

CASMI2016から学ぶ化合物構造推定


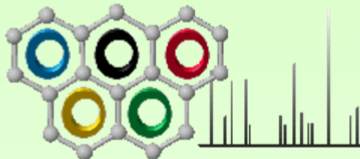
理研CSRS（リーダー・有田正規）・理研IMS（リーダー・有田誠）
津川 裕司

CASMI contest

Critical Assessment of Structure Nomenclature x

← → ↻ ⓘ www.casmi-contest.org/2017/challenges.shtml

🔍 ☆ 🔄 ⋮



CASMI 2017
Important Dates
Contest Rules
Challenge Data
About the Team

CASMI 2016

CASMI 2014

CASMI 2013

CASMI 2012

News
July 09th, 2017
We fixed the intensities in the TSV archive for challenges 046-243.

Overview | Challenges 1-45 | Challenges 46-243

CASMI 2017 will focus on testing current mass spectrometric approaches for identification of natural products. This year, CASMI will again have 3 categories based on (subsets of the) 243 challenges. These are:

Category 1
Manual approaches: Challenges 1-45

Category 2
Automated approaches (spectral data only): Challenges 1-243

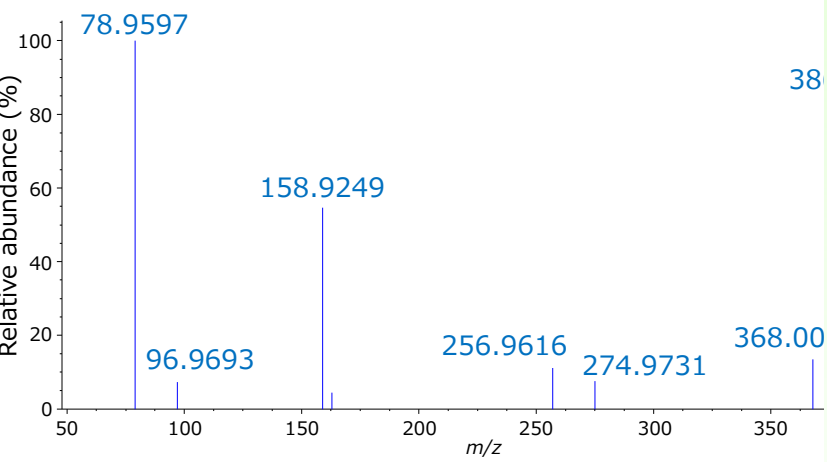
Category 3
Automated approaches (all metadata in): Challenges 1-243

If you are unsure in which category your approach fits, do not hesitate to contact the organisers (email link at the bottom). Challenge data and details on the experiments are available from the navigation menu above.

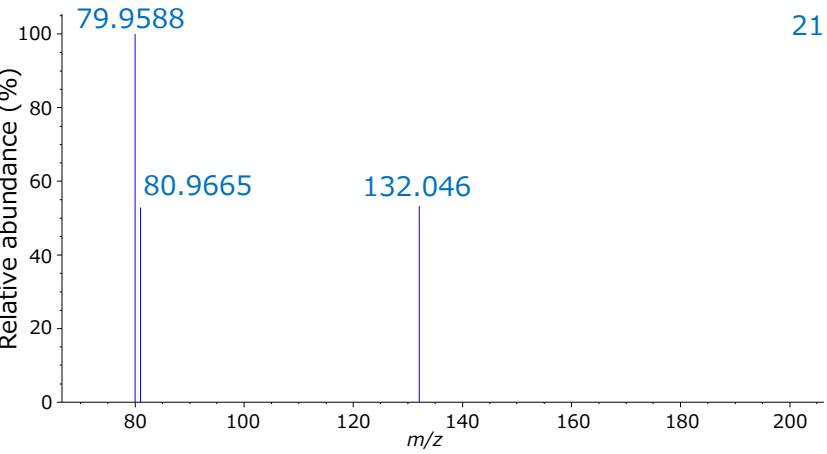
Update 20170522: We have fixed MS1 data for challenges 29, 42 and improved challenges 71, 89, 105, 106 and 144.

このMS/MSスペクトルの元構造は何でしょう？

1. Negative, <10ppm, Human



3. Negative, <10ppm, Human



Challenge Solutions Category 1

Please use the tabs above to navigate to the solutions for Categories 2+3.

The information is also available as [CSV](#) for download

ChallengeName	Compound	Structure
Challenge-001	Dibromophakellin C ₁₁ H ₁₁ Br ₂ N ₅ O PubChem: 42636938 ChemSpider: 9592765	
Challenge-002	Oroidin C ₁₁ H ₁₁ Br ₂ N ₅ O PubChem: 6312649 ChemSpider: 4880362	
Challenge-004	Cytochalasin B C ₂₉ H ₃₇ NO ₅ PubChem: 5311281 ChemSpider: 4470791	

Computational MSとデータベース

MS-FINDER ver. 2.14 D:\20151016-MSFINDER-Evaluation\Bioresource base\HILIC Human Plasma Pos-CNOSP-MS-5mDa-MSMS-25mDa-Rel-2.5

File Analysis Setting Export Help

File navigator

2015_5_31_053_HILIC_IDA_45CE_15CES_S02_b_0.88_848.7653
2015_5_31_053_HILIC_IDA_45CE_15CES_S02_b_0.88_868.7387
2015_5_31_053_HILIC_IDA_45CE_15CES_S02_b_0.88_870.7526
2015_5_31_053_HILIC_IDA_45CE_15CES_S02_b_0.88_872.7688
2015_5_31_053_HILIC_IDA_45CE_15CES_S02_b_0.88_874.7831
2015_5_31_053_HILIC_IDA_45CE_15CES_S02_b_0.88_894.7505
2015_5_31_053_HILIC_IDA_45CE_15CES_S02_b_0.88_920.7692
2015_5_31_053_HILIC_IDA_45CE_15CES_S02_b_0.88_922.7824
2015_5_31_053_HILIC_IDA_45CE_15CES_S02_b_0.88_944.7654
2015_5_31_053_HILIC_IDA_45CE_15CES_S02_b_0.91_669.5577
2015_5_31_053_HILIC_IDA_45CE_15CES_S02_b_1.23_195.0881
2015_5_31_053_HILIC_IDA_45CE_15CES_S02_b_1.38_326.0865
2015_5_31_053_HILIC_IDA_45CE_15CES_S02_b_1.66_308.1721
2015_5_31_053_HILIC_IDA_45CE_15CES_S02_b_1.69_181.0702
2015_5_31_053_HILIC_IDA_45CE_15CES_S02_b_2.25_235.1806
2015_5_31_053_HILIC_IDA_45CE_15CES_S02_b_2.9_61.041
2015_5_31_053_HILIC_IDA_45CE_15CES_S02_b_3.48_247.1798
2015_5_31_053_HILIC_IDA_45CE_15CES_S02_b_3.87_428.3738
2015_5_31_053_HILIC_IDA_45CE_15CES_S02_b_3.97_424.342
2015_5_31_053_HILIC_IDA_45CE_15CES_S02_b_3.9_426.3555
2015_5_31_053_HILIC_IDA_45CE_15CES_S02_b_4.01_400.3391
2015_5_31_053_HILIC_IDA_45CE_15CES_S02_b_4.02_398.3261
2015_5_31_053_HILIC_IDA_45CE_15CES_S02_b_4.07_396.3101
2015_5_31_053_HILIC_IDA_45CE_15CES_S02_b_4.11_372.3103
2015_5_31_053_HILIC_IDA_45CE_15CES_S02_b_4.32_754.5361
2015_5_31_053_HILIC_IDA_45CE_15CES_S02_b_4.44_328.248
2015_5_31_053_HILIC_IDA_45CE_15CES_S02_b_4.44_764.5252
2015_5_31_053_HILIC_IDA_45CE_15CES_S02_b_4.49_742.5392

File information

Name: Zolpidem [M+H]⁺; C19H21N3O; ZAFYA
Scan number: 186
Retention time [min]: 1.66
Precursor m/z [Da]: 308.1721
Precursor type: [M+H]⁺
Ion mode: Positive
Spectrum type: Centroid
Collision energy: 40
Formula:
SMILES:
Intensity: 10236
MS1 peak number: 207
MS2 peak number: 21

Molecular formula finder

Formula	Error [mDa]	Score	Resource	Select
C19H21N3O	3.6387	3.8840	BMDB, ChEBI, DrugBank, F	<input checked="" type="checkbox"/>
C17H25NO2S	-4.2235	3.5330	UNPD, MINE	<input checked="" type="checkbox"/>
C20H21NO2	-7.5946	3.3980	UNPD, MINE, STOFF	<input checked="" type="checkbox"/>
C12H25N3O6	9.5120	3.2180	UNPD, MINE	<input checked="" type="checkbox"/>
C14H21N5O3	-0.3840	3.0520	MINE	<input checked="" type="checkbox"/>
C13H25NO7	-1.7214	2.8270		<input type="checkbox"/>
C12H26N3O4P	1.2694	2.8020		<input type="checkbox"/>
C14H29NO2S2	-0.8524	2.7870		<input type="checkbox"/>
C16H25N3OS	7.0099	2.7860	MINE	<input type="checkbox"/>
C11H25N5O3S	2.9871	2.7550		<input type="checkbox"/>
C8H22N9O2P	-1.4160	2.7510		<input type="checkbox"/>
C17H26NO2P	5.2922	2.7280		<input type="checkbox"/>
C10H17N11O	-3.0694	2.7130		<input type="checkbox"/>
C7H13N11O6	2.3017	2.7000		<input type="checkbox"/>

Isotopic ions

C19H21N3O

Relative Abundance

m/z

MS/MS spectrum

C4H10N
72.0698 from 307.1934 to 235.1237
Dialkylamines

Relative Abundance

m/z

Precursor m/z

Structure finder

Name	Total score	InChIKey
HMDB05023	6.2323	ZAFYATHCZYHLPB-UHFFFAOYSA-N
HMDB15670	5.7611	MWTKBTRZPHJQLH-UHFFFAOYSA-N

Spectrum

Structure

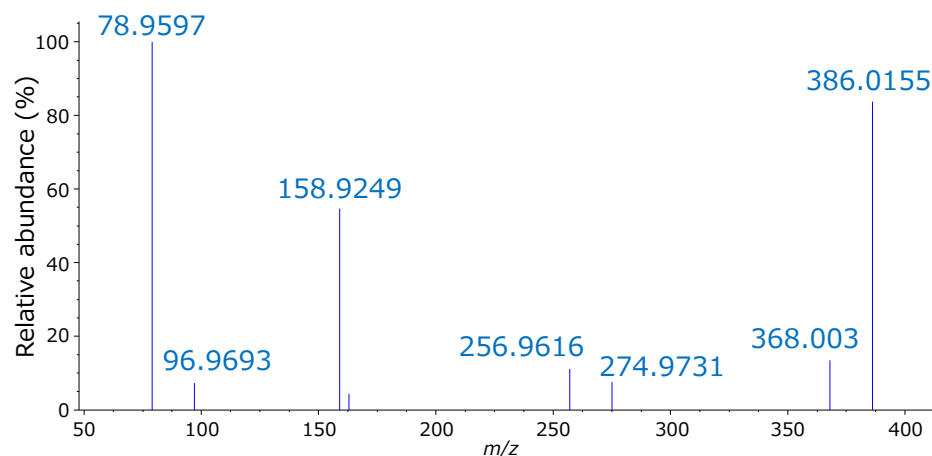
Meta data

話の流れ

1. 組成式の決定方法
2. データベースの活用
3. MS/MSに対する部分構造アサインメント
4. 構造候補の検証

組成式候補の出力

1. Negative, <10ppm, Human



m/z 386.0155 \pm 0.005 as [M-H]⁻



C: 12.00000000	O: 15.9949146	F: 18.99840322
H: 1.00782503	S: 31.9720710	Cl: 34.96885268
N: 14.0030740	P: 30.9737616	Br: 78.9183371
e-: 0.00054857		I: 126.904473

組成式の絞り込み：ルールを活用

✓原子価測の確認

例) $C_8H_7NO_4S$

TA: $8+7+1+4+1 = 21$
OV: $7+1 = 8$
SV: $32+7+3+8+6=56$

元素の総数をTA
原子価が奇数の元素の総数をOV
原子価の総数をSVとしたとき,

$OV * SV = \text{even}$
 $SV \geq 2 * (TA - 1)$

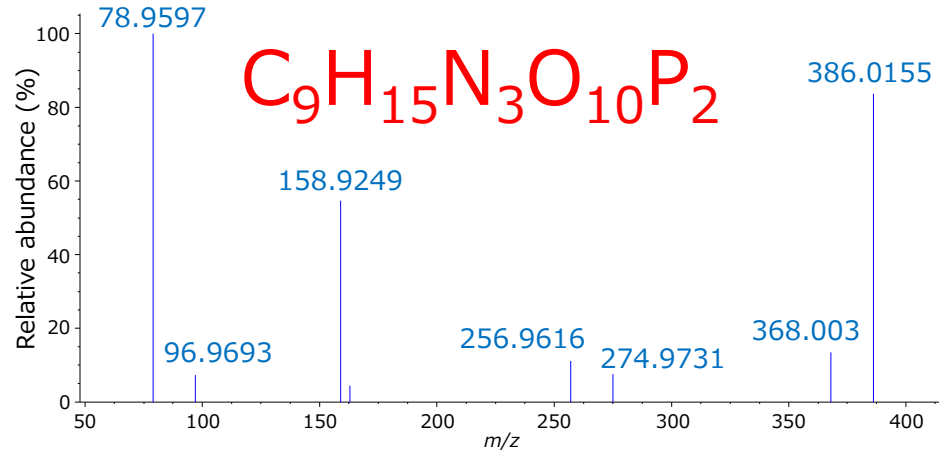
ただし, N(3), O(2), P(5), S(6)

✓経験的閾値

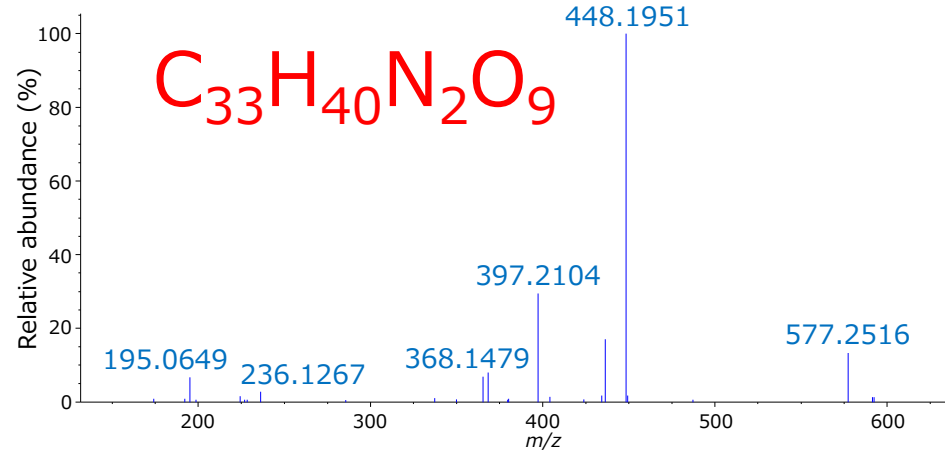
	Mass	H/C	N/C	O/C	P/C	S/C
Max	1999.90	8.00	4.00	10.00	3.00	6.00
Min	50.02	0.00	0.00	0.00	0.00	0.00
Mean	601.03	1.38	0.10	0.32	0.01	0.02
Stdev	349.29	0.43	0.15	0.27	0.04	0.09
Median	512.32	1.39	0.04	0.27	0.00	0.00
Percentile 0.15%	72.05	0.00	0.00	0.00	0.00	0.00
Percentile 0.005%	52.02	0.00	0.00	0.00	0.00	0.00
Percentile 99.85%	1939.74	3.33	1.20	2.20	0.40	1.00
Percentile 99.995%	1998.10	6.00	4.00	6.00	1.83	3.00

組成式の推定・決定 by Computational MS

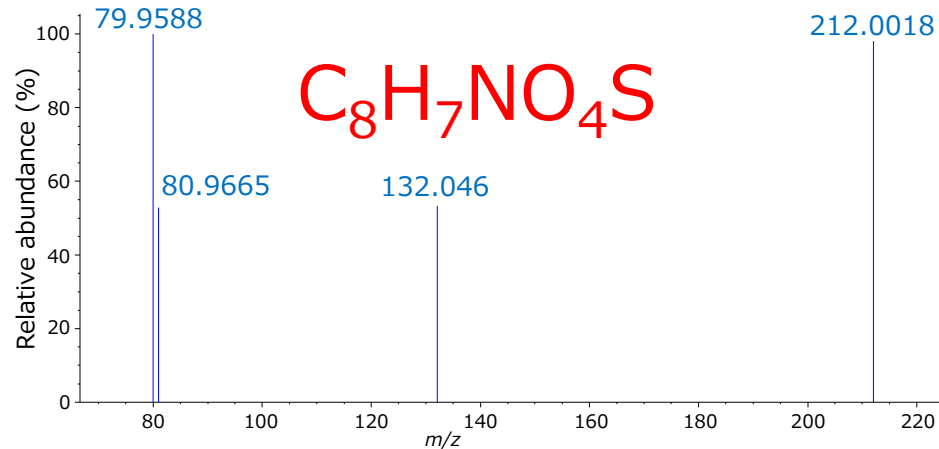
1. Negative, <10ppm, Human



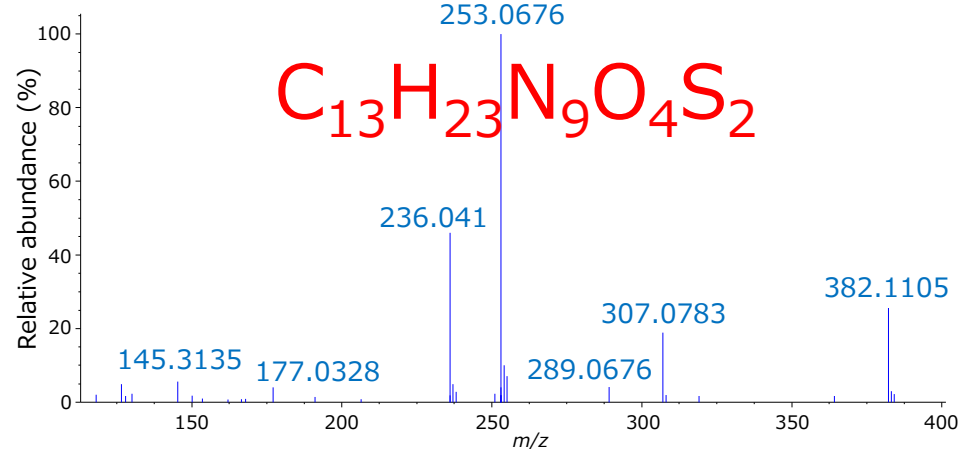
2. Positive, <3ppm, Drug



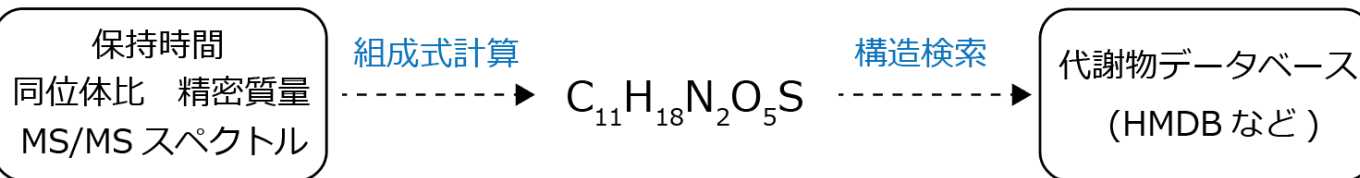
3. Negative, <10ppm, Human



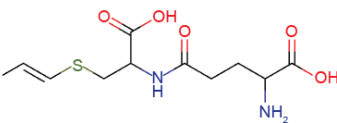
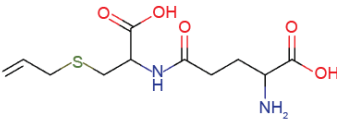
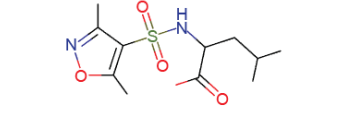
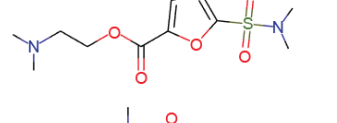
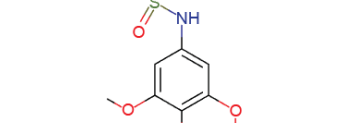
4. Positive, <3ppm, Plant (Onion)



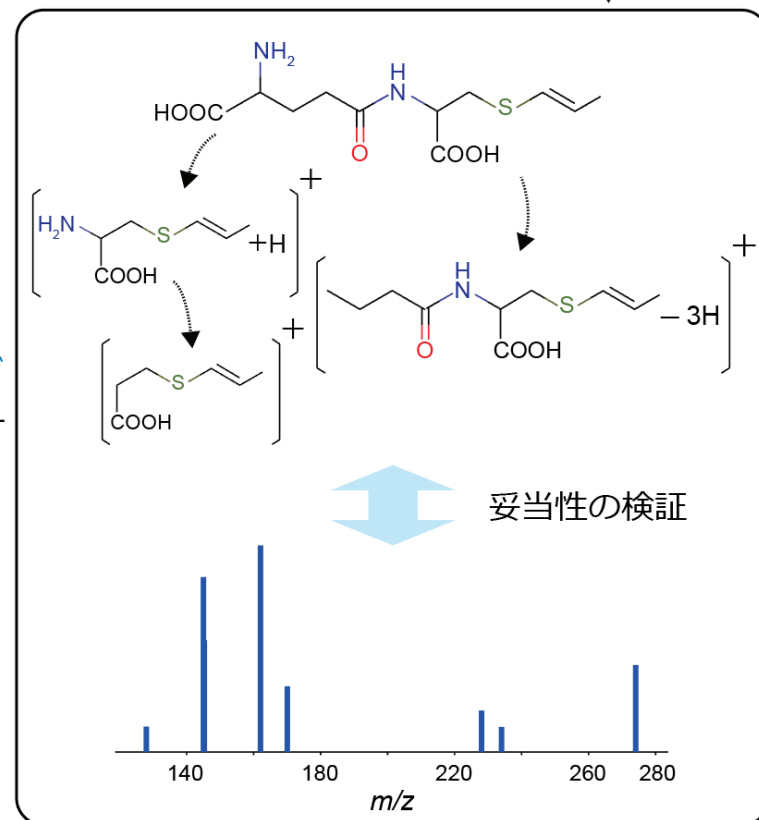
Computational MSとデータベース



In silico フラグメントと
実測 MS/MS の比較

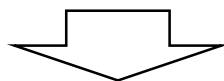
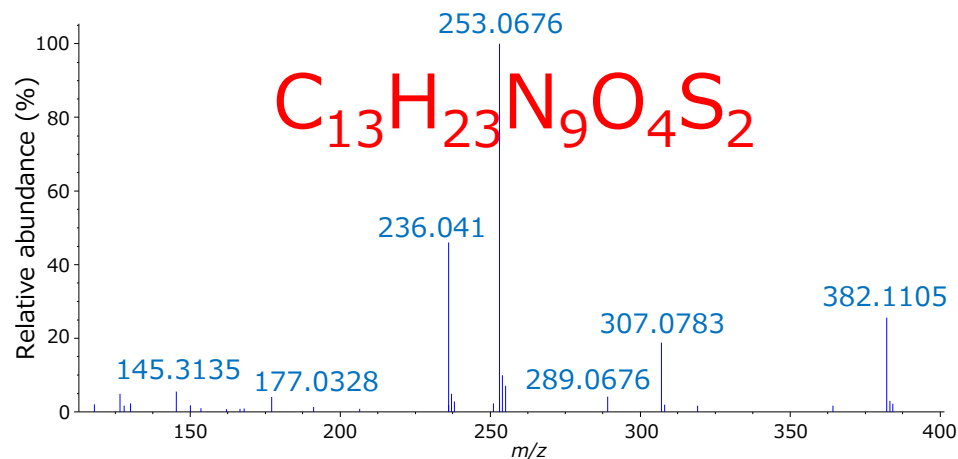
Rank	Candidate
1	
2	
3	
4	
5	

ランキング



組成式に基づく化合物構造の検索

4. Positive, <3ppm, Plant (Onion)



KNaPSAcK (45,852) UNPD (166,995) PlantCyc (3,817)
FooDB (21,943) NANPDB (3,882) **STOFF-IDENT (10,231)**
LipidMAPS (34,466) ChEBI (53,746) DrugBank (6,273)
SMPDB (1,490) YMDB (1,840) ECMDB (1,342)
T3DB (2,499) BMDB (7,232) HMDB (95,924) Urine (3,929)
Saliva (1,129) Fecal (1,073) CSF (360) Serum (23,138)

STOFF-IDENT



STOFF-IDENT Database

Responsible authors: Sylvia Grosse, Thomas Letzel

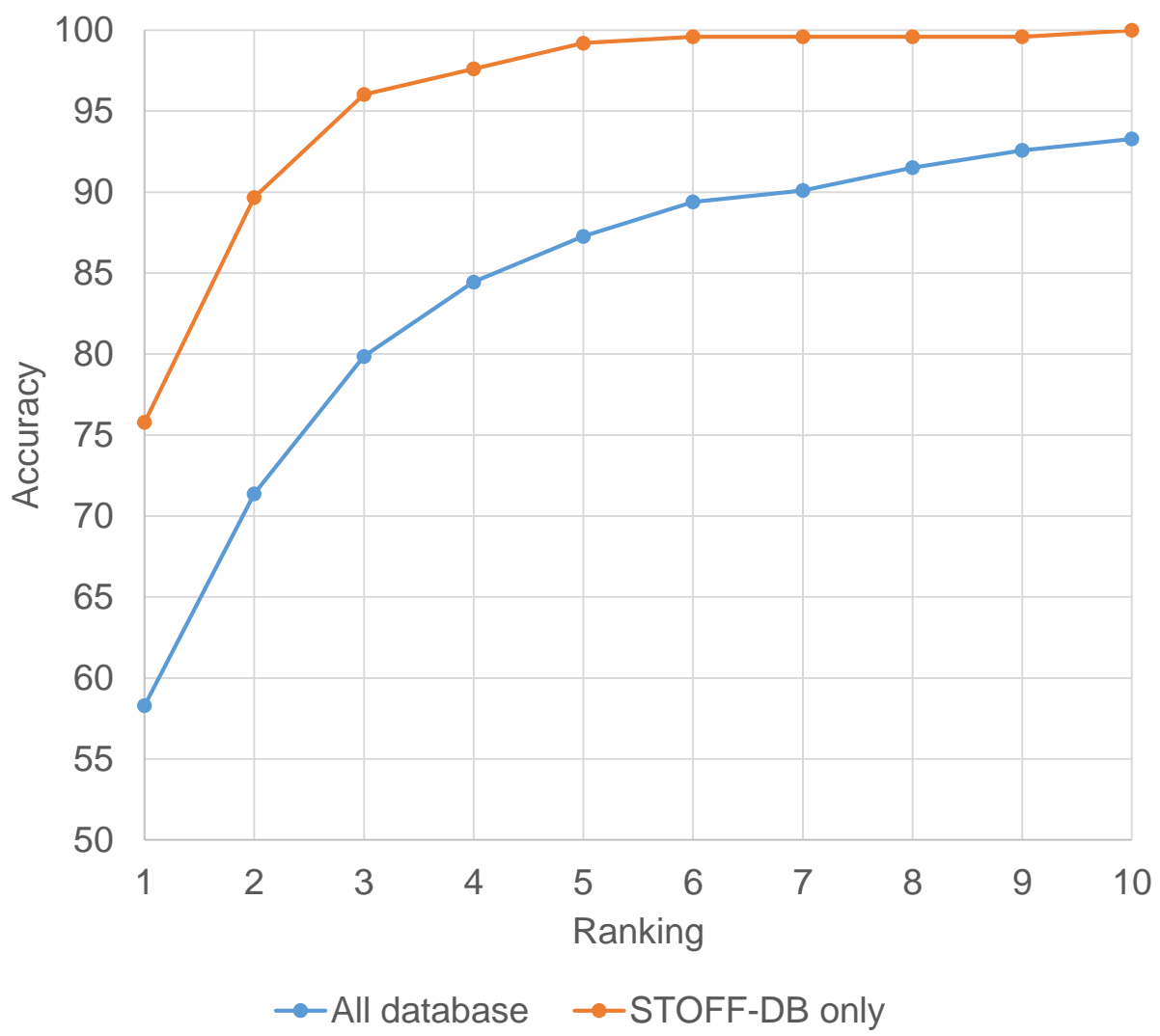
In cooperation with the following institutes: LfU (Marion Letzel, Manfred Sengl), LW (Thomas Lucke, Wolfgang Schulz), HSWT (Frank Lesske, Marco Luthardt, Tobias Placht)

Cite as: Grosse S., Letzel T.: User Manual for STOFF-IDENT Database, 2016 (4.2), 1-33

Primary the compound database 'STOFF-IDENT' is used for identification purposes, more exactly for non-target screening ('Hidden Targets' and/or 'Known Unknown') and suspected-target screening analysis in different parts of water research.¹ This is why just water relevant organic molecules, their transformation products and metabolites that occur in the environment are listed.²

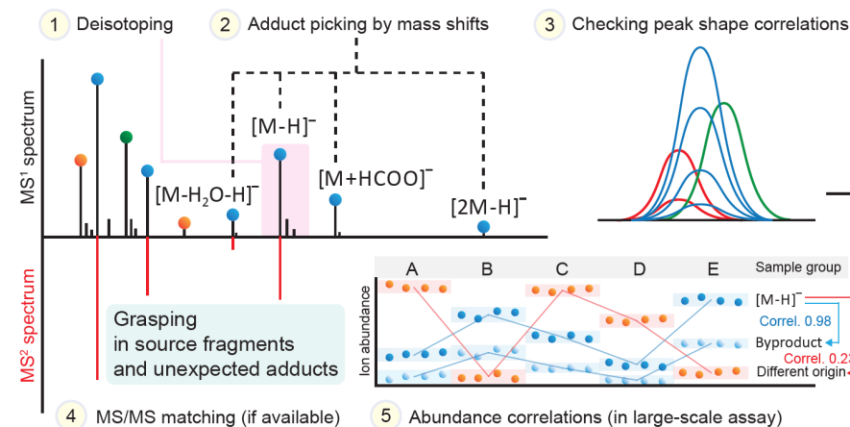
Besides common compound parameters as Name, CAS Number, Formula, InChi key, IUPAC Name, SMILES Code, monoisotopic mass, also physicochemical indications as the logP and the logD value for four different pH values are included.²

MS-FINDERに考えさせた結果の比較

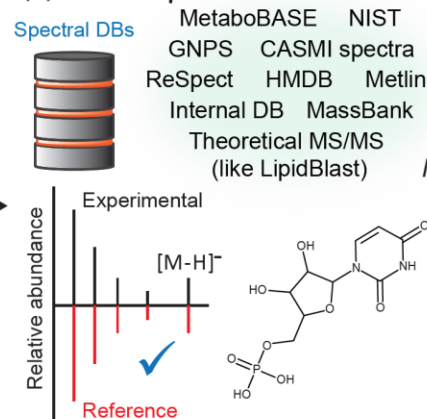


化合物同定に必要な行程まとめ

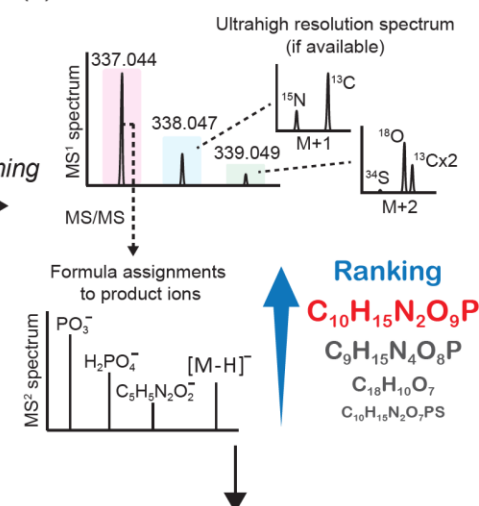
(a) Eliminate false positive peaks



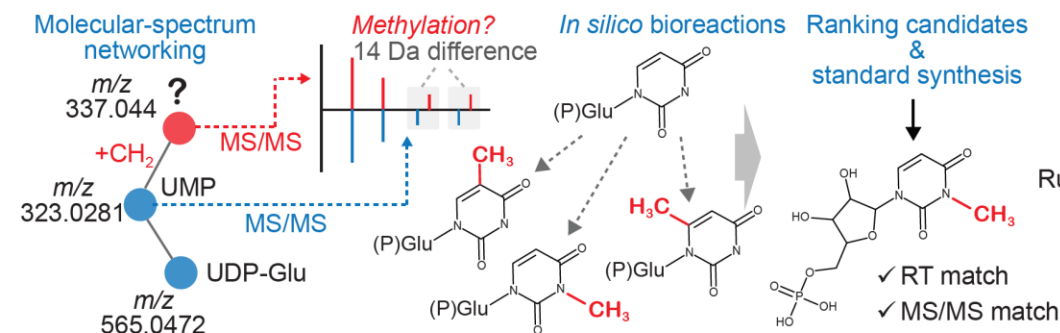
(b) Search spectral libraries



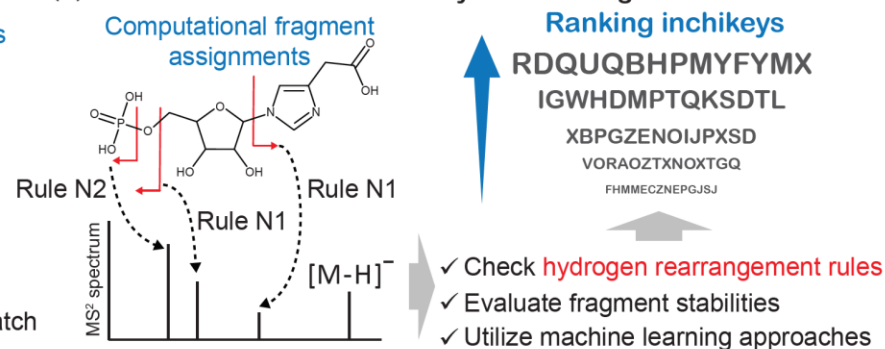
(c) Predict molecular formula



(e) Expand the chemical spaces & Predict the molecular scaffold



(d) Retrieve structures followed by their ranking





RIKEN PRIME website

RIKEN PRIME - Google S

← → ↻ <https://www.google.co.jp/webhp?sourceid=chrome-instant&ion=1&espv=2&ie=UTF-8#q=RIKEN+PRIME>

Google

RIKEN PRIME



All

Images

News

Shopping

Videos

More ▾

Search tools

About 353,000 results (0.36 seconds)

Platform for RIKEN Metabolomics: PRIME

prime.psc.riken.jp/









About PRIME. PRIME, the Platform for RIKEN Metabolomics, is a Web-based service for metabolomics

Computational metabolomics

← → × <file:///C:/Users/hiroshi.tsgawa/Dropbox/PRIME%20homepage/CompMetWeb/index.html>

HOME ABF CONVERTER MS-DIAL MS-FINDER MRMPROBS DATABASES SMETSEARCH AIOUTPUT MASSBANK TO MSP LOWESS NORMALIZATION STATISTICS DERIVATIZER CONTACT

Software tools & Databases

 <p>ABF converter</p> <p>ABF binary format file can be used in MS-DIAL and MRMPROBS, and the converter is freely available at Reifycs Converter Website. The supported vendor's- and common MS format is listed in their homepage.</p>	 <p>MS-DIAL</p> <p>MS-DIAL is a universal program for untargeted metabolomics and lipidomics supporting any type of chromatography/mass spectrometry methods (GC/MS, GC-MS/MS, LC/MS, and LC-MS/MS etc.).</p>	 <p>MS-FINDER</p> <p>MS-FINDER is a universal program for compound annotation supporting EI-MS (GC/MS) and ESI-MS/MS spectra. It aims to provide solutions for 1) spectral searching, 2) formula prediction, and 3) structure elucidation for unknown spectra.</p>	 <p>EI-MS and MS/MS library</p> <p>We released several MSP files including both EI- and MS/MS spectra as a 'start-up kit'. Nowadays, MS-DIAL/MS-FINDER internally have a version of Fiehn lab's GC/MS database, and in silico RT- and MS/MS database for lipidomics.</p>
 <p>MassBank to MSP</p> <p>This program helps you to prepare NIST MSP format libraries applicable to a lot of programs (MS-DIAL, NIST MS Search, AMDIS) from MassBank records.</p>	 <p>MRMPROBS/MRMDIFF</p> <p>MRMPROBS is launched as a universal program for targeted metabolomics supporting not only multiple reaction monitoring (MRM) but also SCAN and data independent MS/MS acquisition (DIA) data.</p>	 <p>MRM database</p> <p>Several conditions/reference libraries for large scaled MRM assay are available.</p>	 <p>Smetsearch</p> <p>The purpose of this project is to develop the methodology for formula 'identification' by ultrahigh resolution MS instrument in combination with labeled experiments.</p>
