Recalibration and Annotation of High Accuracy MS/MS: Implications for non-target screening

Eawag: Swiss Federal Institute of Aquatic Science and Technology

Presenting: Emma Schymanski

Coauthors: Michael Stravs, Heinz Singer and Juliane Hollender

Eawag – Department of Environmental Chemistry

Dübendorf (Zurich), Switzerland

emma.schymanski@eawag.ch

Presentation Overview

- Motivation
- Improve Availability of High Accuracy MS/MS
  - Formula annotation improves interpretation
- Influence of Recalibration and Clean-up on Non-target Analysis
  - Improvements to database search
  - Improved ranking of correct molecular formula
- Future Perspectives for Non-targets and NORMAN MassBank
- CASMI: test your non-target skills!

Motivation

The Catch-22 of MS/MS

- MS is method of choice for environmental investigations
- Many institutes now have access to high accuracy MS/MS (MS^n) instruments
- MS/MS libraries are small and metabolite-dominated
  - ~20,000 compounds compared with 200,000 EI-MS
  - Only LR-MS/MS in NIST MS/MS library
- All environmental institutes are in the same boat => NORMAN MassBank

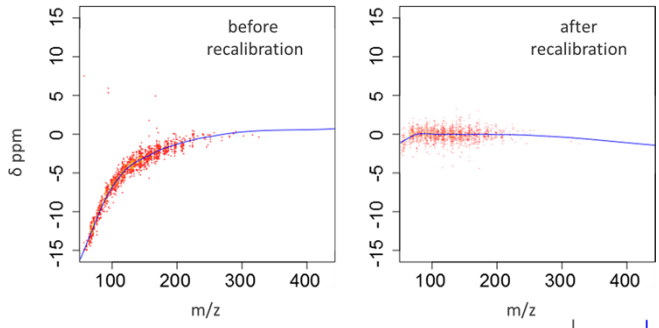
Value-add High-Accuracy Data

- Many existing programs merely extend low-resolution philosophy
  - Ignores much of the additional information gained with HR-MS/MS
- Increasing availability of HR-MS/MS
  - Improves identification via availability for all
  - More test data for development of HR-MS/MS methods

Improving Availability of High Accuracy MS/MS

Development of the RMassBank workflow[1,2]

- Open source workflow to process standard MS/MS spectra
  - Developed on Orbitrap data; also used on Q-Exactive
- Recalibration of MS/MS data to remove MS/MS drift at low *m/z*



[1]Stravs, Schymanski, Singer and Hollender, 2012, *J. Mass Spectrom.*, DOI: 10.1002/jms.3131

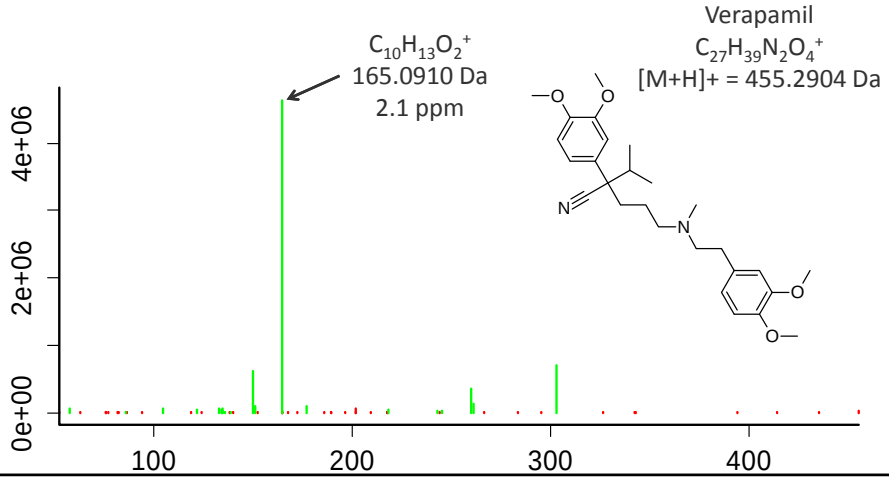
[2]http://www.bioconductor.org/packages/devel/bioc/html/RMassBank.html

RMassBank

Improving Availability of High Accuracy MS/MS

Development of the RMassBank workflow[1,2]

- Clean-up of spectra to remove FT and electronic artefacts, noise



Improving Availability of High Accuracy MS/MS

Development of the RMassBank workflow[1,2]

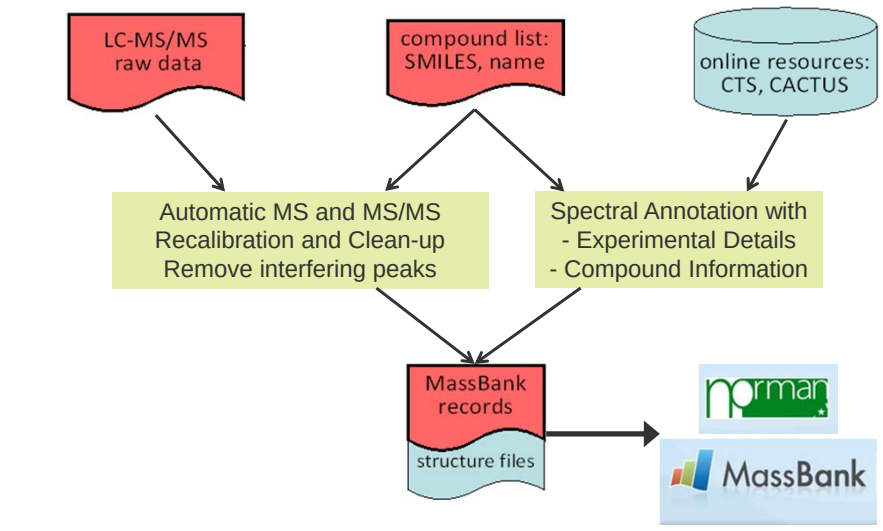
- Annotation of formulas to MS/MS peaks
  - Provides interpretation support within MassBank record already!

```
ACCESSION: EA030301
RECORD_TITLE: Caffeine; LC-ESI-ITFT; MS2; …; [M+H]+
...
PK$ANNOTATION: m/z num {formula mass error(ppm)}
  110.0713 1 C5H8N3+ 110.0713 0.24
  138.0662 1 C6H8N3O+ 138.0662 -0.06
  195.0877 1 C8H11N4O2+ 195.0877 0.14
PK$NUM_PEAK: 3
PK$PEAK: m/z int. rel.int.
  110.0713 33560 21
  138.0662 1570359.5 999
  195.0877 238647.1 151
```

[1]Stravs, Schymanski, Singer and Hollender, 2012, *J. Mass Spectrom.*, DOI: 10.1002/jms.3131

[2]http://www.bioconductor.org/packages/devel/bioc/html/RMassBank.html

RMassBank Workflow – In Summary



Stravs, Schymanski, Singer and Hollender, 2012, *Journal of Mass Spectrometry*. DOI: 10.1002/jms.3131

Improving Availability of High Accuracy MS/MS

Spectra Generated with RMassBank so far

Publically Available RMassBank MS/MS spectra

- http://massbank.normandata.eu/MassBank
- Eawag: 3,102 records (226 compounds)
- UFZ: 3,193 records (158 compounds)

Total Number of "RMassBank Spectra"

- RMassBank records in Uchem-MassBank:
  - 5,312 records (374 compounds) LTQ Orbitrap
  - 153 records (12 compounds) Orbitrap Adducts
  - 1,262 records (151 compounds) Q Exactive Orbitrap
- 3,193 records (158 compounds) from UFZ Orbitrap

>9,900 records of >600 compounds so far!

Why Databases are Useful!

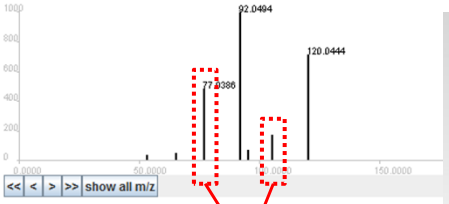
MassBank, RMassBank and “Fail Peaks”

MassBank Record: EZ066612

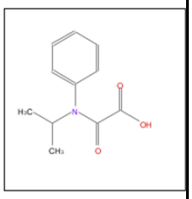
Home | Spectrum | Quick | Peak | Substructure | Browser | Browse | Index | MassBank ID: Go

Propachlor OXA; LC-ESI-ITFT; MS2; 75%; R=15000; [M+H]+

Mass Spectrum



Chemical Structure



```
PK$ANNOTATION:
m/z formula error(ppm)
…
77.0386 C6H5+ -0.35
92.0494 C6H6N+ -0.28
…
105.0447 C6H5N2+ -0.52
…
207.0895 C11H13NO3 [M]
```

Perform a peak search…

Why Databases are Useful

MassBank, RMassBank and “Fail Peaks” – Peak Search

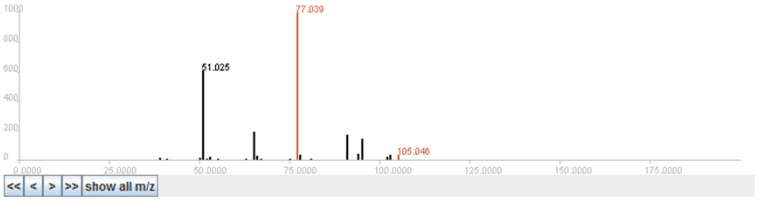
Peak Search Results (Peaks by *m/z* value)

MassBank Record: PB000477

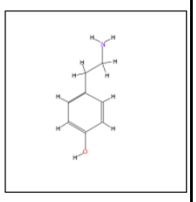
Home | Spectrum | Quick | Peak | Substructure | Prediction | Browser | Batch | Browse | Index | MassBank ID: Go

Tyramine; LC-ESI-QTOF; MS2; CE:50 eV; [M+H]+

Mass Spectrum



Chemical Structure

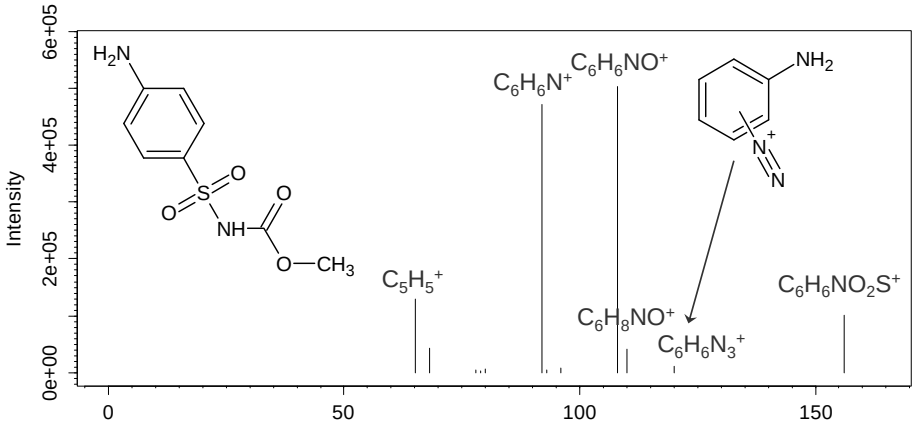


```
ACCESSION: PB000477
RECORD_TITLE: Tyramine; LC-ESI-QTOF; MS2; CE:50 eV; [M+H]+
DATE: 2011.12.06 (Created 2008.01.02)
AUTHORS: Boettcher C, Institute of Plant Biochemistry, Halle, Germany
LICENSE: CC BY-SA
COMMENT: IPB_RECORD: 267
COMMENT: CONFIDENCE: confident structure
```

Why Databases are Useful!

MassBank, RMassBank and More “Fail Peaks”

MassBank record EA015611



Schaffer et al., 2012, *Chemphyschem.*,13, 2688 – mechanisms of N_2 adduct formation

Improving Database Search Results

Experimental Method

- 43 MS/MS spectra extracted from 50 ng spiked sample of Rhine River
  - Uncalibrated query set
  - Recalibrated query set
- 3 Internal Eawag MassBank Databases
  - Uncalibrated, uncleaned records (UUC)
  - Recalibrated, uncleaned records (RUC)
  - Recalibrated, cleaned records (RC)
  - PLUS all other “MassBank” data
- Performed TWO database queries
  - Uncalibrated query set *versus* all MassBank data (incl. UUC, RUC & RC)
  - Recalibrated query set *versus* all MassBank data (incl. UUC, RUC & RC)

Improving Database Search Results

Results – ***Uncalibrated*** Query Set: 43 Spectra

Clean-up makes a big difference!

	Uncalibrated Query Spectra		
Database	**UUC**	**RUC**	**RC**
No. Spectra at Rank 1	8	4	29
Average Rank	3.4	3.4	2.3
Median Rank	3	3	1
Lowest Rank	11	8	9

Best results with Recalibrated, Cleaned-up Database Spectra

Improving Database Search Results

Results – ***Recalibrated*** Query Set: 43 Spectra

	Recalibrated Query Spectra		
Database	**UUC**	**RUC**	**RC**
No. Spectra at Rank 1	5	7	29
Average Rank	15	2	1
Median Rank	8	3	1
Lowest Rank	48	7	9

Worse

Slightly Better

Slightly Better

Best results with Recalibrated, Cleaned-up Database Spectra Independent of whether query spectra are calibrated or not

Improving Molecular Formula Calculation

Exact mass and isotope pattern not sufficient for clear formula assignment

- Several algorithms available to also incorporate MS/MS information
  - MOLGEN-MS/MS[1] – score of MS/MS scaled by ppm error 0-1-X
  - Sirius[2] 1.0[2] – Fragmentation Trees with ColorCodingDP
  - MEF[3] and Mzmine 2[4] – results not shown / not used

Experimental

- 60 MS/MS at 45 % C. E. with C, H, N, O, P, S, Cl, Br, I
  - No explicit number of element restrictions used
- Uncalibrated and recalibrated MS/MS used for formula calculation
  - Uncalibrated at 10 ppm & 5 ppm, Sirius & MOLGEN-MS/MS
  - Recalibrated at 10 ppm & 5 ppm, Sirius & MOLGEN-MS/MS

[1] http://molgen.de/?src=documents/download;
[2] http://bio.informatik.uni-jena.de/sirius2/download/
[3] https://github.com/miquelrojascherto/samsn;
[4] http://mzmine.sourceforge.net/

Improving Molecular Formula Calculation

Results – Formula Calculation: 60 MS/MS; 5 ppm; C,H,N,O,P,S,Cl,Br,I*

	MOLGEN-MS/MS		Sirius[2]	
	Uncal. →	Recal.	Uncal. →	Recal.
Top 1	34 →	50	35 →	52
Top 5	55	58	48	57
Avg. Rank	2.3	***1.5***	1.6	***1.2***
Lowest Rank	20	12	15	8
Total Results	60	60	29	58
Formula absent/crash	-	-	6 / 5	0 / 2

Best results with Recalibrated Spectra irrespective of the program used

* Includes “big” compounds e.g. propaquizafop, 443.1248 Da, 150 possible formulas at 5 ppm

Implications for Non-target Screening

RMassBank Workflow

- Reduces manual work associated with bulk creation of many records
- Creates high quality MS/MS spectra with formula annotation
- Tested extensively on Orbitrap
  - Extension to other instruments in future: NORMAN collaborations
- Over 3,000 MS/MS already publically available; >9,000 total

Recalibration of MS/MS Data

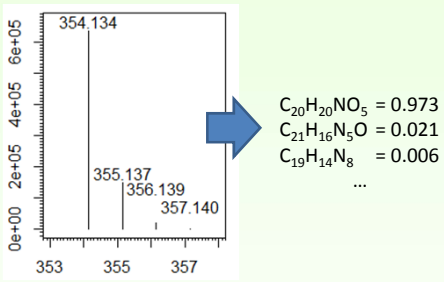
- Recalibration reduces instrument specificity => improve *all* search results
- Improved database search results *irrespective* of quality of query data
- Improved molecular formula assignment using recalibrated data
- Formula annotation (through recalibration) value-adds the spectra
  - Provides valuable information for non-target interpretation

Critical Assessment of Small Molecule Identification

or CASMI for short... http://casmi-contest.org

- Small molecule identification competition – we give you the data and
- You tell CASMI…

Molecular Formula … and …



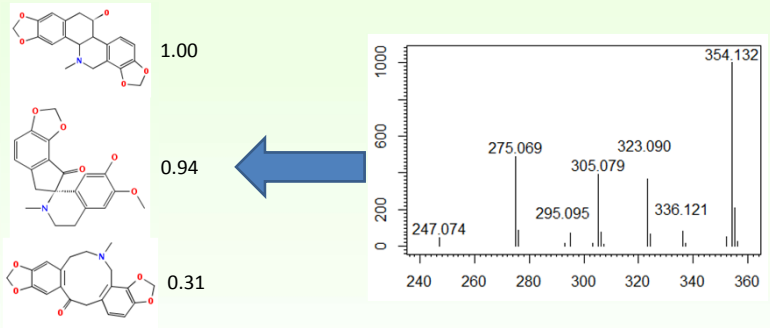
$C_{20}H_{20}NO_5$ = 0.973
$C_{21}H_{16}N_5O$ = 0.021
$C_{19}H_{14}N_8$ = 0.006
…

Critical Assessment of Small Molecule Identification

or CASMI for short... http://casmi-contest.org

- Small molecule identification competition – we give you the data and
- You tell CASMI…

Molecular Formula … and … Molecular Structure



Critical Assessment of Small Molecule Identification

or CASMI for short... http://casmi-contest.org

- Participate in as many or as few challenges as you like
  http://casmi-contest.org/challenges.shtml
- Submit article to a special CASMI issue of ***Metabolites***, by 31 March 2013
- See the website for more details:
  - Join our mailing lists
  - The true winner will be science
- CASMI is coordinated by Steffen Neumann and Emma Schymanski
- Advisory Board: Lloyd Sumner, Susan Richardson, Sandra Perez

Acknowledgements

- Marie-Curie Post Doctoral Fellowship (E. Schymanski),
- Eawag Discretionary Funds (M. Stravs)
- Coauthors: ***Michael Stravs***, Heinz Singer, Juliane Hollender
- R & MassBank help: Steffen Neumann, Michael Gerlich, Carsten Kuhl, (IPB)
- Discussions: Matthias Ruff, Martin Loos (Eawag)
- Discussions: Tobias Schulze, Martin Krauss, Werner Brack (UFZ)
- Programs: Markus Meringer (MOLGEN-MS/MS), Florian Rasche (Sirius), Miguel Rojas-Cherto & Egon Willighagen (MEF)
- MassBank & Naming Rights: Prof. Takaaki Nishioka
- NORMAN Association

RMassBank

Any Questions?

Emma Schymanski, Michael Stravs, Heinz Singer, Juliane Hollender

massbank@eawag.ch (ES, MS)
emma.schymanski@eawag.ch

DOI: 10.1002/jms.3131
All Details Contained Within!