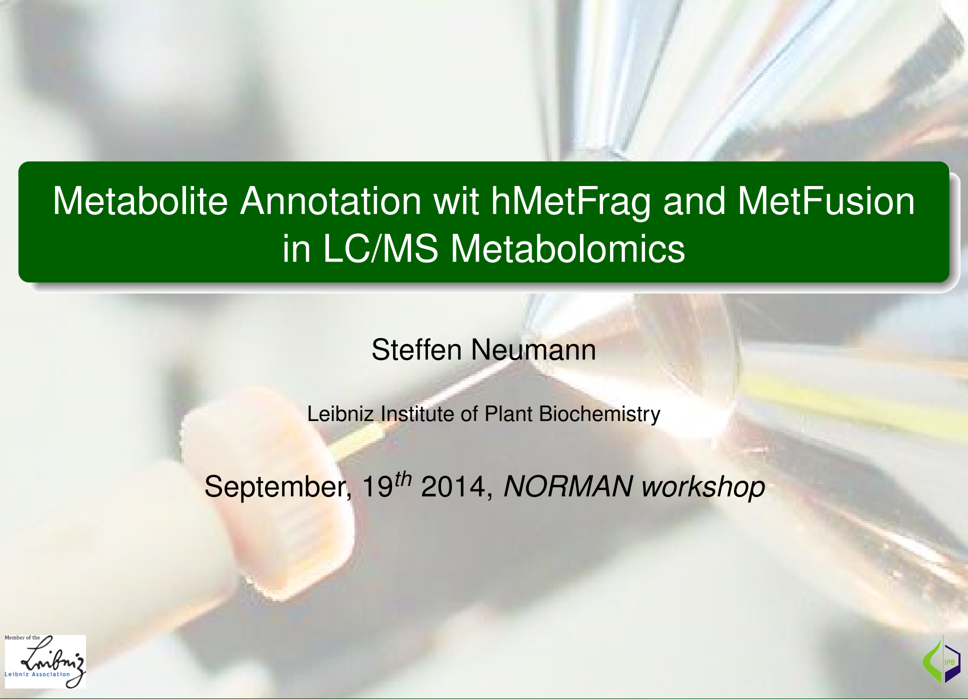


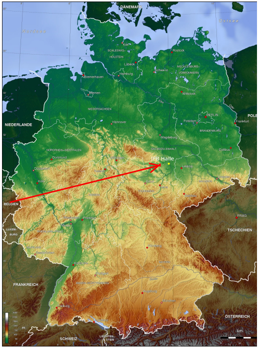
Metabolite Annotation wit hMetFrag and MetFusion in LC/MS Metabolomics

Steffen Neumann

Leibniz Institute of Plant Biochemistry

September, 19th 2014, *NORMAN workshop*

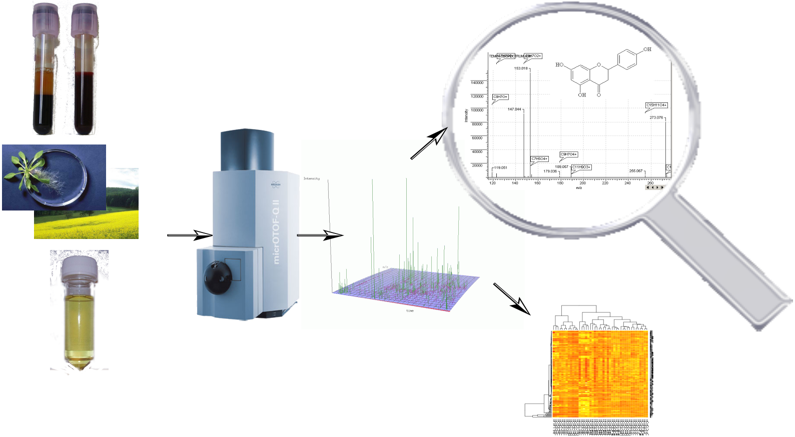
Leibniz Institute of Plant Biochemistry



- Established 1958 by Kurth Mothes
- Plant diversity, development and adaptation
- Plant production, -protection and biologically active compounds
- About 180 Researchers, including 5-15 Bioinformaticians

Member of the Leibniz Association

Metabolomics – The Pipeline



Outline

1. Identification of Molecules from LC-MS/MS

2. Critical Assessment of Small Molecule Identification

3. Summary & Outlook

Outline

1. Identification of Molecules from LC-MS/MS

2. Critical Assessment of Small Molecule Identification

3. Summary & Outlook

What *is* Identification ??

What *is* Identification ??

"Proposed minimum reporting standards for chemical analysis"
of the Chemical Analysis Working Group (CAWG)
of the Metabolomics Standards Initiative (MSI) for non-novel compounds

1. Identified compound:
   retention time/index and mass spectrum, retention time and NMR spectrum, accurate mass and tandem MS, accurate mass and isotope pattern, full ^{1}H and/or ^{13}C NMR, 2-D NMR spectra of in-house measured authentic reference compound Optionally (esp. for unambiguous stereo configuration) selective solvent extraction, retention time, m/z, photodiode array spectra, λ_{max} and ϵ_{max}, chemical derivatization, isotope labeling, 2D NMR, IR spectra, etc.
2. Putative compound:
   without chemical reference standards, based upon physico-chemical properties and/or spectral similarity with (public/commercial) spectral libraries
3. Putative compound class
4. Unknown compounds – but can be differentiated and quantified

Member of the Leibniz Association

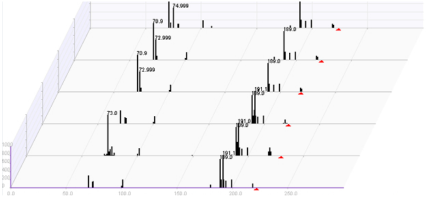
And how confident are you ?

Emma Levels

1. *The* compound structure
2. (Few) Isomers
3. Compound class
4. Molecular formula
5. Accurate Mass

Schymanski *et al.*, Environmental Science & Technology (2014)

Identification *with* reference spectra: MassBank

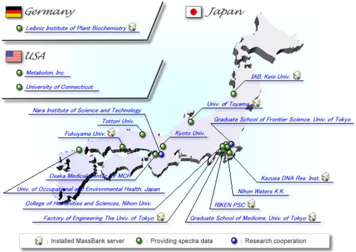


Other (ESI) libraries:

- Metlin
- HMDB
- MMCD
- NIST
- . . .

- `http://www.massbank.jp/`
- Open Data, Open Consortium
- IPB Halle first European server: `msbi.ipb-halle.de/MassBank/`

≈ 4 275 compounds with ≈27 500 MS^2 spectra (6/2013)

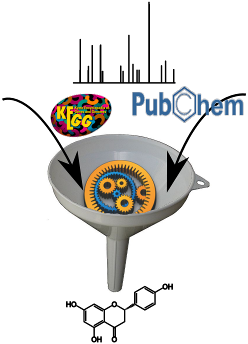


Germany
Leibniz Institute of Plant Biochemistry
USA
Metabolon, Inc.
University of Connecticut
Japan
IAB, Keio Univ.
Univ. of Toyama
Graduate School of Frontier Science, Univ. of Tokyo
Nara Institute of Science and Technology
Tottori Univ.
Fukuyama Univ.
Kyoto Univ.
Osaka Medical Center for MCH
Univ. of Occupational and Environmental Health, Japan
College of Humanities and Sciences, Nihon Univ.
Factory of Engineering The Univ. of Tokyo
Graduate School of Medicine, Univ. of Tokyo
RIKEN PSC
Nihon Waters K.K.
Kazusa DNA Res. Inst.
: Installed MassBank server
● : Providing spectra data
● : Research cooperation

Horai, H., *et al.* MassBank: a public repository for sharing mass spectral data for life sciences. Journal of Mass Spectrometry (2010)

Identification ***without*** reference spectra

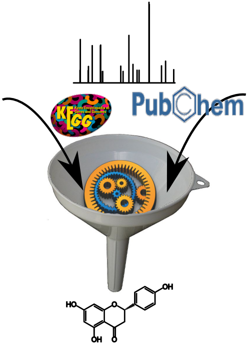
- Spectral libraries (even MassBank) inherently incomplete
- General purpose compound databases:
  - KEGG Compound: 14 067
  - PubChem: 27 million
  - ChemSpider: 25 million
- Known molecular structures
- But: no spectra search → “known unknowns”



Wolf, S., Schmidt, S., Müller-Hannemann, M., and Neumann, S.
In silico fragmentation for computer assisted identification of metabolite mass spectra. BMC Bioinf. (2010)

Identification *without* reference spectra

- Spectral libraries (even MassBank) inherently incomplete
- General purpose compound databases:
  - KEGG Compound: 14 067
  - PubChem: 27 million
  - ChemSpider: 25 million
- Known molecular structures
- But: no spectra search → “known unknowns”

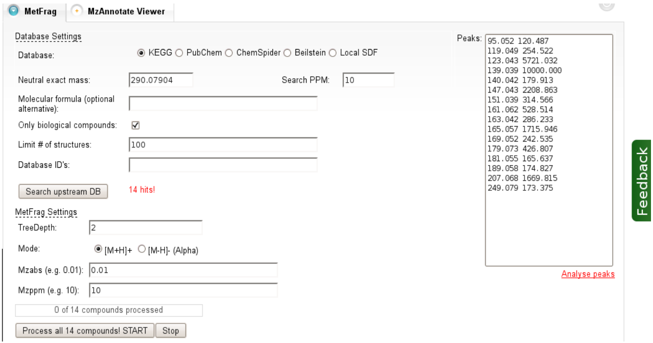


`msbi.ipb-halle.de/MetFrag/` provides this search:

1. Search compound database for precursor mass
2. *In-silico* fragmentation of molecular structure
3. Score measured vs. “predicted” peaks

Wolf, S., Schmidt, S., Müller-Hannemann, M., and Neumann, S.
In silico fragmentation for computer assisted identification of metabolite mass spectra. BMC Bioinf. (2010)

MetFrag: an example



Database Settings

Database: ● KEGG ○ PubChem ○ ChemSpider ○ Beilstein ○ Local SDF

Neutral exact mass: 290.07904 Search PPM: 10

Molecular formula (optional alternative):

Only biological compounds: ☑

Limit # of structures: 100

Database ID's:

Search upstream DB 14 hits!

MetFrag Settings

TreeDepth: 2

Mode: ● [M+H]+ ○ [M-H]- (Alpha)

Mzabs (e.g. 0.01): 0.01

Mzppm (e.g. 10): 10

0 of 14 compounds processed

Process all 14 compounds! START Stop

Peaks:

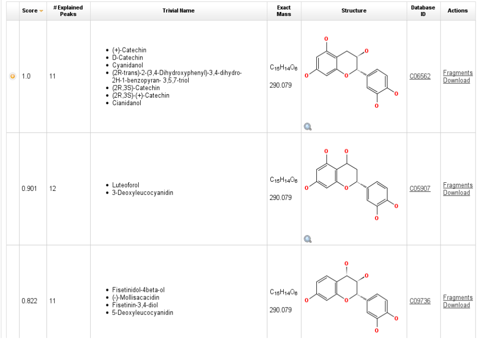
```
95.052 120.487
119.049 254.522
123.043 5721.032
139.039 10000.000
140.042 179.913
147.043 2208.863
151.039 314.566
161.062 528.514
163.042 286.233
165.057 1715.946
169.052 242.535
179.073 426.807
181.055 165.637
189.058 174.827
207.068 1669.815
249.079 173.375
```

Analyse peaks

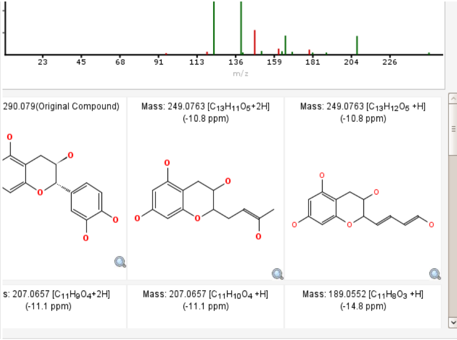
- Search 290.08Da with 10ppm
- 14 KEGG hits
- MetFrag takes ca. 10sec
- Results ordered by score
- Details / Fragment view
- Excel/SDF download
- Feedback form
- Local version available

Wolf, S., Schmidt, S., Müller-Hannemann, M., and Neumann, S.
In silico fragmentation for computer assisted identification of metabolite mass spectra. BMC Bioinf. (2010)

MetFrag: an example



	Score	# Explained Peaks	Trivial Name	Exact Mass	Structure	Database ID	Actions
	1.0	11	• (+)-Catechin • D-Catechin • Cyanidanol • (2R-trans)-2-(3,4-Dihydroxyphenyl)-3,4-dihydro-2H-1-benzopyran- 3,5,7-triol • (2R,3S)-Catechin • (2R,3S)-(+)-Catechin • Cianidanol	$C_{15}H_{14}O_6$ 290.079		C06562	Fragments Download
	0.901	12	• Luteoforol • 3-Deoxyleucocyanidin	$C_{15}H_{14}O_6$ 290.079		C05907	Fragments Download
	0.822	11	• Fisetinidol-4beta-ol • (-)-Mollisacacidin • Fisetinin-3,4-diol • 5-Deoxyleucocyanidin	$C_{15}H_{14}O_6$ 290.079		C09736	Fragments Download



290.079(Original Compound)

Mass: 249.0763 [$C_{13}H_{11}O_5$+2H] (-10.8 ppm)

Mass: 249.0763 [$C_{13}H_{12}O_5$ +H] (-10.8 ppm)

s: 207.0657 [$C_{11}H_9O_4$+2H] (-11.1 ppm)

Mass: 207.0657 [$C_{11}H_{10}O_4$ +H] (-11.1 ppm)

Mass: 189.0552 [$C_{11}H_8O_3$ +H] (-14.8 ppm)

- Search 290.08Da with 10ppm
- 14 KEGG hits
- MetFrag takes ca. 10sec
- Results ordered by score

- Details / Fragment view
- Excel/SDF download
- Feedback form
- Local version available

Member of the Leibniz Leibniz Association

Wolf, S., Schmidt, S., Müller-Hannemann, M., and Neumann, S.
In silico fragmentation for computer assisted identification of metabolite mass spectra. BMC Bioinf. (2010)

IPB

Rest of the world

There are other tools for Metabolite annotation/identification, incomplete list and some literature:
Since before 2010

- ACD Fragmenter (ACD)
- MassFrontier (HighChem), Grant *et al.* Anal. Chem (2008)

After 2010

- MAGMa (Ridder *et al.,*) CASMI 2013
- CFM (Allen *et al.,*) CASMI 2013
- Rational Numbers (Sweeny *et al.,*) CASMI 2013
- SIRIUS (Böcker *et al.,*)
- MassFrontier (HighChem), (Viant *et al.*)

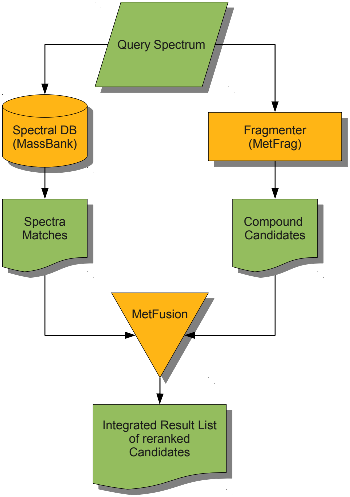
Schymanski, E.L. and Neumann, S. CASMI: And the Winner is . . . Metabolites (2013) / Nishioka *t al.*, Winners of CASMI2013: Automated Tools and Challenge Data, J Mass Spectrometry (2014) / Böcker *et al.*, J. Cheminf

Integrating MassBank & MetFrag ?

Both tools are great ! Both aid in identifying compounds:

- MassBank: **reference** spectra
  - Actual, real measurements under real conditions
  - Limited coverage, metabolite you look for is (almost) always missing
  - Results based on *spectral* similarity
- MetFrag: **large** compound DB
  - Upstream *exact mass* or *molecular formula* search
  - But: A prediction is a prediction, nothing else
  - Molecular re-arrangements mostly ignored

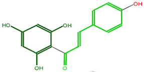
Integrating MassBank & MetFrag !

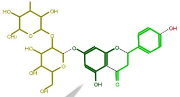


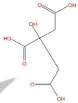
- MetFusion: **Combine** results:
  - Parallel queries in MassBank and MetFrag
  - Pairwise chemical similarities between result sets
- Calculate *integrated* score
- Major improvement of identification power and accuracy
- `msbi.ipb-halle.de/MetFusion/`

Gerlich M., Neumann S., MetFusion: integration of compound identification strategies
Journal of Mass Spectrometry 48 (3), 291-298

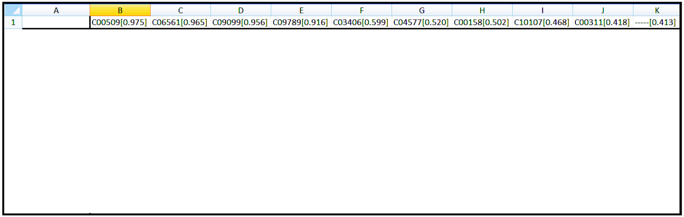
Integrating MassBank + MetFrag: Chemical *Similarity*



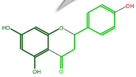


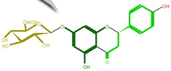


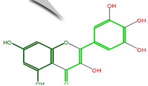
MassBank



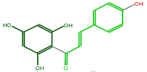
	A	B	C	D	E	F	G	H	I	J	K
1		C00509[0.975]	C06561[0.965]	C09099[0.956]	C09789[0.916]	C03406[0.599]	C04577[0.520]	C00158[0.502]	C10107[0.468]	C00311[0.418]	-----[0.413]

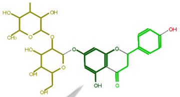


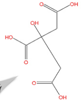




Integrating MassBank + MetFrag: Chemical *Similarity*

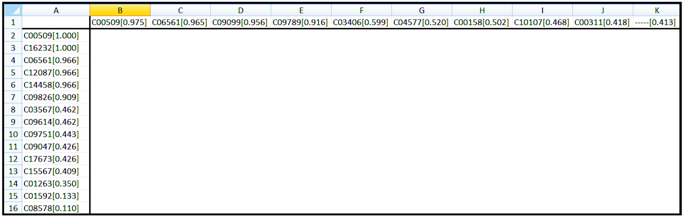




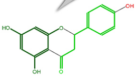


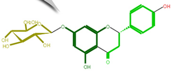
MassBank

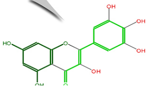
MetFrag



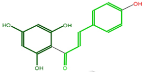
	A	B	C	D	E	F	G	H	I	J	K
1		C00509[0.975]	C06561[0.965]	C09099[0.956]	C09789[0.916]	C03406[0.599]	C04577[0.520]	C00158[0.502]	C10107[0.468]	C00311[0.418]	-----[0.413]
2	C00509[1.000]										
3	C16232[1.000]										
4	C06561[0.966]										
5	C12087[0.966]										
6	C14458[0.966]										
7	C09826[0.909]										
8	C03567[0.462]										
9	C09614[0.462]										
10	C09751[0.443]										
11	C09047[0.426]										
12	C17673[0.426]										
13	C15567[0.409]										
14	C01263[0.350]										
15	C01592[0.133]										
16	C08578[0.110]										

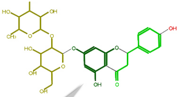


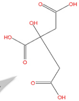




Integrating MassBank + MetFrag: Chemical *Similarity*

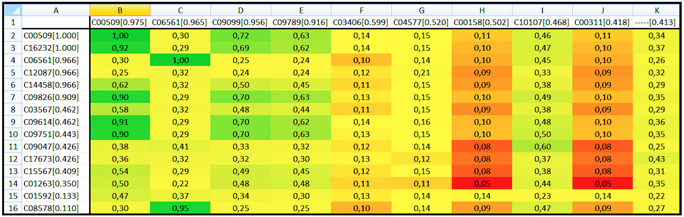




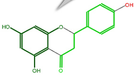


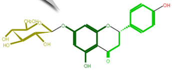
MassBank

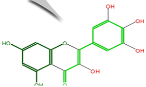
MetFrag



	A	B	C	D	E	F	G	H	I	J	K
1		C00509[0.975]	C06561[0.965]	C09099[0.956]	C09789[0.916]	C03406[0.599]	C04577[0.520]	C00158[0.502]	C10107[0.468]	C00311[0.418]	-----[0.413]
2	C00509[1.000]	1,00	0,30	0,72	0,63	0,14	0,15	0,11	0,46	0,11	0,34
3	C16232[1.000]	0,92	0,29	0,69	0,62	0,14	0,15	0,10	0,47	0,10	0,37
4	C06561[0.966]	0,30	1,00	0,25	0,24	0,10	0,14	0,10	0,45	0,10	0,26
5	C12087[0.966]	0,25	0,32	0,24	0,24	0,12	0,21	0,09	0,33	0,09	0,32
6	C14458[0.966]	0,62	0,32	0,50	0,45	0,11	0,15	0,09	0,38	0,09	0,29
7	C09826[0.909]	0,90	0,29	0,70	0,63	0,13	0,15	0,10	0,49	0,10	0,35
8	C03567[0.462]	0,58	0,32	0,48	0,44	0,11	0,15	0,09	0,38	0,09	0,29
9	C09614[0.462]	0,91	0,29	0,70	0,62	0,14	0,16	0,10	0,48	0,10	0,36
10	C09751[0.443]	0,90	0,29	0,70	0,63	0,13	0,15	0,10	0,50	0,10	0,35
11	C09047[0.426]	0,38	0,41	0,33	0,32	0,12	0,14	0,08	0,60	0,08	0,25
12	C17673[0.426]	0,36	0,32	0,32	0,30	0,13	0,12	0,08	0,37	0,08	0,43
13	C15567[0.409]	0,54	0,29	0,49	0,45	0,12	0,15	0,08	0,38	0,08	0,31
14	C01263[0.350]	0,50	0,22	0,48	0,48	0,11	0,11	0,05	0,44	0,05	0,35
15	C01592[0.133]	0,47	0,37	0,34	0,30	0,13	0,14	0,14	0,23	0,14	0,22
16	C08578[0.110]	0,30	0,95	0,25	0,25	0,10	0,14	0,09	0,47	0,09	0,27



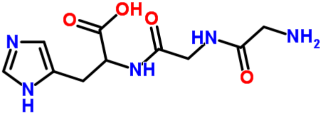




Chemical similarity color-coded: high . . . medium . . . low

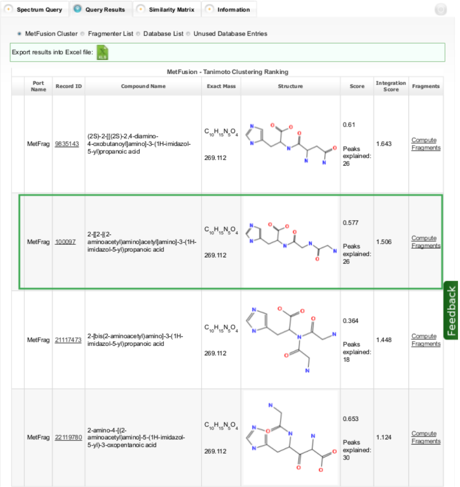
MetFusion: Example Gly-Gly-His

- Example: tripeptide Gly-Gly-His
- Nominal mass spectrum from NIST MS/MS library
- PubChem: 211 candidates
- MassBank: only His and Gly spectra
- MetFrag: correct ranked 23rd



Gerlich M., Neumann S., MetFusion: integration of compound identification strategies
Journal of Mass Spectrometry 48 (3), 291-298, 2013

MetFusion: Example Gly-Gly-His



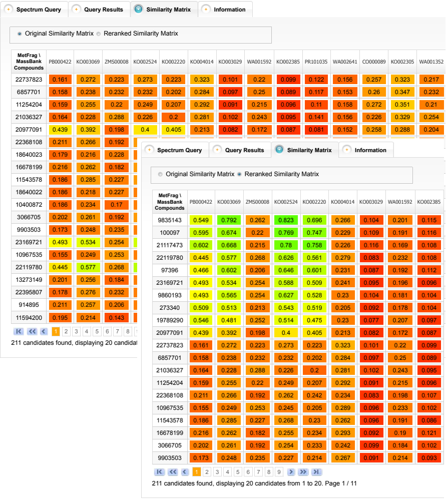
Spectrum Query
Query Results
Similarity Matrix
Information
● MetFusion Cluster ○ Fragmenter List ○ Database List ○ Unused Database Entries
Export results into Excel file:
Feedback

MetFusion - Tanimoto Clustering Ranking

	Port Name	Record ID	Compound Name	Exact Mass	Structure	Score	Integration Score	Fragments
	MetFrag	9835143	(2S)-2-[[(2S)-2,4-diamino-4-oxobutanoyl]amino]-3-(1H-imidazol-5-yl)propanoic acid	$C_{10}H_{15}N_5O_4$ 269.112		0.61 Peaks explained: 26	1.643	Compute Fragments
	MetFrag	100097	2-[[2-[(2-aminoacetyl)amino]acetyl]amino]-3-(1H-imidazol-5-yl)propanoic acid	$C_{10}H_{15}N_5O_4$ 269.112		0.577 Peaks explained: 26	1.506	Compute Fragments
	MetFrag	21117473	2-[bis(2-aminoacetyl)amino]-3-(1H-imidazol-5-yl)propanoic acid	$C_{10}H_{15}N_5O_4$ 269.112		0.364 Peaks explained: 18	1.448	Compute Fragments
	MetFrag	22119780	2-amino-4-[(2-aminoacetyl)amino]-5-(1H-imidazol-5-yl)-3-oxopentanoic acid	$C_{10}H_{15}N_5O_4$ 269.112		0.653 Peaks explained: 30	1.124	Compute Fragments

Gerlich M., Neumann S., MetFusion: integration of compound identification strategies
Journal of Mass Spectrometry 48 (3), 291-298, 2013

MetFusion: Example Gly-Gly-His



Spectrum Query | Query Results | Similarity Matrix | Information

○ Original Similarity Matrix ○ Reranked Similarity Matrix

MetFrag \ MassBank Compounds	PB000422	KO003069	ZMS00008	KO002524	KO002220	KO004014	KO003029	WA001592	KO002385	PR101035	WA002641	CO000089	KO002305	WA001352
22737823	0.161	0.272	0.223	0.273	0.223	0.323	0.101	0.22	0.099	0.122	0.156	0.257	0.323	0.217
6857701	0.158	0.238	0.232	0.232	0.202	0.284	0.097	0.25	0.089	0.117	0.153	0.26	0.347	0.232
11254204	0.159	0.255	0.22	0.249	0.207	0.292	0.091	0.215	0.096	0.11	0.158	0.272	0.351	0.21
21036327	0.164	0.228	0.288	0.226	0.2	0.281	0.102	0.243	0.095	0.141	0.156	0.226	0.329	0.254
20977091	0.439	0.392	0.198	0.4	0.405	0.213	0.082	0.172	0.087	0.081	0.152	0.258	0.288	0.204
22368108	0.211	0.266	0.192											
18640023	0.179	0.216	0.228											
16678199	0.216	0.262	0.182											
11543578	0.186	0.285	0.227											
18640022	0.186	0.218	0.227											
10400872	0.186	0.234	0.17											
3066705	0.202	0.261	0.192											
9903503	0.173	0.248	0.235											
23169721	0.493	0.534	0.254											
10967535	0.155	0.249	0.253											
22119780	0.445	0.577	0.268											
13273149	0.201	0.256	0.184											
22395807	0.178	0.276	0.232											
914895	0.211	0.257	0.206											
11594200	0.195	0.214	0.143											

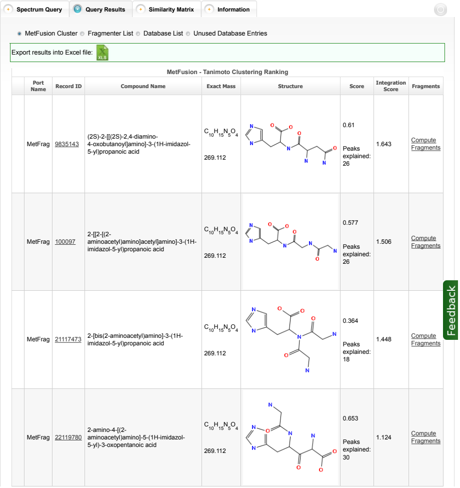
211 candidates found, displaying 20 candidat

Spectrum Query | Query Results | Similarity Matrix | Information

○ Original Similarity Matrix ● Reranked Similarity Matrix

MetFrag \ MassBank Compounds	PB000422	KO003069	ZMS00008	KO002524	KO002220	KO004014	KO003029	WA001592	KO002385
9835143	0.549	0.792	0.262	0.823	0.696	0.266	0.104	0.201	0.115
100097	0.595	0.674	0.22	0.769	0.747	0.229	0.109	0.191	0.116
21117473	0.602	0.668	0.215	0.78	0.758	0.226	0.116	0.169	0.108
22119780	0.445	0.577	0.268	0.626	0.561	0.279	0.083	0.232	0.108
97396	0.466	0.602	0.206	0.646	0.601	0.231	0.087	0.192	0.112
23169721	0.493	0.534	0.254	0.588	0.509	0.241	0.095	0.196	0.096
9860193	0.493	0.565	0.254	0.627	0.528	0.23	0.104	0.181	0.104
273340	0.509	0.513	0.213	0.543	0.519	0.205	0.092	0.178	0.104
19789290	0.546	0.481	0.252	0.514	0.475	0.23	0.077	0.207	0.097
20977091	0.439	0.392	0.198	0.4	0.405	0.213	0.082	0.172	0.087
22737823	0.161	0.272	0.223	0.273	0.223	0.323	0.101	0.22	0.099
6857701	0.158	0.238	0.232	0.232	0.202	0.284	0.097	0.25	0.089
21036327	0.164	0.228	0.288	0.226	0.2	0.281	0.102	0.243	0.095
11254204	0.159	0.255	0.22	0.249	0.207	0.292	0.091	0.215	0.096
22368108	0.211	0.266	0.192	0.262	0.242	0.234	0.083	0.198	0.107
10967535	0.155	0.249	0.253	0.245	0.205	0.289	0.096	0.233	0.102
11543578	0.186	0.285	0.227	0.268	0.23	0.262	0.096	0.191	0.086
16678199	0.216	0.262	0.182	0.255	0.234	0.293	0.092	0.19	0.121
3066705	0.202	0.261	0.192	0.254	0.233	0.242	0.099	0.184	0.102
9903503	0.173	0.248	0.235	0.227	0.214	0.267	0.091	0.214	0.093

211 candidates found, displaying 20 candidates from 1 to 20. Page 1 / 11



Spectrum Query | Query Results | Similarity Matrix | Information

● MetFusion Cluster ○ Fragmenter List ○ Database List ○ Unused Database Entries

Export results into Excel file:

MetFusion - Tanimoto Clustering Ranking

	Port Name	Record ID	Compound Name	Exact Mass	Structure	Score	Integration Score	Fragments
	MetFrag	9835143	(2S)-2-[[(2S)-2,4-diamino-4-oxobutanoyl]amino]-3-(1H-imidazol-5-yl)propanoic acid	$C_{10}H_{15}N_5O_4$ 269.112		0.61 Peaks explained: 26	1.643	Compute Fragments
	MetFrag	100097	2-[[2-[(2-aminoacetyl)amino]acetyl]amino]-3-(1H-imidazol-5-yl)propanoic acid	$C_{10}H_{15}N_5O_4$ 269.112		0.577 Peaks explained: 26	1.506	Compute Fragments
	MetFrag	21117473	2-[bis(2-aminoacetyl)amino]-3-(1H-imidazol-5-yl)propanoic acid	$C_{10}H_{15}N_5O_4$ 269.112		0.364 Peaks explained: 18	1.448	Compute Fragments
	MetFrag	22119780	2-amino-4-[(2-aminoacetyl)amino]-5-(1H-imidazol-5-yl)-3-oxopentanoic acid	$C_{10}H_{15}N_5O_4$ 269.112		0.653 Peaks explained: 30	1.124	Compute Fragments

Gerlich M., Neumann S., MetFusion: integration of compound identification strategies
Journal of Mass Spectrometry 48 (3), 291-298, 2013

MetFusion: Evaluation

- 345 compounds, 89 to 837 Da
- flavonoids, isoflavonoids, steroids, amino acids, carboxylic acids, polyketids, prenol and sterol lipids, glucosides, drugs, toxins, alcohols, carbohydrates
- 1062 spectra from MassBank, all QTOF
- Median 707 candidates from PubChem
- Leave-*some*-out MassBank “pruning”:

		Max. chemical similarity to MassBank
		=1
MetFusion	Rank	1
	RRP	1

Gerlich M., Neumann S., MetFusion: integration of compound identification strategies
Journal of Mass Spectrometry 48 (3), 291-298, 2013

MetFusion: Evaluation

- 345 compounds, 89 to 837 Da
- flavonoids, isoflavonoids, steroids, amino acids, carboxylic acids, polyketids, prenol and sterol lipids, glucosides, drugs, toxins, alcohols, carbohydrates
- 1062 spectra from MassBank, all QTOF
- Median 707 candidates from PubChem
- Leave-*some*-out MassBank "pruning":

		Max. chemical similarity to MassBank	
		<1	=1
MetFusion	Rank	4	1
	RRP	0.993	1

Gerlich M., Neumann S., MetFusion: integration of compound identification strategies
Journal of Mass Spectrometry 48 (3), 291-298, 2013

MetFusion: Evaluation

- 345 compounds, 89 to 837 Da
- flavonoids, isoflavonoids, steroids, amino acids, carboxylic acids, polyketids, prenol and sterol lipids, glucosides, drugs, toxins, alcohols, carbohydrates
- 1062 spectra from MassBank, all QTOF
- Median 707 candidates from PubChem
- Leave-*some*-out MassBank "pruning":

		Max. chemical similarity to MassBank		
		<0.9	<1	=1
MetFusion	Rank	**7**	4	1
	RRP	**0.991**	0.993	1

Gerlich M., Neumann S., MetFusion: integration of compound identification strategies
Journal of Mass Spectrometry 48 (3), 291-298, 2013

MetFusion: Evaluation

- 345 compounds, 89 to 837 Da
- flavonoids, isoflavonoids, steroids, amino acids, carboxylic acids, polyketids, prenol and sterol lipids, glucosides, drugs, toxins, alcohols, carbohydrates
- 1062 spectra from MassBank, all QTOF
- Median 707 candidates from PubChem
- Leave-*some*-out MassBank “pruning”:

		Max. chemical similarity to MassBank				
		<0.7	<0.8	**<0.9**	<1	=1
MetFusion	Rank	10	8	**7**	4	1
	RRP	0.986	0.990	**0.991**	0.993	1

Gerlich M., Neumann S., MetFusion: integration of compound identification strategies
Journal of Mass Spectrometry 48 (3), 291-298, 2013

Outline

1. Identification of Molecules from LC-MS/MS

2. Critical Assessment of Small Molecule Identification

3. Summary & Outlook

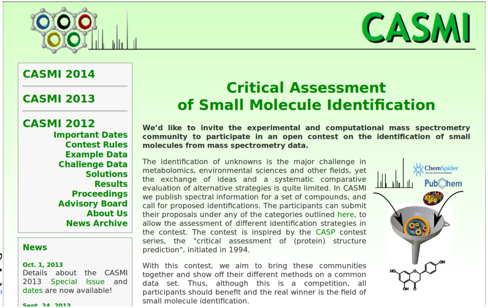
What is CASMI ?

We invited the experimental and computational mass spectrometry community to participate in an open **contest** on the **identification** of small molecules from mass spectrometry data:

Critical **A**ssessment of **S**mall **M**olecule **I**dentification

---

www.casmi-contest.org



CASMI

CASMI 2014

CASMI 2013

CASMI 2012

Important Dates
Contest Rules
Example Data
Challenge Data
Solutions
Results
Proceedings
Advisory Board
About Us
News Archive

News

Oct. 1, 2013
Details about the CASMI 2013 Special Issue and dates are now available!

Sept. 24, 2013

Critical Assessment of Small Molecule Identification

We'd like to invite the experimental and computational mass spectrometry community to participate in an open contest on the identification of small molecules from mass spectrometry data.

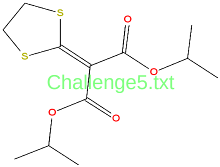
The identification of unknowns is the major challenge in metabolomics, environmental sciences and other fields, yet the exchange of ideas and a systematic comparative evaluation of alternative strategies is quite limited. In CASMI we publish spectral information for a set of compounds, and call for proposed identifications. The participants can submit their proposals under any of the categories outlined here, to allow the assessment of different identification strategies in the contest. The contest is inspired by the CASP contest series, the "critical assessment of (protein) structure prediction", initiated in 1994.

With this contest, we aim to bring these communities together and show off their different methods on a common data set. Thus, although this is a competition, all participants should benefit and the real winner is the field of small molecule identification.

E. Schymanski, S. Neumann

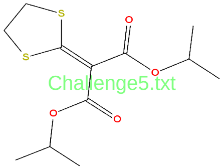
CASMI as an exercise for post-processing analysis

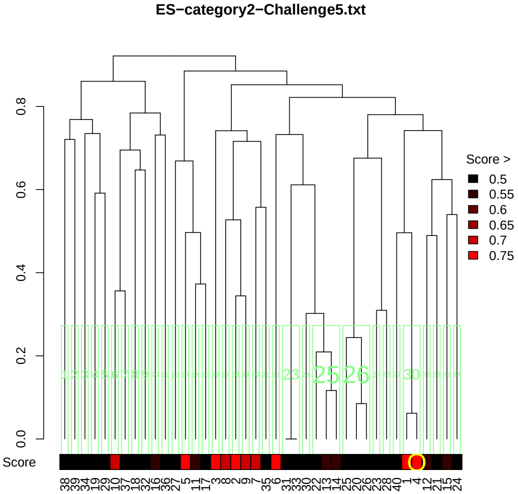
Isoprothiolane



CASMI as an exercise for post-processing analysis

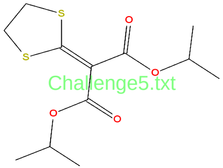
Isoprothiolane

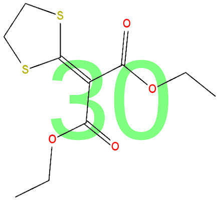


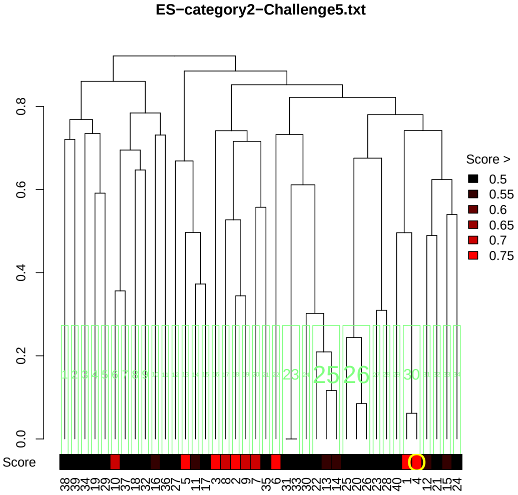


CASMI as an exercise for post-processing analysis

Isoprothiolane

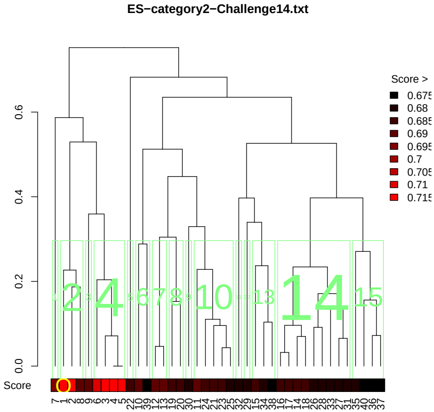


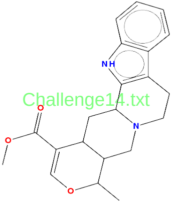


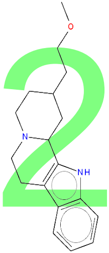


CASMI as an exercise for post-processing analysis

Tetrahydroalstonine

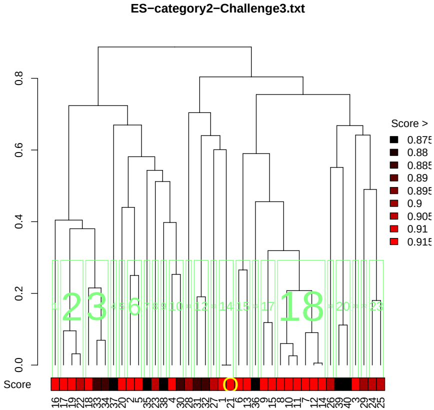


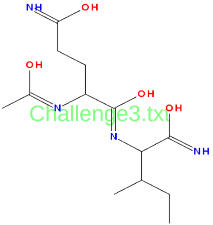


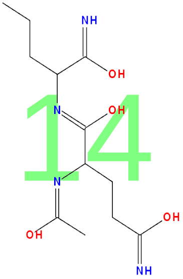


CASMI as an exercise for post-processing analysis

Acetyl-Gln-Leu-amide







KEGG extrapolation

The higher the similarity between an unknown and the reference spectrum, the better the identification result.

⇒ How similar are compounds between MassBank and KEGG ?

Number of KEGG compounds for which a MassBank record with a chemical similarity greater or equal the threshold exists:

Similarity	>0.7	>0.8	**>0.9**	=1.0
KEGG Entries	5 513	4 068	**2 690**	1 470

MetFusion has median rank 7 on test data if reference spectra with >0.9 chemical similarity are available. **If** that generalises to all KEGG compounds, we expect for 1 345 metabolites rank 7 or better.

Gerlich M., S Neumann S., MetFusion: integration of compound identification strategies
Journal of Mass Spectrometry 48 (3), 291-298

Outline

1 Identification of Molecules from LC-MS/MS

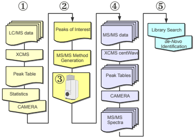
2 Critical Assessment of Small Molecule Identification

3 Summary & Outlook

Summary

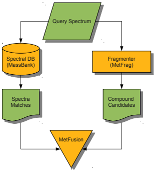
MetShot aproach

- High-quality MS/MS spectra
- Biologically relevant features
- Signal processing with XCMS and CAMERA



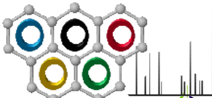
Metabolite identification

- With reference spectra (MassBank)
- With *in silico* tools (MetFrag)
- With a combination thereof (MetFusion)



CASMI 2012 – 2013 – 2014

- Open contest for structure elucidation
- Good results, but manual methods don't scale
- Becoming a regular contest series



Thanks to . . .

© 2002-2004 by Stadt Halle (Saale) Kartengrundlage FB Vermessung und Geodaten

MSBI @ IPB Halle:

- Christoph Ruttkies (MetFrag)
- Michael Gerlich (MetFusion)
- D. Schober, S. Mönchgesang

Acknowledgements:

- Profs. Nishioka, Arita (MassBank Japan)
- Emma Schymanski (Eawag, Zürich)
- Funding: DFG grant “ChemFrag”,
  EU FP7 “COSMOS”,
  EU FP7 “SOLUTIONS”

Alumni (excerpt):

- Dr. Sebastian Wolf (now Bruker Biospin)
- Dr. Ralf Tautenhahn (SCRIPPS, now ThermoFischer)
- Carsten Kuhl (now Bruker Biospin)
- Björn Egert (now Max Rübner Institute)

DFG

SEVENTH FRAMEWORK PROGRAMME

Member of the Leibniz Association

IPB

Only supplemental slides beyond this point