

Identification of Metabolites tools and methods

Ways to identify compounds

- Accurate mass - Normally first step to finding the compound
 - database search using small ppm window
 - Adducts :-{
- Retention time - running a standard or local method
 - database lookup
 - Drift :-{
- Fragmentation patterns -
 - Data - incomplete databases

Accurate mass

- Features must be characterised to known neutral mass
- Mass Spectrometers have errors on mass - leads to larger search space.
- Too many isobaric compounds

BMC Bioinformatics



Research article

Open Access

Metabolomic database annotations via query of elemental compositions: Mass accuracy is insufficient even at less than 1 ppm

Tobias Kind* and Oliver Fiehn

Retention time

- Retention time against standards is a powerful tool
 - Must have the standards
 - Have to keep the same gradient for database lookup method
 - Columns and LC methods are changing
 - Don't always use LC method - Rapid injection methods

Fragmentation

- Fragmentation is a (sort of) 2D representation of the compound
- Lots of issues
 - Specificity of isolation window
 - Scoring algorithms
 - Access to standards



Matching spectra to compounds

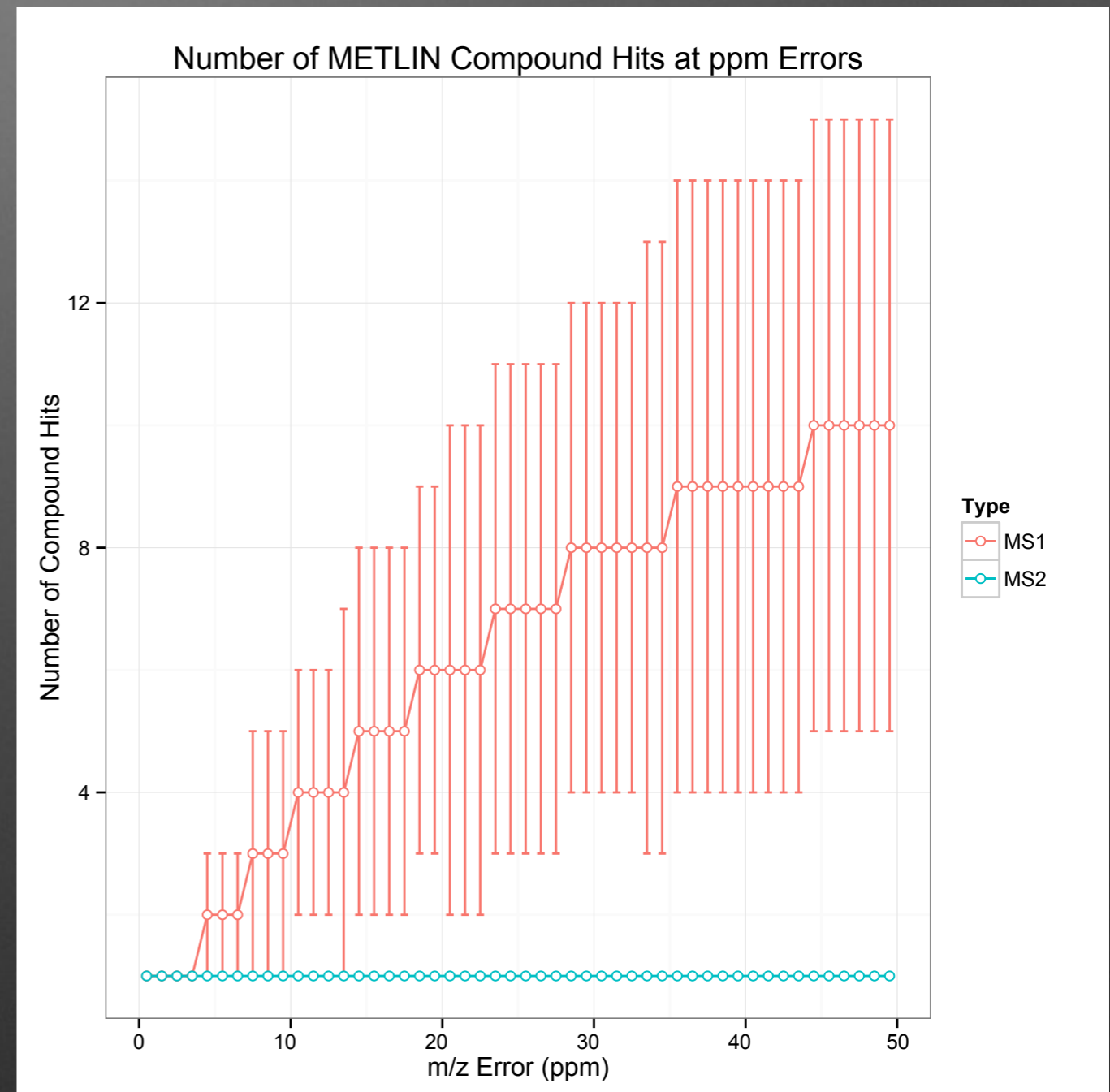
Scoring methods

- Either matching to a standard or a database
- Many different metrics some are more sensitive
 - Cosine similarity suffers from low resolution issues and should be low sensitivity and specificity
 - X-Rank - a very cool algorithm looks at peaks that matter the most
 - Many more

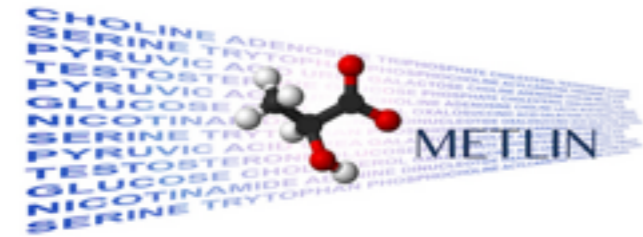
$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

Why it works

- Precursor selection with fragmentation data is highly specific
- The scoring is only needed for ranking purposes



Databases



Statistics

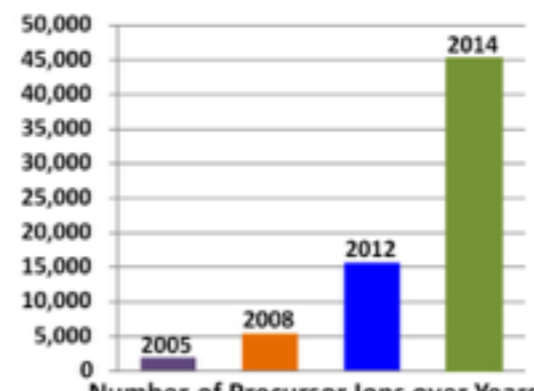
- # Metabolites: 240,588
- # High Resolution MS/MS Spectra: 66,764
- # Metabolites w/ High Resolution MS/MS: 12,855

[example](#) | [details...](#)

Functionality

- Single & Batch** Precursor Ion (m/z) searching
- Single & Multiple** Fragment Ion (m/z) searching
- Neutral Loss** searching
- De Novo** Fragment Characterization

NIST Tandem Mass Spectral Library 2014



9,345 Compounds
45,298 Precursor Ions
234,284 Spectra
 ~90% Positive Ion Spectra
 ~10% Negative Ion Spectra

Instrument Type	Precursor Ions
Ion Trap	>40,000
Collision Cell (QTOF, QQQ, HCD)	>14,000

Number of Precursor Ions over Years

ReSpect for Phytochemicals

RIKEN MSn spectral database for phytochemicals (ReSpect) is a collection of literature and in-house MSn spectra data for research on plant metabolomics. Now, a total of **9017** spectra records are available.

Latest News

2013.11.11 5 spectra was added (data update) ► [Change Log](#)

Publication

'RIKEN tandem mass spectral database (ReSpect) for phytochemicals: A plant-specific MS/MS-based data resource and database' (Phytochemistry)



MassBank High Quality Mass Spectral Database

Database Service

- Statistics
- Publications
- Download
- Manuals
- About MassBank
- Contact
- Consortium Members
- Site Map
- User Restrictions
- Copyright
- Japanese Site

Consortium Members

Last updated Sep 27, 2013

[EU](#)
 NORMAN MassBank
 Leibniz Inst Plant Biochem
 Chem Biol, Max Planck Inst
 Hamamatsu Cant Environ Res (JPR)
[UK](#)
 ICB, Risc U
 PSC, RIKEN
 Nihon Univ K.A.
 Great Sch Pharm Sci, Kyoto U & Res Inst Prod Dev
 College Life Health, BK, Qufu U
 College Chem Eng, Tsinghua U
 College Chem Eng, Tsinghua U
 College Chem Eng, Tsinghua U
 Inst Natl Med, U Tokyo
 Fac Agr, Yaman U
 Great Sch Eng, U Tokyo
 Fac Life Sci Biotech, Fukuoka U
 Sch Med, UOEH
 Great Sch Intern Sci, NAOE
 Osaka Med Center (OMC)
 Great Sch Eng, Osaka U
 PSC, research group
 JRC, UK
 RIKEN, Science Inc.
[USA](#)
 U Colorado
 Washington State U
 Boise State U
[Brazil](#)
 Oswaldo Cruz Foundation, State Wisc. Servis
[China](#)
 Inst Materia Med, CAMS & PUMC

eawag
 normal
 NAIST
 RIKEN

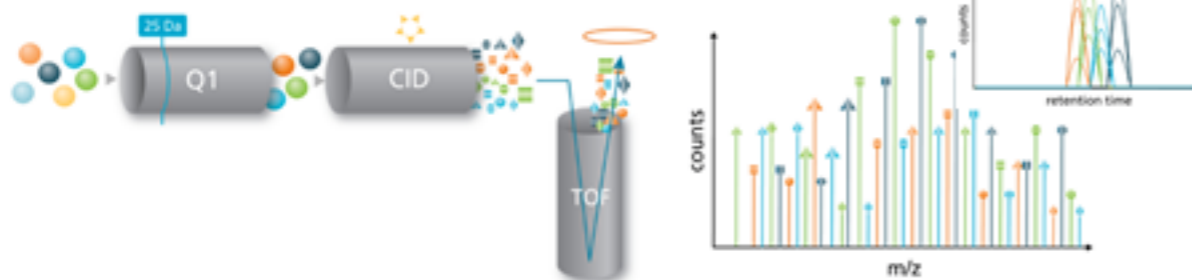
Very hard to tell how many unique compounds Mass Bank has

Methods to matching MS-DIAL

Targeted Quantitative Workflows

Top Two Techniques

SWATH™ Acquisition



MRM Workflow



3 © 2014 AB SCIEX

AB SCIEX

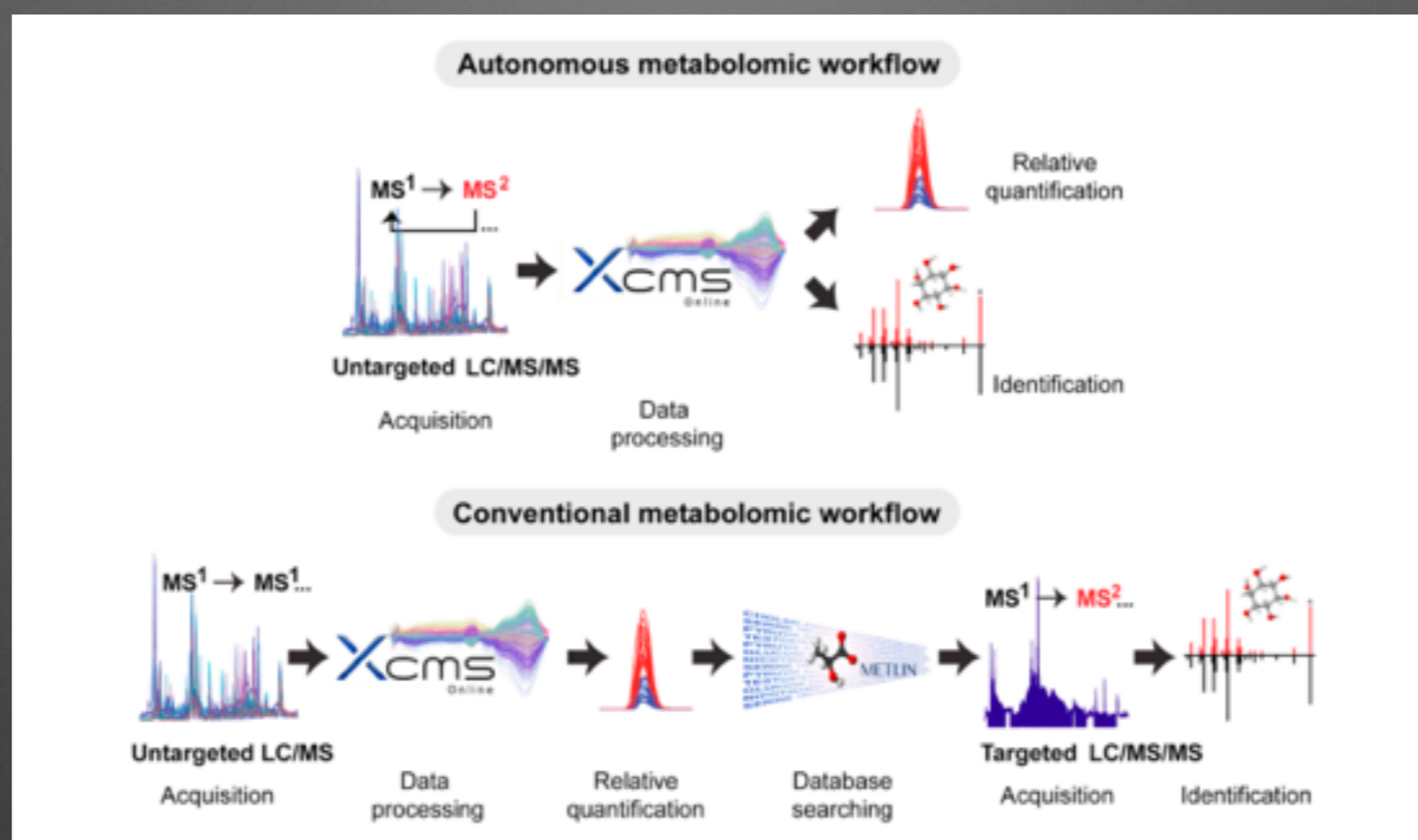
Powerful Tools for Omics Research

MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis

Hiroshi Tsugawa^{1,2}, Tomas Cajka³, Tobias Kind³,
Yan Ma³, Brendan Higgins⁴, Kazutaka Ikeda^{5,6},
Mitsuhiro Kanazawa⁷, Jean VanderGheynst⁴,
Oliver Fiehn^{3,8} & Masanori Arita^{1,9}

Data-independent acquisition (DIA) in liquid chromatography (LC) coupled to tandem mass spectrometry (MS/MS) provides comprehensive untargeted acquisition of molecular data. We provide an open-source software pipeline, which we call MS-DIAL, for DIA-based identification and quantification of small molecules by mass spectral deconvolution. For a reversed-phase LC-MS/MS analysis of nine algal strains, MS-DIAL using an enriched LipidBlast library identified 1,023 lipid compounds, highlighting the chemotaxonomic relationships between the algal strains.

Methods to matching autonomous



Autonomous Metabolomics for Rapid Metabolite Identification in Global Profiling

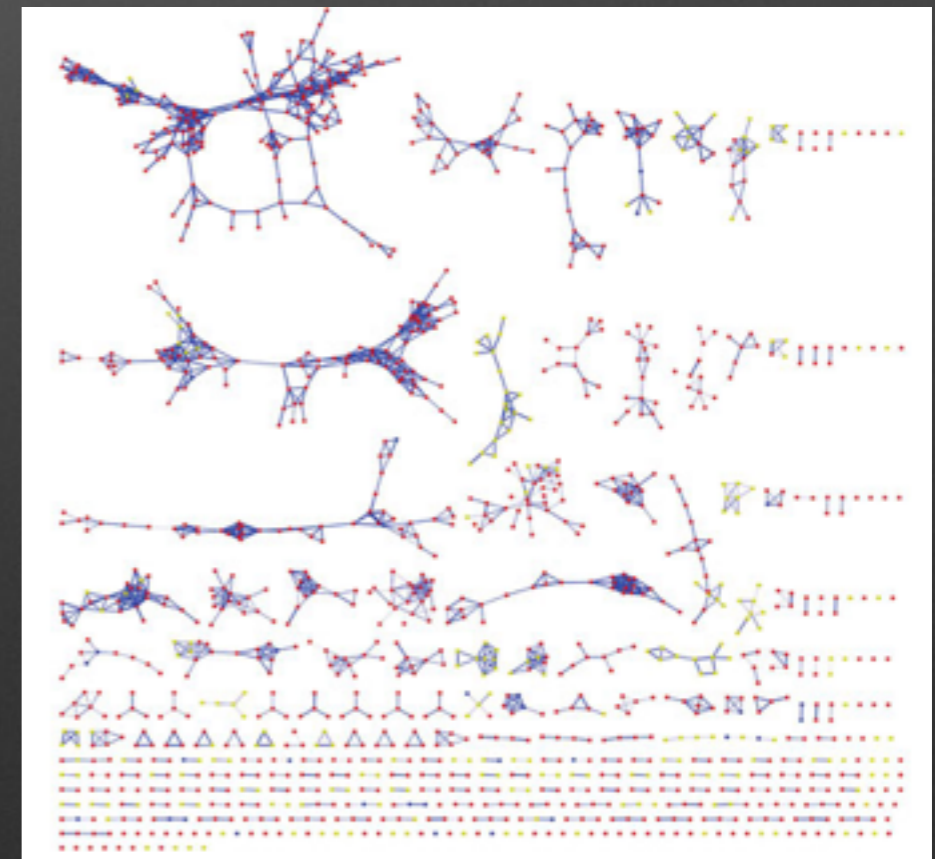
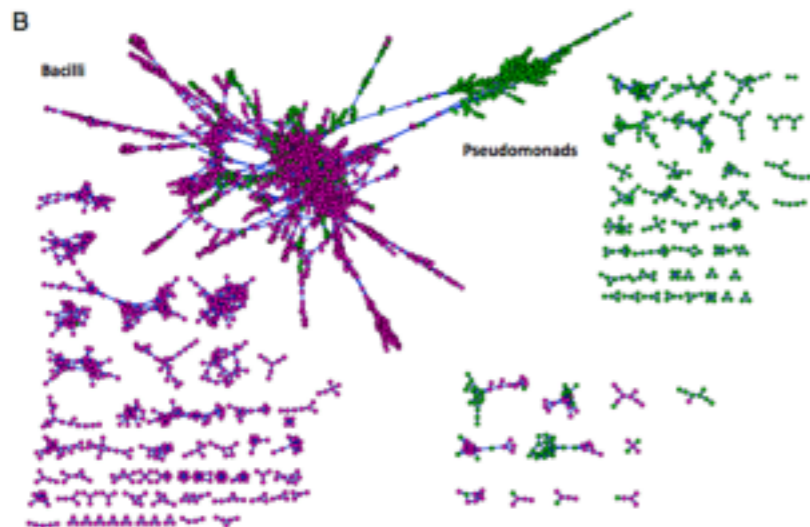
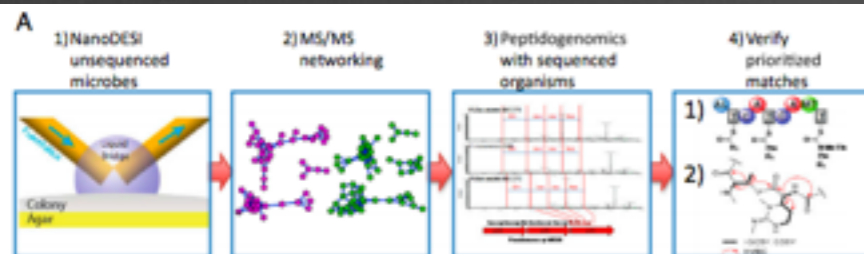
H. Paul Benton,[†] Julijana Ivanisevic,[†] Nathaniel G. Mahieu,[‡] Michael E. Kurczy,[†] Caroline H. Johnson,[†] Lauren Franco,[§] Duane Rinehart,[†] Elizabeth Valentine,[#] Harsha Gowda,^{†,¶} Baljit K. Ubhi,[∇] Ralf Tautenhahn,^{†,||} Andrew Gieschen,[⊥] Matthew W. Fields,[§] Gary J. Patti,^{*,‡} and Gary Siuzdak*

Methods to matching Similarity matching

MS/MS networking guided analysis of molecule and gene cluster families

Don Duy Nguyen^{a,1}, Cheng-Hsuan Wu^{a,1}, Wilna J. Moree^{b,1}, Anne Lamsa^c, Marnix H. Medema^d, Xiling Zhao^a,
Ronnie G. Gavilan^{e,f}, Marystella Aparicio^e, Librada Atencio^e, Chanaye Jackson^e, Javier Ballesteros^e, Joel Sanchez^e,
Jeramie D. Watrous^a, Vanessa V. Phelan^b, Corine van de Wiel^{a,b}, Roland D. Kersten^g, Samina Mehnaz^h, René De Motⁱ,
Elizabeth A. Shank^j, Pep Charusanti^k, Harish Nagarajan^k, Brendan M. Duggan^b, Bradley S. Moore^{b,g},
Nuno Bandeira^{b,l,m}, Bernhard Ø. Palsson^k, Kit Pogliano^c, Marcelino Gutiérrez^e, and Pieter C. Dorrestein^{a,b,g,n,2}

^aDepartment of Chemistry and Biochemistry, University of California at San Diego, La Jolla, CA 92093; ^bSkaggs School of Pharmacy and Pharmaceutical



De-novo methods

- General method is to take a structure of a molecule and break bonds (either randomly or rule based) to match to fragments masses.

MS Fragmentor



Mass-MetaSite



Mass Frontier/
MetWorks



MetFrag



MetFrag

In silico fragmentation for computer assisted identification of metabolite mass spectra

MetFrag | MzAnnotate Viewer | About / News

Database Settings

Database: KEGG PubChem ChemSpider Local SDF

Neutral exact mass: Search PPM:

Molecular formula:

Only biological compounds:

Limit # of structures:

Database ID's:

15 hits!

MetFrag Settings

Mode: [M+H] [M-H] [M]

Charge: pos. neg.

Mzabs (e.g. 0.01):

Mzppm (e.g. 10):

Parent ion: Neutral

Peaks:

```
119.051 467.616
123.044 370.662
147.044 6078.145
153.019 10000.0
179.036 141.192
189.058 176.358
273.076 10000.000
274.083 318.003
```

[View spectrum](#)

Download complete table: [Generate output files](#)

Score	# Explained Peaks	Trivial Name	Exact Mass	Structure	Database ID	Actions
1.0	5	<ul style="list-style-type: none">Naringenin chalcone2',4,4',6'-TetrahydroxychalconeIsoalpurpolChalconaringenin	C ₁₉ H ₁₂ O ₅ 272.0685		C06561	Fragments Download
0.989	5	<ul style="list-style-type: none">p-Coumaroyltriacetic acid lactone	C ₁₉ H ₁₂ O ₅ 272.0685		C12087	Fragments Download

MIDAS

A DATABASE-SEARCHING ALGORITHM FOR METABOLITE IDENTIFICATION IN METABOLOMICS

[HOME](#)

[SUBMIT A JOB](#)

[LOCATE JOB](#)

[DOWNLOADS](#)

NEW JOB SUBMISSION



Job Submitted

Job Id **8bdf8437-50fd-4ea3-816d-94e62a8e651f**

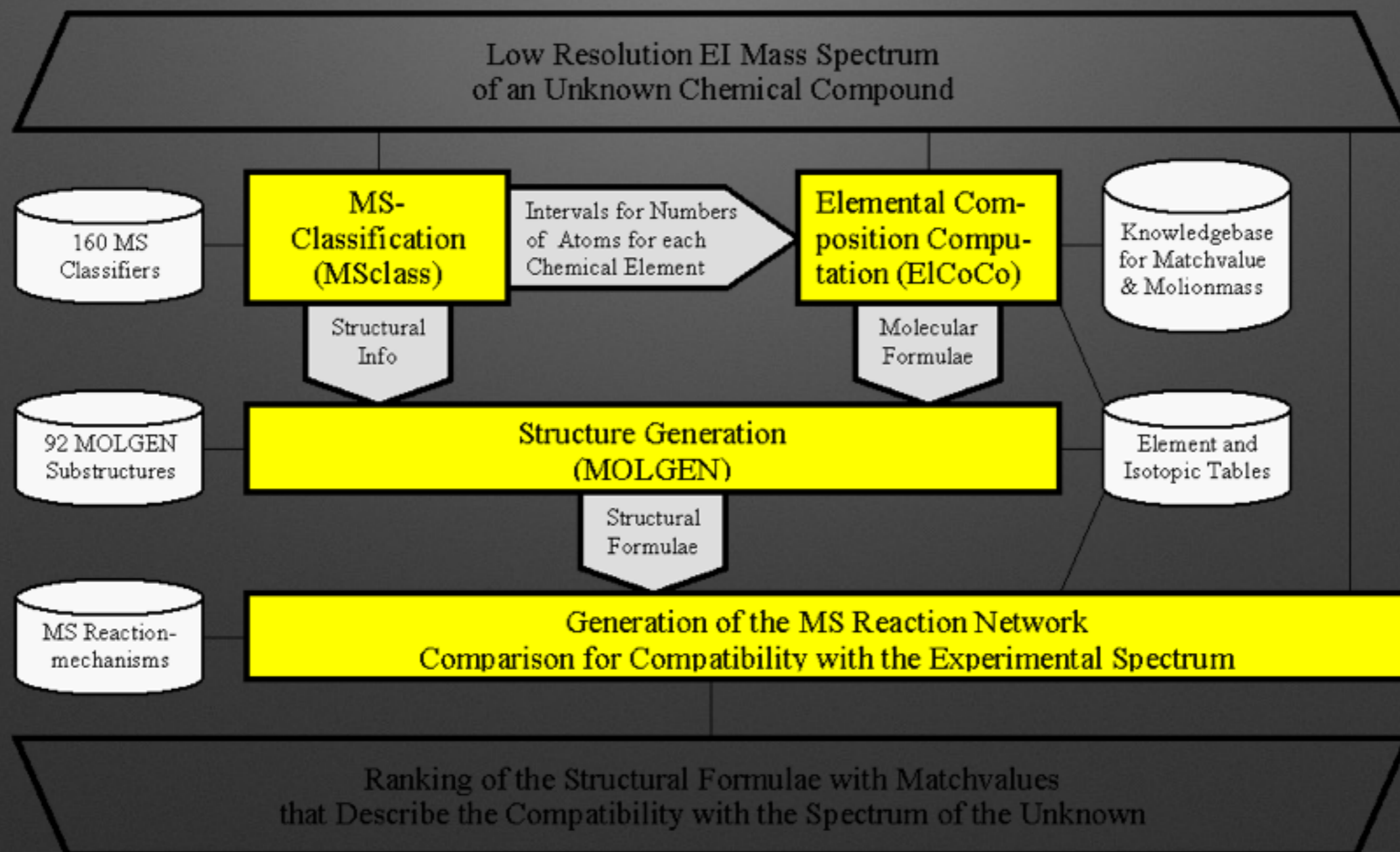
Email sent to hpbenton@gmail.com with URL to track progress.

CURRENT STATUS

Job Submitted → Job In Queue → Job In Execution → Job Execution Complete

Copyright © ORNL 2014

MolGen - EI spectra



Personalised Databases

- Crowd sourcing MS/MS data
 - mzCloud - run by thermo collection of user based fragmentation data
 - Metlin - Using XCMS Online each time someone searches MS/MS data the spectra is kept in their space. If approved by user as compound X, next time search on user's personal MS/MS database.


Manually Curated Data

	Compounds 2 771	Trees 3 853	Spectra 178 228	Anotations 3 058 730	QM Models	more ...
---	---------------------------	-----------------------	---------------------------	--------------------------------	------------------	----------

Thank you! Questions?

Suggested reading material for more !!

Scheubert et al. *Journal of Cheminformatics* 2013, 5:12
<http://www.jcheminf.com/content/5/1/12>



Journal of
Cheminformatics

REVIEW **Open Access**

Computational mass spectrometry for small molecules

Kerstin Scheubert^{1*}, Franziska Hufsky^{1,2} and Sebastian Böcker¹