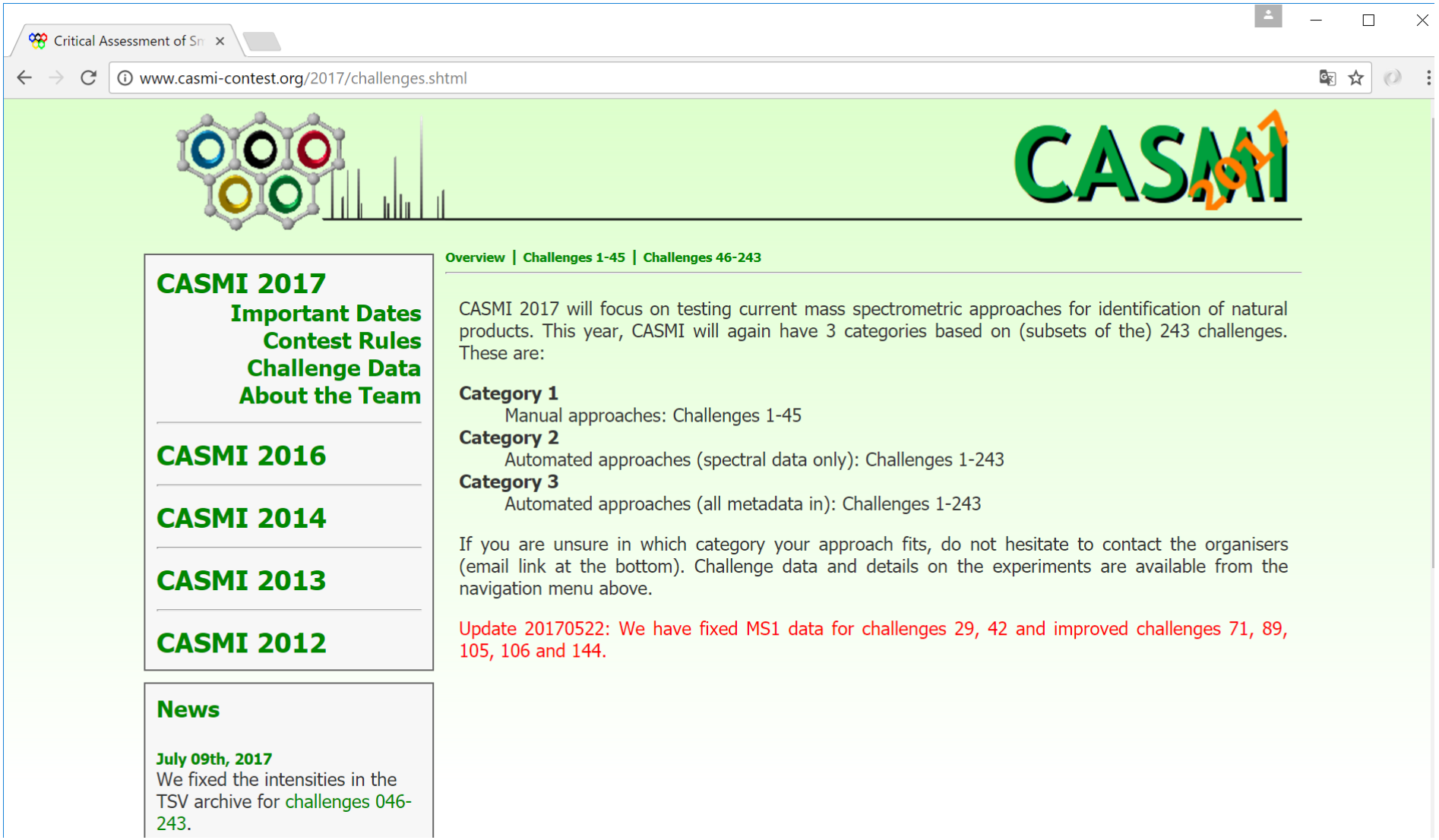
CASMI2016から学ぶ化合物構造推定

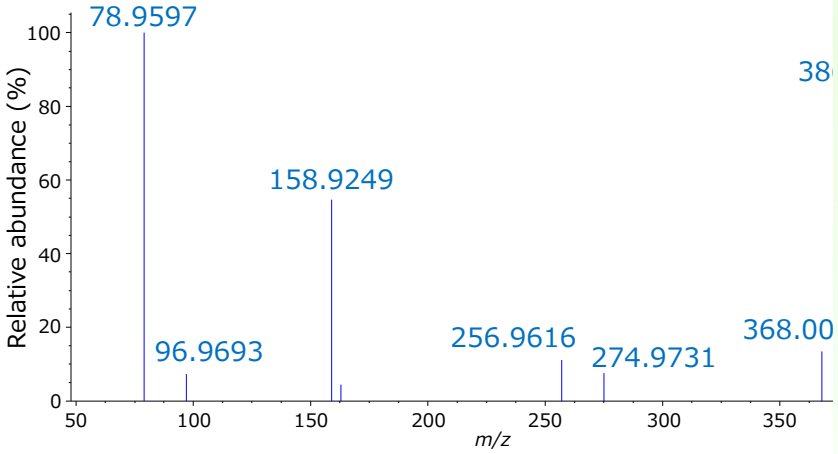
理研CSRS（リーダー・有田正規）・理研IMS（リーダー・有田誠）
津川 裕司

CASMI contest

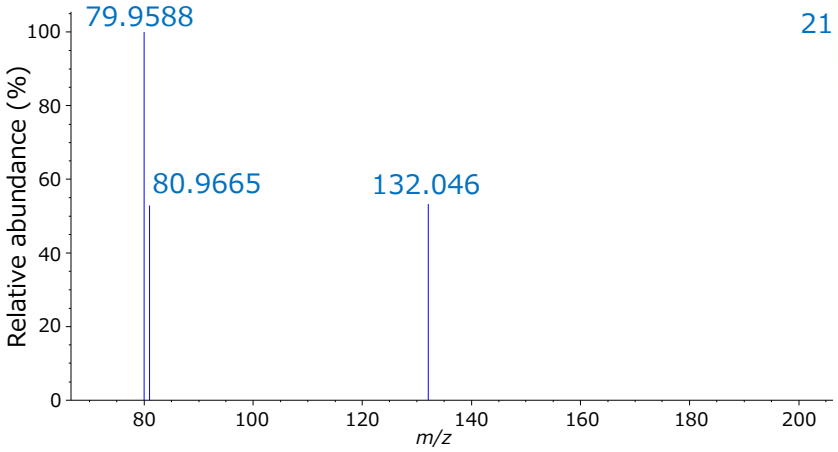


このMS/MSスペクトルの元構造は何でしょう？

1．Negative, <10ppm, Human



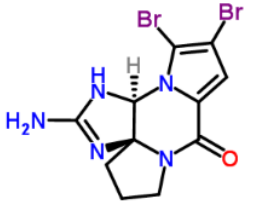
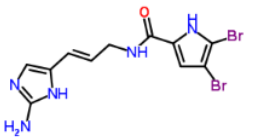
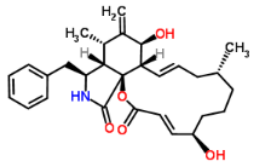
3．Negative, <10ppm, Human



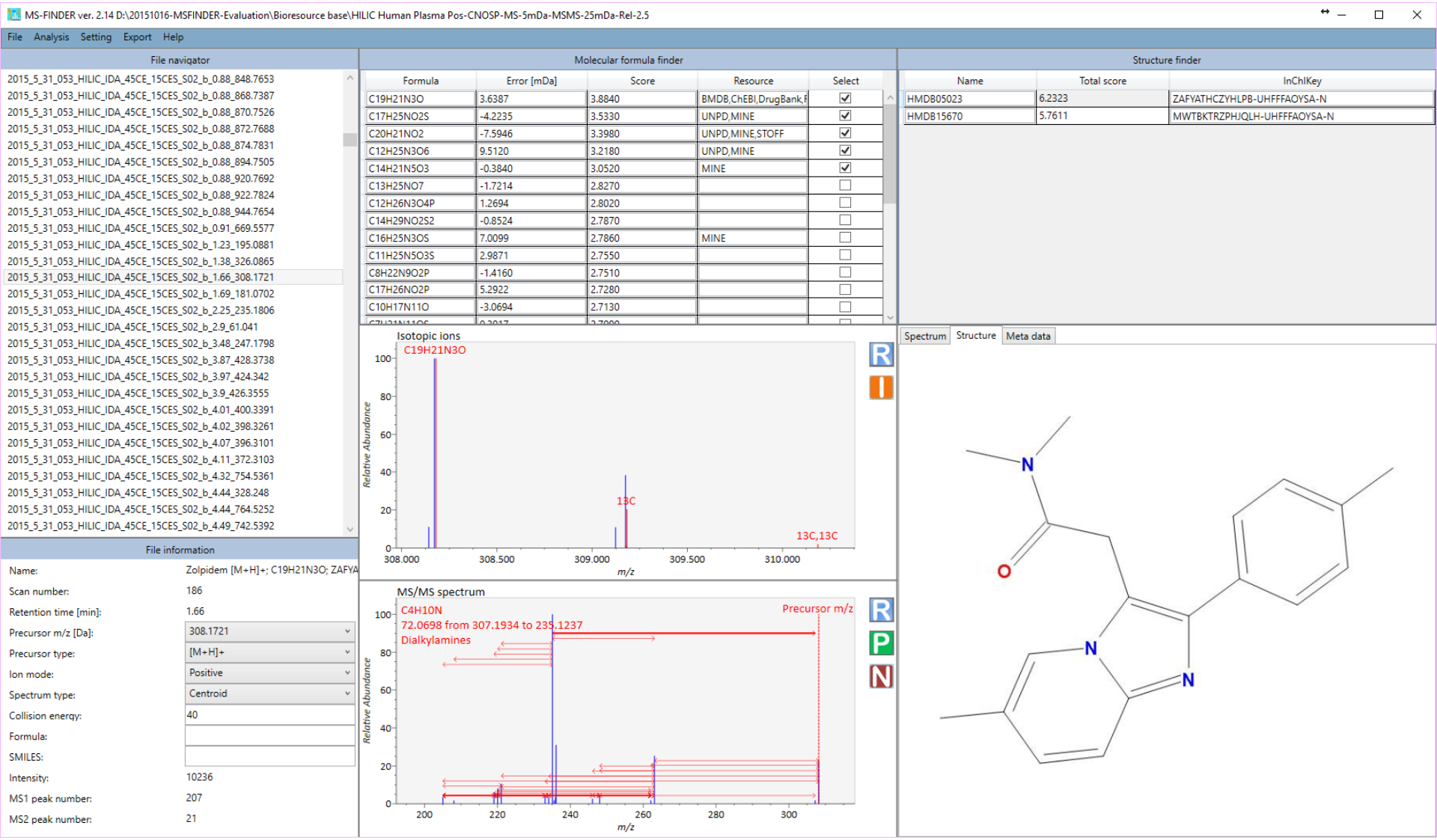
Challenge Solutions Category 1

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 $C_{11}H_{11}Br_2N_5O$ PubChem: 42636938 ChemSpider: 9592765	
Challenge-002	Oroidin $C_{11}H_{11}Br_2N_5O$ 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



話の流れ

1. 組成式の決定方法

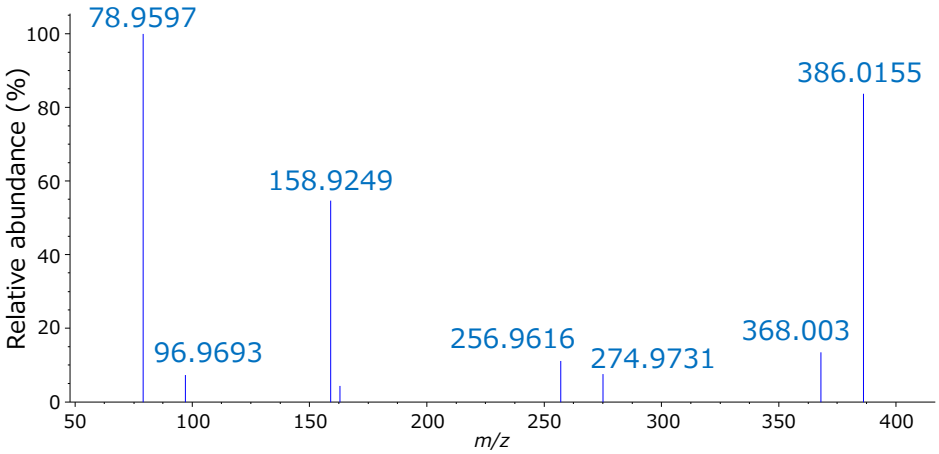
2. データベースの活用

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±0.005 as $[M-H]^-$

$C_xH_yN_zO_uS_vP_wF_aCl_bBr_cI_d$

C: 12.0000000　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：ルールの活用

✓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 = even
SV >= 2* (TA - 1)

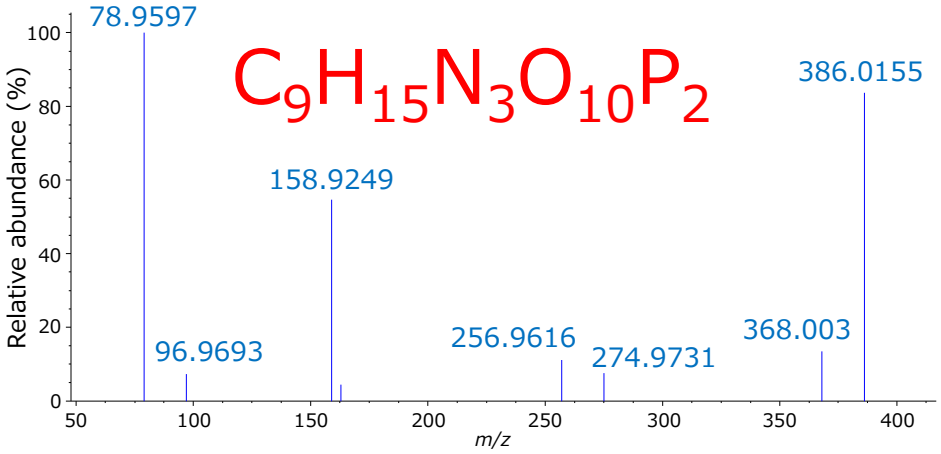
ただし, N(3), O(2), P(5), S(6)

✓経験的閾値

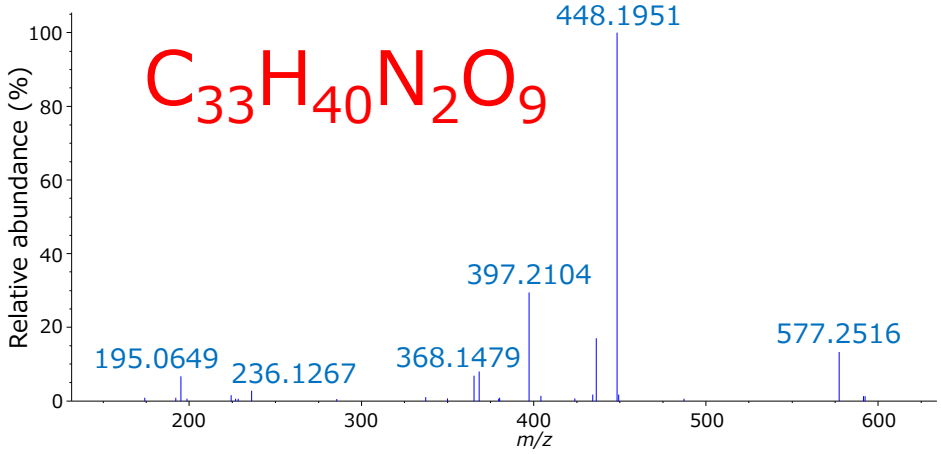
	Mass	H/C	N/C	O/C	P/C	S/C
Max	1999.90	8.00	4.00	10.00	3.00	6.00
Min	50.02	0.00	0.00	0.00	0.00	0.00
Mean	601.03	1.38	0.10	0.32	0.01	0.02
Stdev	349.29	0.43	0.15	0.27	0.04	0.09
Median	512.32	1.39	0.04	0.27	0.00	0.00
Percentile 0.15%	72.05	0.00	0.00	0.00	0.00	0.00
Percentile 0.005%	52.02	0.00	0.00	0.00	0.00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

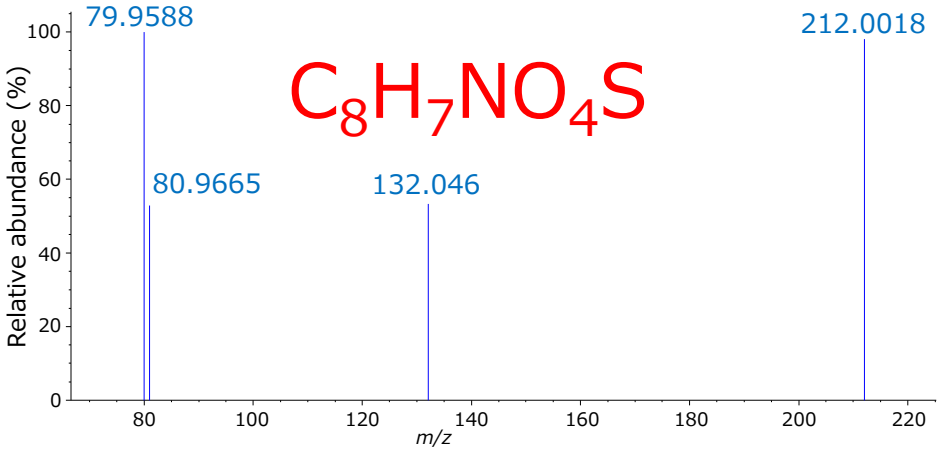
1．Negative, <10ppm, Human



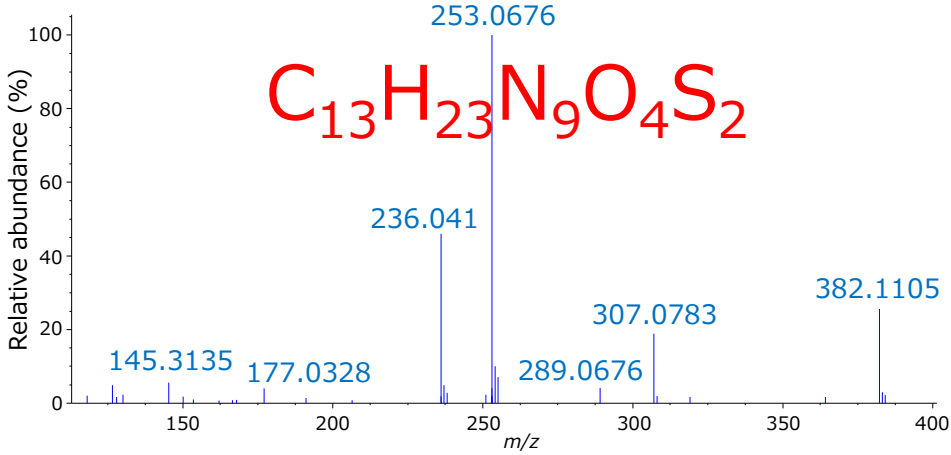
2．Positive, <3ppm, Drug



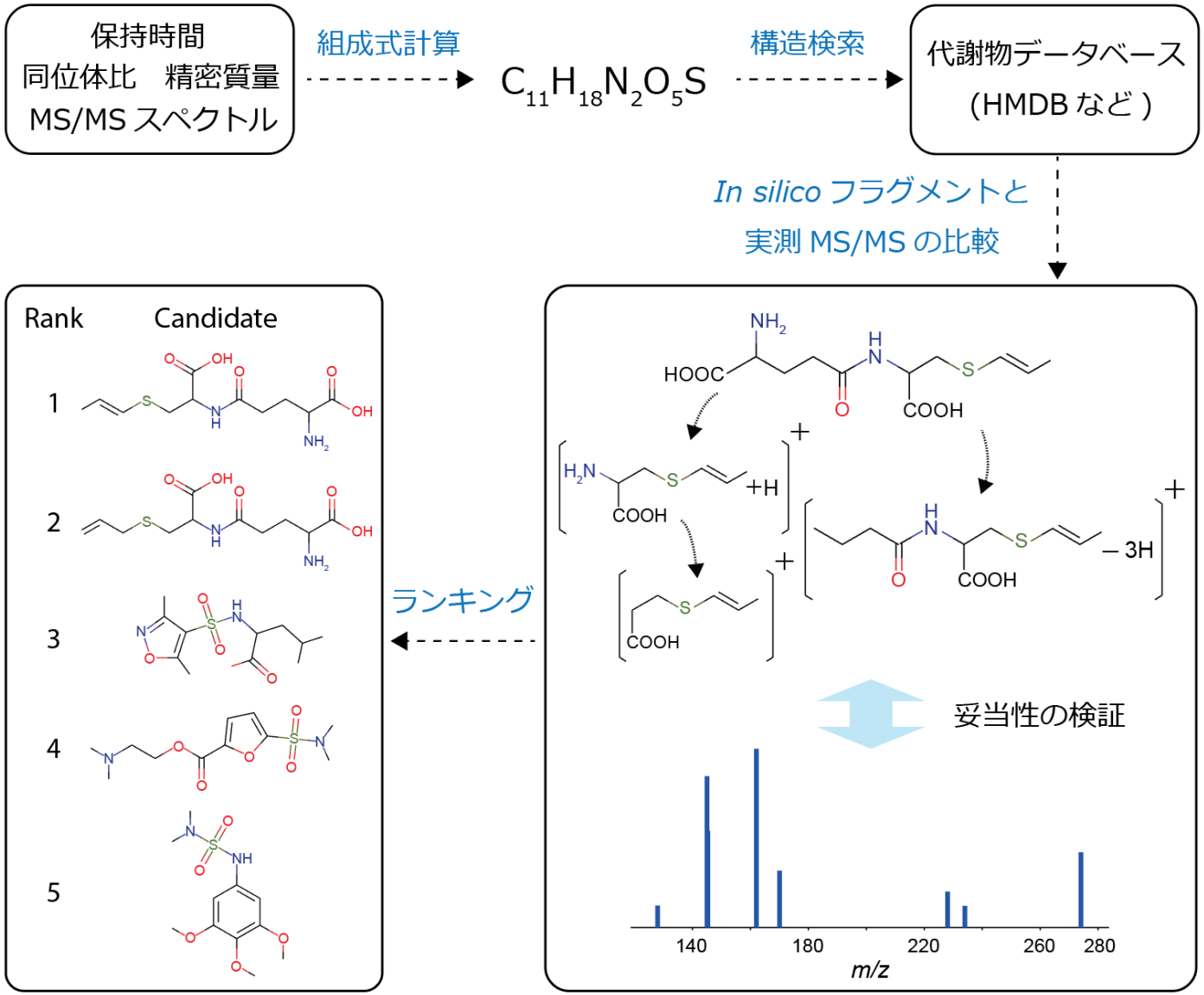
3．Negative, <10ppm, Human



4．Positive, <3ppm, Plant (Onion)

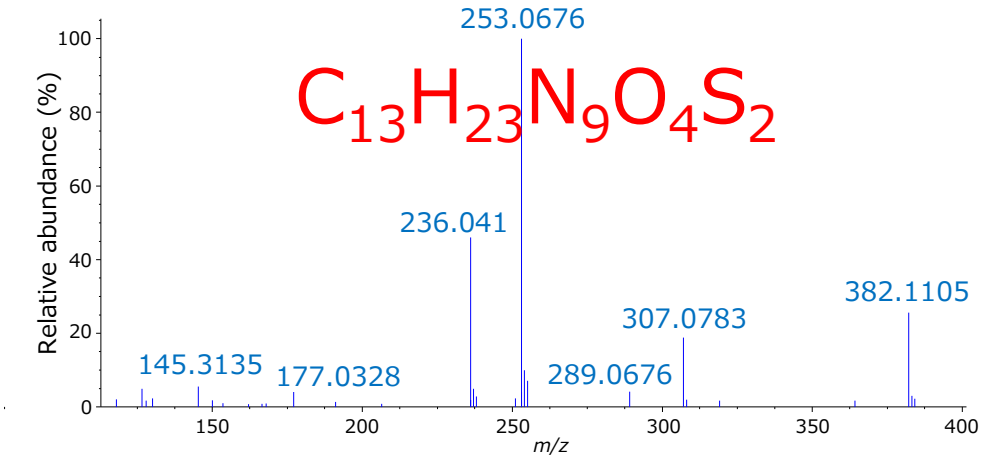


Computational MSとデータベース



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

?TOFF !DENT

STOFF-IDENT Database

Responsible authors: Sylvia Grosse, Thomas Letzel

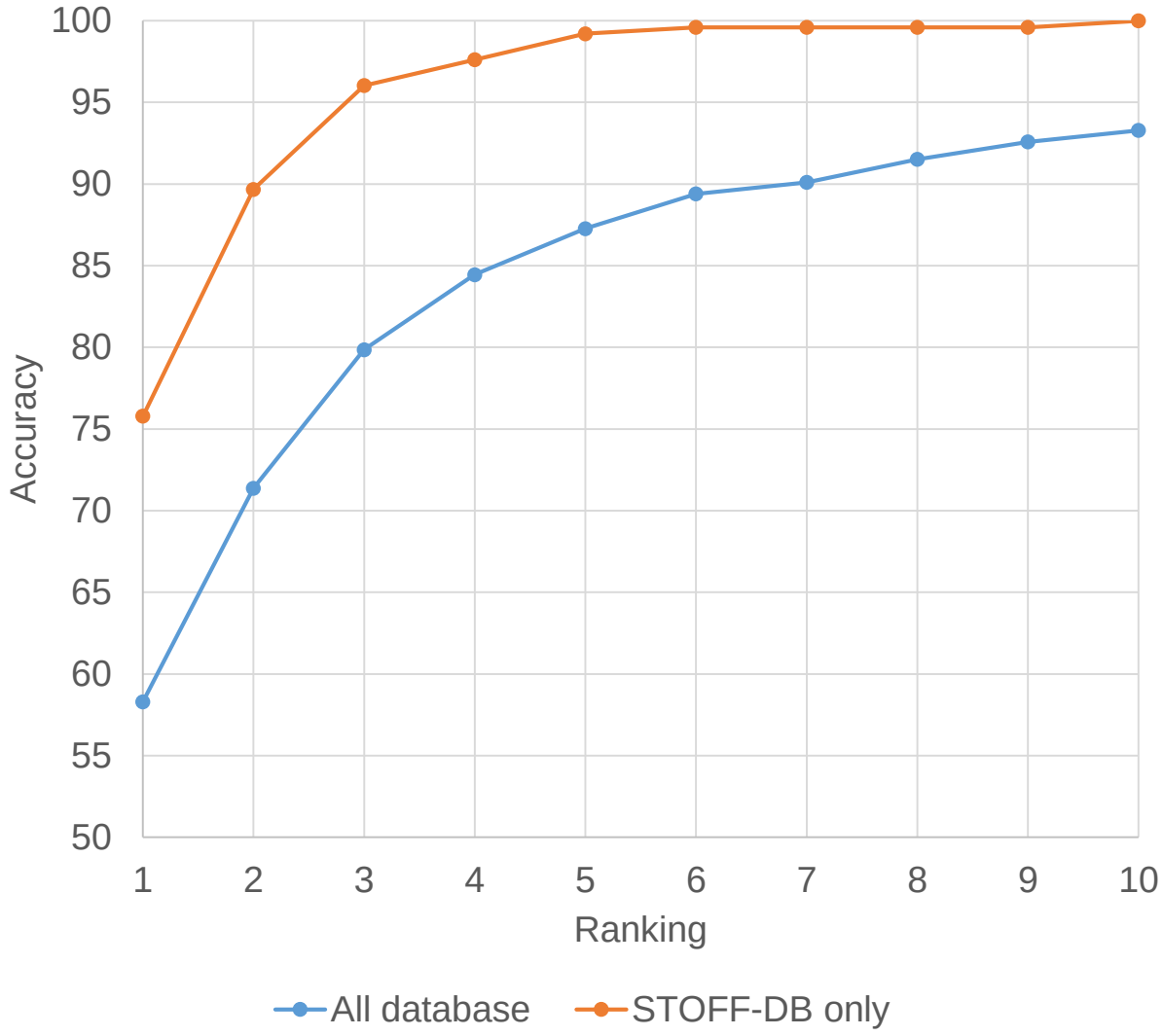
In cooperation with the following institutes: LfU (Marion Letzel, Manfred Sengl), LW (Thomas Lucke, Wolfgang Schulz), HSWT (Frank Lesske, Marco Luthardt, Tobias Placht)

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

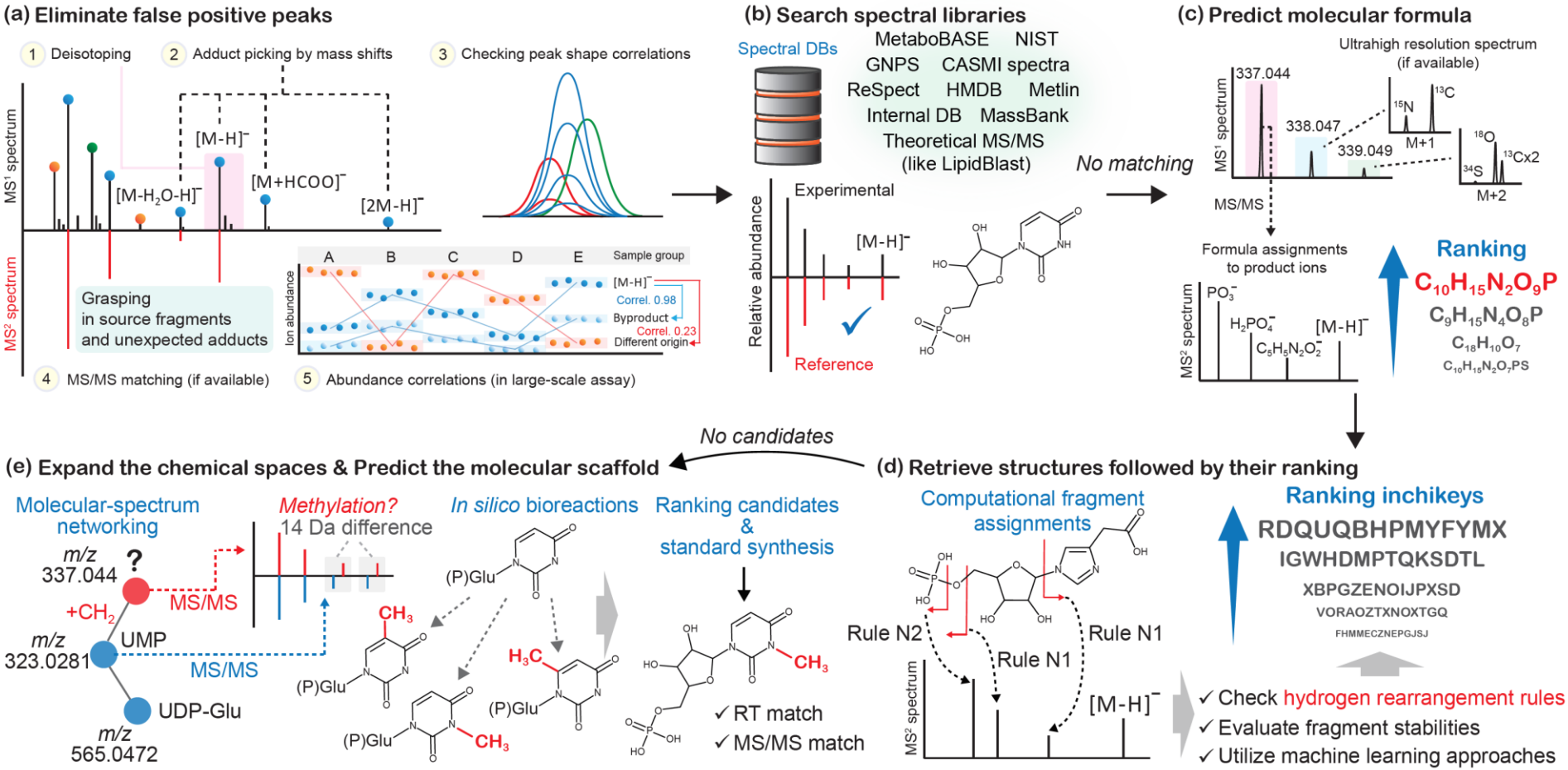
Primary the compound database 'STOFF-IDENT' is used for identification purposes, more exactly for non-target screening ('Hidden Targets' and/or 'Known Unknown') and suspected-target screening analysis in different parts of water research.[1] This is why just water relevant organic molecules, their transformation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



RIKEN PRIMe website