

**EDAEEMERGE**

**eawag**  
aquatic research

**Candidate Selection via  
*in silico* fragmentation and candidate search**

**Eawag: Swiss Federal Institute of Aquatic Science and Technology**

Emma Schymanski  
Marie Curie Inter-European Postdoctoral Fellow

Plus many others who I have worked with...

**Candidate Selection via *in silico* fragmentation**

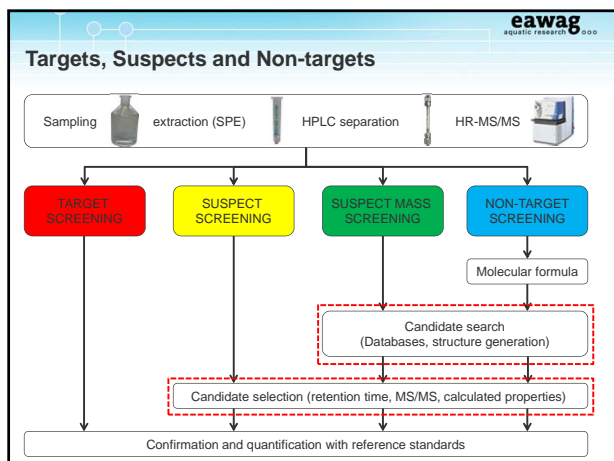
- Use MS/MS to search ChemSpider (or PubChem) with MetFrag
  - How does MetFrag work?
- Expand this to search ChemSpider (or PubChem) and a Mass Spectral Database with MetFusion
  - What does MetFusion do extra to MetFrag?
- "Bond Disconnection" versus "Rule-Based Fragmentation"
  - MetFrag
  - Mass Frontier

**ChemSpider**  
The free chemical database

**MetFrag**

**MassBank**

**MetFusion**



**Basics of Candidate Selection**

Molecules split into fragments in the mass spectrometer

- Use these as "clues" to identify unknowns

- Many "rules" for fragmentation in standard mass spectrometry text books
  - Manual interpretation is long, many rules and rearrangements
  - Most rule-based programs need candidates prior to fragmenting, e.g. Mass Frontier, ACD MS Fragmenter, FID
- Need a "CASE" system
  - MetFrag: compound database and "bond disconnection" fragmentation
  - MOLGEN-MS: structure generation and "rule-based" fragmentation – but only for GC-MS so far...

**MetFrag: <http://msbi.ipb-halle.de/MetFrag/>**

**MetFrag**  
In silico fragmentation for computer assisted identification of metabolite mass spectra

Database Settings: ☒ KEGG ☒ PubChem ☒ ChemSpider

Neutral exact mass: 272.0547

Molecular formula:

Only biological compounds: ☒

Limit # of structures: 100

Database IDs:

Search upstream DB:

MetFrag Settings  
Mode: ☒ [M+H]<sup>+</sup> ☐ [M+H]<sup>0</sup> ☐ [M]<sup>-</sup>  
Charge: ☒ pos ☐ neg  
Mzabs (e.g. 0.01): 0.01  
Mzppm (e.g. 10): 10

Peaks:

Calculate

S. Wolf, S. Schmidt, M. Müller-Hannemann, S. Neumann, BMC Bioinformatics 2010, 11:148

**MetFrag Unknown Example 1**  
EDA of River Elbe with Blue Rayon; c/o Christine Gallampois

Signal: [M+H]<sup>+</sup> = 293.1092 at 20.01 min

Calculate formula (MOLGEN-MS/MS<sup>1</sup>)

- Match formulas based on
  - ppm difference
  - Isotope pattern deviation
  - Fragment assignment
- C<sub>19</sub>H<sub>17</sub>PO is best fit

MS	MS/MS
293.1092	520745.8
294.1125	108585.3
91.0539	63505.5
201.0465	64053.3
215.0620	8374.6
219.0576	6768.7
233.0733	7224.4
237.1126	8704.7
265.1008	5939.7

Perform candidate search with C<sub>19</sub>H<sub>17</sub>PO

- MetFrag<sup>2</sup> with PubChem – 23 hits

<sup>1</sup>M. Meringer, S. Reinker, J. Zhang, A. Müller: MATCH (2011) 65:259. <sup>2</sup>S. Wolf et al. BMC Bioinformatics (2010) 11:148.

### MetFrag Unknown Example 1

MetFrag  
In silico fragmentation for computer assisted identification of metabolite mass spectra

Database: KEGG, PubChem, ChEMBL, Local SDF

Neutral exact mass: 293.1092  
Molecular formula: C<sub>19</sub>H<sub>17</sub>O<sub>3</sub>P  
Only biological compounds: ☐  
Limit # of structures: 1000  
Database IDs:

Search upstream DB: 23 hits

MetFrag Settings:  
Mode: ☒ [M+H]<sup>+</sup> ☐ [M+H]<sup>+</sup> ☐ [M]<sup>+</sup>  
Charge: ☒ pos. ☐ neg.  
Mzabs (e.g. 0.01): 0.005  
Mzppm (e.g. 10): 5

Process all 23 compounds/START Step View spectrum

### MetFrag Unknown Example 1

Log

Download complete table Generate output files

Rank	Score	Exact Mass	Structure	Database ID	Actions
0	3	292.1017		12550550	Download
1	3	292.1017		300120	Fragment Download

Feedback

### MetFrag Unknown Example 1

Individual Fragments...

Fragments

Mass: 292.1017 [C<sub>19</sub>H<sub>17</sub>O<sub>3</sub>P]<sup>+</sup> (Original Compound)

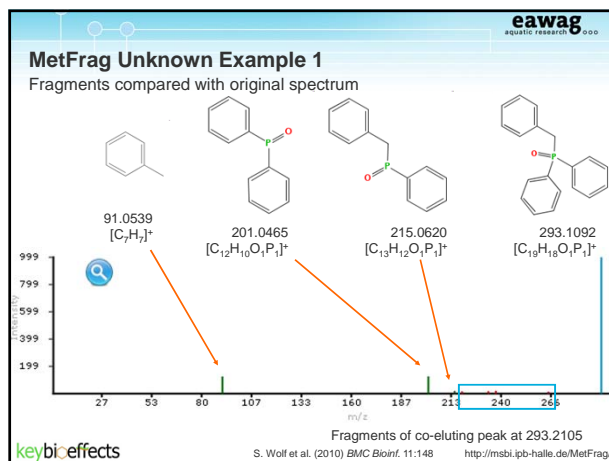
Mass: 215.062 [C<sub>13</sub>H<sub>12</sub>O<sub>2</sub>P]<sup>+</sup> (0.0 ppm)

Mass: 215.062 [C<sub>13</sub>H<sub>12</sub>O<sub>2</sub>P]<sup>+</sup> (0.0 ppm)

Mass: 215.062 [C<sub>13</sub>H<sub>12</sub>O<sub>2</sub>P]<sup>+</sup> (0.0 ppm)

Mass: 201.0464 [C<sub>12</sub>H<sub>10</sub>O<sub>2</sub>P]<sup>+</sup> (0.5 ppm)

Mass: 91.0539 [C<sub>7</sub>H<sub>7</sub>]<sup>+</sup> (3.3 ppm)



### MetFrag Unknown Example 1

Generate and download files...as xls or sdf

Download complete table

MetFragResults\_1349073807740.xls [Compatibility Mode]

Rank	Score	# of Peaks	Molecular Exact Mass	Database	XlogP	AlogP	Peaks	Ex Image	Name	Smiles
0	3	3	292.1017	12550550	2.3570000	2.3570000	1.0539 990.0 201.04	1-diphenyl	CC1=CC=CC(C1P(=O)(O)C2=CC=CC=C2)C3=CC=CC=C3	
1	3	3	292.1017	300120	2.3570000	2.3570000	1.0539 990.0 201.04	1-diphenyl	CC1=CC=CC(C1P(=O)(O)C2=CC=CC=C2)C3=CC=CC=C3	
2	3	3	292.1017	2550554	2.3570000	2.3570000	1.0539 990.0 201.04	1-diphenyl	CC1=CC=CC(C1P(=O)(O)C2=CC=CC=C2)C3=CC=CC=C3	
3	3	3	292.1017	76293	2.2463999	2.2463999	1.0539 990.0 201.04	benzylip	O=P(C1=CC=CC=C1)C2=CC=CC=C2	
4	3	3	292.1017	298154	2.9785000	2.9785000	1.0539 990.0 201.04	1-diphenyl	C1=CC=CC(C1COP(=O)(O)C2=CC=CC=C2)C3=CC=CC=C3	
5	3	3	292.1017	2933766	3.1621000	3.1621000	1.0539 990.0 201.04	1-diphenyl	C1=CC=CC(C1COP(=O)(O)C2=CC=CC=C2)C3=CC=CC=C3	
6	3	3	292.1017	21433748	3.1621000	3.1621000	1.0539 990.0 201.04	1-diphenyl	C1=CC=CC(C1COP(=O)(O)C2=CC=CC=C2)C3=CC=CC=C3	
7	3	3	292.1017	21433754	3.1621000	3.1621000	1.0539 990.0 201.04	1-diphenyl	C1=CC=CC(C1COP(=O)(O)C2=CC=CC=C2)C3=CC=CC=C3	
8	3	3	292.1017	7690578	2.9785000	2.9785000	1.0539 990.0 201.04	1-diphenyl	C1=CC=CC(C1COP(=O)(O)C2=CC=CC=C2)C3=CC=CC=C3	
9	3	3	292.1017	27327721	3.4235000	3.4235000	1.0539 990.0 215.06	2-diphenyl	COC1=CC=CC(C1COP(=O)(O)C2=CC=CC=C2)C3=CC=CC=C3	
10	3	3	292.1017	5242049	2.8960000	2.8960000	1.0539 990.0 215.06	2-diphenyl	COC1=CC=CC(C1COP(=O)(O)C2=CC=CC=C2)C3=CC=CC=C3	
11	3	3	292.1017	2754311	3.5755000	3.5755000	1.0539 990.0 215.06	1-diphenyl	CC1=CC=CC(C1COP(=O)(O)C2=CC=CC=C2)C3=CC=CC=C3	
12	3	3	292.1017	4407809	3.0416999	3.0416999	215.062 131.0	1-methoxy	COCC1=CC=CC(C1COP(=O)(O)C2=CC=CC=C2)C3=CC=CC=C3	
13	3	3	292.1017	603401	3.0416999	3.0416999	215.062 131.0	1-methoxy	COCC1=CC=CC(C1COP(=O)(O)C2=CC=CC=C2)C3=CC=CC=C3	
14	3	3	292.1017	2775967	3.0416999	3.0416999	215.062 131.0	1-methoxy	COCC1=CC=CC(C1COP(=O)(O)C2=CC=CC=C2)C3=CC=CC=C3	
15	3	3	292.1017	74963484	5.031	5.031	215.062 131.0	1-methoxy	COCC1=CC=CC(C1COP(=O)(O)C2=CC=CC=C2)C3=CC=CC=C3	
16	3	3	292.1017	11119874	2.4	2.1780000	1.0539 990.0 215.06	1-methyl	CC(C)C1=CC=CC(C1COP(=O)(O)C2=CC=CC=C2)C3=CC=CC=C3	
17	3	3	292.1017	44633933	3.1889999	3.1889999	1.0539 990.0 215.06	1-methyl	CC(C)C1=CC=CC(C1COP(=O)(O)C2=CC=CC=C2)C3=CC=CC=C3	
18	3	3	292.1017	71748344	5.009	1.9273999	215.062 131.0	1-methyl	CC(C)C1=CC=CC(C1COP(=O)(O)C2=CC=CC=C2)C3=CC=CC=C3	
19	3	3	292.1017	23281218	3.009	1.9273999	215.062 131.0	1-methyl	CC(C)C1=CC=CC(C1COP(=O)(O)C2=CC=CC=C2)C3=CC=CC=C3	
20	3	3	292.1017							
21	3	3	292.1017							
22	3	3	292.1017							
23	3	3	292.1017							
24	3	3	292.1017							
25	3	3	292.1017							
26	3	3	292.1017							
27	3	3	292.1017							

### MetFrag Score

Considers peak count and bond disconnection energy

Equation to calculate MetFrag score is:

$$S_i = \frac{1}{\max(w)} w_i - \frac{1}{2 \max(e)} e_i$$

where  $w_i = \sum_{f \in F_i} I_f^{0.6} m_f^3$  and  $e_i = \frac{1}{|F_i|} \sum_{f \in F_i} \sum_{b \in B_f} BDE_b$

i.e. this considers:

- Intensity and the mass (m/z) of peaks in the spectrum
- Energy to break the bonds in the fragments
- More details in Wolf et al 2010, BMC Bioinformatics 11:148
- MetFrag is still "research in progress" – scoring & features dynamic

**MetFrag Unknown Example 1**  
Candidate Selection

Log  $K_{ow}$  too high  
Lower MetFrag Score (fragments, BDE)

keybi:effects

Images: <http://msbi.ipb-halle.de/MetFrag/>

**MetFrag Unknown Example 1**  
Confirmation Efforts

BDPO

Benzyldiphenylphosphine oxide  
(Methylphenyl)diphenylphosphine oxide

All fragments predicted using MetFrag  
Co-eluting fragments easy to spot  
MS/MS and RT of BDPO match  
⇒ identity confirmed analytically  
Additional evidence:  
Triphenylphosphine oxide found at RT = 19.80  
⇒ reaction products of phosphines (org. synthesis)

**MetFrag Unknown Example 2**  
EDA of River Elbe with Blue Rayon c/o Christine Gallampois

Signal:  $[M+H]^+ = 233.0963$  at 18.51 min

Fraction N2-8

- Expected log  $K_{ow}$  for this fraction: [1.2 - 3.2]

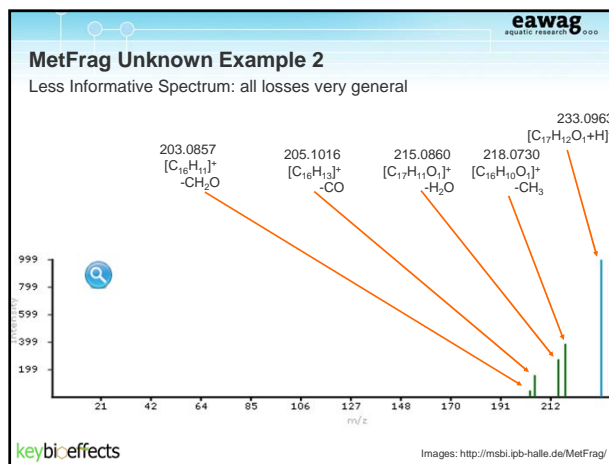
Calculate formula (MOLGEN-MS/MS<sup>1</sup>)

- $C_{17}H_{12}O_1$  clearly best match

Candidate Search:

- MetFrag<sup>2</sup> with ChemSpider and PubChem
- 113 hits

<sup>1</sup>M. Meringer, S. Reinker, J. Zhang, A. Müller: MATCH (2011) 65:259. <sup>2</sup>S. Wolf et al. BMC Bioinformatics (2010) 11:148.



**Real World Examples**  
EDA of River Elbe with Blue Rayon – Unknown Example 2

Candidate selection:

Log  $K_{ow}$  too high  
Lower MetFrag Score (fragments, BDE)

keybi:effects

Images: <http://msbi.ipb-halle.de/MetFrag/>

**MetFrag Unknown Example 2**  
EDA of River Elbe with Blue Rayon

We have our prime candidate from the database:

- 7,9-dimethyl-8H-cyclopenta[a]acenaphthylen-8-one
- MetFrag, database search and predicted properties help with structure selection
- BUT
  - Relies on molecules in database
  - e.g. many isomers of this compound not present

The other problem:

- Purchasing of standards for confirmation – this compound is not "available", although it is in the database
- Purchase is difficult enough for database entries, almost impossible for generated structures...

eawag  
environmental water research group

## In silico fragmentation

Some words of caution

MetFrag has a “bond disconnection” approach

- Very quick, often very good – but can’t do e.g. rearrangements
- Don’t fall into the trap of thinking Rank 1 = correct structure
- Use “common sense” when looking at the fragments: some are strange

Mass Frontier (commercial) – “rule-based” approach


- Given a structure, fragments according to rules (irrespective of spectrum)
- Can view mechanisms for proposed fragments

42.010 | 42.010 | 43.018 | 126.046 | 127.054 | 143.049 | m/z 144.057 | 145.065 | 171.044 | 196.068 |

1 2 1 1 2 3 4 5 ▶ Select possible fragments with m/z 144.057

CC(=O)c1ccccc1  $\rightarrow$  CC(=O)[O-]c1ccccc1  $\xrightarrow{-\cdot\text{CH}_3}$  [O-]C(=O)c1ccccc1  $\xrightarrow{H}$  [O-]C(=O)c1ccccc1  $\xrightarrow{\alpha}$  [O-]C(=O)c1ccccc1

m/z 196.068
m/z 136.068
m/z 136.068
m/z 144.057



## In silico fragmentation

Some words of caution

MetFrag has a “bond disconnection” approach

- o Very quick, often very good – but can’t do e.g. rearrangements
- o Don’t fall into the trap of thinking Rank 1 = correct structure
- o Use “common sense” when looking at the fragments: some are strange

Mass Frontier (commercial) – “rule-based” approach

- o Given a structure, fragments according to rules (irrespective of spectrum)
- o Can view mechanisms for proposed fragments

A general rule to be careful:

- o The more fragmentation steps, the more you explain  
– for the **correct** structure as well as the **incorrect**
- o Ranking of candidates –top 30-40 % (of **all** candidates)
- o See (for those who want more details...)

Schymanski, Meringer, Brack, 2009, Analytical Chemistry, 81, 3608–3617

[illegible][illegible]

**MetFusion Screenshot**

**Spectrum Query** | **Information**

**Selected Spectral Database:** MassBank

**MassBank Parameters**  
<http://www.massbank.jp/>

**MassBank Server:** <http://www.massbank.jp/>

**Number of Results:** 100

**Cutoff threshold of relative intensities:** 5

**Ionization Mode:** positive

**Instruments:**

**Select EI** | **Selected ESI** | **Select Others**

- ☐ B-I
- ☐ B-EBEB
- ☐ GC-EI-TOF
- ☒ ESI-IT-MS/MS
- ☒ ESI-QQ
- ☒ ESI-QqT-MS/MS
- ☒ ESI-QqQ-MS/MS
- ☒ ESI-QqTqT-MS/MS
- ☒ LC-ESI-IT
- ☒ LC-ESI-HFT
- ☒ LC-ESI-HFTT
- ☒ LC-ESI-Q
- ☒ LC-ESI-QQ
- ☒ LC-ESI-QqT
- ☐ APPQ-QQ-MS
- ☐ APPQ-QqQ-MS/MS
- ☐ C-IE
- ☐ FAB-B
- ☐ FAB-EB
- ☐ FAB-EBEB
- ☐ FD-B
- ☐ LC-APPQ-QQ
- ☐ MALDI-TOF
- ☐ MALDI-QqT
- ☐ MALDI-QqTqT

**MetFusion Parameters**

**Filter:** Unique

Peak	Intensity
118.03146	
123.04437	
147.044607	
153.010999	
176.03614	
186.05817	
272.076999	
276.08311	

**Peaks:**

**MetFusion Progress**

**MetFusion Parameters**

**MetFusion is set to work in positive mode**

**cawag**  
academic resources 2020

# MetFusion Screenshot

**MetFrag Parameters**

Upstream DB: ☒ REGG ☐ PubChem ☐ ChemSpider ☐ SDf Upload

Database IDs:

Molecular Formula:

Parent Ion:  Neutral

Exact Mass:

Limit # of Structures:

C,H,N,O,P,S only? ☒

Search PPM:

m/z abs:

m/z ppm:

MetFrag is set to work in positive mode

**eawag**  
aquatic research 000

## The Data “bottleneck” for MetFusion


Not many spectra in databases for environmental compounds...


Original Similarity Matrix Reranked Similarity Matrix

WIA000888	WIA002547	WIA001845	KC0008908	WIA002954	CC000115	WIA001240	WIA000198	WIA001762	WIA000362	WIA001101	WIA000155	WIA001091	WIA001773	WIA002558
0.134	0.254	0.203	0.16	0.158	0.18	0.28	0.129	0.232	0.176	0.192	0.182	0.187	0.208	0.199
0.218	0.186	0.212	0.212	0.224	0.166	0.099	0.10	0.382	0.162	0.173	0.228	0.160	0.238	0.214
0.211	0.163	0.18	0.201	0.209	0.166	0.089	0.103	0.398	0.162	0.164	0.221	0.163	0.233	0.19
0.459	0.139	0.178	0.246	0.274	0.199	0.103	0.166	0.309	0.188	0.184	0.23	0.164	0.266	0.192
0.454	0.192	0.172	0.242	0.263	0.209	0.148	0.16	0.338	0.169	0.166	0.234	0.163	0.266	0.187
0.347	0.164	0.188	0.228	0.267	0.171	0.122	0.165	0.291	0.163	0.162	0.224	0.164	0.265	0.185
0.361	0.158	0.183	0.232	0.258	0.179	0.118	0.167	0.291	0.151	0.148	0.229	0.158	0.22	0.182
0.341	0.168	0.192	0.224	0.262	0.178	0.114	0.166	0.312	0.161	0.147	0.213	0.156	0.223	0.182
0.498	0.128	0.172	0.247	0.262	0.217	0.163	0.161	0.301	0.148	0.165	0.234	0.161	0.285	0.188
0.197	0.207	0.22	0.17	0.164	0.17	0.118	0.164	0.2	0.164	0.178	0.179	0.173	0.188	0.211

Don't benefit from the similarity comparison when the spectra in databases are too dissimilar to the unknowns...

**eawag**  
aquatic research 000

 **ChemSpider**  
The free chemical database

 **PubChem**  
Structure Search

 **MetFrag**

## Practice Session

Candidate Search with Compound Databases  
MetFrag and MetFusion