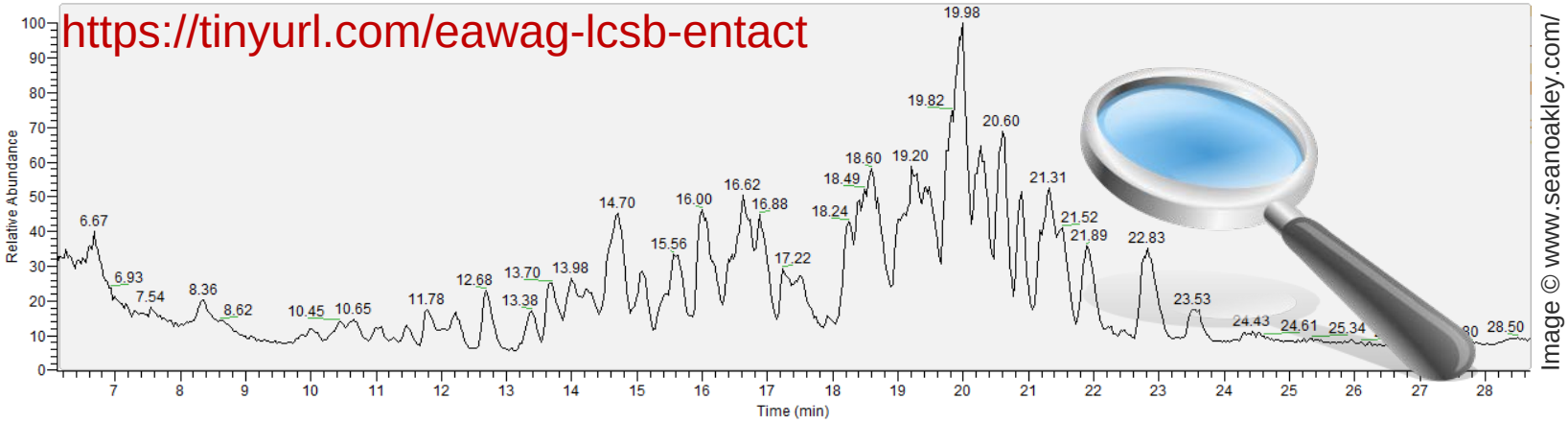
Integrating Eawag, LCSB, MetFrag and CompTox Efforts in ENTACT



Emma L. Schymanski
FNR ATTRACT Fellow and PI in Environmental Cheminformatics
Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg
Email: emma.schymanski@uni.lu

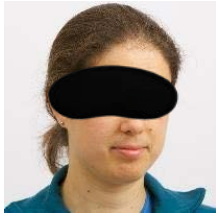
Birgit Beck, Philipp Longrée, Heinz P. Singer, Juliane Hollender (Eawag)
Christoph Ruttkies (IPB Halle), Andrew D. McEachran, Alex Chao, Christopher M. Grulke,
Jon R. Sobus and Antony J. Williams (US EPA)

ENTACT Meeting, Research Triangle Park, NC, August 13-15, 2018

Outline

- Part 1: Eawag Blinded

TraceFinder™
The Power to Know



RMassBank

MetFrag

Chemistry Dashboard | TOXCAST

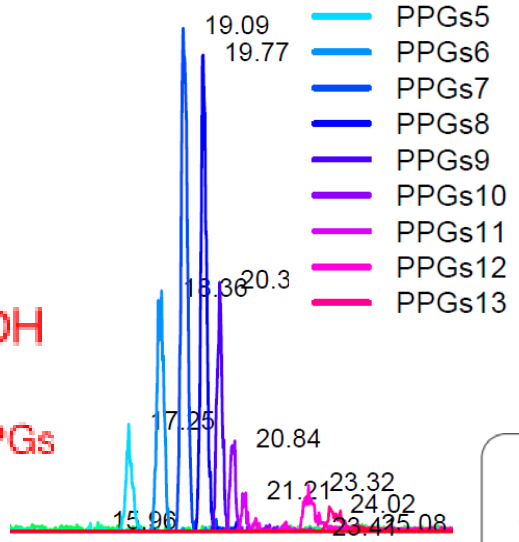
- Part 2: Unblinded

RMassBank

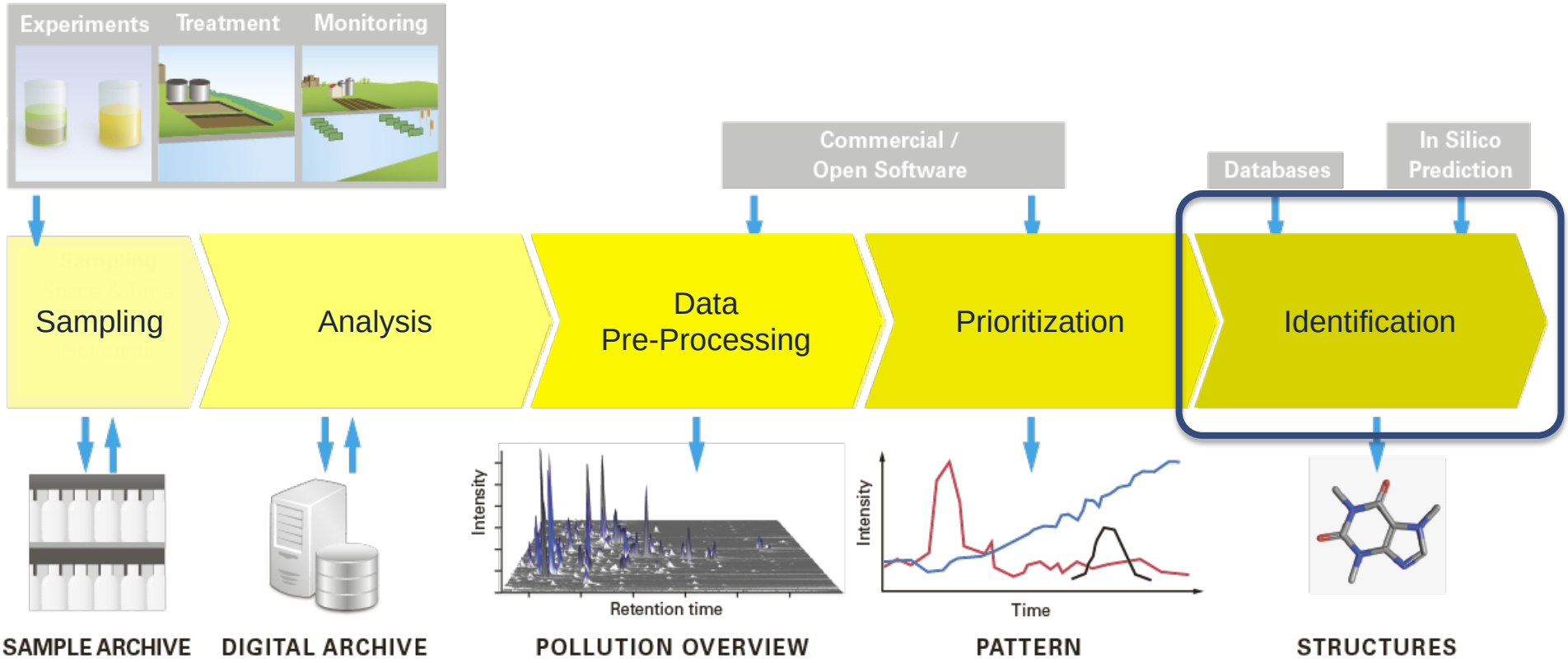
MetFrag

- Part 3: The Three "M"s
  - Making MetFrag MS-Ready
  - Metals or no Metals?
  - Mixtures…

DTXSID9027863 - PPGs

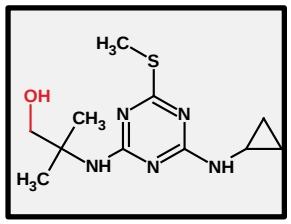
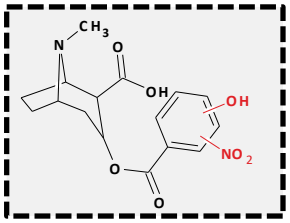


Non-target Screening … and Identification



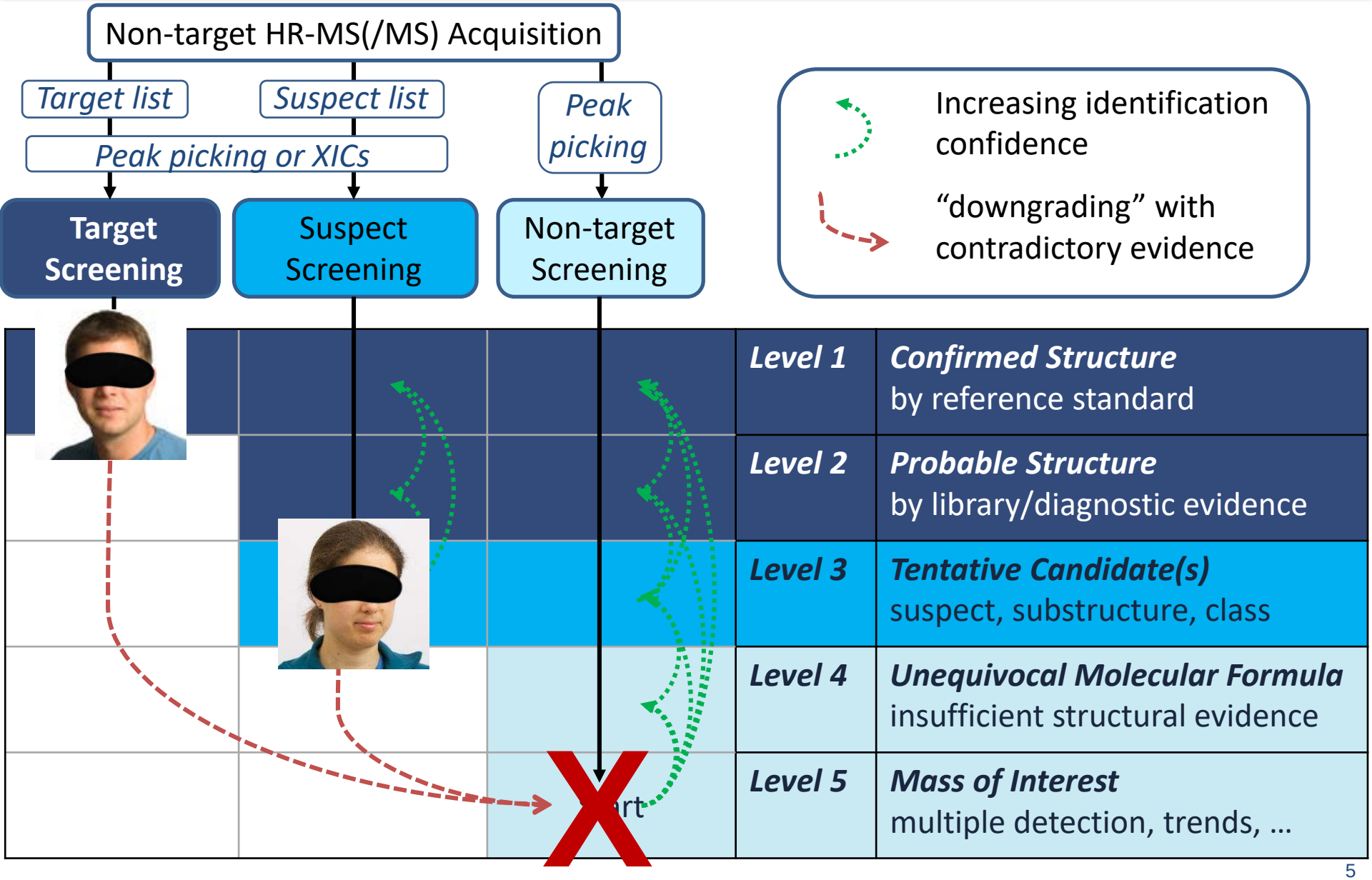
Mod. from Hollender, Schymanski, Singer & Ferguson, 2018, ES&T Feature, 51:20, 11505-11512. DOI: 10.1021/acs.est.7b02184

Identification Approaches and Terminology I

Example	*Identification confidence*	*Minimum data requirements*
	***Level 1: Confirmed structure*** by reference standard	MS, MS^2, RT, Reference Std.
	Level 2: Probable structure a) by library spectrum match b) by diagnostic evidence	MS, MS^2, Library MS^2 MS, MS^2, Exp. data
	***Level 3: Tentative candidate(s)*** structure, substituent, class	MS, MS^2, Exp. data
$C_6H_5N_3O_4$	***Level 4: Unequivocal molecular formula***	MS isotope/adduct
192.0757	***Level 5: Exact mass*** of interest	MS

Schymanski, Jeon, Gulde, Fenner, Ruff, Singer & Hollender (2014) ES&T, DOI: 10.1021/es5002105

Identification Approaches and Terminology II



Schymanski *et al*, 2014, ES&T. DOI: 10.1021/es5002105 & Schymanski *et al*. 2015, ABC, DOI: 10.1007/s00216-015-8681-7

Part 1: Eawag & LCSB ENTACT - Blinded

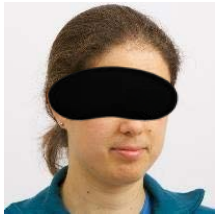
SampleName	FileName_DIA_pos	FileName_DIA_neg	FileName_DD_pos	FileName_DD_neg
Blank_Nanopure	20170106_DIA_pos_01	20170106_DIA_neg_01	20170106_dd_pos_01	20170106_dd_neg_01
11_DMSO_Blank	20170106_DIA_pos_02	20170106_DIA_neg_02	20170106_dd_pos_02	20170106_dd_neg_02
01_BF00173499	20170106_DIA_pos_03	20170106_DIA_neg_03	20170106_dd_pos_03	20170106_dd_neg_03
02_BF00173500	20170106_DIA_pos_04	20170106_DIA_neg_04	20170106_dd_pos_04	20170106_dd_neg_04
03_BF00173501	20170106_DIA_pos_05	20170106_DIA_neg_05	20170106_dd_pos_05	20170106_dd_neg_05
04_BF00173502	20170106_DIA_pos_06	20170106_DIA_neg_06	20170106_dd_pos_06	20170106_dd_neg_06
05_BF00173503	20170106_DIA_pos_07	20170106_DIA_neg_07	20170106_dd_pos_07	20170106_dd_neg_07
06_BF00173504	20170106_DIA_pos_08	20170106_DIA_neg_08	20170106_dd_pos_08	20170106_dd_neg_08
07_BF00173505	20170106_DIA_pos_09	20170106_DIA_neg_09	20170106_dd_pos_09	20170106_dd_neg_09
08_BF00173506	20170106_DIA_pos_10	20170106_DIA_neg_10	20170106_dd_pos_10	20170106_dd_neg_10
09_BF00173507	20170106_DIA_pos_11	20170106_DIA_neg_11	20170106_dd_pos_11	20170106_dd_neg_11
10_BF00173508	20170106_DIA_pos_12	20170106_DIA_neg_12	20170106_dd_pos_12	20170106_dd_neg_12
12_StdMixEawag50ugL	20170106_DIA_pos_13	20170106_DIA_neg_13	20170106_dd_pos_13	20170106_dd_neg_13
12_StdMixEawag50ugL	20170106_DIA_pos_14b	20170106_DIA_neg_14	20170106_dd_pos_14	20170106_dd_neg_14
Blank	20170106_DIA_pos_15			

ENTACT Mixes → Analysis → Data Pre-Processing → Prioritization → Identification

Target Screening

TraceFinder™ The Power to Know

Suspect Screening



RMassBank

Metfrag

Chemistry Dashboard | TOXCAST

https://www.kisspng.com/

LCSB

UNIVERSITÉ DU LUXEMBOURG

Analysis

ENTACT Mixture Measurements

Blank – Nanopure

Blank – DMSO

Waters: Atlantis T3
3µm, 3x150mm,
(with precolumn)

Q Exactive Plus
Orbitrap, Thermo

Mixes 499-508

SampleName	FileName_DIA_pos	FileName_DIA_neg	FileName_DD_pos	FileName_DD_neg
Blank_Nanopure	20170106_DIA_pos_01	20170106_DIA_neg_01	20170106_dd_pos_01	20170106_dd_neg_01
11_DMSO_Blank	20170106_DIA_pos_02	20170106_DIA_neg_02	20170106_dd_pos_02	20170106_dd_neg_02
01_BF00173499	20170106_DIA_pos_03	20170106_DIA_neg_03	20170106_dd_pos_03	20170106_dd_neg_03
02_BF00173500	20170106_DIA_pos_04	20170106_DIA_neg_04	20170106_dd_pos_04	20170106_dd_neg_04
03_BF00173501	20170106_DIA_pos_05	20170106_DIA_neg_05	20170106_dd_pos_05	20170106_dd_neg_05
04_BF00173502	20170106_DIA_pos_06	20170106_DIA_neg_06	20170106_dd_pos_06	20170106_dd_neg_06
05_BF00173503	20170106_DIA_pos_07	20170106_DIA_neg_07	20170106_dd_pos_07	20170106_dd_neg_07
06_BF00173504	20170106_DIA_pos_08	20170106_DIA_neg_08	20170106_dd_pos_08	20170106_dd_neg_08
07_BF00173505	20170106_DIA_pos_09	20170106_DIA_neg_09	20170106_dd_pos_09	20170106_dd_neg_09
08_BF00173506	20170106_DIA_pos_10	20170106_DIA_neg_10	20170106_dd_pos_10	20170106_dd_neg_10
09_BF00173507	20170106_DIA_pos_11	20170106_DIA_neg_11	20170106_dd_pos_11	20170106_dd_neg_11
10_BF00173508	20170106_DIA_pos_12	20170106_DIA_neg_12	20170106_dd_pos_12	20170106_dd_neg_12
12_StdMixEawag50ugL	20170106_DIA_pos_13	20170106_DIA_neg_13	20170106_dd_pos_13	20170106_dd_neg_13
12_StdMixEawag50ugL	20170106_DIA_pos_14b	20170106_DIA_neg_14	20170106_dd_pos_14	20170106_dd_neg_14
Blank	20170106_DIA_pos_15			

Data Independent
Pos & Neg

Data Dependent
Pos & Neg

Eawag Std Mix

Retention times for MetFrag

ENTACT Target Screening

- Target Screening: TraceFinder4.1
- Eawag internal database:
  - 1000 compounds, 1300 with internal standards
  - Adduct, MS/MS and RT on identical chromatography (Atlantis)

- Known limitations:
  - Method not suitable for perfluorinated compounds (PFCs) or quaternary ammonium compounds (QACs) as these do not elute, or elute as very broad peaks
  - Target matching *did not consider presence or absence of our targets in ToxCast* – possible isomer matches

Data Analysis ENTACT Target Screening

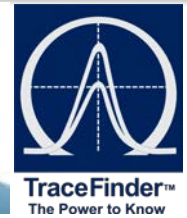
- Target Screening: TraceFinder4.1
- Eawag internal database:
  - 1000 compounds, 1300 with internal standards
  - Adduct, MS/MS and RT on identical chromatography (Atlantis)

Identification confidence	*Minimum data requirements*	*Reported identifications in ENTACT (PL / ES)*
Level 1: Confirmed structure by reference standard	MS, MS^2, RT, Reference Std.	**Level 1: IP 4.5 MS1, RT, MS/MS match**
Level 1: Confirmed structure by reference standard	MS, ~~MS^2,~~ RT, Reference Std.	**Level 1: IP 2 MS1, RT match. NO MS/MS**

Rejected Target Identifications [how to report?]
Present in blank, ≥ 1 of MS1, RT, MS/MS DO NOT match

Similar reporting used in Schymanski *et al*. 2015, ABC, DOI: 10.1007/s00216-015-8681-7

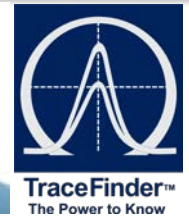
ENTACT Target Screening



- Caveat: DTXSID mapping
  (potential underestimate due to different forms!)

Mix	499	500	501	502	503	504	505	506	507	508
Total Present	95	95	95	95	185	185	365	365	95	365*
Target Pos IP4.5	14	11	8	8	27	20	50	48	2	39
Target Neg IP4.5	3	1	4	3	6	5	9	16	5	3
Target Pos IP2	9	4	8	6	4	6	12	10	1	5
Target Neg IP2	0	0	1	1	6	3	6	10	3	4

ENTACT Target Screening



o Caveat: DTXSID mapping
(potential underestimate due to different forms!)

Mix	499	500	501	502	503	504	505	506	507	508
Total Present	95	95	95	95	185	185	365	365	95	365*
Target Pos IP4.5	14	11	8	8	27	20	50	48	2	39
Target Neg IP4.5	3	1	4	3	6	5	9	16	5	3
Target Pos IP2	9	4	8	6	4	6	12	10	1	5
Target Neg IP2	0	0	1	1	6	3	6	10	3	4
Unique Targets (by DTXSID)	*25*	*14*	*20*	*18*	*40*	*32*	*73*	*81*	*11*	*49*
Present in Mix and Found (by DTXSID)	*20*	*10*	*16*	*14*	*29*	*27*	*63*	*69*	*11*	*42*
% Unique Targets Present & Found In Mix	***21%***	***11%***	***17%***	***15%***	***16%***	***15%***	***17%***	***19%***	***12%***	***12%***

ENTACT Target Screening

TraceFinder™ The Power to Know

- Caveat: DTXSID mapping
  (potential underestimate due to different forms!)

Mix	499	500	501	502	503	504	505	506	507	508
Total Present	95	95	95	95	185	185	365	365	95	365*

- Systematically found more targets than present in mix
- More on this to come later

Unique Targets (by DTXSID)	*25*	*14*	*20*	*18*	*40*	*32*	*73*	*81*	*11*	*49*
Present in Mix and Found (by DTXSID)	*20*	*10*	*16*	*14*	*29*	*27*	*63*	*69*	*11*	*42*
% Unique Targets Present & Found In Mix	***21%***	***11%***	***17%***	***15%***	***16%***	***15%***	***17%***	***19%***	***12%***	***12%***

11 to 21 % of mixes covered with targets

Identification ENTACT Target Screening

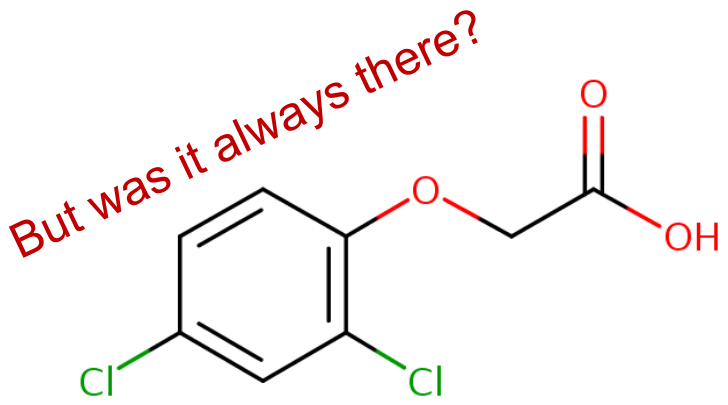
- Total Targets detected (unique): 230
- Total Targets detected & in mixes: 188
- % Targets detected & in mixes: **15 %**

- Rejected Targets: 38-94 positive mode, 17-28 negative

- Targets present in many mixes

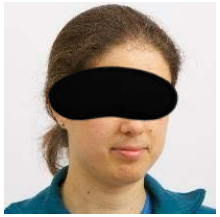
>=6	1
>=4	13
>=3	23
>=2	56
>=1	187

DTXSID0020442



https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID0020442

ENTACT Suspect Screening



- MS Processing: RMassBank
- Screening List: Chemistry Dashboard | TOXCAST

https://github.com/MassBank/RMassBank and https://msbi.ipb-halle.de/MetFragBeta/ and https://massbank.eu/MassBank/

RMassBank Overview

eawag aquatic research

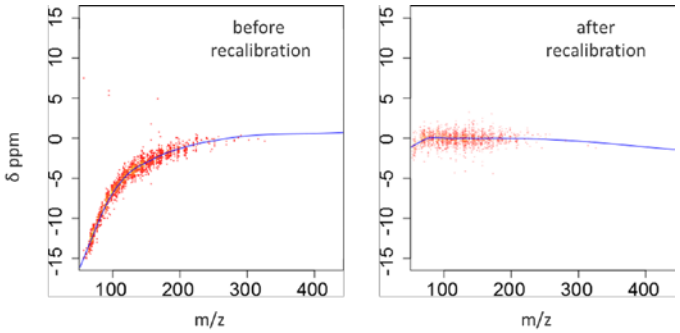
LevelCode	LevelKeyword	COMMENT.CONFIDENCE
1	standard	Reference Standard (Level 1)
1a	standard	Reference Standard (Level 1)
1b	parent	Parent Substance with Reference Standard (Level 1)
1c	confirmed	Identification confirmed with Reference Standard (Level 1)
2	probable	Probable structure, tentative identification (Level 2)
2a	probableLibrary	Probable structure via library match, tentative identification (Level 2a)
2b	probableDiagnostic	Probable structure via diagnostic evidence, tentative identification (Level 2b)
3	tentative	Tentative identification only (Level 3)
3a	tentativeStructure	Tentative identification: most likely structure (Level 3)
3b	tentativeIsomer	Tentative identification: isomers possible (Level 3)
3c	tentativeTPClass	Tentative identification: substance class known (Level 3)
3d	tentativeBestMatch	Tentative identification: best match only (Level 3)
4	formula	Tentative identification: molecular formula only (Level 4)
5	unknown	Tentative identification: structure and formula unknown (Level 5)
5	exactMass	Tentative identification: structure and formula unknown (Level 5)

LC-MS/MS raw data

online resources: CTS, CACTUS

Automatic MS and MS/MS Recalibration and Clean-up Remove interfering peaks

Spectral Annotation with
- Experimental Details
- Compound Information



MassBank records

structure files

nprman

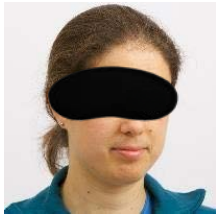
MassBank.eu

https://github.com/MassBank/RMassBank/
http://bioconductor.org/packages/RMassBank/

RMassBank

Stravs *et al*. *(*2013), *J. Mass Spectrom,* 48(1):89-99. DOI: 10.1002/jms.3131

ENTACT Suspect Screening



- MS Processing: RMassBank
- Screening List: Chemistry Dashboard | TOXCAST
- Identification: MetFrag + TOXCAST

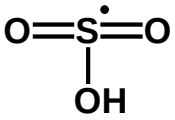
https://github.com/MassBank/RMassBank and https://msbi.ipb-halle.de/MetFragBeta/ and https://massbank.eu/MassBank/

MetFrag Overview

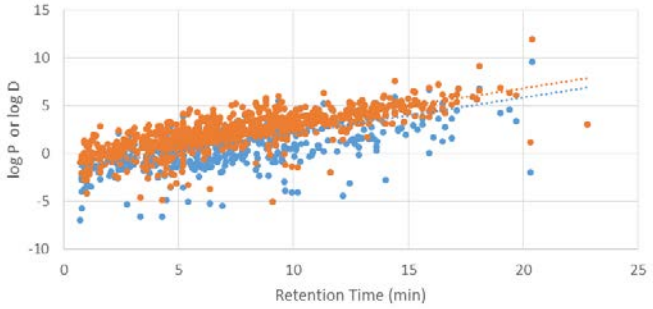

Status: 2010 => 2016

mz [M-H]⁻
213.9637
± 5 ppm

Elements: C,N,S
5 ppm
0.001 Da



RT: 4.54 min
355 InChI/RTs



ChemSpider
Search and

TOXCAST

PubChem OPEN CHEMISTRY DATABASE

MetFrag 2.3

MoNA
MassBank of North America
MassBank.eu

References
External Refs
Data Sources
RSC Count
PubMed Count

Literature Patents

Suspect Lists

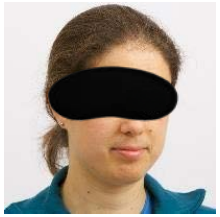
?TOFF !DENT
Chemistry Dashboard

MS/MS

134.0054	339689
150.0001	77271
213.9607	632466

Ruttkies, Schymanski, Wolf, Hollender, Neumann (2016) *J. Cheminf.*, 2016, DOI: 10.1186/s13321-016-0115-9

ENTACT Suspect Screening

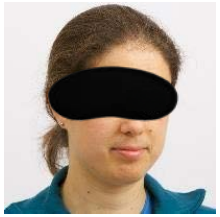


- MS Processing: RMassBank
- Screening List: Chemistry Dashboard | TOXCAST
- Identification: MetFrag + TOXCAST
  - Fragmenter Score
  - Retention Time Score (RTs from Eawag Mix Standards)
  - Offline MetFusion Score
  - Offline Individual MoNA Score
    - Maximum Score: 4
    - NO REFERENCE INFORMATION (irrelevant)

MassBank.eu

https://github.com/MassBank/RMassBank and https://msbi.ipb-halle.de/MetFragBeta/ and https://massbank.eu/MassBank/

ENTACT Suspect Screening



- Known limitations:
  - **TIME TO PROCESS RESULTS**
  - Isobars: RMassBank reports *most intense* peak by default – insufficient time to report and extract *all* matches
  - Computationally inefficient but eliminates peak picking
  - Metals could not be captured by RMassBank
  - Not all adducts were processed; DIA not processed for MS/MS
  - Very “ENTACT-specific” – not a realistic scenario at all
  - Results reflect chosen “boundary conditions”
    - Min 5E5 and max. 400 candidates positive mode, min. 1E5 negative

- Lesson learnt:
  - Level 2a is not yet clearly automatable
    - *This is confirmed by our unblinded target results!*

ENTACT Suspect Screening

eawag aquatic research

Correlates with number of substances … but why? [artefact?]

Mix	499	500	501	502	503	504	505	506	507	508	DMSO	Water
Total Present	95	95	95	95	185	185	365	365	95	365*	0	0
Suspect Pos	393	400	400	400	400	400	400	400	400	400	80	69
Suspect Neg	150	146	160	141	205	237	365	319	171	312	70	66
Total Suspects	543	546	560	541	605	637	765	719	571	712	150	135
Suspects with MSMS	***251***	***276***	***282***	***263***	***352***	***411***	***562***	***542***	***276***	***505***	***76***	***60***
% Suspects with MSMS	***46%***	***51%***	***50%***	***49%***	***58%***	***65%***	***73%***	***75%***	***48%***	***71%***	***51%***	***44%***

ENTACT Suspect Screening

eawag aquatic research

Correlates with number of substances … but why? [artefact?]

Mix	499	500	501	502	503	504	505	506	507	508	DMSO	Water
Total Present	95	95	95	95	185	185	365	365	95	365*	0	0
Suspect Pos	393	400	400	400	400	400	400	400	400	400	80	69
Suspect Neg	150	146	160	141	205	237	365	319	171	312	70	66
Total Suspects	543	546	560	541	605	637	765	719	571	712	150	135
Suspects with MSMS	***251***	***276***	***282***	***263***	***352***	***411***	***562***	***542***	***276***	***505***	***76***	***60***
% Suspects with MSMS	***46%***	***51%***	***50%***	***49%***	***58%***	***65%***	***73%***	***75%***	***48%***	***71%***	***51%***	***44%***
Total Unique Suspects (by FORMULA)	*467*	*484*	*484*	*483*	*503*	*545*	*608*	*596*	*495*	*577*	*143*	*125*
Present in Mix (by FORMULA)	*84*	*86*	*88*	*84*	*165*	*164*	*331*	*323*	*61*	*226*	*0*	*0*
% Present in Mix & Found (by FORMULA)	***88%***	***91%***	***93%***	***88%***	***89%***	***89%***	***91%***	***88%***	***64%***	***62%***	***0%***	***0%***
% (Possible) True Hits	***18%***	***18%***	***18%***	***17%***	***33%***	***30%***	***54%***	***54%***	***12%***	***39%***	***0%***	***0%***
% False Hits	***82%***	***82%***	***82%***	***83%***	***67%***	***70%***	***46%***	***46%***	***88%***	***61%***	***100%***	***100%***

Better coverage … but lower quality … many false hits

ENTACT Suspect Screening

eawag aquatic research

For candidates with multiple matches, many have matching isobars

Mix	499	500	501	502	503	504	505	506	507	508	DMSO	Water
Total Present	95	95	95	95	185	185	365	365	95	365*	0	0
Total Suspects	543	546	560	541	605	637	765	719	571	712	150	135
Only 1 Candidate	281	271	286	275	322	389	484	492	292	385	70	60
Average Candidates	2.35	2.41	2.34	2.41	2.18	2.04	1.82	1.80	2.35	2.16	2.55	2.67
n=1, score >=3.5	***10***	***11***	***12***	***11***	***18***	***26***	***51***	***59***	***8***	***33***	***1***	***4***
n>=1, score >=3.5	***25***	***29***	***41***	***36***	***45***	***47***	***91***	***91***	***22***	***77***	***7***	***10***
Total cand, score >=3.5	31	31	50	43	52	53	102	102	23	90	7	10
n=1, score >=3	107	110	117	108	165	231	321	348	120	239	21	19
n>=1, score>=3	154	168	183	163	239	306	424	423	165	362	37	30

- Lesson learnt:
  - Level 2a is not yet clearly automatable!!!!
  - ToxCast is too big to do high quality suspect screening … too time consuming, too many candidates, too many peaks …

Part 2: Eawag & LCSB ENTACT Unblinded

SampleName	FileName_DIA_pos	FileName_DIA_neg	FileName_DD_pos	FileName_DD_neg
Blank_Nanopure	20170106_DIA_pos_01	20170106_DIA_neg_01	20170106_dd_pos_01	20170106_dd_neg_01
11_DMSO_Blank	20170106_DIA_pos_02	20170106_DIA_neg_02	20170106_dd_pos_02	20170106_dd_neg_02
01_BF00173499	20170106_DIA_pos_03	20170106_DIA_neg_03	20170106_dd_pos_03	20170106_dd_neg_03
02_BF00173500	20170106_DIA_pos_04	20170106_DIA_neg_04	20170106_dd_pos_04	20170106_dd_neg_04
03_BF00173501	20170106_DIA_pos_05	20170106_DIA_neg_05	20170106_dd_pos_05	20170106_dd_neg_05
04_BF00173502	20170106_DIA_pos_06	20170106_DIA_neg_06	20170106_dd_pos_06	20170106_dd_neg_06
05_BF00173503	20170106_DIA_pos_07	20170106_DIA_neg_07	20170106_dd_pos_07	20170106_dd_neg_07
06_BF00173504	20170106_DIA_pos_08	20170106_DIA_neg_08	20170106_dd_pos_08	20170106_dd_neg_08
07_BF00173505	20170106_DIA_pos_09	20170106_DIA_neg_09	20170106_dd_pos_09	20170106_dd_neg_09
08_BF00173506	20170106_DIA_pos_10	20170106_DIA_neg_10	20170106_dd_pos_10	20170106_dd_neg_10
09_BF00173507	20170106_DIA_pos_11	20170106_DIA_neg_11	20170106_dd_pos_11	20170106_dd_neg_11
10_BF00173508	20170106_DIA_pos_12	20170106_DIA_neg_12	20170106_dd_pos_12	20170106_dd_neg_12
12_StdMixEawag50ugL	20170106_DIA_pos_13	20170106_DIA_neg_13	20170106_dd_pos_13	20170106_dd_neg_13
12_StdMixEawag50ugL	20170106_DIA_pos_14b	20170106_DIA_neg_14	20170106_dd_pos_14	20170106_dd_neg_14
Blank	20170106_DIA_pos_15			

Mixes 499-508

ENTACT Mixes → Analysis → Data Pre-Processing → Prioritization → Identification

	A	B	H	I	J
1	Aliquot_Plate_Ba	Aliquot_Well_ID	DTXSID	CASRN	Preferred_Name
92	BF00173499	EPA_MIXTURE_001	DTXSID9032113	107534-96-3	Tebuconazole
93	BF00173499	EPA_MIXTURE_001	DTXSID9034282	122-88-3	4-Chlorophenoxyacetic acid
94	BF00173499	EPA_MIXTURE_001	DTXSID9041653	4707-47-5	Methyl 2,4-dihydroxy-3,6-dimethy
95	BF00173499	EPA_MIXTURE_001	DTXSID9044322	929-73-7	Dodecylamine hydrochloride
96	BF00173499	EPA_MIXTURE_001	DTXSID9045007	98-72-6	Nitarsone
97	BF00173500	EPA_MIXTURE_002	DTXSID0032493	55219-65-3	Triadimenol
98	BF00173500	EPA_MIXTURE_002	DTXSID0034192	10161-33-8	17beta-Trenbolone
99	BF00173500	EPA_MIXTURE_002	DTXSID0037495	1007-28-9	Deisopropylatrazine
100	BF00173500	EPA_MIXTURE_002	DTXSID0038700	97-52-9	2-Methoxy-4-nitroaniline
101	BF00173500	EPA_MIXTURE_002	DTXSID0042454	589-68-4	Tetradecanoic acid, 2,3-dihydroxy
102	BF00173500	EPA_MIXTURE_002	DTXSID0044650	430227	2,2'-[(4-Methylphenyl)imino]dieth
103	BF00173500	EPA_MIXTURE_002	DTXSID0045254	64-85-7	21-Hydroxyprogesterone

All | 499 | 500 | 501 | 502 | 503 | 504 | 505 | 506 | 507 | 508 | Notes | Tabelle1

RMassBank

MetFrag

Unblinded Results

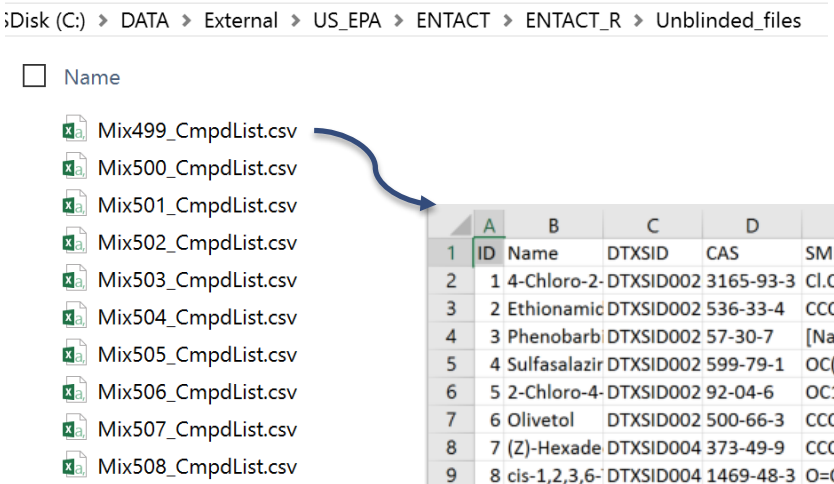
- MS Processing: RMassBank
- Screening List: ***Compound List of Mix Content***
- Identification: MetFrag + ***Compound List of Mix Content***
  - Fragmenter Score
  - Retention Time Score (RTs from Eawag Mix Standards)
  - Offline MetFusion Score
  - Offline Individual MONA Score
    - Maximum Score: 4
    - NO REFERENCE INFORMATION (irrelevant)

MassBank.eu

https://github.com/MassBank/RMassBank and https://msbi.ipb-halle.de/MetFragBeta/ and https://massbank.eu/MassBank/

Data Analysis Compound List Creation

eawag aquatic research



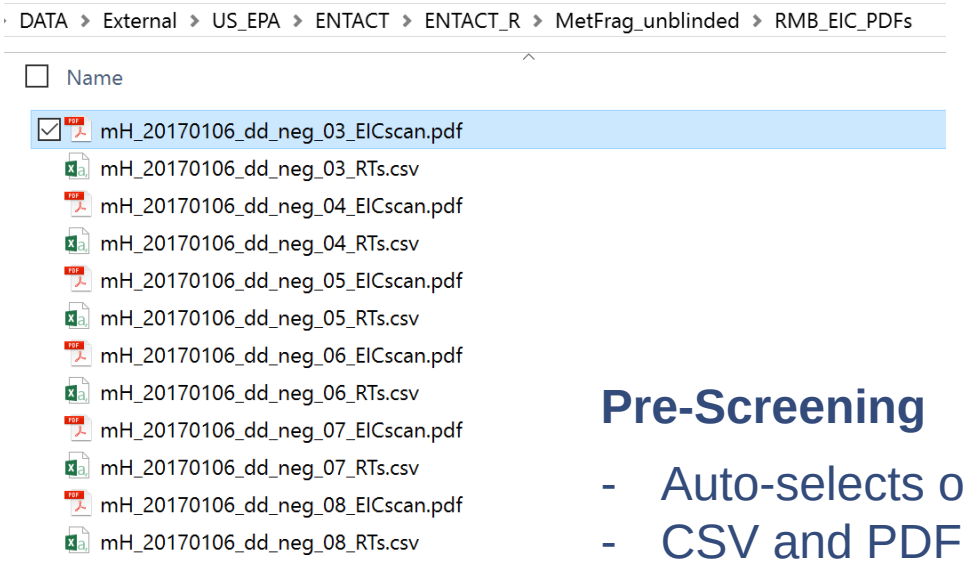
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	ID	Name	DTXSID	CAS	SMILES_DT	DTXCID	SMILES	InChIKey	Formula	mz	RT	Level	Source	
2	1	4-Chloro-2-	DTXSID002	3165-93-3	Cl.CC1=C(N	DTXCID902	CC1=C(N)C=	CXNVOWPI	C7H8ClN	141.0345		1	BF00173499	
3	2	Ethionamic	DTXSID002	536-33-4	CCC1=NC=(	DTXCID005	CCC1=NC=(	AEOCXXJPC	C8H10N2S	166.0565		1	BF00173499	
4	3	Phenobarb	DTXSID002	57-30-7	[Na+].CCC1	DTXCID601	CCC1(C(=O)	DDBREPKU	C12H12N2	232.0848		1	BF00173499	
5	4	Sulfasalazir	DTXSID002	599-79-1	OC(=O)C1=	DTXCID401	OC(=O)C1=	NCEXYHBE	C18H14N4	398.0685		1	BF00173499	
6	5	2-Chloro-4-	DTXSID002	92-04-6	OC1=CC(Cl)	DTXCID502	OC1=CC(Cl)	MXORDJXB	C12H9ClO	204.0342		1	BF00173499	
7	6	Olivetol	DTXSID002	500-66-3	CCCCCC1=C	DTXCID205	CCCCCC1=C	IRMPFYJSH	C11H16O2	180.115		1	BF00173499	
8	7	(Z)-Hexade	DTXSID004	373-49-9	CCCCCC\C=	DTXCID501	CCCCCC\C=	SECPZKHBE	C16H30O2	254.2246		1	BF00173499	
9	8	cis-1,2,3,6-	DTXSID004	1469-48-3	O=C1NC(=C	DTXCID106	O=C1NC(=C	CIFFBTOJC	C8H9NO2	151.0633		1	BF00173499	
10	9	Monopotas	DTXSID004	877-24-7	[K+].OC(=O	DTXCID901	OC(=O)C1=	XNGIFLGAS	C8H6O4	166.0266		1	BF00173499	
11	10	Octylparab	DTXSID004	1219-38-1	CCCCCCCCC	DTXCID002	CCCCCCCCC	RIKCMEDSI	C15H22O3	250.1569		1	BF00173499	
12	11	2-Amino-5-	DTXSID102	121-88-0	NC1=C(O)C	DTXCID206	NC1=C(O)C	DOPJTDJKZ	C6H6N2O3	154.0378		1	BF00173499	
13	12	4-Chloroph	DTXSID102	106-48-9	OC1=CC=C(	DTXCID201	OC1=CC=C(	WXNZTHH	C6H5ClO	128.0029		1	BF00173499	
14	13	Stavudine	DTXSID102	3056-17-5	CC1=CN([C	DTXCID102	CC1=CN(C2	XNKLLVCAF	C10H12N2	224.0797		1	BF00173499	
15	14	Pirimicarb	DTXSID103	23103-98-2	CN(C)C(=O)	DTXCID901	CN(C)C(=O)	YFGYUFNIC	C11H18N4	238.143		1	BF00173499	
16	15	17-Methylt	DTXSID103	58-18-4	C[C@]1(O)(	DTXCID202	CC1(O)CCC	GCKMFJBG	C20H30O2	302.2246		1	BF00173499	
17	16	Tamoxifen	DTXSID103	10540-29-1	CC\C(=C(/C	DTXCID901	CC\C(=C(/C	NKANXQFJ	C26H29NO	371.2249		1	BF00173499	
18	17	Phenyl 1-hy	DTXSID103	132-54-7	OC1=C(C=C	DTXCID901	OC1=C(C=C	QHDYIMW	C17H12O3	264.0786		1	BF00173499	

- Chemicals in each list had a unique ID (0XXX, 1XXX, …. 9XXX)
- Processing done on DTXCID SMILES
- Metals now included in RMassBank [PPGs done separately]
- No retention times given (=> isobars = Level 3!!!!!!)

LCSB

uni.lu UNIVERSITÉ DU LUXEMBOURG

RMassBank Pre-Screening



Pre-Screening

- Auto-selects only those with MS/MS for RMassBank
- CSV and PDF output for quality control
- CSV basis for (part) of results summary
- CAVEAT: only takes most intense peak without RT (=> Level 3 for isobars)
- [this can be addressed but I ran out of time …]

https://github.com/schymane/ReSOLUTION/blob/master/R/RMB_EIC_prescreen.R

Data Analysis RMassBank Pre-Screening

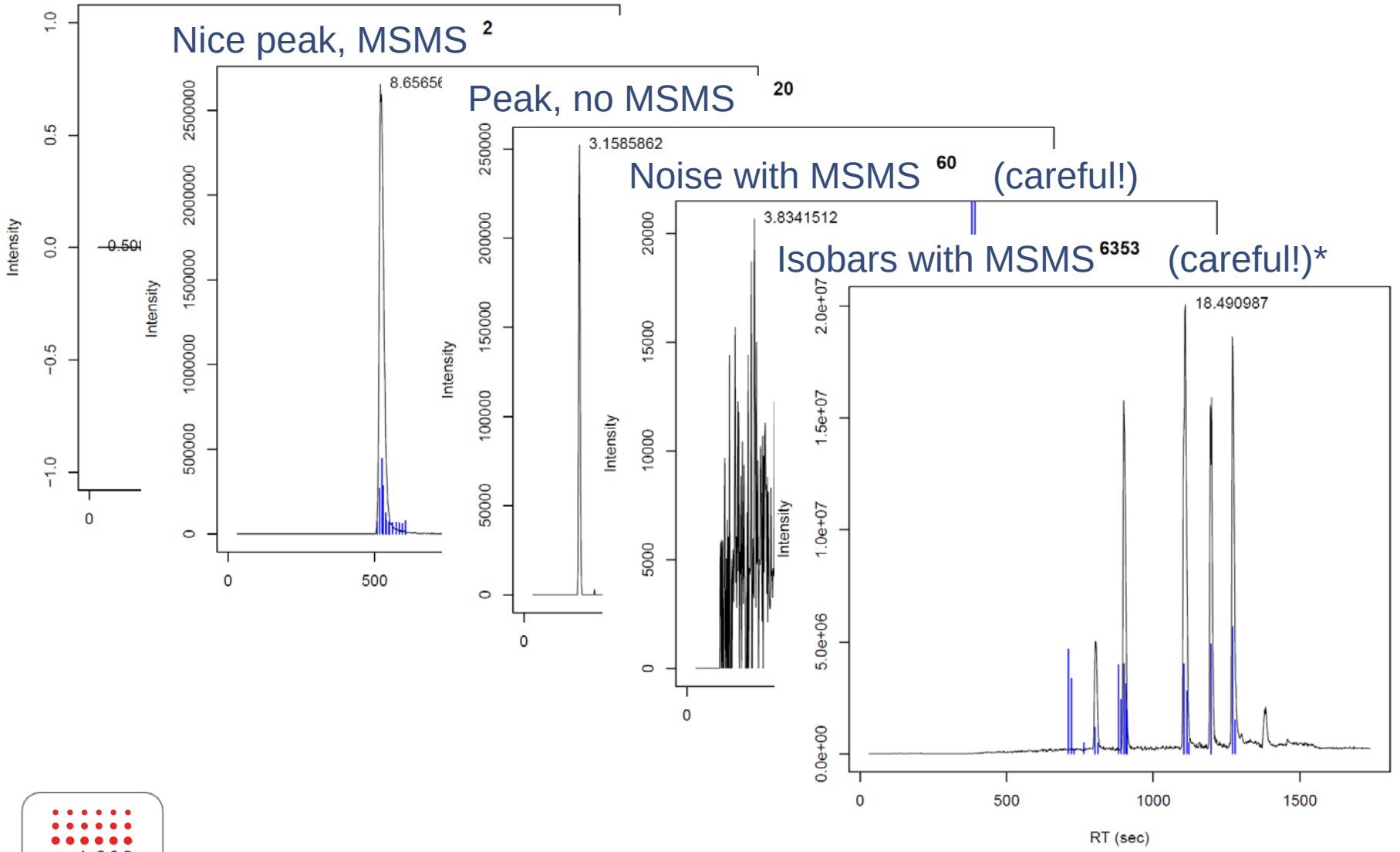
No peak at all 1

Nice peak, MSMS 2

Peak, no MSMS 20

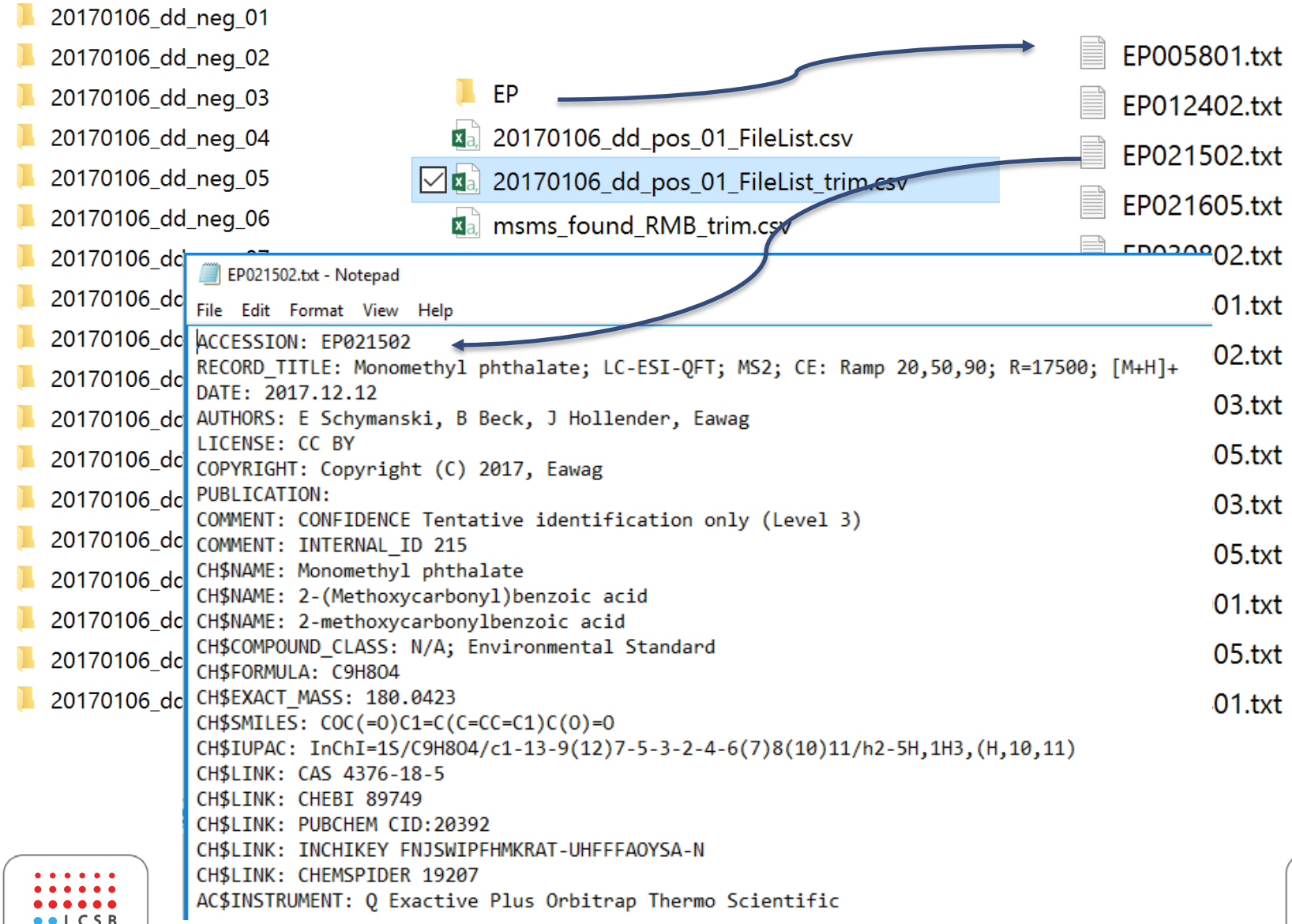
Noise with MSMS 60 (careful!)

Isobars with MSMS 6353 (careful!)*



https://github.com/schymane/ReSOLUTION/blob/master/R/RMB_EIC_prescreen.R

Data Analysis RMassBank Results



https://github.com/MassBank/RMassBank

MetFrag Calculations

- Information auto-retrieved via the unique ID from
  - Individual mixture lists (used also to create summaries)
  - MassBank records created during processing
  - Dummy record (mz=999, I=999) used if MS/MS was absent
- Screening List: ***LocalDB of Mix Content (localCSV)***
- Identification:
  - Fragmenter Score
  - Retention Time Score (RTs from Eawag Mix Standards)
  - Offline MetFusion Score
  - Offline Individual MONA Score
    - Maximum Score: 4
    - NO REFERENCE INFORMATION (irrelevant)

http://c-ruttkies.github.io/MetFrag/projects/metfragcl/ and https://github.com/schymane/ReSOLUTION/

MetFrag Calculations

- Create Config Files (MetFragConfig)
- Run MetFrag (runMetFrag)

```
176     # now, run MetFrag and extract results for reporting into cmpd_info.
177
178     results_filename <- paste0(run_name,"_",cmpdID_char,"_",as.character(i))
179     if (isPos) {
180       config_file <- MetFragConfig(mass = ExactMass, adduct_type = 0, neutralPrecursorMass=TRUE,
181                                    results_filename = results_filename,
182                                    peaklist_path = MetFrag_msms, base_dir = results_run_dir,
183                                    DB = "LocalCSV", localDB_path=localDB,useMonaIndiv = T,useMoNAMetFusion = T,
184                                    IsPosMode = TRUE,filter_by_InChIKey = F,rt_file_path=MetFrag_rt_file,rt_exp=RT)
185     } else {
186       config_file <- MetFragConfig(mass = ExactMass, adduct_type = 0, neutralPrecursorMass=TRUE,
187                                    results_filename = results_filename,
188                                    peaklist_path = MetFrag_msms, base_dir = results_run_dir,
189                                    DB = "LocalCSV", localDB_path=localDB,useMonaIndiv = T,useMoNAMetFusion = T,
190                                    IsPosMode = FALSE,filter_by_InChIKey = F,rt_file_path=MetFrag_rt_file,rt_exp=RT)
191     }
192
193     runMetFrag(config_file, MetFrag_dir, CL_name = "MetFrag2.4.4-msready-CL.jar")
194
195     results_file <- paste0(results_run_dir,"/results/",results_filename,".xls")
```


http://c-ruttkies.github.io/MetFrag/projects/metfragcl/ and https://github.com/schymane/ReSOLUTION/

o Extract Results and Summarize

```
198     #extract results we need:
199     MetFrag_res <- read_excel(results_file)
200
201     cmpd_info$num_poss_IDs[i] <- length(MetFrag_res$Score)
202     cmpd_info$poss_IDs[i] <- paste(MetFrag_res$Name,collapse=";")
203     cmpd_info$poss_ID_scores[i] <- paste(MetFrag_res$Score,collapse=";")
204     cmpd_info$max_Score[i] <- max(MetFrag_res$Score)
205     cmpd_info$n_Score_GE3p5[i] <- length(which(MetFrag_res$Score>=3.5))
206     cmpd_info$n_Score_GE3[i] <- length(which(MetFrag_res$Score>=3))
207     cmpd_info$n_Score_GE2p5[i] <- length(which(MetFrag_res$Score>=2.5))
208     cmpd_info$poss_DTXSIDs[i] <- paste(MetFrag_res$DTXSID,collapse=";")
209     cmpd_info$poss_CAS[i] <- paste(MetFrag_res$CAS,collapse=";")
210     cmpd_info$MoNAScore[i] <- paste(MetFrag_res$OfflineIndividualMoNAScore,collapse=";")
211     cmpd_info$MaxMoNAScore[i] <- max(MetFrag_res$OfflineIndividualMoNAScore)
212
213   }
214
215
216   write.csv(cmpd_info,paste0(results_summary_dir,"/MetFragResultSummary_",run_name,".csv"),row.names = F)
```


http://c-ruttkies.github.io/MetFrag/projects/metfragcl/ and https://github.com/schymane/ReSOLUTION/

Data Analysis MetFrag Calculations

eawag aquatic research

- Voila!

DATA › External › US_EPA › ENTACT › ENTACT_R › MetFr

Name

- MetFragResultSummary_20170106_dd_neg_04.csv
- MetFragResultSummary_20170106_dd_neg_05.csv
- MetFragResultSummary_20170106_dd_neg_06.csv
- MetFragResultSummary_20170106_dd_neg_07.csv
- MetFragResultSummary_20170106_dd_neg_08.csv

	A	B	C	D	E	F	G	H	I	J
1	ID	Name	DTXSID	CAS	SMILES_DT	DTXCID	SMILES	InChIKey	Formula	exactMass
8	7	(Z)-Hexade	DTXSID004	373-49-9	CCCCCC\C=	DTXCID501	CCCCCC\C=	SECPZKHBE	C16H30O2	254.2246
9	8	cis-1,2,3,6-	DTXSID004	1469-48-3	O=C1NC(=C	DTXCID106	O=C1NC(=C	CIFFBTOJCK	C8H9NO2	151.0633
10	9	Monopotas	DTXSID004	877-24-7	[K+].OC(=O	DTXCID901	OC(=O)C1=	XNGIFLGAS	C8H6O4	166.0266
11	10	Octylparab	DTXSID004	1219-38-1	CCCCCCCCC	DTXCID002	CCCCCCCCC	RIKCMEDSE	C15H22O3	250.1569
12	11	2-Amino-5-	DTXSID102	121-88-0	NC1=C(O)C	DTXCID206	NC1=C(O)C	DOPJTDJKZ	C6H6N2O3	154.0378
13	12	4-Chloroph	DTXSID102	106-48-9	OC1=CC=C(	DTXCID201	OC1=CC=C(	WXNZTHH(	C6H5ClO	128.0029
14	13	Stavudine	DTXSID102	3056-17-5	CC1=CN([C	DTXCID102	CC1=CN(C2	XNKLLVCAF	C10H12N2	224.0797
15	14	Pirimicarb	DTXSID103	23103-98-2	CN(C)C(=O)	DTXCID901	CN(C)C(=O)	YFGYUFNIC	C11H18N4	238.143
16	15	17-Methylt	DTXSID103	58-18-4	C[C@]1(O)(	DTXCID202	CC1(O)CCC	GCKMFJBG	C20H30O2	302.2246
17	16	Tamoxifen	DTXSID103	10540-29-1	CC\C(=C(/C	DTXCID901	CC\C(=C(/C	NKANXQFJ	C26H29NO	371.2249
18	17	Phenyl 1-hy	DTXSID103	132-54-7	OC1=C(C=C	DTXCID901	OC1=C(C=C	QHDYIMW	C17H12O3	264.0786
19	18	Propamoca	DTXSID104	24579-73-5	CCCOC(=O)	DTXCID902	CCCOC(=O)	WZZLDXDL	C9H20N2O	188.1525
20	19	Clofibric ac	DTXSID104	882-09-7	CC(C)(OC1=	DTXCID902	CC(C)(OC1=	TXCGAZHT	C10H11ClO	214.0397

	A	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF
1	ID	RT_RMB_p	maxI_RMB	MSMS_RM	msms_avai	CollisionEn	msms_peal	num_poss_	poss_IDs	poss_ID_sc	max_Score	n_Score_GI	n_Score_GI	n_Score_GI	poss_DTXSI	poss_CAS	MoNAScor	RT_RMB	MaxMoNAScore	
8	7	24.10022	210452.6	TRUE	TRUE	Ramp 20, 5	210.0915:2	1	(Z)-Hexade	3	3	0	1	1	DTXSID004	373-49-9	0	19.125	0	
9	8	14.35642	5921155	TRUE	TRUE	Ramp 20, 5	152.0706:4	1	cis-1,2,3,6-	3	3	0	1	1	DTXSID004	1469-48-3	0	14.298	0	
10	9	13.36366	4603213	TRUE	TRUE	Ramp 20, 5	167.0336:1	1	Monopotas	3	3	0	1	1	DTXSID004	877-24-7	0	13.335	0	
11	10	22.08323	720673.9	FALSE	TRUE	Ramp 20, 5	250.0832:3	1	Octylparab	3	3	0	1	1	DTXSID004	1219-38-1	0	12.575	0	
12	11	13.42212	6202431	TRUE	TRUE	Ramp 20, 5	155.0451:6	1	2-Amino-5-	3	3	0	1	1	DTXSID102	121-88-0	0	13.364	0	
13	12	3.66803	65882.92	FALSE	FALSE	N/A	N/A	1	4-Chloroph	1	1	0	0	0	DTXSID102	106-48-9	0		0	
14	13	20.25228	5359125	TRUE	TRUE	Ramp 20, 5	227.0417:2	1	Stavudine	3	3	0	1	1	DTXSID102	3056-17-5	0	20.223	0	
15	14	14.61709	2.2E+08	TRUE	TRUE	Ramp 20, 5	239.1502:3	1	Pirimicarb	3.99747	3.99747	1	1	1	DTXSID103	23103-98-2	0.99747	14.617	0.99747	
16	15	20.45478	62800456	TRUE	TRUE	Ramp 20, 5	303.2316:1	1	17-Methylt	3	3	0	1	1	DTXSID103	58-18-4	0	20.368	0	
17	16	17.82425	18223456	TRUE	TRUE	Ramp 20, 5	372.2322:4	1	Tamoxifen	3	3	0	1	1	DTXSID103	10540-29-1	0	17.766	0	
18	17	22.93052	39799.7	FALSE	TRUE	Ramp 20, 5	217.0361:8	1	Phenyl 1-hy	3	3	0	1	1	DTXSID103	132-54-7	0	20.541	0	
19	18	8.381433	47508344	TRUE	TRUE	Ramp 20, 5	189.1598:7	1	Propamoca	3.97525	3.97525	1	1	1	DTXSID104	24579-73-5	0.97525	8.44	0.97525	
20	19	19.18297	13768.19	FALSE	FALSE	N/A	N/A	1	Clofibric ac	1	1	0	0	0	DTXSID104	882-09-7	0		0	
21	20	17.94073	88652360	TRUE	TRUE	Ramp 20, 5	207.065:12	1	5,7-Dimeth	3	3	0	1	1	DTXSID104	487-06-9	0	17.853	0	
22	21	1.85239	2762502	TRUE	TRUE	Ramp 20, 5	113.0597:9	1	1,3-Cyclohe	3	3	0	1	1	DTXSID104	504-02-9	0	1.929	0	
23	22	8.003695	744583.6	TRUE	TRUE	Ramp 20, 5	258.0432:1	1	Potassium	3	3	0	1	1	DTXSID104	70321-85-6	0	8.004	0	
24	23	8.381433	47236864	TRUE	TRUE	Ramp 20, 5	138.0914:3	1	4-Methoxy	3	3	0	1	1	DTXSID202	102-50-1	0	8.468	0	
25	24	10.03769	44673616	TRUE	TRUE	Ramp 20, 5	230.0551:1	1	Amiloride h	3.00182	3.00182	0	1	1	DTXSID202	2016-88-8	0.00182	10.038	0.00182	
26	25	19.90874	72932112	TRUE	TRUE	Ramp 20, 5	331.0398:1	1	Fenarimol	3	3	0	1	1	DTXSID203	60168-88-9	0	19.937	0	
27	26	20.85829	2.7E+08	TRUE	TRUE	Ramp 20, 5	278.0556:1	1	Triflumizole	3	3	0	1	1	DTXSID203	68694-11-1	0	20.802	0	
28	27	18.43411	42072580	TRUE	TRUE	Ramp 20, 5	381.1608:5	1	CP-457677	3	3	0	1	1	DTXSID204	214535-77	0	18.548	0	
29	28	19.82335	84253376	TRUE	TRUE	Ramp 20, 5	286.1438:1	1	Piperine	3.76577	3.76577	1	1	1	DTXSID302	94-62-2	0.76577	19.88	0.76577	

MetFragResultSummary_20170106_dd_pos_12.csv

SummaryOfMetFragSummaries.xlsx

LCSB

uni.lu UNIVERSITÉ DU LUXEMBOURG

http://c-ruttkies.github.io/MetFrag/projects/metfragcl/ and https://github.com/schymane/ReSOLUTION/

Overview: Negative Mode

- 35-54 % coverage in negative mode!

Mix	499	500	501	502	503	504	505	506	507	508
Total Present	95	95	95	95	185	185	365	365	95	365*
Level1, IP4.5	41	30	38	30	66	62	150	129	24	66
Level1, IP2	6	2	3	3	12	5	27	23	2	5
Level 3	2(1)	2(1)	0	0	4(2)	0	12(9)	2(2)	29(25)	102(70)
CorrectTarget	2	1	3	3	5	4	8	17	5	4
IncorrectTarget	1	0	1	1	7	4	7	9	0	3
- Incorrect Target IP2	0	0	1	1	5	3	4	6	0	2
ND or <1E5	46	61	54	62	102	118	176	210	40	191
% Level 1 IP4.5	***43%***	***32%***	***40%***	***32%***	***36%***	***34%***	***41%***	***35%***	***25%***	***18%***
% Level 1 IP4.5 and 2	***49%***	***34%***	***43%***	***35%***	***42%***	***36%***	***48%***	***42%***	***27%***	***19%***
% All Levels	***51%***	***35%***	***43%***	***35%***	***43%***	***36%***	***51%***	***42%***	***54%***	***39%***
% Correct Eawag Targets	2%	1%	3%	3%	3%	2%	2%	5%	5%	1%

Overview: Positive Mode

eawag aquatic research

- 68-83 % coverage in positive mode

Mix	499	500	501	502	503	504	505	506	507	508
Total Present	95	95	95	95	185	185	365	365	95	365*
Level1, IP4.5	74	74	79	69	146	140	283	280	31	132
Level1, IP2	0	0	0	0	0	0	3	5	0	0
Level 3	2(2)	2(1)	0	0	3(2)	0	16(13)	2(2)	46(34)	174(141)
CorrectTarget	13	10	8	7	25	21	54	50	7	37
IncorrectTarget	10	3	10	8	6	4	7	9	2	6
- Incorrect Target IP2	9	3	8	6	3	2	2	4	2	0
ND or <5E5	19	19	16	26	39	45	63	77	18	50
% Level 1 IP4.5	***78%***	***78%***	***83%***	***73%***	***79%***	***76%***	***78%***	***77%***	***33%***	***36%***
% Level 1 IP4.5 and 2	***78%***	***78%***	***83%***	***73%***	***79%***	***76%***	***78%***	***78%***	***33%***	***36%***
% All Levels	***80%***	***79%***	***83%***	***73%***	***80%***	***76%***	***82%***	***79%***	***68%***	***75%***
% Correct Eawag Targets	14%	11%	8%	7%	14%	11%	15%	14%	7%	10%

Overview: Combined Results

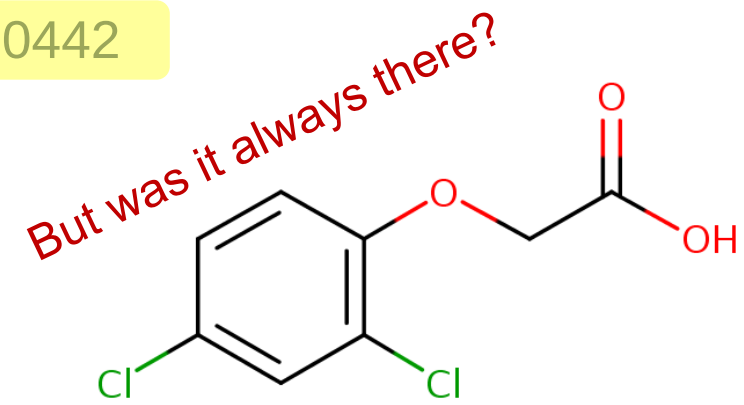
- 85-96 % coverage over both modes
  - Very few Level 1 IP2s remaining
  - Mixes 499-506: 83-91 % Level 1 IP4.5, up to 35 % in both modes

Mix	499	500	501	502	503	504	505	506	507	508
Total Present	95	95	95	95	185	185	365	365	95	365*
Pos										
Level1, IP4.5	82	82	86	79	159	160	321	316	40	139
- In both modes	33	22	31	20	49	42	112	93	15	59
Level1, IP2	1	1	0	2	1	0	5	8	1	1
- In both modes	0	0	0	0	0	0	0	0	0	0
Level 3	2	2	0	0	4	0	18	2	50	190
ND or <5E5	10	10	9	14	21	25	21	38	4	27
% Level 1 IP4.5	*86%*	*86%*	*91%*	*83%*	*86%*	*86%*	*88%*	*87%*	*42%*	*38%*
% Level 1 IP4.5 and 2	*87%*	*87%*	*91%*	*85%*	*86%*	*86%*	*89%*	*89%*	*43%*	*38%*
% All Levels	*89%*	*89%*	*91%*	*85%*	*89%*	*86%*	*94%*	*89%*	*96%*	*90%*
%NDs	*11%*	*11%*	*9%*	*15%*	*11%*	*14%*	*6%*	*10%*	*4%*	*7%*
%Level1 IP4.5 Both	*35%*	*23%*	*33%*	*21%*	*26%*	*23%*	*31%*	*25%*	*16%*	*16%*

De-bugging Targets present in many mixes (or not)

>=6	1
>=4	13
>=3	23
>=2	56
>=1	187

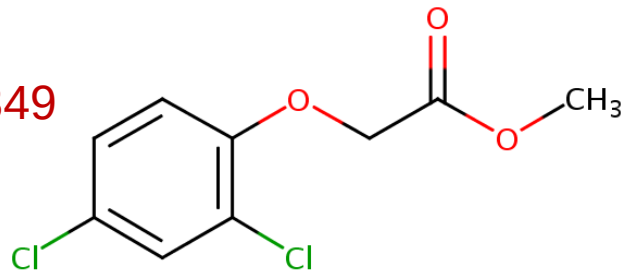
DTXSID0020442



Identified: Mixes 499, 501, 503, 505, 507, 508
Present: Mixes 501, 507
Absent: Mixes 499, 503, 505, 508

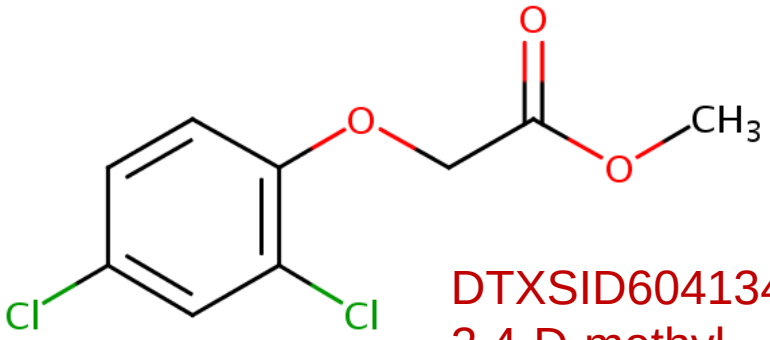
No obvious source yet…
(SMARTS search)

2,4-D-methyl
DTXSID6041349

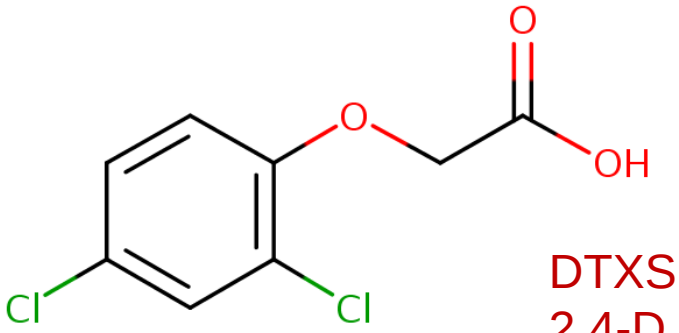


https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID0020442

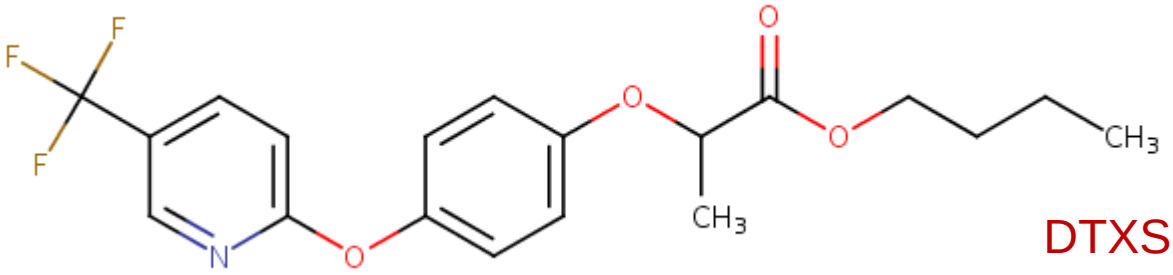
De-bugging "In source" Related Substances



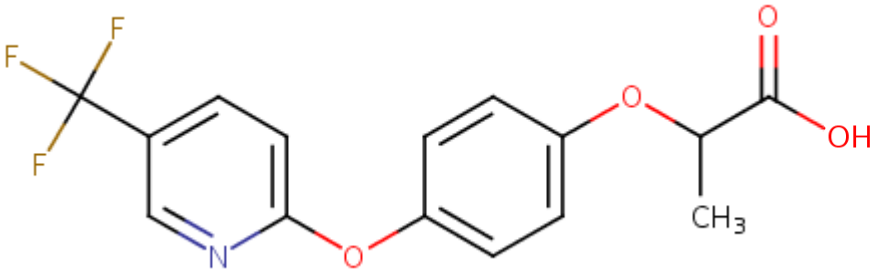
DTXSID6041349
2,4-D-methyl



DTXSID0020442
2,4-D



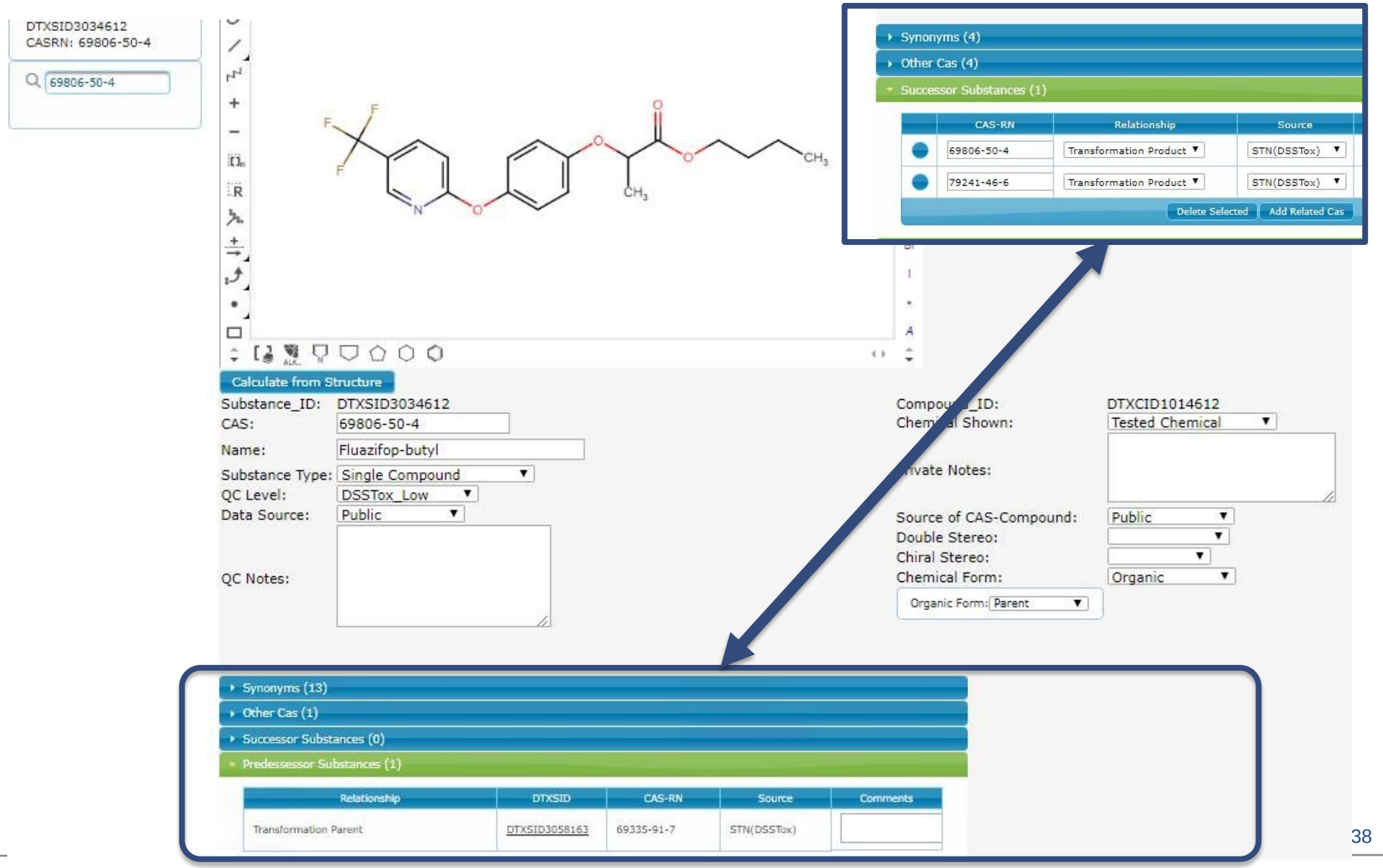
DTXSID3034612
Fluazifop-butyl



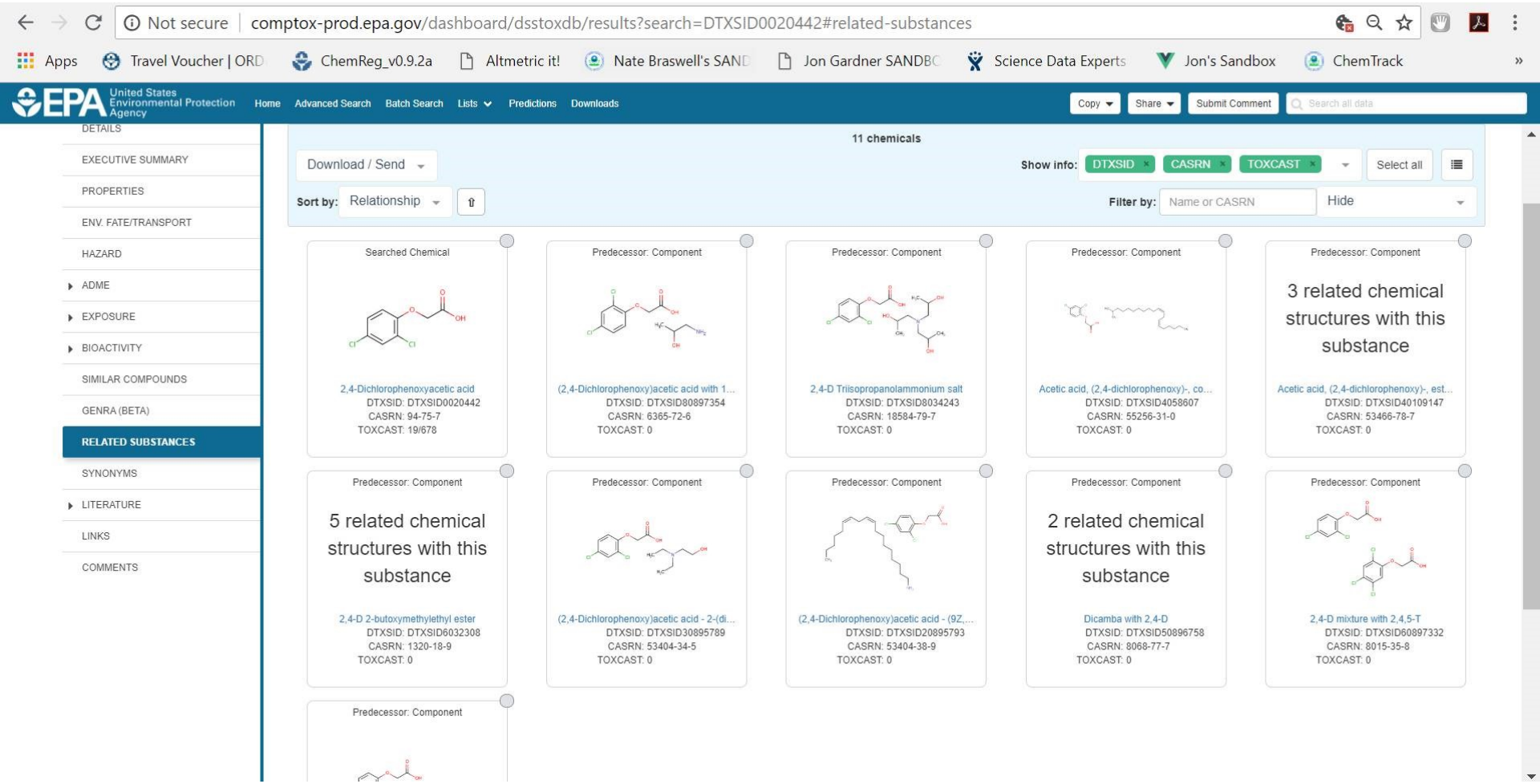
DTXSID0034855
Fluazifop-P-butyl

DTXSID3058163
Fluazifop

Future … CompTox Related Substances

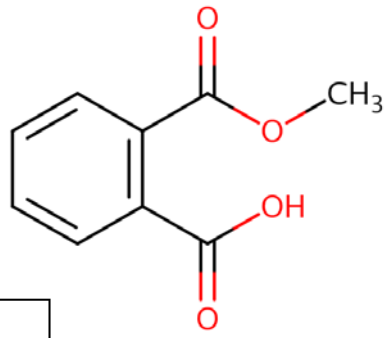


Future … CompTox Related Substances

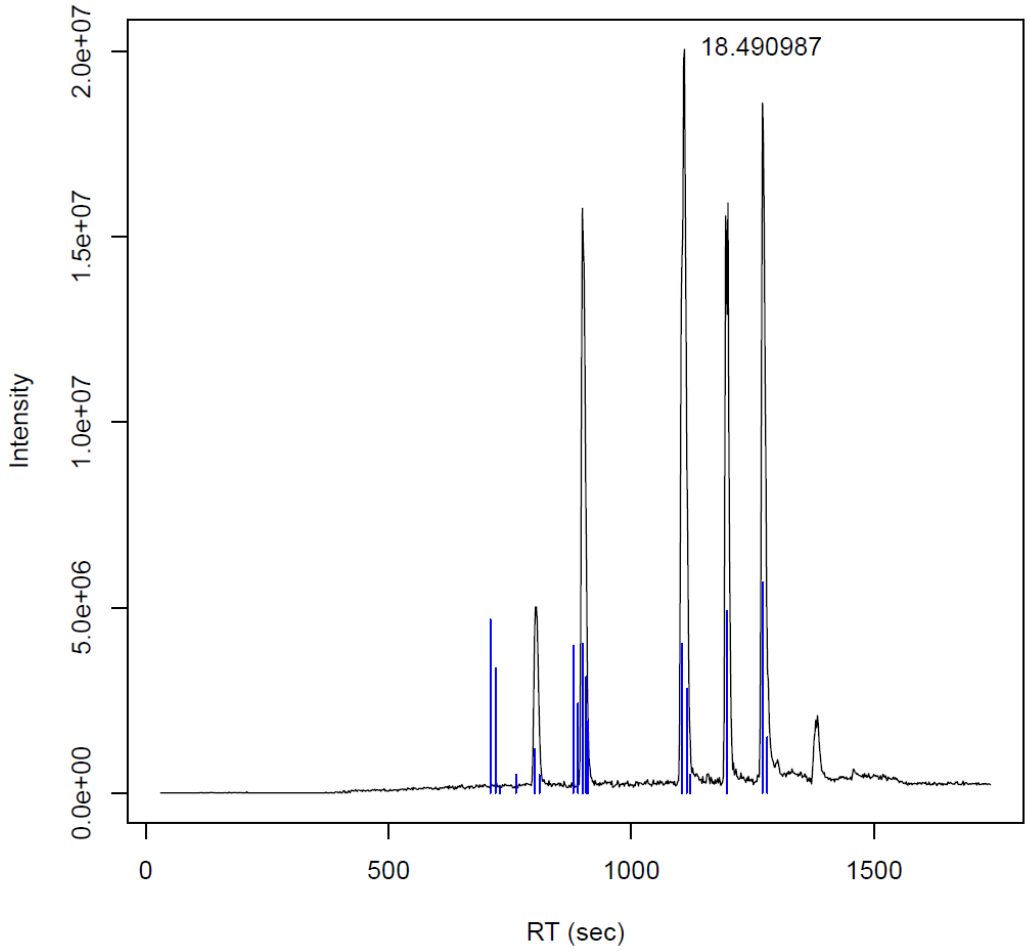


Detecting In Source Fragments

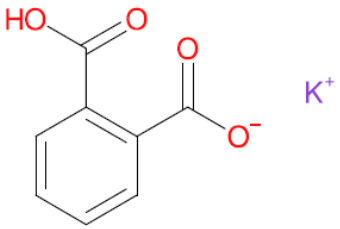
- Mix 505: One candidate ….
  - DTXSID9040001, $C_9H_8O_4$



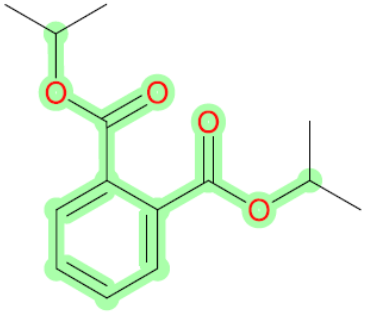
- How many peaks?



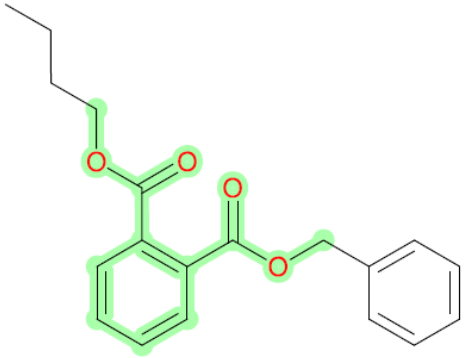
De-bugging Mix 505 – Search “phthalate”, find …



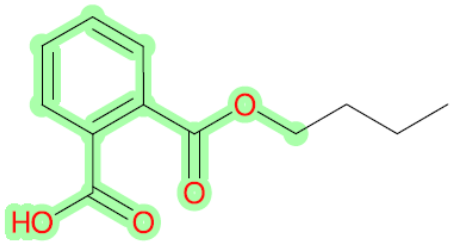
DTXSID0042167



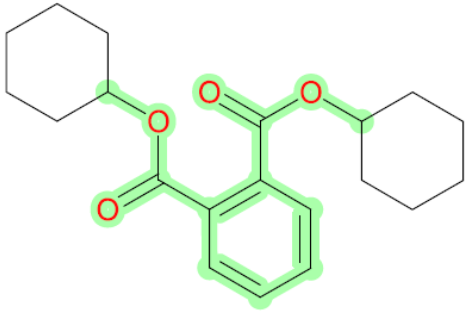
DTXSID2040731



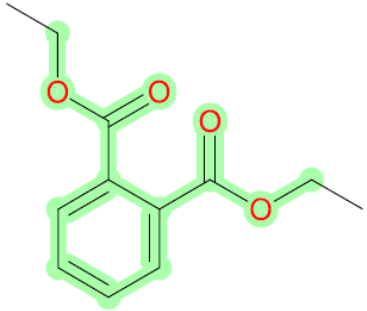
DTXSID3020205



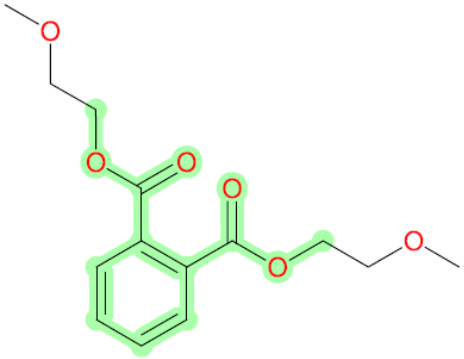
DTXSID4040002



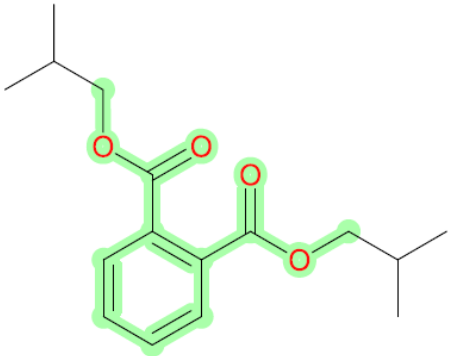
DTXSID5025021



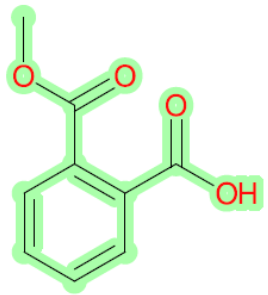
DTXSID7021780



DTXSID8025094

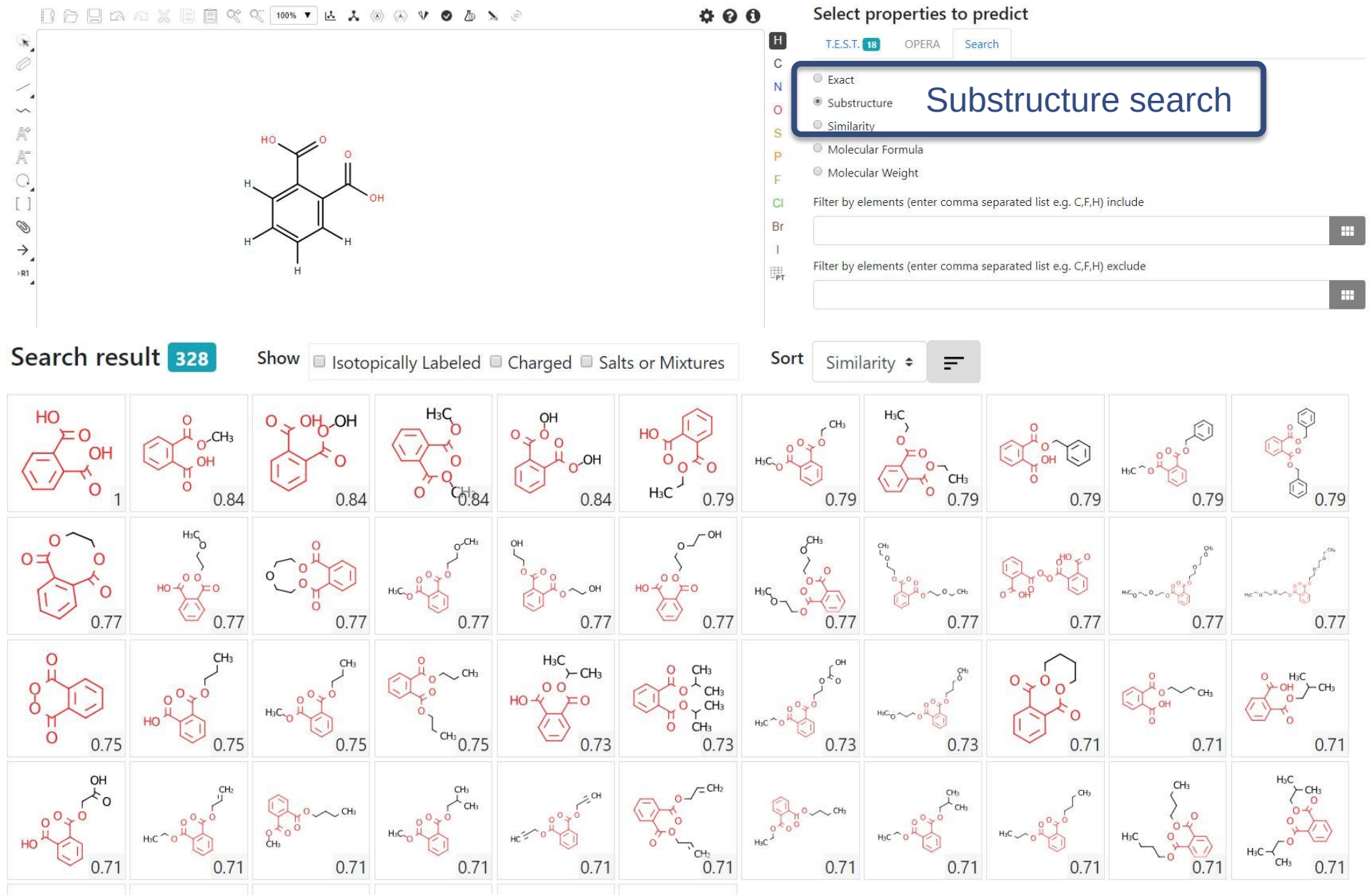


DTXSID9022522



DTXSID9040001

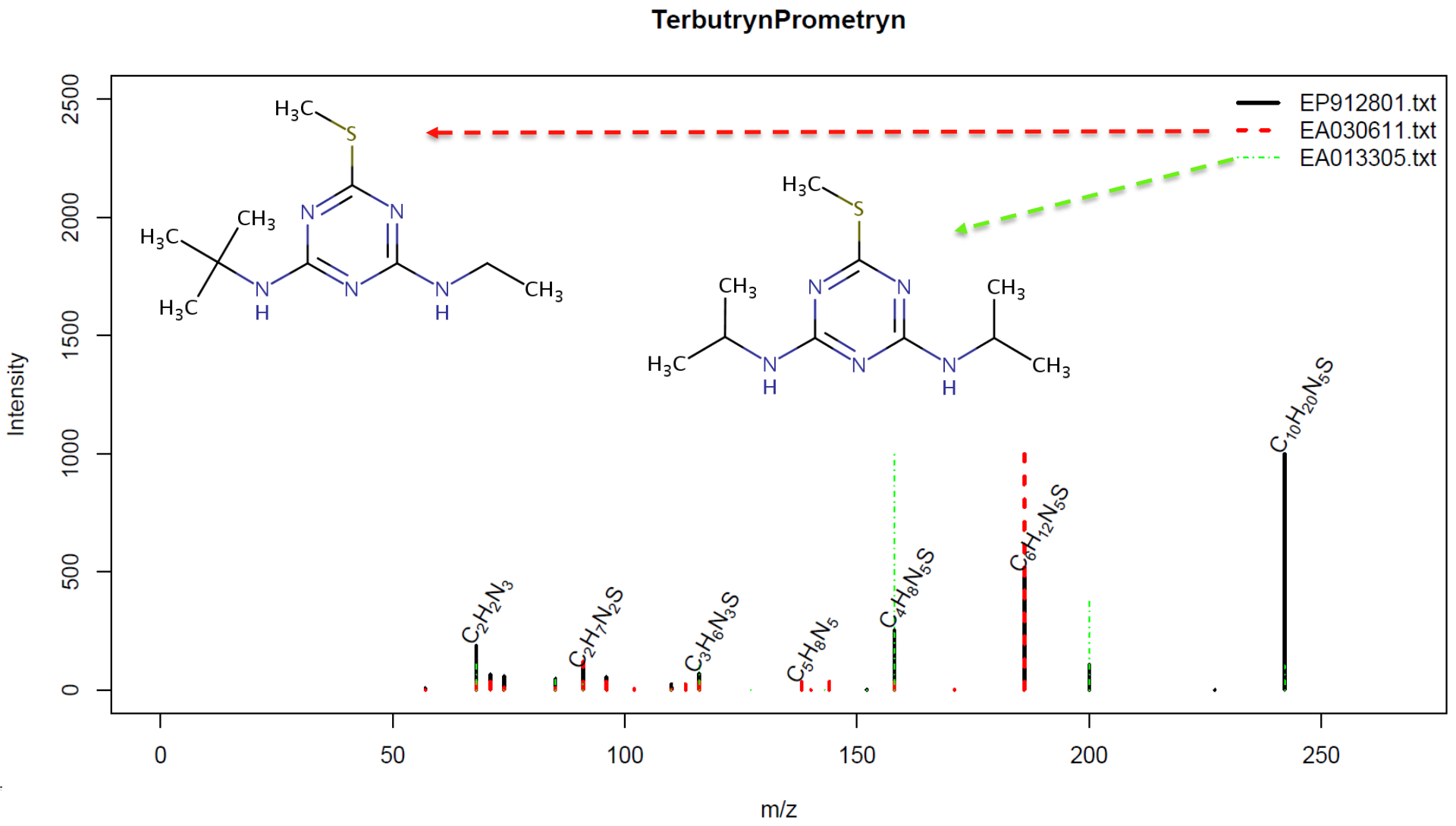
Future … CompTox Substructure Search



De-bugging MS/MS for co-eluting peaks?

Identical retention time for reference standards … co-eluting …

Both detected but only one has the “better” match value ..



De-bugging “Wrong” targets TPs of Mix Content?

Level 1 IP4.5 Target “not included” in Mix 508 – definitely present

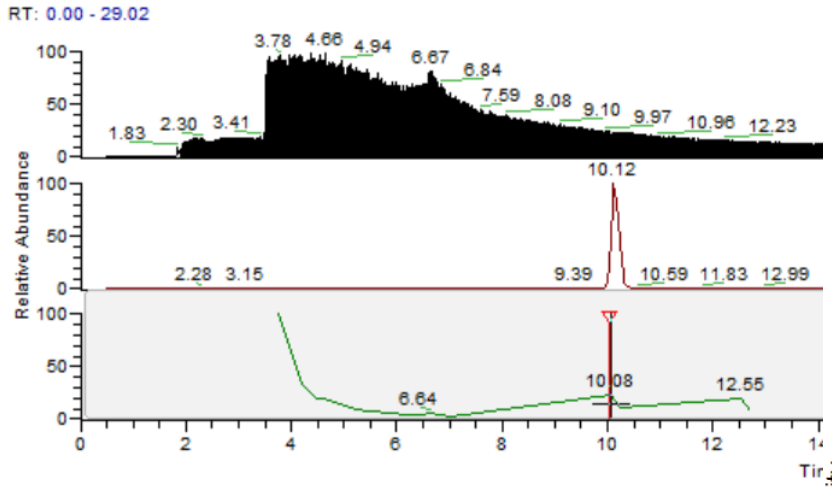
https://massbank.eu/MassBank/jsp/RecordDisplay.jsp?id=EA066805&dsn=Eawag + RT match on Atlantis system

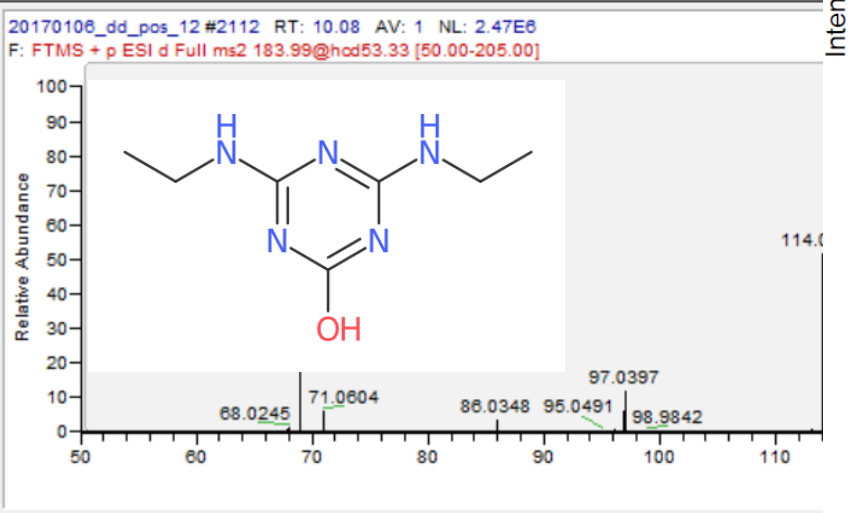
De-bugging "Wrong" targets TPs of Mix Content?

Simazine-2-hydroxy absent in mix … but parent Simazine is present!

TraceFinder™ The Power to Know

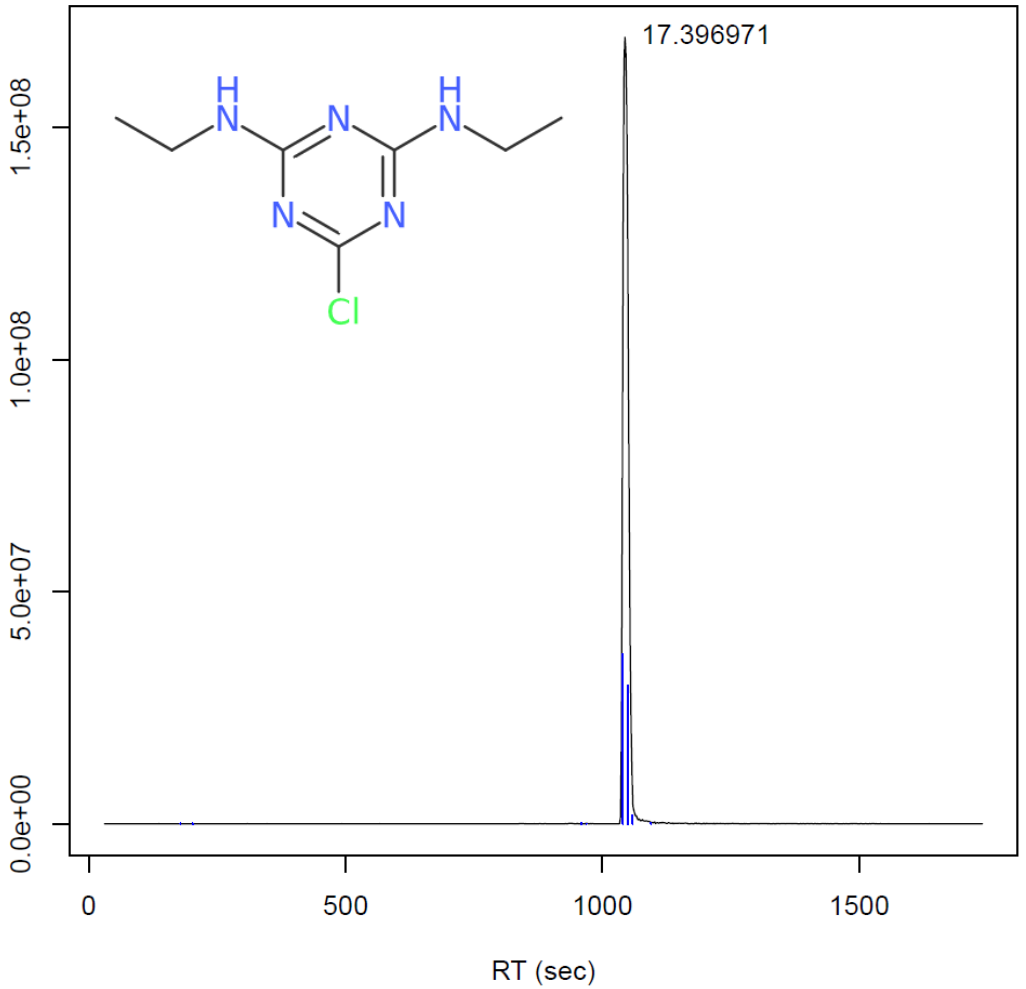
C:\DATA\...\ENTACT\20170106_dd_pos_12 01/08/17 09:52:33 10_BF00173508





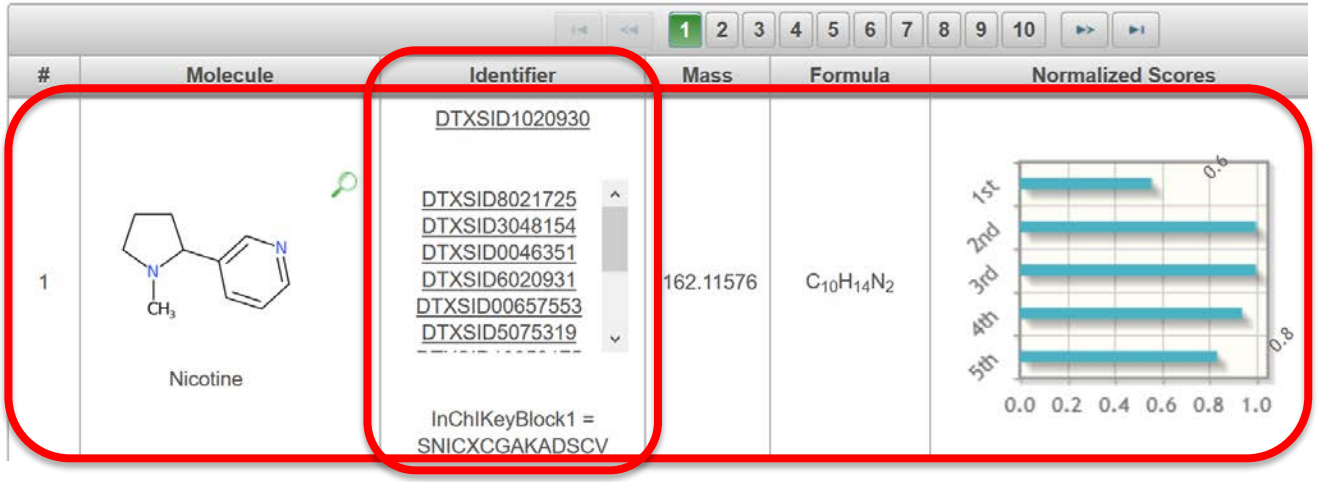
https://massbank.eu/MassBank/jsp/RecordDisplay.jsp?id=

9152

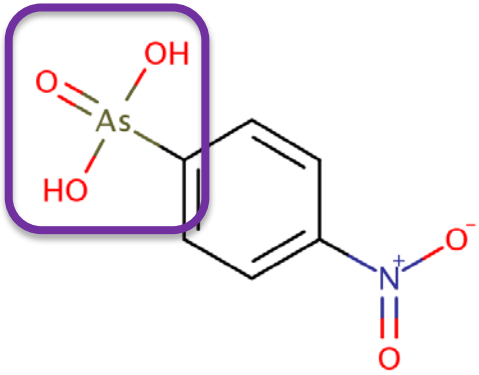


Part 3: MetFrag and CompTox Efforts

- 3Ms: Making MetFrag MS-Ready

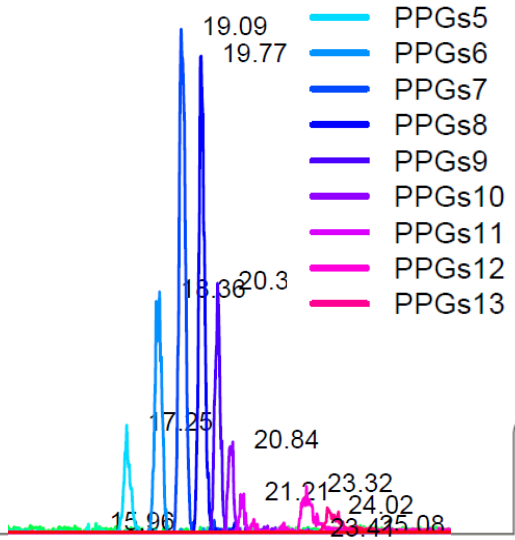


- 2Ms: Metals or no Metals?



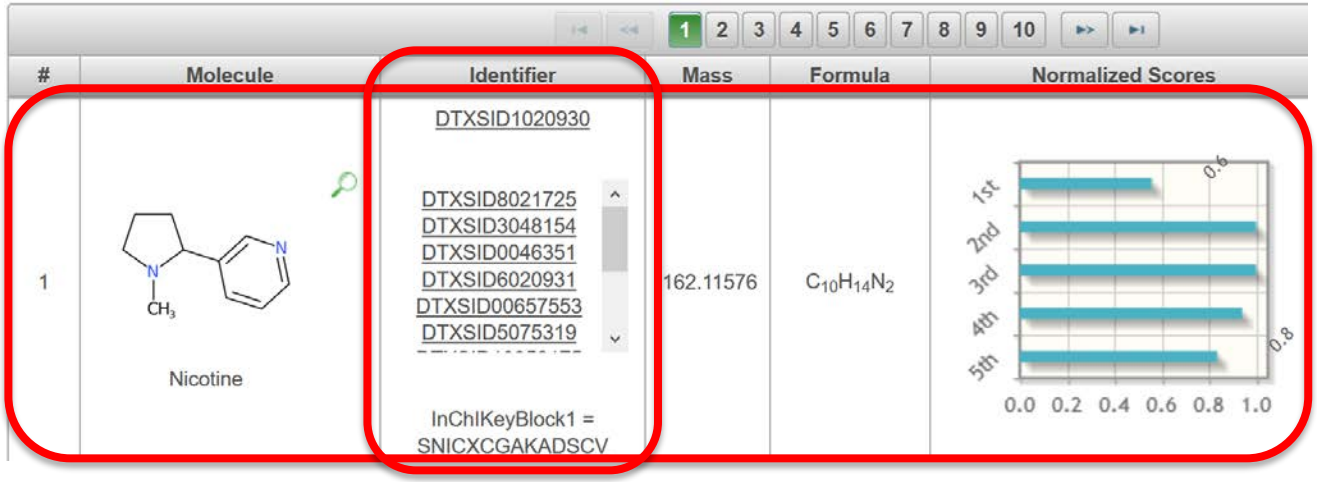
- 1M: Mixtures…

DTXSID9027863 - PPGs



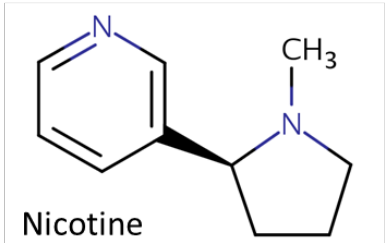
Part 3: MetFrag and CompTox Efforts

- 3Ms: Making MetFrag MS-Ready

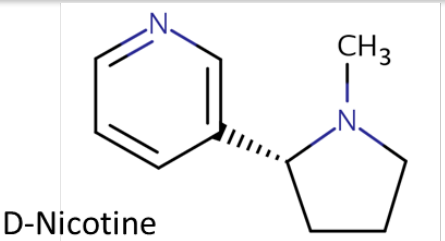


MS-ready & Chemical Metadata

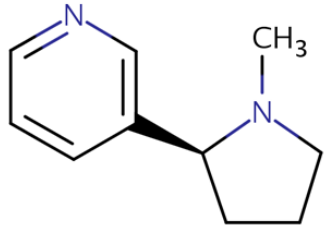
LEGEND: Name, SMILES
DTXSID | InChIKey 1st Block
CAS | **Monoiso. Mass** | logP | **Sources**
Data on: **Toxicity | Exposure | Bioassays**



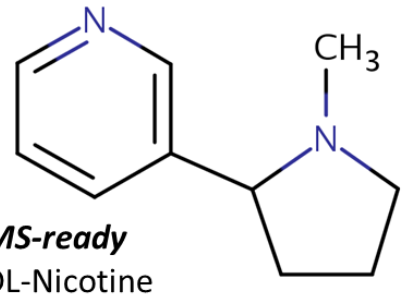
Nicotine
CN1CCC[C@H]1C1=CN=CC=C1
DTXSID1020930| SNICXCGAKADSCV
54-11-5 | **162.1157**| 0.929| **72**
Tox: **yes**| Expo: **yes**| Bioassay: **yes**



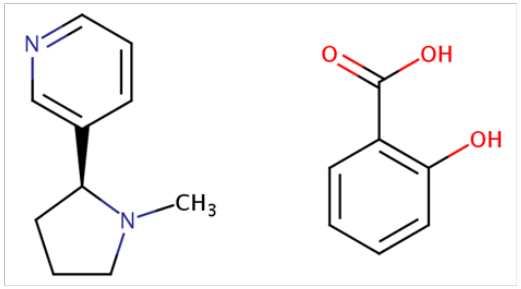
D-Nicotine
CN1CCC[C@@H]1C1=CN=CC=C1
DTXSID0046351| SNICXCGAKADSCV
25162-00-9 | **162.1157**| 0.929| **20**
Tox: **no**| Expo: **yes**| Bioassay: **yes**



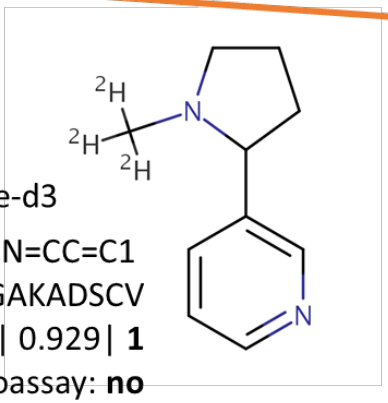
HCl
Nicotine hydrochloride
Cl.CN1CCC[C@H]1C1=CN=CC=C1
DTXSID6020931| HDJBTCAJIMNXEW
2820-51-1 | **198.0924** | 0.929| **9**
Tox: **no**| Expo: **yes**| Bioassay: **yes**



MS-ready
DL-Nicotine
CN1CCCC1C1=CN=CC=C1
DTXSID3048154 | SNICXCGAKADSCV
22083-74-5 | **162.1157**| 0.953| 9
Tox: **yes**| Expo: **no**| Bioassay: **yes**



Benzoic acid, 2-hydroxy-, compd. with
3-[(2S)-1-methyl-2-pyrrolidinyl]pyridine (1:1)
OC(=O)C1=C(O)C=CC=C1.CN1CCC[C@H]1C1=CN=CC=C1
DTXSID5075319| AIBWPBUAKCMKNS
29790-52-1| **300.1474**| 0.929| **6**
Tox: **no**| Expo: **yes**| Bioassay: **no**



DL-Nicotine-d3
[2H]C([2H])([2H])N1CCCC1C1=CN=CC=C1
DTXSID80442666| SNICXCGAKADSCV
69980-24-1| **165.1345**| 0.929| **1**
Tox: **no**| Expo: **no**| Bioassay: **no**

Schymanski & Williams, 2017, ES&T51 (10), pp 5357–5359. DOI: 10.1021/acs.est.7b01908

MS-ready, Dashboard & MetFrag

https://comptox.epa.gov/dashboard/dsstoxdb/batch_search

curare

EPA United States Environmental Protection Agency

Home | Advanced Search | Batch Search | Lists | Predictions | Downloads | Search

Chemistry Dashboard

Aa | Aa | Aa

Step One — Step Two — Step Three — Step Four — Step Five — Step Six

Step Three: Select Download Data or Display Chemicals

Please enter one identifier per line

Select Input Type(s)

- [ ] Chemical Name
- [ ] CASRN
- [ ] InChIKey
- [ ] Skeleton
- [ ] DSSTox Substance ID
- [x] MS-Ready Formula(e)
- [ ] Exact Formula(e)
- [ ] Monoisotopic Mass

Metadata

- [ ] Curation Level Details
- [x] Data Sources
- [x] Assay Hit Count
- [ ] Include links to ACToR reports - SLOW! (BETA)
- [ ] NHANES/Predicted Exposure
- [ ] Include ToxVal Data Availability
- [ ] Number of PubMed Articles
- [ ] Abstract Sifter Input File (Beta)
- [x] MetFrag Input File(Beta)
- [x] IRIS
- [x] PPRTV
- [x] PubChem Data Sources
- [ ] ToxPrint fingerprints
- [x] CPDat Product Occurrence Count

ifiers)

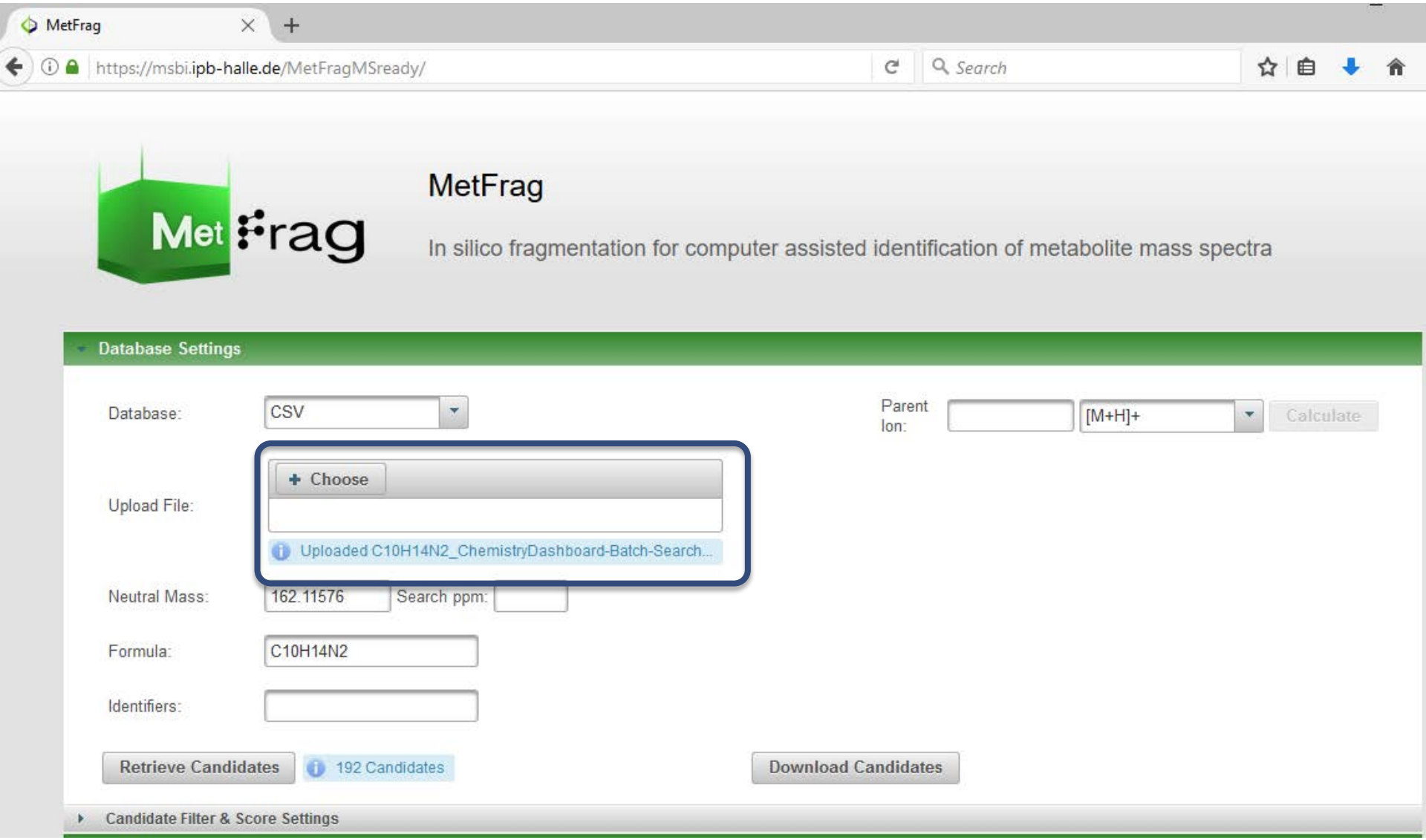
https://comptox.epa.gov/dashboard/dsstoxdb/batch_search

MS-ready Batch Search Output … input for MetFrag

A	B	C	D	E	F	G	H	I	J	K	L	M
INPUT	FOUND_BY	DTXCID_IN	FORMULA_	SMILES_IN	MAPPED_D	PREFERRE	CASRN_D	FORMULA_	SMILES_M/	MS_READY	INCHI_STR	INCHIKEY
C10H14N2	MS Ready F	DTXCID902	C10H14N2	CN1CCCC1	DTXSID102	Nicotine	54-11-5	C10H14N2	CN1CCC[C	CN1CCCC1	InChI=1/C1(	SNICXCGA
C10H14N2	MS Ready F	DTXCID502	C10H14N2	C1CCC(NC	DTXSID904	L-3-(2'-Piper	494-52-0	C10H14N2	C1CC[C@H	C1CCC(NC	InChI=1/C1(	MTXSIJUG
C10H14N2	MS Ready F	DTXCID903	C10H14N2	C1CN(CCN	DTXSID805	Phenylpiper	92-54-6	C10H14N2	C1CN(CCN	C1CN(CCN	InChI=1S/C	YZTJYBJC
C10H14N2	MS Ready F	DTXCID901	C10H14N2	CNC=NC1=	DTXSID103	N'-(2,4-Dime	33089-74-6	C10H14N2	CNC=NC1=	CNC=NC1=	InChI=1S/C	JIIOLEGNE
C10H14N2	MS Ready F	DTXCID902	C10H14N2	CN1CCCC1	DTXSID802	Nicotine sul	65-30-5	C20H30N4C	OS(O)(=O)=	CN1CCCC1	InChI=1/C1(	SNICXCGA
C10H14N2	MS Ready F	DTXCID902	C10H14N2	CN1CCCC1	DTXSID004	D-Nicotine	25162-00-9	C10H14N2	CN1CCC[C	CN1CCCC1	InChI=1/C1(	SNICXCGA
C10H14N2	MS Ready F	DTXCID902	C10H14N2	CN1CCCC1	DTXSID304	DL-Nicotine	22083-74-5	C10H14N2	CN1CCCC1	CN1CCCC1	InChI=1/C1(	SNICXCGA
C10H14N2	MS Ready F	DTXCID901	C10H14N2	CNC=NC1=	DTXSID101	Semiamitraz	51550-40-4	C10H15ClN	Cl.CNC=NC	CNC=NC1=	InChI=1S/C	JIIOLEGNE
C10H14N2	MS Ready F	DTXCID902	C10H14N2	CN1CCCC1	DTXSID602	Nicotine hyc	2820-51-1	C10H15ClN	Cl.CN1CCC	CN1CCCC1	InChI=1/C1(	SNICXCGA
C10H14N2	MS Ready F	DTXCID303	C10H14N2	CC\C(C)=N\	DTXSID905	Phenylhydra	1129-62-0	C10H14N2	CC\C(C)=N\	CCC(C)N=N	InChI=1S/C	BPAJPMUI
C10H14N2	MS Ready F	DTXCID501	C10H14N2	C(N1CCCC	DTXSID901	N-(3-Pyridyl	370-09-2	C10H14N2	C(N1CCCC	C(N1CCCC	InChI=1S/C	YFCWBKE
C10H14N2	MS Ready F	DTXCID503	C10H14N2	CN(C)C=NC	DTXSID406	Methanimida	10278-71-4	C10H14N2	CN(C)C=NC	CN(C)C=NC	InChI=1S/C	VTJDEAHI
C10H14N2	MS Ready F	DTXCID902	C10H14N2	CN1CCCC1	DTXSID507	Benzoic acid	29790-52-1	C17H20N2C	OC(=O)C1=	CN1CCCC1	InChI=1/C1(	SNICXCGA
C10H14N2	MS Ready F	DTXCID903	C10H14N2	C1CN(CCN	DTXSID401	Piperazine,	2210-93-7	C10H15ClN	Cl.C1CN(C	C1CN(CCN	InChI=1S/C	YZTJYBJC
C10H14N2	MS Ready F	DTXCID801	C10H14N2	C1CCN(CC	DTXSID201	4-Piperidylp	2767-90-0	C10H14N2	C1CCN(CC	C1CCN(CC	InChI=1S/C	MTPBUCC
C10H14N2	MS Ready F	DTXCID903	C10H14N2	C1CN(CCN	DTXSID401	1-Phenylpip	4004-95-9	C10H16Cl2I	Cl.Cl.C1CN	C1CN(CCN	InChI=1S/C	YZTJYBJC
C10H14N2	MS Ready F	DTXCID304	C10H14N2	CCCC1=CC	DTXSID307	Benzaldehy	93480-03-6	C10H14N2	CCCC1=CC	CCCC1=CC	InChI=1S/C	IAWNEZM
C10H14N2	MS Ready F	DTXCID902	C10H14N2	CN1CCCC1	DTXSID202	(S)-Nicotine	88660-51-9	C22H30N2C	CN1CCC[C	CN1CCCC1	InChI=1/C1(	SNICXCGA
C10H14N2	MS Ready F	DTXCID902	C10H14N2	CN1CCCC1	DTXSID308	Pyridine, 3-[	6019-02-9	C10H16Cl2I	Cl.Cl.CN1C	CN1CCCC1	InChI=1/C1(	SNICXCGA
C10H14N2	MS Ready F	DTXCID502	C10H14N2	C1CCC(NC	DTXSID201	Anabasine (	31945-06-9	C10H16Cl2I	Cl.Cl.C1CC	C1CCC(NC	InChI=1/C1(	MTXSIJUG
C10H14N2	MS Ready F	DTXCID401	C10H14N2	C=C(CCCC	DTXSID801	5-Methylene	50592-61-5	C10H14N2	C=C(CCCC	C=C(CCCC	InChI=1S/C	KKWHJFK
C10H14N2	MS Ready F	DTXCID301	C10H14N2	CC1=C2NC	DTXSID802	5,8-Dimethy	66102-39-4	C10H14N2	CC1=C2NC	CC1=C2NC	InChI=1S/C	DJVBBBRH
C10H14N2	MS Ready F	DTXCID002	C10H14N2	CN(C)\C=N\	DTXSID503	Formamidin	2305-75-1	C10H14N2	CN(C)\C=N\	CN(C)C=NC	InChI=1S/C	NCXQXIRX
C10H14N2	MS Ready F	DTXCID006	C10H14N2	CC1=CC2=	DTXSID401	Quinoxaline,	10579-68-7	C10H14N2	CC1=CC2=	CC1=CC2=	InChI=1S/C	DPSZEBZ
C10H14N2	MS Ready F	DTXCID901	C10H14N2	CN1CCN(C	DTXSID001	Quinoxaline,	2427-06-7	C10H14N2	CN1CCN(C	CN1CCN(C	InChI=1S/C	XGJGSZQ
C10H14N2	MS Ready F	DTXCID502	C10H14N2	C1CCC(NC	DTXSID301	(S)-3-(2-Pip	37520-45-9	C22H20N8C	C1CC[C@H	C1CCC(NC	InChI=1/C1(	MTXSIJUG
C10H14N2	MS Ready F	DTXCID501	C10H14N2	CC1=C(C)C	DTXSID202	5H-1-Pyrind	62732-46-1	C10H14N2	CC1=C(C)C	CC1=C(C)C	InChI=1S/C	XCACCLN
C10H14N2	MS Ready F	DTXCID102	C10H14N2	CCN\C(=N/(	DTXSID902	Benzenecar	64593-93-7	C10H15ClN	Cl.CCN\C(=	CCN\C(=N/(	InChI=1S/C	XPJYNRPF
C10H14N2	MS Ready F	DTXCID502	C10H14N2	CN\C(=N/C)	DTXSID802	Benzenecar	64594-06-5	C10H15ClN	Cl.CN\C(=N	CN\C(=N/C)	InChI=1S/C	NWPZITRZ
C10H14N2	MS Ready F	DTXCID301	C10H14N2	CCN1CCNC	DTXSID302	Quinoxaline,	73855-46-6	C10H14N2	CCN1CCNC	CCN1CCNC	InChI=1S/C	KIDFEKJW
C10H14N2	MS Ready F	DTXCID903	C10H14N2	C1CN(CCN	DTXSID102	Iron, (ethan	75079-25-3	C22H28FeN	[Fe++].[O-]C	C1CN(CCN	InChI=1S/C	YZTJYBJC
C10H14N2	MS Ready F	DTXCID902	C10H14N2	CN1CCCC1	DTXSID102	Pyridine, 3-(	80604-52-0	C10H12T2N	[3H]C1([3H]	-	InChI=1/C1(	SNICXCGA
C10H14N2	MS Ready F	DTXCID601	C10H14N2	CN(C)C1=C	DTXSID302	p-Dimethyla	82027-08-5	C10H14N2	CN(C)C1=C	CN(C)C1=C	InChI=1S/C	FSTZGHIB

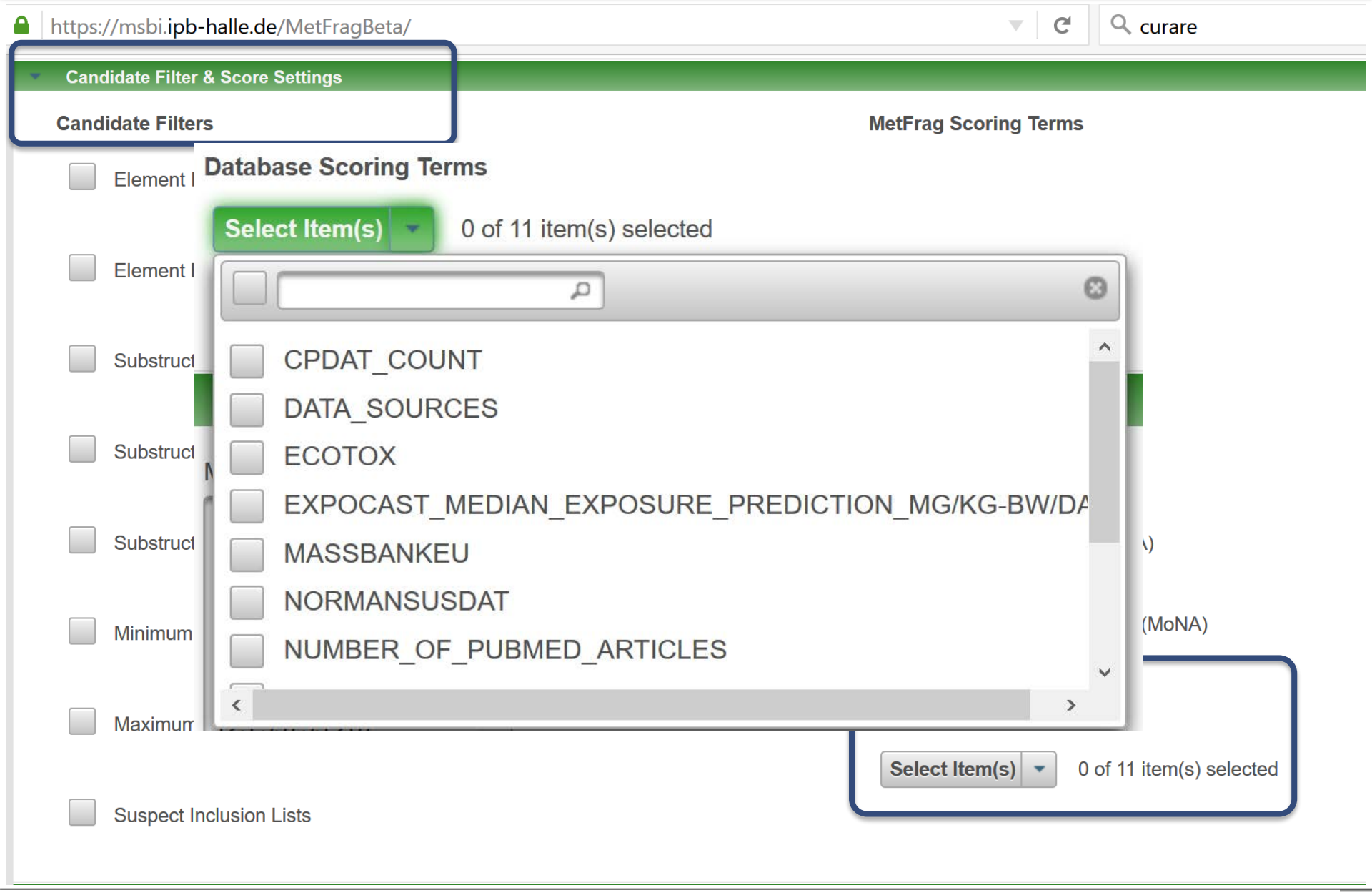
https://comptox.epa.gov/dashboard/dsstoxdb/batch_search

Uploading MS-ready Batch Search to MetFragBeta



https://msbi.ipb-halle.de/MetFragBeta/

Make sure you choose your metadata!



https://msbi.ipb-halle.de/MetFragBeta/

Take an example Mass Spectrum from MassBank

MetFrag

https://msbi.ipb-halle.de/MetFragMSready/

Search

Fragmentation Settings & Processing

Mzppm: 5

Mzabs: 0.001

Mode: [M+H]+

Tree depth: 2

Group candidates

MS/MS Peak list

```
117.0571 43503063 349
118.0649 392125 3
119.0726 230767.3 1
120.0806 21877380.1 175
130.065 116905175 937
131.073 880898.8 7
132.0806 124509966.1
999
134.0962 389380.4 3
146.0961 136376.5 1
163.1227 38578752.8 309
```

Show Spectrum

Process Candidates

192 Candidates processed

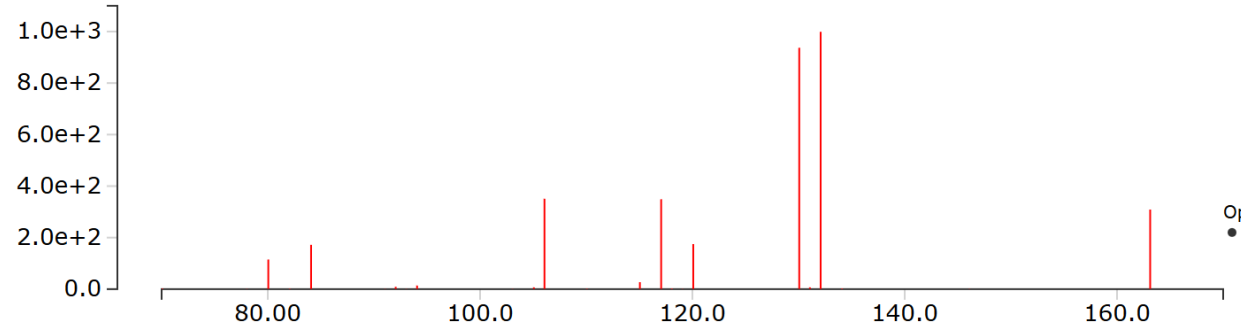
Download Parameters

https://massbank.eu/MassBank/jsp/RecordDisplay.jsp?id=EQ300804&dsn=Eawag

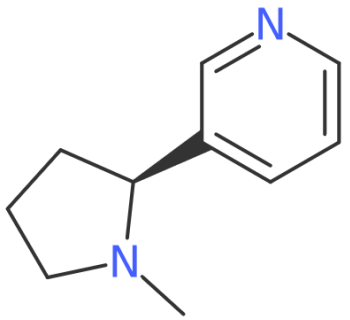
Search

Nicotine; LC-ESI-QFT; MS2; CE: 60; R=35000; [M+H]+

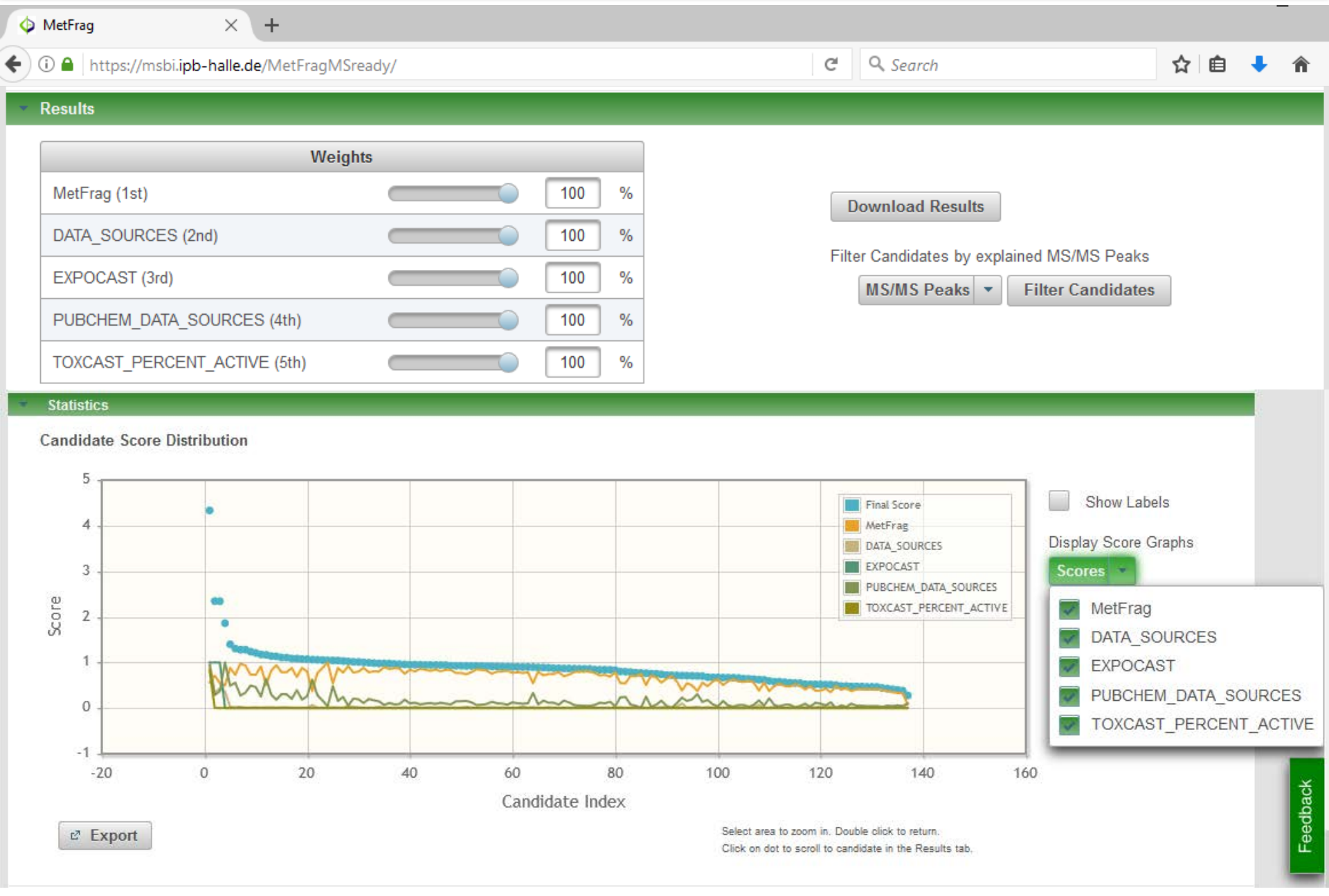
Mass Spectrum



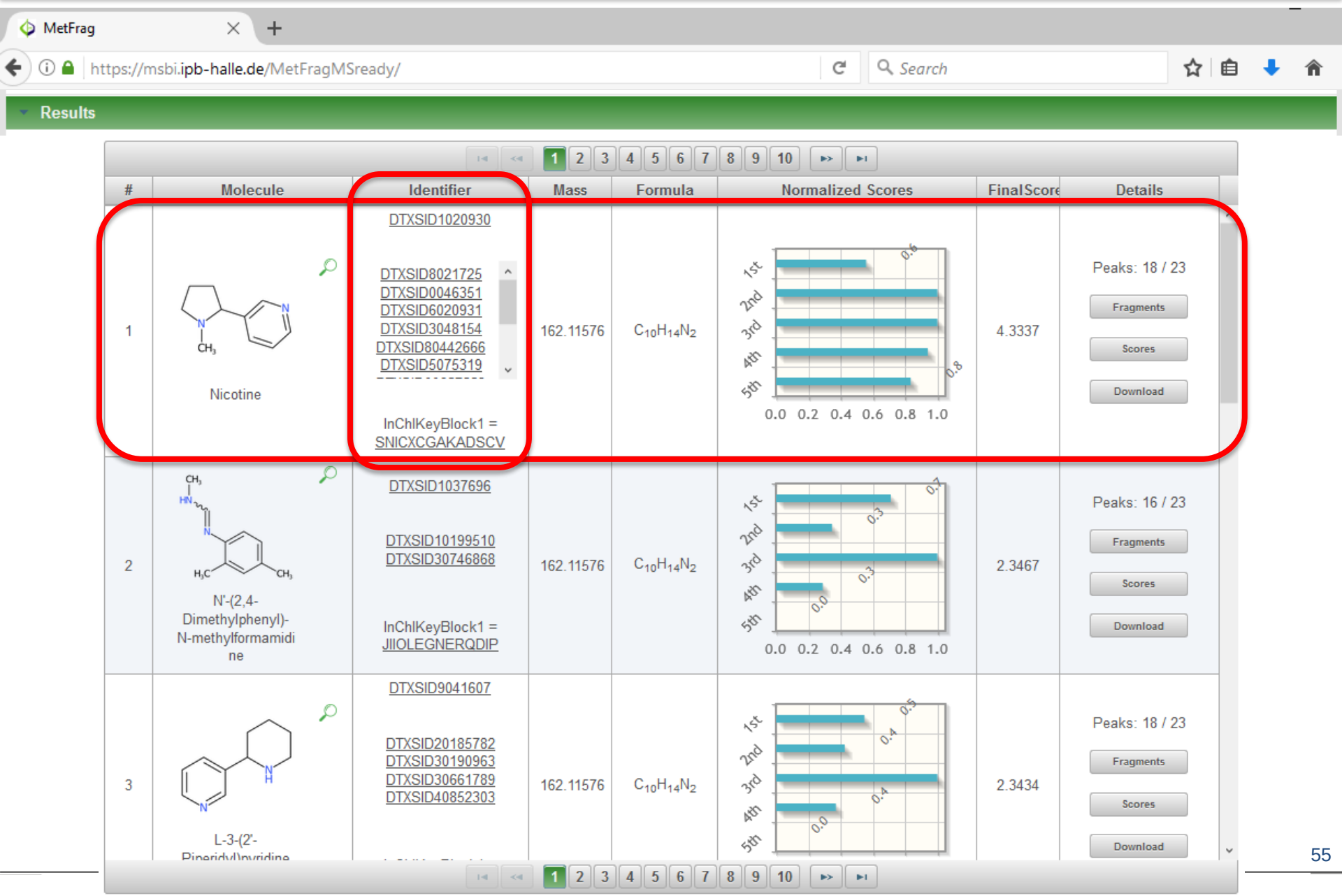
Chemical Structure



…and you have Dashboard results in MetFrag



Default: Candidates grouped by InChIKey First Block



Don’t want them grouped? Deselect “Group Candidates”

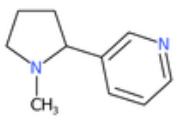
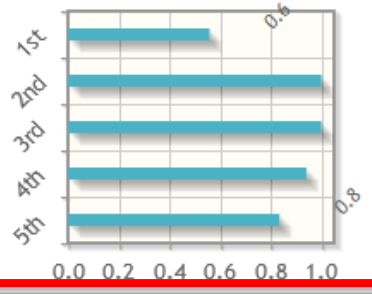
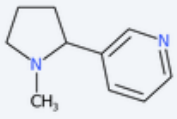
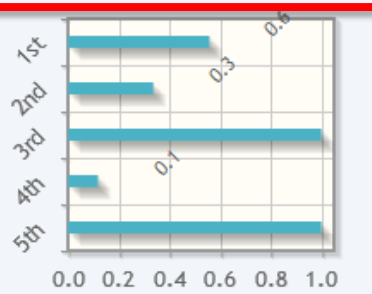
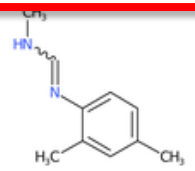
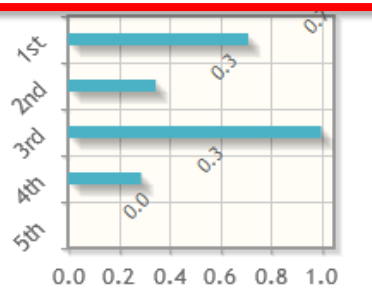
MetFrag

https://msbi.ipb-halle.de/MetFragMSready/

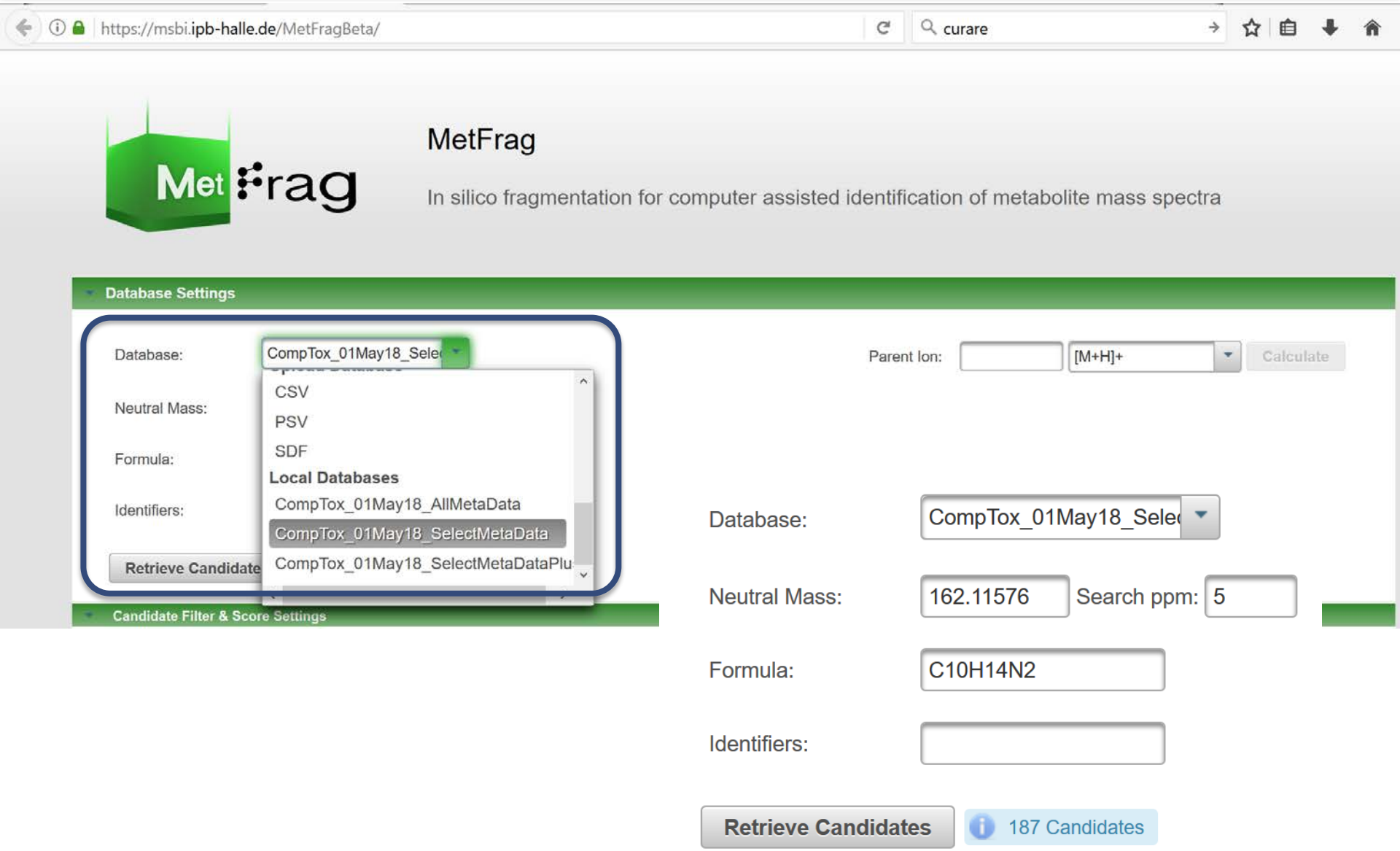
Search

Results

1 2 3 4 5 6 7 8 9 10

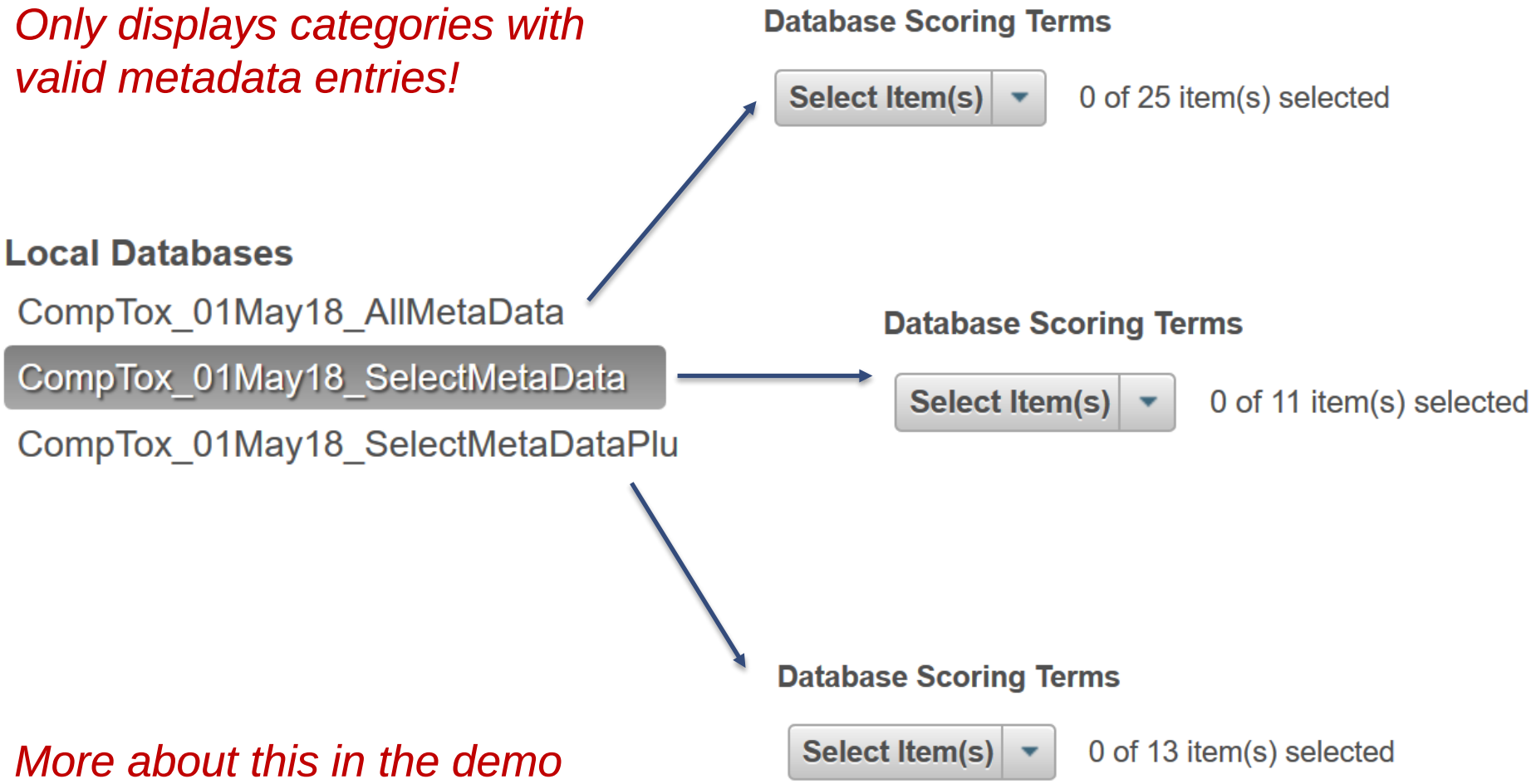
#	Molecule	Identifier	Mass	Formula	Normalized Scores	FinalScor	Details
1	 Nicotine	DTXSID1020930 InChIKeyBlock1 = SNICXCGAKADSCV	162.11576	$C_{10}H_{14}N_2$		4.3337	Peaks: 18 / 23 Fragments Scores Download
2	 Nicotine sulfate	DTXSID8021725 InChIKeyBlock1 = SNICXCGAKADSCV	162.11576	$C_{10}H_{14}N_2$		3.0136	Peaks: 18 / 23 Fragments Scores Download
3	 N'-(2,4-Dimethylphenyl)-N-methylformamidine	DTXSID1037696 InChIKeyBlock1 = JIIOLEGNERQDIP	162.11576	$C_{10}H_{14}N_2$		2.3467	Peaks: 16 / 23 Fragments Scores Download

Want to do this in MetFrag directly?



Metadata Selection Options

Only displays categories with valid metadata entries!



More about this in the demo session tomorrow!

Many Scoring Options Now Combined in MetFrag

https://msbi.ipb-halle.de/MetFragBeta/ Search

Results

Weights		
MetFrag (1st)	100	%
ExactSpectralSimilarity (2nd)	100	%
NORMANSUSDAT (3rd)	100	%
PUBCHEM_DATA_SOURCES (4th)	100	%
TOXCAST_PERCENT_ACTIVE (5th)	100	%

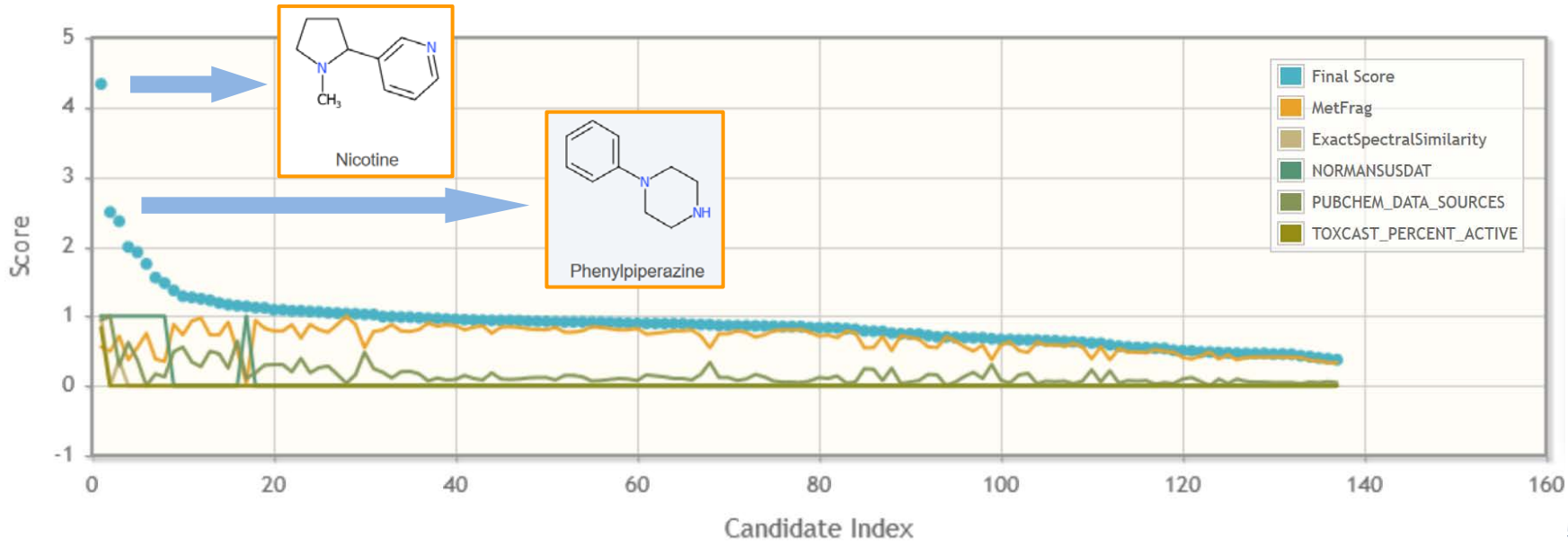
MetFrag → MassBank.eu

norman suspects

EPA Chemistry Dashboard

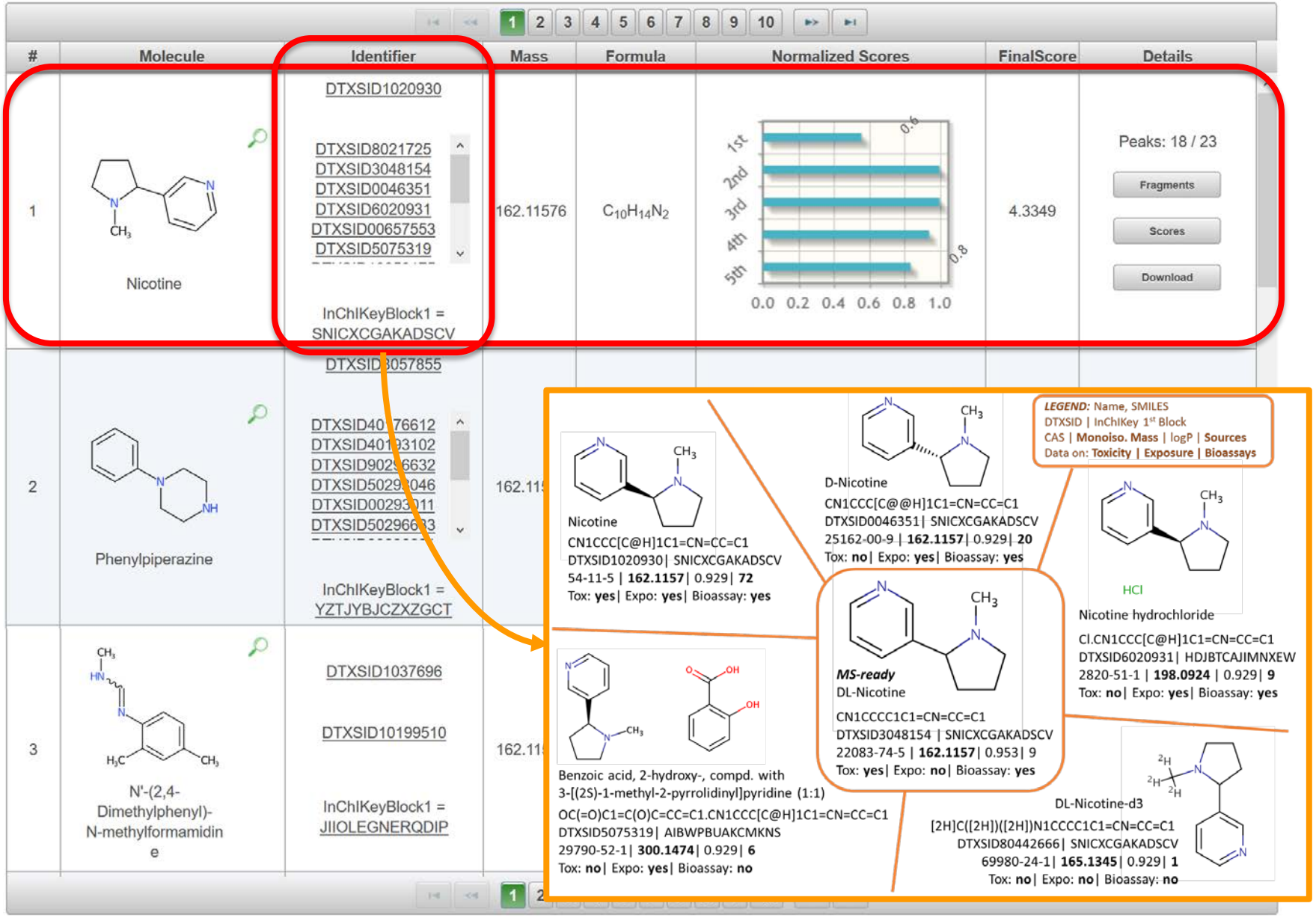
Statistics

Candidate Score Distribution



https://msbi.ipb-halle.de/MetFragBeta/ AND https://comptox.epa.gov/dashboard/dsstoxdb/batch_search (MetFrag Export)

…with the links back to the original mixtures …

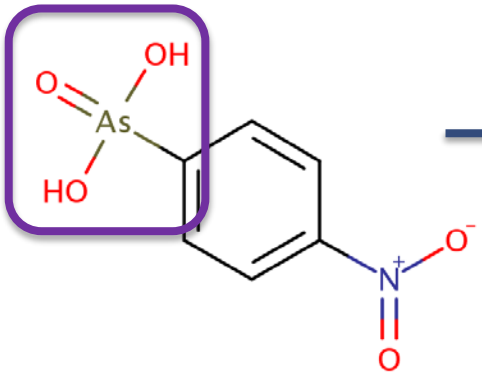


Example from: Schymanski & Williams, 2017, ES&T, 51 (10), pp 5357–5359. DOI: 10.1021/acs.est.7b01908

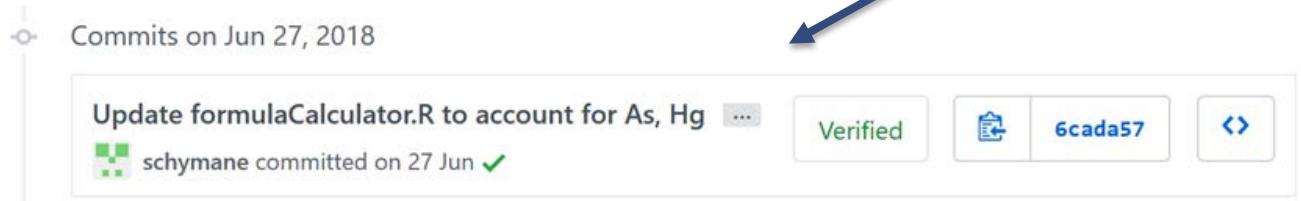
Part 3: Metals or no metals?

- Do we see them in MS?
- Not in QSAR-Ready form
- Fail most Fragmenters

o 2Ms: Metals or no Metals?



Blinded run: Error in RMassBank

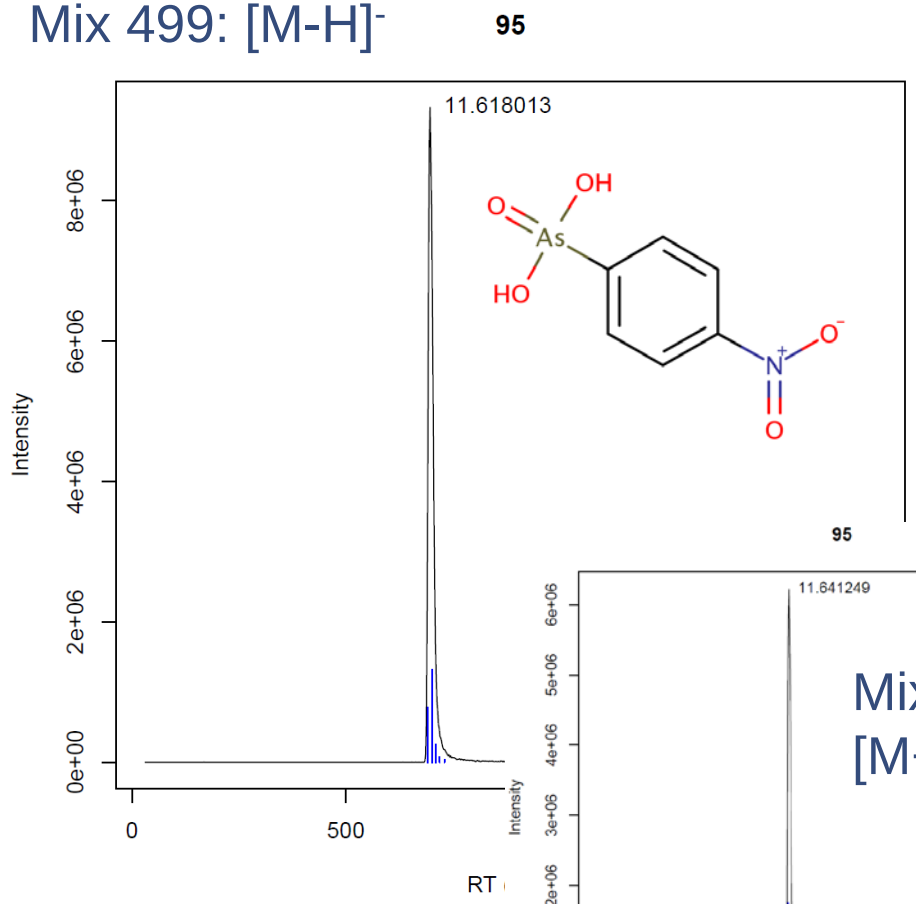


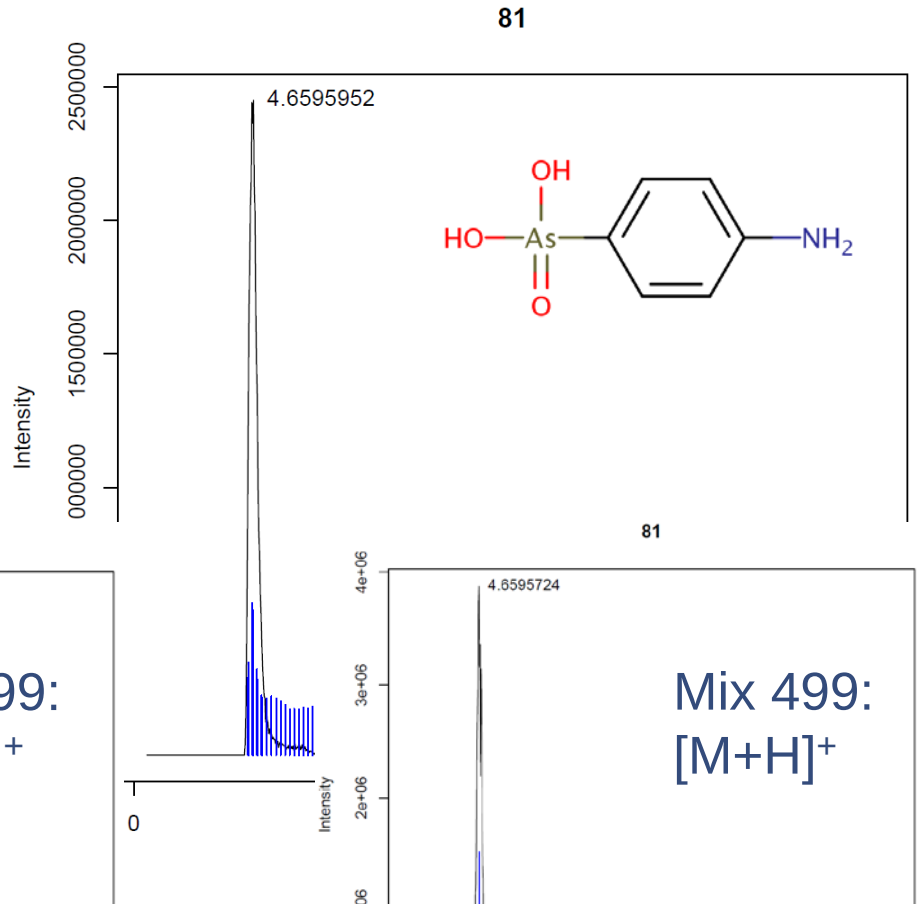
https://github.com/MassBank/RMassBank/commits/master

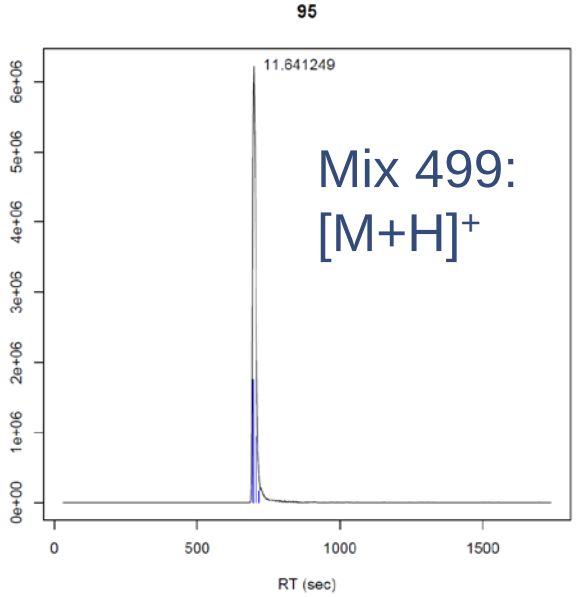
As-containing Compounds Clear in Both Modes

eawag aquatic research

Mix 499: $[M-H]^-$





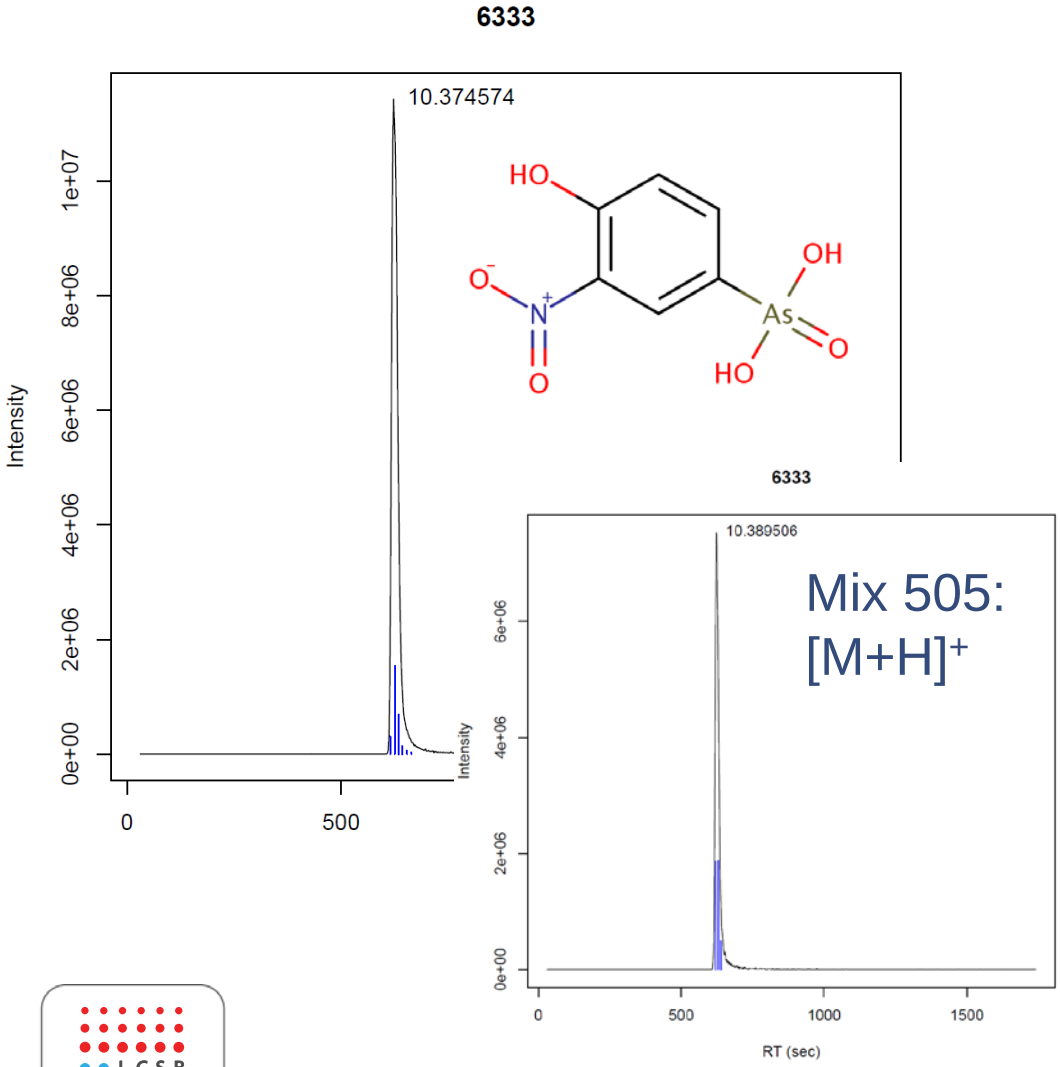


Mix 499: $[M+H]^+$

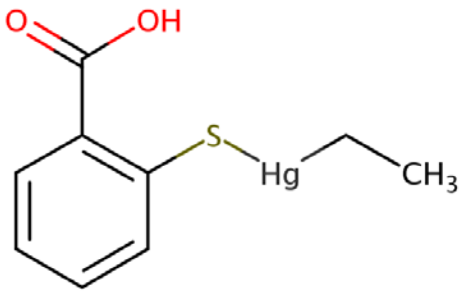
Mix 499: $[M+H]^+$

As-containing Compounds Clear in Both Modes

Mix 505: $[M-H]^-$

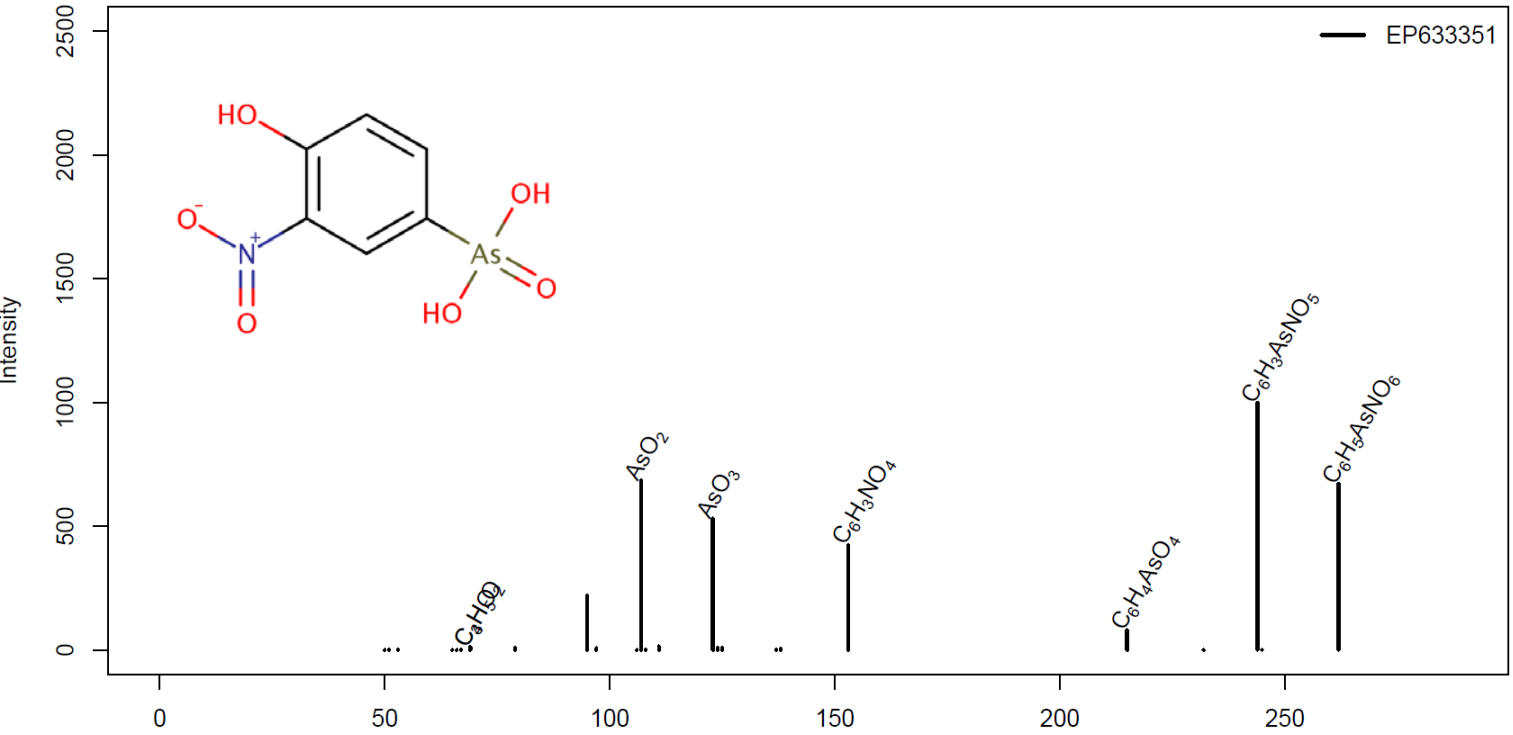


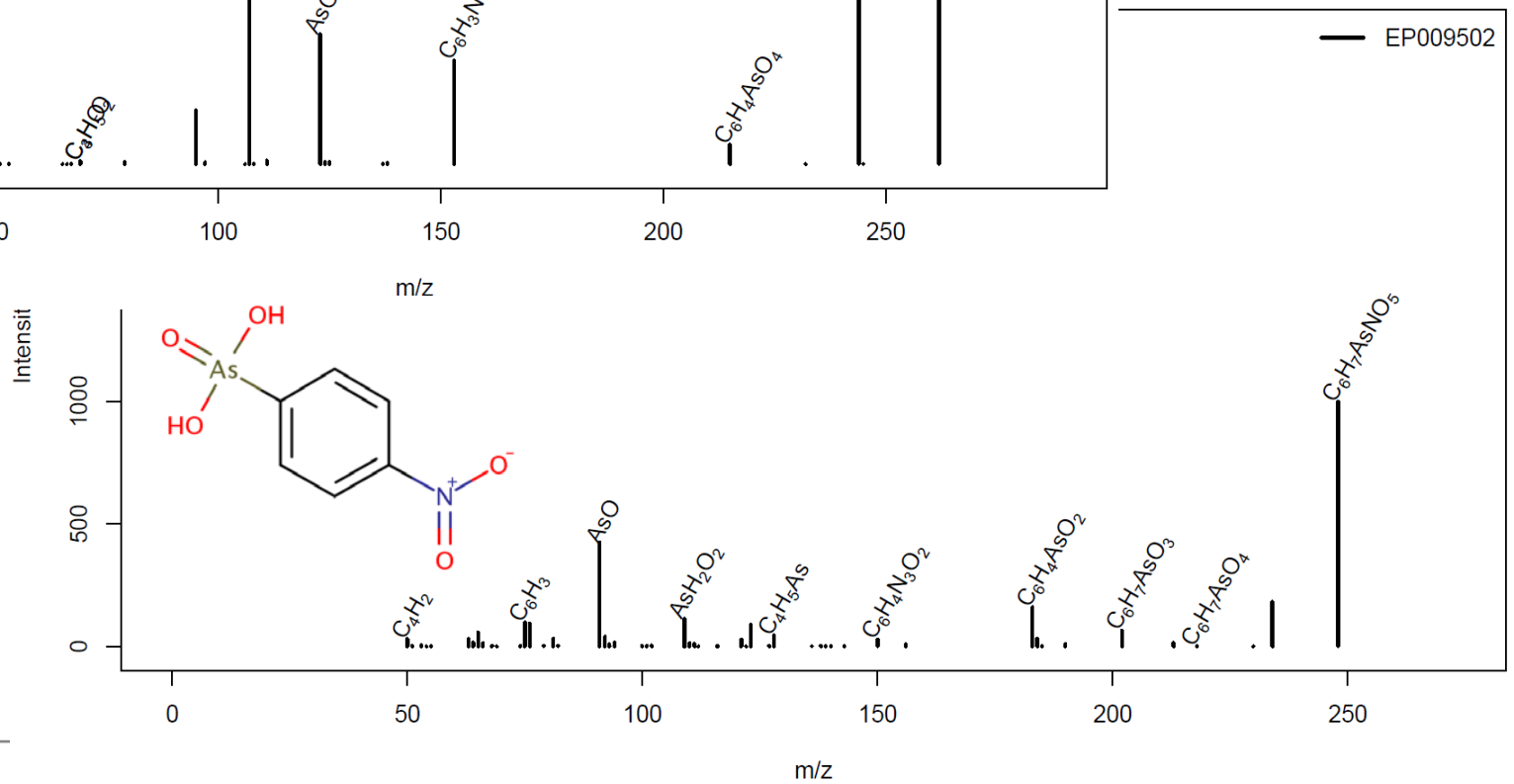
…but the Hg-containing Thimerosal was not observed [possible in source fragment?]



As-containing Annotated MS/MS Spectra …

Plotting functions available in ReSOLUTION

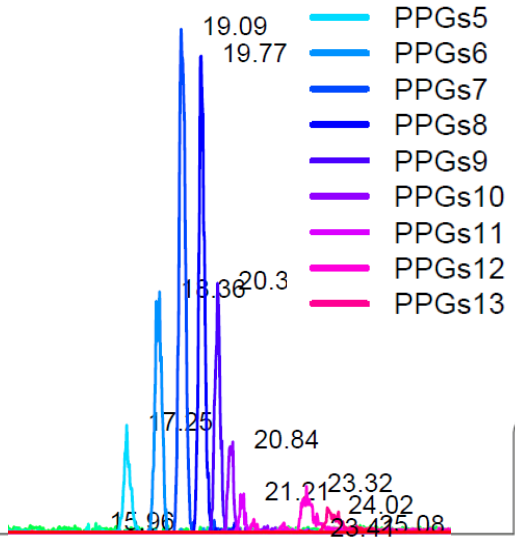




Part 3: Mixtures…

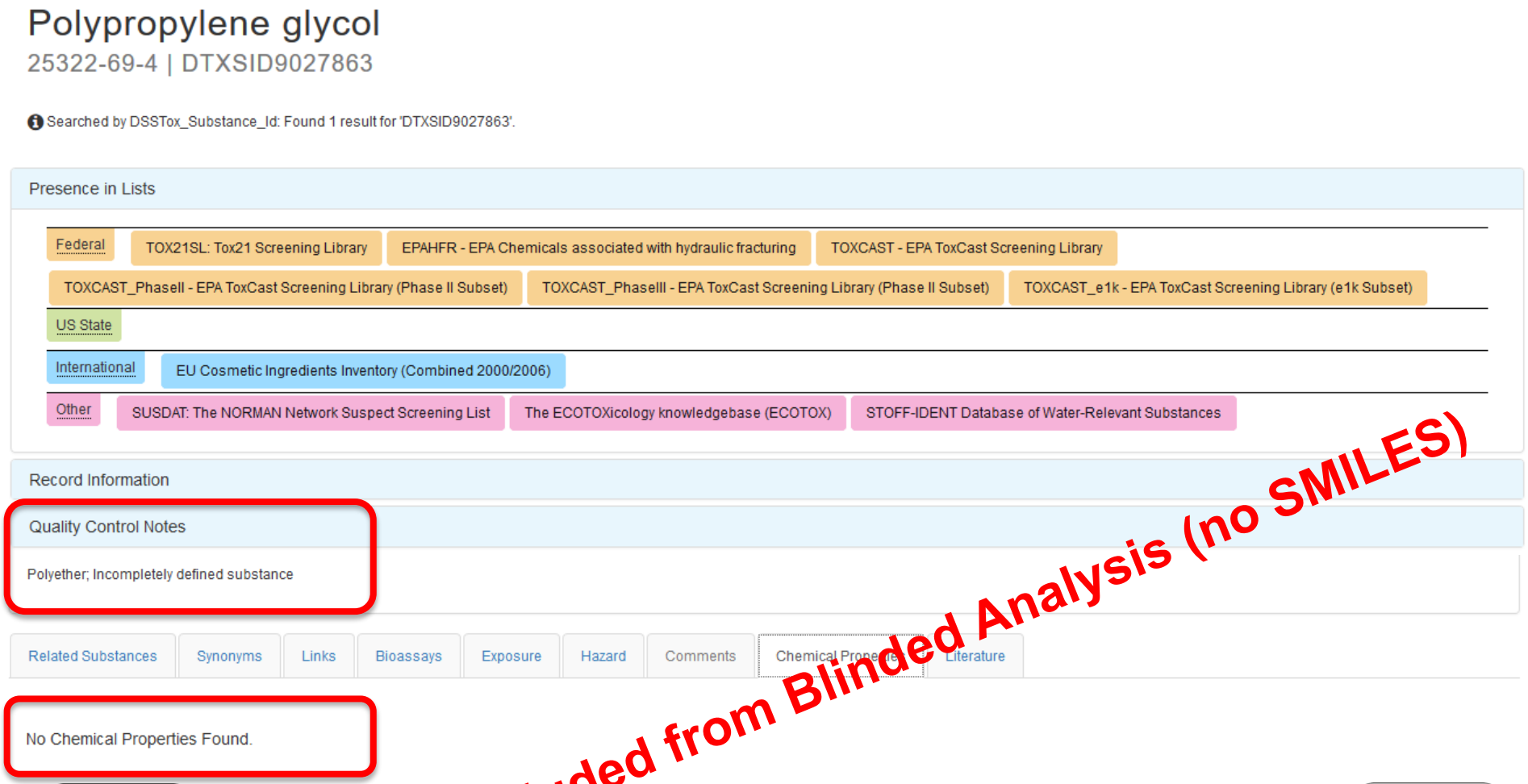
- 1M: Mixtures…

DTXSID9027863 - PPGs



PPGs – Capturing PPGs in Mix 508

- Starting point for ENTACT … no structures, little information



Excluded from Blinded Analysis (no SMILES)

https://comptox.epa.gov/dashboard/dsstoxdb/results?utf8=%E2%9C%93&search=DTXSID9027863 [public version 16/07/2018]

PPGs – Capturing Mixtures

- Dear Tony and Chris …

  do you have more information?

Substance Identifier "25322-69-4 " > substances (1) > **25322-69-4**

SUBSTANCE DETAIL | Get References | Get Reactions | Get Commercial Sources

Return

CAS Registry Number 25322-69-4

~35,243 ~60

$(C_3 H_6 O)_n H_2 O$

Poly[oxy(methyl-1,2-ethanediyl)], α-hydro-ω-hydroxy-

Incompletely Defined Substance, Polymer

Polymer Class Terms

Polyether

Melting Point (Experimental)

Value: 50-70 °C

Boiling Point (Experimental)

Value: >300 °C

Density (Experimental)

Value: 1.002-1.007 g/cm3

Other Names

Glycols, polypropylene (8CI)

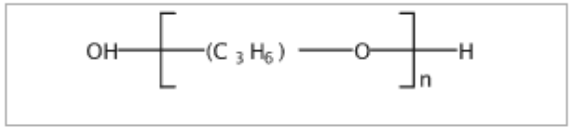
1,2-Epoxypropane polymer

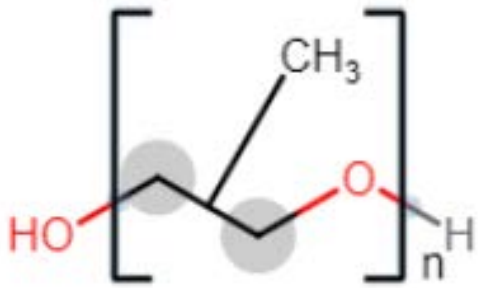
1,2-Propanediol, homopolymer

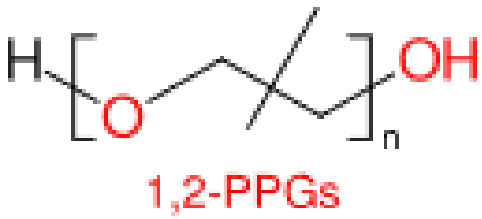
1,2-Propylene glycol-propylene oxide polymer

3600K

View more...







1,2-PPGs

[H]OCCO.C* |lp:1:2,4:2,m:6:3.2,Sg:n:5,1,2,3::ht|

PPGs – Capturing Mixtures

- Generating base case SMILES in RChemMass
- https://github.com/schymane/RChemMass/

```
23  # try with one 1,2 variant
24  genSmiles12PPG1 <- "[R]O"
25  R1toN_12PPG1 <- "(OC(C)C)"
26  nR1toN_PPG <- "(0-15)"
27
28  smiles_12PPG1_0to16 <- buildSmiles(genSmilesPPG,R1toN_12PPG1,nR1toN_PPG)
29  smiles_12PPG1_0to16
30  write.table(smiles_12PPG1_0to16,"smiles_12PPG1_0to16.txt",row.names = F, col.names = F,quote = F)
31
```

C:/DATA/External/US_EPA/ENTACT/ENTACT_R/PPGs_selectFiles/

```
> smiles_12PPG1_0to16 <- buildSmiles(genSmilesPPG,R1toN_12PPG1,nR1toN_PPG)
> smiles_12PPG1_0to16
 [1] "OC(C)CO"
 [2] "OC(C)COC(C)CO"
 [3] "OC(C)COC(C)COC(C)CO"
 [4] "OC(C)COC(C)COC(C)COC(C)CO"
 [5] "OC(C)COC(C)COC(C)COC(C)COC(C)CO"
 [6] "OC(C)COC(C)COC(C)COC(C)COC(C)COC(C)CO"
 [7] "OC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)CO"
 [8] "OC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)CO"
 [9] "OC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)CO"
[10] "OC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)CO"
[11] "OC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)CO"
[12] "OC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)CO"
[13] "OC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)CO"
[14] "OC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)CO"
[15] "OC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)CO"
[16] "OC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)COC(C)CO"
```

PPGs – Capturing Mixtures

o Work in Progress – Related Structures

Polypropylene glycol
25322-69-4 | DTXSID9027863
Searched by CAS-RN.

18 chemicals

Download / Send | Sort by: Relationship | DTXSID | CASRN | TOXCAST | Filter by: Name or CASRN | Hide | Select all

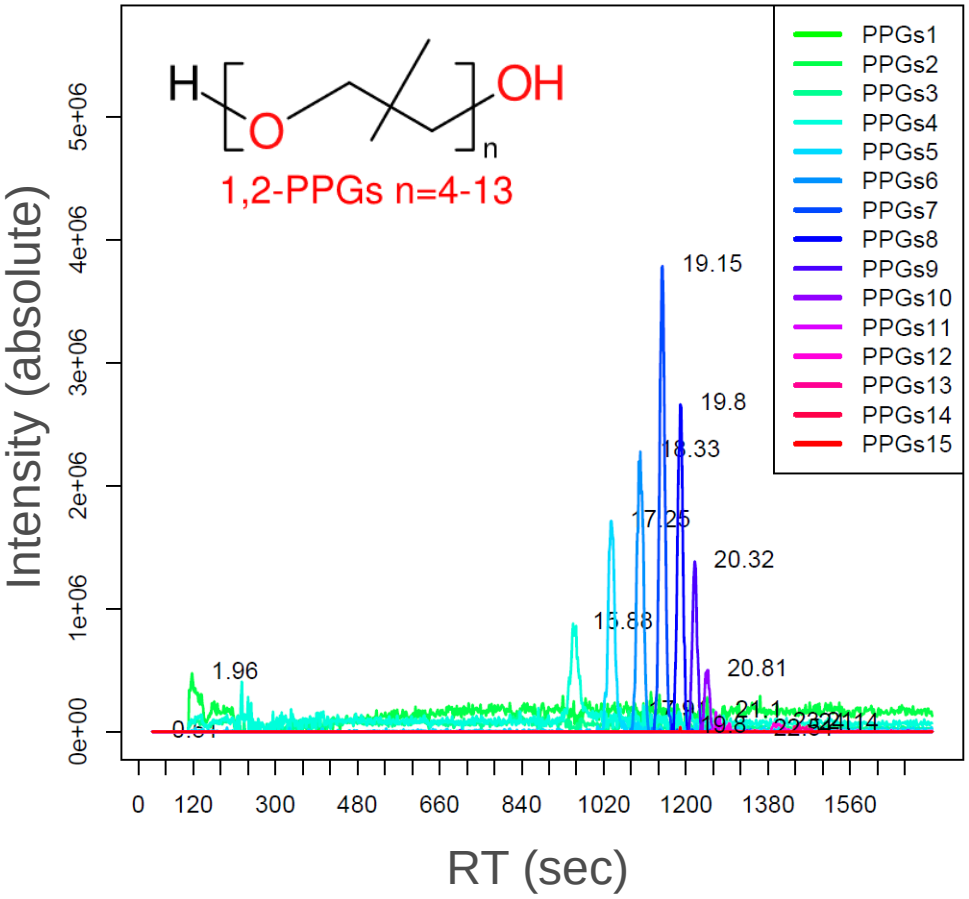
Searched Chemical	Component	Component	Component	Component	Component
Polypropylene glycol DTXSID: DTXSID9027863 CASRN: 25322-69-4 TOXCAST: 3/299	2,5,8,11,14,17,20-heptamethyl-3,6,9,12... DTXSID: DTXSID80903904 CASRN: 45308-36-9 TOXCAST: 0	2,5,8,11,14,17,20,23,26,29,32,35,38,41... DTXSID: DTXSID10903912 CASRN: NOCAS_903912 TOXCAST: 0	2-[2-[2-(2-Hydroxypropoxy)propoxy]pro... DTXSID: DTXSID7029324 CASRN: 24800-25-7 TOXCAST: 0	1-propanol, 2-(2-hydroxypropoxy)- DTXSID: DTXSID30861722 CASRN: 106-62-7 TOXCAST: 0	2-[2-(2-Hydroxypropoxy)propoxy]propa... DTXSID: DTXSID90871342 CASRN: 45096-22-8 TOXCAST: 0
Component	Component	Component	Component	Component	Component
2,5,8,11-tetramethyl-3,6,9,12-tetraoxap... DTXSID: DTXSID00903901 CASRN: 21482-12-2 TOXCAST: 0	2,5,8,11,14-pentamethyl-3,6,9,12,15-pe... DTXSID: DTXSID60903902 CASRN: 74388-92-4 TOXCAST: 0	2,5,8,11,14,17-hexamethyl-3,6,9,12,15,... DTXSID: DTXSID20903903 CASRN: 14382-16-4 TOXCAST: 0	Tripropylene glycol DTXSID: DTXSID7027837 CASRN: 24800-44-0 TOXCAST: 1/298	2,5,8,11,14,17,20,23-octamethyl-3,6,9,1... DTXSID: DTXSID40903905 CASRN: 2172326-56-4 TOXCAST: 0	2,5,8,11,14,17,20,23,26-nonamethyl-3,... DTXSID: DTXSID00903906 CASRN: NOCAS_903906 TOXCAST: 0
Component	Component	Component	Component	Component	Monomer
2,5,8,11,14,17,20,23,26,29-decamethyl-... DTXSID: DTXSID60903907 CASRN: NOCAS_903907 TOXCAST: 0	2,5,8,11,14,17,20,23,26,29,32-undeca... DTXSID: DTXSID20903908 CASRN: NOCAS_903908 TOXCAST: 0	2,5,8,11,14,17,20,23,26,29,32,35-dode... DTXSID: DTXSID60903909 CASRN: NOCAS_903909 TOXCAST: 0	2,5,8,11,14,17,20,23,26,29,32,35,38-tri... DTXSID: DTXSID90903910 CASRN: NOCAS_903910 TOXCAST: 0	2,5,8,11,14,17,20,23,26,29,32,35,38,41... DTXSID: DTXSID50903911 CASRN: NOCAS_903911 TOXCAST: 0	1,2-Propylene glycol DTXSID: DTXSID0021206 CASRN: 57-55-6 TOXCAST: 11/539

https://comptox.epa.gov/dashboard/dsstoxdb/results?utf8=%E2%9C%93&search=DTXSID9027863 [devel version]

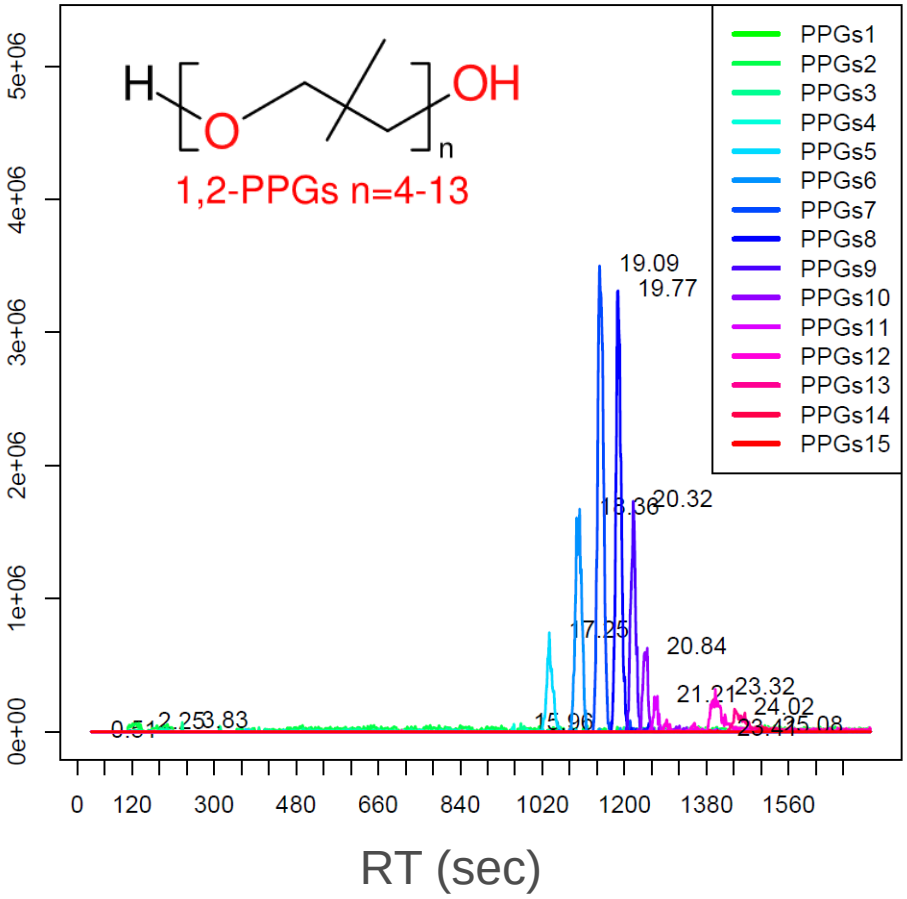
PPGs – Mix 508 Eawag Results

- Now we can start screening!

$[M+H]^+$



$[M+NH_4]^+$

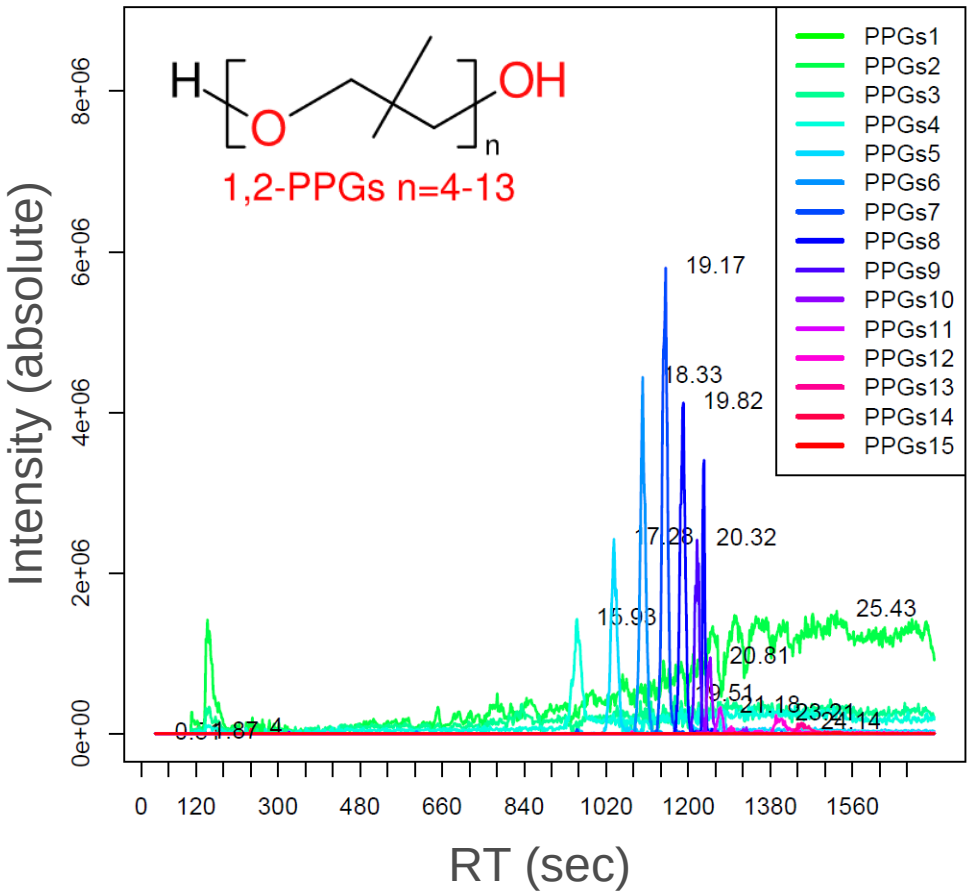


https://github.com/schymane/RChemMass/ and https://github.com/schymane/ReSOLUTION

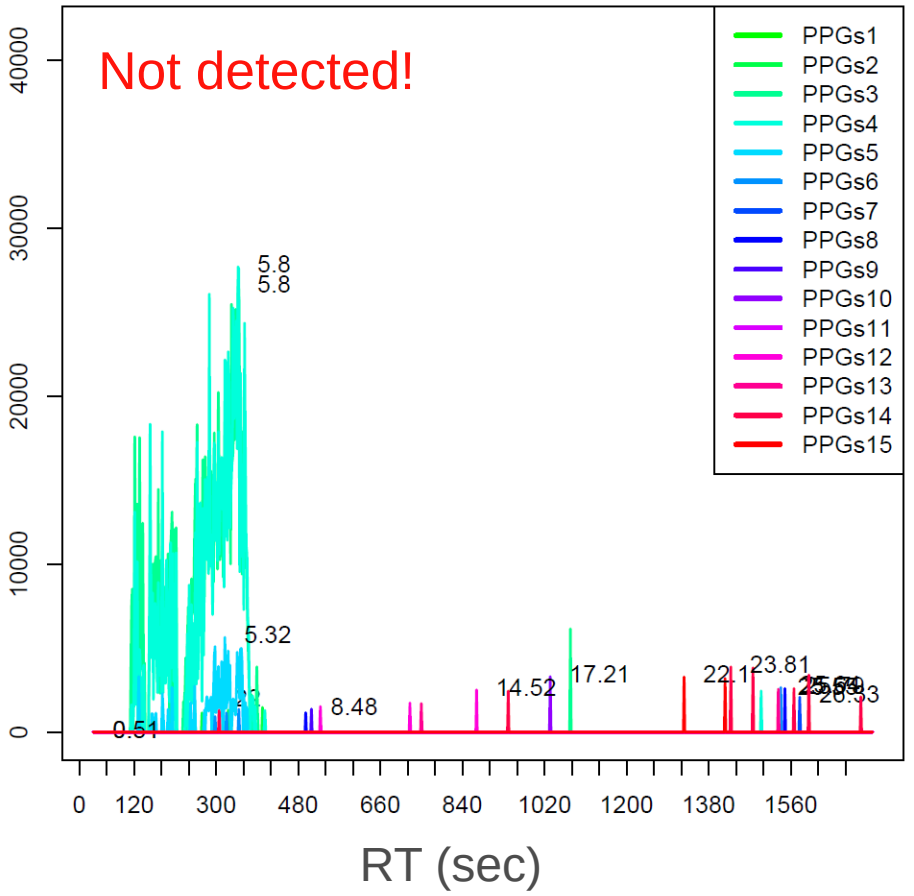
PPGs – Mix 508 Eawag Results

- More screening results…

$[M+Na]^+$



$[M-H]^-$



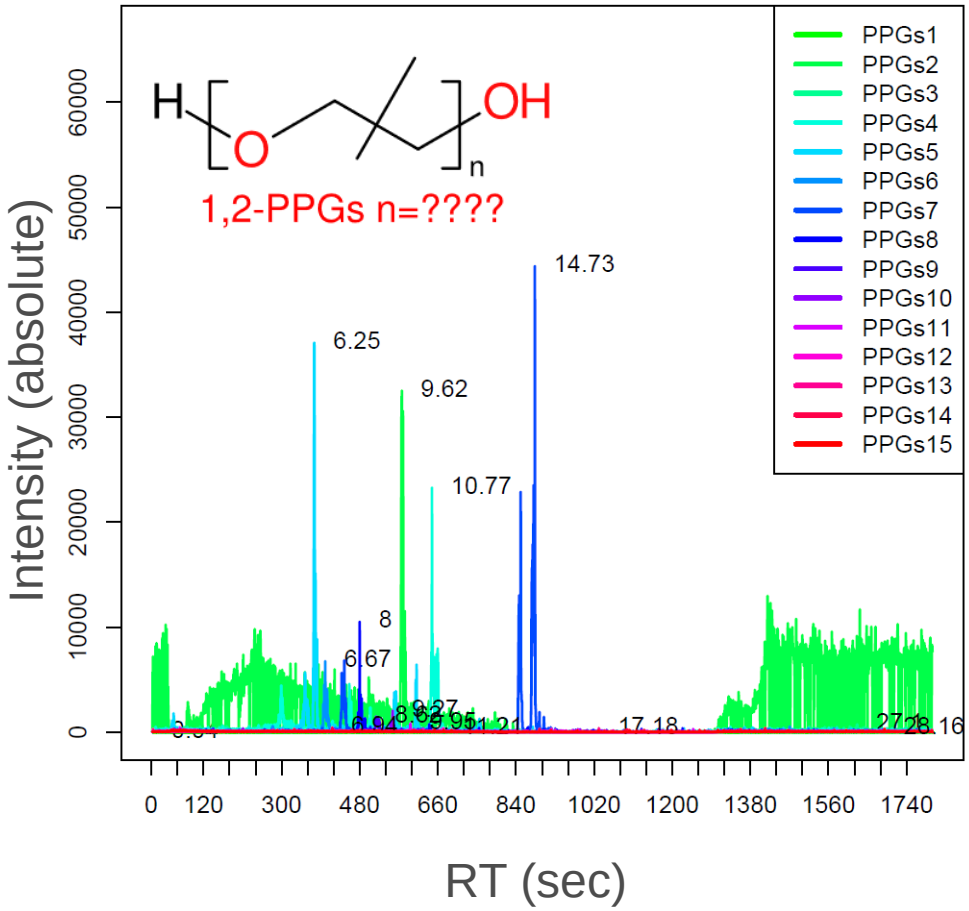
https://github.com/schymane/RChemMass/ and https://github.com/schymane/ReSOLUTION

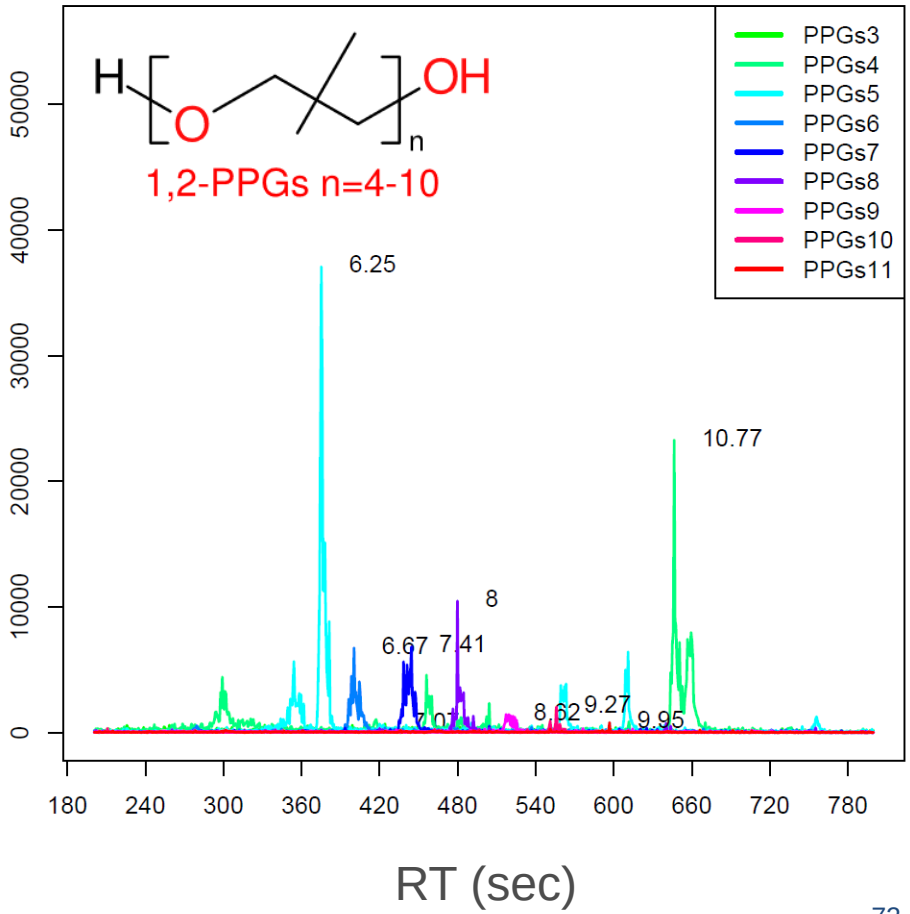
PPGs – Mix 508 EPA Results

- Jon, what did you guys see?

$[M+H]^+$

Cleaned up – reduced RT and PPG range
- Not as consistent as Eawag data

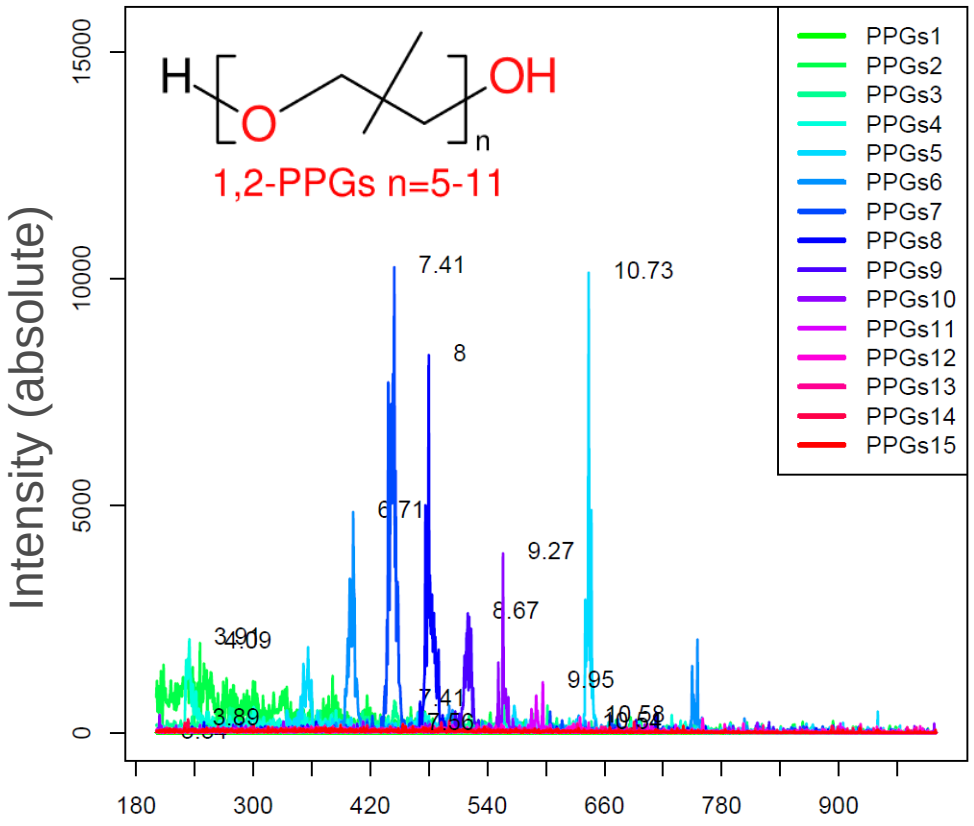




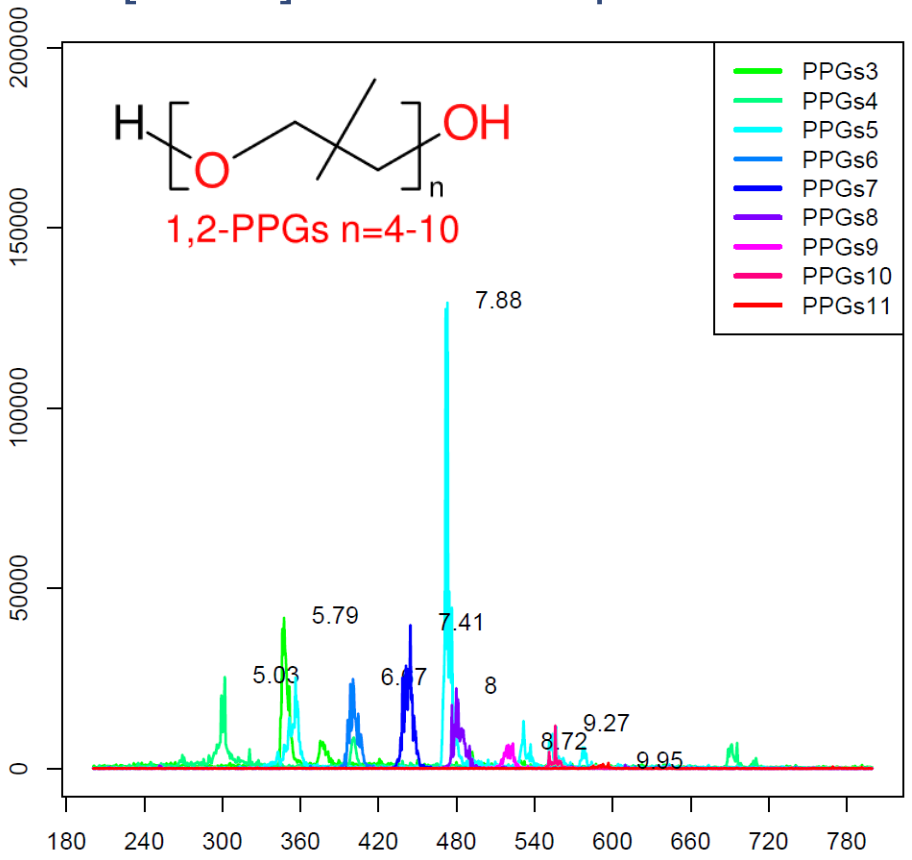
PPGs – Mix 508 EPA Results

- Patterns not as consistent … but obviously there!

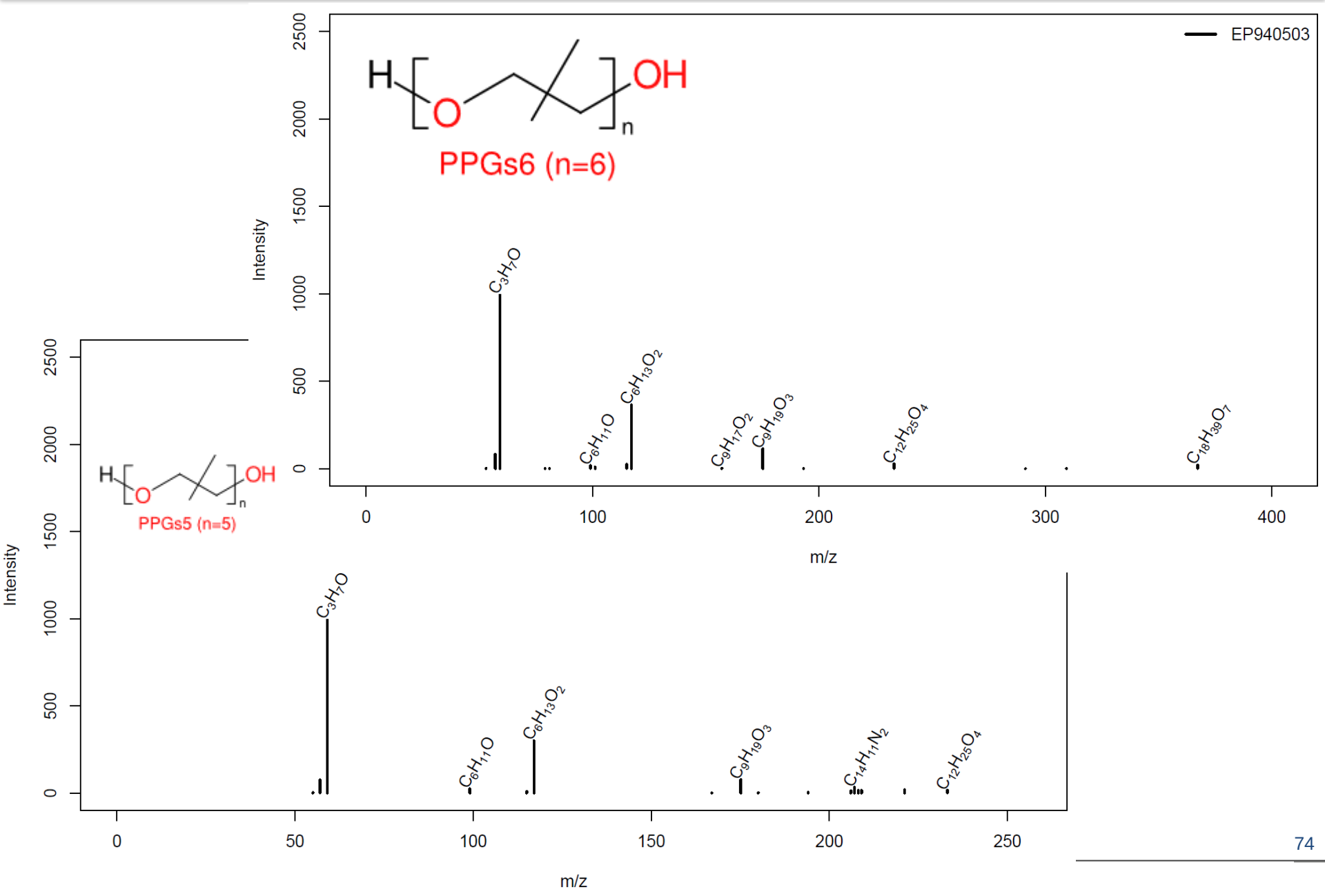
$[M+NH_4]^+$



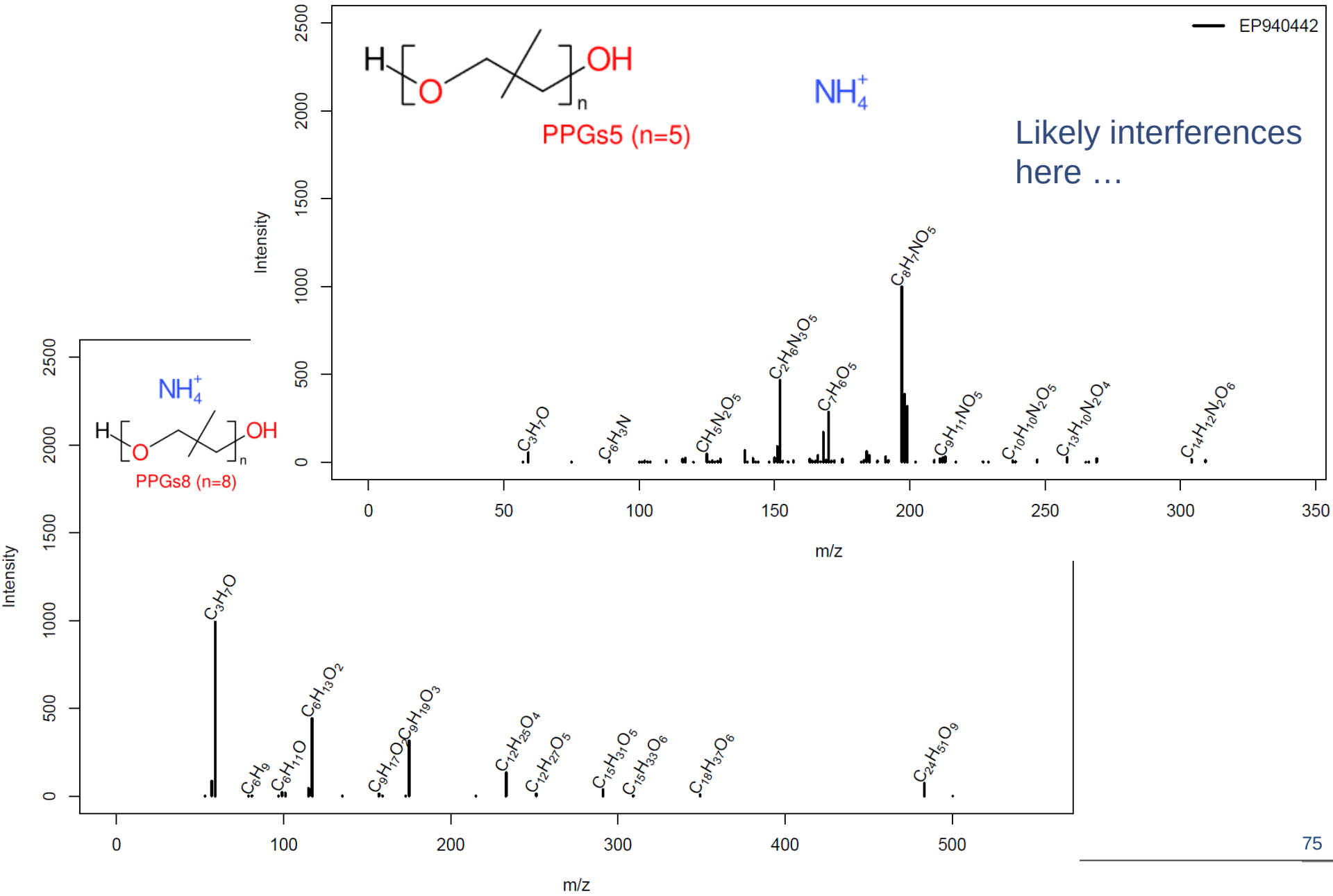
$[M+Na]^+$ - cleaned up



PPGs – MS/MS


EP940503
PPGs6 (n=6)
Intensity
m/z
C_3H_7O
$C_6H_{11}O$
$C_6H_{13}O_2$
$C_9H_{17}O_2$
$C_9H_{19}O_3$
$C_{12}H_{25}O_4$
$C_{18}H_{39}O_7$
PPGs5 (n=5)
C_3H_7O
$C_6H_{11}O$
$C_6H_{13}O_2$
$C_9H_{19}O_3$
$C_{14}H_{11}N_2$
$C_{12}H_{25}O_4$

PPGs – MS/MS


EP940442
PPGs5 (n=5)
NH_4^+
Likely interferences here …
Intensity
m/z
C_3H_7O
C_6H_3N
$CH_5N_2O_5$
$C_2H_6N_3O_5$
$C_7H_6O_5$
$C_8H_7NO_5$
$C_9H_{11}NO_5$
$C_{10}H_{10}N_2O_5$
$C_{13}H_{10}N_2O_4$
$C_{14}H_{12}N_2O_6$
PPGs8 (n=8)
C_6H_9
$C_6H_{11}O$
$C_6H_{13}O_2$
$C_9H_{17}O_3$
$C_9H_{19}O_3$
$C_{12}H_{25}O_4$
$C_{12}H_{27}O_5$
$C_{15}H_{31}O_5$
$C_{15}H_{33}O_6$
$C_{18}H_{37}O_6$
$C_{24}H_{51}O_9$

Take Home Messages

- Target/Suspect Screening: Surprising “bottleneck”
  - Level 1 and 2a not yet clear!
- Suspect Screening even with “medium” lists is dangerous!
  - Large suspect lists will lead to incredible false positive rates
- Outcomes: Special cases
  - MS-ready / mixture linking in MetFrag now a reality – with metadata!
  - We *do* observe organometallics in LC-MS
  - Semi-automated methods to screen polymer mixtures like PPGs
- Workflow weaknesses to address
  - (1) in source fragments and (2) co-eluting MS/MS … match values?

Call For Participants! http://normanews.eu/

- Global emerging contaminant early warning exercise
- Retrospective suspect screening with HR-MS

- Expression of Interest by November 1, 2018
- Submit novel suspects by November 1, 2018
- [Participant Meeting, Greece: November 28, 2018]
- Submit raw data and results: May 1, 2019

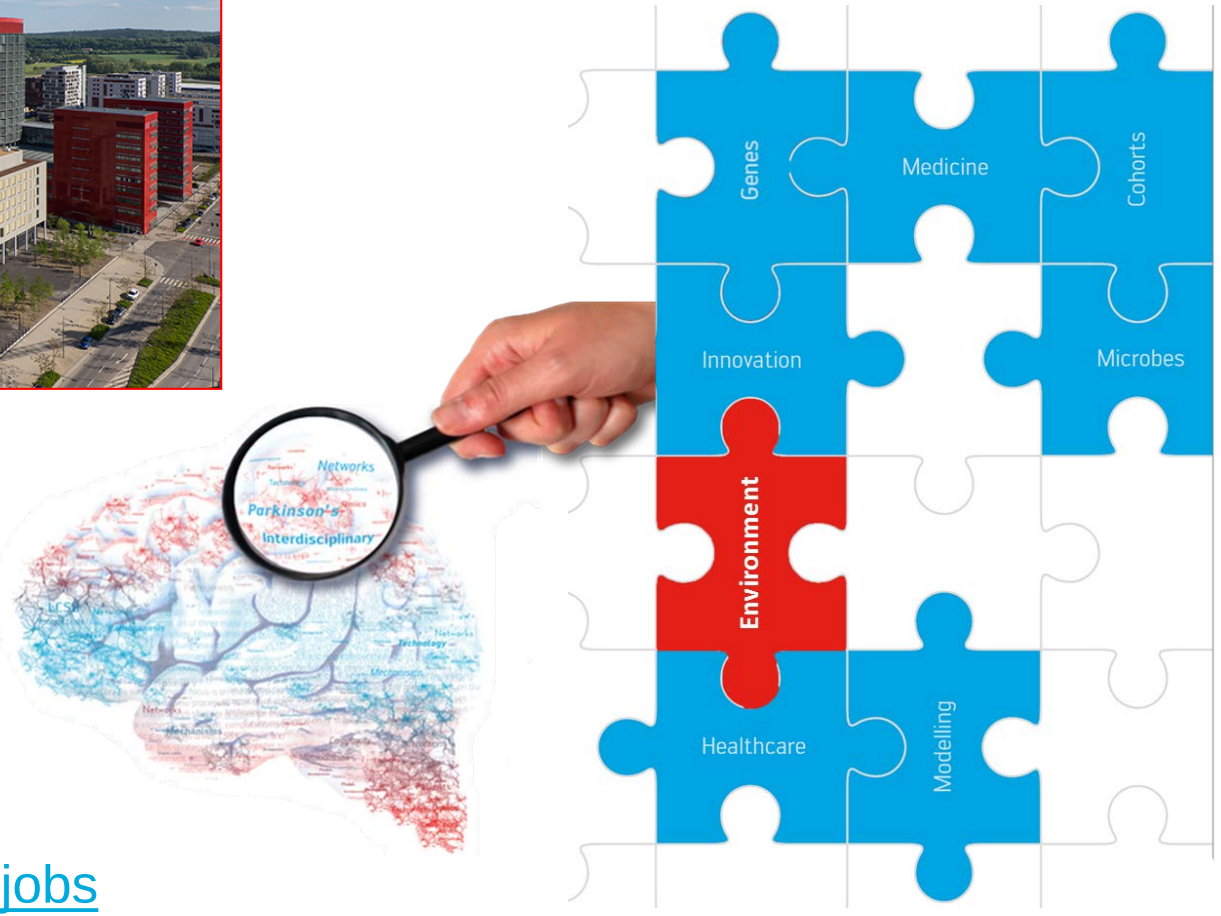
- Contact: Kevin Thomas: kevin.thomas@uq.edu.au

NormaNEWS1: Alygizakis *et al* 2018, ES&T DOI: 10.1021/acs.est.8b00365

Looking for NTS work? Join us!

LUXEMBOURG

LET'S MAKE IT HAPPEN



Contact
(shared address):
metabolomics.jobs@uni.lu

More Information …
Ask me or:
https://tinyurl.com/lcsb-eci-jobs
https://tinyurl.com/lcsb-metabolomics-jobs

Acknowledgements

solutions

EU Grant 603437

eawag aquatic research

EPA

MassBank.eu

MetFrag

Luxembourg National Research Fund

emma.schymanski@uni.lu

Further Information:

https://massbank.eu/MassBank/

https://github.com/MassBank/RMassBank

https://comptox.epa.gov/dashboard/

https://msbi.ipb-halle.de/MetFragBeta/

https://github.com/schymane/RChemMass/

https://github.com/schymane/ReSOLUTION/

Updates: https://www.researchgate.net/project/Supporting-Mass-Spectrometry-Through-Cheminformatics