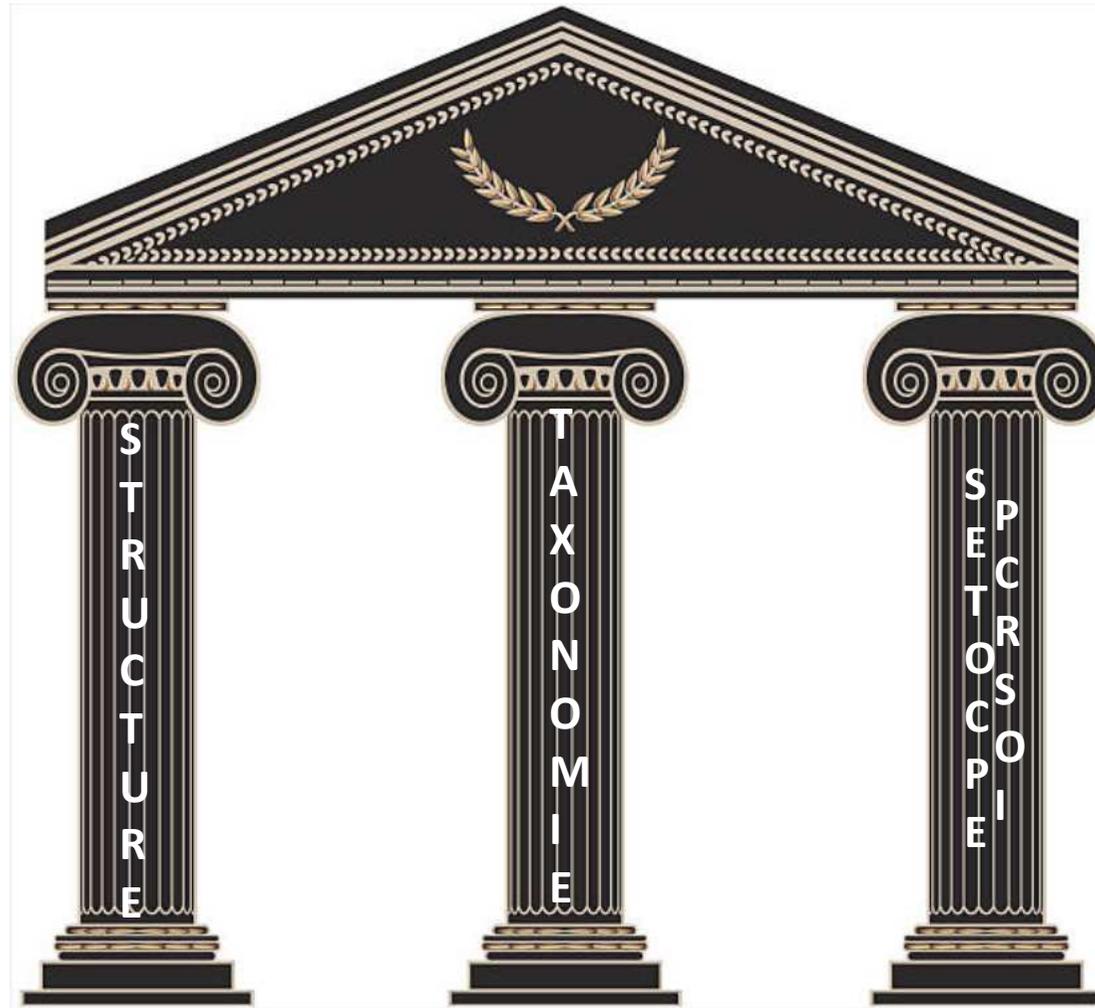
- Propositions de formules brutes à partir de $[M+H]^+$



Comment identifier les composés connus ?



Les 3 piliers de la déréplication





Créer une base de données contenant des structures de molécules isolées dans des Amaryllidaceae

- PubChem (103 millions de composés)
- Se restreindre au produits naturels
 - Dictionary of Natural Products
 - Bases de données spécialisées, Dictionary of Alkaloids
 - CH-NMR-NP (JEOL)
 - ZINC « Natural Products » (?)
 - COCONUT : doi:10.20944/preprints201912.0332.v1
 - ...



Structure + Taxonomie

- KNApSACK
 - [Plant Cell Physiol.](#) 2012 Feb;53(2):e1.
 - http://www.knapsackfamily.com/knapsack_core/top.php

Select by ...

ALL Types Organism Metabolite Molecular formula
 C_ID CAS_ID INCHI-KEY INCHI-CODE SMILES

last update	2020/01/06
metabolite	51179 entries
metabolite-species pair	116314 entries
species	22943 entries



KNApSACK, exemple de recherche

Fichier Édition Affichage Historique Marque-pages Outils Aide

KNApSACK Core System × KNApSACK Metabolite Information × KNApSACK Metabolite Information × +

www.knapsackfamily.com/knapsack_core/result.p



input type = **metabolite** , input word = **crinine**

Number of matched data :7

C ID	CAS ID	Metabolite	Molecular formula	Mw
C00024357	80665-67-4	6alpha-Hydroxy crinine	C16H17NO4	287.11575804
C00024358	80665-68-5	6beta-Hydroxy crinine	C16H17NO4	287.11575804
C00024372	23367-61-5	(-)-Cherylline (S)-(-)-Cherylline Cherylline Cheryllin Crinine (C17 alkaloid) Crinine	C17H19NO3	285.13649348
C00024384	510-67-8	Crinine Crinidine	C16H17NO3	271.12084342
C00024416	93452-26-7	3-O-Acetyl crinine Krepowine O-Acetyl crinine Krepowine	C18H19NO4	313.1314081
C00025196	4684-32-6	Picrinine Deacetyldeformopicraline	C20H22N2O3	338.16304258
C00027665	82260-04-6	12-Demethoxytabernulosine 10-Methoxy picrinine	C21H24N2O4	368.17360727

All rights reserved. © 2007 NARA INSTITUTE of SCIENCE and TECHNOLOGY



Automatisation de la recherche dans KNApSACK

- Etablissement d'une liste de genres pour une famille
 - Exemple : tous les genres d'Amaryllidaceae (*Amaryllis*, *Narcissus*, ...)
 - A partir de wikipedia, ou du « NCBI taxonomy browser »
- Recherche des paires (composé, *Genre espèce*)
 - Exemple : (C00001576, *Clivia miniata*)
- Construction des paires (composé, liste de *Genre espèce*)
 - Exemple : (C00001567, *Zephyranthes carinata* | *Zephyranthes grandiflora*)
- Recherche des données relatives à chaque composé
 - Formule brute, SMILES, InChI, InChIKey, Masse molaire, ...
- *KNApSACK* contient environ 50.000 composés (uniques ?)
 - et n'est donc pas exhaustif (300.000 composés organiques naturels ?)



Production des structures 2D

- Générateur de coordonnées 2D à partir des SMILES
- RDKit (rdkit.org/)
- Python (python.org/)



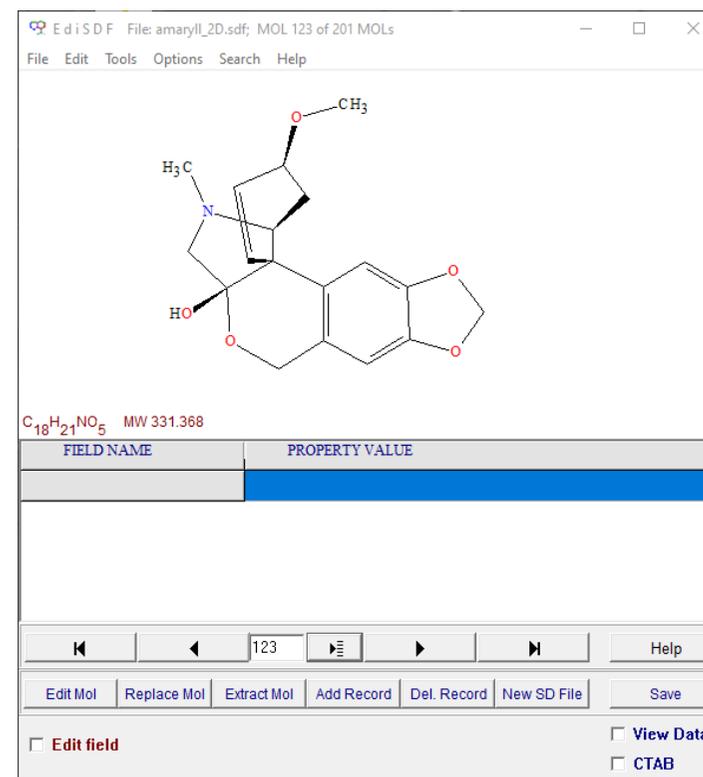
Open-Source Cheminformatics
and Machine Learning



- Visualisation : EdiSDF (gratuit)

Tazettine

c12c(cc3c(c1)[C@]14[C@](OC3)(CN([C@H]1C[C@@H](C=C4)OC)C)O)OCO2



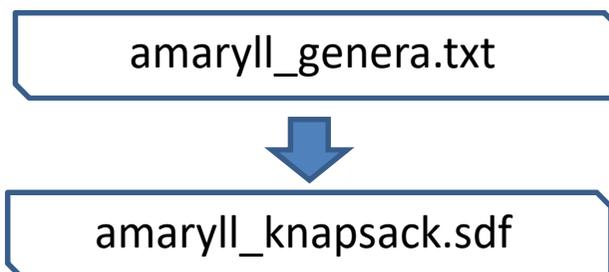


Spectroscopie : La RMN du ^{13}C

- Pourquoi la RMN du ^{13}C ?
 - Un carbone -> 1 pic et 1 pic -> 1 carbone (sauf symétrie ou accident)
 - Pics fins par rapport à la largeur spectrale, superpositions rares
 - Sensibilité : RMN 600 (150 MHz pour ^{13}C) et cryo-sonde ^{13}C
- Comment ?
 - Données de la littérature.
 - Difficiles à collecter, incomplètes, pas toujours très fiables
 - Utiliser des données prédites
 - **ACD/CNMR Predictor**
 - CSEARCH/NMRPREDICT
 - ChemDraw
 - NMRShiftDB

KnapsackSearch : Structure + Taxonomie + Spectroscopie

- <https://github.com/nuzillard/KnapsackSearch/>



- Utilisable tant que le site web de KNAPSAcK ne change pas le format des réponses aux requêtes.
- Le fichier *familyname_knapsack.sdf* contient
 - La structure 2D (avec indication des configurations des carbones asymétriques) des composés trouvés pour l'ensemble des genres.
 - La ou les organismes vivants qui ont produit chaque composé
 - Les δ du ^{13}C prédits par NMRShiftDB pour chaque composé



Déréplication « naïve » de l'extrait 2

- Principe, script « matchPP_MFs.py »
 - Lister les δ expérimentaux en RMN du ^{13}C , sans les intensités
 - Créer 6 fichiers *Formule.sdf* à partir de *amaryll_knapsack.sdf*, chacun correspondant à une formule brute issue de l'analyse UPLC-HRMS
 - Pour chaque composé de chaque fichier *Formule.sdf*:
 - Rechercher le nombre N_{ok} de valeurs de δ prédites qui « correspondent » à la liste des δ expérimentaux (du mélange)
 - Calculer le score $N_{\text{ok}} / N_{\text{C}}$, où N_{C} est le nombre d'atomes de carbone du composé
 - La correspondance des δ prédits avec les δ expérimentaux est définie à partir d'une différence considérée comme acceptable ou non (1,5 ppm)
 - Trier le contenu de chaque fichier *Formule.sdf* par score décroissant
 - Regarder les premières structures de chaque fichier...



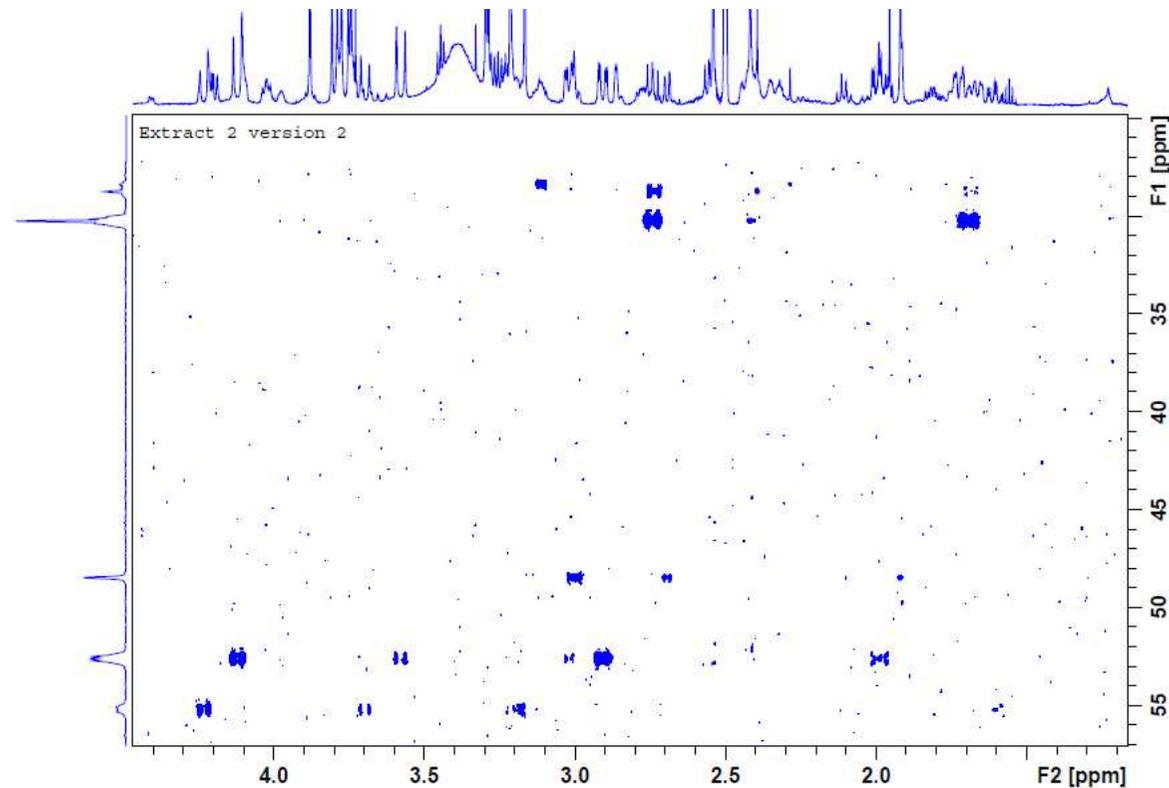
Déréplication « naïve » de l'extrait 2

Formule	Nombre de composés	Composé	Score	Formule	Nombre de composés	Composé	Score
C ₁₆ H ₁₇ NO ₃	3	Crinine	1.000	C ₁₈ H ₂₁ NO ₅	12	Pseudolycorine 1-acetate	1.000
		Vittatine	1.000			Pseudolycorine 2-acetate	1.000
C ₁₇ H ₁₉ NO ₄	11	Crinamine	1.000			Steinbergine	1.000
		Haemanthamine	1.000			Tazettine	1.000
		Hippamine	1.000			Criwelline	1.000
		Montanine	1.000	C ₁₉ N ₂₃ NO ₅	2	Albomaculine	0.947
C ₁₈ H ₂₁ NO ₄	5	Norpluviine 1-acetate	1.000	C ₁₉ H ₂₅ NO ₅	2	Ungvedine	0.895
		Oduline O-Me	1.000				



Extrait 2 de *Urceolina peruviana*, HMBC ^1H - ^{15}N

- Les alcaloïdes d'Amaryllidacea ne comportent (en général) qu'un seul atome d'azote
 - Le spectre HMBC ^1H - ^{15}N permet de compter les composés majoritaires





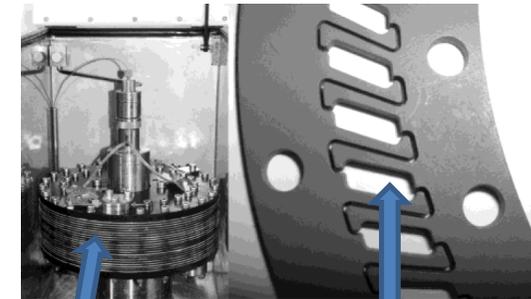
Déréplication « CAMEL »

- CAMEL : CARActérisation de MELanges
- *Anal. Chem.* **2014**, 86, 2955-2962. doi: 10.1021/ac403223f
- Méthode :
 - Fractionnement par CPC/EPC (10-20 fractions)
 - Spectres de RMN du ^{13}C
 - « Peak picking » et « bucketting »
 - Recherche de groupes de δ qui ont le même profil chromatographique
 - Recherche, groupe par groupe, des composés qui correspondent
- La procédure CAMEL est à l'origine de la création de la société Nat'Explore (<https://nat-explore.com/>)
- La base de données associée à CAMEL contient environ 4000 composés et a été créée avec **CNMR Predictor**

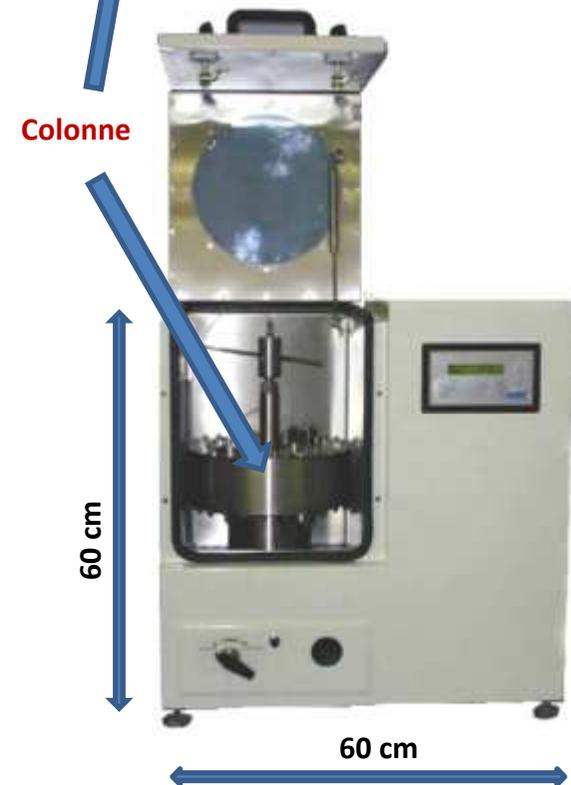
CPC / EPC

- **Extraction/Chromatographie de Partage Centrifuge**
- Partition des analytes entre deux phases liquides
- La colonne est constituée de centaines de cellules de partage
- La phase stationnaire est maintenue par la force centrifuge
- Les analytes sont injectés en tête de colonne
- La phase mobile percole au travers de la phase stationnaire
- Pas d'adsorption irréversible sur une phase solide
- Tout ce qui rentre finit par sortir
- Chromatographie en mode élution (graduée) et déplacement
- Débit élevé, typiquement 20 mL/mn
- Il est possible d'injecter 5 g dans une colonne de 200 mL

- **Technique préparative**



Cellule de partage



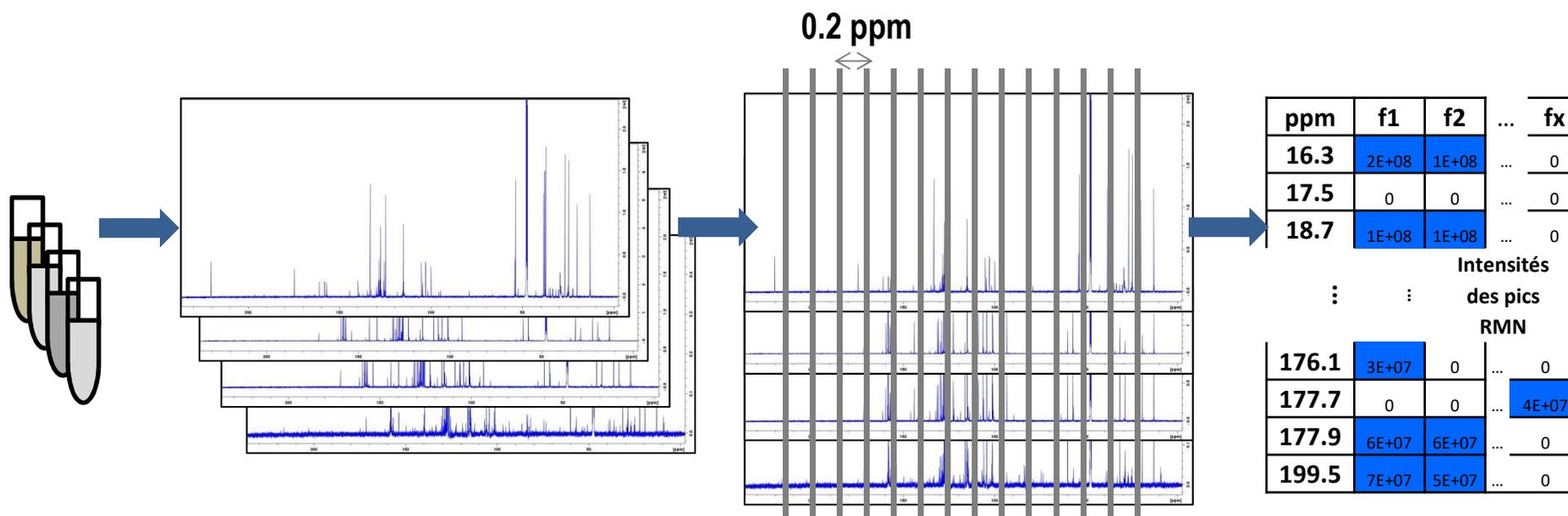
Colonne

60 cm

60 cm



Déréplication « CAMEL »

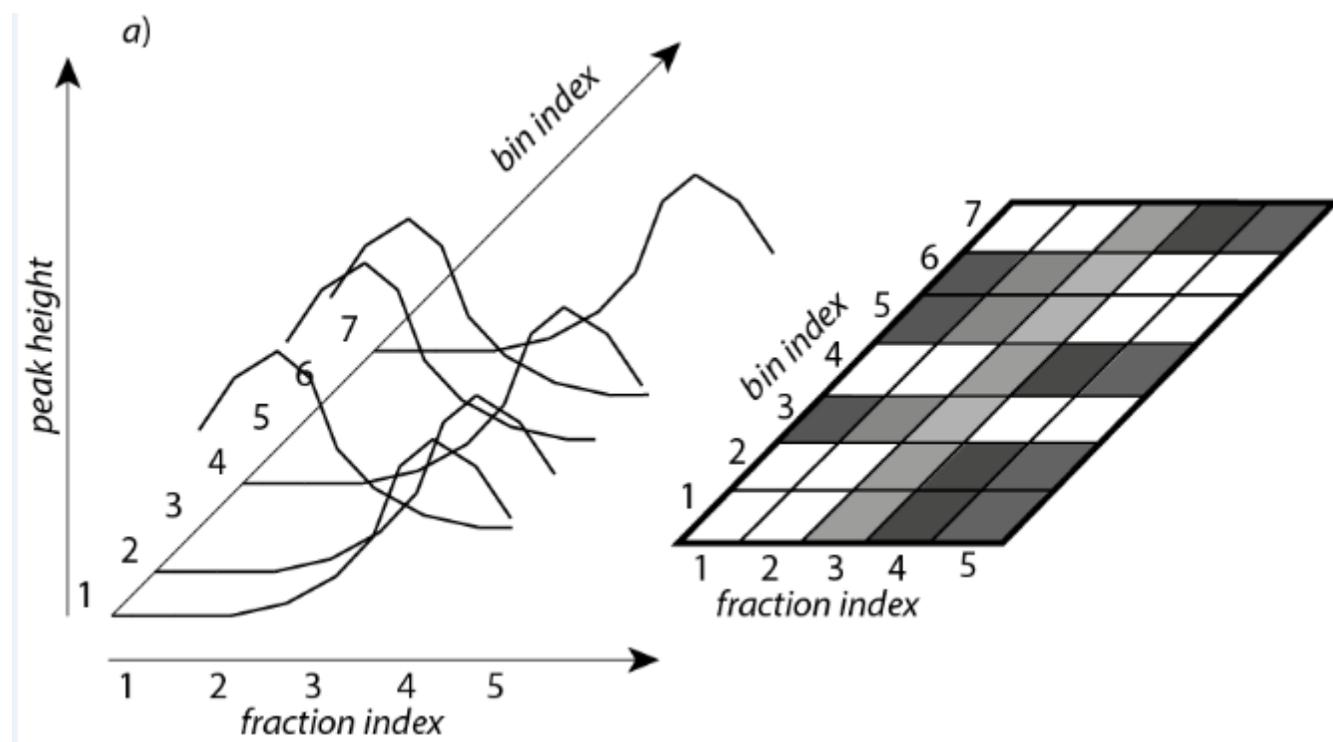


Analyse RMN ¹³C

- Peak picking
- Conversion de format de fichier
- Alignement par bucketting



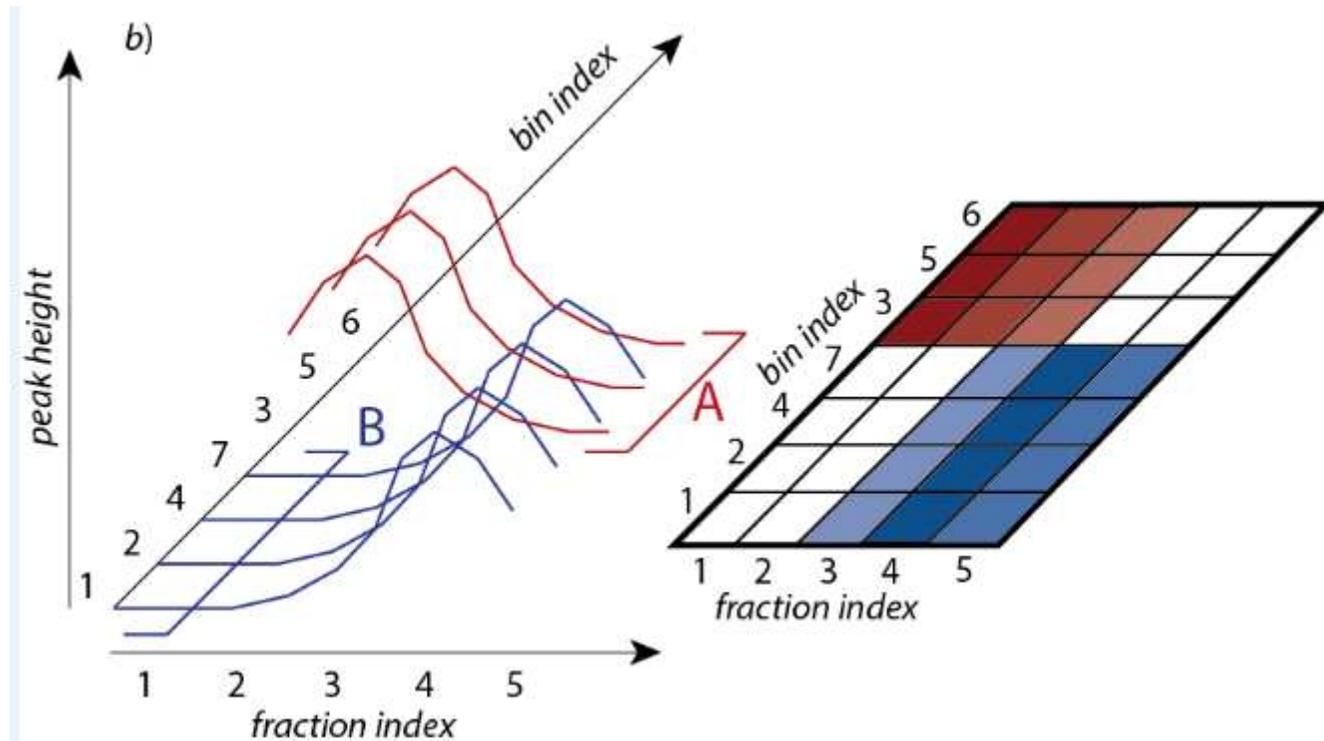
Agrégation (« clustering ») des profils chromatographiques



AVANT



Agrégation (« clustering ») des profils chromatographiques

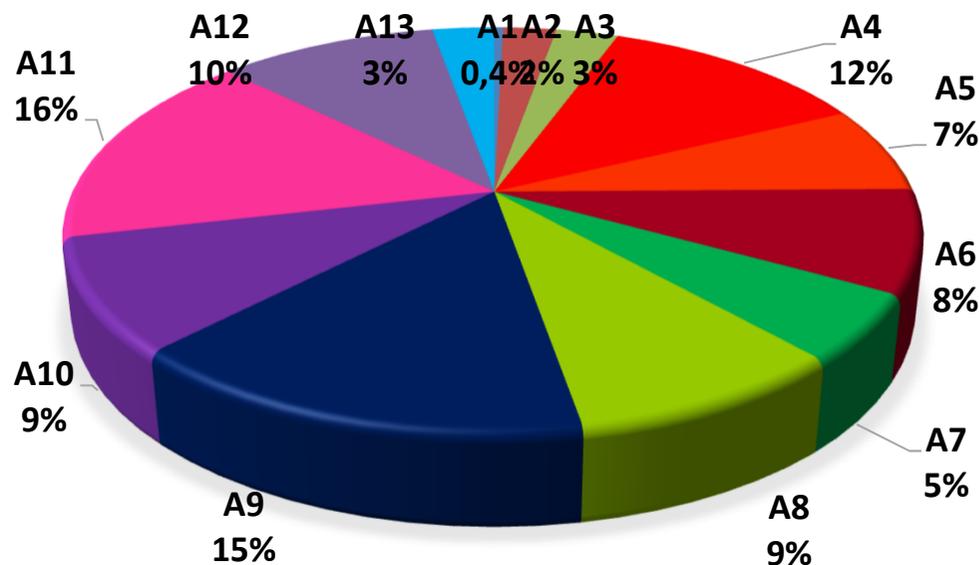


APRÈS

« CAMEL » sur l'extrait 3 de *U. peruviana*

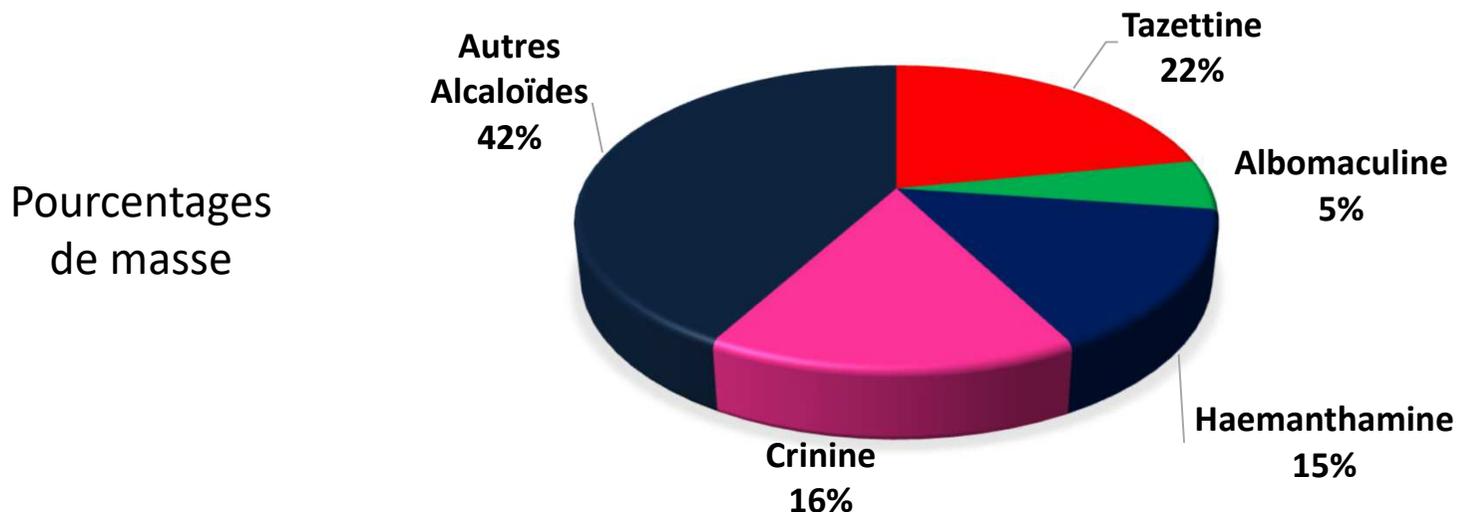
- Obtention de 13 fractions par EPC, mode « pH-zone refining »
 - Système biphasique de solvants MtBE, CH₃CN et H₂O, 5:2:3 (v/v)
 - Extrait injecté, **1 g**, en phase stationnaire aqueuse (acidifiée par H₂SO₄ 10 mM)
 - Déplacement des alcaloïdes dans la phase mobile organique (alcalinisée par NEt₃, 8mM) en fonction des pK_a des coefficients de partage de la forme neutre (base)
 - Les fractions de phase mobile sont analysées par CCM et regroupées
 - Fractions A1 à A13.

Pourcentages
de masse



« CAMEL » sur l'extrait « 3 » de *U. peruviana*

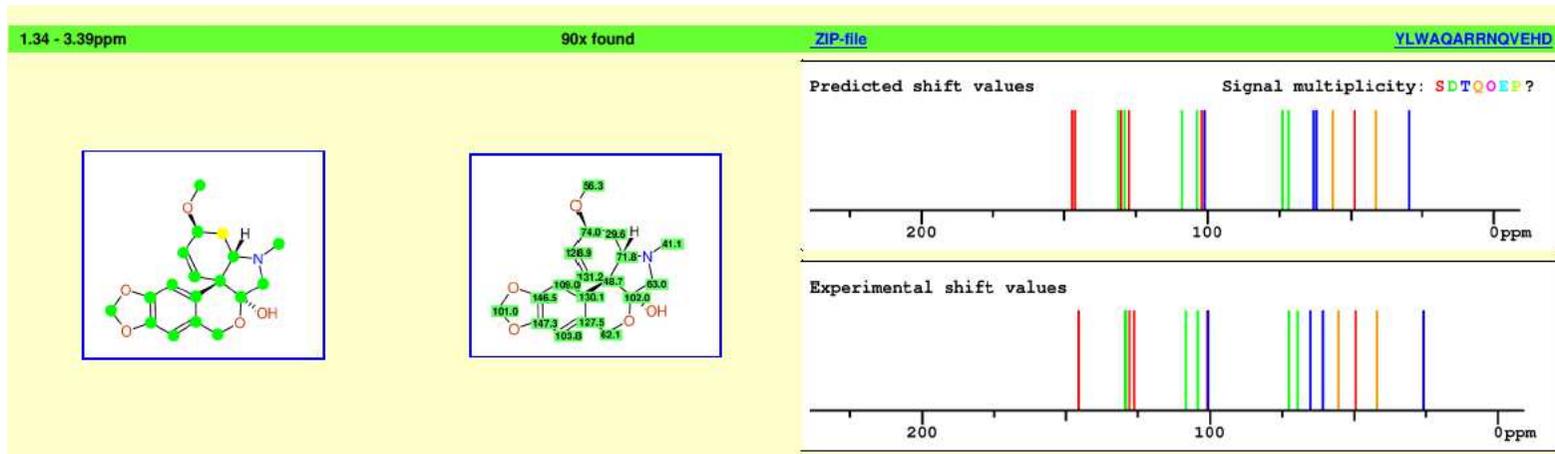
- Analyse par RMN
 - ^1H , ^{13}C , ^1H - ^1H COSY, ^1H - ^{13}C HSQC, ^1H - ^{13}C HMBC, ^1H - ^1H ROESY
 - Les fractions A3 à A5 sont identiques et sont « presque » pures
 - Les fractions A7 et A9 sont « presque » pures
 - La fraction A11 comporte un produit très majoritaire
 - **Détermination directe de la structure des composés des fractions A4, A7, A9, A11**
 - Les fractions A2, A6, A8, A10, A11 et A12 sont plus complexes (transitions)





Composé des fractions A3 à A5

- 18 valeurs de δ relevées dans le spectre de RMN du ^{13}C de A4
- Le spectre HSQC aide à associer une multiplicité (*s*, *d*, *t*, *q*) à chaque valeur de δ
- Recherche dans CSEARCH (web).

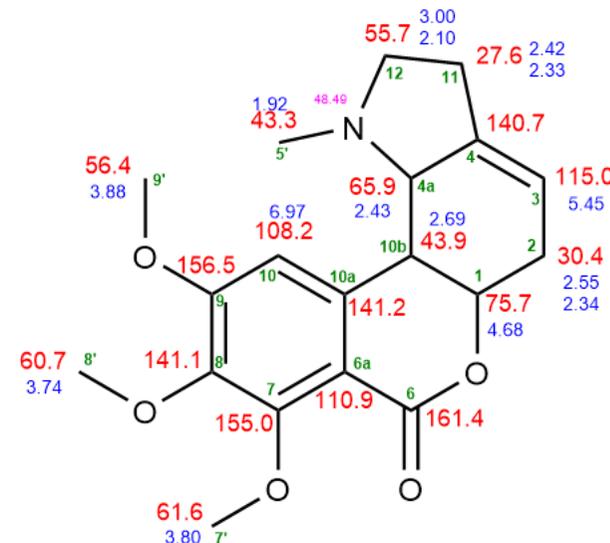
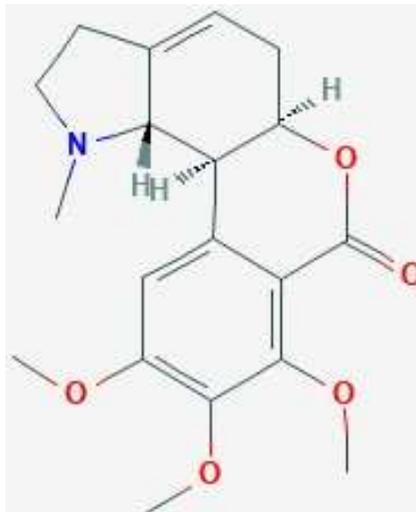


- La structure de la Tazettine est classée comme la plus probable



Composé de la fraction A7

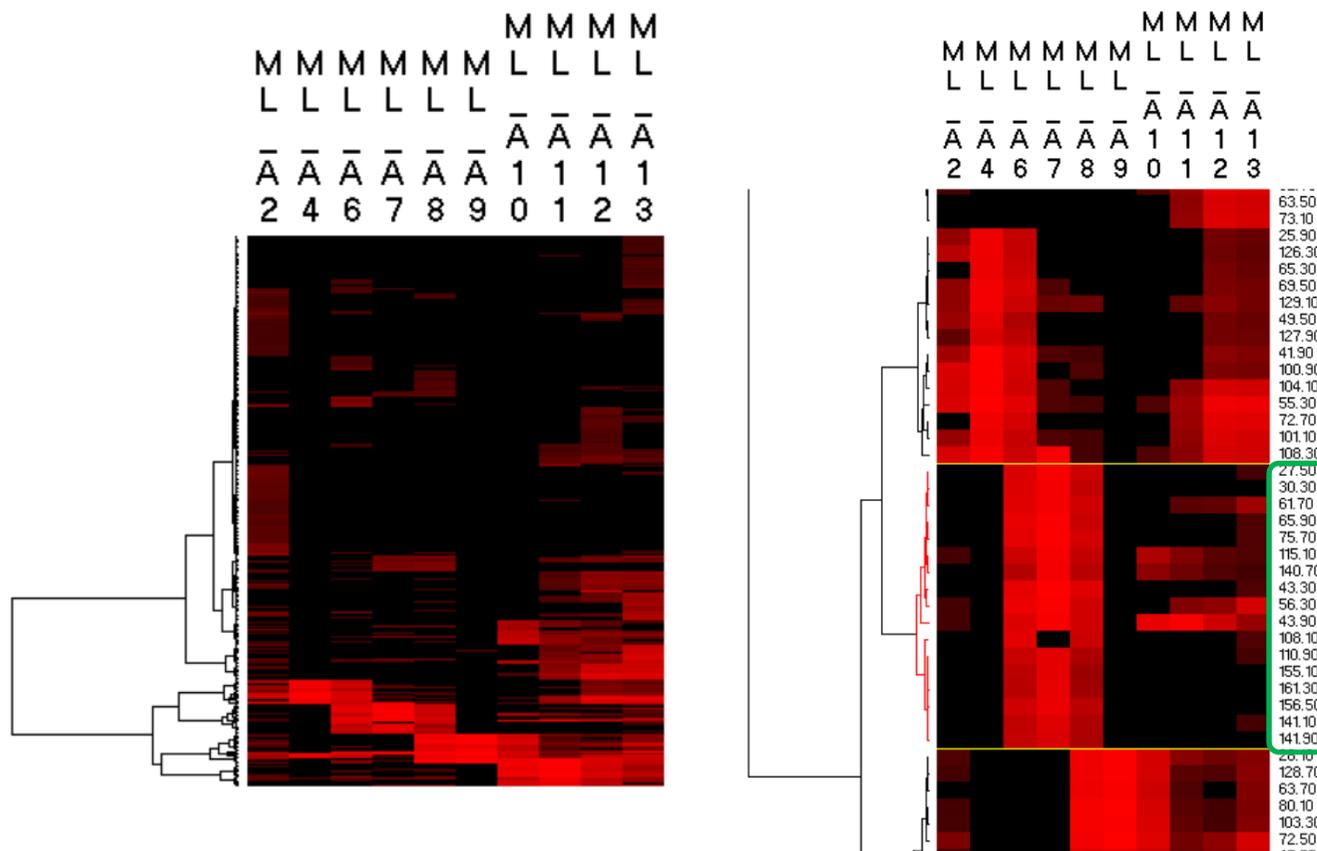
- 19 valeurs de δ et de multiplicités relevées dans le spectre de RMN du ^{13}C
- Recherche dans CSEARCH (web) peu concluante.
- KNApSACK contient 2 composés $\text{C}_{19}\text{H}_{23}\text{NO}_5$ et 2 composés $\text{C}_{19}\text{H}_{25}\text{NO}_5$.
- Le composé contient 3 groupes $\text{CH}_3\text{-O-Aryl}$
- Un seul composé correspond dans KNApSACK : l'albomaculine
- Validation par les autres spectres de RMN (1D et 2D)





« CAMEL », *Extrait 3, Classification par profil chromatographique*

- Classification par « PermutMatrix », *Bioinformatics* **2005**, 21, 1280-1281.



Ensemble de valeurs de δ associées au même profil Chromatographique

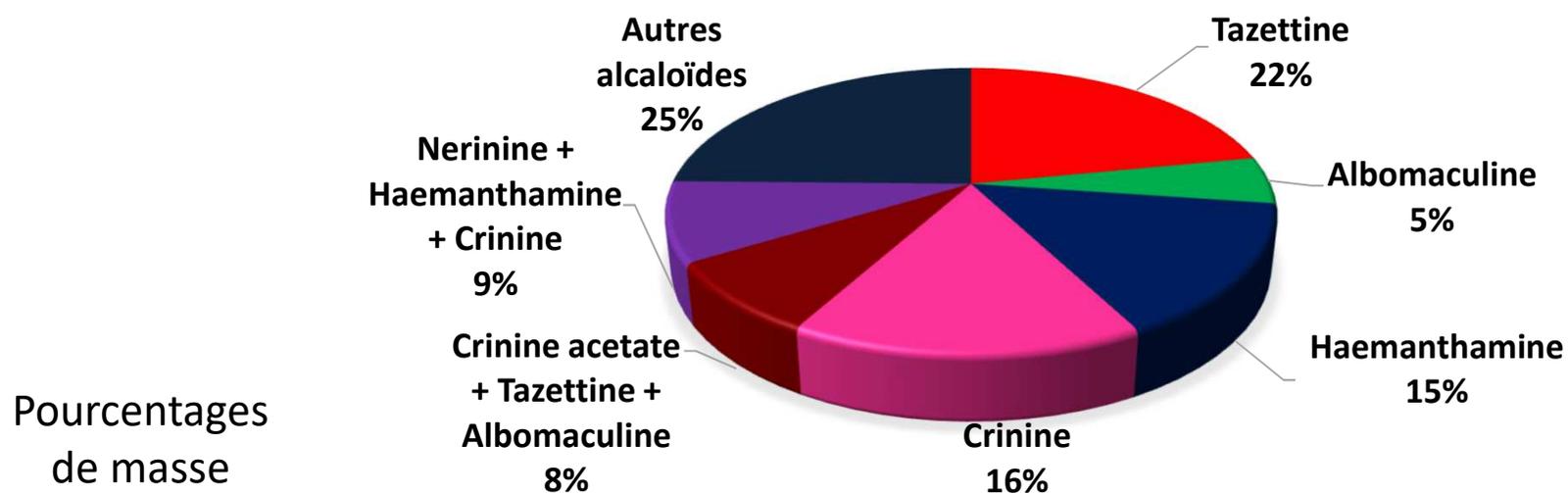
=>

Recherche avec CNMR Predictor



« Work in progress »

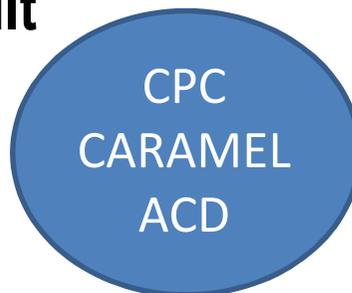
- Construction d'une base de données des composés isolés des Amaryllidaceae, avec déplacements chimiques prédits par **CNMR Predictor**
- Etude des composés minoritaires de l'extrait 3, présents dans les fractions A2 et A10 à A13
- Compléter la liste des composés identifiés dans *Urceolina peruviana*, connus (7) et inconnus (0) à ce jour.





REMERCIEMENTS

- Pr Jean-Hugues Renault
- Dr Jane Hubert
- Dr Alexis Kotland
- Nicolas Borie



- Dr Simon Rémy
- Carine Machado
- Dominique Harakat



- Agathe Martinez
- Anthony Robert



Ministère de l'Enseignement supérieur,
de la Recherche et de l'Innovation





Natural Product Team in Reims

