

Déréplication (PARTIE II)

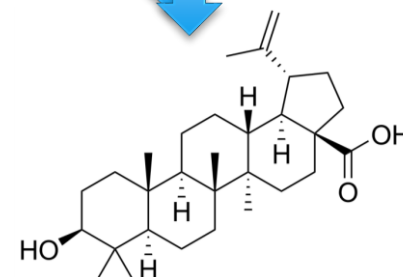
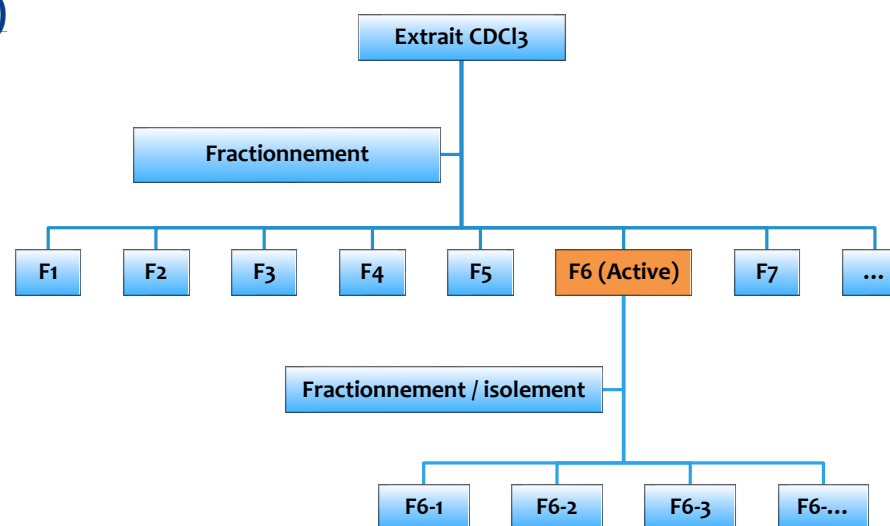
Déréplication et Bases de Données

Samuel Bertrand

Alphitonia zizyphoides (feuilles)

Activité Antifongique lors d'un screening (*C. albicans*)

Stratégie par micro-fractionnement

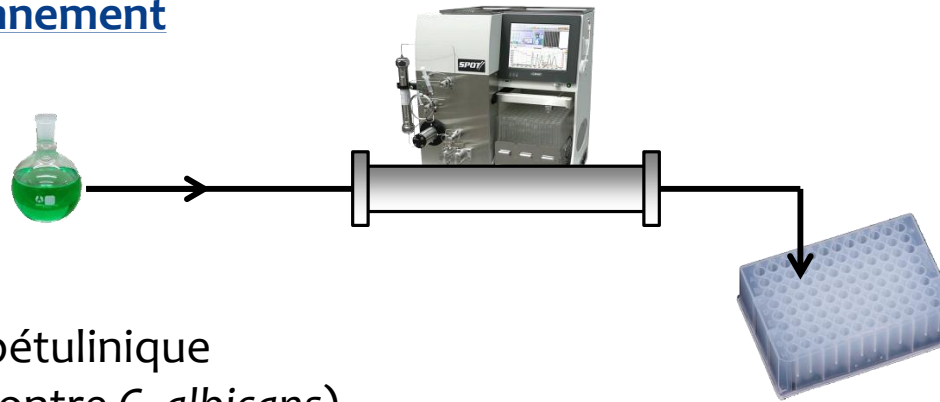


Acide bétulinique
(actif contre *C. albicans*)

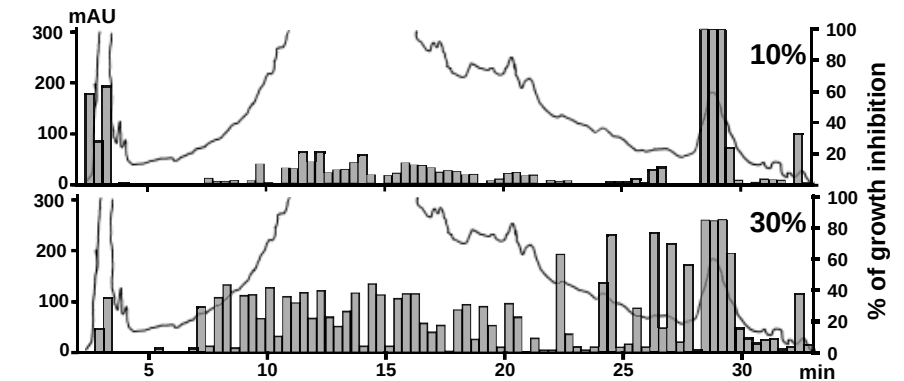
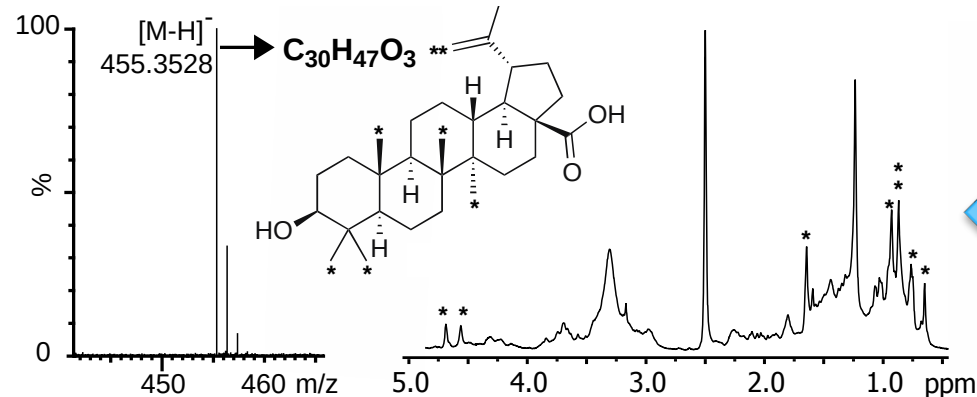
Alphitonia zizyphoides (feuilles)

Activité Antifongique lors d'un screening (*C. albicans*)

Stratégie par micro-fractionnement



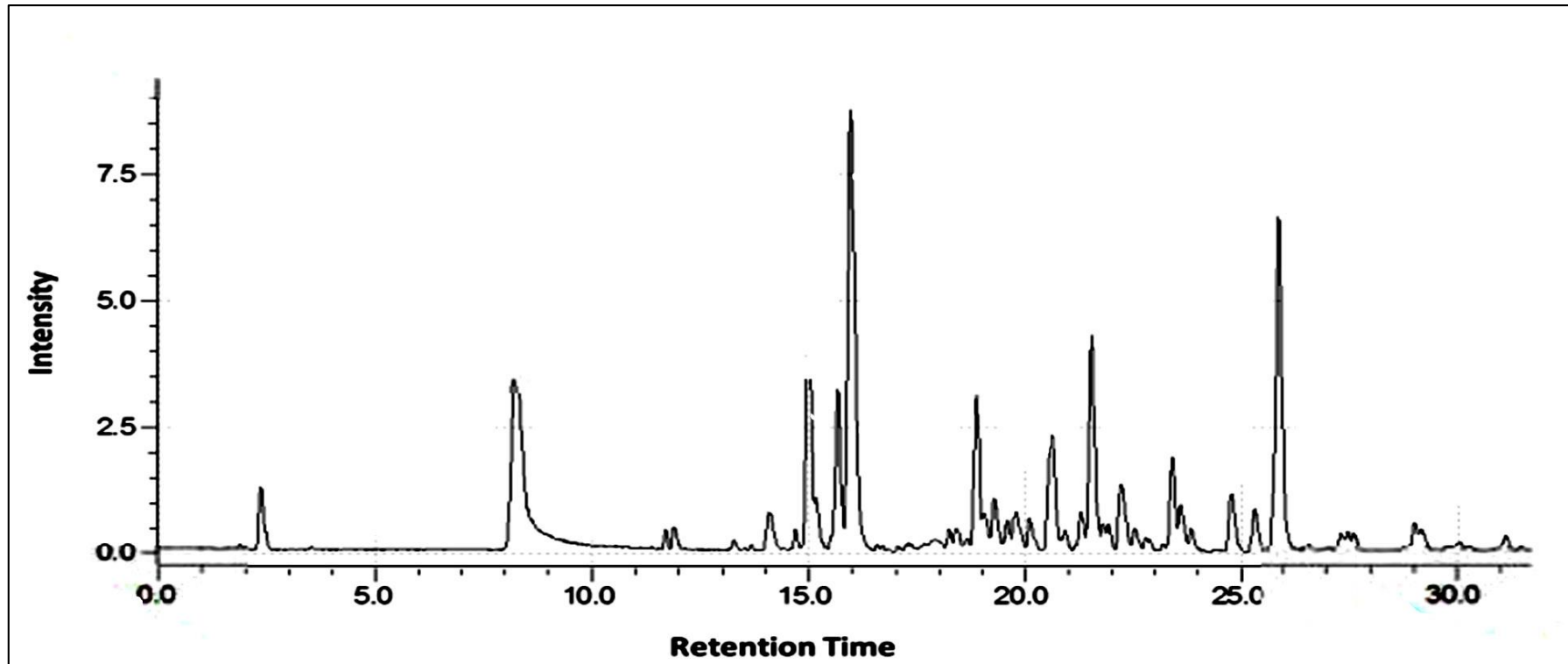
Acide bétulinique
(actif contre *C. albicans*)



Analyse
des puits
(LC-MS & RMN)

Identification de composés dans un chromatogramme

Identification de chaque pic.





CSI: CRIME SCENE INVESTIGATION

Anal Bioanal Chem
DOI 10.1007/s00216-013-7582-x

RESEARCH PAPER

Aggressive dereplication using UHPLC–DAD–QTOF screening extracts for up to 3000 fungal secondary m

Andreas Klitgaard · Anita Iversen · Mikael R. Andersen ·
Thomas O. Larsen · Jens Christian Frisvad ·
Kristian Fog Nielsen

Received: 11 September 2013 / Revised: 3 December 2013 / Accepted: 14 December 2013
© The Author(s) 2014. This article is published with open access at Springerlink.com

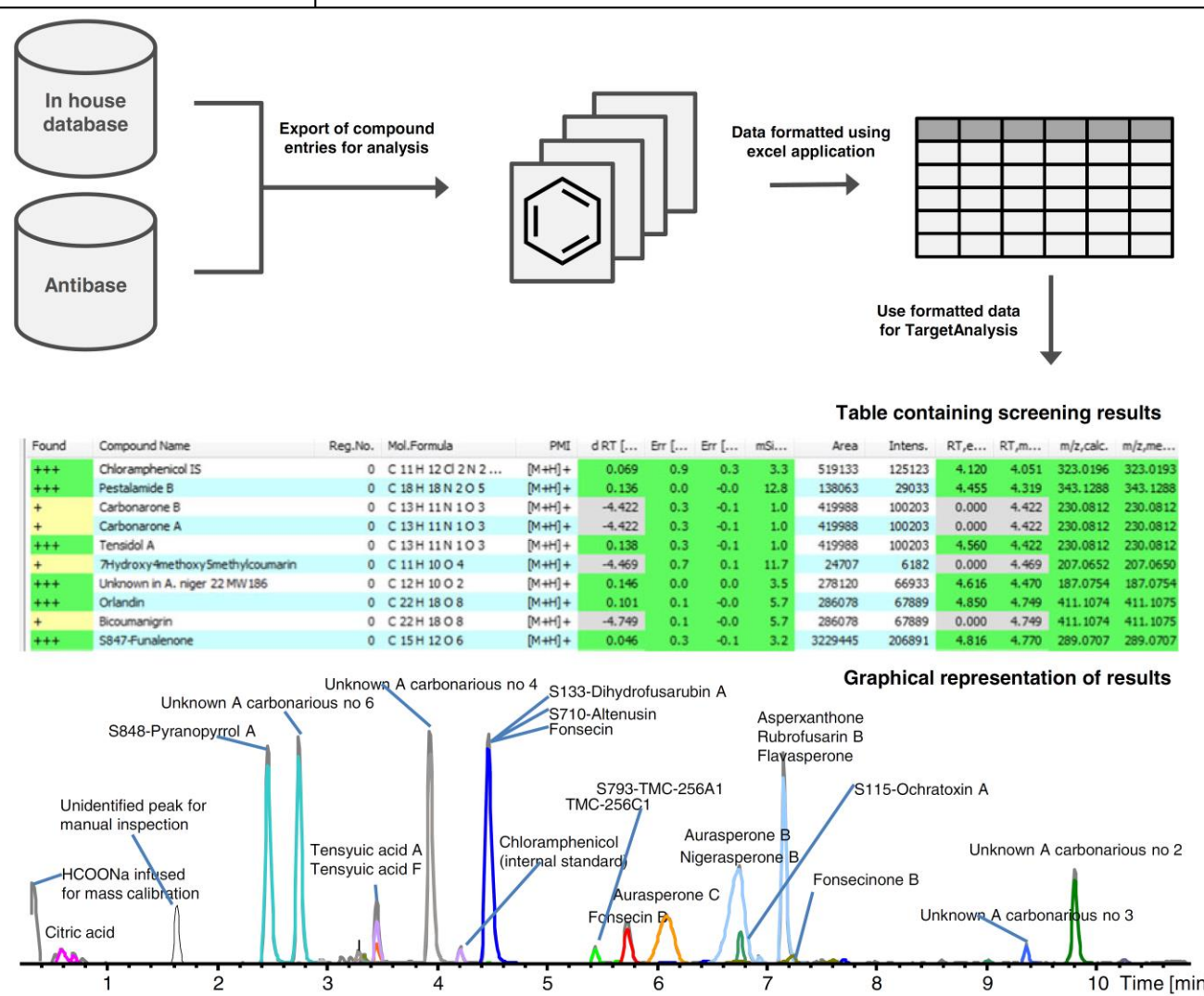
Abstract In natural-product drug discovery, finding new compounds is the main task, and thus fast dereplication of known compounds is essential. This is usually performed by manual liquid chromatography–ultraviolet (LC–UV) or visible light-mass spectroscopy (Vis–MS) interpretation of detected peaks, often assisted by automated identification of previously identified compounds. We used a 15 min high-performance liquid chromatography–diode array detection (UHPLC–DAD)–high-resolution MS method (electrospray ionization (ESI)⁺ or ESI[−]), followed by 10–60 s of automated data analysis for up to 3000 relevant elemental compositions. By overlaying automatically generated extracted-ion chromatograms

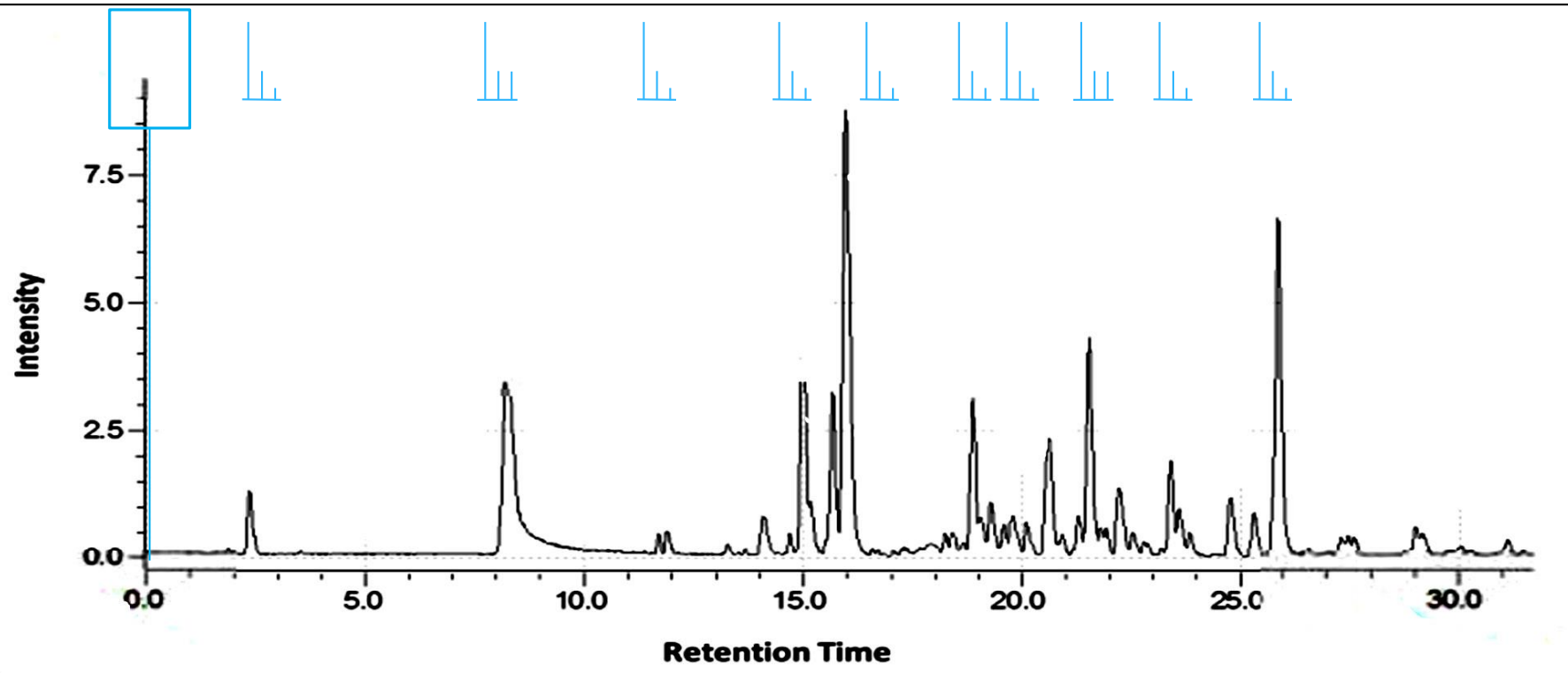
was investigated on reference standard was used on extracts of *Aspergillus melanoconidium*, revealing biomarkers for both species.

Keywords Metabolomics · Mycology · UPLC · Polyketide · Nonribosomal

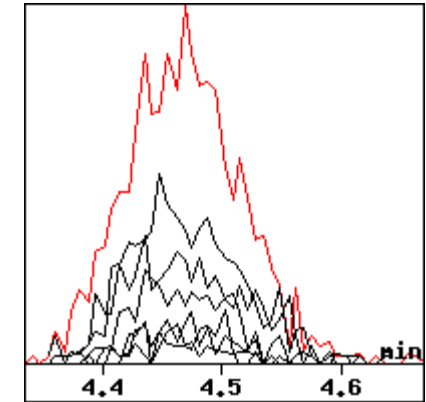
Introduction

Fungi are an immense source of diverse natural products that

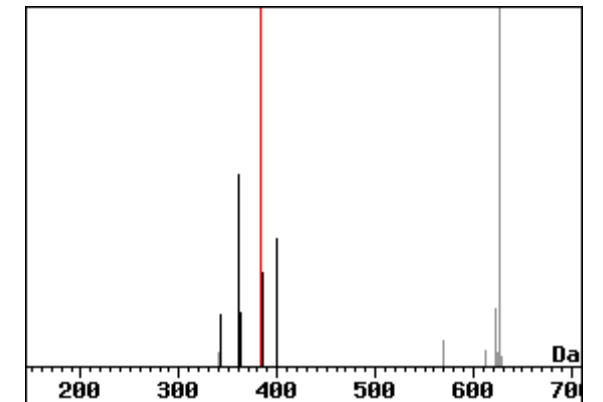




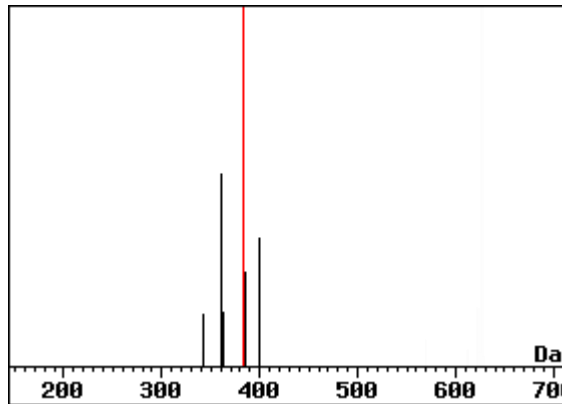
Regroupement des ions



Pseudo-spectre



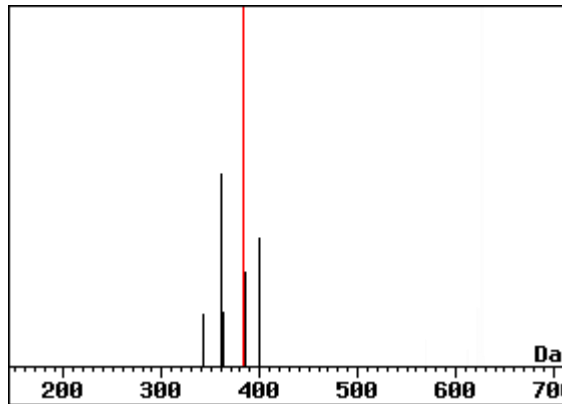
Quel composé ?



Quelle est la structure du composé ?

C'est pas si simple !!

Quel composé ?

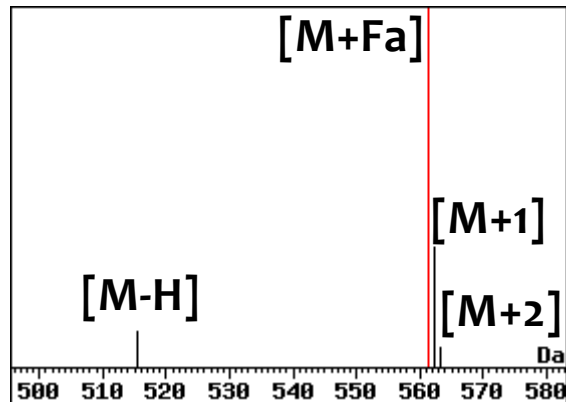


Quelle est la structure du composé ?

Utilisation des données de mass haute résolution

Stratégie générale

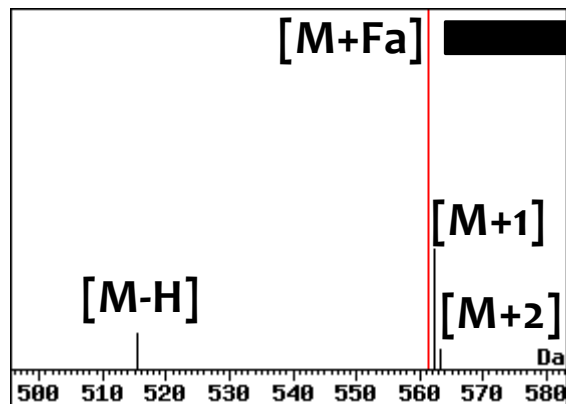
Interprétation du
spectre



Stratégie générale

Interprétation du
spectre

Détermination de
la formule brute

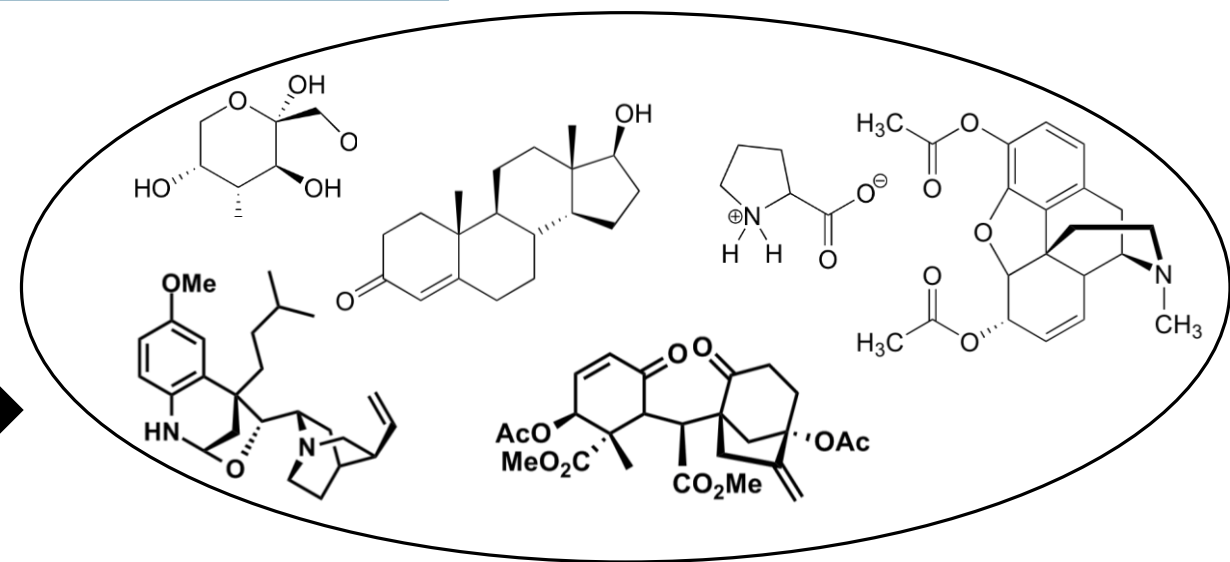
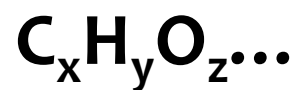
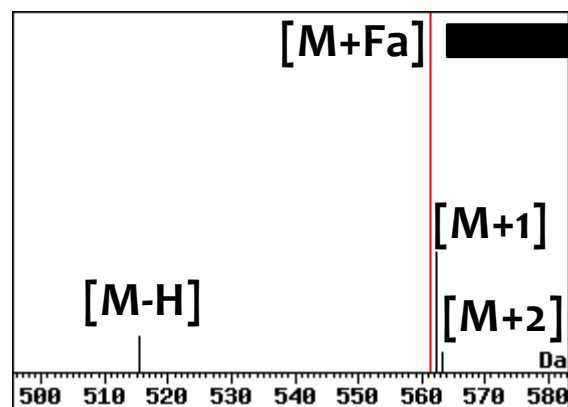


Stratégie générale

Interprétation du spectre

Détermination de la formule brute

Recherche dans les bases de données



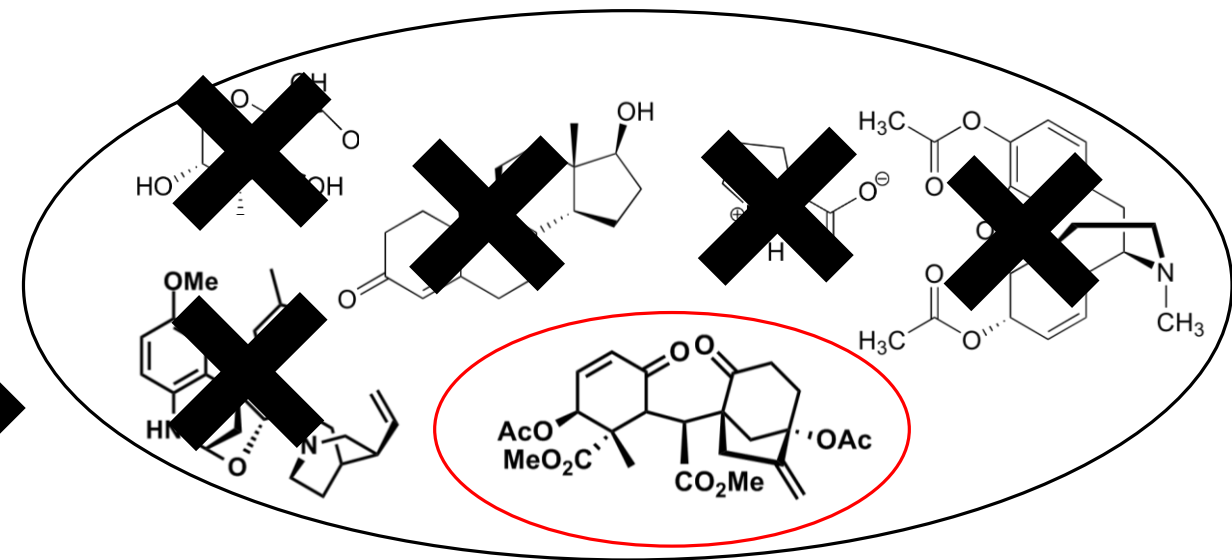
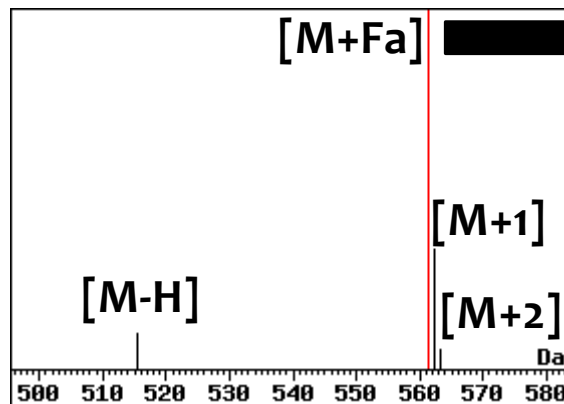
Stratégie générale

Interprétation du spectre

Détermination de la formule brute

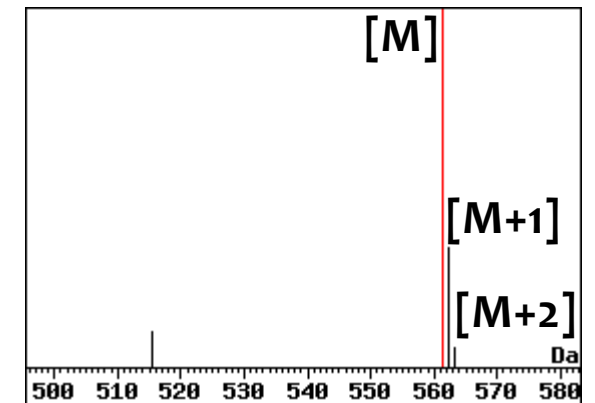
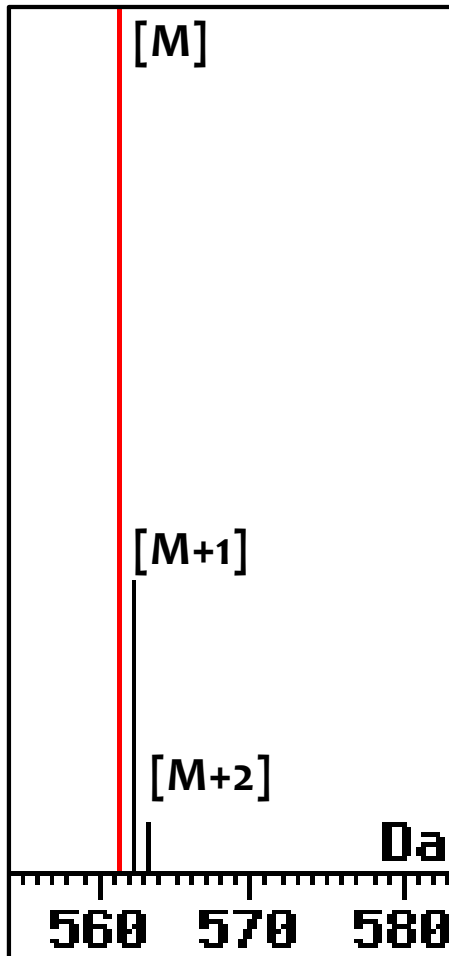
Recherche dans les bases de données

Confirmation de l'identification



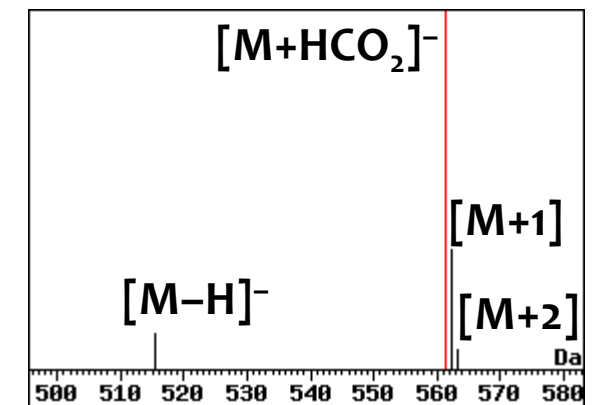
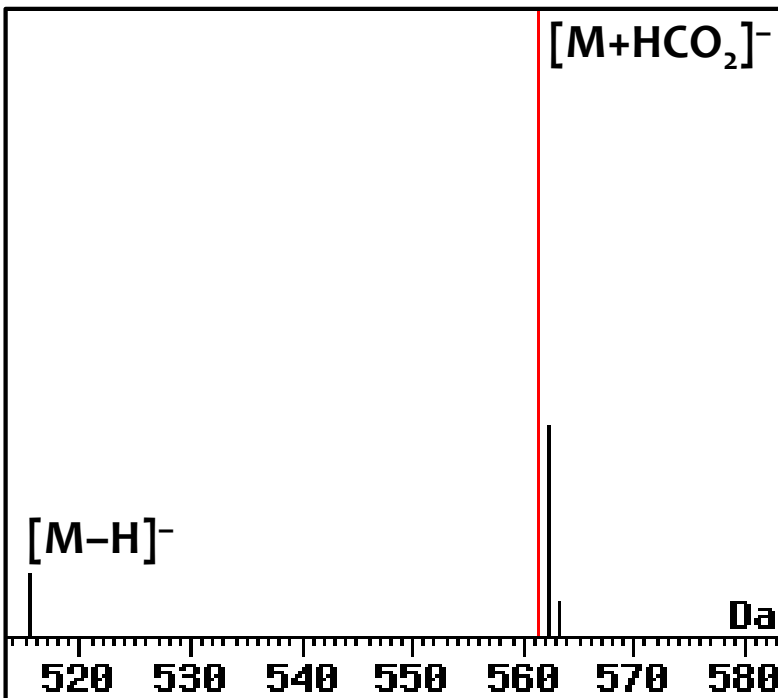
Interprétation de spectres MS¹

1- recherche des isotopes



Interprétation de spectres MS¹

- 1- recherche des isotopes
- 2- recherche des adduits



Interprétation de spectres MS¹

Table 2. Common Mass “Jumps” (Δ) Observed in Electrospray MS and the Frequency of Adducts in 719 Reference Standards

positive electrospray	relative to [M + H] ⁺ Δ (amu) ^a	most intense (%)	observed (%)
[M + H] ⁺		60.2	88.7
[M + Na] ⁺ , very stable adduct ^b	21.9820	0.8	21.0
[M + NH ₄] ⁺ ^c	17.0265	8.6	20.0
[M + H + MeCN] ⁺ , neutral adduct ^d	41.0265	7.8	26.7
[M + Na + MeCN] ⁺ , very stable adduct ^{b,d}	63.0085	0.4	20.2
[M - H + Na + HCOO] ⁺ ^g			
[M + K] ⁺ , very stable adduct ^b	37.9559	ND	1.0
[M + K + MeCN] ⁺ , very stable adduct ^{b,d}	78.9824	ND	ND
[M - H + 2Na] ⁺ (acids, phenols, and enoles)	43.9640	ND	0.2
[M - H + Fe] ⁺ , isotope <i>m/z</i> 2 lower ca. 4% ^e	53.9193	ND	0.2
Exchanging adducts			
[M + Na] ⁺ to [M + NH ₄] ⁺	4.9555	8.9 ^f	
[M + Na] ⁺ to [M + K] ⁺	15.9739	ND	
[M - H ₂ O + H] ⁺ to [M + Na] ⁺	39.9926	6.4	
[M - H ₂ O + H] ⁺ to [M + MeCN + Na] ⁺ ^d	81.0191	7.1 ^d	
[M - H ₂ O + H] ⁺ to [M + NH ₄] ⁺	35.0371	7.5 ^d	
[M - 2H ₂ O + H] ⁺ to [M + Na] ⁺	58.0032	1.4	
[M + NH ₄] ⁺ to [M + H - CH ₃ COOH] ⁺ ^d	77.0476	2.9 ^d	
[M + Na+MeCN] ⁺ to [M + H - CH ₃ COOH] ⁺ ^d	123.0296	2.6 ^d	
Neutral losses			
H ₂ O	18.0106	4.9	33.0
2H ₂ O	36.0212	0.1	8.3
CO ₂	43.9898	ND	4.9
CH ₃ CHO	44.0262	1.0	1.8

positive electrospray	relative to [M + H] ⁺ Δ (amu) ^a	most in- tense (%)	observed (%)
Neutral losses			
HCOOH	46.0055	1.0	5.3
CH ₃ COOH	60.0211	4.0	6.8
NH ₃ ^c	17.0265	ND	1.5
negative electrospray	relative to [M - H] ⁻ Δ (amu) ^a	most intense (%)	observed (%)
[M - H] ⁻		56.7	69.4
[M + HCOO] ^{-g}	46.0055	17.2	36.4
[M + CH ₃ COO] ^{-h}	60.0211	ND ^h	ND ^h
[M - 2H + Na] ⁻ (acids, phenols, and enoles)	21.9820	ND	4.5
[M + Cl] ⁻ (A + 2 isotope)	35.9767	ND	13.6
[M - H + H ₂ O] ⁻ , e.g., opening of lactones	18.0106	ND	2.6
[M - H + HCOONa] ⁻	67.9874	0.10	9.9
[M - H + CH ₃ COONa] ^{-h}	82.0031	ND ^h	ND ^h
Exchanging adducts			
[M + HCOO] ⁻ to [M + Cl] ^{-g}	10.0288	12.2 ^g	
H ₂ O	18.0106	ND	2.6
CO ₂	43.9898	0.3	13.6
CH ₃	15.0235	ND	2.6
HCOOH	46.0055	ND	0.6
CH ₃ COOH	60.0211	ND	2.2

^aMass shift. ^bIncreased by high in-source fragmentation settings. ^cLoss or addition of NH₃ can usually be determined looking for [M + Na]⁺. ^dIf MeCN is used as solvent. MeCN forms NH₃ upon acidic hydrolysis. ^eSiderophores and artifact from Fe²⁺ liberation in ESI⁺. ^fOften seen in peptides >1000 Da. ^gIf HCOOH is added as buffer. ^hNot observed since acetate was not added.

Metabolomics Fiehn Laboratory UC Davis, Genome Center

Tobias Kind

Table 1. Monoisotopic exact masses of molecular ion adducts often observed in ESI mass spectra

Ion name	Ion mass	Charge	Mult	Mass	Your M here: Result:	Your M+X or M-X Reverse:
1. Positive ion mode					661.966177	804.38365
M+3H	M/3 + 1.007276	3+	0,33	1,007276	221,662668	267,113941
M+2H+Na	M/3 + 8.334590	3+	0,33	8,334590	228,989982	259,788627
M+H+2Na	M/3 + 15.7661904	3+	0,33	15,766190	236,421583	252,355026
M+3Na	M/3 + 22.989218	3+	0,33	22,989218	243,644610	245,131999
M+2H	M/2 + 1.007276	2+	0,5	1,007276	331,990365	401,174549
M+H+NH ₄	M/2 + 9.520550	2+	0,5	9,520550	340,503639	392,661275
M+H+Na	M/2 + 11.998247	2+	0,5	11,998247	342,981336	390,183578
M+H+K	M/2 + 19.985217	2+	0,5	19,985217	350,968306	382,196608
M+ACN+2H	M/2 + 21.520550	2+	0,5	21,520550	352,503639	380,661275
M+2Na	M/2 + 22.989218	2+	0,5	22,989218	353,972307	379,192607
M+2ACN+2H	M/2 + 42.033823	2+	0,5	42,033823	373,016912	360,148002
M+3ACN+2H	M/2 + 62.547097	2+	0,5	62,547097	393,530186	339,634728
M+H	M + 1.007276	1+	1	1,007276	662,973453	803,356374
M+NH ₄	M + 18.033823	1+	1	18,033823	680,000000	786,329827
M+Na	M + 22.989218	1+	1	22,989218	684,955395	781,374432
M+CH ₃ OH+H	M + 33.033489	1+	1	33,033489	694,999666	771,330161
M+K	M + 38.963158	1+	1	38,963158	700,929335	765,400492
M+ACN+H	M + 42.033823	1+	1	42,033823	704,000000	762,329827
M+2Na+H	M + 44.971160	1+	1	44,971160	706,937337	759,392490
M+Ac	M + 59.0133	1+	1	59,013300	720,979477	745,350350
M+IsoProp+H	M + 61.06534	1+	1	61,065340	723,031517	743,298310
M+ACN+Na	M + 64.015765	1+	1	64,015765	725,981942	740,347885
M+2K+H	M + 76.919040	1+	1	76,919040	738,885217	727,444610
M+DMSO+H	M + 79.02122	1+	1	79,021220	740,987397	725,342430
M+2ACN+H	M + 83.060370	1+	1	83,060370	745,026547	721,303280
M+IsoProp+Na+H	M + 84.05511	1+	1	84,055110	746,021287	720,308540
2M+H	2M + 1.007276	1+	2	1,007276	1324,939630	1607,720024
2M+NH ₄	2M + 18.033823	1+	2	18,033823	1341,966177	1590,693477
2M+Na	2M + 22.989218	1+	2	22,989218	1346,921572	1585,738082
2M+3H ₂ O+2H	2M + 28.02312	2+	2	28,023120	1351,955474	1580,704180
2M+K	2M + 38.963158	1+	2	38,963158	1362,895512	1569,764142
2M+ACN+H	2M + 42.033823	1+	2	42,033823	1365,966177	1566,693477
2M+ACN+Na	2M + 64.015765	1+	2	64,015765	1387,948119	1544,711535

Source:
Huang N.; Siegel M.M.1; Kruppa G.H.; Laukien F.H.
J Am Soc Mass Spectrom 1999, 10, 1166–1173
Automation of a Fourier transform ion cyclotron resonance mass spectrometer for acquisition, analysis, and e-mailing of high-resolution exact-mass electrospray ionization mass spectral data

[http://dx.doi.org/10.1016/S1044-0305\(99\)00089-6](http://dx.doi.org/10.1016/S1044-0305(99)00089-6)

Use Molecular Weight Calculator
<http://www.alchemistmatt.com/mwtwin.html>
<http://ncrr.pnl.gov/software/>

(actually this task should be done by every good LC-MS software automatically, directly integrated, no questions asked)

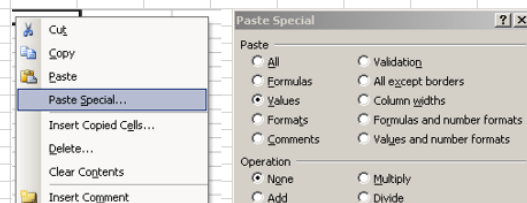
Example:
1) Find Adduct
Taxol, C₄₇H₅₁NO₁₄, M=853.33089
Enter 853.33089 in green box read M+22.9, m/z=876.320108

2) Reverse take 12 Tesla-FT-MS result out of MS m/z=876.330 suspect M+Na adduct, read M=853.340782, enter this value into formula finder with 2 ppm mass accuracy (CHNSOP enabled) get some thousand results, compare isotopic pattern

ppm Calculator
Experimental M here 663.461170
Calculated M here 663.4542
Result [ppm] 10.505513

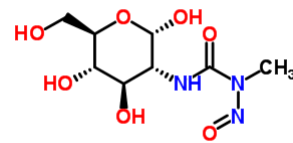
ppm Calculator Example (copy/paste value)
Experimental M here 787.42502
Calculated M here 787.42443
Result [ppm] 0.74927769

Copy the values by Copy/Paste Special + Value

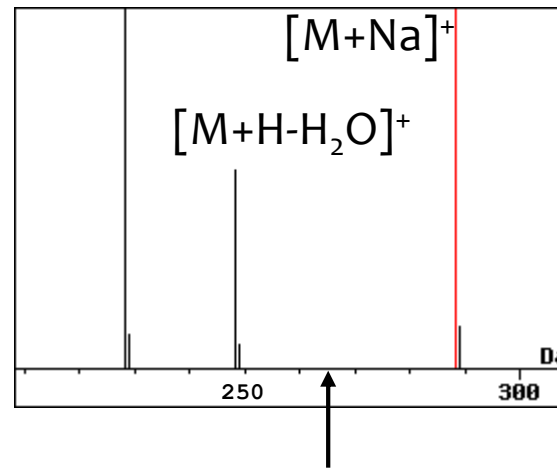


Interprétation de spectres MS¹

C'est pas si simple !!



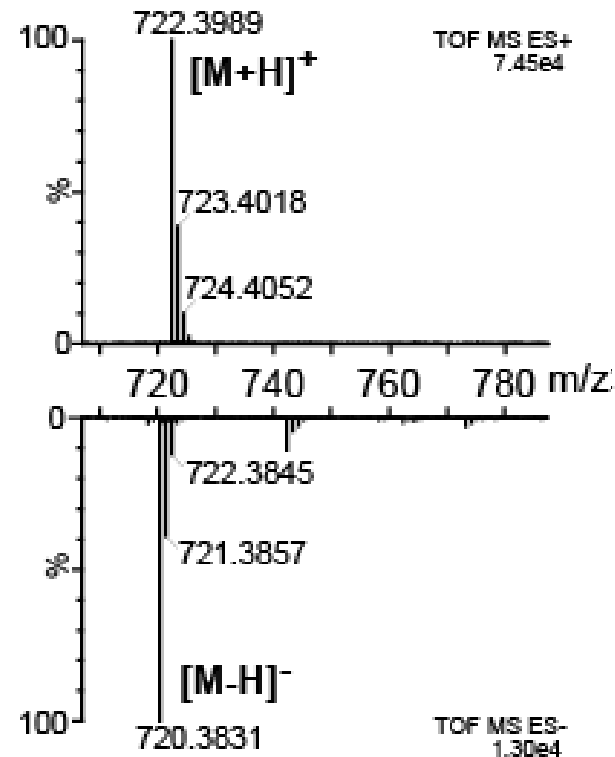
Streptozocin



Pas de $[M+H]^+$!

Interprétation de spectres MS¹

Combinaison ionisation négative et positive

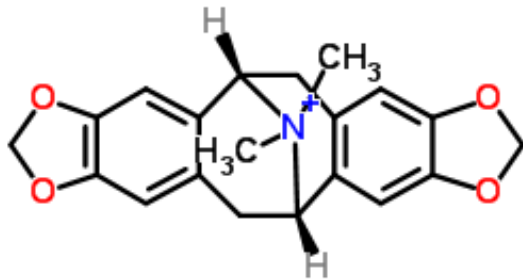


Aide à l'identification de $[M]$

Interprétation de spectres MS¹

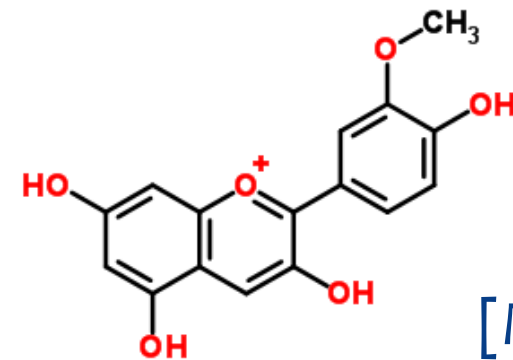
C'est pas si simple !!

Californidine



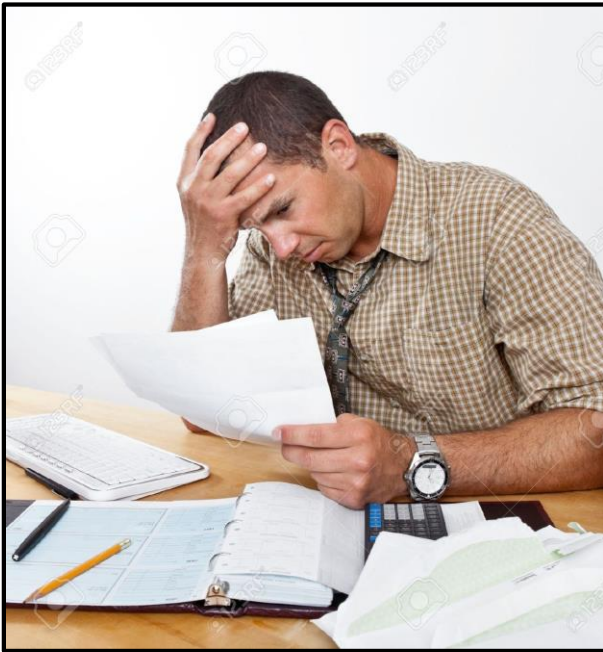
[M]⁺

peonidin



[M]⁺

Interprétation de spectres MS¹



CAMERA: An Integrated Strategy for Compound Spectra Extraction and Annotation of Liquid Chromatography/Mass Spectrometry Data Sets

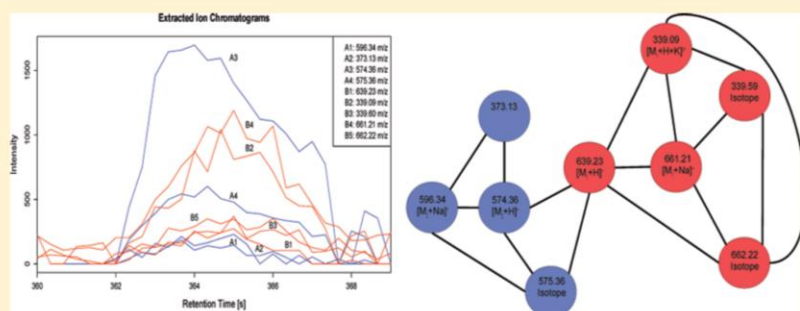
Carsten Kuhl,^{*,†} Ralf Tautenhahn,[‡] Christoph Böttcher,[†] Tony R. Larson,[§] and Steffen Neumann^{*,†}

[†]Department of Stress and Developmental Biology, Leibniz Institute of Plant Biochemistry, Weinberg 3, 06120 Halle (Saale), Germany

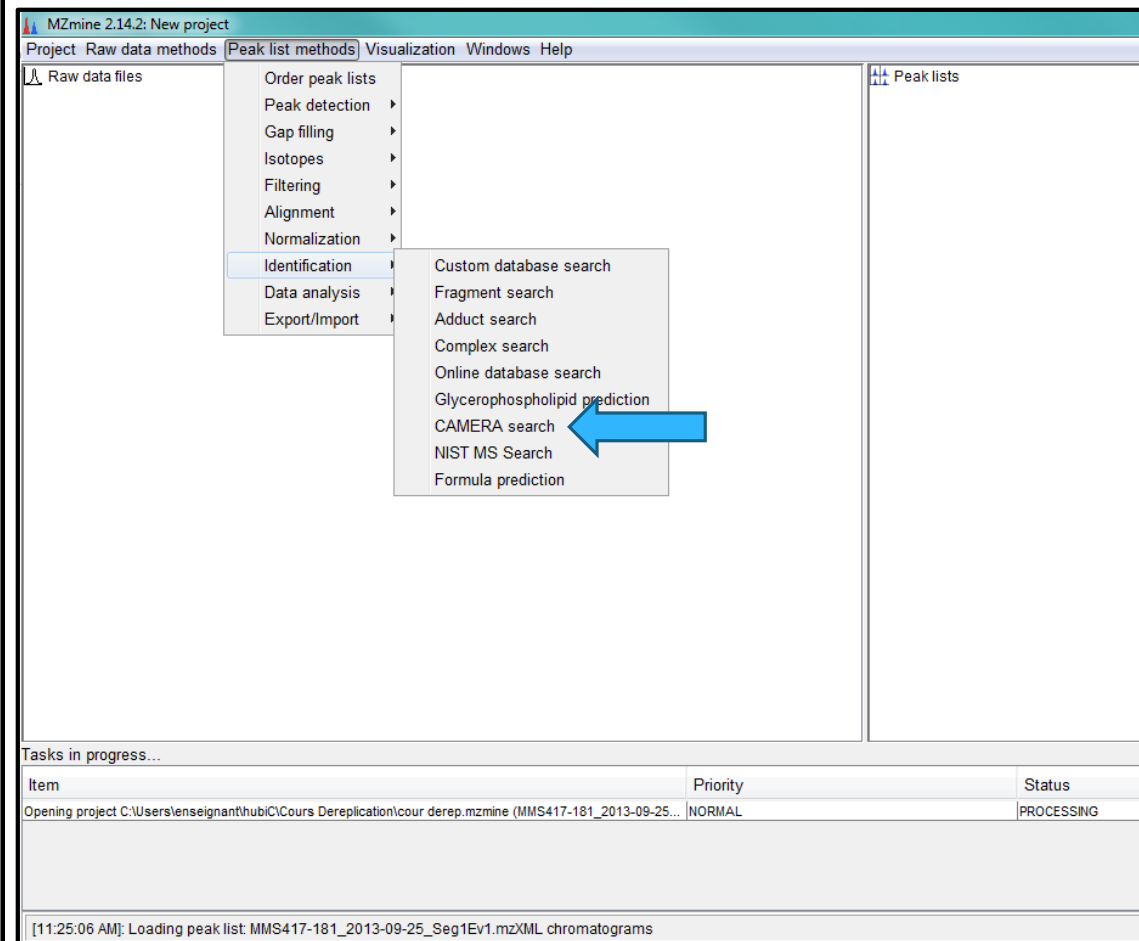
[‡]Department of Chemistry and Molecular Biology, Center for Metabolomics, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, United States

[§]Centre for Novel Agricultural Products, Department of Biology, University of York, United Kingdom

Supporting Information



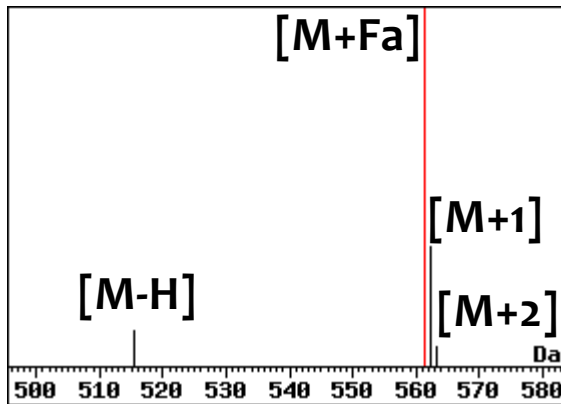
ABSTRACT: Liquid chromatography coupled to mass spectrometry is routinely used for metabolomics experiments. In contrast to the fairly routine and automated data acquisition steps, subsequent compound annotation and identification require extensive manual analysis and thus form a major bottleneck in data interpretation. Here we present CAMERA, a Bioconductor package integrating algorithms to extract compound spectra, annotate isotope and adduct peaks, and propose the accurate compound mass even in highly complex data. To evaluate the algorithms, we compared the annotation of CAMERA against a manually defined annotation for a mixture of known compounds spiked into a complex matrix at different concentrations. CAMERA



Item	Priority	Status
Opening project C:\Users\lenseignant\hubi\Cours Dereplication\cour derep.mzmine (MMS417-181_2013-09-25...	NORMAL	PROCESSING

[11:25:06 AM]: Loading peak list: MMS417-181_2013-09-25_Seg1Ev1.mzXML chromatograms

Détermination de la formule brute



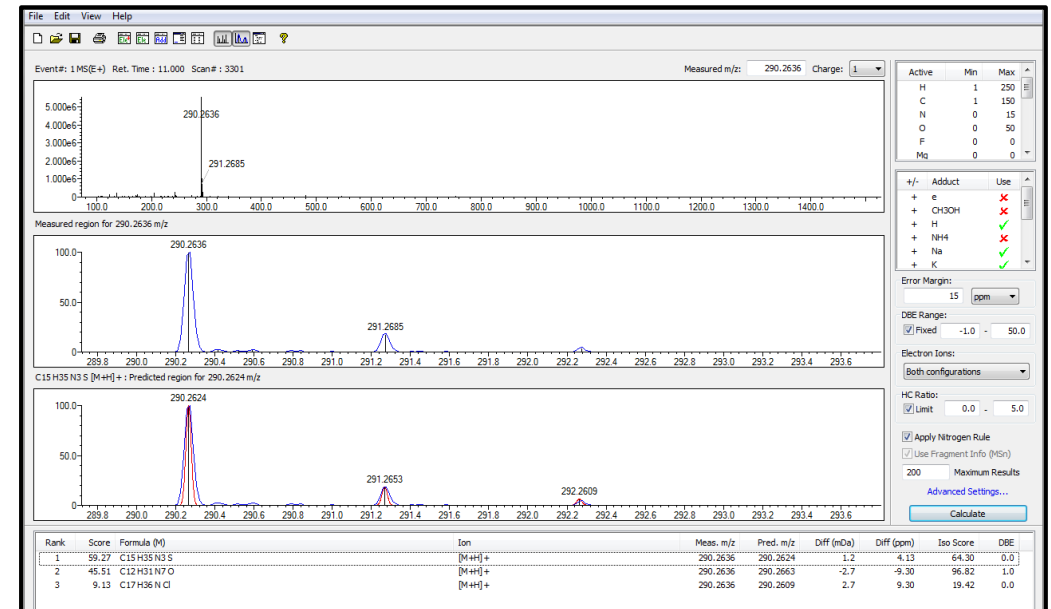
À partir de la masse exacte !

Seul une combinaison limitée peut expliquer m/z

Détermination de la formule brute

Quel logiciel utiliser :

- logiciels constructeurs
- Molgen-MS
- Rdisop
- MZmine
- **Sirius²**



Détermination de la formule brute

Comment ça fonctionne ?

1- définition de l'ensemble formules possible dans la limite définie par l'utilisateur (CHONPS...)

Table 1: Restrictions for number of elements during formula generation for small molecules based on examination of the DNP and Wiley mass spectral databases. For each element, the higher count was taken for denominating the element restriction rule #1

Mass Range [Da]	Library	C max	H max	N max	O max	P max	S max	F max	Cl max	Br max	Si max
< 500	DNP	29	72	10	18	4	7	15	8	5	
	Wiley	39	72	20	20	9	10	16	10	4	8
< 1000	DNP	66	126	25	27	6	8	16	11	8	
	Wiley	78	126	20	27	9	14	34	12	8	14
< 2000	DNP	115	236	32	63	6	8	16	11	8	
	Wiley	156	180	20	40	9	14	48	12	10	15
< 3000	DNP	162	208	48	78	6	9	16	11	8	

Détermination de la formule brute

Comment ça fonctionne ?

- 1- définition de l'ensemble formules possible dans la limite définie par l'utilisateur (CHONPS...)
- 2- calcule de l'erreur sur m/z (en ppm)
- 3- sélection des formules avec une erreur inférieure à la limite de l'appareil

Type	Mass Accuracy
FT-ICR-MS	0.1 - 1 ppm
Orbitrap	0.5 - 1 ppm
Magnetic Sector	1 - 2 ppm
TOF-MS	3 - 5 ppm
Q-TOF	3 - 5 ppm
Triple Quad	3 - 5 ppm
Linear IonTrap	50-200 ppm

Détermination de la formule brute

Comment ça fonctionne ?

- 1- définition de l'ensemble formules possible dans la limite définie par l'utilisateur (CHONPS...)
- 2- calcul de l'erreur sur m/z (en ppm)
- 3- sélection des formules avec une erreur inférieure à la limite de l'appareil

Table. Number of possible molecular formula

molecular mass [Da]	10 ppm	5 ppm	3 ppm	1 ppm	0.1 ppm
150	2	1	1	1	1
200	3	2	2	1	1
300	24	11	7	2	1
400	78	37	23	7	1
500	266	115	64	21	2
600	505	257	155	50	5
700	1046	538	321	108	10
800	1964	973	599	200	20
900	3447	1712	1045	345	32

Détermination de la formule brute

Comment ça fonctionne ?

- 1- définition de l'ensemble formules possible dans la limite définie par l'utilisateur
- 2- calcul de l'erreur sur m/z (en ppm)
- 3- sélection des formules avec une erreur inférieure à la limite de l'appareil
- 4- utilisation de la ressemblance des ratios isotopiques

BMC Bioinformatics



Open Access

Research article

Metabolomic database annotations via query of elemental compositions: Mass accuracy is insufficient even at less than 1 ppm
Tobias Kind* and Oliver Fiehn

Address: University of California Davis, Genome Center, 451 E. Health Sci Dr., Davis, CA 95616, USA

Email: Tobias Kind* - tkind@ucdavis.edu; Oliver Fiehn - ofiehn@ucdavis.edu

* Corresponding author

Published: 28 April 2006

BMC Bioinformatics 2006, 7:234 doi:10.1186/1471-2105-7-234

Received: 22 December 2005

Accepted: 28 April 2006

This article is available from: <http://www.biomedcentral.com/1471-2105/7/234>

© 2006 Kind and Fiehn; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

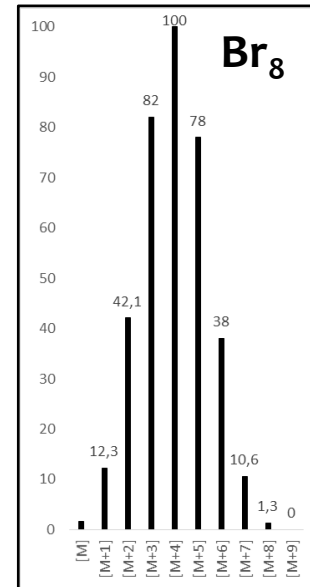
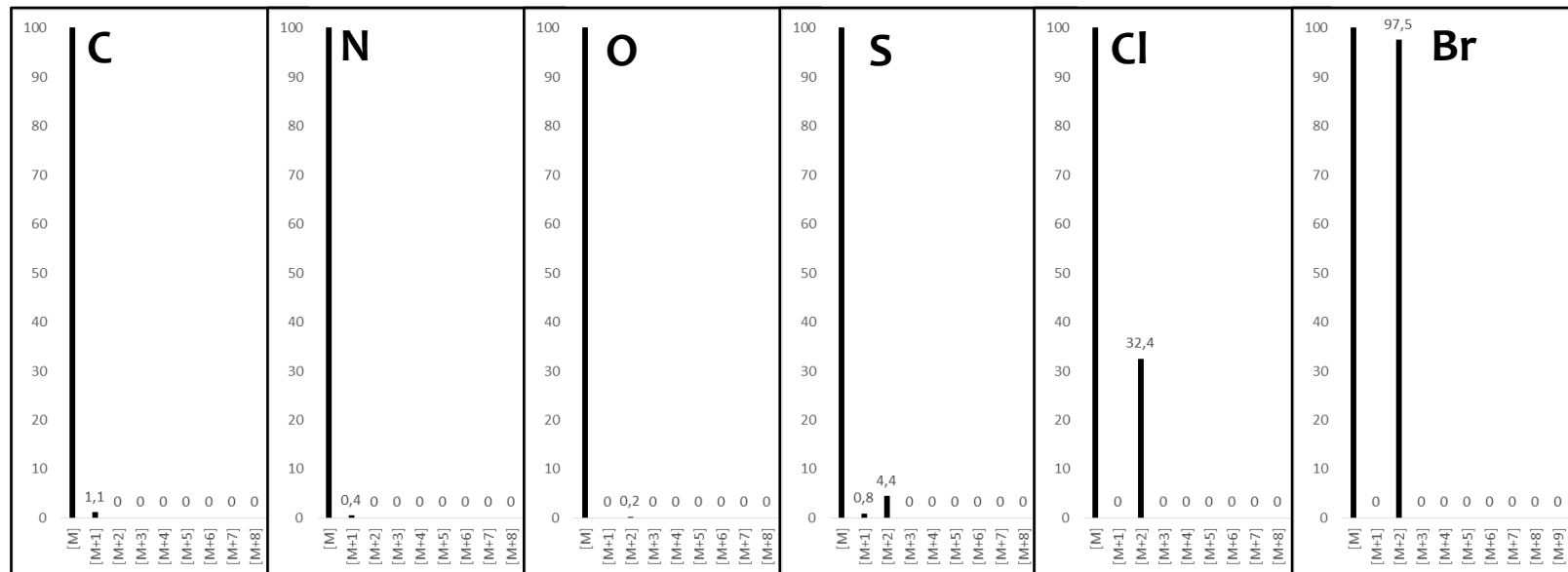
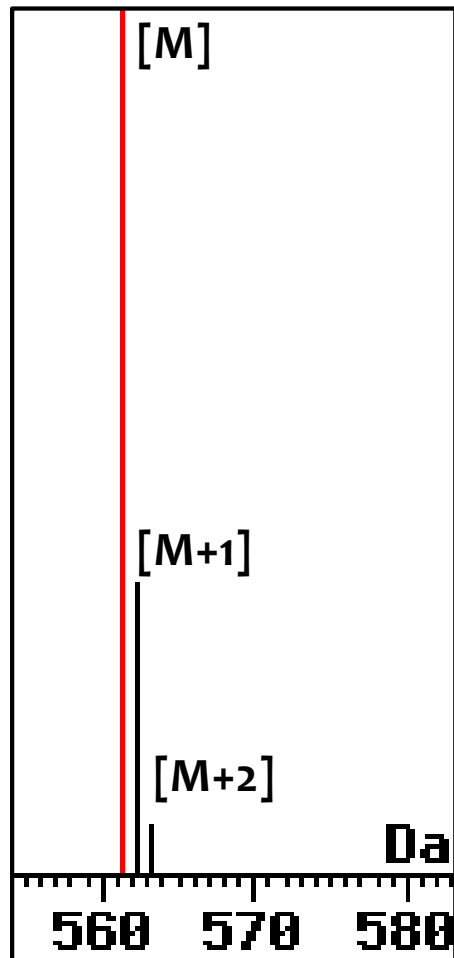
Abstract

Background: Metabolomic studies are targeted at identifying and quantifying all metabolites in a given biological context. Among the tools used for metabolomic research, mass spectrometry is one of the most powerful tools. However, metabolomics by mass spectrometry always reveals a high number of unknown compounds which complicate in depth mechanistic or biochemical understanding. In principle, mass spectrometry can be utilized within strategies of *de novo* structure elucidation of small molecules, starting with the computation of the elemental composition of an unknown metabolite using accurate masses with errors <5 ppm (parts per million). However even with very high mass accuracy (<1 ppm) many chemically possible formulae are obtained in higher mass regions. In automatic routines an additional orthogonal filter therefore needs to be applied in order to reduce the number of potential elemental compositions. This report demonstrates the

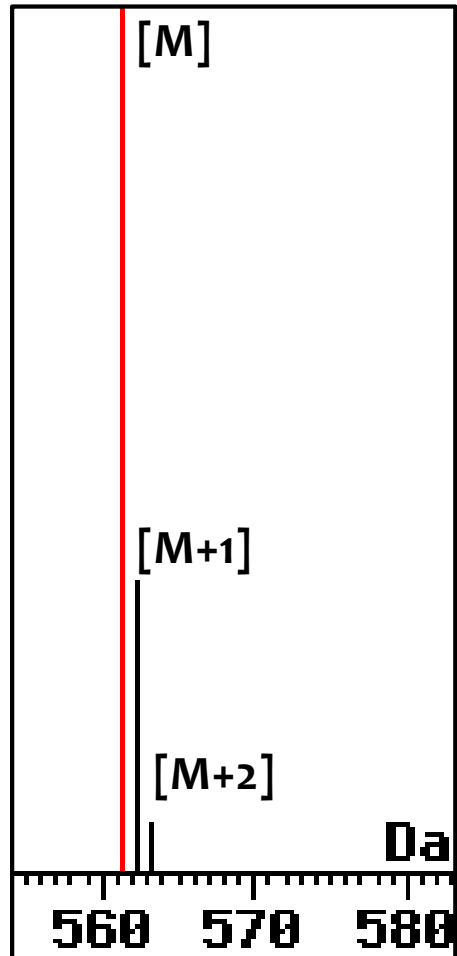
Détermination de la formule brute

$[M+1]$ dépend du nombre de C

$[M+2]$ dépend du nombre de C, S, Cl, Br



Détermination de la formule brute

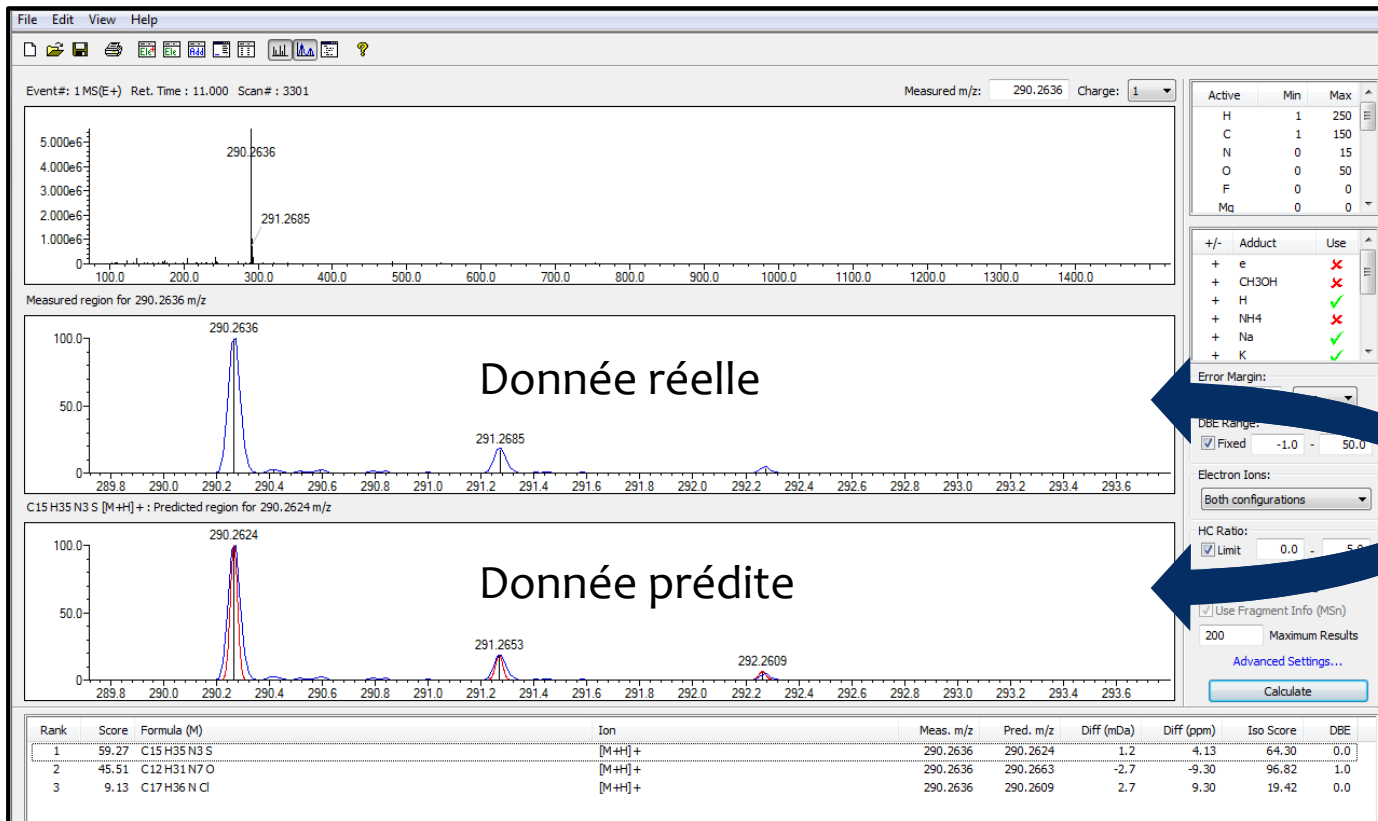


$[M+1]$ dépend du nombre de ^{13}C

$[M+2]$ dépend du nombre de ^{13}C , ^{34}S , ^{37}Cl , ^{81}Br

avec $[M+2]$ → détermination de la présence/nombre de S, Cl, Br

Détermination de la formule brute



Similitude des
ratios isotopiques

Détermination de la formule brute

Comment ça fonctionne ?

- 1- définition de l'ensemble formules possible dans la limite définie par l'utilisateur
- 2- calcul de l'erreur sur m/z (en ppm)
- 3- sélection des formules avec une erreur inférieure à la limite de l'appareil
- 4- utilisation de la ressemblance des ratios isotopiques

Approche utilisée par l'ensemble des logiciels.

molecular mass [Da]	2% isotopic abundance accuracy		5% isotopic abundance accuracy	
	10 ppm	5 ppm	3 ppm	3 ppm
150	2	1	1	1
200	3	2	2	1
300	24	11	7	1
400	78	37	23	2
500	266	115	64	3
600	505	257	155	4
700	1046	538	321	10
800	1964	973	599	13
900	3447	1712	1045	18

Que faire quand il y a plusieurs formules possibles ?

➤ Utiliser les données de fragmentation

Anal. Chem. **2007**, *79*, 1187–1197

Molecular Formula Analysis by an MS/MS/MS Technique To Expedite Dereplication of Natural Products

Yasuo Konishi,^{*,†} Taira Kiyota,^{†,‡} Cristina Draghici,[†] Jin-Ming Gao,^{†,§} Faustinus Yeboah,^{†,||} Stéphane Acoca,^{†,⊥} Suwatchai Jarussophon,^{†,¶} and Enrico Purisima[†]

Biotechnology Research Institute, National Research Council Canada, 6100 Royalmount Avenue, Montréal, Québec

Research Article

Received: 14 September 2011

Revised: 29 June 2012

Accepted: 2 July 2012

Published online in Wiley Online Library

Rapid Commun. Mass Spectrom. **2012**, *26*, 2275–2286
(wileyonlinelibrary.com) DOI: 10.1002/rcm.6340

Fragmentation trees for the structural characterisation of metabolites

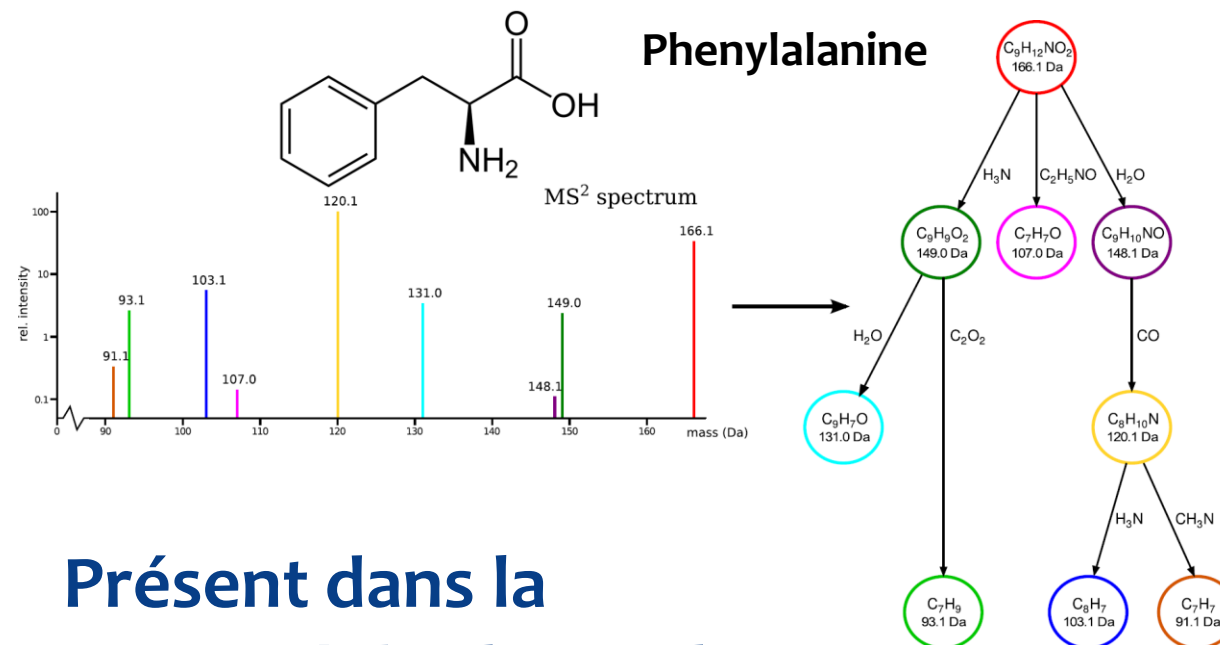
Piotr T. Kasper^{1,2}, Miguel Rojas-Chertó^{1,2}, Robert Mistrik³, Theo Reijmers^{1,2}, Thomas Hankemeier^{1,2} and Rob J. Vreeken^{1,2*}

¹Netherlands Metabolomics Centre, Einsteinweg 55, Leiden, The Netherlands

²Leiden/Amsterdam Centre for Drug Research (LACDR), Leiden University, Einsteinweg 55, Leiden, The Netherlands

³HighChem. Ltd., Bratislava, Slovakia

Metabolite identification plays a crucial role in the interpretation of metabolomics research results. Due to its sensitivity and widespread implementation, a favourite analytical method used in metabolomics is electrospray mass spectrometry.



Présent dans la majorité des logiciels.

Que faire quand il y a plusieurs formules possibles ?

➤ Utiliser la « logique » de la chimique ➡ « Seven Golden Rules »

BMC Bioinformatics



Research article

Open Access

Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry

Tobias Kind and Oliver Fiehn*

Address: University of California Davis, Genome Center, 451 E. Health Sci. Dr., Davis, CA 95616, USA

Email: Tobias Kind - tkind@ucdavis.edu; Oliver Fiehn* - ofiehn@ucdavis.edu

* Corresponding author

Published: 27 March 2007

Received: 16 December 2006

BMC Bioinformatics 2007, 8:105 doi:10.1186/1471-2105-8-105

Accepted: 27 March 2007

This article is available from: <http://www.biomedcentral.com/1471-2105/8/105>

© 2007 Kind and Fiehn; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Structure elucidation of unknown small molecules by mass spectrometry is a challenge despite advances in instrumentation. The first crucial step is to obtain correct elemental compositions. In order to automatically constrain the thousands of possible candidate structures, rules need to be developed to select the most likely and chemically correct molecular formulas.

Results: An algorithm for filtering molecular formulas is derived from seven heuristic rules: (1) restrictions for the number of elements, (2) LEWIS and SENIOR chemical rules, (3) isotopic patterns, (4) hydrogen/carbon ratios, (5) element ratio of nitrogen, oxygen, phosphorus, and sulphur versus carbon, (6) element ratio probabilities and (7) presence of trimethylsilylated compounds. Formulas are ranked according to their isotopic patterns and subsequently constrained by presence in public chemical databases. The seven rules were developed on 68,237 existing molecular formulas and were validated in four experiments. First, 432,968 formulas covering five million PubChem database entries were checked for consistency. Only 0.6% of these compounds did not pass all rules. Next, the rules were shown to effectively reducing the complement all eight billion theoretically possible C, H, N, S, O, P-formulas up to 2000 Da to only 623 million most probable elemental compositions. Thirdly, 6,000 pharmaceutical, toxic and natural compounds were selected from DrugBank, TSCA and DNP databases. The correct formulas were retrieved as top hit at 80–99% probability when assuming data acquisition with complete resolution of unique compounds and 5% absolute isotope ratio deviation and 3 ppm mass accuracy. Last, some exemplary

Les 7 règles pour filtrer les formules brutes dans le cas des molécules neutres:

- 1) Limiter le nombre d'éléments pour générer les formules brutes
- 2) Vérifier LEWIS et SENIOR
- 3) Vérifier les ratio isotopiques
- 4) Vérifier le rapport H/C
- 5) Vérifier le rapport N/C, O/C, P/C, S/C
- 6) Vérifier la probabilité du rapport H/C, N/C, O/C, P/C, S/C
- 7) Vérifier la présence de -TMS (GC-MS)

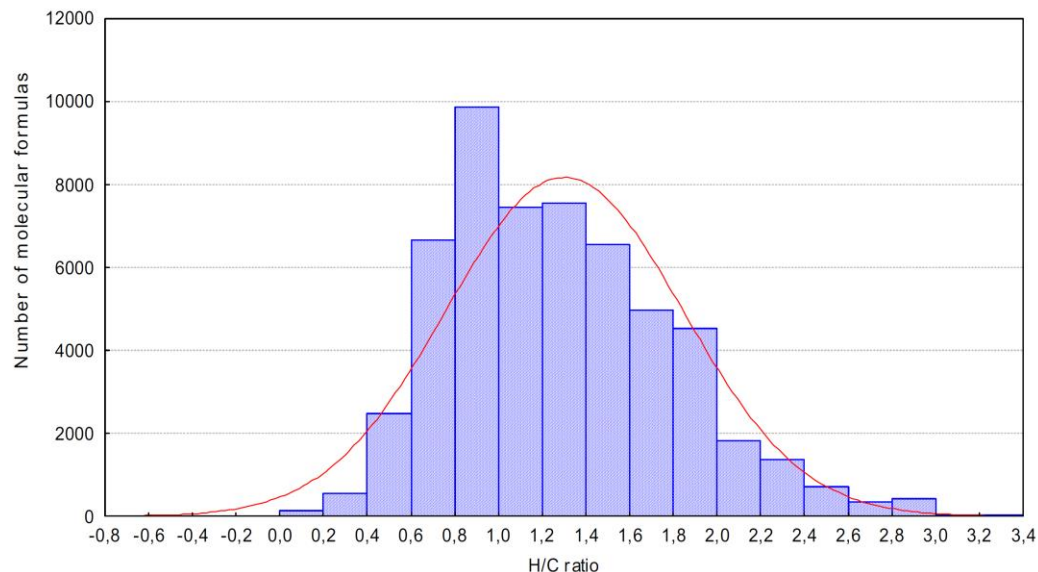


Figure 2
Hydrogen/Carbon ratio (H/C) for 42,000 diverse molecules (containing C, H, N, S, O, P, F, Cl, Br, I, Si) taken from the Wiley mass spectral library.

Table 2: Common element ratios obtained from 45,000 formulas comprising the Wiley mass spectral database for the mass range 30 Da – 1500 Da

Element ratios	Common range (covering 99.7%)	Extended range (covering 99.99%)	Extreme range (beyond 99.99%)
H/C	0.2–3.1	0.1–6	< 0.1 and 6–9
F/C	0–1.5	0–6	> 1.5
Cl/C	0–0.8	0–2	> 0.8
Br/C	0–0.8	0–2	> 0.8
N/C	0–1.3	0–4	> 1.3
O/C	0–1.2	0–3	> 1.2
P/C	0–0.3	0–2	> 0.3
S/C	0–0.8	0–3	> 0.8
Si/C	0–0.5	0–1	> 0.5

Que faire quand il y a plusieurs formules possibles ?

➤ Utiliser la « logique » de la chimique ➡ « Seven Golden Rules »

Table 6: Single performance of each rule of the seven rules from a total of 696 formulas comprising the elements CHNSOP and Si, calculated from GC-TOF data of sorbitol TMS₆

Rules for molecular formula filtering	Single application of each rule
1) heuristic restrictions for number of elements	not used (smart H option instead)
2) perform LEWIS and SENIOR check	can remove 420 candidates
3) isotopic pattern filter at 5% error	can remove 668 candidates
isotopic pattern filter at 10% error	can remove 632 candidates
isotopic pattern filter at 20% error	can remove 462 candidates
4) H/C ratio check (hydrogen/carbon ratio)	can remove 56 candidates
5) NOPS ratio check (N, O, P, S/C ratios)	can remove 51 candidates
6) heuristic HNOPS probability check	can remove 180 candidates
7) TMS check	can remove 432 candidates
combined 10 candidates left	

**Attention seulement 99,99%
des molécules naturelles**

Détermination de la formule brute

*Pas de panique ! C'est souvent
présent dans les logiciels !*



Recherche dans les bases de données

Est-ce que cette formule existe dans les bases de données ?



Attention la formule brut de l'ion ne se trouve pas dans les bases de données !

Recherche dans les bases de données

Est-ce que cette formule existe dans les bases de données ?

On a bien travaillé donc on a la
formule brut de la molécule.



Name	Type of NP	Number of entries	Complementary information					Simulated data	SOAP or API for automated search
			MS ² spectra	UV spectra	NMR spectra	Biological sources	Pathway		
AntiBase	Secondary metabolites	42,000	Yes	Yes	¹³ C	Yes	No	No	No
ChEBI	Primary and secondary metabolites	32,000	No	No	No	Yes	No	No	Yes
Dictionary of Marine Natural Products	Marine secondary metabolites	30,000	No	Yes	No	Yes	No	No	No
Dictionary of Natural Product	Secondary metabolites	260,000	No	Yes	No	Yes	No	No	No
E. coli Metabolome Database (ECMDB)	Primary and secondary metabolites	3,000	Yes	No	Yes	Yes	Yes	No	No
Golm Metabolome	Primary and secondary metabolites	2,200	No	No	No	Yes	No	No	Yes
Human Metabolome database (HMDB)	Primary metabolites and lipids	42,000	Yes	No	No	Yes	No	No	Yes
KNApSack	Secondary metabolites	51,000	No	No	No	Yes	No	No	Yes
Kyoto Encyclopedia of Genes and Genomes (KEGG)	Primary and secondary metabolites	17,000	No	No	No	Yes	Yes	No	Yes
Lipid Maps	Lipids	38,000	Yes	No	No	No	Yes	No	Yes
LipidBlast	Lipids	119,000	Yes	No	No	No	No	Yes	No
Madison Metabolomics Consortium Database	Primary and secondary metabolites	20,000	No	No	Yes	Yes	No	No	No
MarinLit	Secondary metabolites	5,000	No	Yes	¹ H	Yes	No	No	No
MassBank	Primary and secondary metabolites	15,000	Yes	No	No	No	No	No	Yes
MetaboLights	Primary and secondary metabolites	11,000	Yes	No	Yes	Yes	Yes	No	No
MetaCyc	Primary and secondary metabolites	11,000	No	No	No	Yes	Yes	No	Yes
MetLin	Primary and secondary metabolites	240,000	Yes	No	No	No	No	No	No
MFSearcher	Primary and secondary metabolites with a focus on flavonoids	-	No	No	No	No	No	No	Yes
MZedDB	Primary and secondary metabolites	-	No	No	No	No	No	No	Yes
NIST12	Primary and secondary metabolites	7,000	Yes	No	No	No	No	No	No
Plant Metabolic Network (PMN)	Primary and secondary metabolites	4,000	No	No	No	No	Yes	No	Yes
Seaweed Metabolite Database	Secondary metabolites from seaweeds	1,000	No	No	No	Yes	No	No	No
SuperNatural II	Primary and secondary metabolites	325,000	No	No	No	No	Yes	No	No
RIKEN	Secondary metabolites	3,500	Yes	No	No	No	No	No	No
Universal Natural Product Database	Secondary metabolites	250,000	No	No	No	Yes	No	No	No
Yeast Metabolome database (YMDB)	Primary and secondary metabolites	2,000	Yes	No	Yes	Yes	Yes	No	No

25 bases de données

Cette liste n'est pas exhaustive !!!

Recherche dans les bases de données



PubChem

BioAssay 2 Compound 2 Substance 2

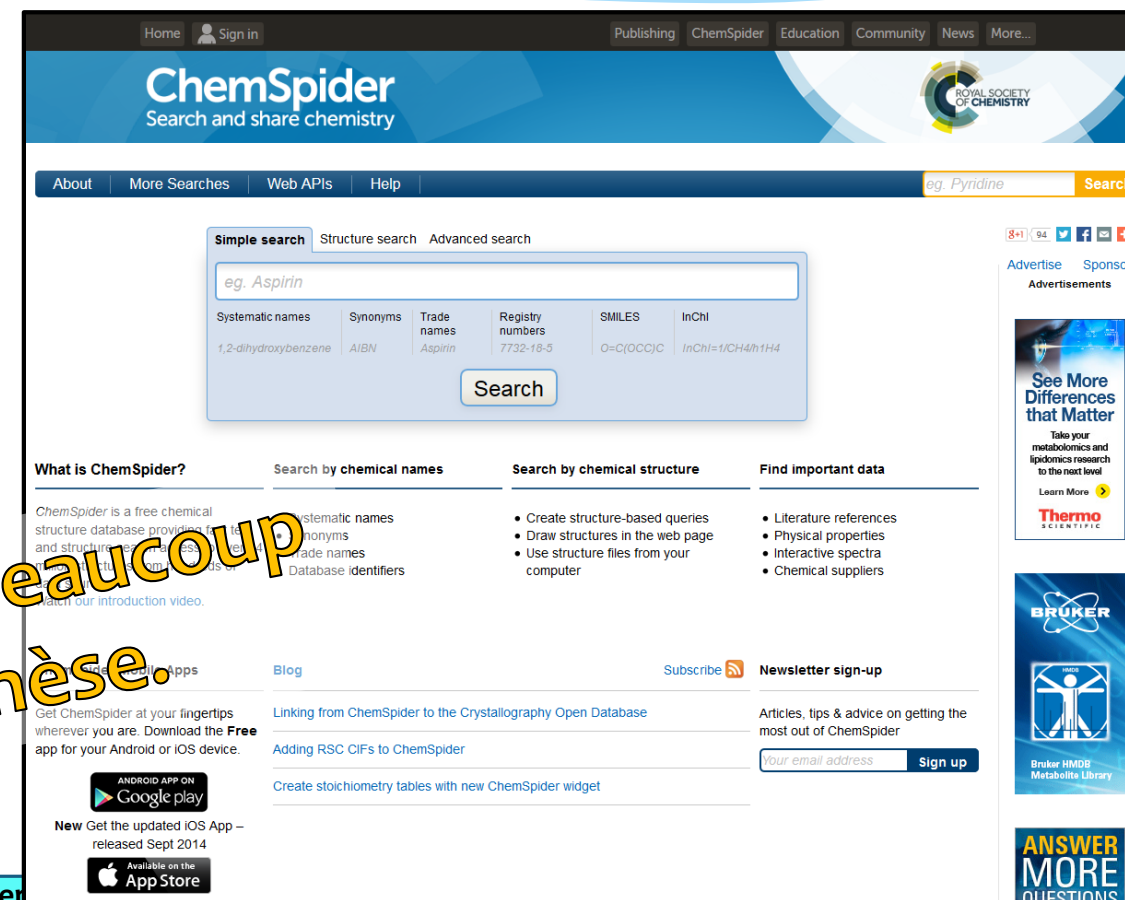
Go Limits Advanced

Try the new PubChem Search

New The PubChem Substance Record page is now released. [Read more...](#)

Write to Helpdesk | Disclaimer | Privacy Statement | Accessibility | Data Citation Guidelines
National Center for Biotechnology Information
NLM | NIH | HHS

- BioActivity Summary
- BioActivity Databale
- BioActivity SAR
- BioActivity DataDicer
- Structure Search
- 3D Conformer Tools
- Structure Clustering
- Classification
- Upload
- Download
- PubChem FTP



ChemSpider
Search and share chemistry

Home Sign in Publishing ChemSpider Education Community News More...

About More Searches Web APIs Help

eg. Pyridine Search

Simple search Structure search Advanced search

eg. Aspirin

Systematic names	Synonyms	Trade names	Registry numbers	SMILES	InChI
1,2-dihydroxybenzene	Aspirin	Aspirin	7732-18-5	O=C(OCC)C	InChI=1/CH4H1H4

Search

What is ChemSpider?

Search by chemical names

Search by chemical structure

Find important data

ChemSpider is a free chemical structure database providing fast access to systematic names, synonyms, trade names, and database identifiers.

- Create structure-based queries
- Draw structures in the web page
- Use structure files from your computer

- Literature references
- Physical properties
- Interactive spectra
- Chemical suppliers

Get ChemSpider at your fingertips wherever you are. Download the Free app for your Android or iOS device.

Android App on Google play

New Get the updated iOS App – released Sept 2014

Available on the App Store

Blog

Linking from ChemSpider to the Crystallography Open Database

Adding RSC CIFs to ChemSpider

Create stoichiometry tables with new ChemSpider widget

Subscribe Newsletter sign-up

Articles, tips & advice on getting the most out of ChemSpider

Your email address Sign up

See More Differences that Matter

Take your metabolomics and lipidomics research to the next level

Learn More

Thermo Scientific

BRUKER

BRUKER HMDS Metabolite Library

ANSWER MORE QUESTIONS

Très complète mais avec beaucoup de composés de synthèse.

Recherche dans les bases de données

Attention !!

- Les bases s'interrogent séparément !
- Elle n'ont pas toujours les même données:
 - Présence de MS^2 , UV, source biologique, ... (OU PAS !!!)
 - Structure sous forme de mol, SMILES, InChI, ... (OU PAS !!!)
 - Identifiant: CAS, InChIKey, noms, ...
- Peuvent contenir des erreurs



Recherche dans les bases de données

Exemple de problème

ChemSpider

Search and share chemistry

About More Searches Web APIs Help

Californidine

ChemSpider ID: **28941458**

Molecular Formula: **C₂₀H₂₀NO₄**

Average mass: 338.377 Da

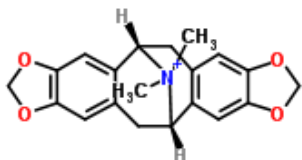
Monoisotopic mass: 338.138672 Da

Systematic name

(1S,12S)-23,23-Dimethyl-5,7,16,18-tetraoxa-23-azoniahexacyclo[10.10.1.0^{2,10}.0^{4,8}.0^{13,21}.0^{15,19}]tricos-2(10),3,8,13(21),14,19-hexaene

► SMILES and InChIs

► Cite this record



? 2D 3D Save Zoom

⊕ - Charge

⚙ - 2 of 2 defined stereocentres

PubChem OPEN CHEMISTRY DATABASE

Search Compounds

Compound Summary for CID 45266443

Download

Print

Share

Californidine

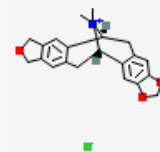
Read about the new page format and features



Literature



Bioactivities



PubChem CID: 45266443

Chemical Names: californidine; ChEMBL538357; More...

Molecular Formula: **C₂₁H₂₂ClNO₃**

Molecular Weight: 371.85728 g/mol

InChI Key: GLCGPUDWOPXBAY-HLRBRJAUSA-M

Modify Date: 2015-05-02

Create Date: 2010-05-07

Contents

1 2D Structure

2 3D Conformer

3 Identification

1 2D Structure

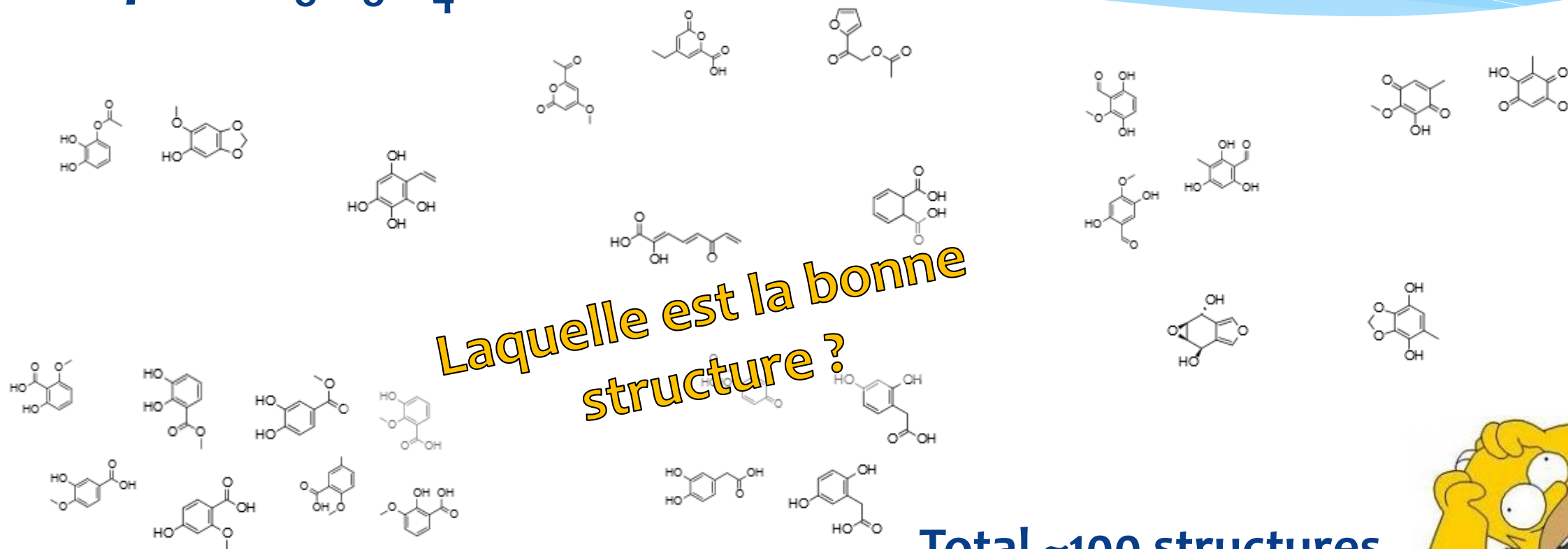
Search

Download

Get Image

Recherche dans les bases de données

Exemple : $C_8H_8O_4$



Laquelle est la bonne structure ?

Total ~100 structures



Discrimination/Confirmation de l'identification

Le plus simple :

- 1- achat du(des) composé(s) possible(s)
- 2- analyse du(des) composé(s) dans les même conditions

Il est souvent difficile d'avoir tous les standards pour les produits naturels.

Discrimination/Confirmation de l'identification

Comment supprimer les identifications incohérentes ?

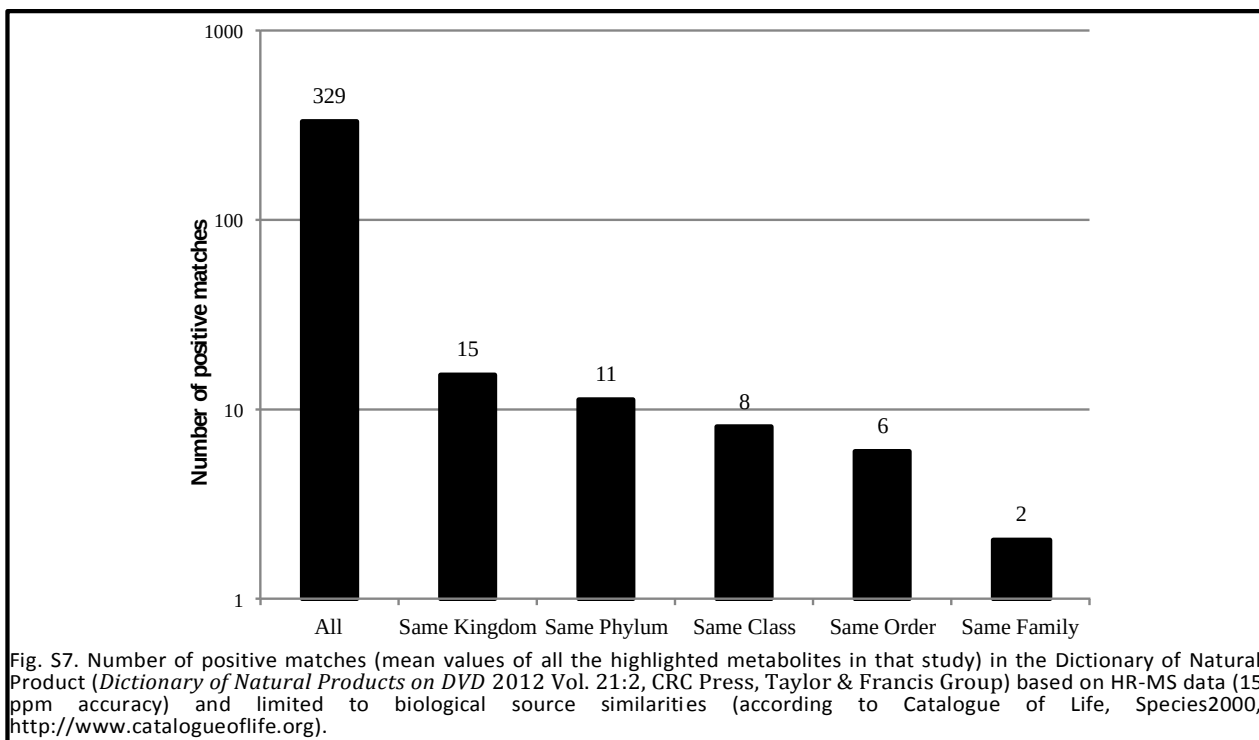
- 1- éliminer les source biologiques anormales
- 2- simuler les spectres MS^2
- 3- simuler les temps de rétention

1- éliminer les source biologiques anormales

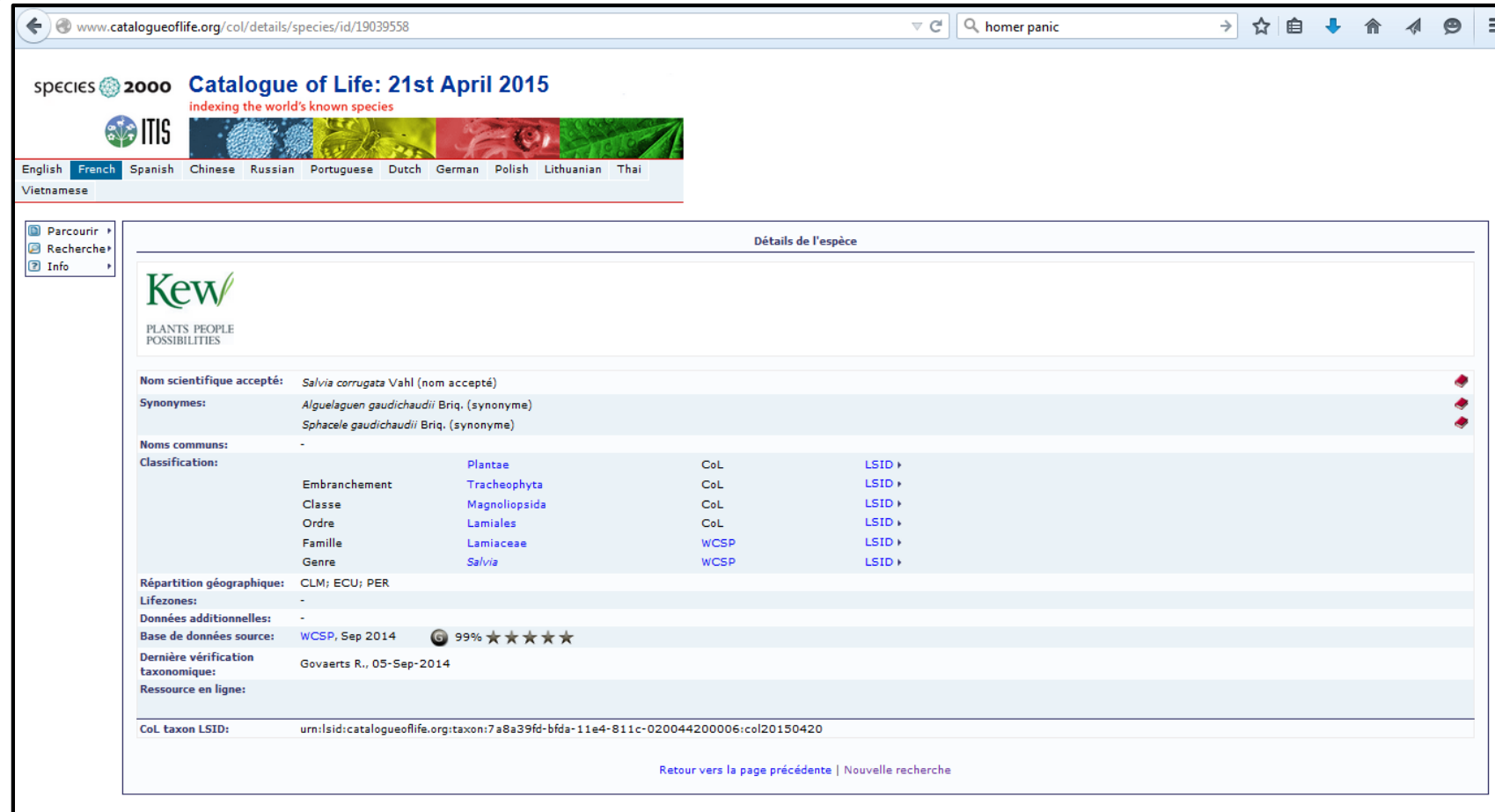
Comparaison de notre source avec celle(s) rapporté(s) pour le composé

	<i>Fusarium oxysporum</i>	<i>Salvia corrugata</i>
Règne	Fungi	Plantae
Embranchement	Ascomycota	Tracheophyta
Classe	Sordariomycetes	Magnoliopsida
Ordre	Hypocreales	Lamiales
Famille	Nectriaceae	Lamiaceae
Genre	<i>Fusarium</i>	<i>Salvia</i>

Bertrand S. *et al.* Molecular BioSystems **2014** 10(9) 2289-2298; <http://www.catalogueoflife.org/>



Données accessibles par exemple sur catalogue of life (<http://www.catalogueoflife.org/>)



The screenshot shows the 'Catalogue of Life: 21st April 2015' website. The page is in French and displays the details for the species *Salvia corrugata* Vahl. The page includes a navigation menu on the left with options like 'Parcourir', 'Recherche', and 'Info'. The main content area is titled 'Détails de l'espèce' and features the Kew logo. The species name is listed as *Salvia corrugata* Vahl (nom accepté). Synonyms include *Alvaelaguen gaudichaudii* Briq. and *Sphacele gaudichaudii* Briq. The classification is shown in a table with columns for taxonomic rank, name, and source. The distribution is listed as CLM; ECU; PER. The page also includes a 'Base de données source' section with a 99% rating and a 'Dernière vérification taxonomique' date of 05-Sep-2014. At the bottom, there is a 'Col. taxon LSID' and a 'Ressource en ligne' link.

species 2000 Catalogue of Life: 21st April 2015
indexing the world's known species

English **French** Spanish Chinese Russian Portuguese Dutch German Polish Lithuanian Thai Vietnamese

Parcourir
Recherche
Info

Détails de l'espèce

Kew
PLANTS PEOPLE
POSSIBILITIES

Nom scientifique accepté: *Salvia corrugata* Vahl (nom accepté)

Synonymes:
Alvaelaguen gaudichaudii Briq. (synonyme)
Sphacele gaudichaudii Briq. (synonyme)

Noms communs: -

Classification:				
Embranchement	Plantae	CoL	LSID	
	Tracheophyta	CoL	LSID	
Classe	Magnoliopsida	CoL	LSID	
Ordre	Lamiales	CoL	LSID	
Famille	Lamiaceae	WCSP	LSID	
Genre	Salvia	WCSP	LSID	

Répartition géographique: CLM; ECU; PER

Lifezones: -

Données additionnelles: -

Base de données source: WCSP, Sep 2014 99% ★★★★★

Dernière vérification taxonomique: Govaerts R., 05-Sep-2014

Ressource en ligne:

Col. taxon LSID: urn:lsid:catalogueoflife.org:taxon:7a8a39fd-bfda-11e4-811c-020044200006:col20150420

[Retour vers la page précédente](#) | [Nouvelle recherche](#)

Données accessibles aussi via Pubmed

NCBI Taxonomy Browser

Entrez PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books

Search for complete name ☒ lock

Display levels using filter: none

Salvia corrugata

Taxonomy ID: 268901
 Inherited blast name: eudicots
 Rank: species
 Genetic code: [Translation table 1 \(Standard\)](#)
 Mitochondrial genetic code: [Translation table 1 \(Standard\)](#)
 Other names:
 synonym: **Salvia corrugata Vahl**

[Lineage\(full \)](#)
[cellular organisms](#); [Eukaryota](#); [Viridiplantae](#); [Streptophyta](#); [Streptophytina](#); [Embryophyta](#); [Tracheophyta](#); [Euphyllophyta](#); [Spermatophyta](#); [Magnoliophyta](#); [Mesangiospermae](#); [eudicotyledons](#); [Gunneridae](#); [Pentapetalae](#); [asterids](#); [lamiids](#); [Lamiales](#); [Lamiaceae](#); [Nepetoideae](#); [Mentheae](#); [Salvia](#)

Comments and References:

 [IPNI](#)
 Name verified on date of entry into database in: The International Plant Names Index (2004 and onward).

 [TROPICOS](#)
 Name verified on date of entry into taxonomy database in: W3TROPICOS, the Missouri Botanical Garden's VAST (VAscular Tropicos) nomenclatural database, vers. 1.5, Missouri Botanical Garden, St. Louis, MO, 63166-0299, USA.

External Information Resources (NCBI LinkOut)

LinkOut	Subject	LinkOut Provider
search GBIF	taxonomy/phylogenetic	Global Biodiversity Information Facility
Salvia corrugata	taxonomy/phylogenetic	The International Plant Names Index
search W3TROPICOS	taxonomy/phylogenetic	Vascular Tropicos
Salvia corrugata	taxonomy/phylogenetic	electronic Plant Information Centre
Wikipedia	taxonomy/phylogenetic	iPhylo

Notes:
 Groups interested in participating in the LinkOut program should visit the [LinkOut home page](#).
 A list of our current non-bibliographic LinkOut providers can be found [here](#).
 To see LinkOut links in this lineage click [here](#)

Discrimination/Confirmation de l'identification

Comment simuler les spectres MS^2

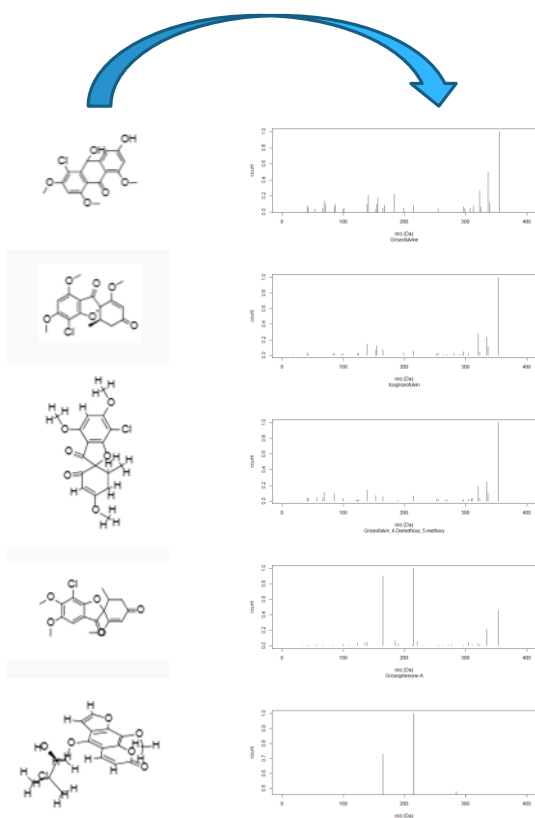


MAGMa

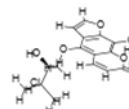
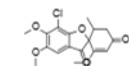
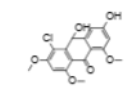
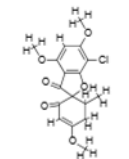
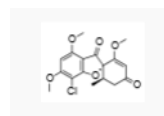


Discrimination/Confirmation de l'identification

Simulation



Comparaison
avec le spectre
mesuré



Plus similaire
Plus probable

Moins similaire
Moins probable



MetFrag

In silico fragmentation for computer assisted identification of metabolite mass spectra

MetFrag MzAnnotate Viewer About / News

Database Settings

Database: ☒ KEGG ☐ PubChem ☐ ChemSpider ☐ Local SDF

Neutral exact mass: Search PPM:

Molecular formula:

Only biological compounds: ☒

Limit # of structures:

Database ID's:

Search upstream DB

MetFrag Settings

Mode: ☒ [M+H] ☐ [M-H] ☐ [M]

Charge: ☒ pos. ☐ neg.

Mzabs (e.g. 0.01):

Mzppm (e.g. 10):

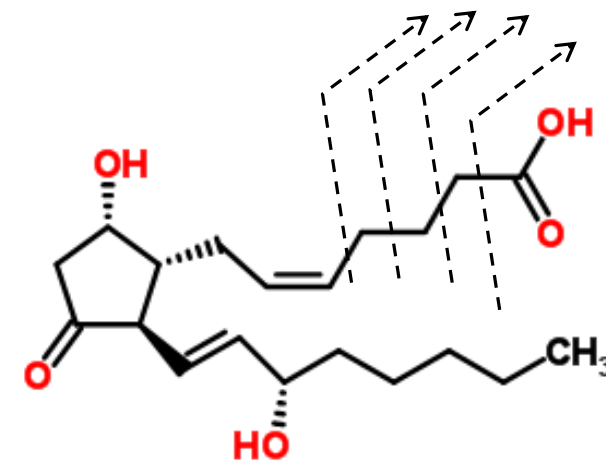
Parent ion: Neutral

Peaks:

- 119.051 467.616
- 123.044 370.662
- 147.044 6078.145
- 153.019 10000.0
- 179.036 141.192
- 189.058 176.358
- 273.076 10000.000
- 274.083 318.003

[View spectrum](#)

Log



Peu de sélectivité
dans la fragmentation

Automatisable

<http://cfmid.wishartlab.com/>

Spectrum Query
Information

Selected Spectral Database:
MassBank

MassBank Parameters
http://massbank.ufl.de/MassBank/

MassBank Server:
MassBank EU

Number of Results:
100

Cutoff threshold of relative intensities:
5

Ionization Mode:
positive

MS Levels:
☒ all ☒ MS ☒ MS2 ☒ MS3 ☒ MS4

Instruments:

Select EI

☐ EI-B
☐ EI-EBEB
☐ GC-EI-QQ
☐ GC-EI-TOF

Deselect ESI

☒ CE-ESI-TOF
☒ ESI-FTICR
☒ ESI-ITFT
☒ ESI-ITTOF
☒ HPLC-ESI-TOF
☒ LC-ESHT
☒ LC-ESHTFT
☒ LC-ESIHTTOF
☒ LC-ESI-Q
☒ LC-ESI-QFT
☒ LC-ESI-QIT
☒ LC-ESI-QQ
☒ LC-ESI-QTOF
☒ LC-ESI-TOF
☒ UPLC-ESI-QTOF

Select Others

☐ APCHITFT
☐ APCHITTOF
☐ CI-B
☐ FAB-B
☐ FAB-EB
☐ FAB-EBEB
☐ FD-B
☐ FI-B
☐ LC-APCI-Q
☐ LC-APCI-QTOF
☐ LC-APPI-QQ
☐ MALDI-QIT
☐ MALDI-TOF
☐ MALDI-TOFTOF

MetFrag Parameters

Upstream DB:
☒ KEGG ☐ PubChem ☐ ChemSpider ☐ SDF Upload

Database IDs:

Molecular Formula:

Parent Ion:
272.06847
Neutral

Exact Mass:
272.06847

Limit # of Structures:
500

C,H,N,O,P,S only?
☒

MetFrag is set to work in positive mode

MetFusion Parameters

Filter
☒

Unique:

Peaks:
119.051 46
123.044 37
147.044 607
153.019 999
179.036 14
189.058 17
273.076 999
274.083 31

MetFusion Progress: - 0%



Limitation aux fragments existant dans MassBank, Metlin, GMD

Pas trop de prise en compte du type d'adduits dans la fragmentation !

<http://msbi.ipb-halle.de/MetFusion/>

ans les bases de données

Confirmation de l'identification

[Home](#)
[Help](#)

Workspace







MS Data

Format: mzXML

Enter MS data in a Tree format or mzXML

Examples: Chlorogenic acid (Mass Tree) Chlorogenic acid (Formula Tree)

or

Upload MS/MS data file Browse...

Filter options:

MS level ≤ 10

Noise filter > 5000

MS1 scan number

Molecules

Database Upload Draw

Format: smiles

Enter SDF, or smiles strings followed by space and name on each line

or

Upload structures from file Browse...

☒ Metabolize

Only first molecule will be metabolized

Scenario

Predefined scenarios: Drugs Polyphenols + Add transformation

Transformation type	Steps	
phase1	2	↑ ↓ -
phase2	1	↑ ↓ -

Parameter settings

Ionisation mode: ☐ Negative ☒ Positive

Substructure options:

Bond dissociations ≤ 3

Additional small losses ≤ 1

Accuracy:

Relative (ppm) ≤ 5

Absolute (Da) ≤ 0.001

Precursor m/z (Da) ≤ 0.005

Intensity thresholds:

MS¹ (abs.) ≥ 0

MSⁿ>1 (% of base peak) ≥ 5

Très adapté aux fragmentations successives

<http://www.emetabolomics.org/magma>

CFM-ID

Competitive Fragmentation Modeling for Metabolite Identification

Compound Identification

Determines the compounds that most closely match a given spectra. The spectra for each candidate compound (in the provided list) are predicted using a pre-trained model and compared to the input spectra. The top candidates are ranked according to how closely they match and returned in a list.

Candidate Compounds

OR

When querying a database, the parent ion mass and adduct type will be used to approximate the molecular weight of the query compound. This molecular weight will then be used to select the most likely candidate compounds from the database (using the candidate mass tolerance), against which the query spectra is compared. The maximum candidate compound size is 200 atoms. [Load an example](#) or [another example](#).

Spectra OR

The spectra should be represented as a list of peaks with the format 'mass intensity' on each line, and can be entered directly into the corresponding energy level boxes below. Multiple energy levels are optional; only one is required.

Input Spectra Text

Low Energy

Medium Energy

High Energy

Number of Results The number of results to return. Leave blank to return all results.

Mass Tolerance The mass tolerance to use when matching peaks within the spectrum comparison.

Scoring Function

Uniquement les adduits que
CFM-ID a appris

Plus sélectif.

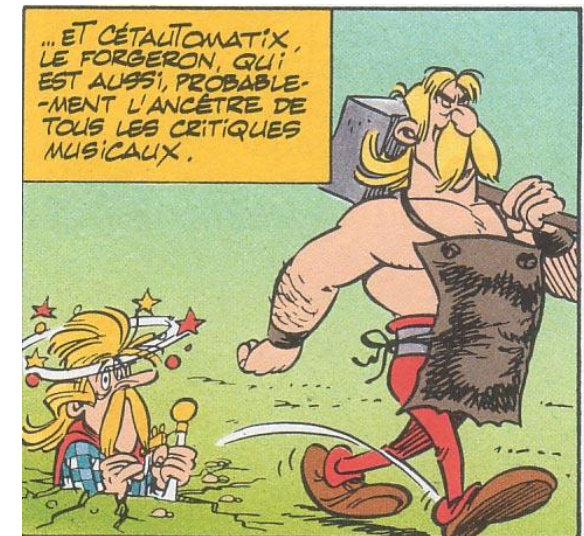
Automatisable

<http://cfmid.wishartlab.com>

Discrimination/Confirmation de l'identification

Que choisir ?

Testez et faites vous votre idée !



Phytochemistry 108 (2014) 196–207

Contents lists available at ScienceDirect

Phytochemistry

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



ELSEVIER



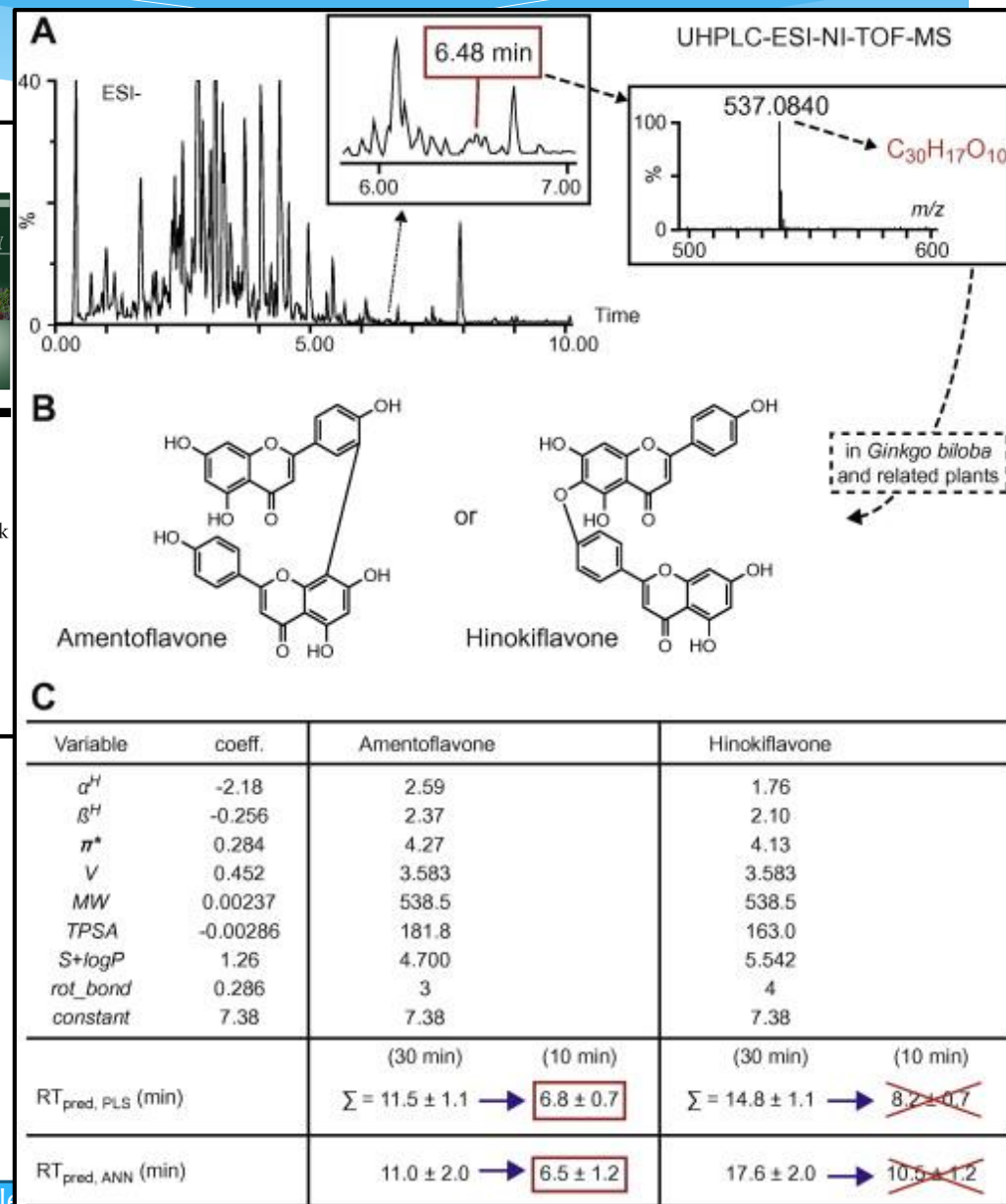
Retention time prediction for dereplication of natural products ($C_xH_yO_z$) in LC-MS metabolite profiling

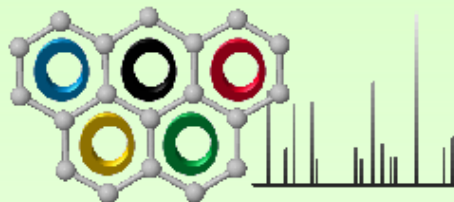
Philippe J. Eugster, Julien Boccard, Benjamin Debrus, Lise Bréant, Jean-Luc Wolfender*, Sophie Martel, Pierre-Alain Carrupt

School of Pharmaceutical Sciences, EPGL, University of Geneva, University of Lausanne, 30, Quai Ernest-Ansermet, CH-1211 Geneva 4, Switzerland



L'obtention de modèle générique est complexe !





CASMI 2014

CASMI 2016

CASMI 2014

[Important Dates](#)
[Contest Rules](#)
[Example Data](#)
[Challenge Data](#)
[Solutions](#)
[Results](#)
[About the Team](#)

CASMI 2013

CASMI 2012

News

March 18th, 2015

The [results and winners](#) of CASMI-2014 are now up.

March 3rd, 2015

The solutions for CASMI-2014 are now available on the [website](#).

January 15th, 2015

Critical Assessment of Small Molecule Identification

The experimental and computational mass spectrometry community is invited to participate in the third round of an open contest on the identification of small molecules from mass spectrometry data. This contest follows two previous and successful contests in 2012 and 2013.

In many different scientific disciplines, including metabolomics and natural products research, we acquire data from complex samples. The annotation or identification of single compounds from these complex samples is a difficult task and requires robustness in the process and reporting of the result and methods applied.

In the current and previous contests we wish to exchange ideas on how to annotate and identify small molecules and perform a comparative evaluation of alternative strategies. In these contests mass spectral or mass spectral and chromatographic data, as well as sample origin is provided and from which participants propose annotations.

CASMI 2014 is the third round of CASMI and is organised by Dr Warwick Dunn (University of Birmingham), Dr Dejan Nikolic (University of Illinois at Chicago) and Professor Lloyd Sumner (The Samuel Roberts Noble Foundation). Support is provided by the Metabolite Identification task group of the [Metabolomics Society](#), the [American Society of Mass Spectrometry](#) and the [Mass Spectrometry, Metabolomics & Proteomics Facility of the University of Illinois at Chicago](#).

The aim is to bring all mass spectral communities together to show off their different methods on a common data set. Thus, although this is a competition, all participants should benefit and the real aim is to bring the field forward through identification.

Niveau 1 - Identification certaine

Obtenue par comparaison avec un standard dans les mêmes conditions analytiques (deux propriétés).

Niveau 2 - Identification probable

Par comparaison avec des spectres de banques spectrales externes.

Niveau 3 - Identification partielle

Présentant des similarités avec des spectres de composés connus ou des spectres *in silico*.

Niveau 4 - Composé non--identifié

Composé inconnu mais dont la présence est décelable.

Metabolomics
DOI 10.1007/s11306-014-0656-8

OPINION

Metabolite identification: are you sure? And how do your peers gauge your confidence?

Darren J. Creek · Warwick B. Dunn · Oliver Fiehn · Julian L. Griffin · Robert D. Hall · Zhentian Lei · Robert Mistrik · Steffen Neumann · Emma L. Schymanski · Lloyd W. Sumner · Robert Trengove · Jean-Luc Wolfender

© Springer Science+Business Media New York 2014

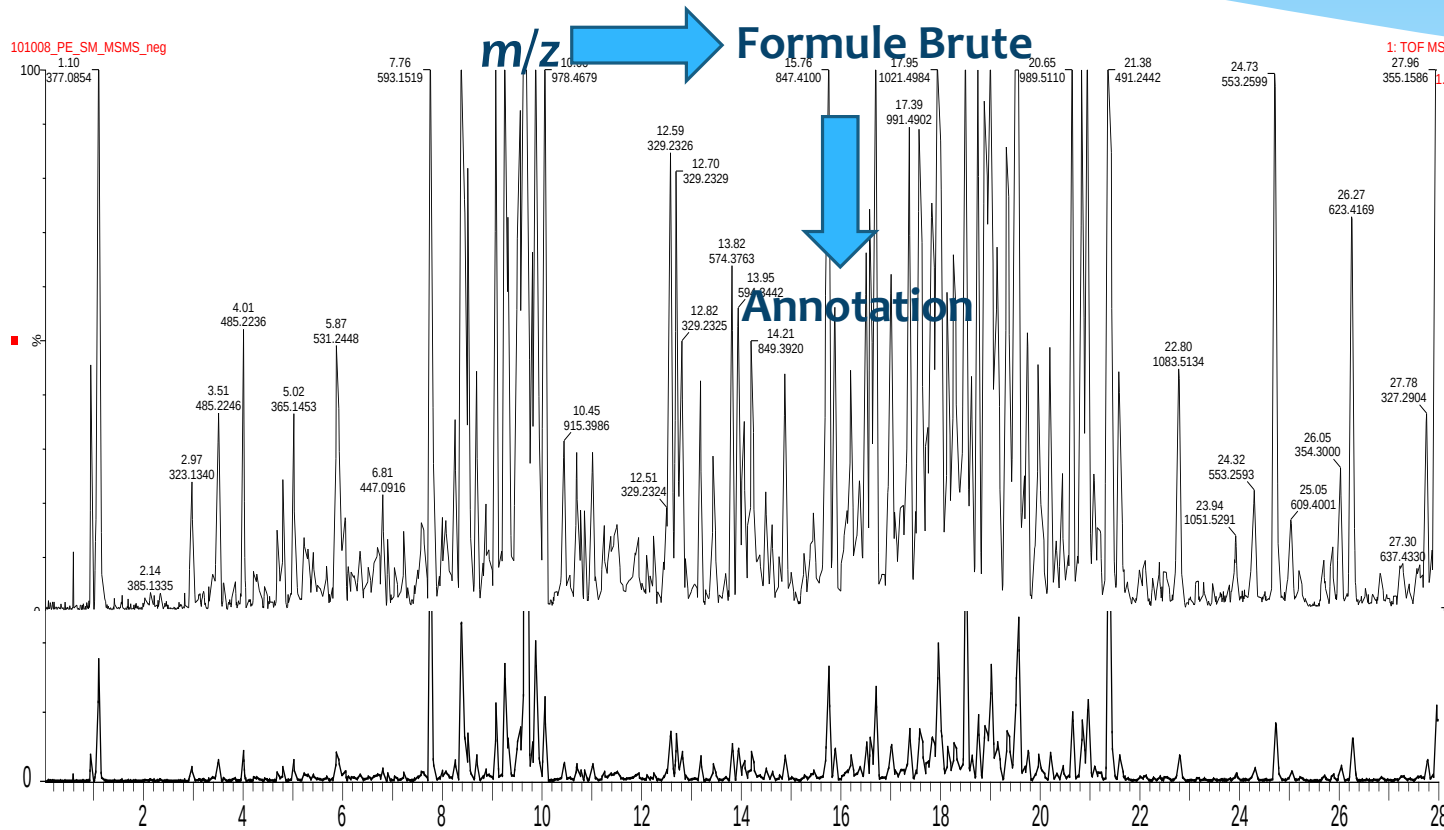
1 The identification challenge

Metabolomics is still faced with several significant challenges which currently limit its full scientific potential. The identification of metabolites is essential to convert analytical data into meaningful biological knowledge. However, identification confidence can vary widely because the process of identification is complex and dependent on the analytical platform and robustness of the methods applied, as well as the databases and resources used. Confident and unequivocal structure identification requires significant effort, which is multiplied dramatically in non-targeted metabolomics studies where 10–100s of metabolites can be deemed as biologically important and require identifica-

spectroscopy (NMR) or integrated MS–NMR strategies (Dunn et al. 2013; Kind and Fiehn 2010; van der Hoven et al. 2011) provide much information for the identification of metabolites (e.g. 1D/2D-NMR and MS/MS).

Relatively high confidence can only be obtained when comparing multiple physicochemical properties of an authentic pure chemical standard to those of the metabolite of interest observed under identical analytical conditions. However, isomers are a significant problem as they often behave differently in biological systems but similarly in analytical platforms with respect to mass, chromatographic retention time and NMR chemical shift. Chiral chromatography or judicious alterations of the column chemistry can help, but is typically not performed for general

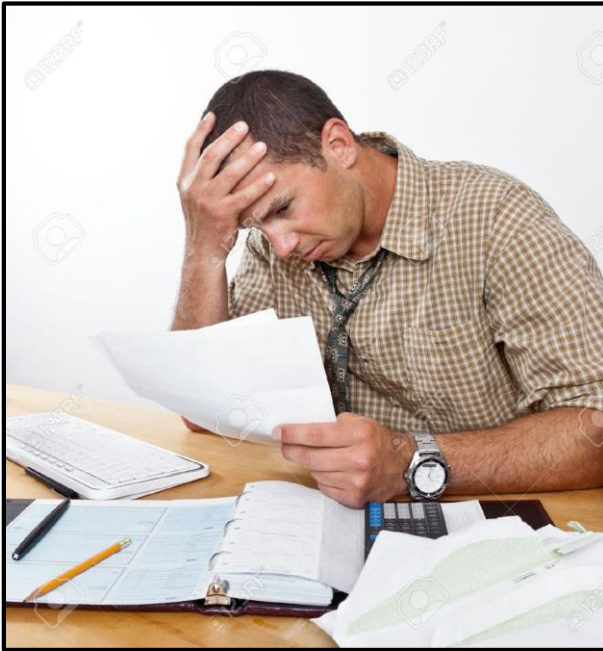
La déréplication



- * 1 pic → 1 jour
- * 40 pics → 40 jours !!

Et avec des profiles haute résolution ?

La déréplication...



Trucs et astuces pour aller plus vite !

Comment faire pratiquement ?

1- chercher dans les bases de données avec m/z puis filtrer

2- utiliser MZmine 2



Trucs et astuces pour aller plus vite !

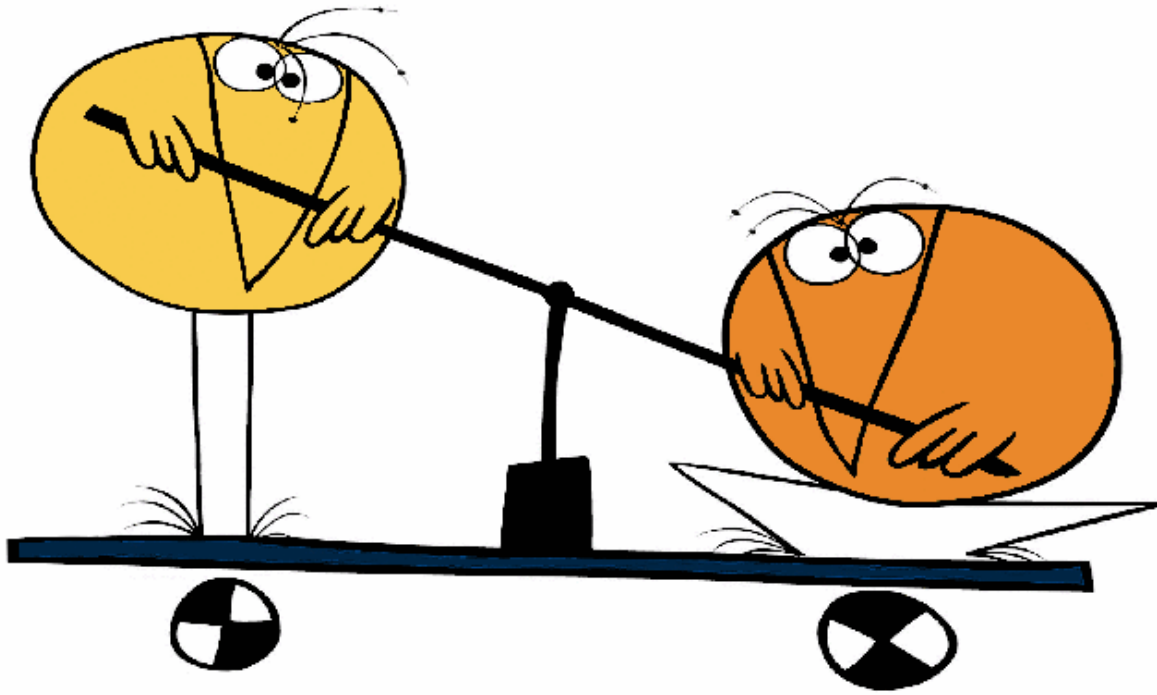
Comment faire en pratique ?

1- chercher dans les bases de données avec m/z puis filtrer

2- utiliser MZmine 2

3- automatiser le processus





N'ayez pas peur ca va bien se passer !



Que faire si le composé n'existe pas dans les bases de données ?

JOURNAL OF
NATURAL
PRODUCTS

Article

pubs.acs.org/jnp

Molecular Networking as a Dereplication Strategy

Jane Y. Yang,^{†,‡} Laura M. Sanchez,^{‡,§} Christopher M. Rath,[‡] Xueting Liu,[§] Paul D. Boudreau,[‡] Nicole Bruns,[‡] Evgenia Glukhov,[‡] Anne Wodtke,[‡] Rafael de Felicio,^{‡,||} Amanda Fenner,[‡] Weng Ruh Wong,[∇] Roger G. Linington,[∇] Lixin Zhang,[§] Hosana M. Debonisi,^{||} William H. Gerwick,[‡] and Pieter C. Dorrestein^{*,†,‡}

[†]Department of Chemistry & Biochemistry, University of California San Diego, La Jolla, California 92093, United States

[‡]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, California 92093, United States

[§]Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, 100190, China

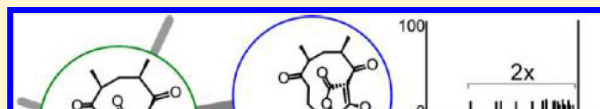
[‡]Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093, United States

^{||}Núcleo de Pesquisas Produtos Naturais e Sintéticos, Departamento de Física e Química, Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Universidade de São Paulo, Avenida Do Café, s/n, Campus Universitário, CEP 14040-903, Ribeirão Preto, São Paulo, Brazil

[∇]Department of Chemistry and Biochemistry, University of California Santa Cruz, Santa Cruz, California 95064, United States

 Supporting Information

ABSTRACT: A major goal in natural product discovery programs is to rapidly dereplicate known entities from complex biological extracts. We demonstrate here that molecular



Merci pour votre attention !

mms
mer molécules santé



mer et littoral
institut universitaire
FR CNRS 3473

UNIVERSITÉ DE NANTES

ThalassOMICS
ThalassOMICS



Biogenouest

Analyse structurale
et métabolomique

Plate-Forme
Corsaire
CORSAIRE