



# Bases de données pour l'identification des métabolites spécialisés par RMN du $^{13}\text{C}$

**Jean-Marc Nuzillard**

*Institut de Chimie Moléculaire de Reims (ICMR), UMR CNRS 7312, SFR CAP Santé,  
Université de Reims Champagne-Ardenne, France*

[jm.nuzillard@univ-reims.fr](mailto:jm.nuzillard@univ-reims.fr)

## Analyse des mélanges de composés organiques

- Intérêt grandissant pour les sources de matières premières renouvelables issues de **plantes**
- Médicaments inspirés de **plantes**
- Ingrédients cosmétiques issus de **plantes**
- **Mais**: Les plantes produisent rarement des composés chimiquement purs
- **L'analyse des mélanges joue un rôle central en chimie des plantes.**

Identification des  
composés connus

=

Déréplication structurale

=

Davantage de temps consacré  
à l'étude des composés nouveaux



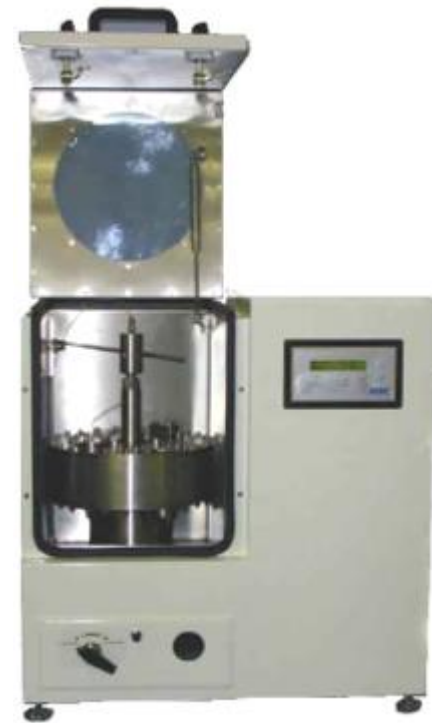
## PROCÉDURE « CAMEL »

- Fractionnement d'un extrait par Chromatographie de Partage Centrifuge (CPC) → 10 à 15 fractions
- RMN du  $^{13}\text{C}$  des fractions
- “Binning” des spectres (1200 fenêtres de 0.2 ppm de large)
- Tableau des intensités : chaque ligne correspond au déplacement chimique associé au centre de la fenêtre et chaque colonne à une fraction chromatographique
- Permutation des lignes pour regrouper les déplacements chimiques ayant des profils chromatographiques identiques
- Détermination de clusters de déplacements chimiques attribuables à un seul composé
- **Identification des composés à partir d'une base de données**



## CPC

- **Centrifugal Partition Chromatography**
- Partition des analytes entre deux phases liquides
- La colonne contient un grand nombre de cellules de partage connectées en série
- La phase stationnaire reste en place par la force centrifuge
- Les analytes sont injectés en tête de colonne
- La phase mobile percole à travers la phase stationnaire
- **Pas d'absorption sur une phase stationnaire solide**
- Elution isocratique ou graduée, "pH-zone refining", échange d'ions
- Tout ce qui rentre finit par sortir, d'un côté ou de l'autre
- **Débits élevés 20 mL/mn**
- Injection de 5g dans une colonne de 200 mL
  
- **Technique préparative**



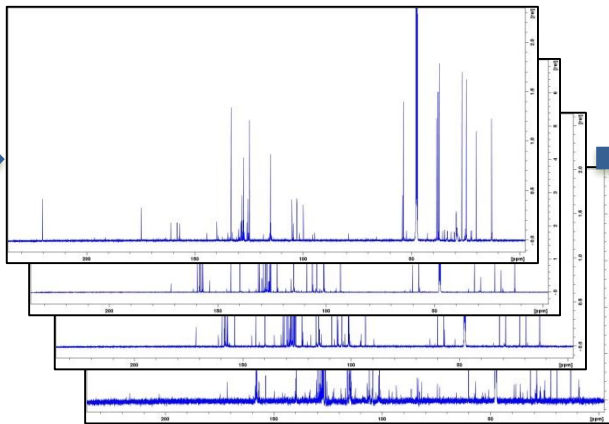
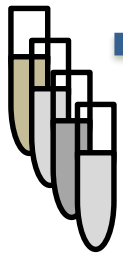


## RMN du $^{13}\text{C}$

- Les fractions de CPC sont analysées par RMN du  $^{13}\text{C}$
- Un carbone, un pic (sauf symétrie ou accident)
- Probabilité limitée de superposition de pics
- Sensibilité faible, a priori
- 600 MHz, cryosonde, bobine  $^{13}\text{C}$  refroidie
- RMN du  $^1\text{H}$ 
  - Il n'y a pas toujours assez de  $^1\text{H}$  à observer
  - Spectres complexes à cause des couplages homonucléaires
- Autres alternatives:
  - RMN  $^1\text{H}$  1D "pure shift" (difficile...)
  - HSQC
  - HMBC

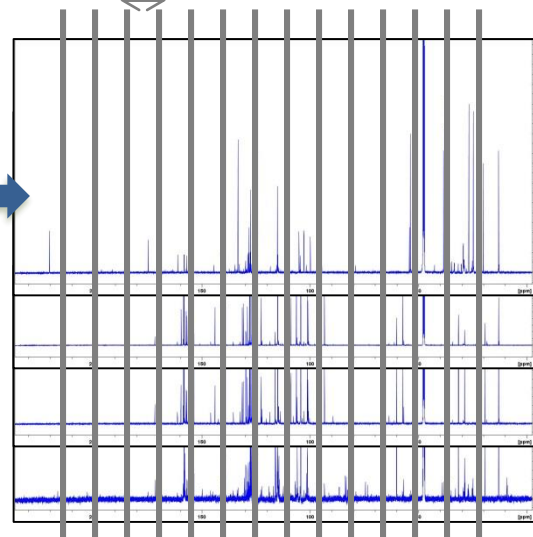
# OBTENTION DU TABLEAU DE DONNÉES DE RMN

Fractions



Analyse par RMN du  $^{13}\text{C}$

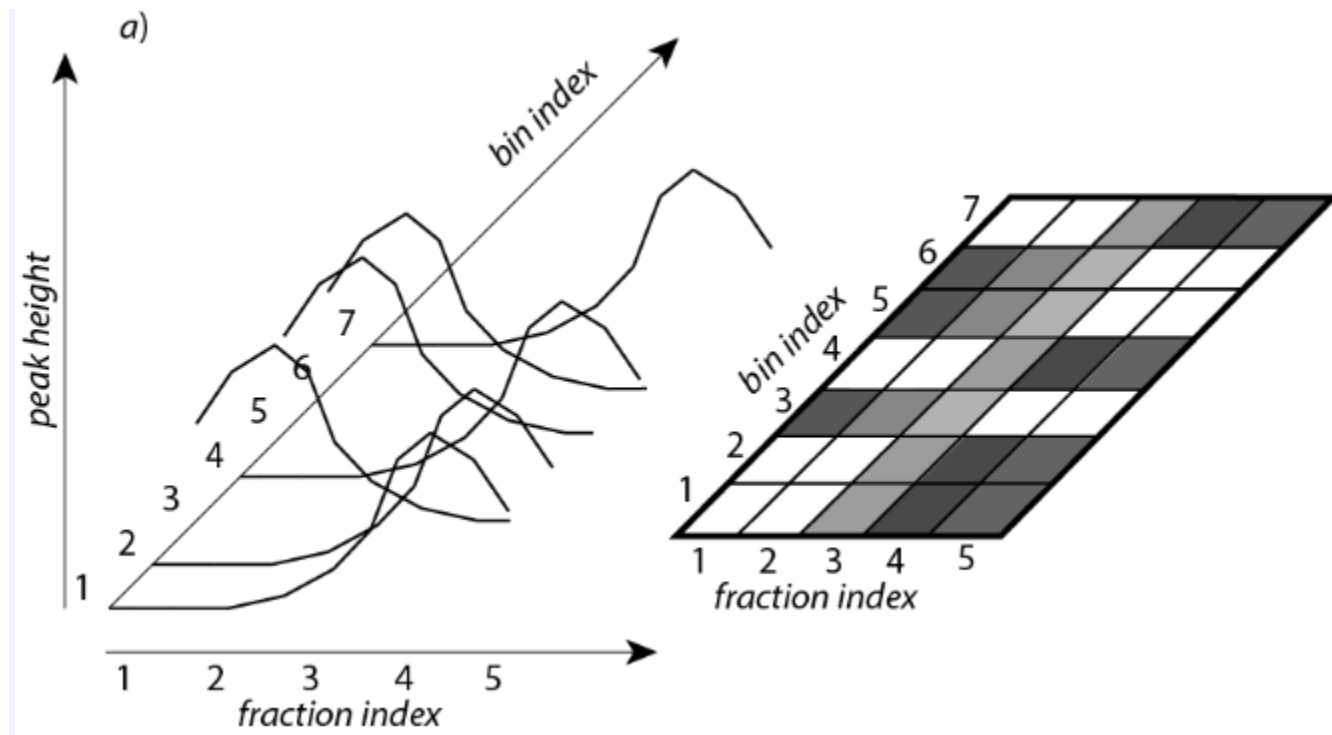
0.2 ppm



ppm	f1	f2	...	fx
16.3	2E+08	1E+08	...	0
17.5	0	0	...	0
18.7	1E+08	1E+08	...	0
⋮	⋮	Intensités des signaux		
176.1	3E+07	0	...	0
177.7	0	0	...	4E+07
177.9	6E+07	6E+07	...	0
199.5	7E+07	5E+07	...	0

- Peak picking automatique
- Conversion de fichier
- Alignement

# DONNÉES BRUTES SIMPLIFIÉES

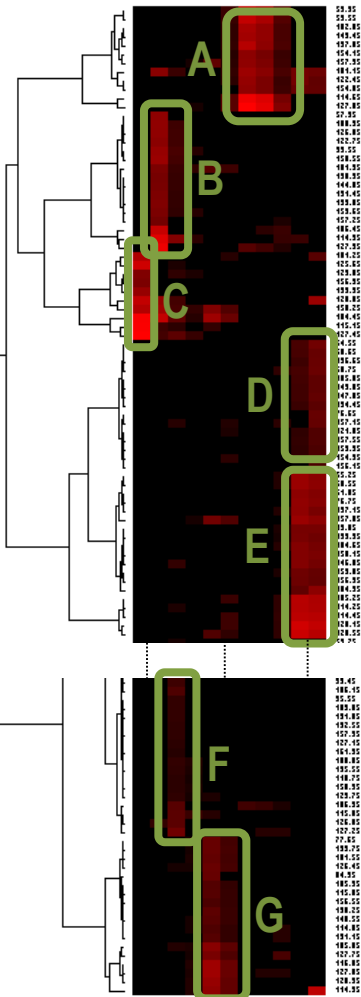






# DONNÉES RÉ-ORDONNÉES PAR CLASSIFICATION ASCENDANTE HIÉRARCHIQUE (CAH)

11 fractions

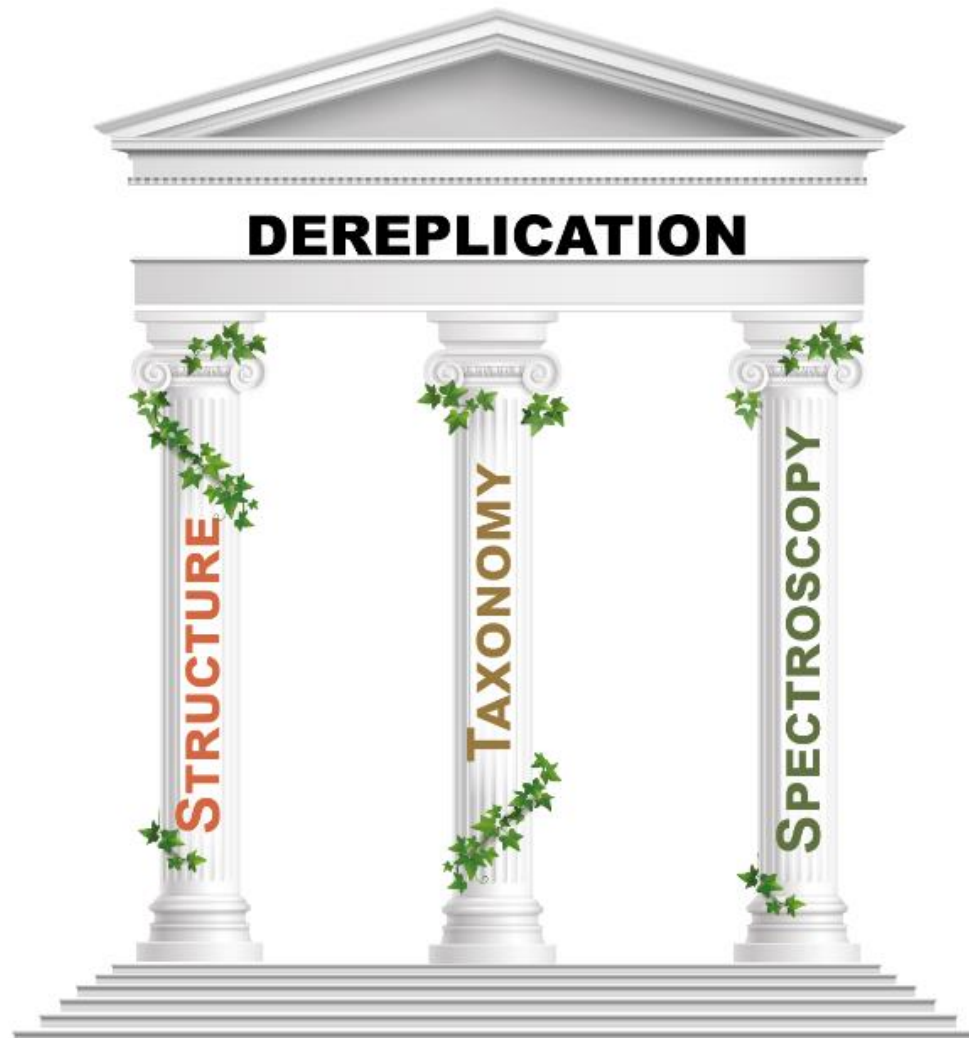


} Groupe de valeurs de déplacements chimiques correspondant à un **composé pur**

- Si une partie de la liste des déplacements chimiques est celle d'un **composé connu**, comment en trouver la structure ?
- La même question se pose lorsqu'un composé a été isolé à l'état pur.
- Y répondre nécessite de disposer de **bases de données**.

A collage of microbiology-related images including petri dishes with bacterial cultures, a petri dish with green mold, and a petri dish with green liquid.

# LES TROIS PILLIERS DE LA DÉRÉPLICATION STRUCTURALE





## STRUCTURE (+ TAXONOMIE)

- Il est nécessaire de chercher des structures de composés connus parmi les composés **naturels** (300k ?)
- **ISDB** [zenodo.org/record/5607264](https://zenodo.org/record/5607264) Environ 200k structures + MS
- **KNAPSAcK** [www.knapsackfamily.com/knapsack\\_core/top.php](http://www.knapsackfamily.com/knapsack_core/top.php) est une première réponse. Environ 50k composés. Fermée.
- **COCONUT** [coconut.naturalproducts.net/](http://coconut.naturalproducts.net/) est plus complète que KS mais ne contient pas que structures de produits naturels (407k composés uniques). Ouverte
- **LOTUS** [lotus.naturalproducts.net/](http://lotus.naturalproducts.net/) poursuit le même but que COCONUT mais est plus sélective (276k composés) et inclut nativement des données de taxonomie biologique et chimiques (classes de substances naturelles) ainsi que des liens vers wikidata [www.wikidata.org/wiki/](http://www.wikidata.org/wiki/). Ouverte



## STRUCTURE + RMN du $^{13}\text{C}$

- Les données **expérimentales** de RMN du  $^{13}\text{C}$  sont **dispersées** dans la littérature chimique, **partielles**, parfois **peu fiables**.
- Il vaut mieux utiliser des données **prédites**.
- Il faut disposer de prédicteurs rapides, autonomes, fiables.
- 200.000 secondes = 2 jours, 7 heures, 33 minutes et 20 secondes
- **Solution 1** : [nmrshiftdb.nmr.uni-koeln.de](http://nmrshiftdb.nmr.uni-koeln.de) avec le code en Java qui permet de réaliser des prédictions localement. Très rapide.
- **Solution 2** : [www.acdlabs.com](http://www.acdlabs.com) et
  - “détournement” du vérificateur de déplacements chimiques expérimentaux (rapide, automatique, 2 composés par seconde)
  - prédiction précise (lente, manuelle, 1 minute par composé).



## DEUX STRATEGIES DE DÉRÉPLICATION

### ➤ **Stratégie non ciblée :**

- La plus grande collection possible de structures est accompagnée de données spectroscopiques prédites et d'informations taxonomiques, une fois pour toutes
- Les résultats d'une recherche de structure sont triés sur des critères taxonomiques si nécessaire

### ➤ **Stratégie ciblée :**

- Les structures sont d'abord triées sur les critères taxonomiques et les **données spectroscopiques** ne sont prédites **à la demande** que sur un petit nombre de structures
- Disposer d'une base de données ciblée, de taille réduite, accélère le travail de recherche et donne immédiatement des propositions de structure pertinentes.



## STRATÉGIE CIBLÉE

- **KnapsackSearch (KS)** [github.com/nuzillard/KnapsackSearch](https://github.com/nuzillard/KnapsackSearch)
  - Pour une famille donnée, fournir une liste de genres
  - Wikipédia est ton ami
  - Un script en Python
    - interroge KNApPSAcK sur tous les genres
    - Collecte les structures trouvées dans la famille
    - Prédit les déplacements chimiques avec nmrshiftdb2
    - Crée un fichier .sdf lisible par ACD/Labs
  - Nécessite une connexion internet
  - Dépend de la structure des données renvoyées par le site de KNApSAcK (en HTML). Solution non pérenne.
  - Dépend de RDKit – <https://www.rdkit.org/>



## STRATÉGIE CIBLÉE

- **VersaDB** [https://github.com/simremy/versadb\\_tk](https://github.com/simremy/versadb_tk)
  - Développé par Simon Remy et Julien Cordonnier, ICMR
  - Interface graphique
  - Sélection par la taxonomie botanique et chimique
  - Interroge à distance LOTUS
  - Données de RMN prédites par nmrshiftdb2
  - Données de MS<sup>2</sup> prédites par CFM-ID 4.0
  - Dépend d'une connexion internet
  - Utilise RDKit
  - Publication soumise...





## STRATÉGIE NON CIBLÉE

- Historique : SISTEMAT, Pr. VdP Emerenciano, Brésil
  - Début des années 1990.
  - Les Trois Piliers étaient déjà en place
  - Logiciels fermés, non maintenus
  
- « Predicted NMR of Natural Products » (PNMRNP), ICMR
  - Créée à partir d'ISDB et complétée par nmrshiftdb2
  - Preuve de principe, usage déconseillé en pratique



## STRATÉGIE NON CIBLÉE

### ➤ **ACD\_LOTUS**

- Structures de LOTUS et complétées par ACD/Labs
  - Utilise l'algorithme de vérification rapide pour la prédiction
- Utilise la librairie RDKit et des outils créés pour KS et PNMRNP
- 218.478 structures dans LOTUS v9
- [zenodo.org/record/7124055](https://zenodo.org/record/7124055) fichier .sdf zippé
- Les données ont été intégrées à nmrshiftdb2 par Stefan Kuhn
- Le fichier .sdf public peut être importé dans ACD/Labs
  - L'utilisation de **acd\_lotusv9.NMRUDB** ne nécessite pas de connexion internet



## EXEMPLE, ACD\_LOTUS

- Sélection de la structure d'un métabolite spécialisé connu
- Copie de la liste des déplacements chimiques du  $^{13}\text{C}$  associés
- Recherche des composés qui correspondent
  - Avec le gestionnaire de bases de données ACD/Labs
  - Avec nmrshiftdb2

## EXEMPLE, ACD\_LOTUS – ELLIPTICINE

- Ellipticine dans la Biological Magnetic Resonance data Bank
- [bmr.io/metabolomics/metabolomics\\_standards.php?dataset=metabolomics](http://bmr.io/metabolomics/metabolomics_standards.php?dataset=metabolomics)

Filter by dataset

NMRFAM highly curated metabolites

Go

FAM : Facility at Madison (Wisconsin)

Jump to molecules beginning with: A B C D E F G H I J K L M N O P Q R S T U V X Z



6-ethylmercaptapurine	(E)-3,7,11,15-tetramethylhexadec-2-en-1-ol	(E)-6-(4-hydroxy-6-methoxy-7-methyl-3-oxo-1H-2-benzofuran-5-yl)-4-methylhex-4-enoic acid
(~)-Eburnamonine	EGTA	elaidic acid
Ellagic acid	<b>Ellipticine</b>	Emodin
(-)-Epicatechin	epichlorohydrin	Epinephrine
epsilon-caprolactam	Epsilon-caprolactone	ergocalciferol

Ellipticine bmse001340 - Data

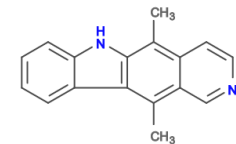
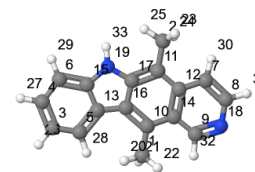
### Ellipticine synonyms

Wikipedia: Ellipticine

Go

Search Wikipedia

Graphical representations:





# ELLIPTICINE

## Set 1

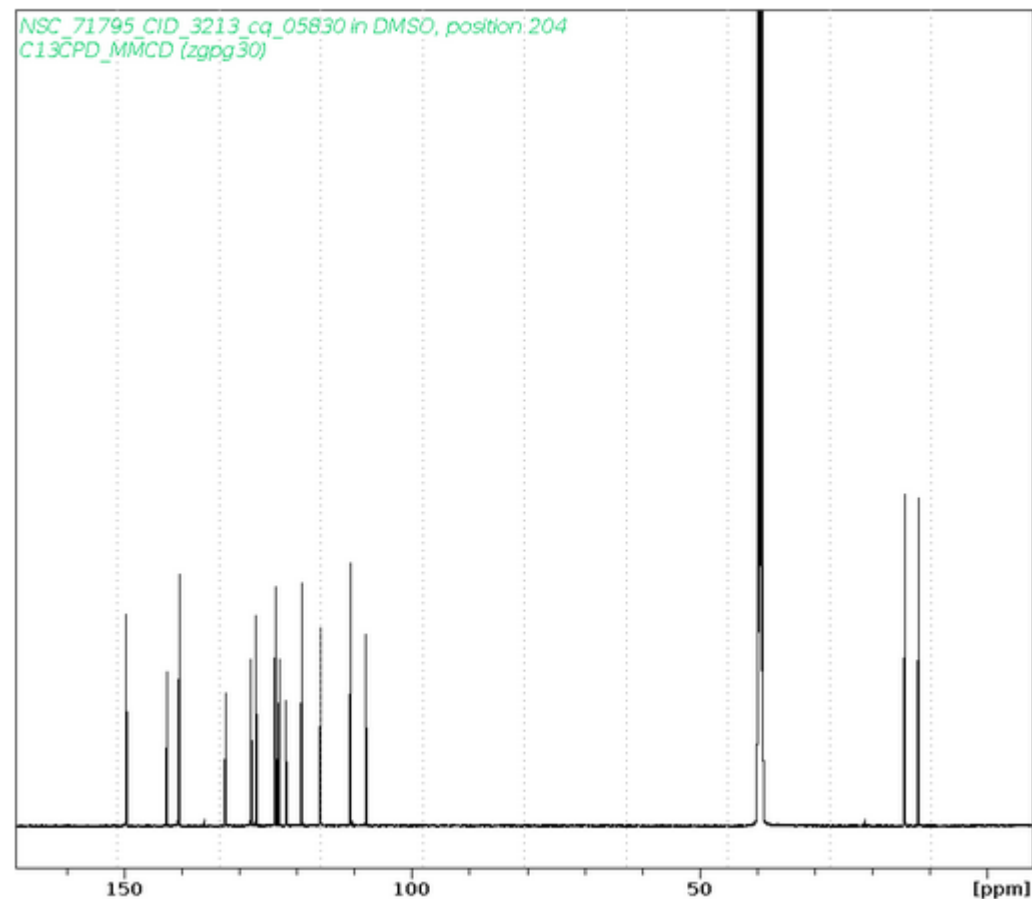
Sample: 5mg in DMSO, ref: TMS  
Conditions: temperature: 298K  
Spectrometer: Bruker Avance III - 500MHz

Atom ID	Value	Ambiguity Code
C1	14.3692	1
C2	11.9824	1
C3	119.191	4
C4	127.1373	4
C5	123.8352	4
C6	110.7067	4
C7	115.8997	1
C8	140.5188	1
C9	149.7257	1
C10	128.066	1
C11	132.4729	1
C12	108.0425	1
C13	123.1352	1
C14	121.9748	1
C15	142.6759	1
C16	123.3902	1
C17	140.5486	1

Valeurs expérimentales

## 3: 1D 13C

Sample: 5mg in DMSO, ref: TMS  
Conditions: temperature: 298K  
Spectrometer: Bruker Avance III - 500MHz



Download time domain data: [bmse001340\\_3.zip](#)

# ELLIPTICINE

## Cible

14.3692  
11.9824  
119.1910  
127.1373  
123.8352  
110.7067  
115.8997  
140.5188  
149.7257  
128.0660  
132.4729  
108.0425  
123.1352  
121.9748  
142.6759  
123.3902  
140.5486

## Ouverture du fichier acd\_lotusv9.NMRUDB

ACD/C-H NMR Predictors and DB: Database Window - [C:\USERS\JMN\DOCUMENTS\CNRS22\LOTUSV9\_ACD\ACD\_LOTUSV9.NMRUDB]

Database View Record Search Lists Table Training Options ACD/Labs Help

Atom No. 13C Shift 1H Shift

Atom No.	13C Shift	1H Shift
1	23.02	
2	28.17	
3	22.76	
4	39.74	
5	24.06	
6	36.46	
7	35.97	
8	18.86	
9	56.56	
10	28.36	
11	24.48	
12	56.91	
13	32.16	
14	32.12	
15	121.49	
16	141.08	
17	42.44	
18	71.59	
20	31.72	
21	37.55	
22	36.67	
23	19.46	
24	50.52	
25	21.29	
26	40.07	

Formula: C<sub>27</sub>H<sub>46</sub>O  
FW: 386.6535  
SMILES\_ID: Q43656\_2  
Predicted 13C shifts: 1[1] 23.02 ; 2[2] 28.17 ; 3[3] 22.76 ; 4[4] 39.74 ; 5[5] 24.06 ; 6[6] 36.46 ; 7[7] 35.97 ; 8[8] 18.86 ; 9[9] 56.56 ; 10[10] 28.36 ; 11[11] 24.48 ; 12[12] 56.91 ; 13[13] 32.16 ; 14[14] 32.12 ; 15[15] 121.49 ; 16[16] 141.08 ; 17[17] 42.44 ; 18[18] 71.59 ; 20[20] 31.72 ; 21[21] 37.55 ; 22[22] 36.67 ; 23[23] 19.46 ; 24[24] 50.52 ; 25[25] 21.29 ; 26[26] 40.07

ID: 1 A: 1/218478 B: 218478 Last Updated: 28/09/2022 16:27 Exclude from Calculations Single DB

1-ChemSketch 2-CNMR Spectrum 3-HNMR Spectrum 4-History 5-Database



# ELLIPTICINE

Fichier Édition Affichage Historique Marque-pages Outils Aide

Wikidata

https://www.wikidata.org/wiki/Wikidata:Main\_Page

English Not logged in Talk Contributions Create account Log in

Main Page Discussion Read View source View history

**ellipticine (Q10359556)**  
chemical compound  
Search for pages containing Q10359556

**WIKIDATA**

Main page  
Community portal  
Project chat  
Create a new Item  
Recent changes  
Random Item  
Query Service  
Nearby  
Help  
Donate

Lexicographical data  
Create a new Lexeme  
Recent changes  
Random Lexeme

Tools  
What links here  
Related changes  
Special pages  
Permanent link

open collaborative  
**Welcome to Wikidata**  
the free knowledge base with 102,507,809 data items that anyone can edit.  
Introduction • Project Chat • Community Portal • Help  
Want to help translate? [Translate the missing messages.](#)  
free structured  
multilingual

**Welcome!**  
Wikidata is a free and open knowledge base that can be read and edited by both humans and machines.  
Wikidata acts as central storage for the **structured data** of its Wikimedia sister

**Learn about data**  
New to the wonderful world of data? [Develop and improve your data literacy through content](#) designed to get you up to speed and feeling comfortable with the fundamentals in no time.



# ELLIPTICINE

Eichier Édition Affichage Historique Marque-pages Outils Aide

ellipticine - Wikidata

https://www.wikidata.org/wiki/Q10359556

Autres marque-pages

English Not logged in Talk Contributions Create account Log in

Item Discussion Read View history Search Wikidata

## ellipticine (Q10359556)

chemical compound [edit](#)

NSC-71795 | Elliptisine | NSC 71795 | ellipticines

[In more languages](#)

### Statements

instance of	<a href="#">chemical compound</a> <a href="#">edit</a>
	<a href="#">0 references</a> <a href="#">+ add reference</a>
	<a href="#">carbazole alkaloid</a> <a href="#">edit</a>
	<a href="#">1 reference</a>
	<a href="#">type of chemical entity</a> <a href="#">edit</a>
	<a href="#">0 references</a>

# ELLIPTICINE

found in taxon



Ochrosia elliptica

edit

▼ 5 references

stated in

The biogenetic, synthetic and biochemical aspects of ellipticine, an antitumor alkaloid

stated in

Fluorodensitometric assay of 6H-pyrido[4,3-b]-5,11-dimethylcarbazoles (ellipticine and derivatives) biosynthesized by Ochrosia elliptica cultures in vitro

stated in

Antitumor alkaloids in callus cultures of Ochrosia elliptica

stated in

Alkaloid Production by Ochrosia elliptica Cell Suspension Cultures

stated in

Alkaloids of Ochrosia elliptica Labill.1

+ add reference



Ochrosia moorei

edit

▶ 1 reference

# ELLIPTICINE



Item [Discussion](#)

## Alkaloids of *Ochrosia elliptica* Labill.1 (Q104937747)

scientific article published on 10 March 2005

### Identifiers

DOI

10.1021/JA01517A031 CR

edit

▼ 0 references

+ add reference



+ add value


+ add statement


# ELLIPTICINE

## Alkaloids of *Ochrosia elliptica* Labill.<sup>1</sup>

Sidney Goodwin, A. F. Smith, and E. C. Horning

 Cite this:  *J. Am. Chem. Soc.* 1959, 81, 8, 1903–1908

Publication Date: April 1, 1959 

<https://doi.org/10.1021/ja01517a031> 

© American Chemical Society

[RIGHTS & PERMISSIONS](#)  Subscribed

 ACS Legacy Archives

Article Views

513

Altmetric


3

Citations

202

[LEARN ABOUT THESE METRICS](#)



PDF (788 KB) 

open URL 

April 20, 1959

ALKALOIDS OF *Ochrosia elliptica* LABILL.

1903

[CONTRIBUTION FROM THE LABORATORY OF CHEMISTRY OF NATURAL PRODUCTS, NATIONAL HEART INSTITUTE, NATIONAL INSTITUTES OF HEALTH]

## Alkaloids of *Ochrosia elliptica* Labill.<sup>1</sup>

BY SIDNEY GOODWIN, A. F. SMITH<sup>2</sup> AND E. C. HORNING

RECEIVED JULY 28, 1958

Four alkaloids have been isolated from leaves of *Ochrosia elliptica* Labill. One has been shown to be identical with isoreserpiline. The other three, ellipticine, methoxyellipticine and elliptinine, have been characterized and certain features of their structures have been suggested.



# ELLIPTICINE – RECHERCHE CIBLÉE

## Cibler les alcaloïdes d'Apocynaceae

**Search Data**

Data Name(s):	Condition:	Value:
<input checked="" type="checkbox"/> BIO_FAMILY	Includes	Apocynaceae

Case Sensitive     Use Wildcards (\*, ?)

**Search Message**

5549 hits found for your query

**Search Data**

Data Name(s):	Condition:	Value:
<input checked="" type="checkbox"/> npclassifier_01pathway	Includes	Alkaloid

Case Sensitive     Use Wildcards (\*, ?)

**Search Message**

2376 hits found for your query

# ELLIPTICINE

Search by Chemical Shifts

Enter CNMR Query Shifts (example: 32.0 128.1..130)

14.3692  
11.9824  
119.1910  
127.1373  
122.555

CNMR Shifts Looseness Factor (+/-): 4.00

Minimum Number of CNMR Query Shifts to Match: 17

Enter HNMR Query Shifts (example: 3.2 6.1..6.4 10s[10])

HNMR Shifts Looseness Factor (+/-): 0.00

Minimum Number of HNMR Query Shifts to Match:

Search through Unambiguous Assignment

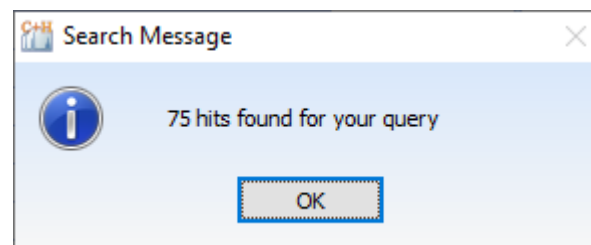
Do not Match One Chemical Shift for Several Shift Queries

Do not Sort Result

Sort Result by HQI Based on Minimal Distances

Sort Result by HQI Based on Shifts Distribution

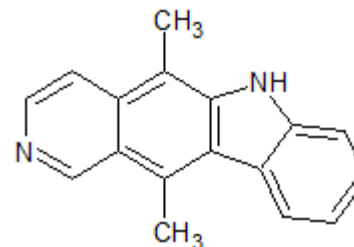
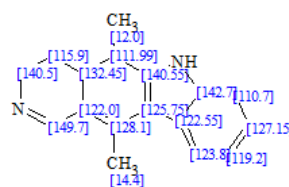
OK Cancel Help



ACD/C+H NMR Predictors and DB: Database Window - [C:\USERS\JMN\DOCUMENTS\CNRS22\LOTUSV9\_ACD\ACD\_LOTUSV9.NMRUDB]

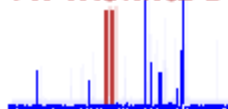
Database View Record Search Lists Table Training Options ACD/Labs Help

Atom No.	<sup>13</sup> C Shift	<sup>1</sup> H Shift
1	12.0	
2	111.99	
3	132.45	
4	115.9	
5	140.5	
7	149.7	
8	122.0	
9	128.1	
10	14.4	
11	125.75	
12	140.55	
14	142.7	
15	110.7	
16	127.15	
17	119.2	
18	123.8	
	122.55	



# EXEMPLE, NMRSHIFTDB2 – ELLIPTICINE

NMRShiftDB



**Current usage is:**

Registered Users: 2138

Structures: 258158

Spectra: Measured 53954, calculated 396566

## Impressum

[Home](#) | [Search](#) | [Results](#) | [Quick Check](#) | [Predict](#) | [Assignment](#)

Search by Spectrum

Switch to expert search mode

## Browse all structures

Input list ⓘ: [Input format](#)

14.3692  
11.9824  
119.1910  
127.1373  
123.8352  
110.7067  
115.8997  
140.5188  
149.7257  
128.0660  
132.4729

Spectrum type to search:

13C ▾

Subspectrum

Complete

Search by spectrum

# ELLIPTICINE

total similarity 13C  
 spectrum search -- 14.3692|11.9824|119....

[Spectral Data](#) [Additional Data](#) [Download](#)

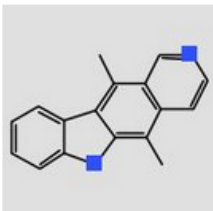
Results: 300

Browse: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) >>

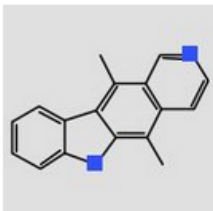
Search for complete spectrum: Similarity measure for the complete spectrum in this record is 92.92.

Type: 13C

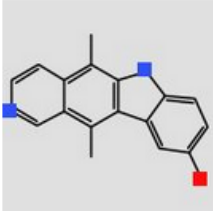
## Next structure



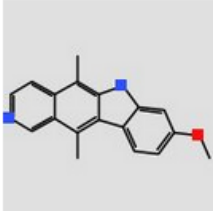
Similarity: 90.15 %



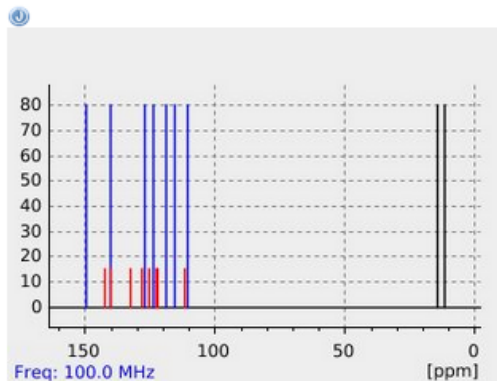
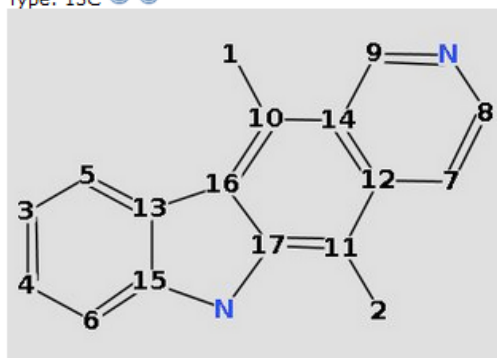
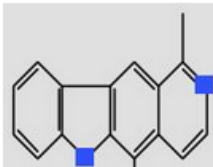
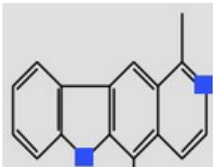
Similarity: 82.55 %



Similarity: 71.02 %



Similarity: 69.05 %



Atom	Mult.(coupling const.)	Prediction Shift	Input Shift	Diff. M-I	Prediction 0 hose_lotusv7
1	Q	14.40	14.3692	0.03	14.32
2	Q	12.00	11.9824	0.02	11.6
3	D	119.20	119.191	0.01	120.19
4	D	127.15	127.1373	0.01	125.25
5	D	123.80	123.3902	0.41	123.06
6	D	110.70	108.0425	2.66	111.2
7	D	115.90	115.8997	0.00	115.7
8	D	140.50	140.5188	0.02	142.2
9	D	149.70	149.7257	0.03	150.2
10	S	128.10	128.066	0.03	128.34
11	S	111.99	110.7067	1.28	111.0
12	S	132.45	132.4729	0.02	130.38
13	S	122.55	123.1352	0.59	121.95
14	S	122.00	121.9748	0.03	125.6
15	S	142.70	142.6759	0.02	140.95
16	S	125.75	123.8352	1.91	124.83
17	S	140.55	140.5486	0.00	136.36

Threshold is 5.88





## MESSAGE À RETENIR

- La base de données acd\_lotusv9 au format SDF est disponible sur <https://zenodo.org/record/7124055>
- Le fichier acd\_lotusv9.sdf peut être importé comme base de données dans les logiciels de la société ACD/Labs
- Les structures et les déplacements chimiques présents dans acd\_lotusv9 sont inclus dans <https://nmrshiftdb.nmr.uni-koeln.de/>
- La validation du processus d'identification des composés connus a été menée sur un ensemble de 58 métabolites spécialisés et le résultat sera prochainement soumis pour publication.

# L'équipe "Substances Naturelles" de l'ICMR

