

ACTIONS
MARIE CURIE

EDA EMERGE

eawag
aquatic research

Where to start identification?

Turning big lists into, well, smaller lists!

Eawag: Swiss Federal Institute of Aquatic Science and Technology

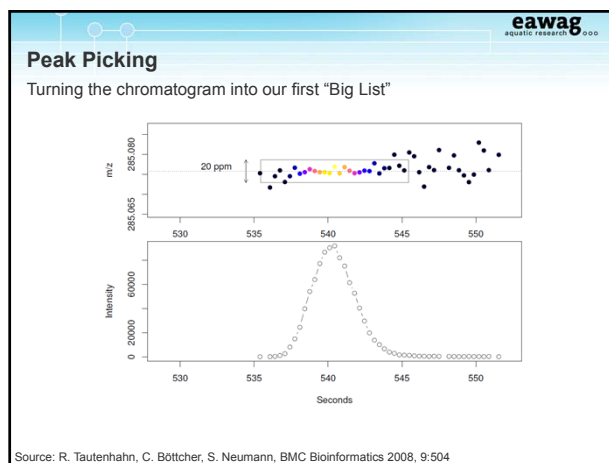
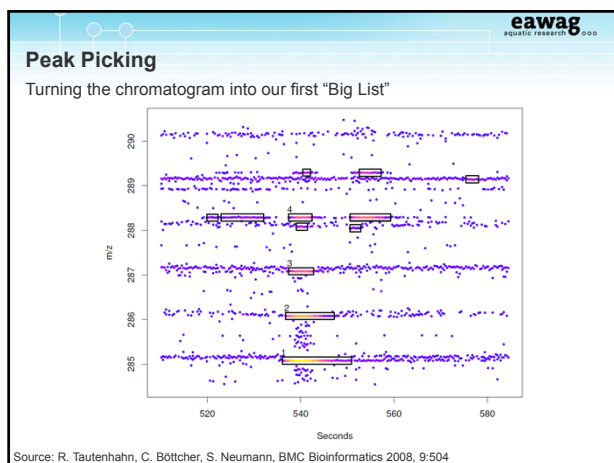
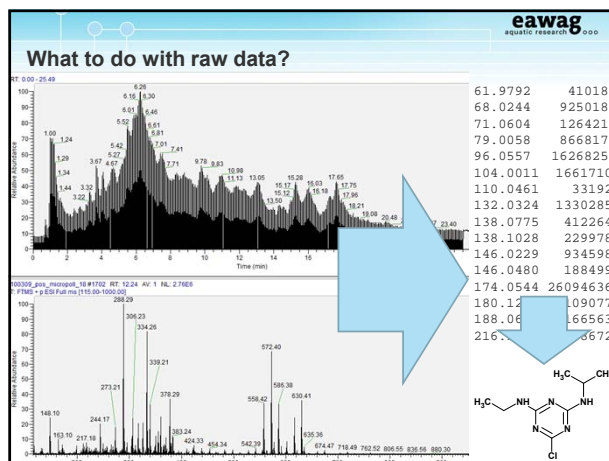
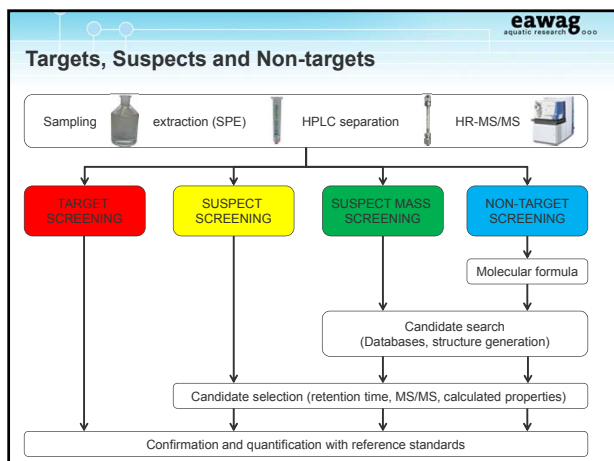
Emma Schymanski
Marie Curie Inter-European Postdoctoral Fellow

Plus many others who have worked with...

eawag
aquatic research

An overview of screening

- Removing unwanted peaks
 - Background and noise reduction
- Target Screening
- Suspect Screening
 - "Known" suspects
 - "Unknown" suspects
 - "Related" suspects
- Detecting "wanted" peaks
 - Peak grouping



Peak Picking

Turning the chromatogram into our first "Big List"

- Several options for peak picking
 - XCMS and centWave¹
 - mzMine (Martin can tell you more about this one)²
 - For Thermo data: Formulator³
- Result is something like this (from Formulator output):

Centroid	m/z	Intensity	Peak S/N	Scan Num	RT (min)	Start Scan	Start RT (min)	End Scan	End RT (min)	Mass Cov
441	01136	53942	811.68	638	4.49	624	4.39	660	4.65	0.995
395	00576	11012	16.53	638	4.49	638	4.49	645	4.54	0.995
161	02187	7221	12.45	1289	9.11	1281	9.05	1290	9.12	0.99
256	06439	38623	60.32	869	6.13	771	5.43	876	6.18	0.984
207	01198	221324	357.77	778	5.48	723	5.09	918	6.49	0.994
369	02123	8295	12.76	1849	13.11	1842	13.06	1849	13.11	0.994
328	12169	9155	14.26	1205	8.51	1205	8.51	1212	8.56	0.99
171	01197	549903	924.68	491	3.45	450	3.17	561	3.94	0.99
171	06629	394062	654.46	820	5.77	168	1.18	1191	8.42	0.981
181	59138	248443	4200.96	435	3.06	315	2.21	588	4.13	0.993
178	05411	638750	1072.03	533	3.74	513	3.61	561	3.94	0.994
557	2436	97739	142.82	1772	12.55	1752	12.41	1785	12.64	0.993
429	05251	170230	254.37	1401	9.91	1380	9.76	1422	10.06	0.994

¹R. Tautenhahn, C. Böttcher, S. Neumann, BMC Bioinformatics 2008, 9:504; www.bioconductor.org
²http://mzmine.sourceforge.net/ ³http://www.eawag.ch/forschung/uchem/software/thermo (download & requirements)

Making the "Big List" Smaller

Overview of the steps

- Blank/Control Subtraction
- Noise and "spark" removal
- Peak grouping
 - Removing peaks but also...
 - Providing information for identification efforts

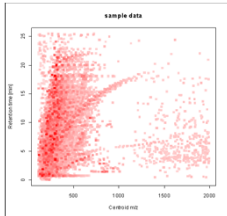
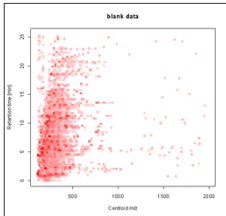
I'll demonstrate this here with examples from our data using envMass

There are many other options!

Making the "Big List" Smaller

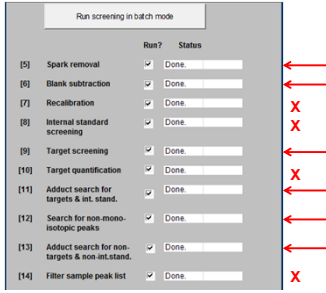
Background / Control subtraction

- Step 1: Upload "sample" and "blank"
 - "blank" = laboratory blank, trip blank, control sample... whatever you would like to subtract from the sample

Making the "Big List" Smaller

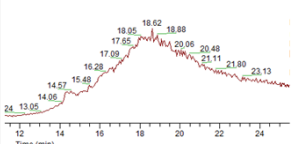
enviMass workflow in one screen shot



Making the "Big List" Smaller

Spark removal

- Spark = electronic/measurement artefacts
 - Peaks that aren't real "peaks"
 - Too short;
 - i.e. Only 1-2 "sticks"
 - Too long;



Spark removal

Remove sparks Stop spark removal

0.03 remove < .JRT (min) (lower bound)

5 remove > .JRT (min) (upper bound)

☒ use scan number:

3 remove < .J scan number (lower bound)

 remove > .J scan number (upper bound)

☒ use only lower retention time

☐ use only upper retention time

Number of samples removed: 2804

Making the "Big List" Smaller

Blank Subtraction

- Remove "all" substances in the blank sample
 - Mass threshold
 - Retention time threshold
 - Intensity threshold
- CAREFUL
 - Laboratory contaminants can also be environmental contaminants!
 - It is very important to adjust the thresholds to your particular investigation
 - If in doubt – try different combinations

Making the "Big List" Smaller

Blank Subtraction

Remove "all" substances in the blank sample

- Mass threshold

Blank subtraction

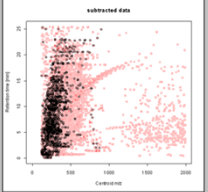
Run blank subtraction Skip blank subtraction

Tolerances:

$\Delta m/z$ ☒ ΔRT [min] ΔRT [min]

Intensity of blank: times smaller than sample

Number of samples removed:



Making the "Big List" Smaller

Target Screening

- Screen for substances where we have
 - Mass and formula - and thus the isotope distribution!
 - Retention time - and thus we can define a RT range
- Make sure you define appropriate thresholds:
 - Mass
 - Retention time
 - Intensity

Tolerances:

$\Delta m/z$ ☒ ΔRT listed vs. measured [min] ΔRT within scan [min] ΔRT within scan [min]

% Intensity

Intensity cutoff (default=5000)

Score weights for summary score:

Score 1: hits after rescaling to monoisotopic peak

Score 2: score 1 + tolerance in intensity

Score 3: score 2 after subtraction of blind candidates

Making the "Big List" Smaller

Target Screening

Target screening Stop target screening

Status: Calculations finished

standards with monoisotopic hits:

Tolerances:

$\Delta m/z$ ☒ ΔRT listed vs. measured [min] ΔRT within scan [min] ΔRT within scan [min]

% Intensity

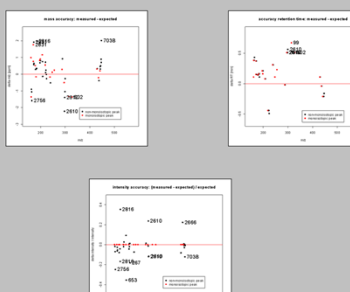
Intensity cutoff (default=5000)

Score weights for summary score:

Score 1: hits after rescaling to monoisotopic peak

Score 2: score 1 + tolerance in intensity

Score 3: score 2 after subtraction of blind candidates



Making the "Big List" Smaller

Target Screening (examples from ESI-)

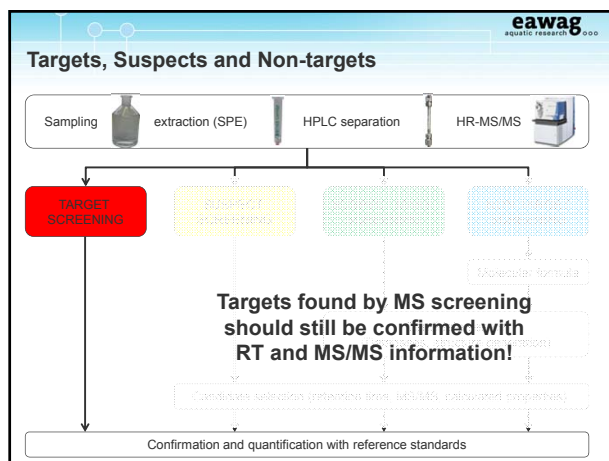
ID	compound name	chemical formula	Mon. mass	retention time
24	Bromocynyl	C ₇ H ₅ BrN ₂ O	274.8581	8.5
87	Isopropyl	C ₃ H ₇ NO	370.8384	9.7
99	SUM Alachlor-ESA / Acetochlor-ESA (103)	C ₁₄ H ₂₁ N ₂ O ₅ S	315.1140	7.8
100	Alachlor-OXA	C ₁₄ H ₁₉ N ₂ O ₄ S	265.1314	7.7
102	SUM Acetochlor-OXA / Alachlor-OXA (100)	C ₁₄ H ₁₉ N ₂ O ₄ S	265.1314	7.7
103	Acetochlor-ESA	C ₁₄ H ₂₁ N ₂ O ₅ S	315.1140	7.8

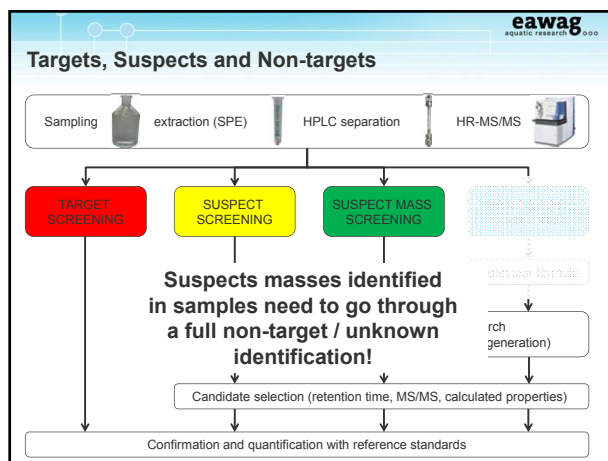
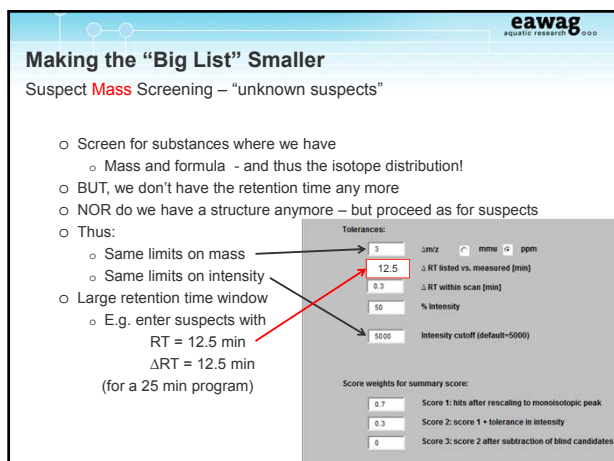
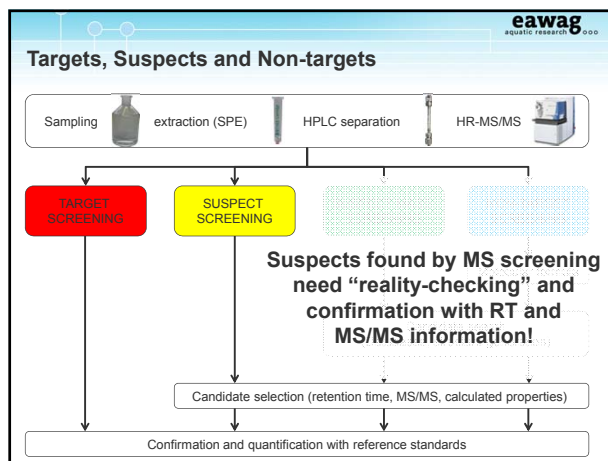
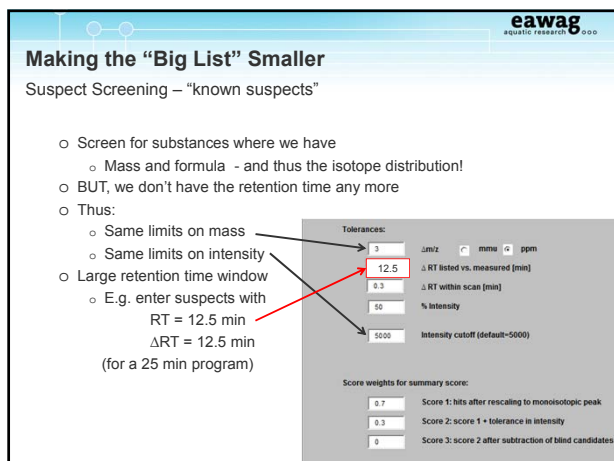
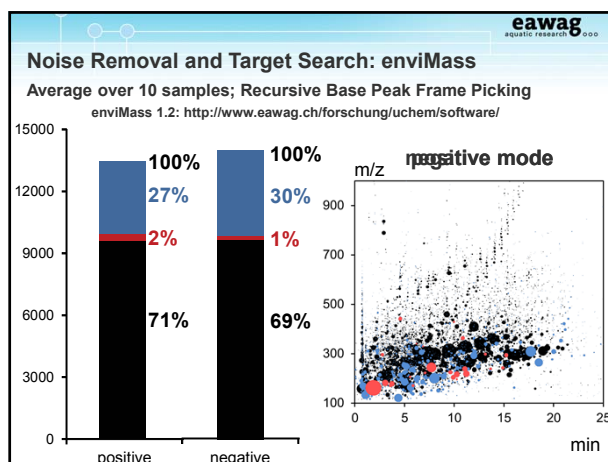
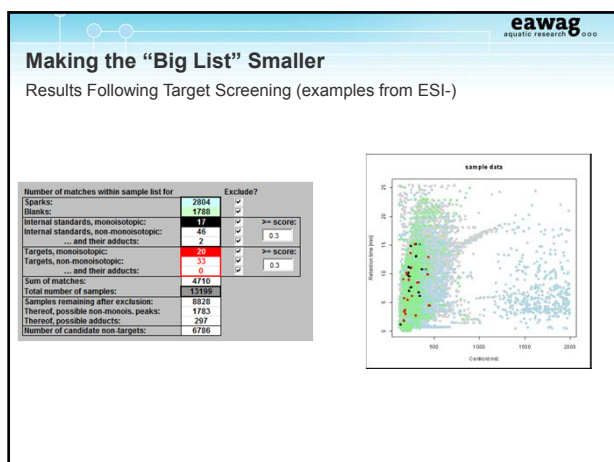
Isotopic m/z	Isotopic abundance
275.84884, 273.85089, 277.84679, 276.8522, 274.85424, 278.85015, 279.8535	1.0, 5.13993, 0.486388, 0.07571, 0.038914, 0.036824, 0.001195
369.82315, 370.8265, 371.82986	1.0, 0.07571, 0.002457
314.10679, 315.11015, 316.10259, 317.10594	1.0, 15.142, 0.044742, 0.006775
264.12416, 265.12751, 266.13087, 267.13176	1.0, 15.142, 0.010645, 0.001245
264.12416, 265.12751, 266.13087, 267.13176	1.0, 15.142, 0.010645, 0.001245
314.10679, 315.11015, 316.10259, 317.10594	1.0, 15.142, 0.044742, 0.006775

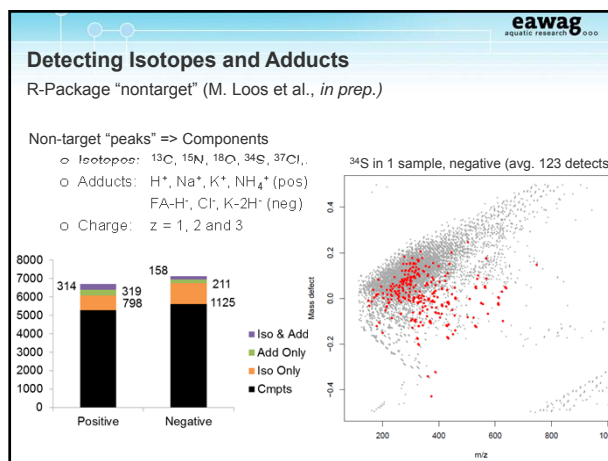
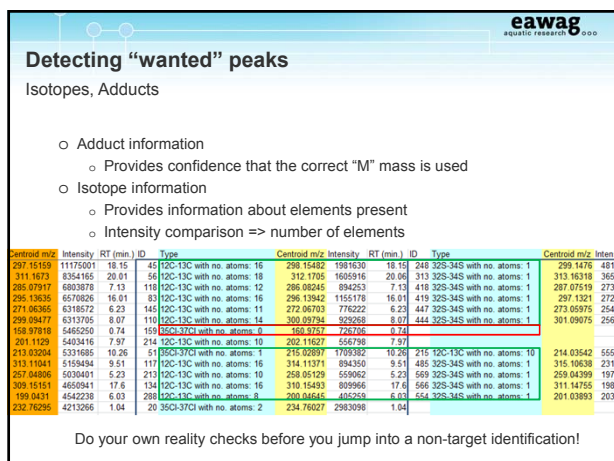
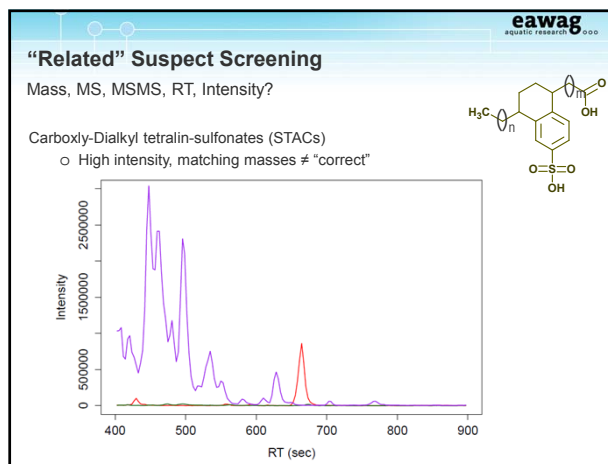
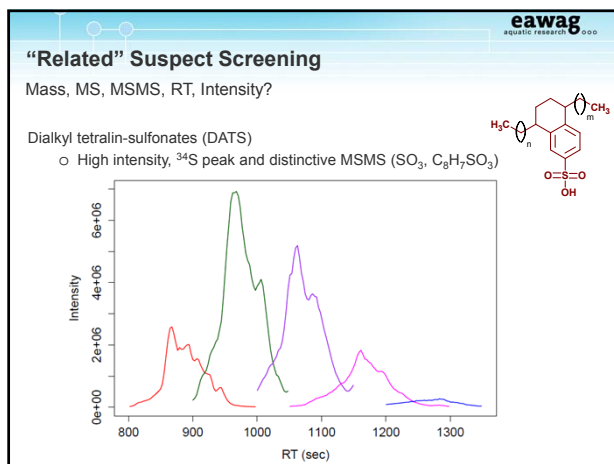
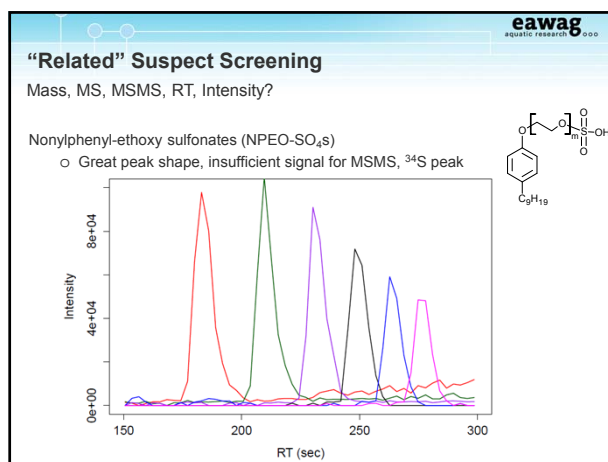
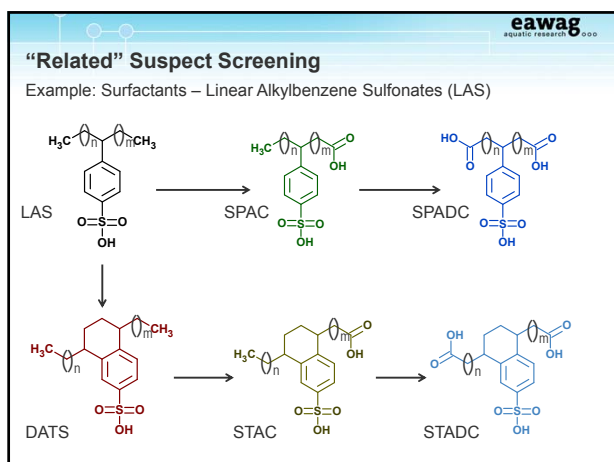
Making the "Big List" Smaller

Target Screening (examples from ESI-)

Monoisotopic hits	Isotopic hits #	delta m/z (ppm)	delta RT	delta intensity	score 1	score 2	score 3	sum score
0	0	No monoisotopic hit found: no isotopic pattern fit conducted.						
0	5043	-1.337	0.67	0	1 of 1	1 of 1	1 of 1	3
0	0	Not used for screening!						
0	0	No monoisotopic hit found: no isotopic pattern fit conducted.						
0	0	Not used for screening!						
0	0	No monoisotopic hit found: no isotopic pattern fit conducted.						
0	0	No monoisotopic hit found: no isotopic pattern fit conducted.						
0	0	No monoisotopic hit found: no isotopic pattern fit conducted.						
0	0	No monoisotopic hit found: no isotopic pattern fit conducted.						
0	0	No monoisotopic hit found: no isotopic pattern fit conducted.						
0	0	No monoisotopic hit found: no isotopic pattern fit conducted.						
0	7818	0.121	0.11	0	1 of 1	1 of 1	1 of 1	3
0	6305 / 6233	0.279 / 0.91697	-0.02 / -0.02	0 / 0.015	2 of 2	2 of 2	2 of 2	6
0	0	Not used for screening!						
0	8082 / 8053	-0.167 / -0.51486	0.31 / 0.31	0 / -0.074	2 of 3	2 of 3	2 of 3	6
0	0	No monoisotopic hit found: no isotopic pattern fit conducted.						
0	5223	0.504	0.16	0	1 of 1	1 of 1	1 of 1	3
0	0	No monoisotopic hit found: no isotopic pattern fit conducted.						
0	0	No monoisotopic hit found: no isotopic pattern fit conducted.						
0	8537 / 8579 / 8661 / 8603	0.1475 / 0.71202, 44 / -0.44 / -0.44 / -0.012 / -0.043 / -0.0			6 of 6	6 of 6	6 of 6	18







eawag
aquatic research

Practice Session

Background Removal, Target & Suspect Screening

