

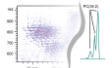
Metabolite identification

Michael Witting

Research Unit Analytical BioGeoChemistry, Helmholtz
Zentrum München

&

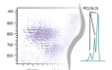
Chair of Analytical Foodchemistry, TU München



Outline

A path to metabolite identification

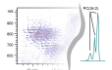
- Introduction
 - Analytical Technologies
 - Terminology in metabolite identification
- Metabolite identification
 - GC-MS MetID
 - EI spectral deconvolution
 - GC-APCI and exact mass
 - LC-MS MetID
 - Metabolite / Ion annotation
 - Formula calculation
 - Utilization of isotopes for formula calculation
 - MS/MS and RT
- New approaches for MetID
 - Ion mobility
 - *in silico* approaches
- *de novo* identification



Outline

A path to metabolite identification

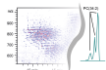
- Introduction
 - Analytical Technologies
 - Terminology in metabolite identification
- Metabolite identification
 - GC-MS MetID
 - EI spectral deconvolution
 - GC-APCI and exact mass
 - LC-MS MetID
 - Metabolite / Ion annotation
 - Formula calculation
 - Utilization of isotopes for formula calculation
 - MS/MS and RT
- New approaches for MetID
 - Ion mobility
 - *in silico* approaches
- *de novo* identification



Outline

A path to metabolite identification

- Introduction
 - Analytical Technologies
 - Terminology in metabolite identification
- Metabolite identification
 - GC-MS MetID
 - EI spectral deconvolution
 - GC-APCI and exact mass
 - LC-MS MetID
 - Metabolite / Ion annotation
 - Formula calculation
 - Utilization of isotopes for formula calculation
 - MS/MS and RT
- New approaches for MetID
 - Ion mobility
 - *in silico* approaches
- *de novo* identification



Introduction

Analytical Technologies

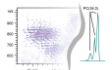
- Metabolomics utilizes different analytical chemistry methods for analysis of metabolites
- These methods include
 - **Mass Spectrometry (MS)** without or with prior metabolite separation
 - **Gas Chromatography (GC)**
 - **Liquid Chromatography (LC)**
 - **Capillary Electrophoresis (CE)**
 - **Nuclear Magnetic Resonance (NMR)**
- Each method has its specific advantages and disadvantages
- **No method can cover all metabolites!**
- Certain overlap between the different methods exists



Outline

A path to metabolite identification

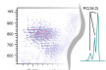
- Introduction
 - Analytical Technologies
 - Terminology in metabolite identification
- Metabolite identification
 - GC-MS MetID
 - EI spectral deconvolution
 - GC-APCI and exact mass
 - LC-MS MetID
 - Metabolite / Ion annotation
 - Formula calculation
 - Utilization of isotopes for formula calculation
 - MS/MS and RT
- New approaches for MetID
 - Ion mobility
 - *in silico* approaches
- *de novo* identification



Introduction

Terminology

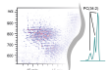
- Following cases have to be consider once it comes to metabolite identification
 - *Known metabolites* = known structure, present in matrix, detected by method of choice
 - *Known unknown metabolites* = know structure, shown to be not present in in matrix so far, detected by method of choice, e.g. comparison of MS/MS spectra
 - *Unknown metabolites* = *de novo* identification of truly novel compounds



Introduction

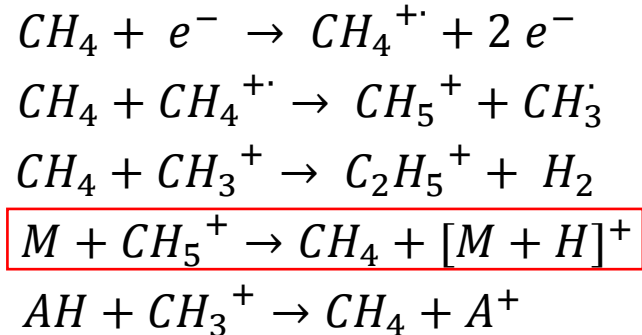
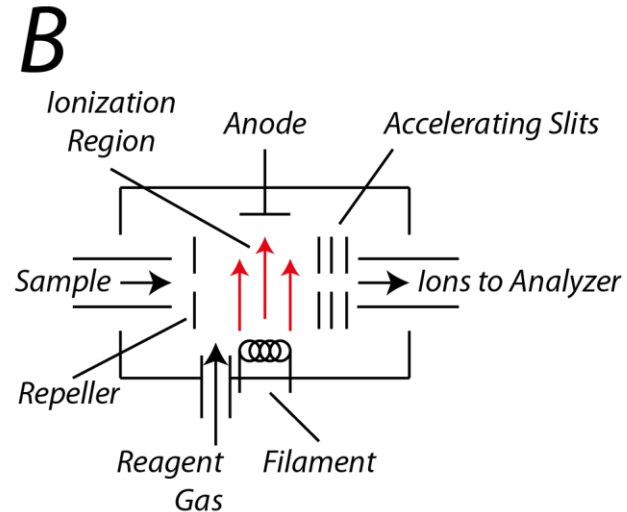
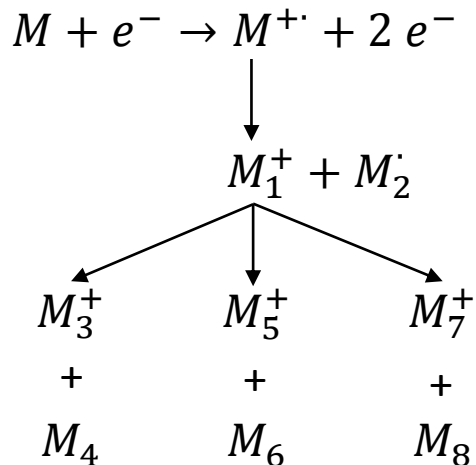
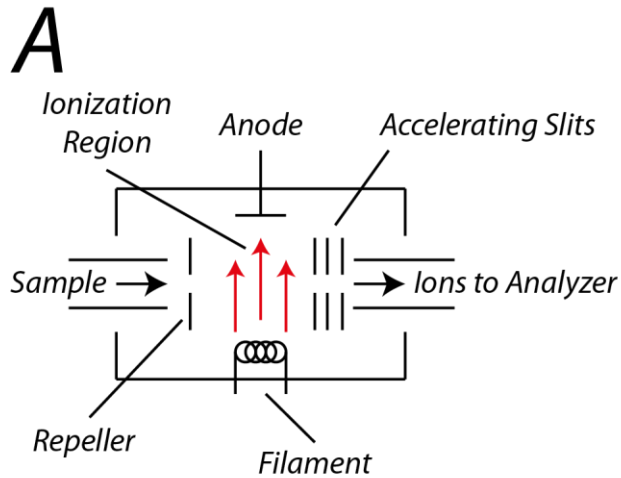
Terminology

- Not every single metabolite gives a single peak
- GC-MS
 - Different derivatization products
 - In source fragments (EI is a hard ionization technique)
- Although ESI is a soft ionization technique many signals are generated in LC-MS, SFC-MS and CE-MS for a single metabolite
 - Isotopes, adducts and multimers
 - In source fragments
 - Different charge states
- NMR
 - Different shifts and coupling
- It is important to differentiate between
 - Annotation = potential metabolites, e.g. based on comparison with exact masses
 - Identification = proven metabolites



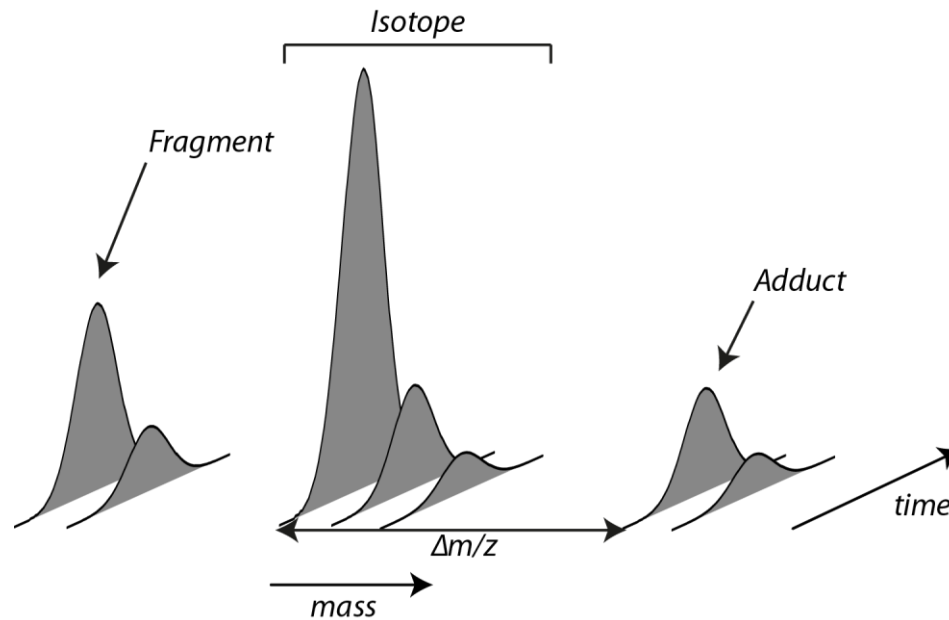
Introduction

Terminology – GC-MS multiple peaks



Introduction

Terminology – LC-MS multiple peaks

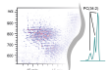


Introduction

Terminology – LC-MS multiple peaks

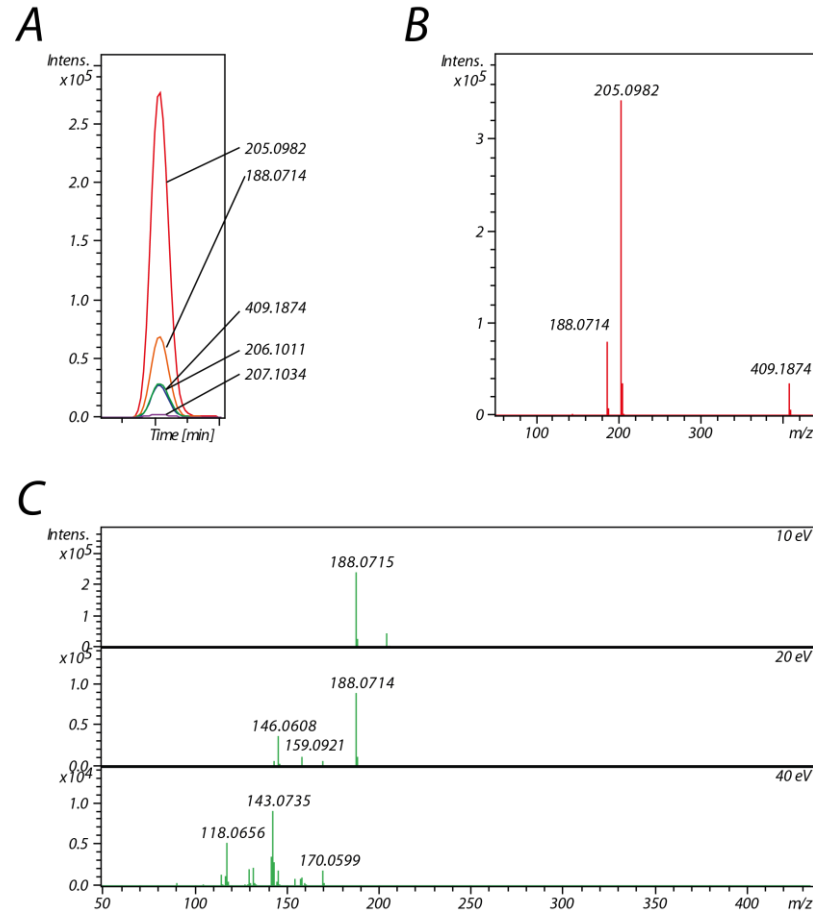
M+3H	M/3 + 1.007276	3+	0.33	1.007276
M+2H+Na	M/3 + 8.334590	3+	0.33	8.334590
M+H+2Na	M/3 + 15.7661904	3+	0.33	15.766190
M+3Na	M/3 + 22.989218	3+	0.33	22.989218
M+2H	M/2 + 1.007276	2+	0.50	1.007276
M+H+NH4	M/2 + 9.520550	2+	0.50	9.520550
M+H+Na	M/2 + 11.998247	2+	0.50	11.998247
M+H+K	M/2 + 19.985217	2+	0.50	19.985217
M+ACN+2H	M/2 + 21.520550	2+	0.50	21.520550
M+2Na	M/2 + 22.989218	2+	0.50	22.989218
M+2ACN+2H	M/2 + 42.033823	2+	0.50	42.033823
M+3ACN+2H	M/2 + 62.547097	2+	0.50	62.547097
M+H	M + 1.007276	1+	1.00	1.007276
M+NH4	M + 18.033823	1+	1.00	18.033823
M+Na	M + 22.989218	1+	1.00	22.989218
M+CH3OH+H	M + 33.033489	1+	1.00	33.033489
M+K	M + 38.963158	1+	1.00	38.963158
M+ACN+H	M + 42.033823	1+	1.00	42.033823
M+2Na-H	M + 44.971160	1+	1.00	44.971160
M-IsoProp+H	M + 61.06534	1+	1.00	61.065340
M+ACN+Na	M + 64.015765	1+	1.00	64.015765
M+2K-H	M + 76.919040	1+	1.00	76.919040
M+DMSO+H	M + 79.02122	1+	1.00	79.021220
M+2ACN+H	M + 83.060370	1+	1.00	83.060370
M-IsoProp+Na+H	M + 84.05511	1+	1.00	84.055110
2M+H	2M + 1.007276	1+	2.00	1.007276
2M+NH4	2M + 18.033823	1+	2.00	18.033823
2M+Na	2M + 22.989218	1+	2.00	22.989218
2M+K	2M + 38.963158	1+	2.00	38.963158
2M+ACN+H	2M + 42.033823	1+	2.00	42.033823
2M+ACN+Na	2M + 64.015765	1+	2.00	64.015765

M-3H	M/3 - 1.007276	3-	0.33	-1.007276
M-2H	M/2 - 1.007276	2-	0.50	-1.007276
M-H2O-H	M - 19.01839	1-	1.00	-19.01839
M-H	M - 1.007276	1-	1.00	-1.007276
M+Na-2H	M + 20.974666	1-	1.00	20.974666
M+Cl	M + 34.969402	1-	1.00	34.969402
M+K-2H	M + 36.948606	1-	1.00	36.948606
M+FA-H	M + 44.998201	1-	1.00	44.998201
M+Hac-H	M + 59.013851	1-	1.00	59.013851
M+Br	M + 78.918885	1-	1.00	78.918885
M+TFA-H	M + 112.985586	1-	1.00	112.985586
2M-H	2M - 1.007276	1-	2.00	-1.007276
2M+FA-H	2M + 44.998201	1-	2.00	44.998201
2M+Hac-H	2M + 59.013851	1-	2.00	59.013851
3M-H	3M - 1.007276	1-	3.00	1.007276
M-3H	M/3 - 1.007276	3-	0.33	-1.007276
M-2H	M/2 - 1.007276	2-	0.50	-1.007276
M-H2O-H	M - 19.01839	1-	1.00	-19.01839
M-H	M - 1.007276	1-	1.00	-1.007276
M+Na-2H	M + 20.974666	1-	1.00	20.974666
M+Cl	M + 34.969402	1-	1.00	34.969402
M+K-2H	M + 36.948606	1-	1.00	36.948606
M+FA-H	M + 44.998201	1-	1.00	44.998201
M+Hac-H	M + 59.013851	1-	1.00	59.013851
M+Br	M + 78.918885	1-	1.00	78.918885
M+TFA-H	M + 112.985586	1-	1.00	112.985586
2M-H	2M - 1.007276	1-	2.00	-1.007276
2M+FA-H	2M + 44.998201	1-	2.00	44.998201
2M+Hac-H	2M + 59.013851	1-	2.00	59.013851



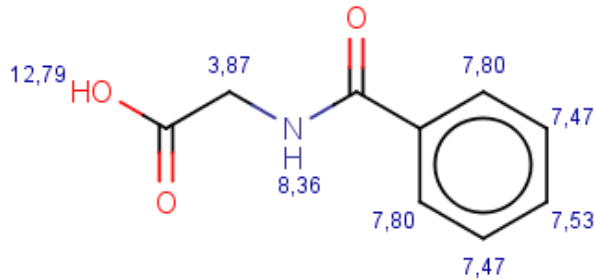
Introduction

Terminology – LC-MS multiple peaks

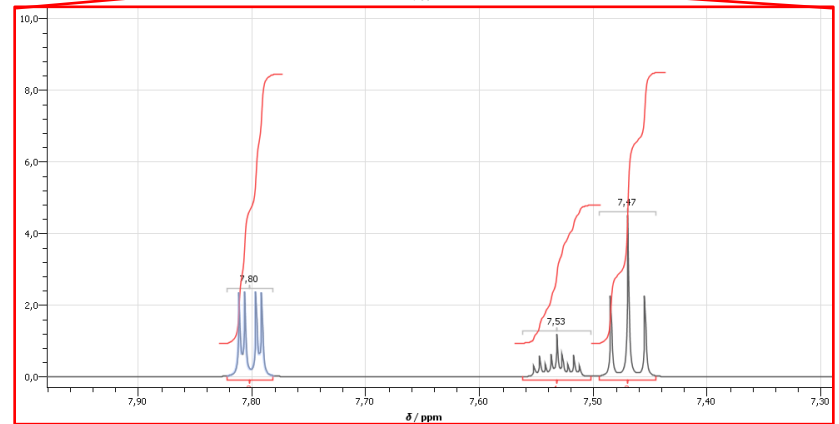
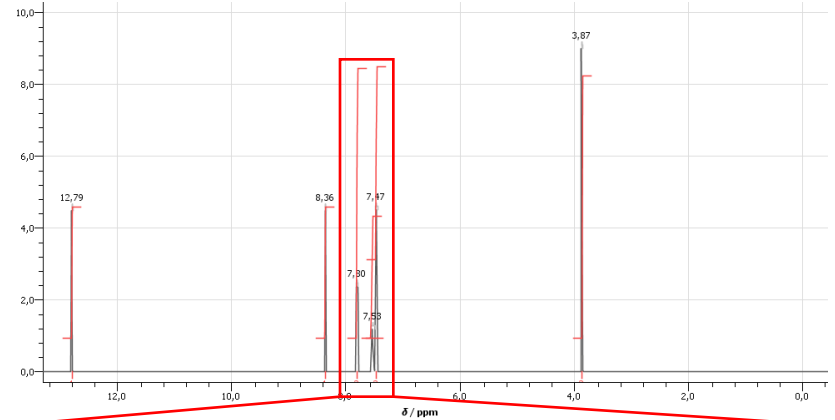


Introduction

Terminology – NMR multiple peaks



Chemical shift	Net intensity	Multiplet information
3.87 ppm	2	s
7.47 ppm	2	t
7.53 ppm	1	n
7.80 ppm	2	q
8.36 ppm	1	s
12.79 ppm	1	s

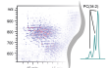


Introduction

MSI levels of identification

Level	Confidence of identity	Level of evidence
1	Confidently identified compounds	Comparison of two or more orthogonal properties with an authentic standard analyzed under identical analytical conditions
2	Putatively annotated compounds	Based upon physicochemical properties and/or spectral similarity with public/commercial spectral libraries, without reference to authentic chemical standards
3	Putatively annotated compound classes	Based upon characteristic physicochemical properties of a chemical class of compounds or by spectra similarity to known compounds of chemical class
4	Unknown compounds	Although unidentified and unclassified, these metabolites can still be differentiated and quantified based upon spectra data

Sumner et al., *Metabolomics*. 2007 September ; 3(3): 211–221



Introduction

MSI levels of identification

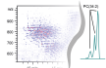
- Sumner et al. suggested a scoring system for metabolite identification
- Based on different scores for different analytical techniques you „quantitatively“ report your confidence
- The sum of individual scores would give the final score
- If data is compared to an authentic standard the score is multiplied with 2, if compared to a public database the score is multiplied with 1.5
- Example 1: UPLC-MS with accurate mass and comparison with an authentic standard
- $(1.5 + 1.0) \times 2 = 5.0$
- Example 2: GC-MS with high resolution RI, nominal mass and match to NIST library
- $(1.5 + 0.5 + 1.5) \times 1.5 = 5.25$

Table 1 Suggested quantitative scoring system for summarizing a metabolite identification confidence

IR absorbance spectrum	0.5
UV absorbance spectrum	0.5
Retention time ($\pm 2.5\%$)	1.0
High resolution retention time ($\pm 0.5\%$, $W_{1/2} < 10$ s)	1.5
High resolution retention index ($\pm 0.5\%$, $RI \pm 25$, $W_{1/2} < 10$ s)	2.0
Nominal mass of parent ion	0.5
Accurate mass of parent ion (< 5 ppm)	1.0
Molecular formula based upon accurate m/z and isotope pattern	1.0
Confident EI Spectral match to commercial library	1.0
Tandem mass spectrum	1.5
Accurate mass tandem mass spectrum	2.0
^1H 1D NMR	2.0
^1H 2D NMR	3.0
$^1\text{H} \times ^{13}\text{C}$ 2D NMR	4.0

IR infrared spectroscopy, *UV* ultraviolet, *Rt* retention time, *RI* retention index, $W_{1/2}$ peak width at half-height, m/z mass-to-charge ratio, *EI* electron ionization, *1D* one-dimensional, *2D* two-dimensional, *NMR* nuclear magnetic resonance spectroscopy

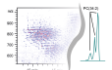
Sumner et al., *Metabolomics* (2014) 10:1047–1049



Outline

A path to metabolite identification

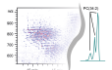
- Introduction
 - Analytical Technologies
 - Terminology in metabolite identification
- Metabolite identification
 - GC-MS MetID
 - EI spectral deconvolution
 - GC-APCI and exact mass
 - LC-MS MetID
 - Metabolite / Ion annotation
 - Formula calculation
 - Utilization of isotopes for formula calculation
 - MS/MS and RT
- New approaches for MetID
 - Ion mobility
 - *in silico* approaches
- *de novo* identification



Outline

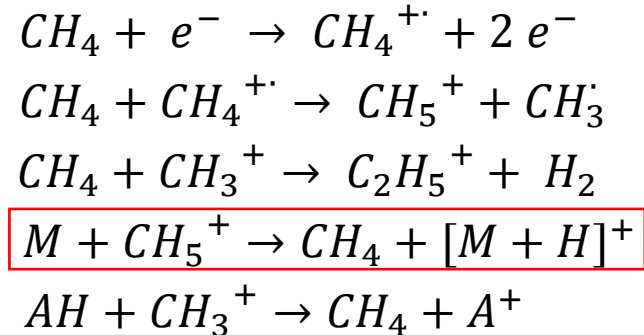
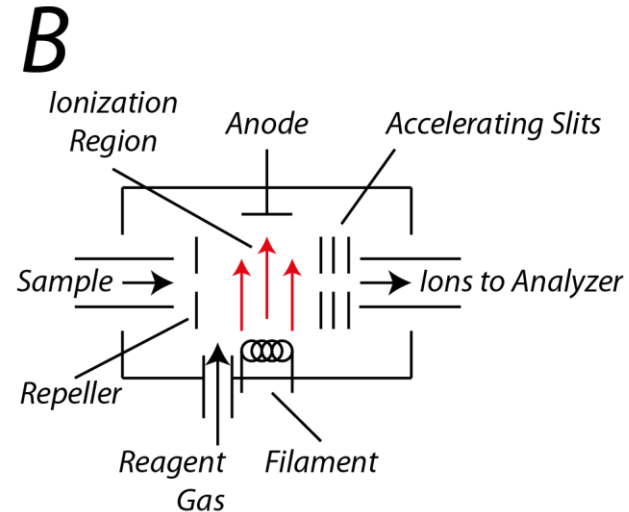
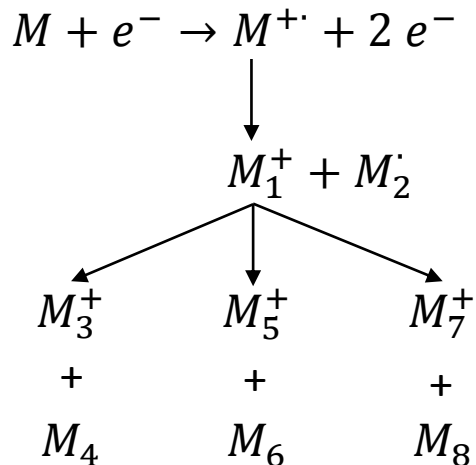
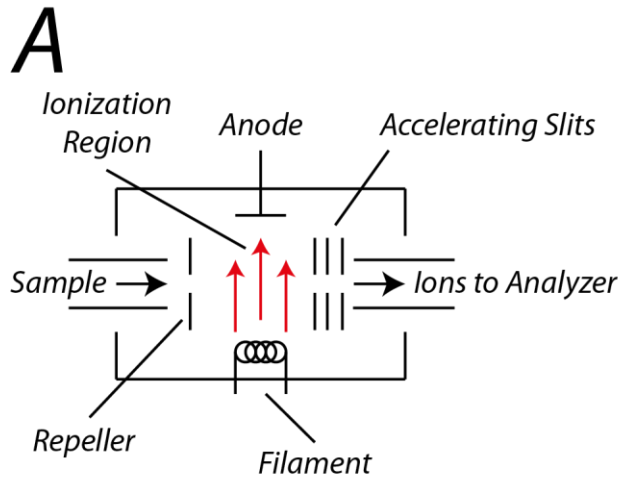
A path to metabolite identification

- Introduction
 - Analytical Technologies
 - Terminology in metabolite identification
- Metabolite identification
 - GC-MS MetID
 - EI spectral deconvolution
 - GC-APCI and exact mass
 - LC-MS MetID
 - Metabolite / Ion annotation
 - Formula calculation
 - Utilization of isotopes for formula calculation
 - MS/MS and RT
- New approaches for MetID
 - Ion mobility
 - *in silico* approaches
- *de novo* identification



Metabolite Identification

GC-MS MetID – EI spectral deconvolution



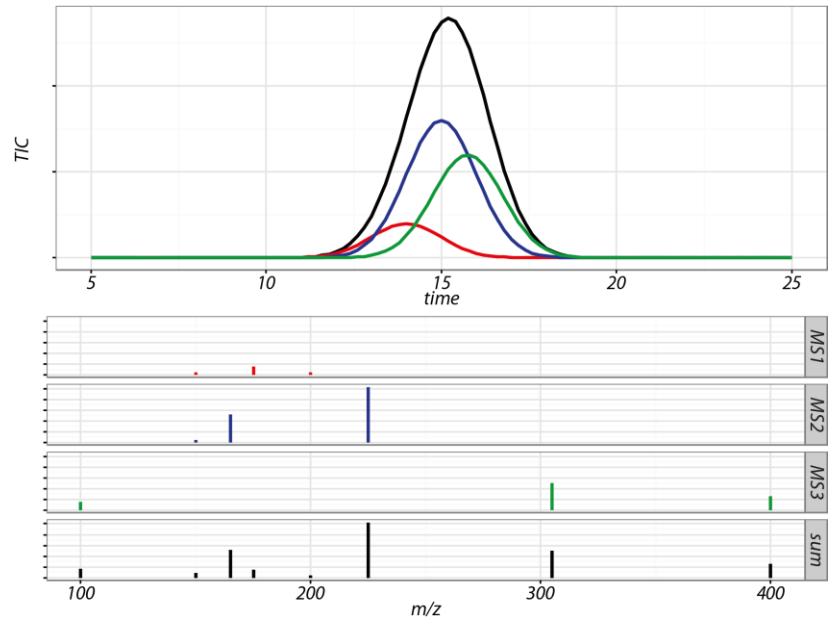
Metabolite Identification

GC-MS MetID – EI spectral deconvolution

- Fragmentation-rich ionization methods like EI lead to mixed mass spectra
- Mixed spectra have to be deconvoluted into individual compound spectra
- AMDIS is a software capable of deconvoluting GC-MS runs



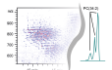
- Deconvoluted compounds and compound spectra can be searched in EI libraries, e.g. NIST, Golm Metabolome Database, Fiehn library, etc..



Outline

A path to metabolite identification

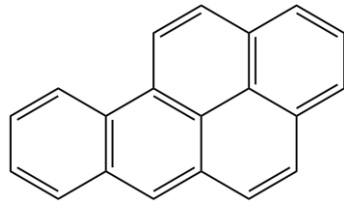
- Introduction
 - Analytical Technologies
 - Terminology in metabolite identification
- Metabolite identification
 - GC-MS MetID
 - EI spectral deconvolution
 - GC-APCI and exact mass
 - LC-MS MetID
 - Metabolite / Ion annotation
 - Formula calculation
 - Utilization of isotopes for formula calculation
 - MS/MS and RT
- New approaches for MetID
 - Ion mobility
 - In silico approaches



Metabolite Identification

GC-MS MetID – GC-APCI and exact mass

Analytics of LC
unamenable
compounds



Benzo(a)pyren

Libraries (NIST)

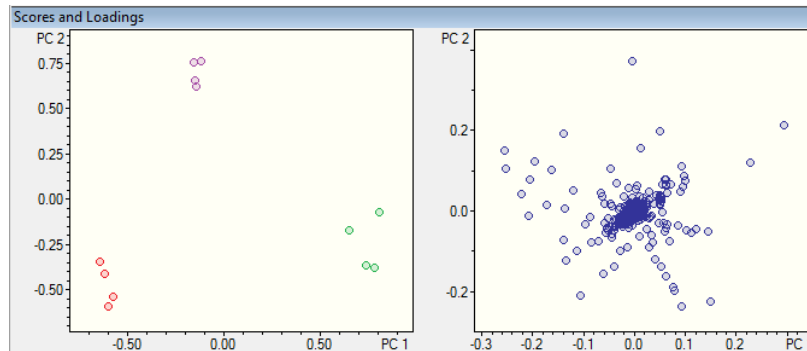
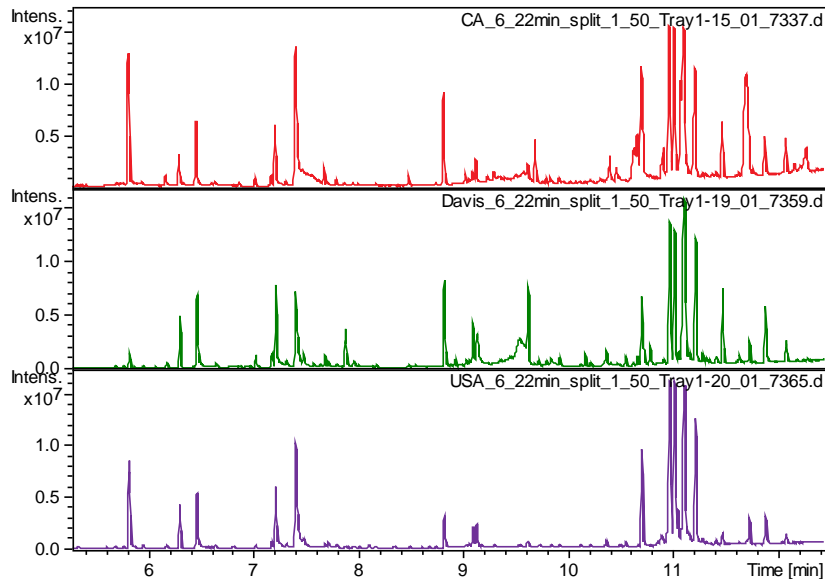
```
Estimated non-polar retention index (normal scale):  
Value: 1238.40  
Confidence Interval (Absolute): 41(0%) 176(0%) to  
  
Reference Index  
I. Value: 1252.40  
Column Type: Capillary  
Column Class: Standard non-polar  
Active Phase: TQ14.10M-05  
Column Length: 100 m  
Carrier Gas: He  
Column Diameter: 0.25 mm  
Phase Thickness: 0.25 µm  
Data Type: Science JE  
Program Type: Ramp  
Start T: 150 C  
End T: 250 C  
Ramp Rate: 5 K/min  
Start Time: 10 min  
End Time: 25 min  
Source: Gaudier, H.H.; Lopez, A.R.; Com, S.C.; Tomasco, C.J.;  
Fukushima, K.C.R.; Miguel, M.D.; Santos, A.V.; Amor, C.G.; Miguel,  
C.G., Physicochemical and structural evaluation of benzofluoranthene  
from volcanic Martiana Mt. (Papasovos), P. New. Chem. Soc., 18(1),  
2007, 184-188.
```

No hit in EI library search?



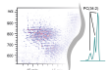
Metabolite Identification

GC-MS MetID – GC-APCI and exact mass



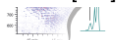
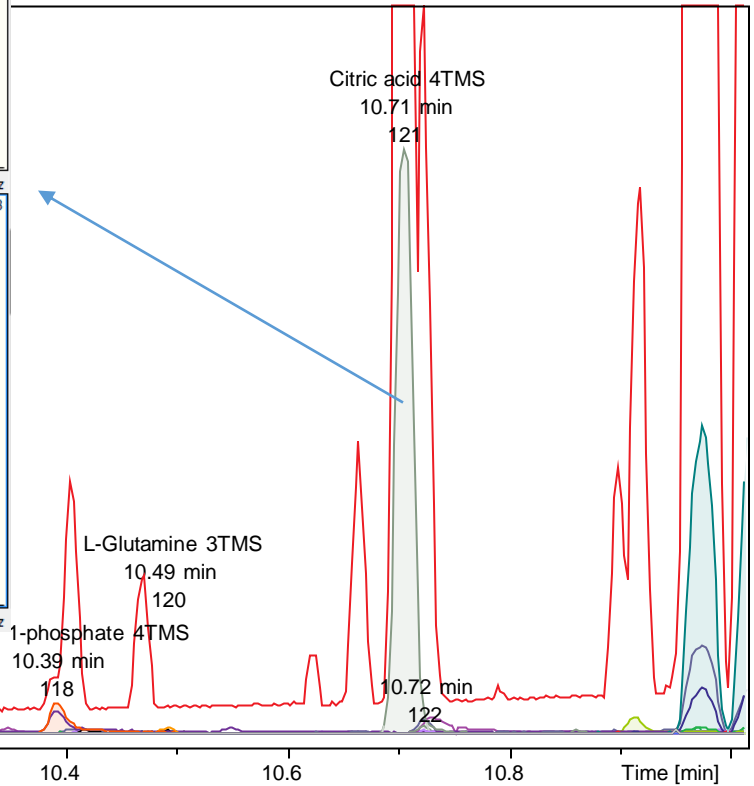
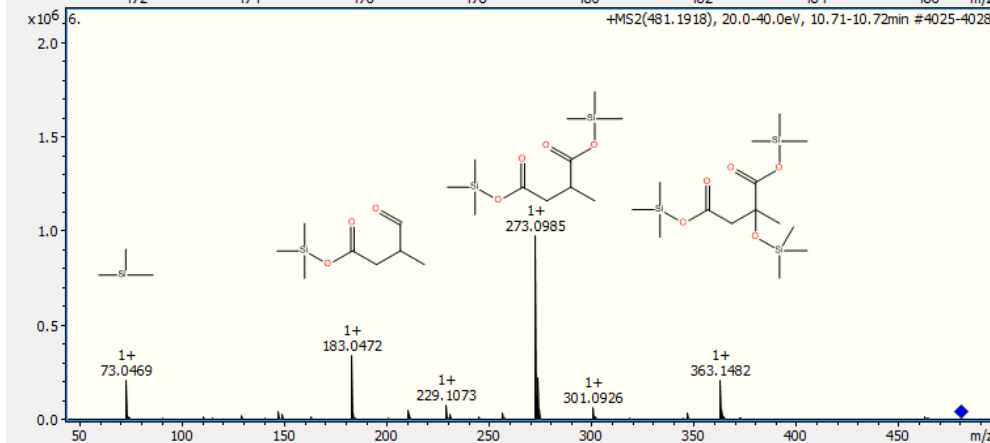
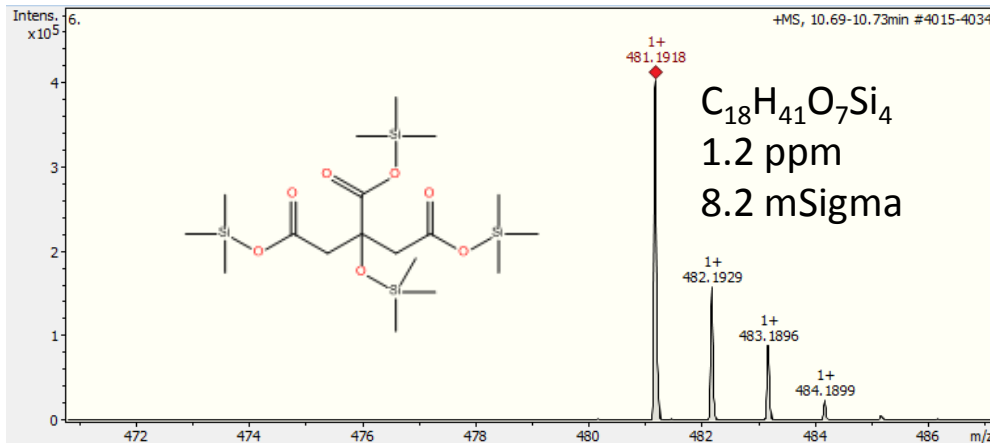
GC-APCI-TOF/MS coupling

- Soft ionization
- Molecular ion preserved
- High resolution data
- ID by SmartFormula



Metabolite Identification

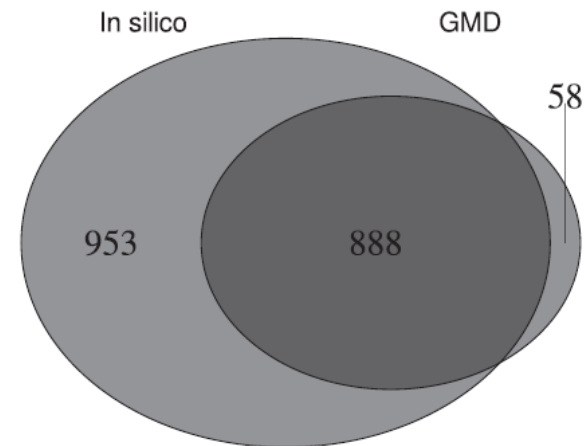
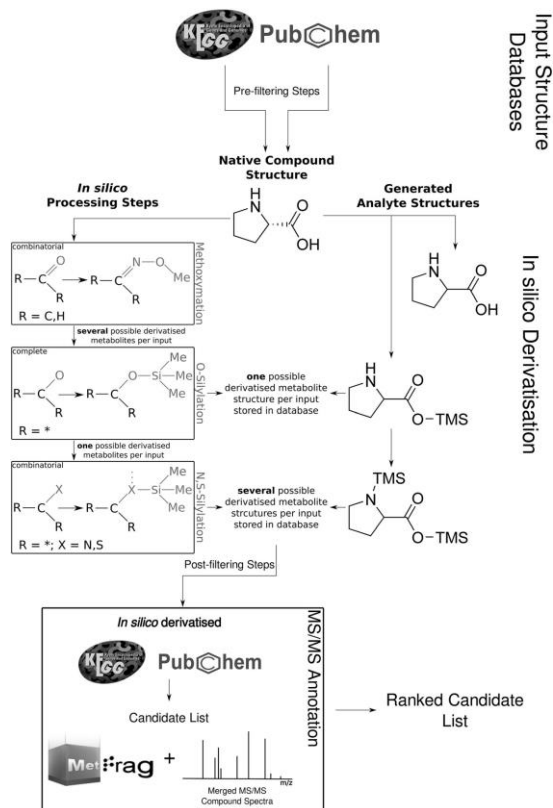
GC-MS MetID – GC-APCI and exact mass



Metabolite Identification

GC-MS MetID – GC-APCI and exact mass

- GC-MS spectral databases still contain only a limited number of APCI-QToF spectra
- Ruttkies et al. developed a workflow based on in silico derivatization of metabolites and in silico fragmentation to identify potential candidate structures

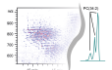


Ruttkies et al., Rapid Commun. Mass Spectrom. 2015, 29, 1521–1529

Outline

A path to metabolite identification

- Introduction
 - Analytical Technologies
 - Terminology in metabolite identification
- Metabolite identification
 - GC-MS MetID
 - EI spectral deconvolution
 - GC-APCI and exact mass
 - LC-MS MetID
 - Metabolite / Ion annotation
 - Formula calculation
 - Utilization of isotopes for formula calculation
 - MS/MS and RT
- New approaches for MetID
 - Ion mobility
 - *in silico* approaches
- *de novo* identification



Metabolite Identification

LC-MS MetID - *Metabolite / Ion annotation*

HelmholtzZentrum münchen
German Research Center for Environmental Health

mips
munich information center
for protein sequences

Home | Start a new run | Job status | Examples | Documentation

MassTRIX: Mass TRANslator into Pathways
This is MassTRIX reloaded, the 3rd version of MassTRIX.

The jobs on the old server remain still available at this [link](#).
Should you encounter any unexpected behaviour, please let us know!

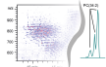
252.04568
449.3549

D-gluconate
D-Glucose-6P
D-Glucono-1,5-lactone-6P
6-Phospho-D-gluconate
D-Ribulose-5P
CO₂

1.1.1.49
3.1.1.31
1.1.1.44
5.1.3.1
5.3.6.1

Welcome to MassTRIX !

MassTRIX annotates metabolites in high precision mass spectrometry data.

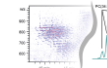


Metabolite Identification

LC-MS MetID - *Metabolite / Ion annotation*

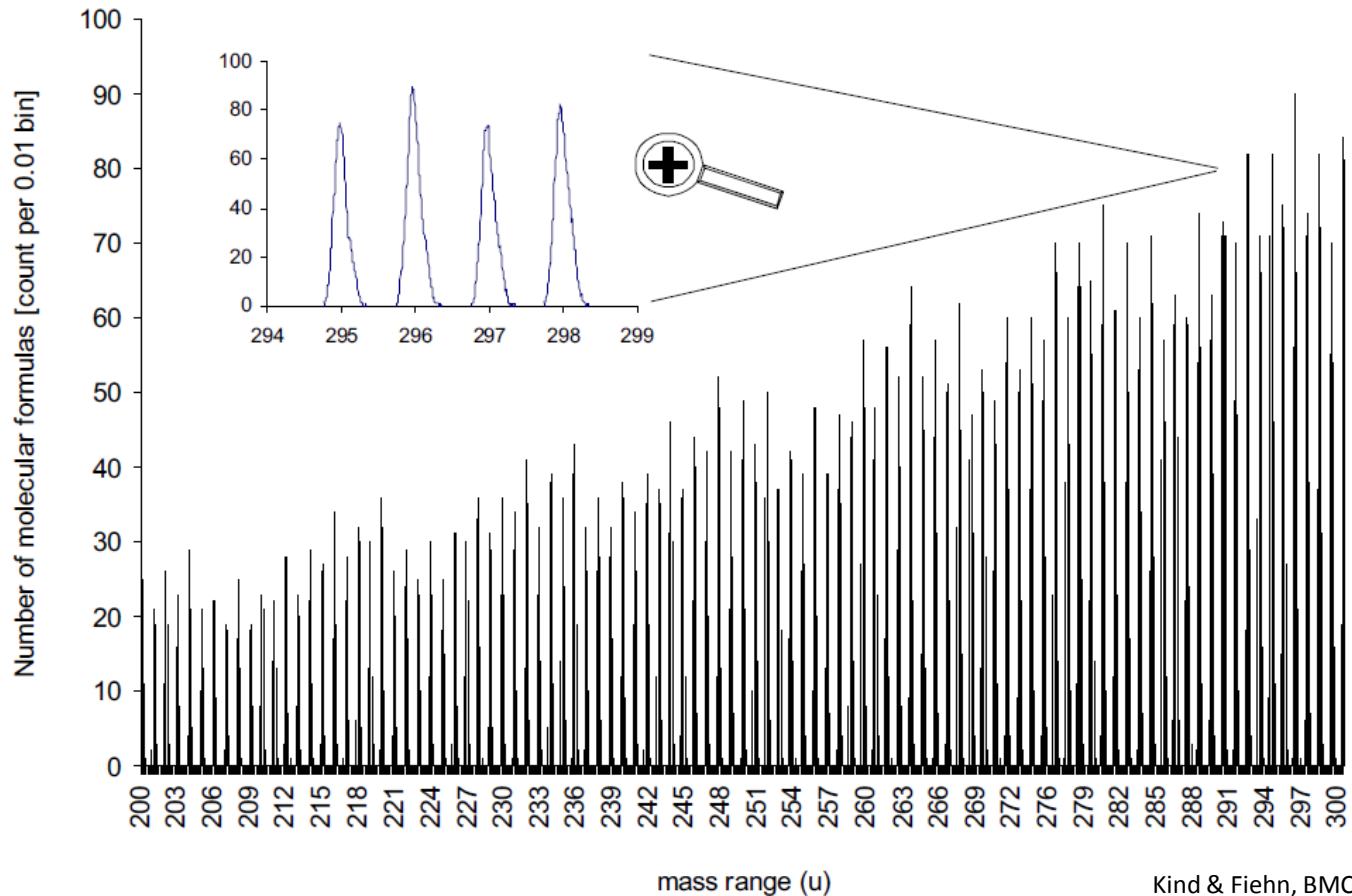
Table 1: Example of a molecular formula search for $C_{15}H_{12}O_7$ in different chemical databases. Search date: July 2007

Database name	Compounds found	Total database entries
Chemical Abstracts (CAS)	181	24,000,000
Beilstein Database (MDL)	166	8,000,000
Dictionary of Natural Products (DNP)	129	170,000
PubChem (NIH)	19	800,000
Available Chemicals Directory (MDL)	6	400,000
ChemIDplus (NIH)	6	370,000
KEGG (Kyoto University)	3	13,000
NIST05 (NIST mass spectral database)	2	163,000
MOLGEN molecular isomer generator (allowing 2 benzene groups; 1 ether group, 1 keto group; 5 hydroxy groups)	788,000	-



Metabolite Identification

LC-MS MetID - *Metabolite / Ion annotation*



Kind & Fiehn, BMC Bioinformatics 2006, 7:234

Metabolite Identification

LC-MS MetID - *Metabolite / Ion annotation*

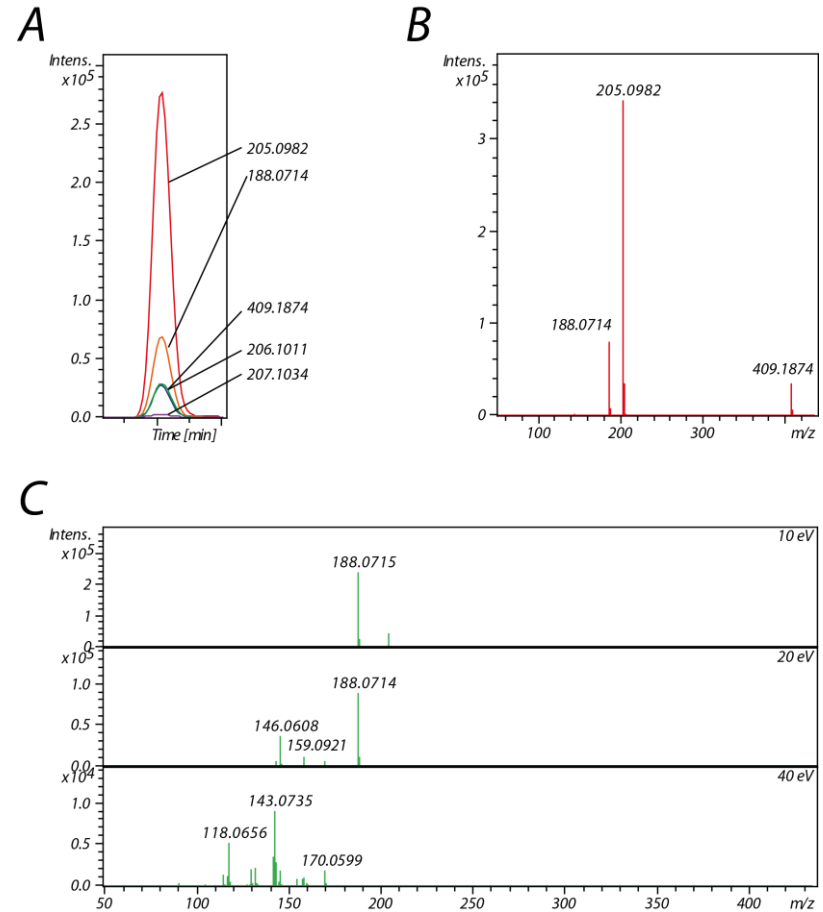
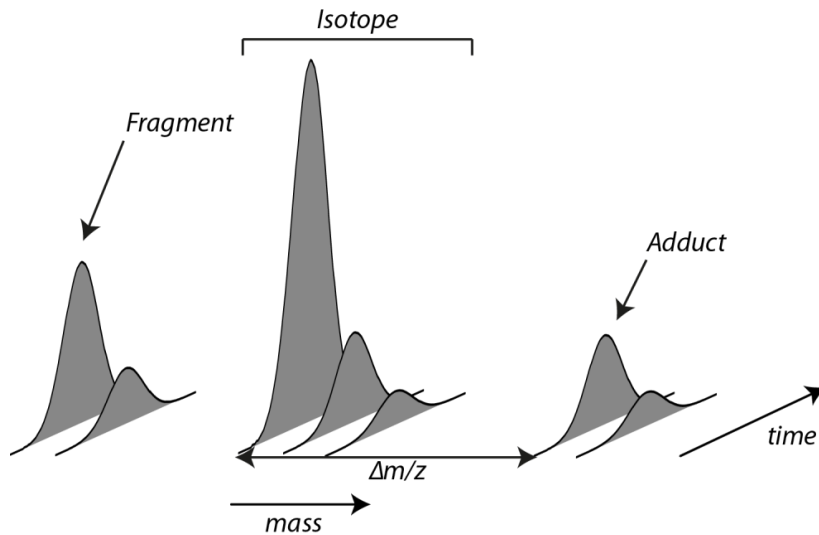
Table 3: Number of possible molecular formulas at different levels of mass accuracy and the impact of isotopic abundance accuracy. A mass spectrometer capable of 3 ppm but with 2% correct isotopic pattern outperforms even a (non-existing) mass spectrometer with 0.1 ppm mass accuracy! The results are computed for randomly selected targets, so single results vary but the trend remains. LEWIS and SENIOR check was applied. Candidates with unrelated high element counts were already excluded

molecular mass [Da]	without isotope abundance information					2% isotopic abundance accuracy	5% isotopic abundance accuracy
	10 ppm	5 ppm	3 ppm	1 ppm	0.1 ppm	3 ppm	5 ppm
150	2	1	1	1	1	1	1
200	3	2	2	1	1	1	1
300	24	11	7	2	1	1	6
400	78	37	23	7	1	2	13
500	266	115	64	21	2	3	33
600	505	257	155	50	5	4	36
700	1046	538	321	108	10	10	97
800	1964	973	599	200	20	13	111
900	3447	1712	1045	345	32	18	196

Kind & Fiehn, BMC Bioinformatics 2006, 7:234

Metabolite Identification

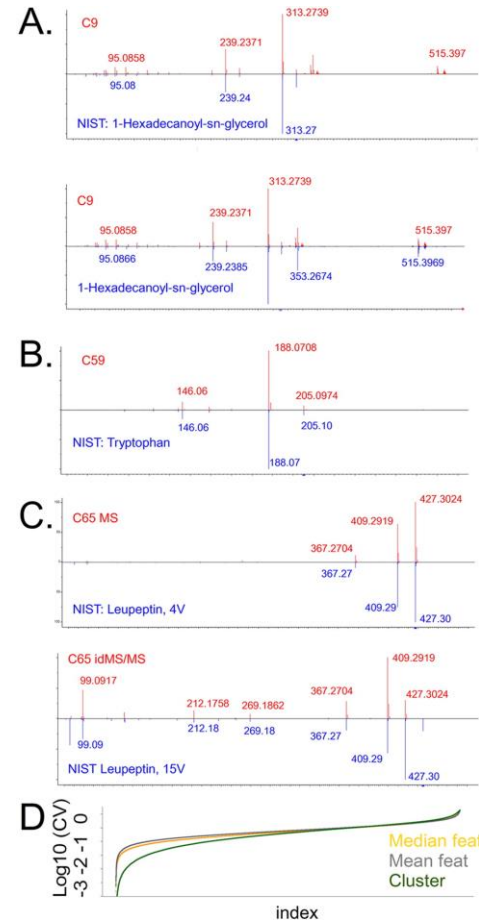
LC-MS MetID - *Metabolite / Ion annotation*



Metabolite Identification

LC-MS MetID - *Metabolite / Ion annotation*

- MS signals from the same metabolite co-elute and co-vary
- Correlation analysis can be used to identify in-source fragments, multimers
- The yielded mass spectra can be used for search in different spectral databases
- Based on the same approach, indiscriminant MS/MS can be used to generate fragmentation for every eluting compound

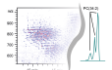


Broeckling et al., Anal. Chem., 2014, 86 (14), pp 6812–6817

Outline

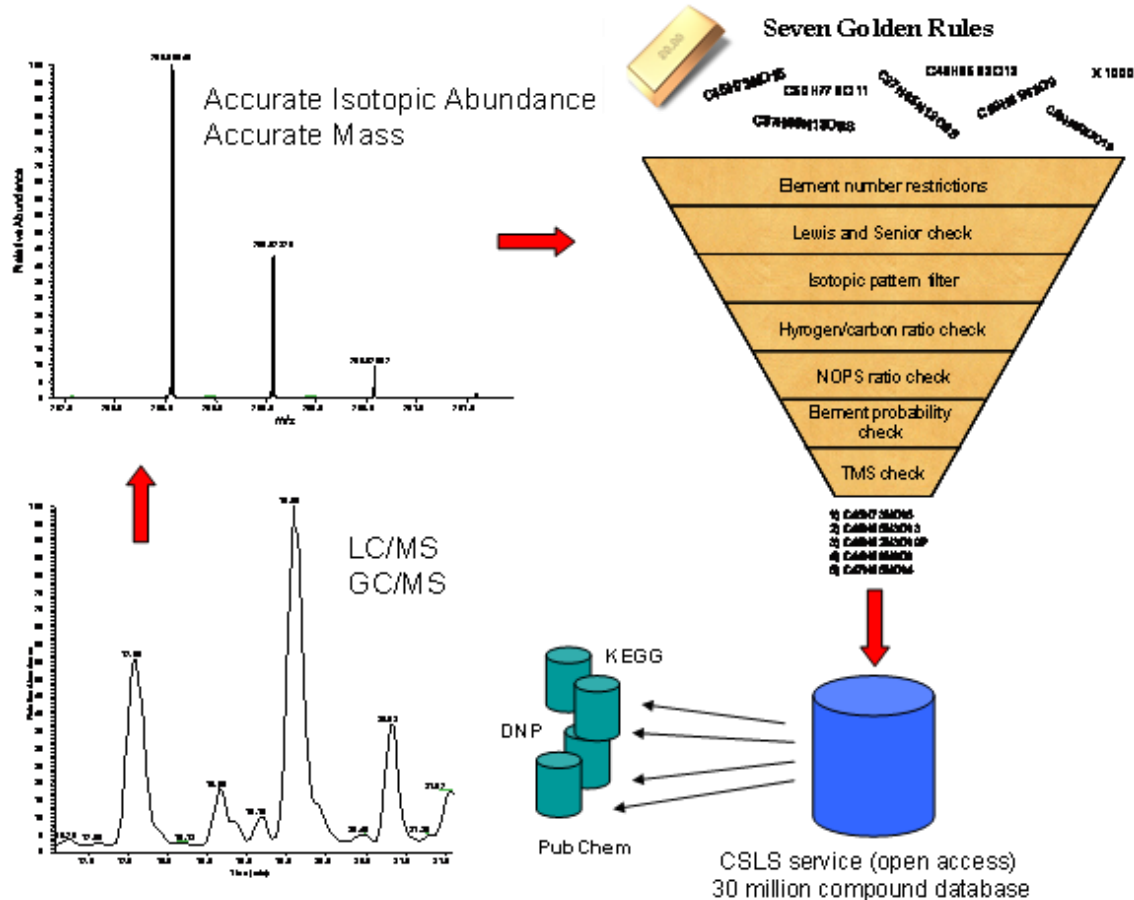
A path to metabolite identification

- Introduction
 - Analytical Technologies
 - Terminology in metabolite identification
- **Metabolite identification**
 - GC-MS MetID
 - EI spectral deconvolution
 - GC-APCI and exact mass
 - LC-MS MetID
 - Metabolite / Ion annotation
 - **Formula calculation**
 - Utilization of isotopes for formula calculation
 - MS/MS and RT
- New approaches for MetID
 - Ion mobility
 - *in silico* approaches
- *de novo* identification

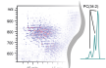


Metabolite Identification

LC-MS MetID – Formula calculation



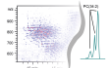
Kind & Fiehn, BMC Bioinformatics 2007, 8:105



Metabolite Identification

LC-MS MetID – Formula calculation

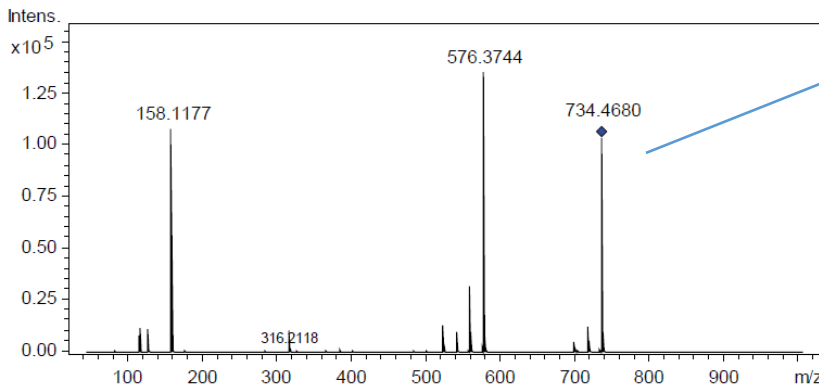
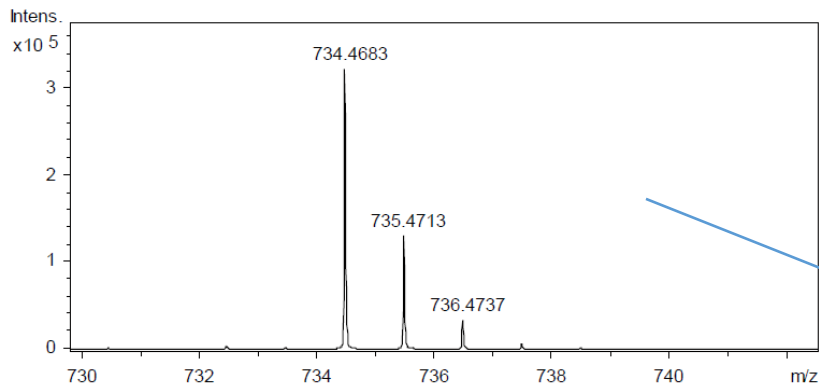
1. Apply heuristic restrictions for number of elements during formula generation
2. Perform LEWIS and SENIOR check
 - LEWIS rule: In its simplest form, the LEWIS rule demands that molecules consisting of main group elements, especially carbon, nitrogen and oxygen, share electrons in a way that all atoms have completely filled s, p-valence shells ('octet rule').
 - SENIOR rule: i) The sum of valences or the total number of atoms having odd valences is even. ii) The sum of valences is greater than or equal to twice the maximum valence. iii) The sum of valences is greater than or equal to twice the number of atoms minus 1.
3. Perform isotopic pattern filter
4. Perform H/C ratio check (hydrogen/carbon ratio)
5. Perform NOPS ratio check (N, O, P, S/C ratios)
6. Perform heuristic HNOPS probability check (H, N, O, P, S/C high probability ratios)
7. Perform –TMS check (for GC-MS if a silylation step is involved)



Metabolite Identification

LC-MS MetID – Formula calculation

- Formula calculation can be enhanced by using results from MS/MS
- Sum of fragment sum formula +/- neutral losses = parent sum formula



Parameters... Swap Spectra Create Compound Calculate Filter Simulate Export Help

+MS (52-28.56) #C(28)

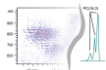
m/z	Intensity	FWHM	S/N
734.4683	322294.8	0.0249	11041.5
735.4713	128034.4	0.0247	4443.3
736.4737	32298.3	0.0233	1179.2
737.4760	5968.2	0.0233	127.1

+MS(734.4683, 0.4-6.0min) #C(56)

#2	Intensity	FWHM	S/N
117.1120	821.4	0.0363	94.8
117.0750	10422.3	0.0353	10421.1
116.0791	626.3	0.0333	603.5
158.1177	107951.2	0.0309	112905.0
158.1180	7996.2	0.0377	9460.8
160.1202	866.0	0.0304	801.4
176.1277	824.7	0.0375	715.7
194.1394	211.5	0.1122	272.8
246.2118	10094.9	0.0142	11492.4
316.2190	15124.6	0.0128	17875.4
316.2169	187.8	0.0121	211.2
316.2089	227.4	0.1114	287.6
347.2215	157.8	0.0128	168.9
365.2219	596.6	0.0124	679.2
366.2296	127.5	0.1124	167.4
365.2427	974.9	0.0138	1045.8

SumFormula	mp calk	enf[sum]	enf[par]	ndsigma	eConf	Com...	Com...	Com...
[C37H66NO]23	734.4685	0.2	0.3	0.2	even	16.0	0.5	0.4

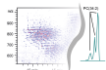
SumFormula	SumFormula Loss	m/z1	enf1	Delta SF	Inten...	m/z calk	enf...	mSz
[C37H66NO]12	H2O	18.0141	-0.8		12283	716.4...	1.0	0
[C37H66NO]13	H4O2	36.0218	-0.7		4569	688.4...	0.9	10
[C37H66NO]12	C8H14O3	128.1211	-0.2		12049	576.3...	-0.1	20
[C37H66NO]9	C8H18O4	176.1311	-0.1		31494	528.3...	0.3	12
[C37H66NO]8	C8H18O5	194.1311	-0.2		9751	548.3...	0.4	28
[C37H66NO]7	C8H20O6	212.1311	-0.2		10759	522.3...	0.2	14
[C37H66NO]5	C21H32O8	418.2111	0.1		10070	316.2...	0.1	15
[C8H16NO]2	C29H52O11	576.3...	0.3		107931	158.1...	-0.1	5



Outline

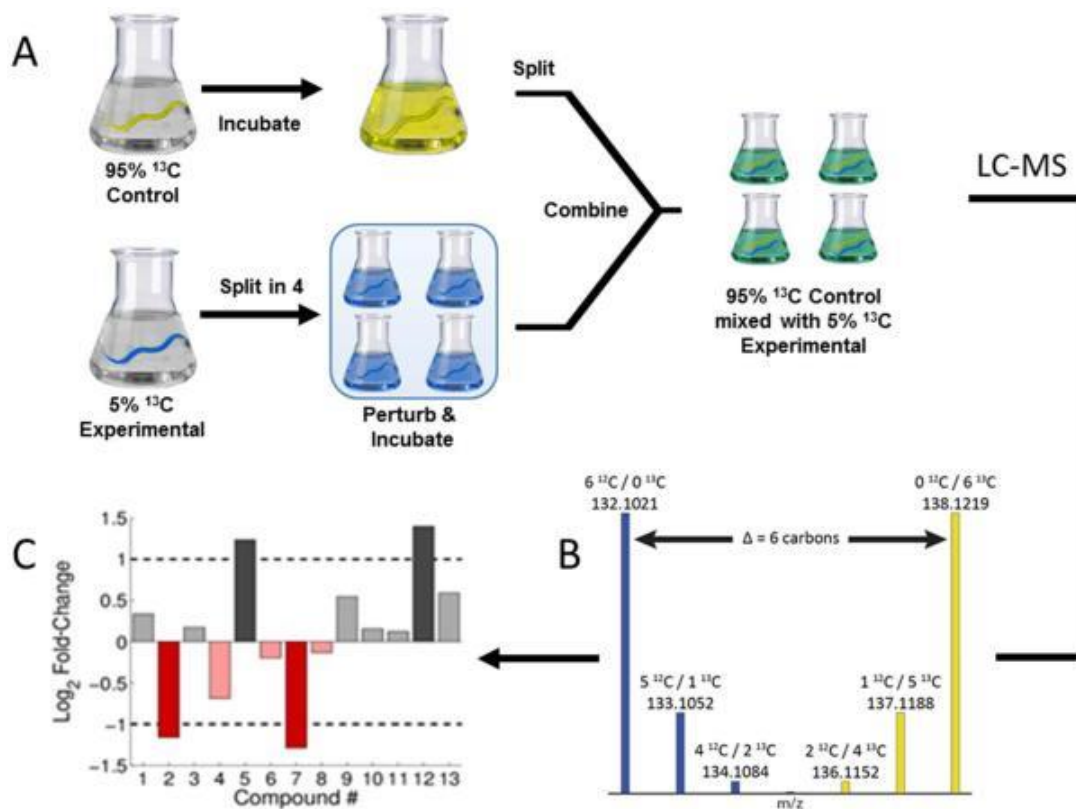
A path to metabolite identification

- Introduction
 - Analytical Technologies
 - Terminology in metabolite identification
- Metabolite identification
 - GC-MS MetID
 - EI spectral deconvolution
 - GC-APCI and exact mass
 - LC-MS MetID
 - Metabolite / Ion annotation
 - Formula calculation / seven golden rules
 - Utilization of isotopes
 - MS/MS and RT
- New approaches for MetID
 - Ion mobility
 - *in silico* approaches
- *de novo* identification



Metabolite Identification

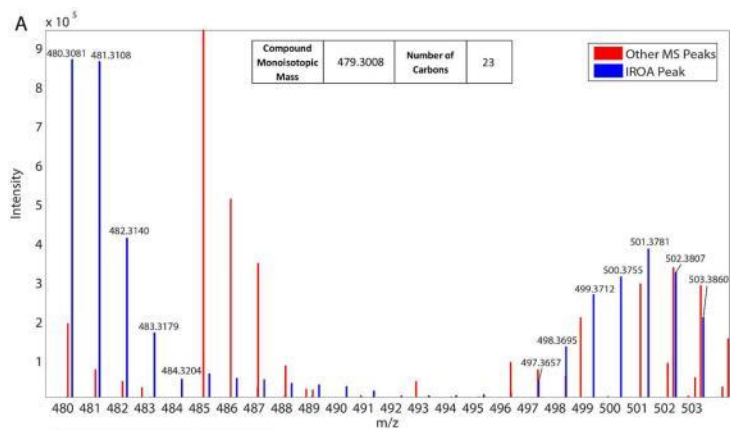
LC-MS MetID – Utilization of isotopes



Stupp et al., Anal. Chem. 2013, 85, 11858–11865

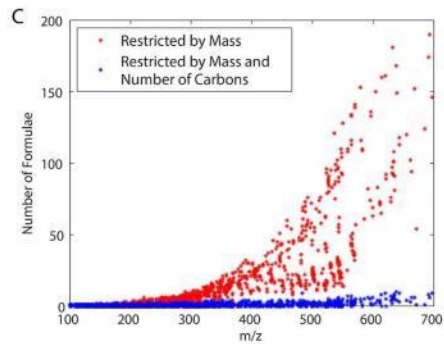
Metabolite Identification

LC-MS MetID – Utilization of isotopes



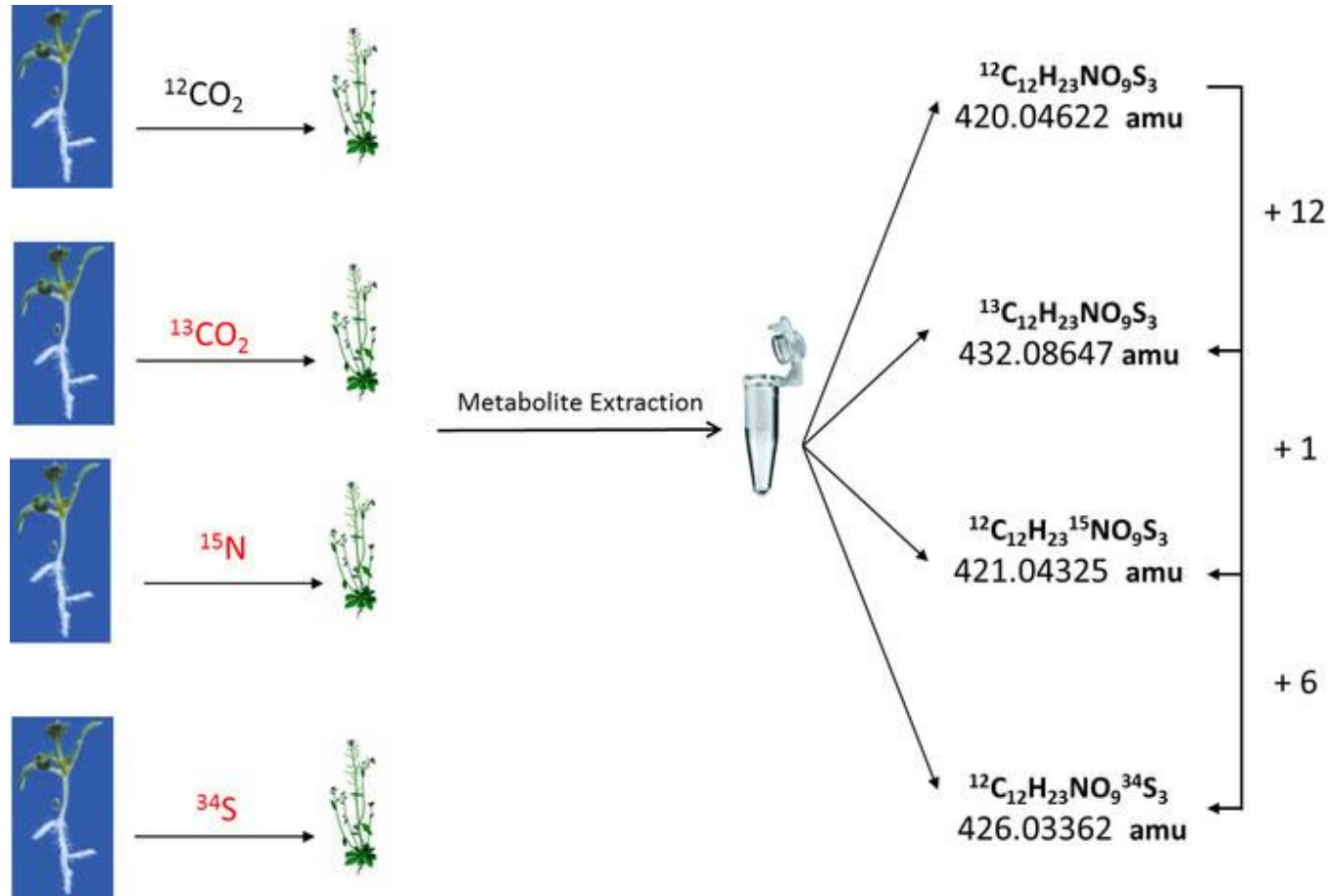
B

Molecular Formula	Formula Mass	PPM Error
C ₁₈ H ₁₇ N ₃ O ₂	479.3009	0.17
C ₁₈ H ₁₄ N ₃ O ₂ S ₂	479.3009	0.17
C ₁₈ H ₁₂ N ₃ OP ₂ S ₂	479.3010	0.38
C ₁₉ H ₁₂ N ₃ P ₃	479.3006	0.46
C ₁₈ H ₁₀ N ₃ S ₄	479.3006	0.46
C ₁₉ H ₁₀ N ₃ OP ₂	479.3005	0.67
C ₁₇ H ₁₀ NO ₃ P	479.3012	0.79
C ₁₉ H ₁₂ N ₃ O ₂	479.3004	0.88
C ₁₉ H ₁₀ N ₃ O ₂	479.3004	0.88
C ₁₇ H ₁₃ N ₃ O ₃	479.3014	1.21
C ₁₇ H ₁₇ N ₃ O ₂ P ₂	479.3002	1.29
C ₁₈ H ₁₂ N ₃ O ₂ S	479.3002	1.29
C ₁₈ H ₁₀ N ₃ O ₂ P ₂	479.3015	1.42
C ₁₉ H ₁₇ N ₃ S	479.3016	1.63
C ₁₉ H ₁₇ N ₃ O ₂	479.3000	1.71
C ₁₇ H ₁₄ N ₃ P ₃	479.3000	1.71
C ₁₉ H ₁₃ N ₃ P ₂ S ₂	479.3000	1.71



Metabolite Identification

LC-MS MetID – Utilization of isotopes

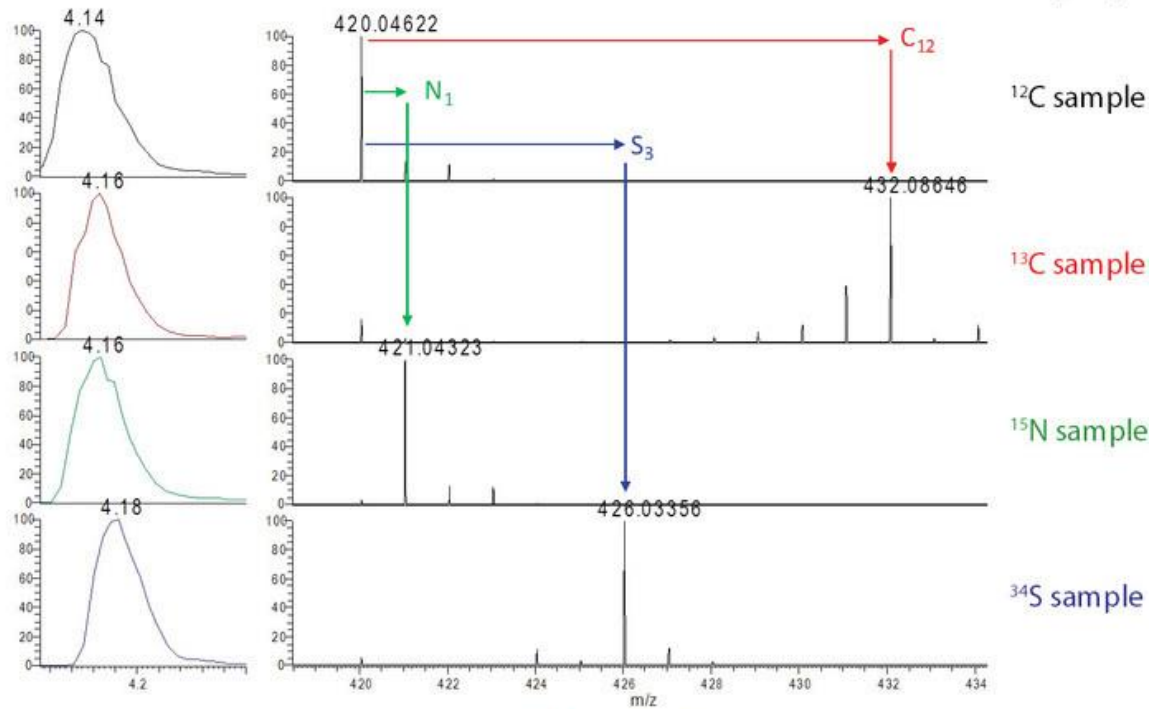
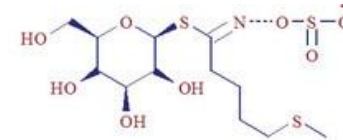


Patrick Gialalisco et al., The Plant Journal (2011) 68, 364–376

Metabolite Identification

LC-MS MetID – Utilization of isotopes

4-Methylthiobutyl glucosinolate (Glucoerucin)
 $[M - H]^-$



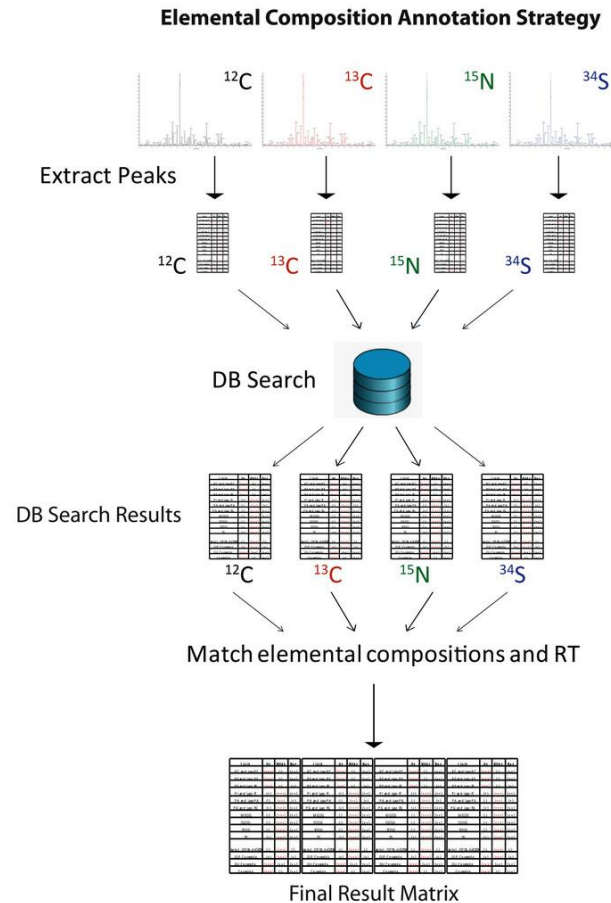
Average RT shift 0.6 sec



Patrick Gialalisco et al., *The Plant Journal* (2011) 68, 364–376

Metabolite Identification

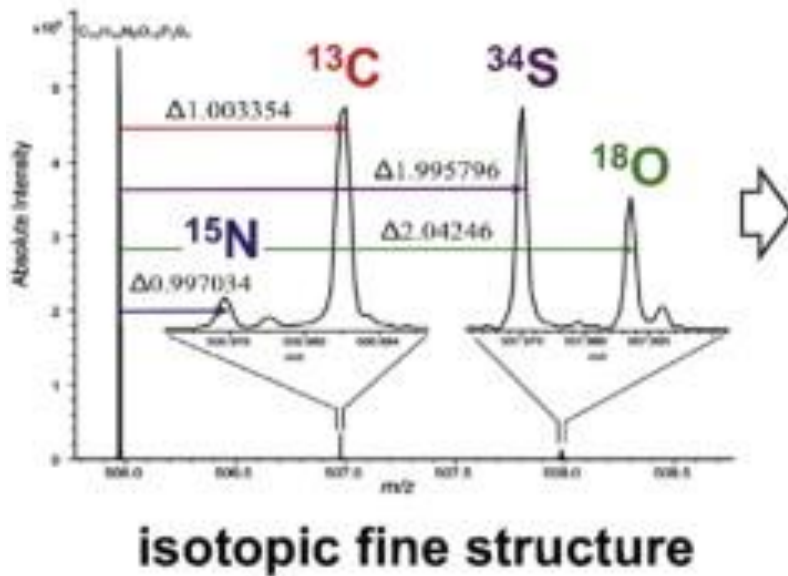
LC-MS MetID – Utilization of isotopes



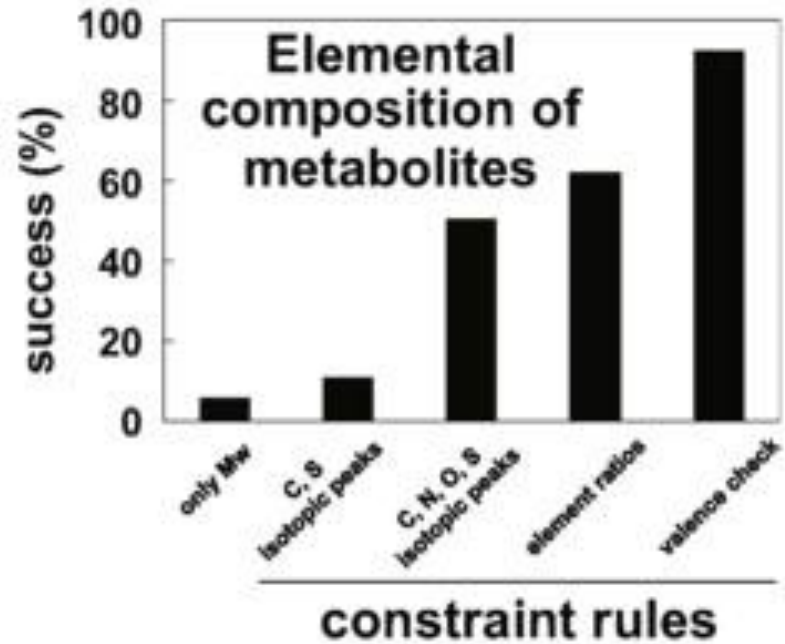
Patrick Gialalisco et al., The Plant Journal (2011) 68, 364–376

Metabolite Identification

LC-MS MetID – Utilization of isotopes



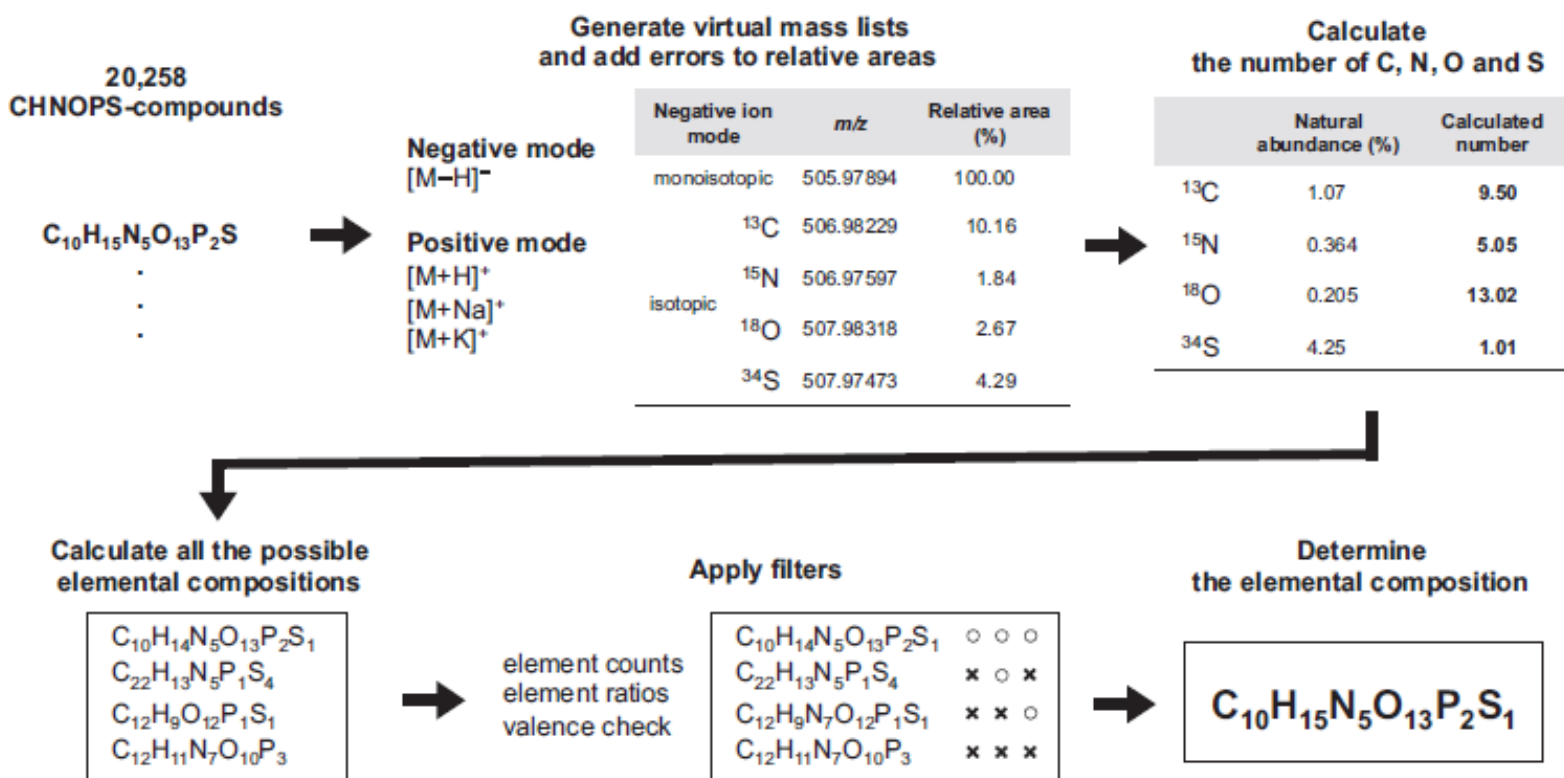
isotopic fine structure



T. Nagao et al., Analytica Chimica Acta 813 (2014) 70–76

Metabolite Identification

LC-MS MetID – Utilization of isotopes

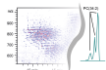


T. Nagao et al., Analytica Chimica Acta 813 (2014) 70–76

Metabolite Identification

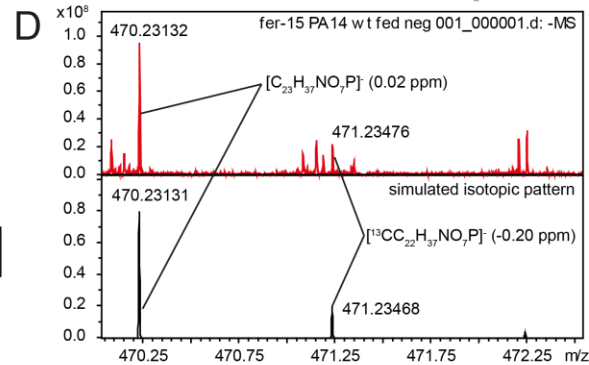
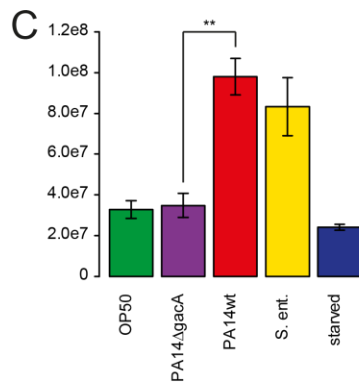
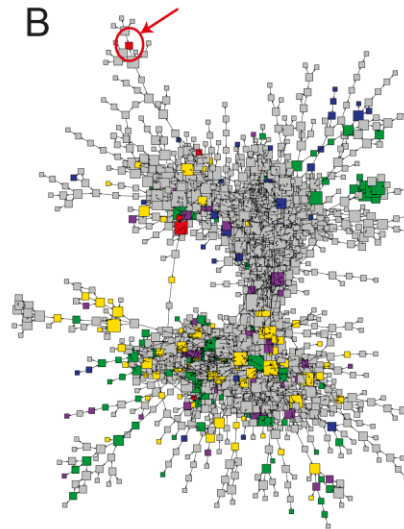
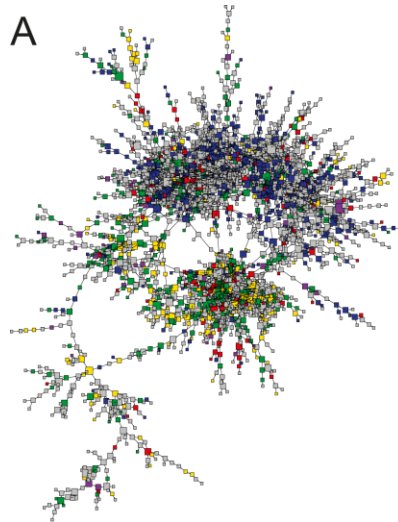
LC-MS MetID – Utilization of isotopes

Element	Isotope	Abundance (%)	Mass	Massdifference
H, hydrogen	¹ H	99.985	1.007825	
C, carbon	¹² C	98.93	12.000000	
	¹³ C	1.07	13.003355	1.003355
N, nitrogen	¹⁴ N	99.632	14.003074	
	¹⁵ N	0.368	15.000109	0.997035
O, oxygen	¹⁶ O	99.757	15.994915	
	¹⁷ O	0.038	16.999132	1.004217
	¹⁸ O	0.205	17.999160	2.004245
P, phosphorus	³¹ P	100	30.973762	
S, sulfur	³² S	94.93	31.972071	
	³³ S	0.76	32.971459	0.999388
	³⁴ S	4.29	33.967867	1.995796
Cl, chlorine	³⁵ Cl	75.78	34.968853	
	³⁷ Cl	24.22	36.965903	1.99705



Metabolite Identification

LC-MS MetID – Utilization of isotopes

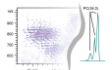


Witting et al., Anal Bioanal Chem (2015) 407:1059–1073

Outline

A path to metabolite identification

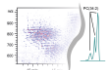
- Introduction
 - Analytical Technologies
 - Terminology in metabolite identification
- Metabolite identification
 - GC-MS MetID
 - EI spectral deconvolution
 - GC-APCI and exact mass
 - LC-MS MetID
 - Metabolite / Ion annotation
 - Formula calculation
 - Utilization of isotopes for formula calculation
 - MS/MS and RT
- New approaches for MetID
 - Ion mobility
 - *in silico* approaches
- *de novo* identification



Metabolite Identification

LC-MS MetID – MS/MS and RT

- Metabolomics Tandem MS databases
 - Metlin
 - MassBank of Japan
 - MassBank of Europe
 - MassBank of North America (Mona)
 - Global Natural Products Social Molecular Networking (GNPS)
 - Human Metabolome Database (HMDB)
 - RIKEN MSn spectral database (Respect)
- Overlap between all spectra (Splash <http://splash.fiehnlab.ucdavis.edu/>)
- *in silico tandem* solutions
 - MetFrag / MetFusion
 - CSI:FingerID
 - CFM-ID
 - MAGMa

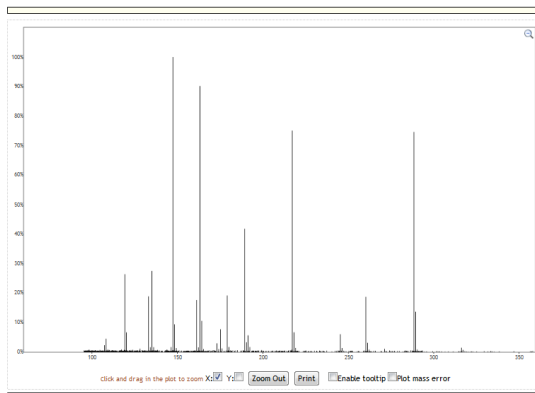


Metabolite Identification

LC-MS MetID – MS/MS and RT



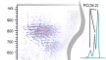
GNPS Library Spectrum CCMSLIB0000079354



Library Spectrum Information	
Spectrum ID	CCMSLIB0000079354
Compound Name	*ML5001158449-01(2S)-2-(((2S)-3-methyl-2-(((2S)-3-oxo-2-phenylpropan-2-yl-2,4-dihydroquinoline-1-carbonyl)amino)butano)amino)-3-phenylpropanoic acid
PI	Dorrestein
Data Collector	VFP/LMS
CAS Number	N/A
Original Submitter	gnpslib
Most Recent Revisor	gnpslib
Library Quality	Gold Spectrum
Smiles	CC(C)C[C@H](NC(=O)N[C@@H](C)C(C)C(=O)NC2C=CC=CC1C(C)=O)N[C@@H](C)C2C=CC=CC1C(=O)O
InChI	N/A
Structure	
Precursor <i>m/z</i>	481.245
Exact Mass	480.237

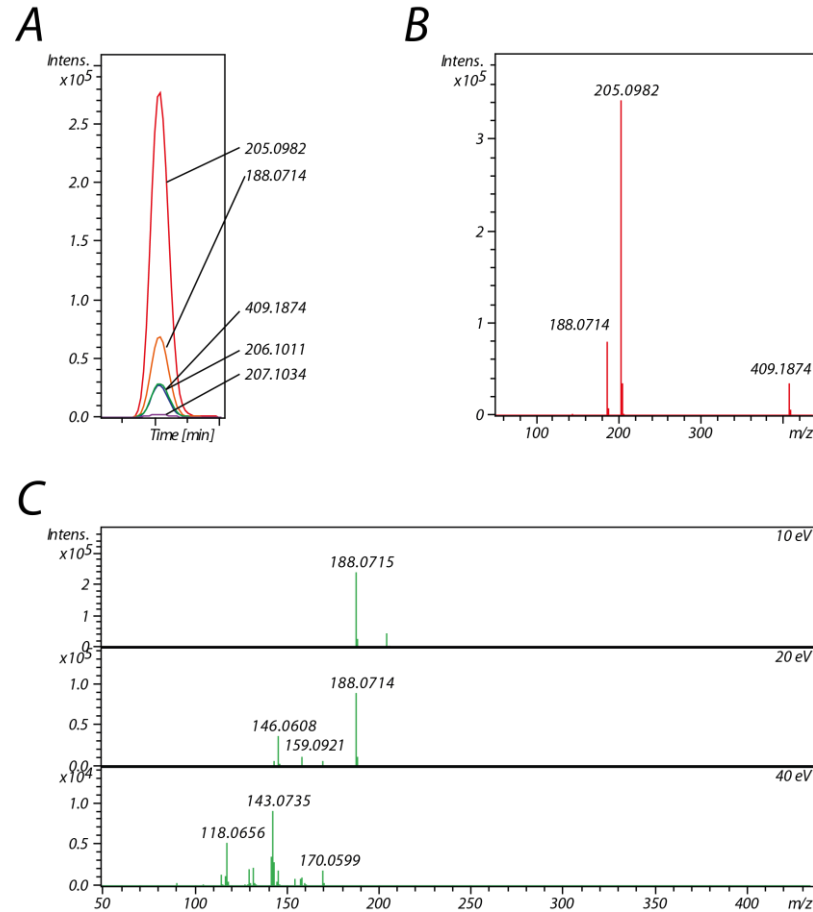
https://gnps.ucsd.edu/ProteoSAFe/gnpslibraryspectrum.jsp?SpectrumID=CCMSLIB0000079354&#{}

<http://mona.fiehnlab.ucdavis.edu/spectra/browse>



Metabolite Identification

LC-MS MetID – MS/MS and RT



Metabolite Identification

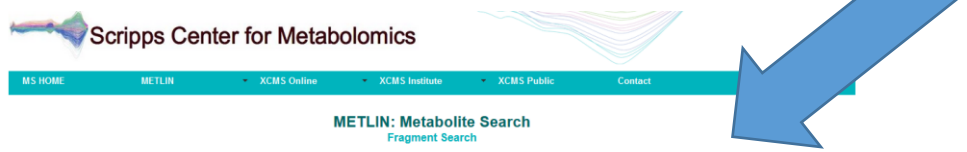
LC-MS MetID – MS/MS and RT

```

1 CHARGE=1+, 2+ and 3+
2 ### D:\Users\michael.witting\Desktop\Trypto10eVpos.mgf
3 ### Y:\msk\is\data\2016_1\Michael\20160211_MRM_DB_pos\MRM_1A5_1-A_5_01_5202.d\A
4 ### MRM_1A5_1-A_5_01_5202.d
5 ### Instrument: mskXis
6 ### MRM_1A5
7 ### S/N 3.0, RelIntensity 0.0, MinimumIntensity 100, Centroid 80.0, UseCent
8 ### Retain Residuals no, Create neutrally no,
9 ### FullScan Parameters: Adduction (+)H (-)-H, LowMass 250, HighMass 4000,
10 ### Perf. isotope decon yes, Max Charge Istopic decon 0, MW agreement 0.00,
11 ### MaxRes Parameters: Adduction (+)H (-)-H, LowMass 250, HighMass 4000
12 ### Perf. isotope decon yes, Max Charge Istopic decon 0, Abund. cutoff 2.00
13 ### Msn Parameters: Adduction (+)H (-)-H, LowMass 100, HighMass 3000
14 ### Perf. isotope decon yes, Max Charge Istopic decon 0, Abund. cutoff 1.00
15 ### Global Charge Limit: yes 3, Prefer FullScan Result: yes, Export what: 1
16 ### *MS2 (205.0985), 10.0eV, 2.42min #1057
17
18 ###FS: #m/z: 205.0984 #charge 1
19 ###MaxRes:#m/z: 0.00000 #charge=128
20
21 ###MS:
22 BEGIN IONS
23 TITLE=MS2 (205.0985), 10.0eV, 2.42min #1057
24 RTINSECONDS=0
25 SCANS=MS: MSMS:
26 PEPMASS=205.09846
27
28 159.09264 3888 1+
29 188.07206 264768 1+
30 205.09839 51752 1+
31 END IONS
32
33

```

<https://metlin.scripps.edu/index.php>
https://metlin.scripps.edu/fragment_search_multi.php



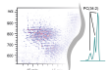
m/z with 30 ppm mass accuracy at Positive Mode Filter

Change Query

Metabolite(s) with containing 2 fragments:

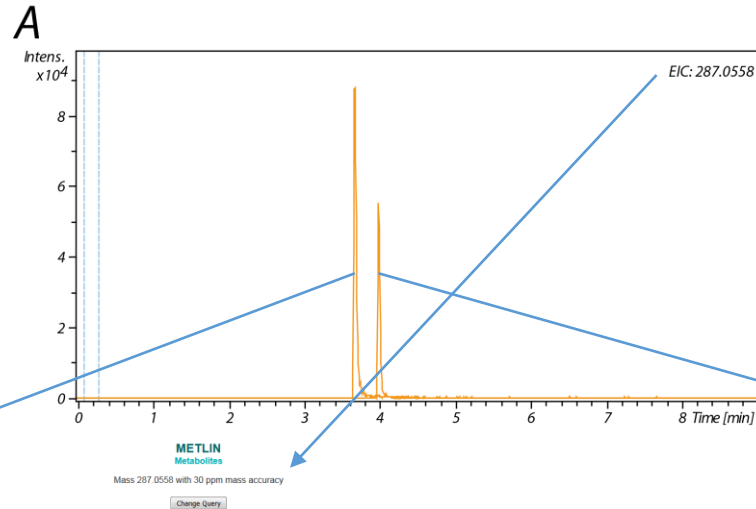
METLIN ID: 33 NAME: L-Tryptophan MASS: 204.0899 [View MS/MS](#) [STRUCTURE](#)

Frag. m/z	Appm	Intensity	CE	Predicted Ion Type	Predicted Fragment Structure
188.0700	10	100.0	10, 20	[M+H] ⁺	
159.0900	16	11.3	10, 20	[M-H+2H] ⁺	



Metabolite Identification

LC-MS MetID – MS/MS and RT



```
###FS: #m/z: 287.05577 #charge 10
###MaxRes:#m/z: 0.00000 #charge-128
###MS:
###MSMS:
BEGIN IONS
TITLE=+MS2 (287.0558), 40.0eV, 3.69min #1672
RTINSECONDS=0
SCANS=MS: MSMS:
PEPMASS=287.05577
```

```
135.04405 1788 1+
137.02522 676 2+
153.01855 8412 1+
161.02339 1490 1+
179.03415 816 1+
186.05838 1012 1+
203.03516 772 1+
213.05480 1352 1+
241.04922 4240 1+
245.04233 564 1+
259.06004 944 1+
269.04502 2540 1+
287.05611 34712 10+
287.11943 984 8+
288.05940 6092 7+
289.05938 1100 1+
END IONS
```

Total: 38 Metabolites

METLIN ID	MASS	Appm	NAME	MS/MS	STRUCTURE
68645	[M+H] ⁺ m/z 287.0558 M 264.0066	0	Ozagrel hydrochloride Formula C ₁₃ H ₁₃ ClN ₂ O ₂ CAS: 78712-43-3	NO	
50019	[M+H] ⁺ m/z 287.0560 M 286.0477	2	3,7,8,4'-Tetrahydroxyflavone Formula C ₁₅ H ₁₀ O ₆ CAS:	NO	
80040	[M+H] ⁺ m/z 287.0560 M 286.0477	2	Datsacetin Formula C ₁₅ H ₁₀ O ₆ CAS:	NO	
49785	[M+H] ⁺ m/z 287.0560 M 286.0477	2	Isoastilarein Formula C ₁₅ H ₁₀ O ₆ CAS:	NO	
43369	[M+H] ⁺ m/z 287.0560 M 286.0477	2	Norartocarpetin Formula C ₁₅ H ₁₀ O ₆ CAS:	NO	

```
###FS: #m/z: 287.05566 #charge 1
###MaxRes:#m/z: 0.00000 #charge-128
###MS:
###MSMS:
BEGIN IONS
TITLE=+MS2 (287.0557), 40.0eV, 4.01min #1824
RTINSECONDS=0
SCANS=MS: MSMS:
PEPMASS=287.05566
```

```
136.01759 716 1+
140.06141 332 1+
147.04301 456 1+
153.01913 4052 1+
153.06592 316 4+
154.04158 312 1+
157.06510 1060 1+
165.01940 2452 1+
178.02877 276 1+
185.06013 1268 1+
188.04982 392 1+
195.04595 280 1+
197.06170 280 1+
201.05693 384 1+
213.05481 2080 11+
229.04913 500 1+
231.06539 1144 1+
241.05134 2036 1+
244.03836 308 1+
258.05319 2328 1+
269.04640 676 1+
285.04744 292 1+
287.05561 13408 1+
287.17779 448 11+
END IONS
```

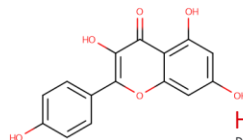
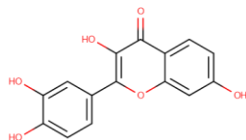
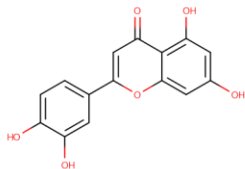
Metabolite Identification

LC-MS MetID – MS/MS and RT

3.69 minutes

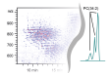
Metabolite(s) with containing 2 fragments:

Frag. m/z	Appm	Intensity	CE	Predicted Ion Type	Predicted Fragment Structure
METLIN ID: 85043 NAME: Demoxepam MASS: 286.0509 View MS/MS STRUCTURE					
241.0520	11	100.0	20.40	-	No structure information is available
269.0470	7	71.7	20.40	-	No structure information is available
METLIN ID: 44206 NAME: Fisetin MASS: 286.0477 View MS/MS STRUCTURE					
287.0520	14	100.0	10.20, 40	-	No structure information is available
241.0480	5	10.8	20.40	[M+H] ⁺	
METLIN ID: 3409 NAME: Luteolin MASS: 286.0477 View MS/MS STRUCTURE					
153.0180	3	100.0	20.40	[M+H] ⁺	
135.0440	0	20.5	40	[M+H] ²⁺	
METLIN ID: 897 NAME: Oxazepam MASS: 286.0509 View MS/MS STRUCTURE					
287.0570	3	100.0	10.20	-	No structure information is available
153.0210	16	18.6	40	[M+H] ²⁺	



4.01 minutes

Frag. m/z	Appm	Intensity	CE	Predicted Ion Type	Predicted Fragment Structure
METLIN ID: 3410 NAME: Kaempferol MASS: 286.0477 View MS/MS STRUCTURE					
213.0530	8	20.8	40	[M+H] ²⁺	
165.0200	3	17.1	20.40	[M] ⁺	
258.0510	8	11.5	40	[M+H] ²⁺	
m/z matches: 1					
METLIN ID: 85043 NAME: Demoxepam MASS: 286.0509 View MS/MS STRUCTURE					
241.0520	2	100.0	20.40	-	No structure information is available
METLIN ID: 44206 NAME: Fisetin MASS: 286.0477 View MS/MS STRUCTURE					
241.0480	13	10.8	20.40	[M+H] ²⁺	
METLIN ID: 897 NAME: Oxazepam MASS: 286.0509 View MS/MS STRUCTURE					
153.0210	12	18.6	40	[M+H] ²⁺	
METLIN ID: 3409 NAME: Luteolin MASS: 286.0477 View MS/MS STRUCTURE					
153.0180	7	100.0	20.40	[M+H] ⁺	



Metabolite Identification

LC-MS MetID – MS/MS and RT

<http://msbi.ipb-halle.de/MetFusion/>

```

###FS:      #m/z: 287.05577 #charge 10
###MaxRes:#m/z: 0.00000 #charge-128
###MS:
###MSMS:
BEGIN IONS
TITLE=+MS2 (287.0558), 40.0eV, 3.69min #1672
RTINSECONDS=0
SCANS=MS:  MSMS:
PEPMASS=287.05577

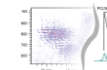
135.04405  1788  1+
137.02522  676  2+
153.01855  8412  1+
161.02339  1480  1+
179.03415  816  1+
185.05838  1012  1+
203.03516  772  1+
213.05480  1352  1+
241.04922  4240  1+
245.04233  564  1+
259.06004  944  1+
269.04502  2540  1+
287.05611  34712  10+
287.11843  984  8+
288.05840  6092  7+
289.05988  1100  1+
END IONS
    
```

The screenshot shows the MetFusion web interface. On the left, there are tabs for 'Spectrum Query' and 'Information'. The 'Spectrum Query' tab is active, showing 'Selected Spectral Database: Nistbank'. Below this, there are sections for 'MassBank Parameters', 'MetFusion Parameters', and 'MetFrag Parameters'. The 'MetFusion Parameters' section includes a 'Filter' dropdown set to 'Unique' and a list of 'Peaks' with their m/z values. The 'MetFrag Parameters' section includes 'Upstream DB' (KEGG, ChemSpider, SDF Upload), 'Database IDs', 'Molecular Formula', 'Parent Ion' (286.0467, Neutral), 'Exact Mass' (286.0467), 'Limit # of Structures' (1000), 'C,H,N,O,P,S only?' (checked), 'Search PPM' (10.0), 'mz abs' (0.01), and 'mz ppm' (10.0). There are 'Start' and 'Reset' buttons at the bottom left. A 'Feedback' button is visible on the right side of the interface.

The screenshot shows the 'Query Results' page of MetFusion. It displays a table with columns: 'Port Name', 'Record ID', 'Compound Name', 'Exact Mass', 'Structure', 'MetFusion - InSitu Clustering Ranking', 'Score', 'Integration Score', and 'Fragments'. The table lists four identified metabolites: Kaempferol, Fisetin, Luteolin, and Daidzein. Each entry includes a chemical structure and a 'Peaks explained' count.

Port Name	Record ID	Compound Name	Exact Mass	Structure	MetFusion - InSitu Clustering Ranking	Score	Integration Score	Fragments
MetFrag	cpd_C0593	Kaempferol	286.048	<chem>O=C1C(=C(OC2=CC(=O)C=C2)C=CC1=O</chem>	0.748	3.56	Peaks explained: 11	Compute Fragments
MetFrag	cpd_C1001	Fisetin	286.048	<chem>O=C1C(=C(OC2=CC(=O)C=C2)C=CC1=O</chem>	0.648	3.532	Peaks explained: 9	Compute Fragments
MetFrag	cpd_C0115	Luteolin	286.048	<chem>O=C1C(=C(OC2=CC(=O)C=C2)C=CC1=O</chem>	0.689	3.355	Peaks explained: 10	Compute Fragments
MetFrag	cpd_C1903	Daidzein	286.048	<chem>O=C1C(=C(OC2=CC(=O)C=C2)C=CC1=O</chem>	0.782	3.23	Peaks explained: 11	Compute Fragments

Gerlich et al., J. Mass Spectrom. 2013, 48, 291–298



Metabolite Identification

LC-MS MetID – MS/MS and RT

Spectrum Query Information

Selected Spectral Database: MassBank

MassBank Parameters
<http://www.massbank.jp/>
MassBank Server: MassBank JP
Number of Results: 100
Cutoff threshold of relative intensities: 5
Ionization Mode: positive
MS Levels: all MS MS2 MS3 MS4
Instruments:

Select EI	Deselect ESI	Select Others
<input type="checkbox"/> EI-B	<input checked="" type="checkbox"/> CE-ESI-TOF	<input type="checkbox"/> APCH-ITFT
<input type="checkbox"/> EI-EBEB	<input type="checkbox"/> ESI-ITFT	<input type="checkbox"/> APCH-ITTOF
<input type="checkbox"/> GC-EI-QQ	<input type="checkbox"/> ESI-ITTOF	<input type="checkbox"/> CI-B
<input type="checkbox"/> GC-EI-TOF	<input checked="" type="checkbox"/> ESI-QTOF	<input type="checkbox"/> FAB-B
	<input type="checkbox"/> LC-ESI-HT	<input type="checkbox"/> FAB-EB
	<input type="checkbox"/> LC-ESI-ITFT	<input type="checkbox"/> FAB-EBEB
	<input type="checkbox"/> LC-ESI-ITTOF	<input type="checkbox"/> FD-B
	<input type="checkbox"/> LC-ESI-Q	<input type="checkbox"/> FI-B
	<input type="checkbox"/> LC-ESI-QFT	<input type="checkbox"/> LC-APCI-QTOF
	<input type="checkbox"/> LC-ESI-QIT	<input type="checkbox"/> LC-APPI-QQ
	<input type="checkbox"/> LC-ESI-QQ	<input type="checkbox"/> MALDI-QIT
	<input checked="" type="checkbox"/> LC-ESI-QTOF	<input type="checkbox"/> MALDI-TOF
	<input checked="" type="checkbox"/> LC-ESI-TOF	<input type="checkbox"/> MALDI-TOFTOF

MetFrag Parameters

Upstream DB: KEGG ChemSpider SDF Upload

Database IDs:

Molecular Formula:

Parent Ion: 287.05577 [M+H]⁺

Exact Mass: 286.048493545

Limit # of Structures: 500

C,H,N,O,P,S only?

Search PPM: 10.0

mz abs: 0.01

mz ppm: 10.0

Start Reset

Spectrum Query Information

Selected Spectral Database: Metlin

Metlin Parameters
<http://metlin.scripps.edu/>
Ionization Mode: positive negative
Collision Energy: 40 eV
Tolerance MSMS (Da): 0.01
Tolerance Precursor (ppm): 5

MetFrag Parameters

Upstream DB: KEGG ChemSpider SDF Upload

Database IDs:

Molecular Formula:

Parent Ion: 287.05577 [M+H]⁺

Exact Mass: 286.048493545

Limit # of Structures: 500

C,H,N,O,P,S only?

Search PPM: 10.0

mz abs: 0.01

mz ppm: 10.0

Start Reset

Gerlich et al., J. Mass Spectrom. 2013, 48, 291–298

Metabolite Identification

LC-MS MetID – MS/MS and RT

3.69 minutes

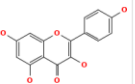
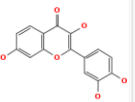
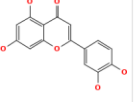
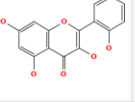
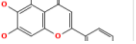
Spectrum Query Query Results Similarity Matrix Information

MetFusion Cluster Fragmenter List Database List Unused Database Entries

Export all results into Excel file

or reranked results only into SD file

MetFusion - Tanimoto Clustering Ranking

Port Name	Record ID	Compound Name	Exact Mass	Structure	Score	Integration Score	Fragments
MetFrag	cpd_C05903	Kaempferol	286.048		0.748	3.317	Peaks explained: 11 Compute Fragments
MetFrag	cpd_C10011	Fisetin	286.048		0.648	3.257	Peaks explained: 9 Compute Fragments
MetFrag	cpd_C01514	Luteolin	286.048		0.689	3.074	Peaks explained: 10 Compute Fragments
MetFrag	cpd_C10036	Datiscetin	286.048		0.782	2.972	Peaks explained: 11 Compute Fragments
MetFrag	cpd_C10184	Scutellarein	286.048		0.585	2.695	Peaks explained: 9 Compute Fragments

4.01 minutes

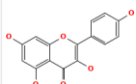
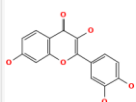
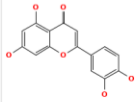
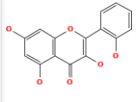
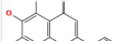
Spectrum Query Query Results Similarity Matrix Information

MetFusion Cluster Fragmenter List Database List Unused Database Entries

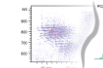
Export all results into Excel file

or reranked results only into SD file

MetFusion - Tanimoto Clustering Ranking

Port Name	Record ID	Compound Name	Exact Mass	Structure	Score	Integration Score	Fragments
MetFrag	cpd_C05903	Kaempferol	286.048		0.758	3.017	Peaks explained: 11 Compute Fragments
MetFrag	cpd_C00011	Fisetin	286.048		0.717	2.969	Peaks explained: 10 Compute Fragments
MetFrag	cpd_C01514	Luteolin	286.048		0.611	2.729	Peaks explained: 9 Compute Fragments
MetFrag	cpd_C10036	Datiscetin	286.048		0.747	2.672	Peaks explained: 11 Compute Fragments
MetFrag	cpd_C10184	Scutellarein	286.048		0.637	2.387	Peaks explained: 9 Compute Fragments

Gerlich et al., J. Mass Spectrom. 2013, 48, 291–298



Metabolite Identification

LC-MS MetID – MS/MS and RT

3.69 minutes

Luteolin
Apigenin
Kaempferol

4.01 minutes

Kaempferol
Maritemein
Fisetin

Datisetin
Kaempferol

Luteolin

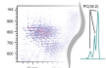
Kaempferol
Datisetin
Fisetin
Luteolin

MetFrag \ MassBank Compounds	PB000743	FIO00087	PB005705	FIO00762	PR020007	FIO00173	PR100243	F
cpd:C10036	0.673	0.783	0.885	0.23	0.258	0.597	0.569	
cpd:C05903	0.752	0.872	1.0	0.225	0.253	0.616	0.587	
cpd:C17786	0.509	0.415	0.434	0.199	0.259	0.365	0.347	
cpd:C10510	0.528	0.494	0.492	0.218	0.268	0.407	0.424	
cpd:C12134	0.501	0.456	0.49	0.213	0.248	0.391	0.417	
cpd:C08720	0.516	0.549	0.495	0.221	0.237	0.428	0.446	
cpd:C01514	1.0	0.742	0.752	0.23	0.256	0.511	0.541	
cpd:C08576	0.576	0.549	0.553	0.22	0.237	0.443	0.467	
cpd:C10041	0.742	1.0	0.872	0.232	0.26	0.593	0.571	
cpd:C10184	0.881	0.664	0.714	0.222	0.262	0.497	0.524	
cpd:C10097	0.85	0.647	0.694	0.224	0.264	0.492	0.518	
cpd:C17810	0.304	0.291	0.282	0.206	0.231	0.273	0.27	

12 candidates found, displaying 12 candidates from 1 to 12. Page 1 / 1

MetFrag \ MassBank Compounds	PB005706	PR100360	FIO00087	PB000743	PR100243	PR101007	PR040005
cpd:C05903	1.0	0.42	0.872	0.752	0.587	0.592	0.253
cpd:C10036	0.885	0.418	0.783	0.673	0.569	0.575	0.258
cpd:C10041	0.872	0.461	1.0	0.742	0.571	0.576	0.26
cpd:C10184	0.714	0.419	0.664	0.881	0.524	0.526	0.262
cpd:C01514	0.752	0.424	0.742	1.0	0.541	0.543	0.256
cpd:C08576	0.553	0.564	0.549	0.576	0.467	0.471	0.237
cpd:C08720	0.495	0.635	0.549	0.516	0.446	0.45	0.237
cpd:C10097	0.694	0.432	0.647	0.85	0.518	0.52	0.264
cpd:C17786	0.434	0.339	0.415	0.509	0.347	0.35	0.259
cpd:C10510	0.492	0.418	0.494	0.528	0.424	0.426	0.268
cpd:C12134	0.49	0.384	0.456	0.501	0.417	0.418	0.248
cpd:C17810	0.282	0.27	0.291	0.304	0.27	0.27	0.231

12 candidates found, displaying 12 candidates from 1 to 12. Page 1 / 1



Metabolite Identification

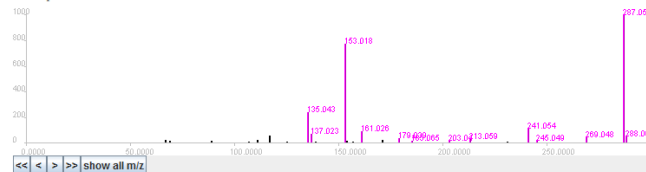
LC-MS MetID – MS/MS and RT

MassBank Record: PB000743

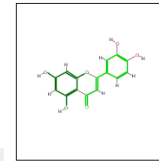
[Home](#) | [Spectrum](#) | [Quick Peak](#) | [Substructure](#) | [Prediction](#) | [Browser](#) | [Batch](#) | [Browse](#) | [Index](#) | MassBank ID:

Luteolin; LC-ESI-QTOF; MS2; CE:40 eV; [M+H]⁺

Mass Spectrum



Chemical Structure



ACCESSION: PB000743
RECORD_TITLE: Luteolin; LC-ESI-QTOF; MS2; CE:40 eV; [M+H]⁺
DATE: 2016.01.19 (Created 2008.03.23, modified 2013.06.04)
AUTHORS: Boettcher C, Institute of Plant Biochemistry, Halle, Germany
LICENSE: [CC BY-SA](#)
COMMENT: IPB_RECORD: 361
COMMENT: CONFIDENCE: confident structure

CHSNAME: Luteolin
CHSNAME: 5,7,3',4'-tetrahydroxy-flavone
CHSCOMPOUND_CLASS: Natural Product; Flavone
CHSFORMULA: [C15H10O6](#)
CHSEXACT_MASS: 286.04774
CHSMILES: O=C1C(=C(C=C1)O)C(=O)C2=C(C=C2)O
CHSIUPAC: InChI=1S/C15H10O6/c16-8-4-11(19)15-12(20)6-13(21-14(15)5-8)7-1-2-9(17)10(18)3-7/h1-6,16-19H
CHSLINK: INCHIKEY [IQPNAANSBPBGFQ-UHFFFAOYSA-N](#)
CHSLINK: KEGG [C01514](#)
CHSLINK: PUBCHEM CID: [5280445](#)

ACQINSTRUMENT: API QSTAR Pulsar 1
ACQINSTRUMENT_TYPE: LC-ESI-QTOF
ACSMASS_SPECTROMETRY: MS TYPE MS2
ACSMASS_SPECTROMETRY: ION MODE POSITIVE
ACSMASS_SPECTROMETRY: COLLISION_ENERGY 40 eV
ACSMASS_SPECTROMETRY: IONIZATION ESI

MSFOCUSSED_ION: PRECURSOR_TYPE [M+H]⁺

PKSFLASH: [splash10-izw000000-e245bb6cbc97af71cd5e](#)

PKSNUM_PEAK: 25

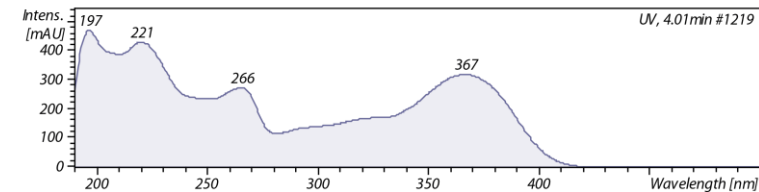
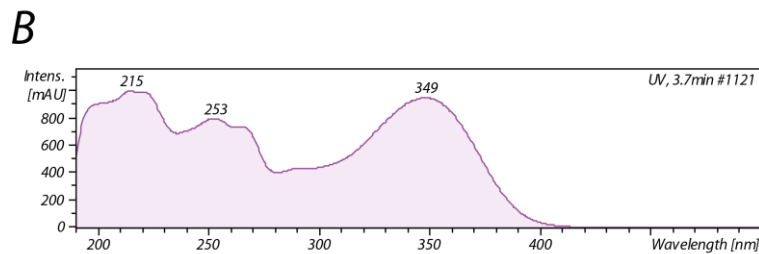
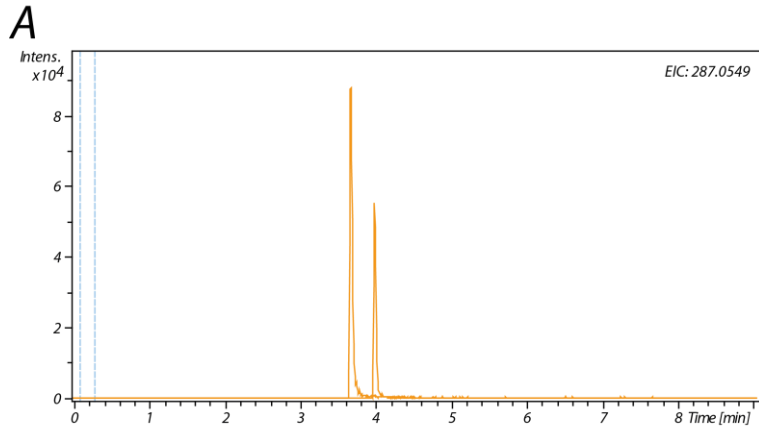
PKSPEAK: m/z int. rel.int.

m/z	int.	rel.int.
67.018	257.309	24
68.996	198.486	18
89.038	161.870	15
107.048	123.989	11
111.007	233.672	22
117.032	564.586	55
125.020	104.437	9
135.043	150	15
153.018	600	60
161.028	100	10
179.009	100	10
203.013	100	10
241.054	100	10
245.049	100	10
269.048	100	10
285.06	100	10
287.055	1000	100

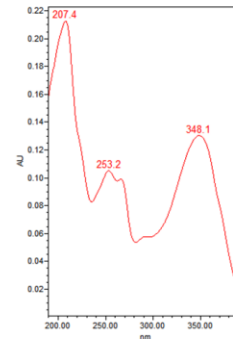
Horai et al., J. Mass. Spectrom. 2010, 45, 703–714.

Metabolite Identification

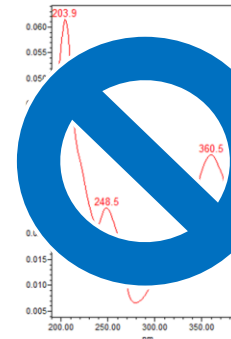
LC-MS MetID – MS/MS and RT



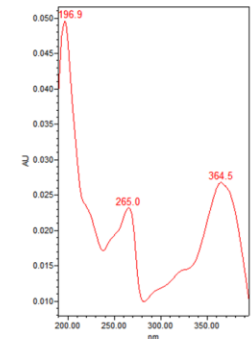
Luteolin



Fisetin



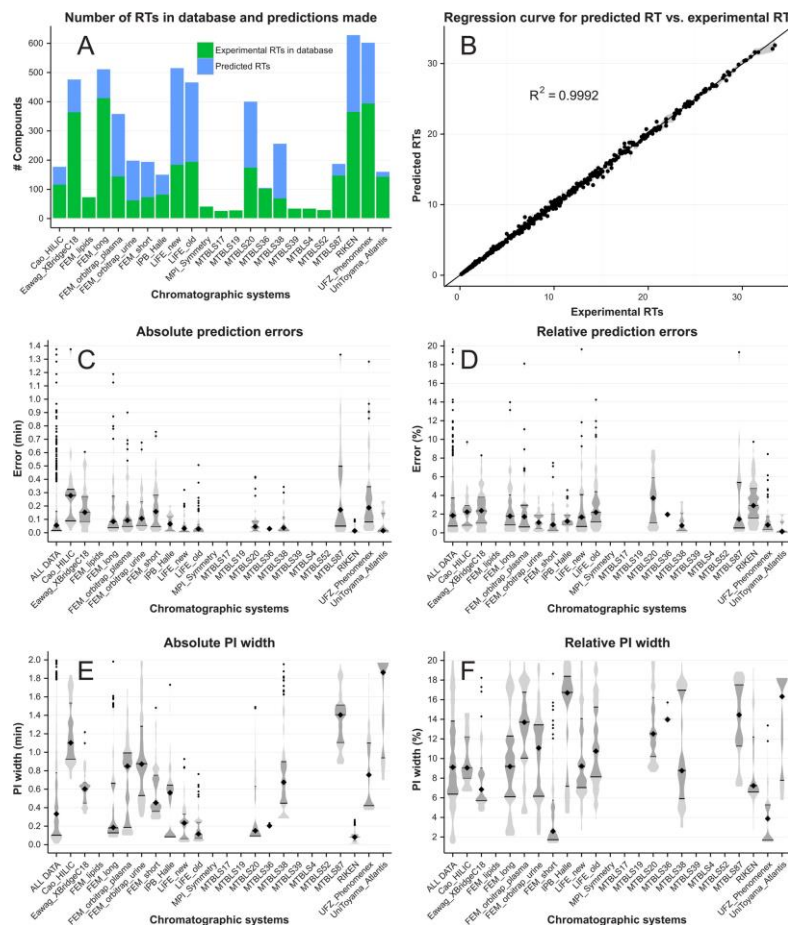
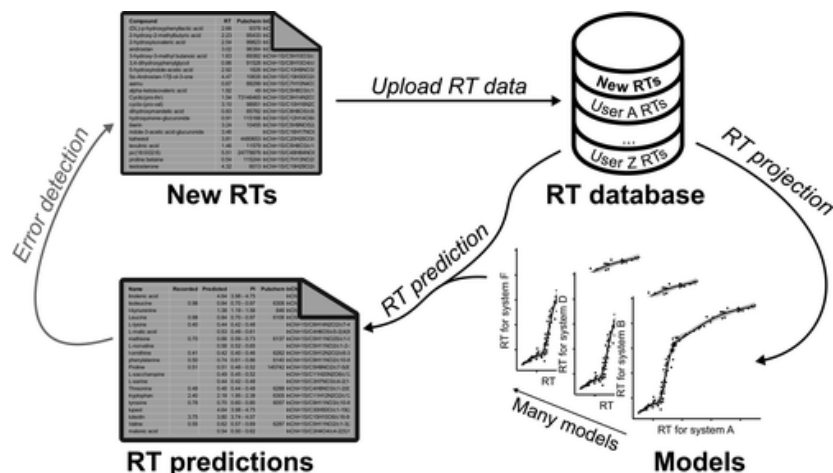
Kaempferol



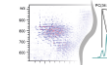
Metabolite Identification

LC-MS MetID – MS/MS and RT

- RTs can be projected between different (comparable) LC-MS setups
- If the same compounds have been measured on both systems a function for projection can be calculated allowing to predict RT of compound measured on one system to another



Stanstrup et al., Anal. Chem., 2015, 87 (18), pp 9421–9428



Metabolite Identification

LC-MS MetID – MS/MS and RT

The screenshot shows the PredRet website interface. At the top, there is a navigation bar with the PredRet logo, the text 'Prediction of retention times in LC systems', and logos for 'THE VELUX FOUNDATIONS' and 'FONDAZIONE EDMUND MACH'. Below the navigation bar, there are tabs for 'PredRet Introduction', 'Retention time prediction', and 'Tools', along with a 'Log Out' button. The main content area is titled 'How to use PredRet' and contains a 'Contents [hide]' section with a list of seven numbered items: 1 Preparing the CSV file, 2 Defining your chromatographic system, 3 Uploading your data, 4 Managing your data, 5 Getting the predictions, 6 Inspecting suspicious experimental values, and 7 Downloading the PredRet database. The 'Preparing the CSV file' section is expanded, showing instructions on how to use the system and a list of required columns: Compound, rt, Method, Pubchem, and Inchi. Below the instructions, a table is shown as an example of a CSV file structure, with columns labeled A, B, C, and D, and rows for 'Method', 'Compound', 'rt', 'Pubchem', and 'inchi'.

PredRet
Prediction of retention times in LC systems

THE VELUX FOUNDATIONS
VILLUM FONDEN × VELUX FONDEN

FONDAZIONE EDMUND MACH

PredRet Introduction | Retention time prediction | Tools | Log Out

How to use PredRet

Contents [hide]

- 1 Preparing the CSV file
- 2 Defining your chromatographic system
- 3 Uploading your data
- 4 Managing your data
- 5 Getting the predictions
- 6 Inspecting suspicious experimental values
- 7 Downloading the PredRet database

Preparing the CSV file

To use the system you first need to upload a CSV file with information about the retention time (RT) of compounds in your own system(s).

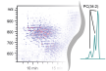
The CSV file you will prepare should contain the following named columns:

- **Compound:** The name of the compound. (column is mandatory)
- **rt:** The RT of the compound in your system. Please use RTs in minutes. (column is mandatory)
- **Method:** The name of the chromatographic method. The method must first be described in the "add or modify system" tab. (this column is mandatory unless a method name is selected in the drop down menu). See below.
- **Pubchem:** The PubChem CID of the compound. (Optional. Either PubChem or InChi must be supplied for each compound)
- **Inchi:** The InChi of the compound. (Optional. Either PubChem or InChi must be supplied for each compound)

When you are done your CSV file should look something like in the following figure.

	A	B	C	D
1	Method	Compound	rt	Pubchem inchi
71	MTBLS36	citric acid	4.399448944	19782904
72	MTBLS36	malic acid	2.440801484	20130941
73	MTBLS36	(R)-3-hydroxy-2-oxo-4-phosphonooxybutanoic acid	9.296789157	21145142
74	MTBLS36	malonic acid	2.640487964	23511544

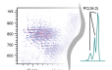
<http://predret.org/how-to-use-predret/>



Outline

A path to metabolite identification

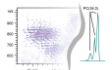
- Introduction
 - Analytical Technologies
 - Terminology in metabolite identification
- Metabolite identification
 - GC-MS MetID
 - EI spectral deconvolution
 - GC-APCI and exact mass
 - LC-MS MetID
 - Metabolite / Ion annotation
 - Formula calculation
 - Utilization of isotopes for formula calculation
 - MS/MS and RT
- New approaches for MetID
 - Ion mobility
 - *in silico* approaches
- *de novo* identification



Outline

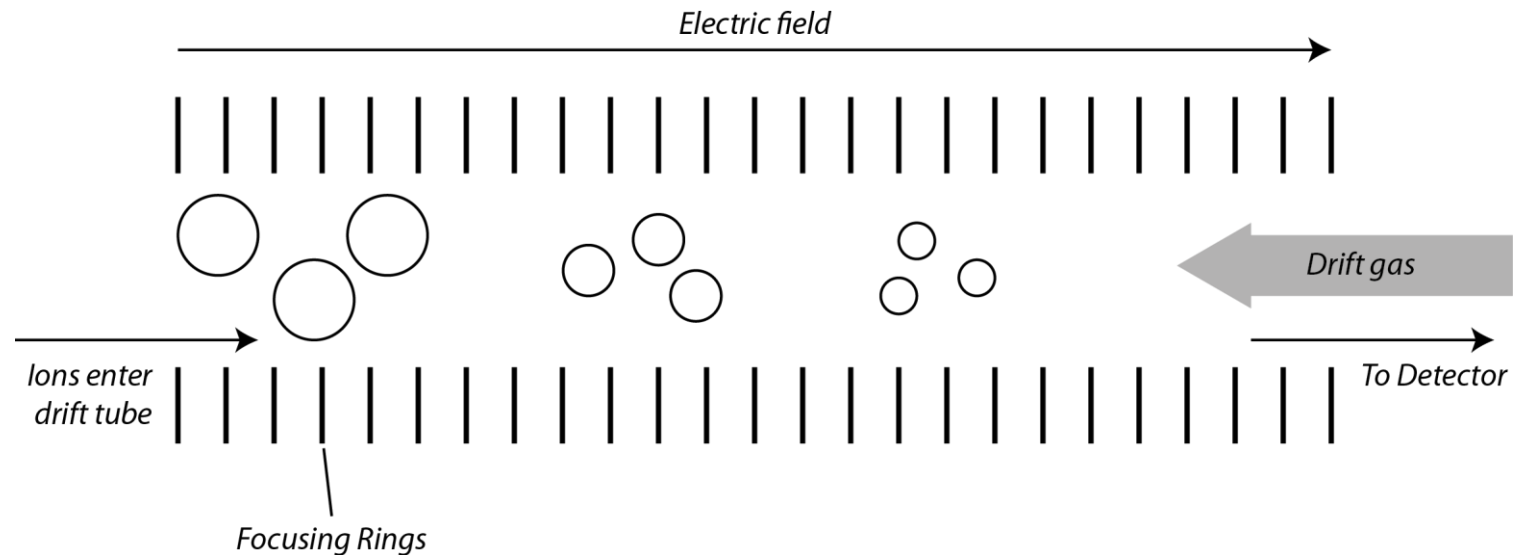
A path to metabolite identification

- Introduction
 - Analytical Technologies
 - Terminology in metabolite identification
- Metabolite identification
 - GC-MS MetID
 - EI spectral deconvolution
 - GC-APCI and exact mass
 - LC-MS MetID
 - Metabolite / Ion annotation
 - Formula calculation
 - Utilization of isotopes for formula calculation
 - MS/MS and RT
- New approaches for MetID
 - Ion mobility
 - *in silico* approaches
- *de novo* identification



New approaches for MetID

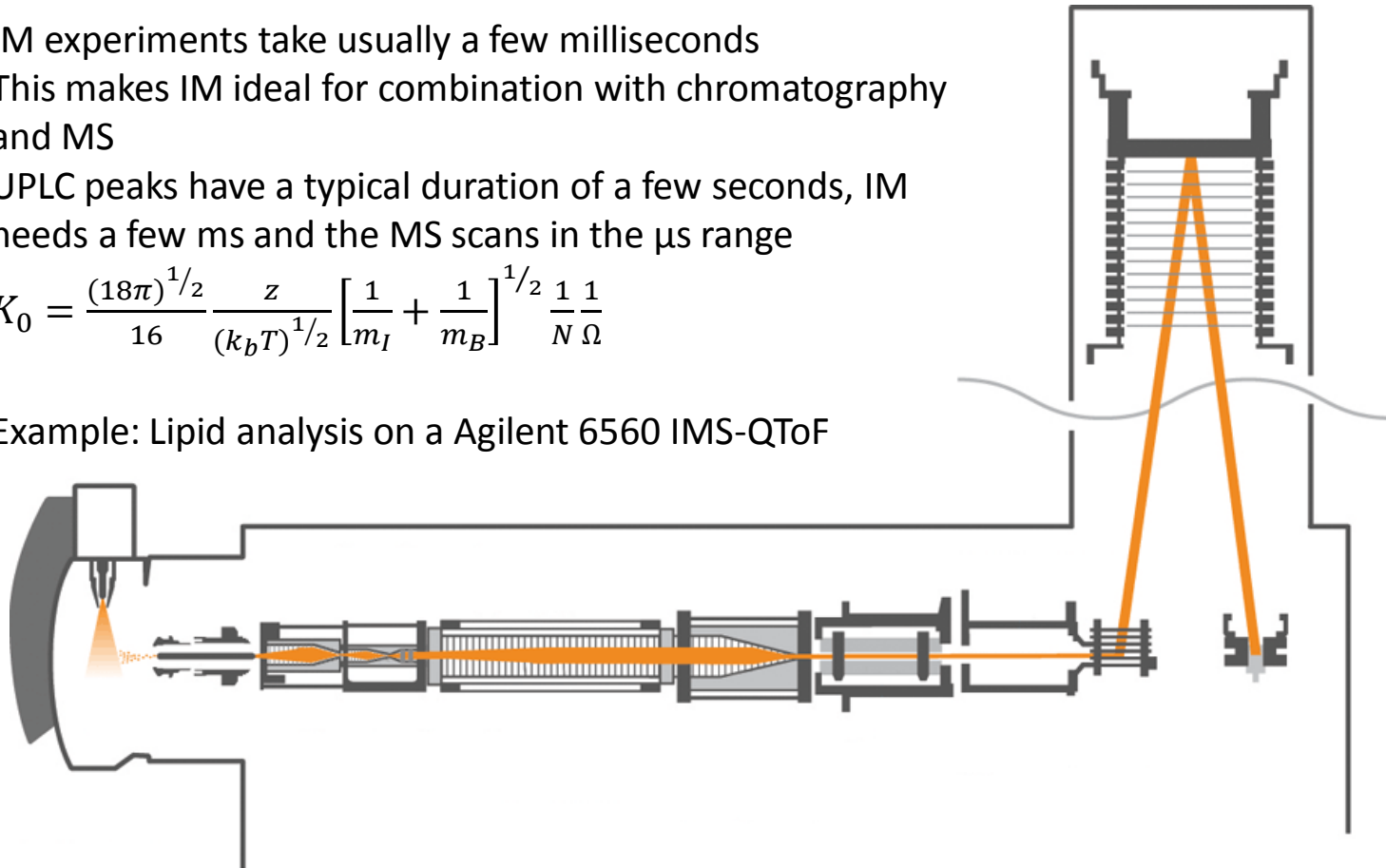
Ion mobility



New approaches for MetID

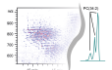
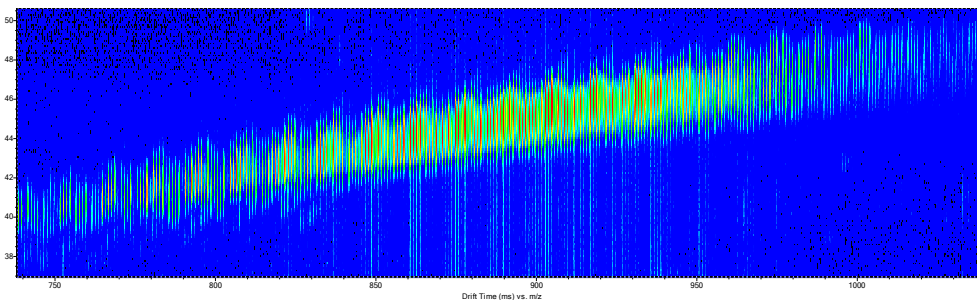
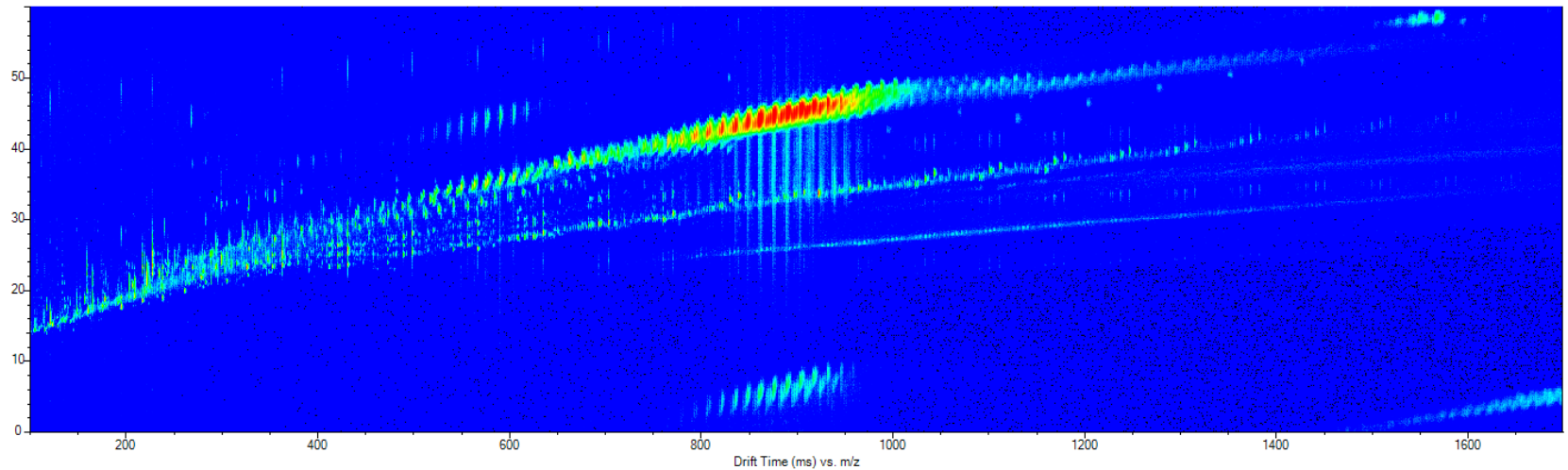
Ion mobility

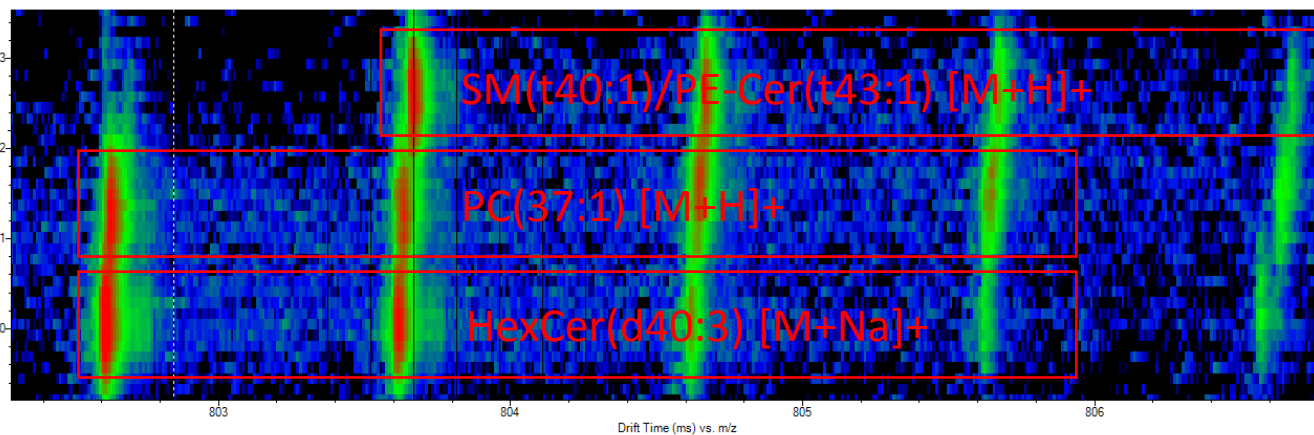
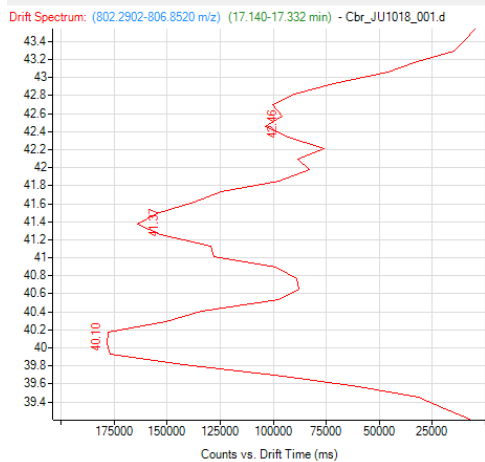
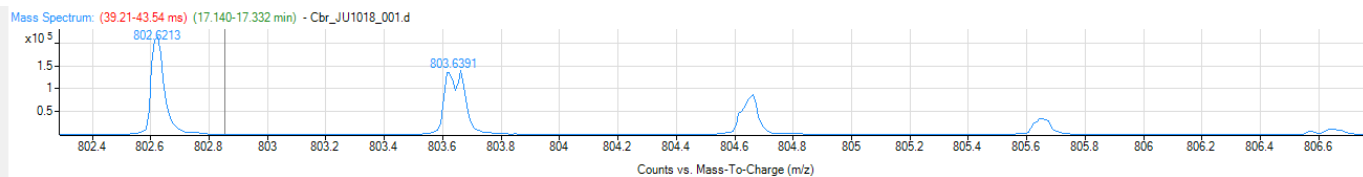
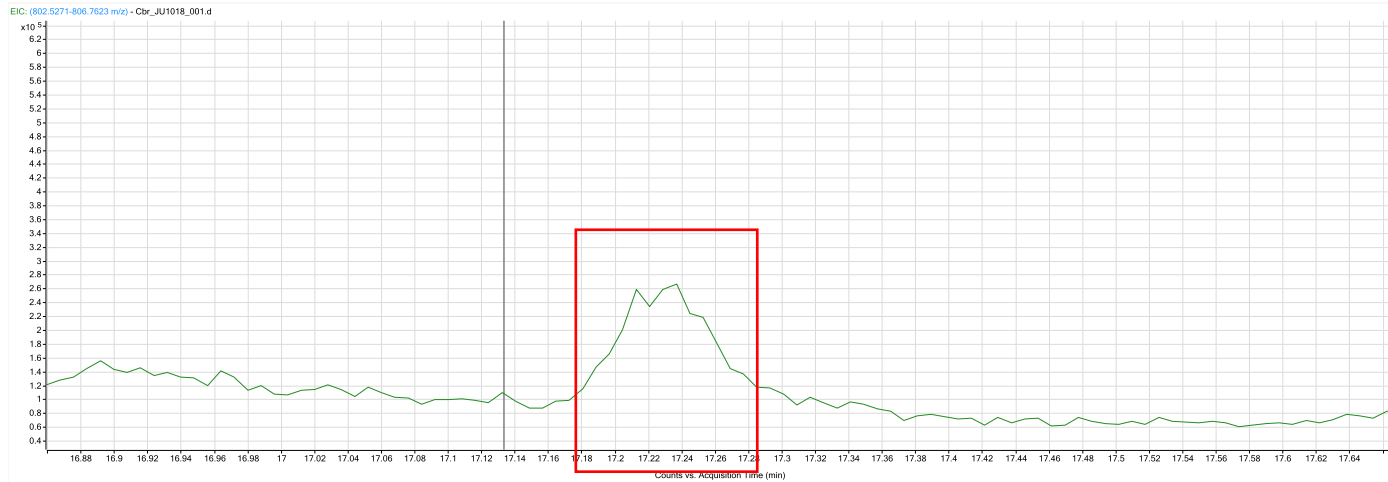
- IM experiments take usually a few milliseconds
- This makes IM ideal for combination with chromatography and MS
- UPLC peaks have a typical duration of a few seconds, IM needs a few ms and the MS scans in the μs range
- $$K_0 = \frac{(18\pi)^{1/2}}{16} \frac{z}{(k_b T)^{1/2}} \left[\frac{1}{m_I} + \frac{1}{m_B} \right]^{1/2} \frac{1}{N} \frac{1}{\Omega}$$
- Example: Lipid analysis on a Agilent 6560 IMS-QToF



New approaches for MetID

Ion mobility

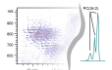




Outline

A path to metabolite identification

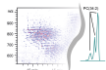
- Introduction
 - Analytical Technologies
 - Terminology in metabolite identification
- Metabolite identification
 - GC-MS MetID
 - EI spectral deconvolution
 - GC-APCI and exact mass
 - LC-MS MetID
 - Metabolite / Ion annotation
 - Formula calculation
 - Utilization of isotopes for formula calculation
 - MS/MS and RT
- New approaches for MetID
 - Ion mobility
 - *in silico* approaches
- *de novo* identification



New approaches for MetID

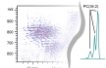
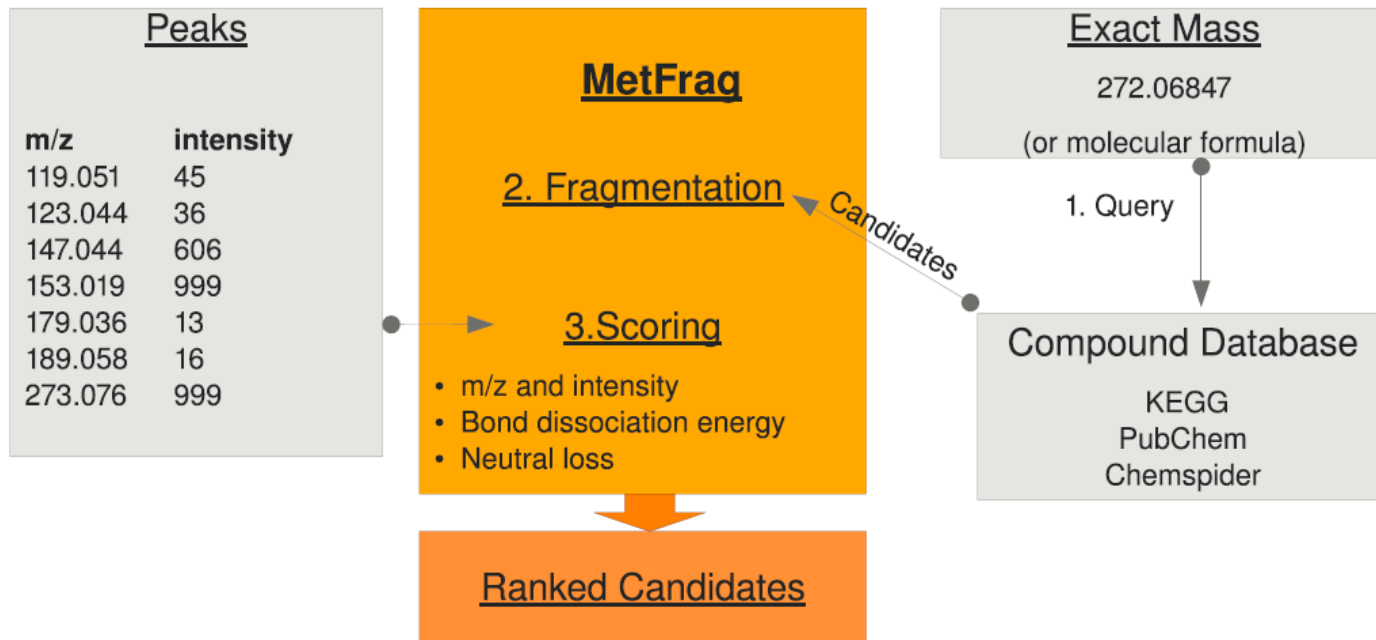
in silico approaches

- *in silico* approaches do not rely on previously acquired MS/MS for the exact molecule, but either matches predicted fragments from simple bond breaking or rule based fragmentation
- A selection of *in silico* tandem solutions are
 - MetFrag / MetFusion (IPB Halle)
 - CSI:FingerID (University of Jena)
 - CFM-ID (Wishart Lab)
 - MAGMa (Netherlands Metabolomics Center)

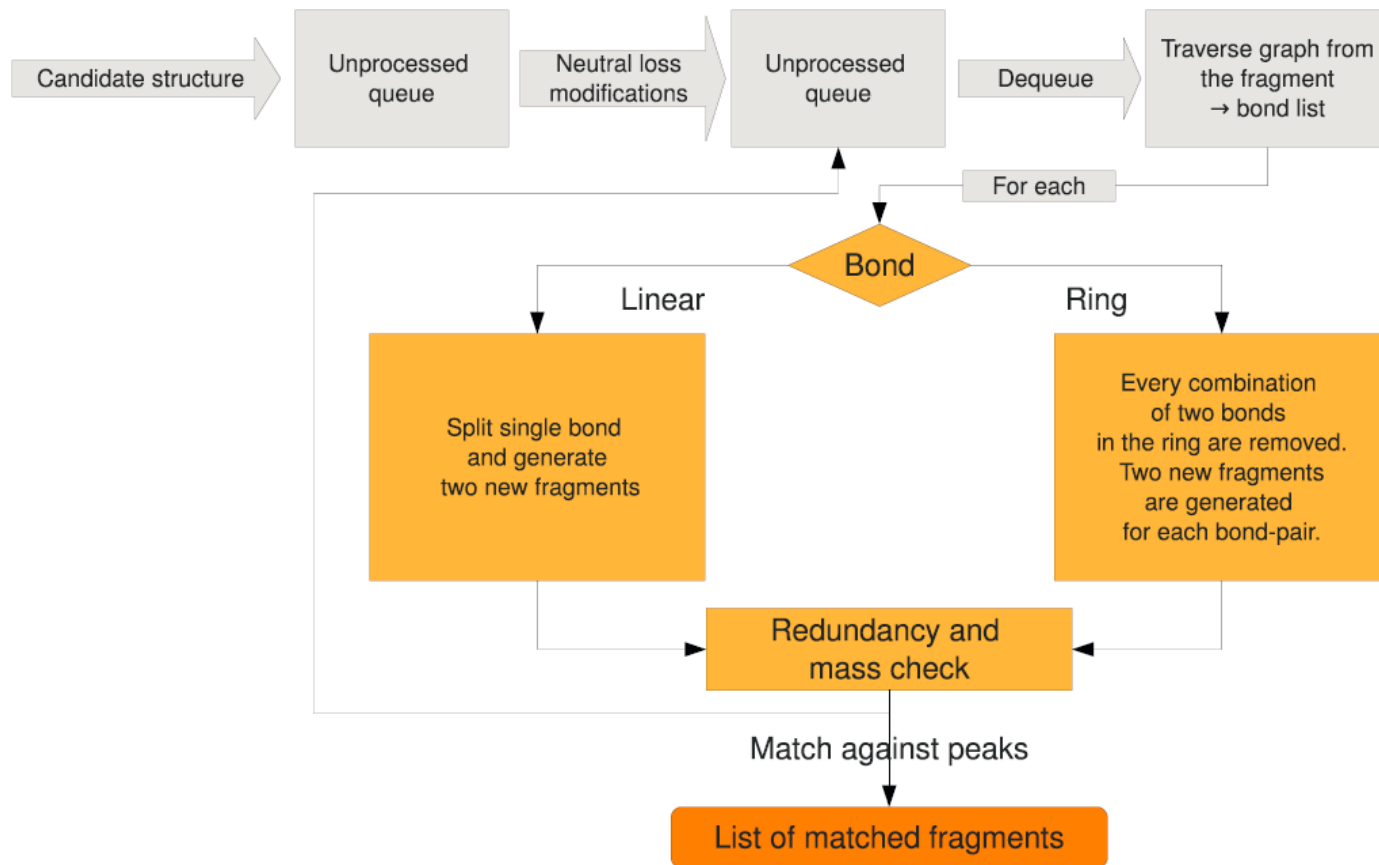


New approaches for MetID

in silico approaches - MetFrag



New approaches for MetID *in silico* approaches - MetFrag



New approaches for MetID *in silico* approaches - MetFrag

MetFrag - Match Predicted Fragments with Mass Spectra

Peaks: 116.051, 467.616, 465, 123.044, 370.653, 265, 147.044, 6076.145, 6066, 153.019, 10000.0, 899, 179.036, 141, 192, 13, 186.058, 176, 258, 16, 273.076, 10000.000, 566, 274.083, 318.003, 30

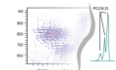
Database Settings
 Neutral exact mass: 272.05847 Search PPM: 30
 Molecular formula (optional alternatives):
 Database: EGG PubChem ChemSpider
 Only biological compounds:
 Limit # of structures: 100
 Database ID's:
MetFrag Settings
 Mode: [M+H]⁺ [M-H]⁻ (Alpha)
 Mz tolerance (e.g. 0.01): 0.01
 Mz range (e.g. 10): 10

Structure **Database ID** **Actions**

	C00502	Show Frags
	C16232	Show Frags

Fragmentation
 Score: 1.0
 Mz: 272.05847 (original compound), 153.019 (C₉H₈O₄), 179.036 (C₉H₈O₄)
 Mz: 186.058 (C₉H₈O₄)
 Mz: 116.051 (C₉H₈O₄)

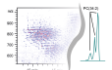
Wolf et al. BMC Bioinformatics 2010, 11:148



Outline

A path to metabolite identification

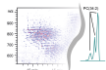
- Introduction
 - Analytical Technologies
 - Terminology in metabolite identification
- Metabolite identification
 - GC-MS MetID
 - EI spectral deconvolution
 - GC-APCI and exact mass
 - LC-MS MetID
 - Metabolite / Ion annotation
 - Formula calculation
 - Utilization of isotopes for formula calculation
 - MS/MS and RT
- New approaches for MetID
 - Ion mobility
 - *in silico* approaches
- *de novo* identification



Metabolite Identification

de novo identification of unknowns

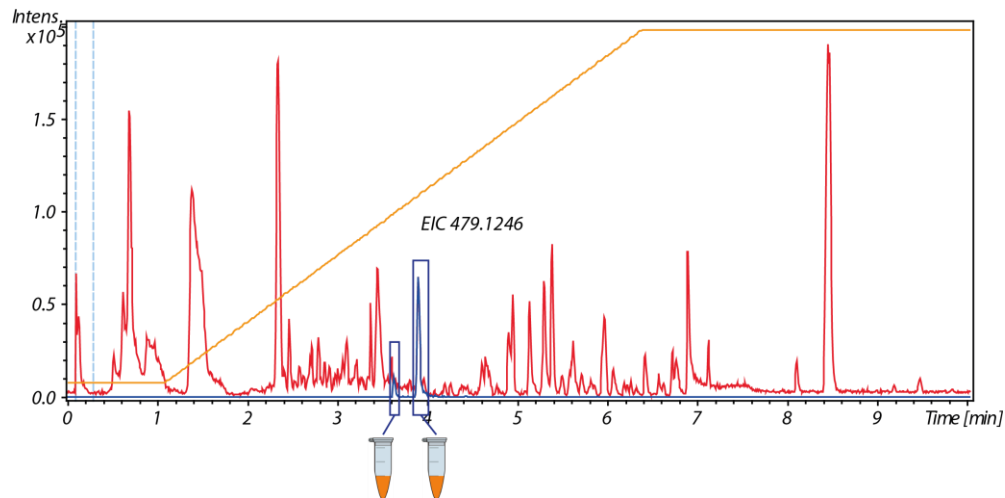
- Complete *de novo* identification is a tedious process and involves several steps
- *de novo* identification should be only conducted if metabolite cannot be identified by any other means
- Complete *de novo* identification requires
 - Full structural elucidation
 - Synthesis of a pure reference substance



de novo identification

Fractionation strategies

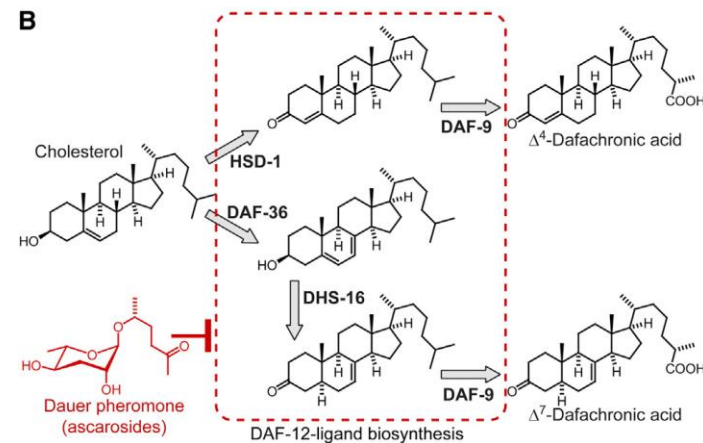
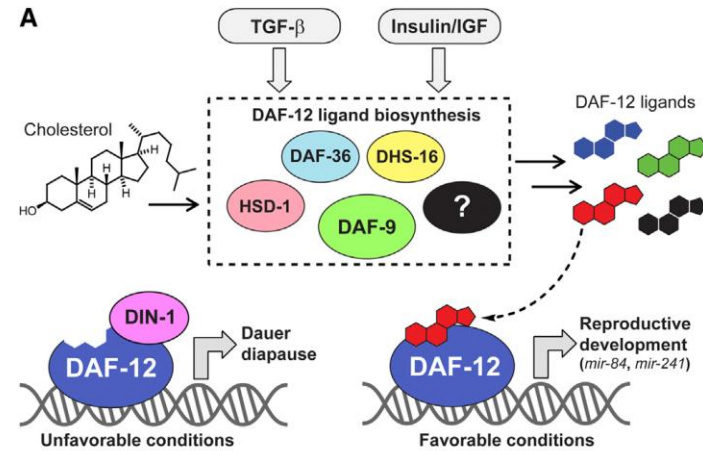
- Enough biomass available?
- Potential activity of metabolite, e.g. hormone?
- Test system available?
- Simple or activity guided fractionation



de novo identification

Fractionation strategies

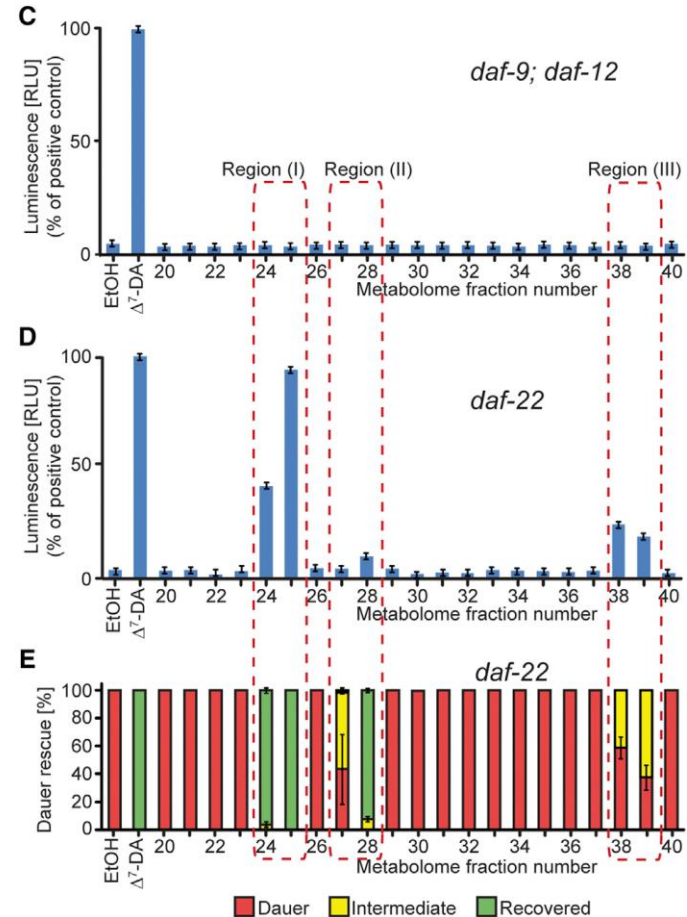
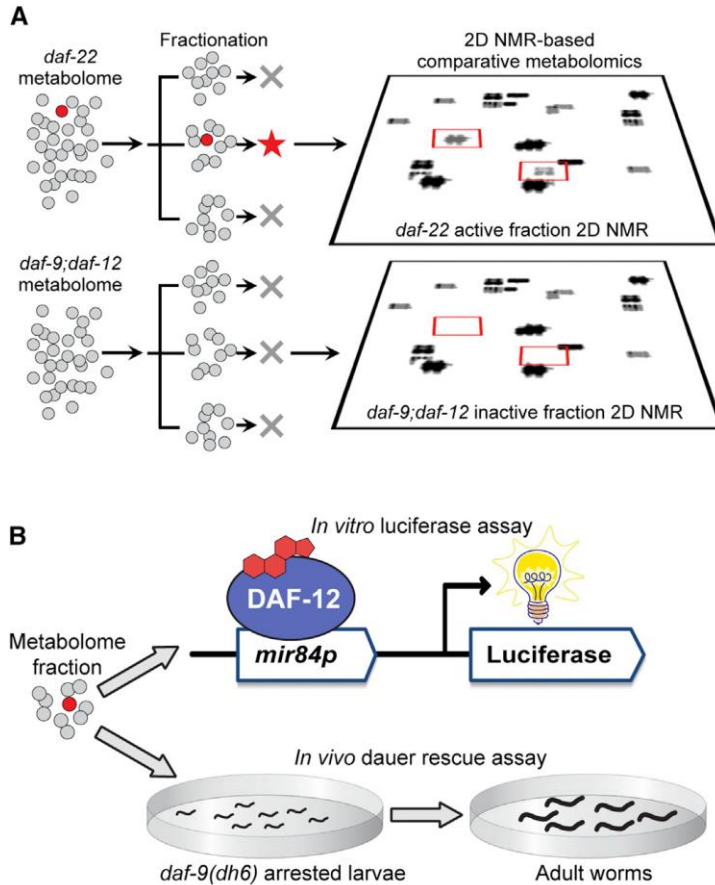
- Mahanti et al. searched for novel ligands of the nuclear hormone receptor DAF-12 in *C. elegans*
- Activity guided fractionation was used in combination with 2D-NMR based metabolomics to find bioactive molecules
- Activity was checked *in vivo* and *in vitro*
- Previously known molecules Δ^4 - and Δ^7 -Dafachronic acid were re-identified together with novel variants
- Synthesis of chemical reference standards confirmed identity



Mahanti et al., Cell Metabolism 19, 73–83, January 7, 2014

de novo identification

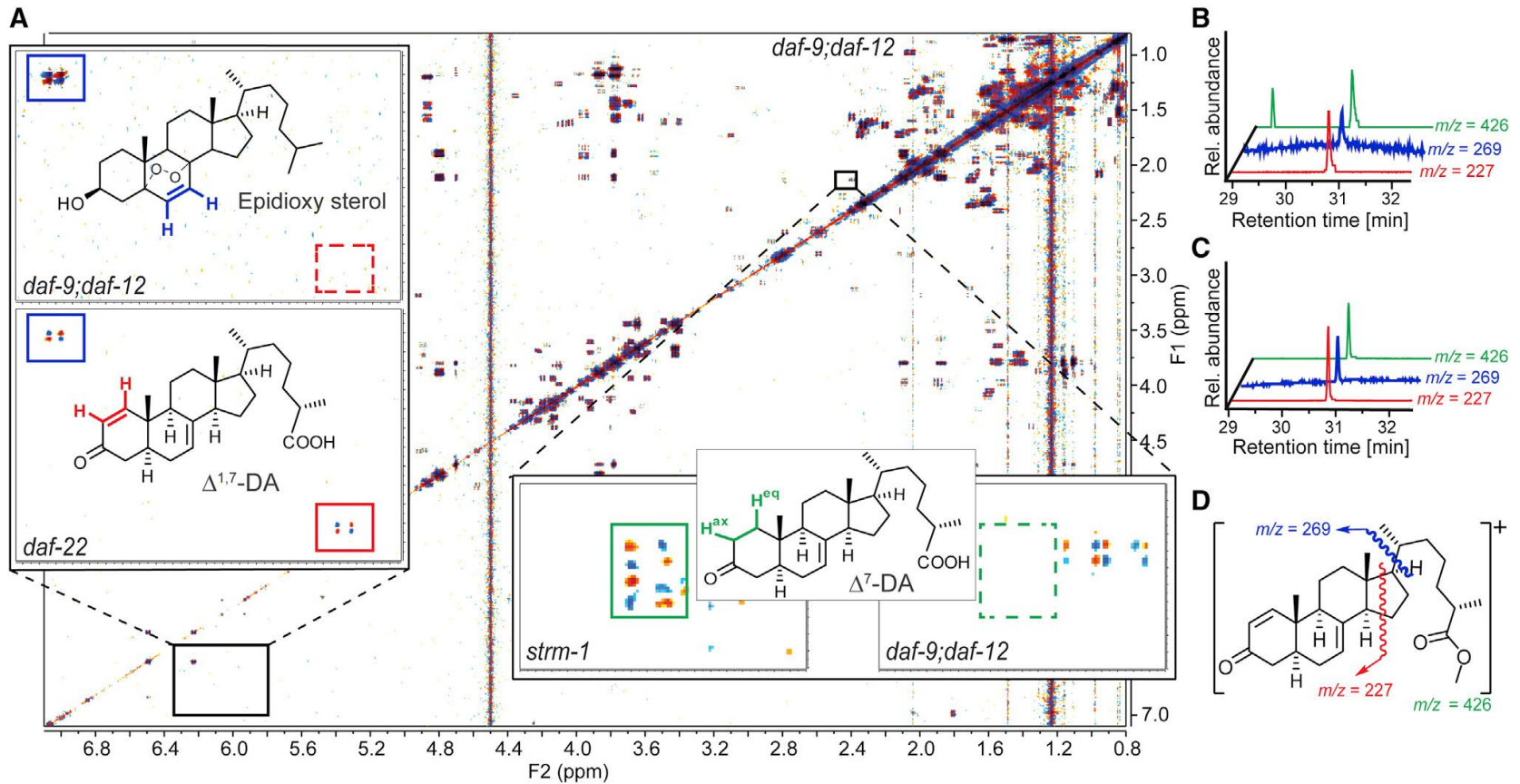
Fractionation strategies



Mahanti et al., Cell Metabolism 19, 73–83, January 7, 2014

de novo identification

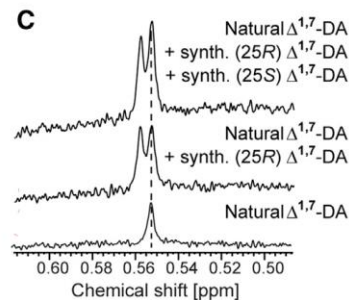
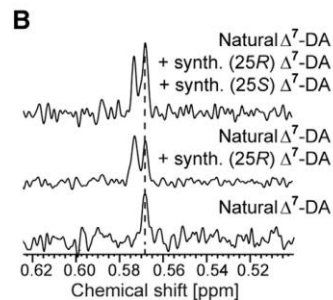
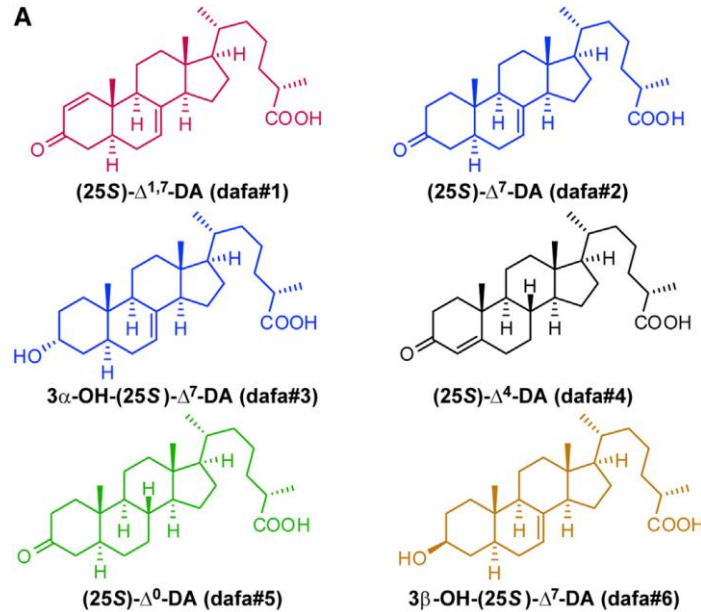
Fractionation strategies



Mahanti et al., Cell Metabolism 19, 73–83, January 7, 2014

de novo identification

Fractionation strategies



Mahanti et al., Cell Metabolism 19, 73–83, January 7, 2014

de novo identification

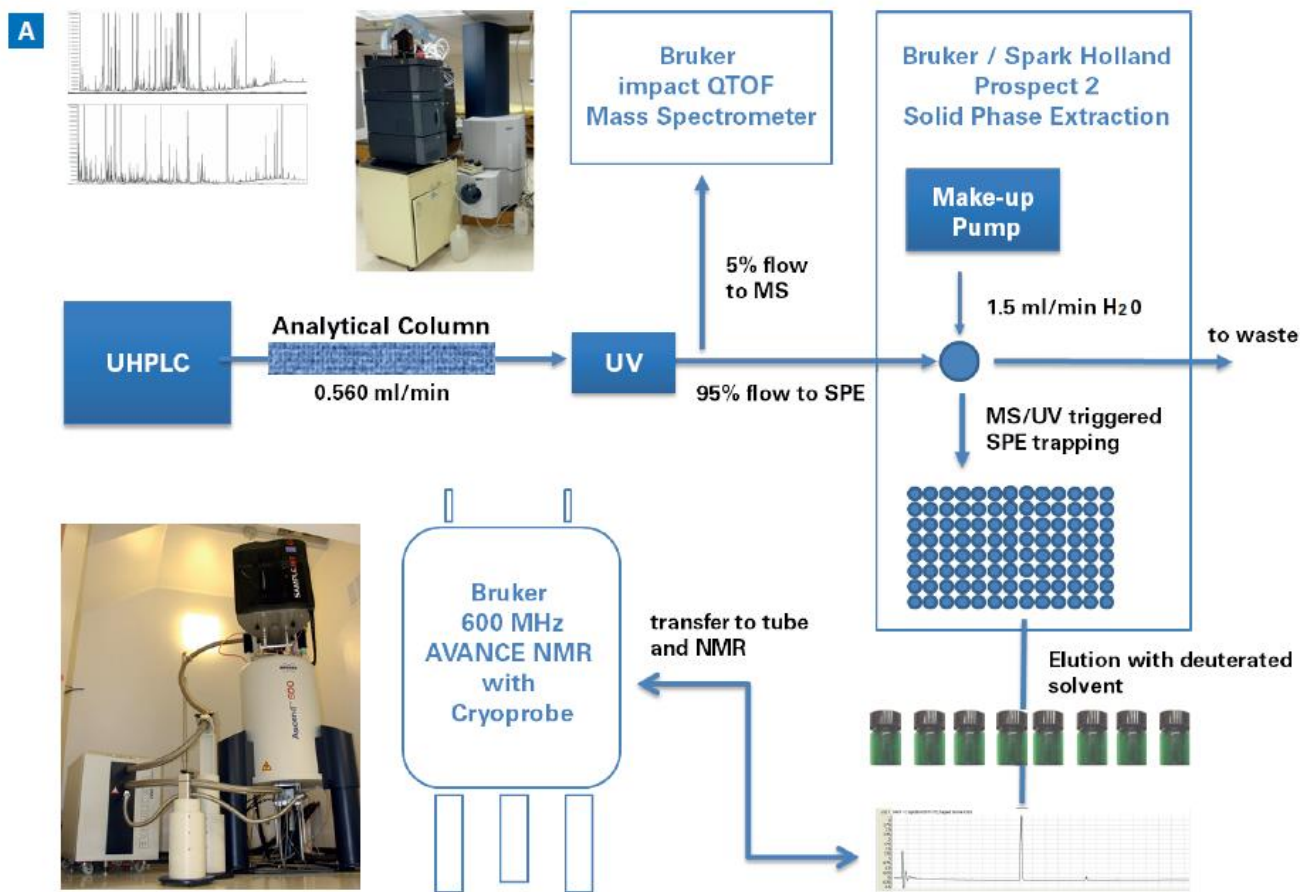
Fractionation strategies



Bruker Metabolic Profiler

de novo identification

Fractionation strategies



Sumner et al. & Bruker

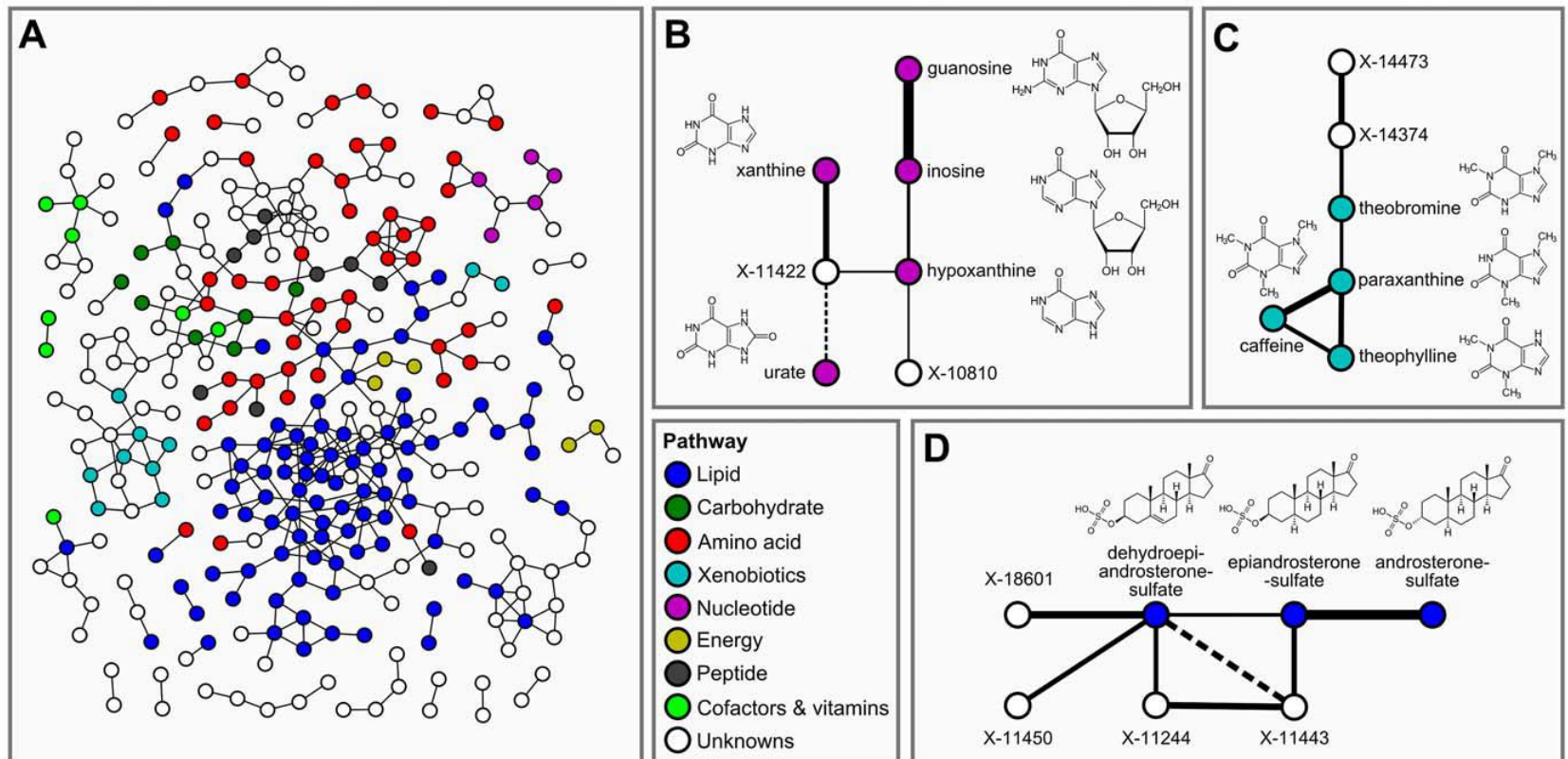
de novo identification

Fractionation strategies

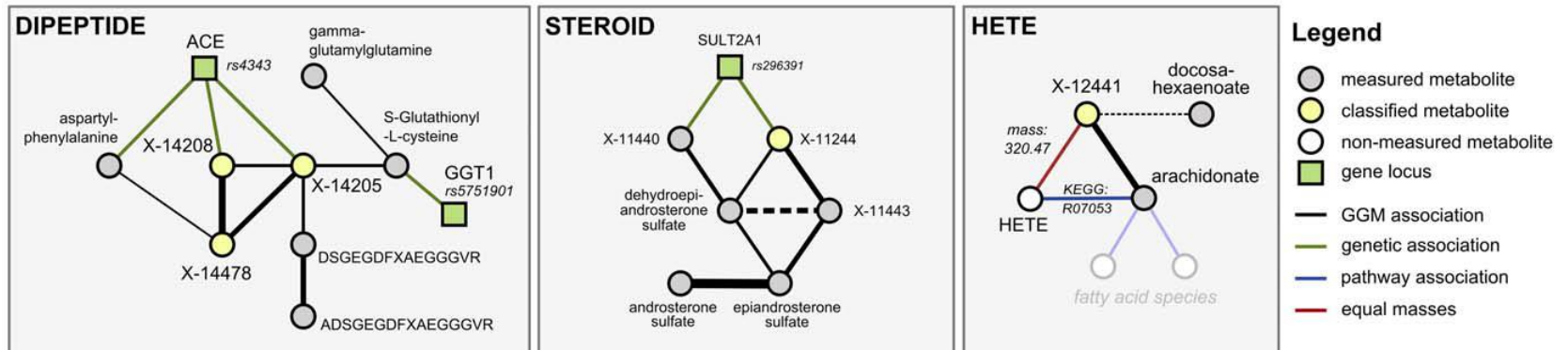
B



de novo identification *in silico* strategies



de novo identification in silico strategies



de novo identification Reporting

- Please report your de novo identified metabolite in (electronically) searchable manner

The screenshot shows the ChEBI search interface. The search bar contains 'dafachronic acid'. Below the search bar, there are navigation links: Home, Advanced Search, Browse, Documentation, Download, Tools, About ChEBI. The search results page displays 'Sie suchten nach All in ChEBI' and 'dafachronic acid in All'. It indicates that more than 18,139 results have been returned but are limited to a maximum of 1,000. The results are displayed in a grid of 9 entries, each with a chemical structure, name, CHEBI ID, stars, formula, mass, and charge.

Name	CHEBI ID	Stars	Formula	Mass	Charge
Δ^1, Δ^7 -dafachronic acid	CHEBI:83137	★★★	C ₂₇ H ₄₂ O ₃	414.62060	0
(25S)- Δ^7 -dafachronic acid	CHEBI:71556	★★★	C ₂₇ H ₄₂ O ₃	414.62060	0
(25S)- Δ^4 -dafachronic acid	CHEBI:71560	★★★	C ₂₇ H ₄₂ O ₃	414.62060	0
Δ^7 -dafachronic acid	CHEBI:78699	★★★	C ₂₇ H ₄₂ O ₃	414.62060	0
Δ^4 -dafachronic acid	CHEBI:78686	★★★	C ₂₇ H ₄₂ O ₃	414.62060	0
stearic acid	CHEBI:28842				
(25R)- Δ^4 -dafachronic acid	CHEBI:71613				

Hamburger to Cow algorithm or "Wishful Thinking"
Requires Jurassic Park Technology



http://fiehnlab.ucdavis.edu/projects/rice_metabolome/hamburger-to-cow-algorithm.png

Thank you! Questions?



PFP Postdoctoral Fellowship Program

DAAD

DFG Deutsche
Forschungsgemeinschaft



Habilitationsfachmentorat:



HelmholtzZentrum münchen
Deutsches Forschungszentrum für Gesundheit und Umwelt

TUM
Technische Universität München

