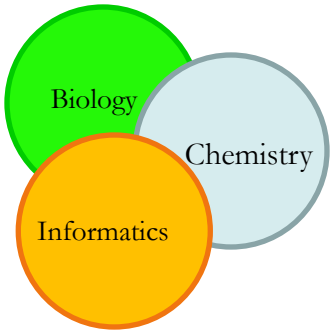


Welcome!

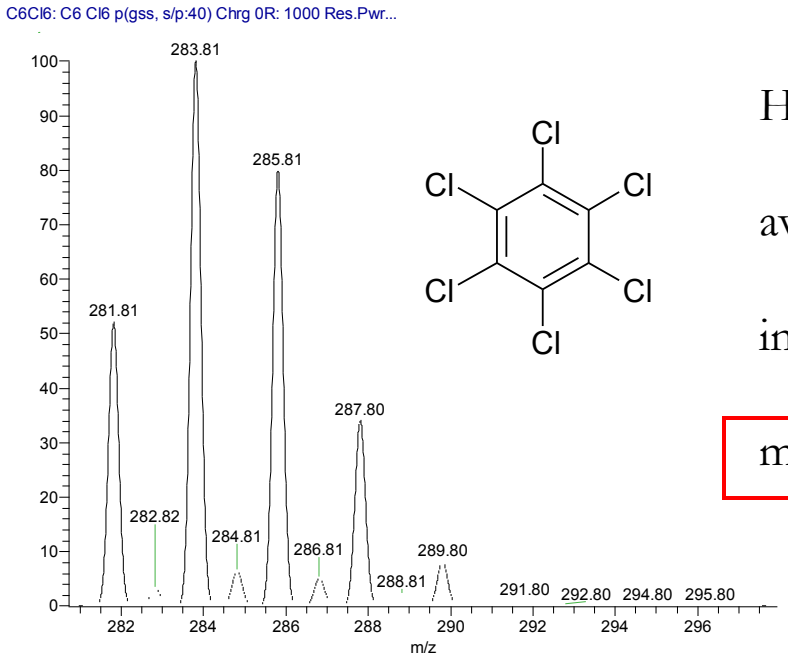
Mass Spectrometry meets Cheminformatics
WCMC Metabolomics Course 2013
Tobias Kind

Course 4: Mass Spectrometry Tools & Concepts

http://fiehnlab.ucdavis.edu/staff/kind

Atomic Mass

Correct unit is [**u**] – unified atomic mass unit or [**Da**] Dalton see SI units
1 u = 1 Da = 1/12th of mass of carbon ^{12}C = 1.66053886 x 10^{-27} kg



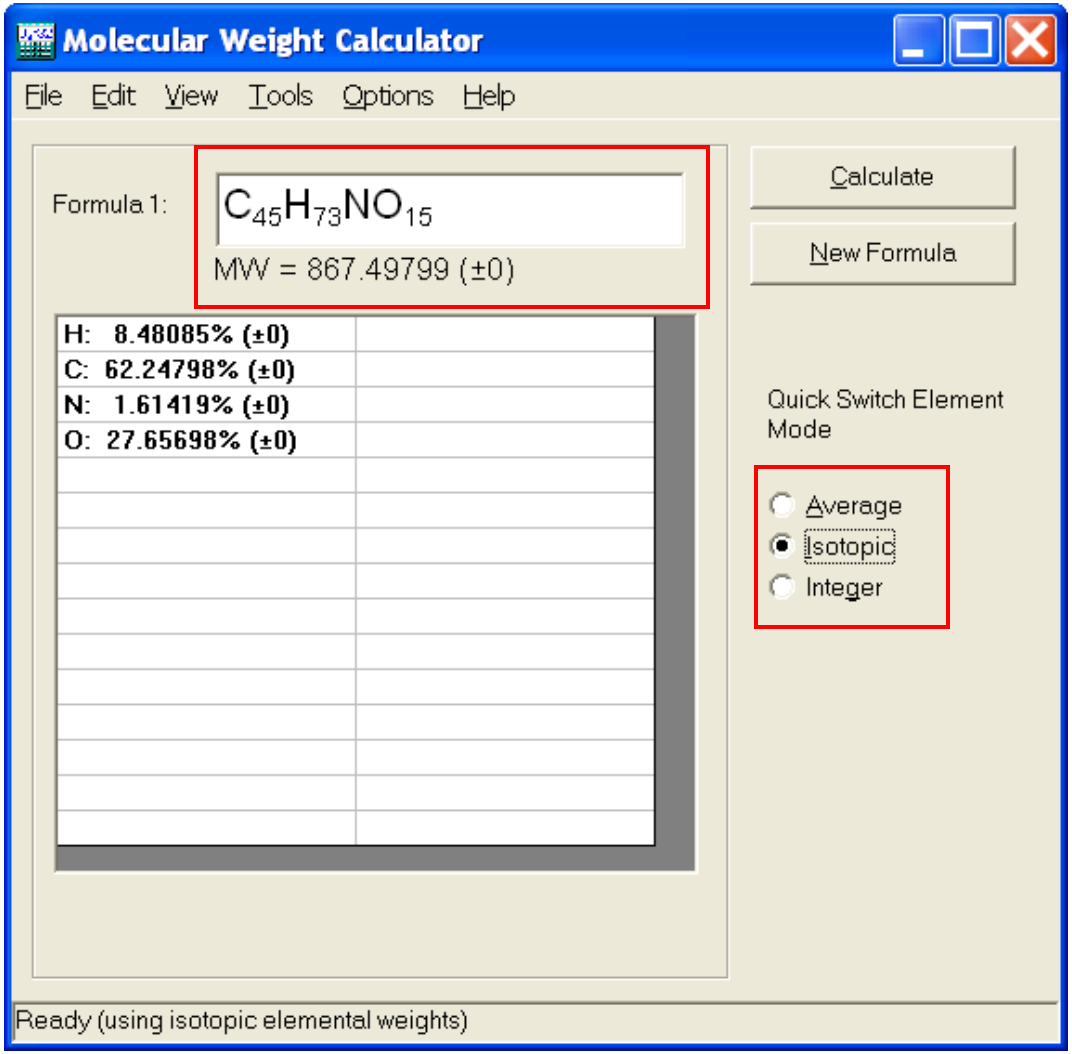
Hexachlorobenzene (C_6Cl_6)

average mass	-	284.7804 u
integer mass	-	282.0 u
monoisotopic mass	-	281.81312 u

Always (always) check molecular masses obtained from databases or publications.
For mass spectrometry the **monoisotopic mass** is used.

InChIKey:CKAPSXZOOQJIBF-UHFFFAOYAV

MWTWIN



Example: Molecular Weight Calculator (Matthew Monroe / PNNL)

Mass Accuracy

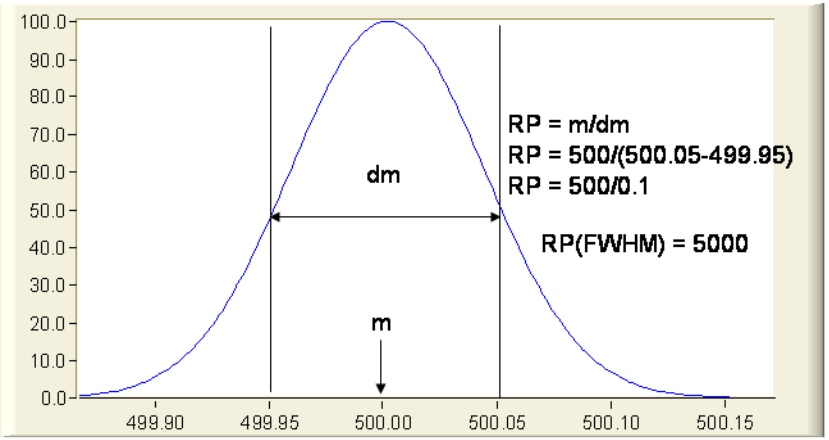
Instruments must be calibrated to obtain high mass accuracy.
In case of FT-ICR-MS mass calibration can be stable over weeks.
Post- mass calibration can be performed if calibrant was run with samples.
Mass of electron becomes important at around 500 Da.

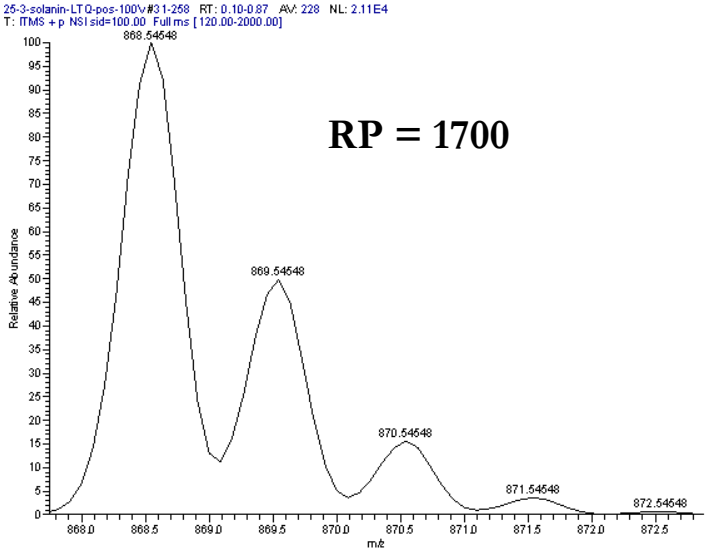
Type	Mass Accuracy
FT-ICR-MS	0.1 - 1 ppm
Orbitrap	0.5 - 1 ppm
Magnetic Sector	1 - 2 ppm
TOF-MS	3 - 5 ppm
Q-TOF	1 - 5 ppm
Triple Quad	3 - 5 ppm
Linear IonTrap	50-200 ppm (10 ppm in Ultra-Zoom)

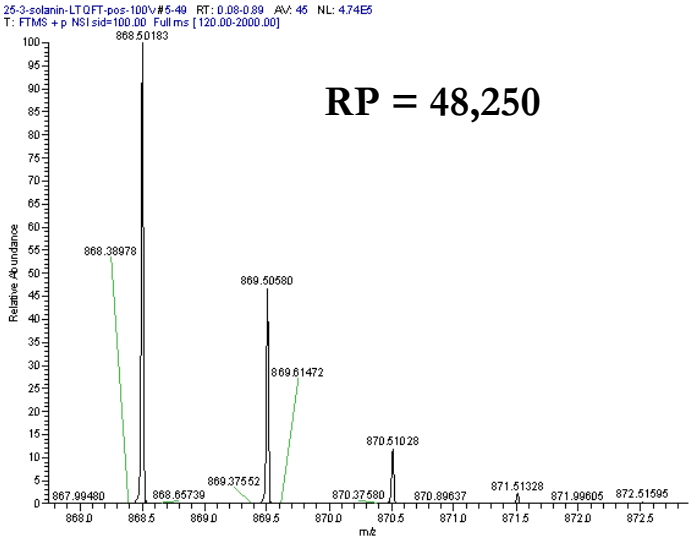
$$ppm = \left(\frac{m_{exp} - m_{calc}}{m_{exp}}\right) * 1E+6$$

m(e-) = 0.00054857990924 u = mass of electron
m(1H+) = 1.00727645199076 u = mass of proton
m(1H) = 1.0078250319

Resolving Power







High resolving power is helpful for separation of species with almost same mass (**isobars**).

High resolving power can **not** be used to distinguish between **structural isomers**.

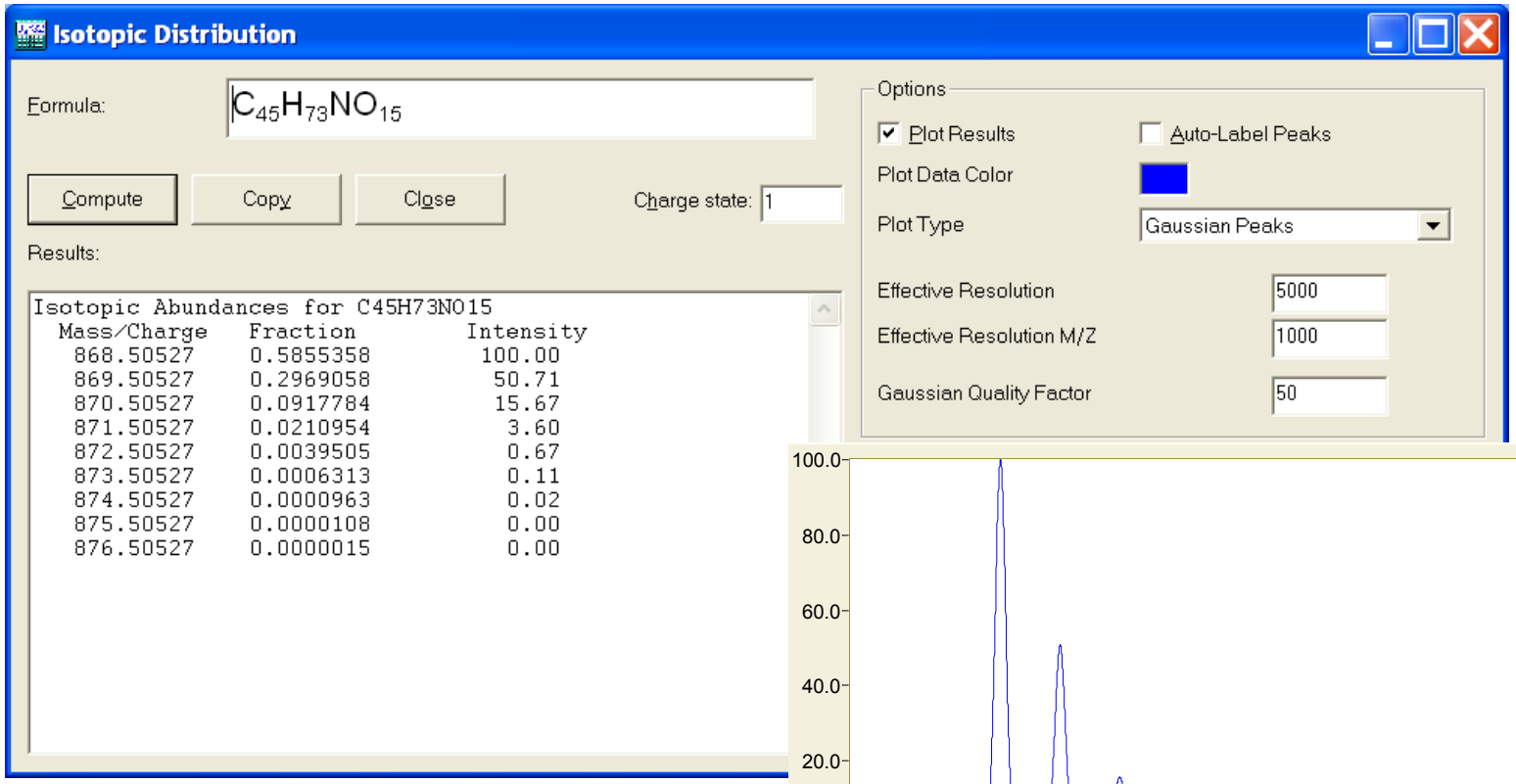
Example:
$C_8H_{10}N_2O$ has 100,082,479 isomers.

Example Solanine (CID=30185)

Isotopic Pattern Generators

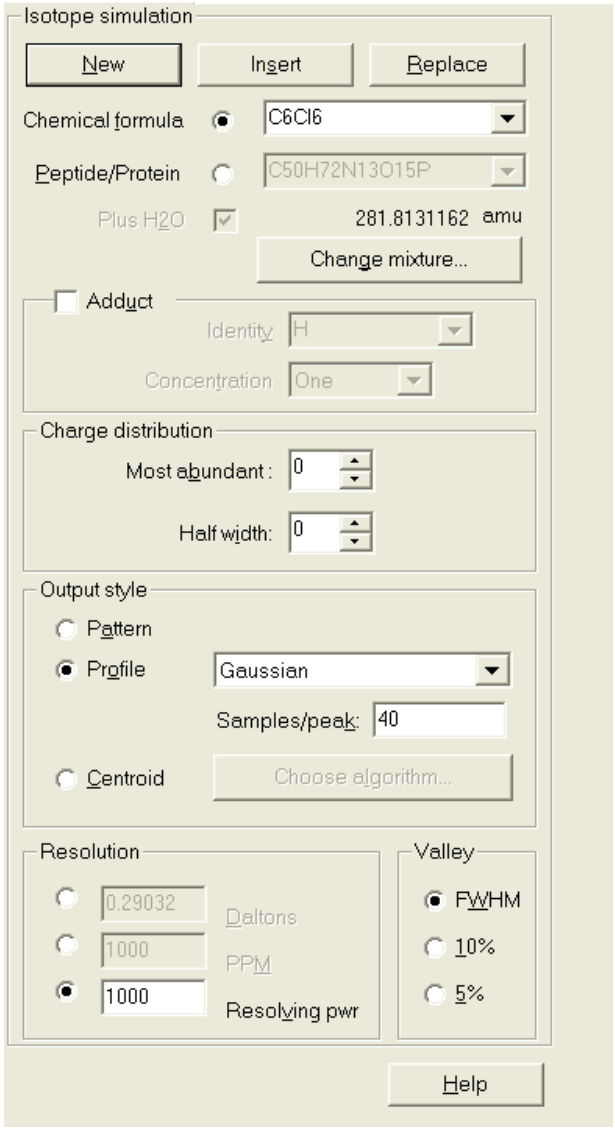
Elements can be a) **monoisotopic** (F, Na, P, I)
b) **polyisotopic** (H, C, N, O, S, Cl, Br)

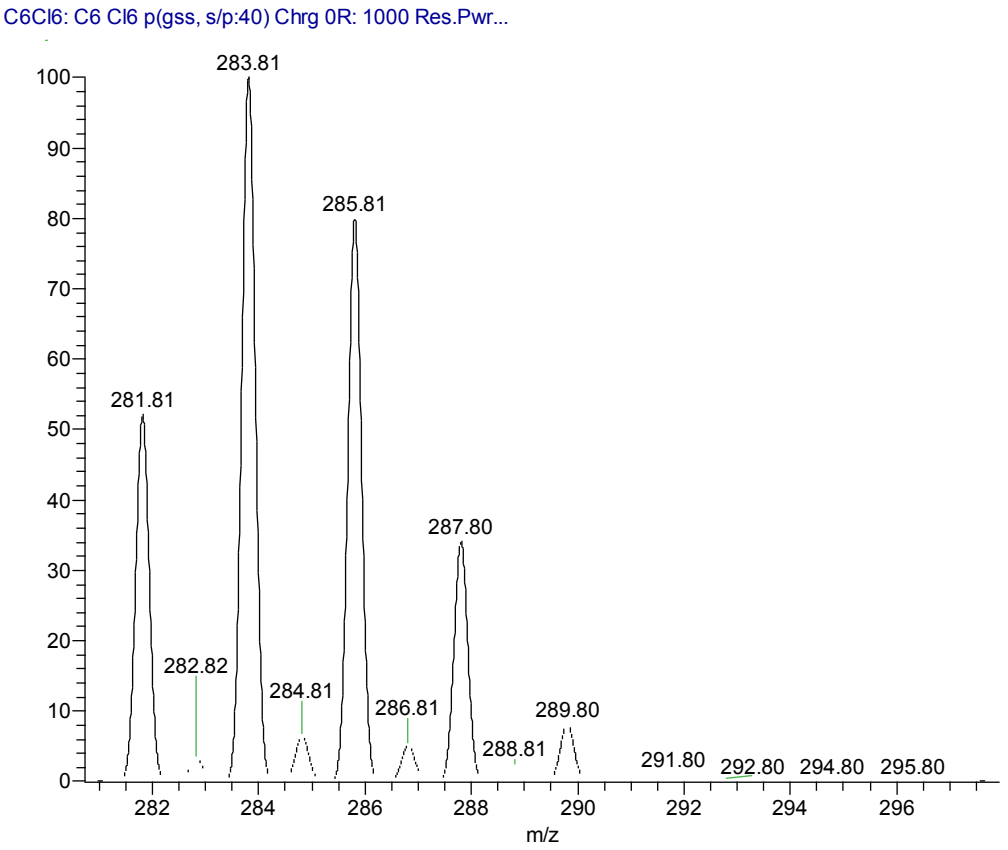
Isotopic pattern generators generate the isotopic abundances for a given mass value. Calculation is very time-consuming and based on Fast Fourier algorithms.



Example: Molecular Weight Calculator (Matthew Monroe / PNNL)

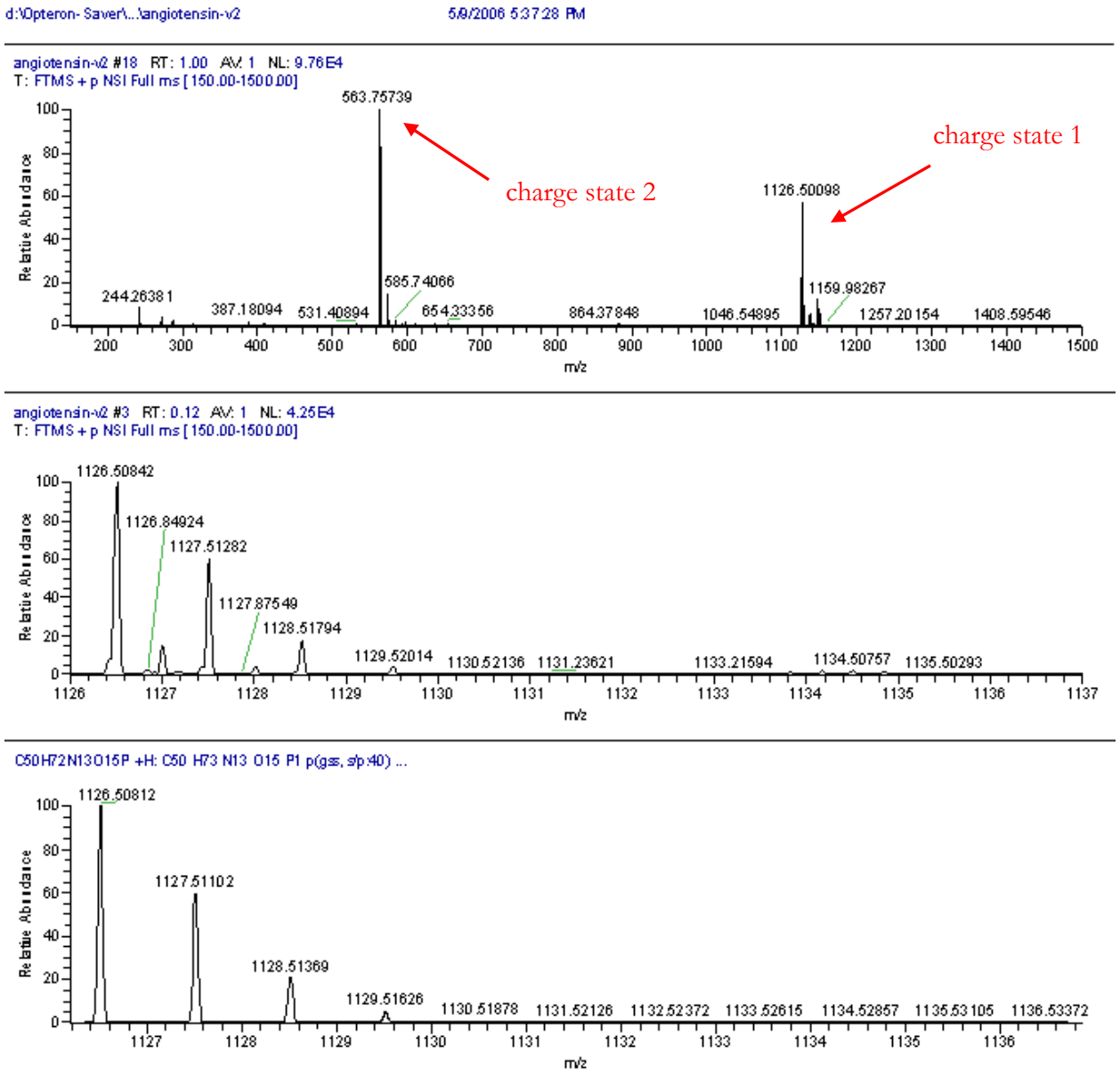
Isotopic pattern generators

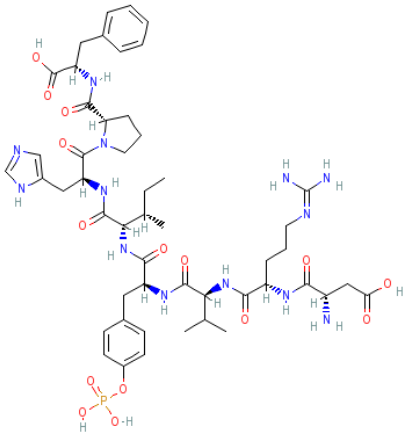




Example form Thermo Xcalibur with a very versatile isotopic pattern generator

Charge states



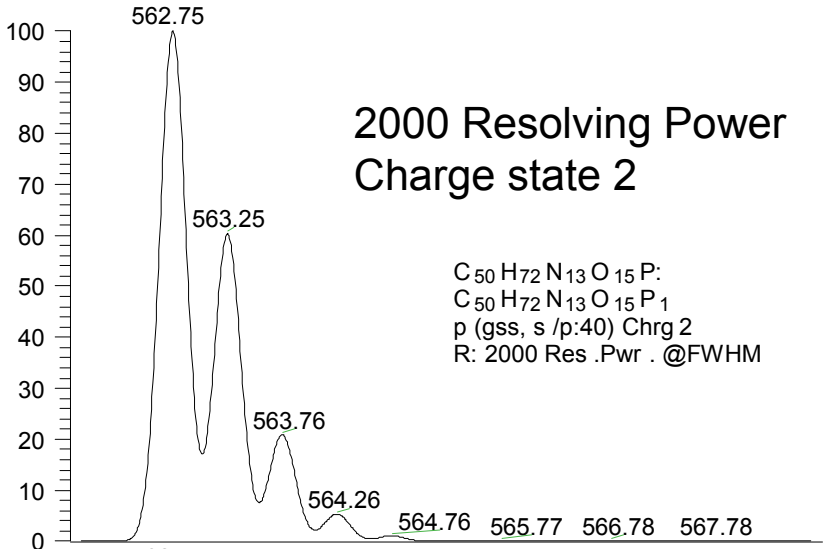


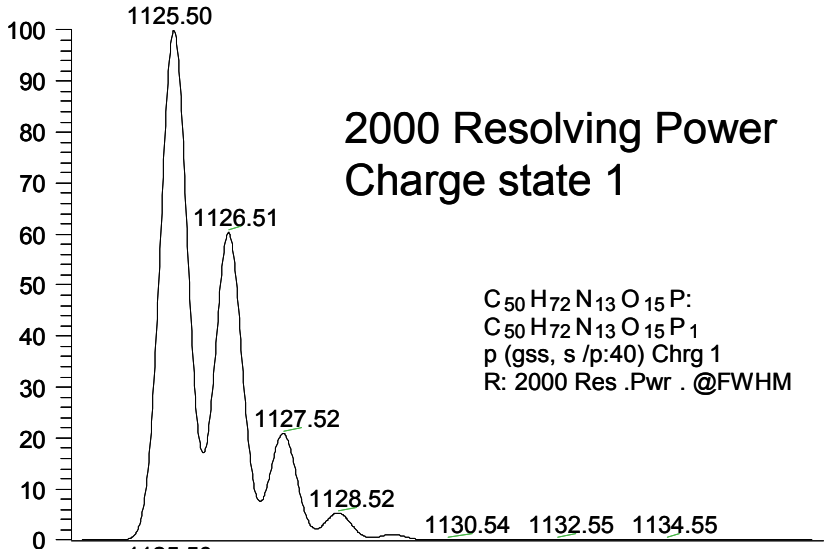
CID: [3081765]

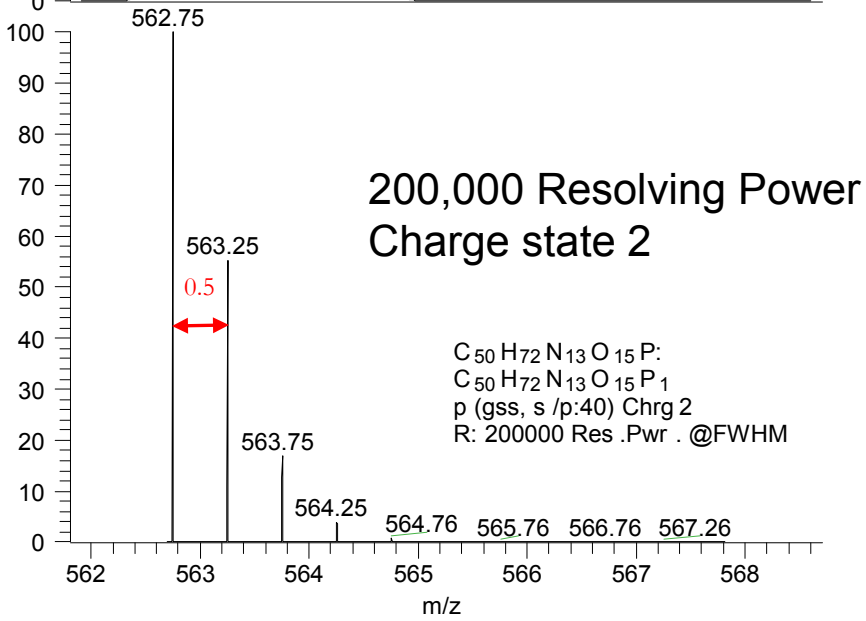
MW = 1125.50082

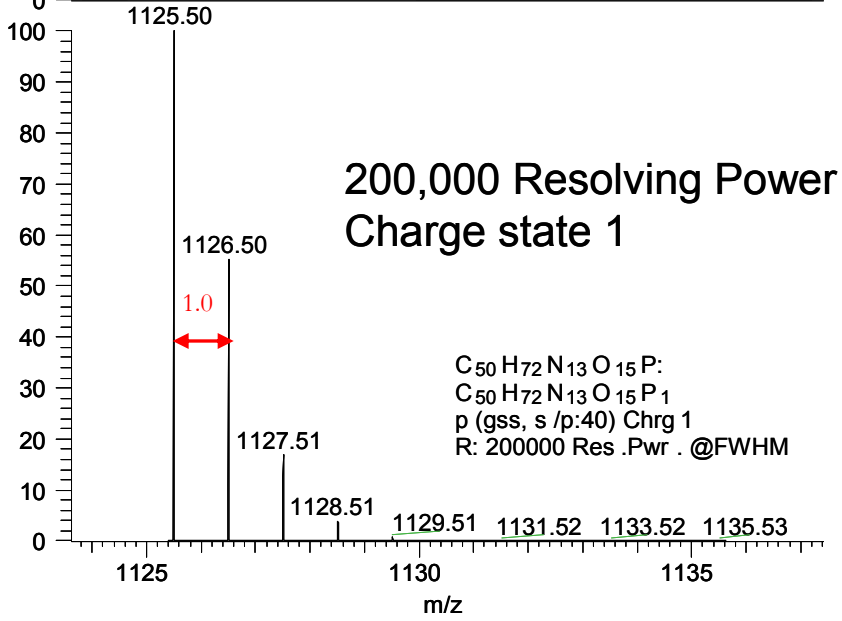
C50H72N13O15P

Different charge states and peak resolutions



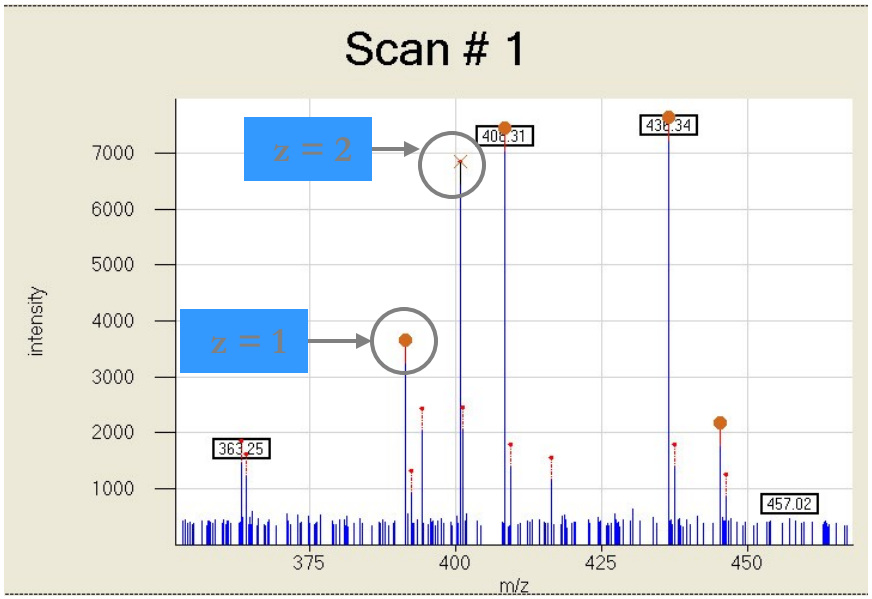


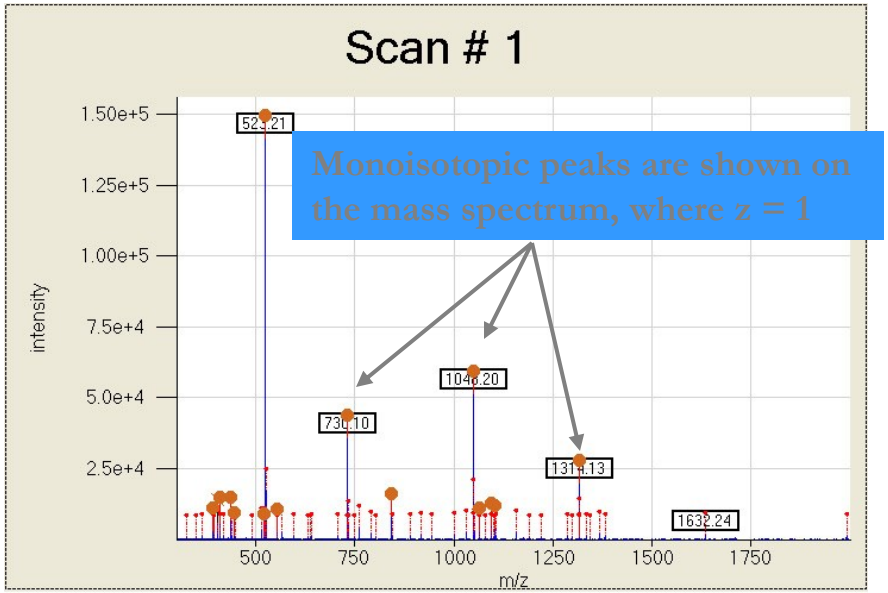


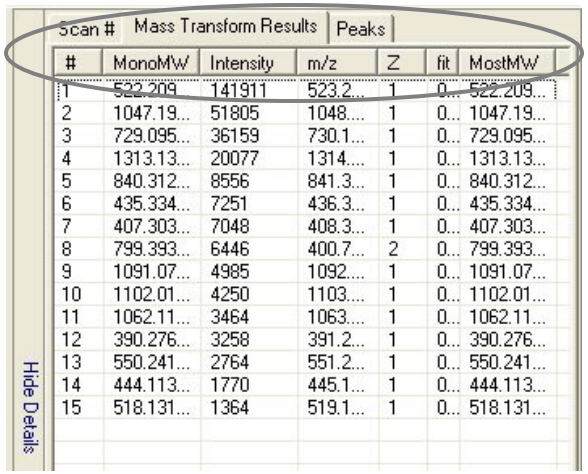


Example of Phosphorylated Angiotensin isotopic pattern without adduct $[M+H]^+$ simulated by Thermo XCalibur

Charge state deconvolution







Scan # | Mass Transform Results | Peaks

Hide Details

#	MonoMW	Intensity	m/z	Z	fit	MostMW
1	522.209...	141911	523.2...	1	0...	522.209...
2	1047.19...	51805	1048....	1	0...	1047.19...
3	729.095...	36159	730.1...	1	0...	729.095...
4	1313.13...	20077	1314....	1	0...	1313.13...
5	840.312...	8556	841.3...	1	0...	840.312...
6	435.334...	7251	436.3...	1	0...	435.334...
7	407.303...	7048	408.3...	1	0...	407.303...
8	799.393...	6446	400.7...	2	0...	799.393...
9	1091.07...	4985	1092....	1	0...	1091.07...
10	1102.01...	4250	1103....	1	0...	1102.01...
11	1062.11...	3464	1063....	1	0...	1062.11...
12	390.276...	3258	391.2...	1	0...	390.276...
13	550.241...	2764	551.2...	1	0...	550.241...
14	444.113...	1770	445.1...	1	0...	444.113...
15	518.131...	1364	519.1...	1	0...	518.131...

Example from **Decon2Ls** Tutorial (PNNL)
A Software Tool for Deconvolution of High Resolution Mass Spectra

Peaks are detected and charge states are automatically calculated for whole datasets

Download Decon2LS

Picture sources: PNNL/ **Decon2Ls** Tutorial

Adduct formation

Adduct formation is observed in ESI, CI and APCI ionization modes (and others).
Adduct detection is problematic for small molecules, can be influenced by solvent selection.
Adduct detection can be automated if two or more adducts are detected in mass spectrum.
Switching polarity (+/-) can be used for confirmation of adduct.

Ion name	Ion mass	Charge	Mult	Mass	Result:
1. Positive ion mode					
M+H	M + 1.007276	1+	1	1.007276	2.007276
M+NH4	M + 18.033823	1+	1	18.033823	19.033823
M+Na	M + 22.989218	1+	1	22.989218	23.989218
M+CH3OH+H	M + 33.033489	1+	1	33.033489	34.033489
1. Negative ion mode					
M-H	M - 1.007276	1-	1	-1.007276	-0.007276
M+Na-2H	M + 20.974666	1-	1	20.974666	21.974666
M+Cl	M + 34.969402	1-	1	34.969402	35.969402

Download the Mass Spectral Adduct Calculator
Check: **MZedDB** Tools for the annotation of High Resolution MS metabolomics data

http://fiehnlab.ucdavis.edu/staff/kind/Metabolomics/MS-Adduct-Calculator/

Adduct formation – expect the unexpected

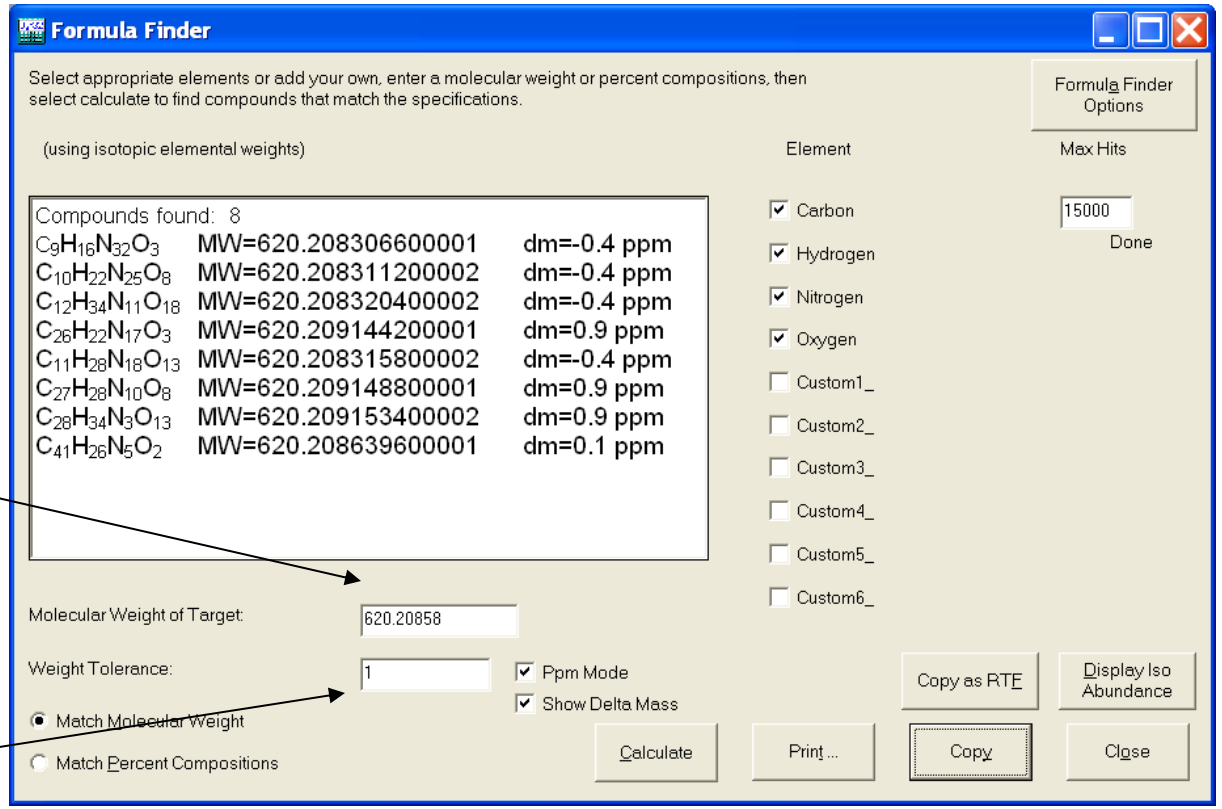
Adduct ion	Percent [%]	Adduct ion	Percent [%]	Adduct ion	Percent [%]	Adduct ion	Percent [%]	Adduct ion	Percent [%]
[M+H]+	62.55381	[M+H-C3H8O]+	0.02667	[M-CCl3]+	0.00381	[M(37Cl)]+.	0.00190	[M-2H+Na]-	0.00127
[M+2H]2+	11.44459	[M-H-H2O-CO2]-	0.02667	[M-H-CO2]-	0.00381	[M-CH3]+	0.00190	[M-H+Co]+	0.00127
[M+H-H2O]+	8.77598	[M-H-H2O-HCO2H]-	0.02667	[M+H-C5H7PO6]+	0.00381	[M+H-C4H11N]+	0.00190	[M+H-(CH3)2NH-C3H6]+	0.00127
[M-H]-	6.25214	[M+H-3H2O]+	0.02540	[M+H-HCl]+	0.00381	[M+H-NO2-CHO]+	0.00190	[M+H-C10H6(OH)N]+	0.00127
[M+Na]+	5.51055	[M+H-CHN]+	0.02540	[M+H-C12H12N2O3]+	0.00381	[M-H-HF]-	0.00190	[M-H+Ni]+	0.00127
[M+H-NH3]+	1.19494	[M+K-3H]2-	0.01905	[M+H-CH3CO2H]+	0.00381	[M(37Cl)+H]+	0.00190	[M-H-H2O-C4H7CO2H]-	0.00127
[M+NH4]+	0.73715	[M+H-(CH3)2NH]+	0.01524	[M+H-CH3]+.	0.00381	[M-H-C6H10O5]-	0.00190	[M+H-OH]+	0.00127
[M-H-H2O]-	0.34604	[M+H-CHNO]+	0.01333	[M+H-H2]+	0.00381	[M+H-H2O-C6H13N]+	0.00190	[M(81Br)+H]+…	0.00127
[M-H+2Na]+	0.32953	[M+H-C2H6O]+	0.01333	[M+H-C3H8NO6P]+	0.00317	[M+H-H2O-H3PO4]+	0.00190	[M-H-CH2O-CH2NH]-	0.00127
[M-H+H2O]-	0.24508	[M+H-CH4O]+	0.01270	[M+H-C5H14NO4P]+	0.00317	[M+H-C5H7PO6-NH3]+	0.00190	[M+H-CO-CONH]+	0.00127
[M+NH4-H2O]+	0.22984	[M+H-C7H13NO3]+	0.01143	[M+Li-(CH3)3N]+	0.00317	[M-H-C5H7PO6]-	0.00190	[M-H-CONH]-	0.00127
[M+H+H2O]+	0.19429	[M+Na-2H]-	0.00952	[M+Li-C5H14NO4P]+	0.00317	[M+H-H2S]+	0.00190	[M+H-C3H4O2]+	0.00127
[M+H+Na]2+	0.18286	[M-H-CH2O]-	0.00952	[M+Cl]-	0.00317	[M+H-H2O-C8H8]+	0.00190	[M+H-C3H6O4]+	0.00127
[M+H+K]2+	0.17524	[M+H-C11H12N2O3]+	0.00952	[M(35Cl)-H]-	0.00317	[M+H-H2O-NH3-C8H8]+	0.00190	[M+Na-H2S]+	0.00127
[M-2H]2-	0.13968	[M+H-C13H16N3O4]+	0.00952	[M(37Cl)-H]-	0.00317	[M+H-H2O-NH3-C8H8-CO]+	0.00190	[M-H+2Na-H2S]+	0.00127
[M+2Na]2+	0.13778	[M+H-C17H25N3O4]+	0.00952	[M-H-C5H7O6P]-	0.00317	[M+H-H2O-NH3]+	0.00190	[M-C5H5Cl]+	0.00127
[M+2H-NH3]2+	0.13714	[M+CH3CO2]-	0.00889	[M+H-C3H7O5P]+	0.00317	[M+H-C3H6]+	0.00190	[M+H-N2]+	0.00127
[M+K]+	0.13651	[M-H2O+Na]+	0.00825	[M-H-C6H6N8O]-	0.00317	[M+HCO2-320]-	0.00190	[M+H-H2O-CO]+	0.00127
[M+H-2H2O]+	0.11810	[M-H+NH3]-	0.00762	[M(81Br)+H]+	0.00317	[M+H-C3H7N]+	0.00190	[M-H-H3PO4]-	0.00127
[M+3H]3+	0.06667	[M+H-C9H9NO]+	0.00762	[M-C4H9]+	0.00317	[M-H-H2]-	0.00190	[M+H+CH3CN]+	0.00127
[M+2H-H2O]2+	0.06476	[M+H-C15H21N2O3]+	0.00762	[M-2H+3Li]+	0.00254	[M-H-C16H30O-H2O]-	0.00190	[M+H-C4H6]+	0.00127
[M]+.	0.05905	[M-2H+3Na]+	0.00698	[M-H-HCl]-	0.00254	[M-H-CH4O]-	0.00190	[M+H-CH3OH]+	0.00127
[M+2Na-H]+	0.05143	[M+HCO2]-	0.00635	[M+2Li-H]+	0.00254	[M+H-C10H8FN3]+	0.00127	[M+H-HCCl3]+	0.00127
[M-H+2K]+	0.05079	[M+H-NO2]+	0.00571	[M+H-C8H10O2]+	0.00254	[M+Li-C3H5NO2]+	0.00127	[M+H-C2H3N3]+	0.00127
[M+H-CO]+	0.04635	[M+H-C6H13NO2]+	0.00571	[M+H-C2Cl4]+	0.00254	[M+Li-H3PO4]+	0.00127	[M+H-C3H6O2]+	0.00127
[M+H-CO2]+	0.04318	[M-H-C3H5NO2]-	0.00508	[M-H-C7H5NO]-	0.00254	[M-2H+3Li-C15H31CO2H]+	0.00127	[M+H-CH2Cl2O]+	0.00127
[M+H-CH2O2]+	0.03810	[M(81Br)-H]-	0.00508	[M+H-C5H11N]+	0.00254	[M-2H+3Na-C3H5NO2]+	0.00127	[M(356)+H-HCl]+	0.00127
[M-H-NH3]-	0.03746	[M+H-HCO2H]+	0.00508	[M+Ba-H]+	0.00254	[M-2H+Na+Co]+	0.00127	[M-C4H4O4S]+	0.00127
[M.Cl]-	0.03556	[M-2H+Li]-	0.00444	[M+H-C14H25NO3]+	0.00254	[M-2H+Li-C3H5NO2]-	0.00127	[M+H-C8H14O3]+	0.00127
[M+Li]+	0.03111	[M+H-CH4]+	0.00444	[M+H-C6H5NO2S]+	0.00254	[M-2H+Li-C16H30O]-	0.00127	[M+H-C2H4]+	0.00127

…around 290 different adducts

Statistics: Adducts in NIST12 MS/MS DB (80,000 spectra)
Most common adducts for LC-MS ([M+H]+ [M+Na]+ [M+NH4]+ [M+acetate]+)

Molecular Formula Generators

Formula generators are used to create molecular formulae from accurate masses.
Input requires 1) accurate isotopic mass (with or without adduct) and
2) error in ppm or mDa (milli Dalton)

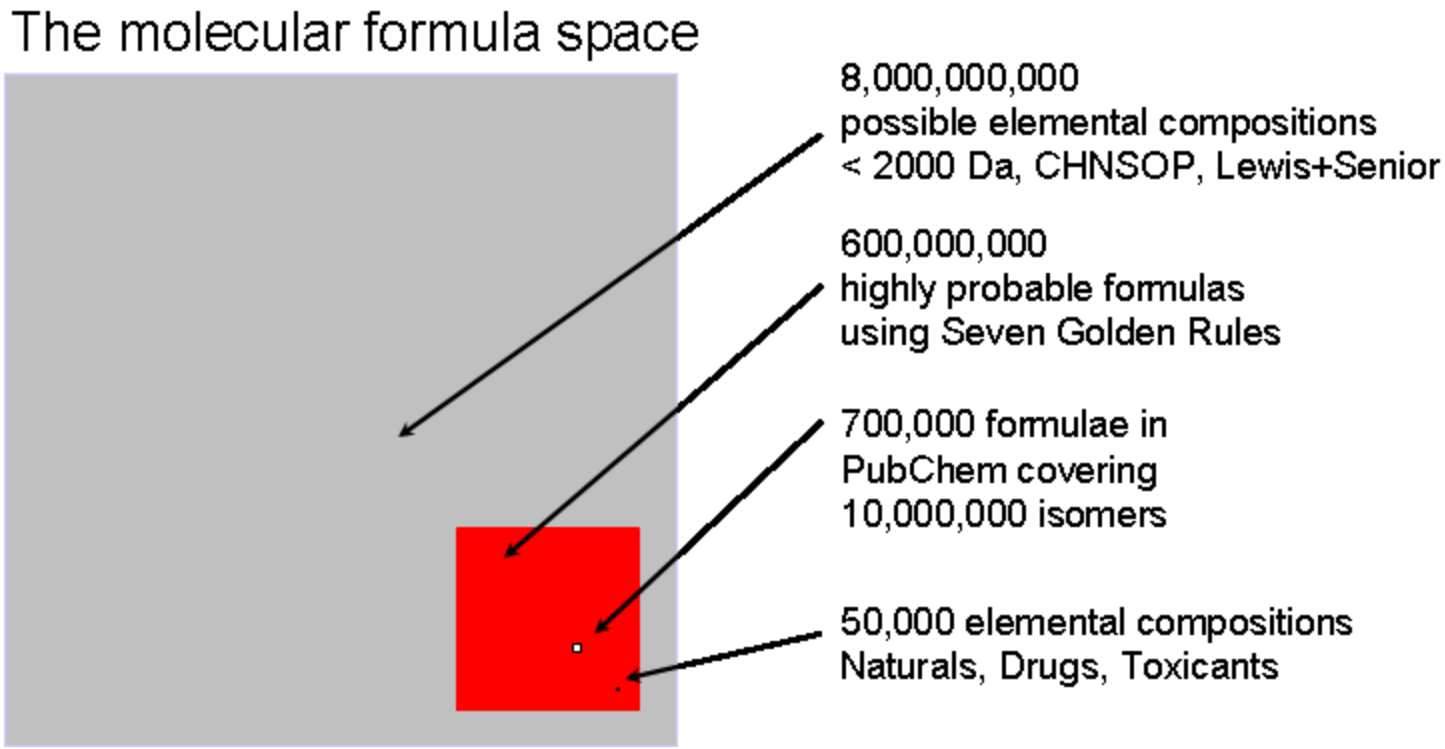


Accurate mass

Mass error ± (ppm or mDa)

Example MWTWIN

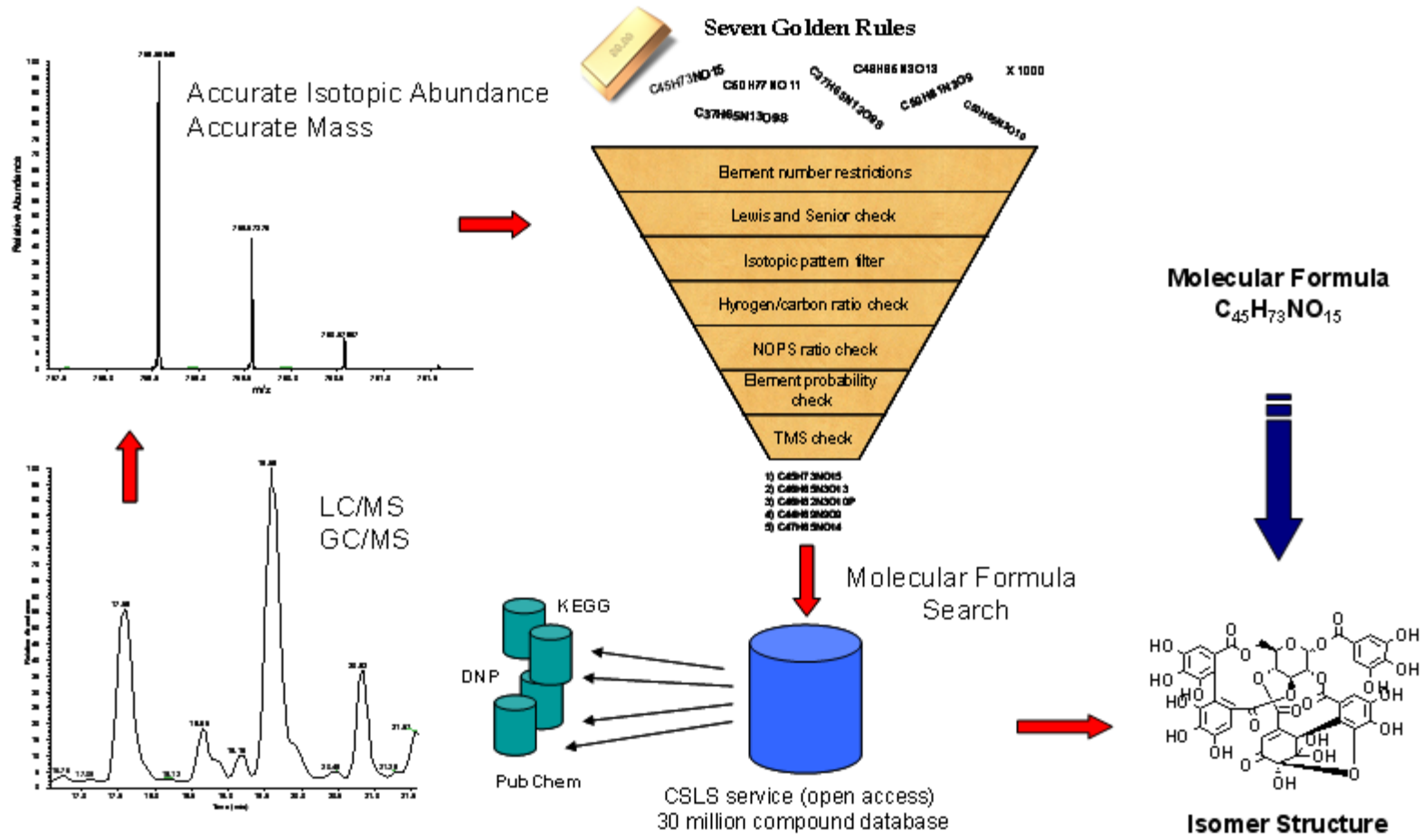
The molecular formula space of small molecules calculated by the Seven Golden Rules



Each molecular formula can expand to billions of structural isomers.
Molecular Formula ≠ Molecular Isomer

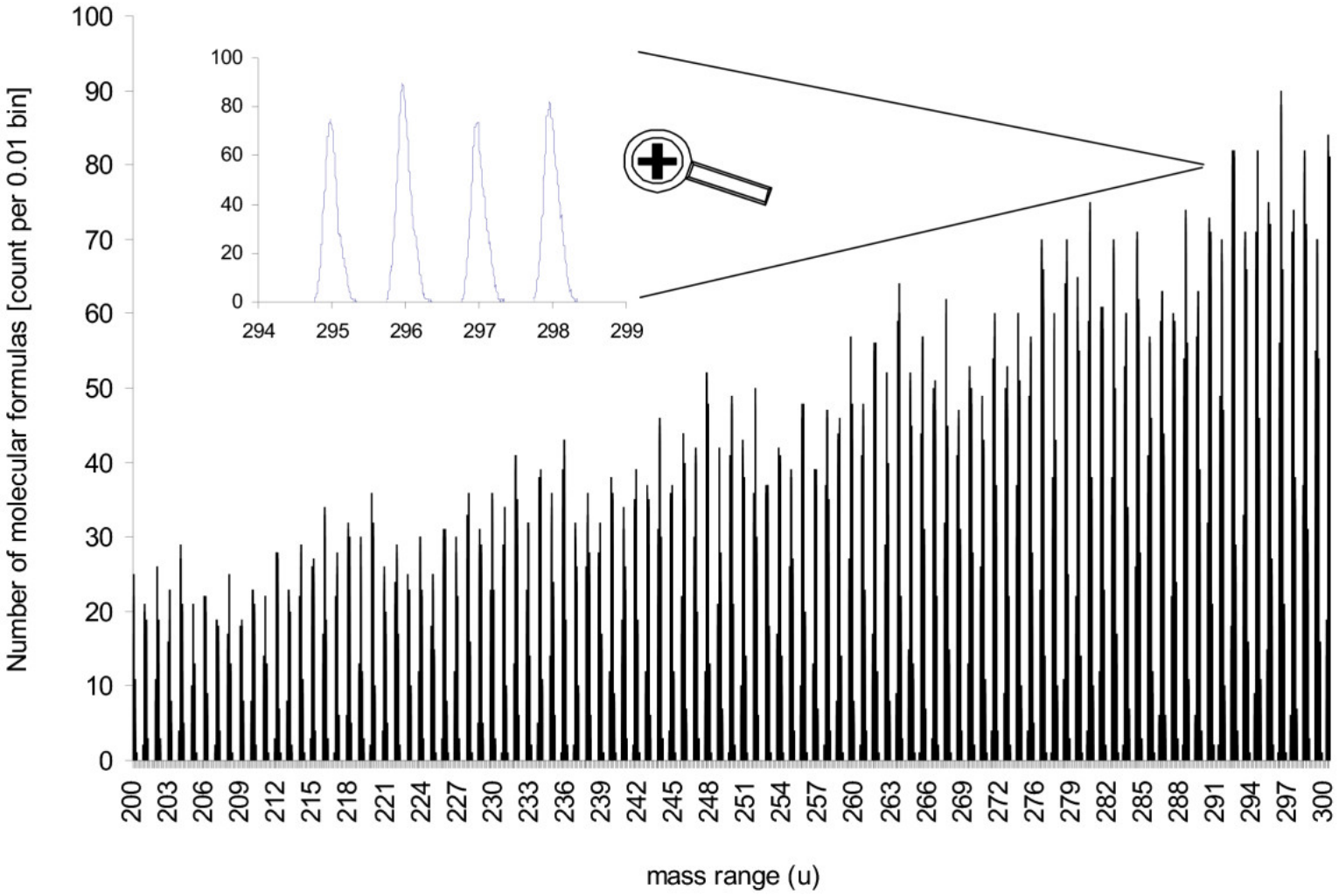
Source http://fiehnlab.ucdavis.edu/projects/Seven_Golden_Rules/Molecular-Formula-Space/

Seven Golden Rules workflow for structure proposals



Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry
BMC Bioinformatics 2007, 8:105 doi:10.1186/1471-2105-8-105
Tobias Kind and Oliver Fiehn

Frequency distribution of molecular formulas



Source: Metabolomic database annotations via query of elemental compositions: Mass accuracy is insufficient even at less than 1 ppm
Kind and Fiehn *BMC Bioinformatics* 2006 **7**:234 doi:10.1186/1471-2105-7-234

Impact of mass accuracy on number of formulas

http://www.biomedcentral.com - Table 3 - Mozilla Firefox

Table 3

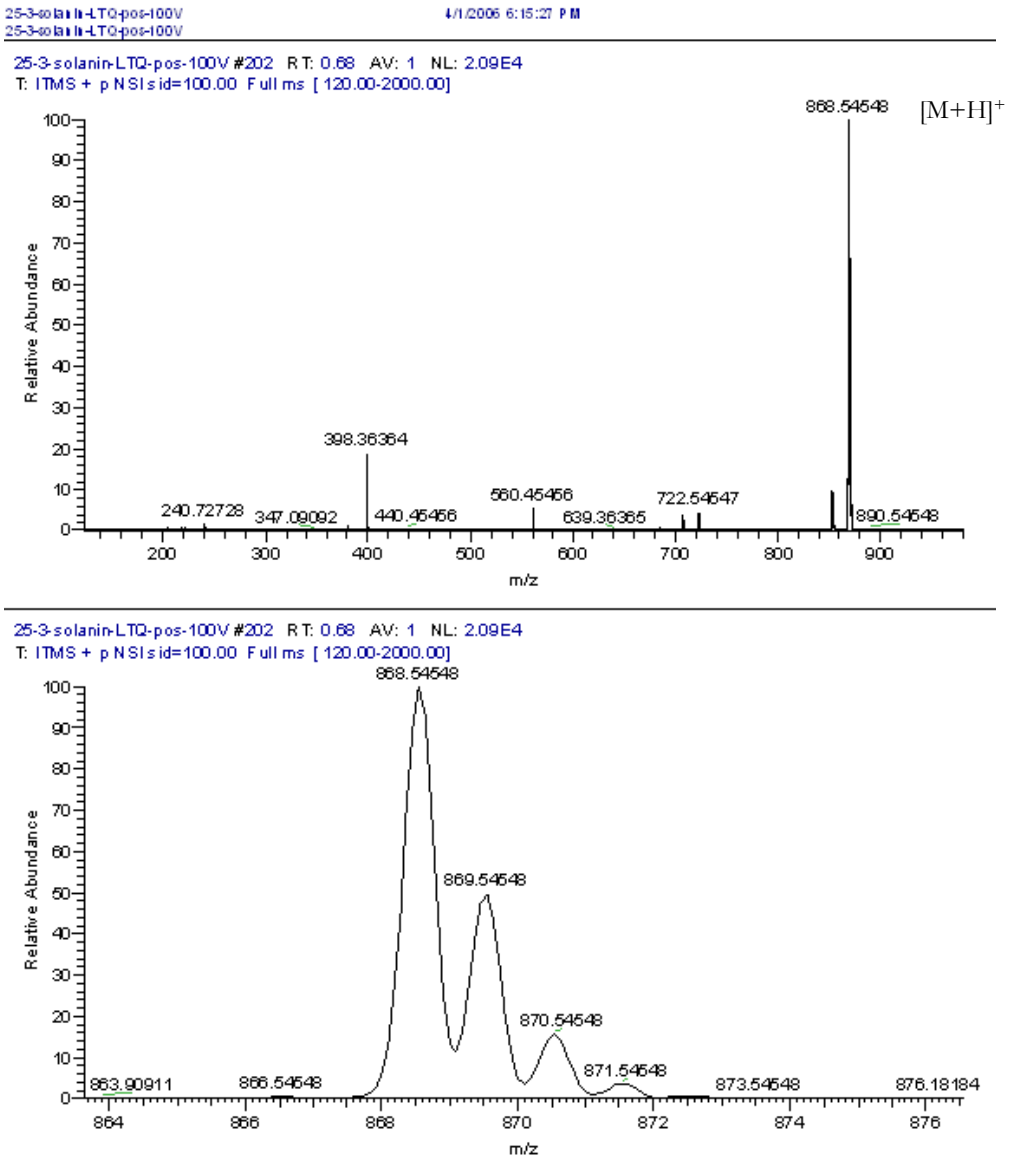
Number of possible molecular formulas at different levels of mass accuracy and the impact of isotopic abundance accuracy. A mass spectrometer capable of 3 ppm but with 2% correct isotopic pattern outperforms even a (non-existing) mass spectrometer with 0.1 ppm mass accuracy! The results are computed 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

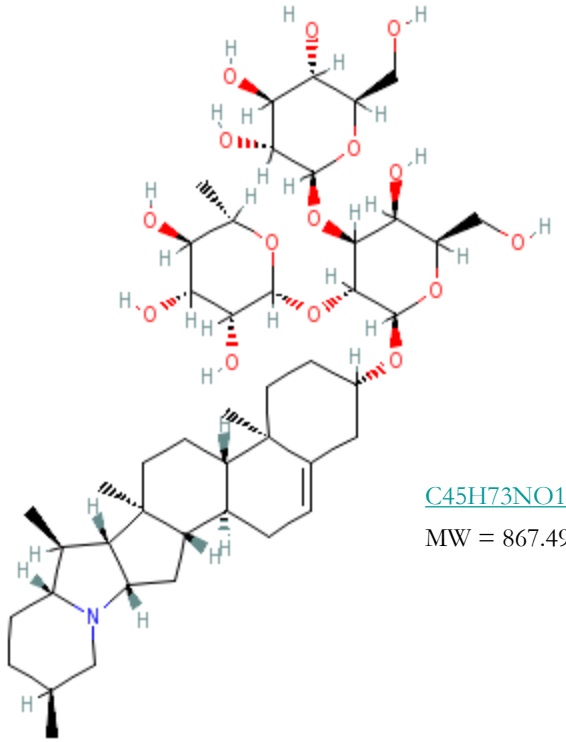
	without isotope abundance information					2% isotopic abundance accuracy	5% isotopic abundance accuracy
molecular mass [Da]	10 ppm	5 ppm	3 ppm	1 ppm	0.1 ppm	3 ppm	5 ppm
150	2	1	1	1	1	1	1
200	3	2	2	1	1	1	1
300	24	11	7	2	1	1	6
400	78	37	23	7	1	2	13
500	266	115	64	21	2	3	33
600	505	257	155	50	5	4	36
700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

Kind and Fiehn *BMC Bioinformatics* 2006 **7**:234 doi:10.1186/1471-2105-7-234

Done

Mass accuracy and isotopic pattern





C45H73NO15

MW = 867.49799

Example:

ESI-MS (+) of Solanine on a LTQ

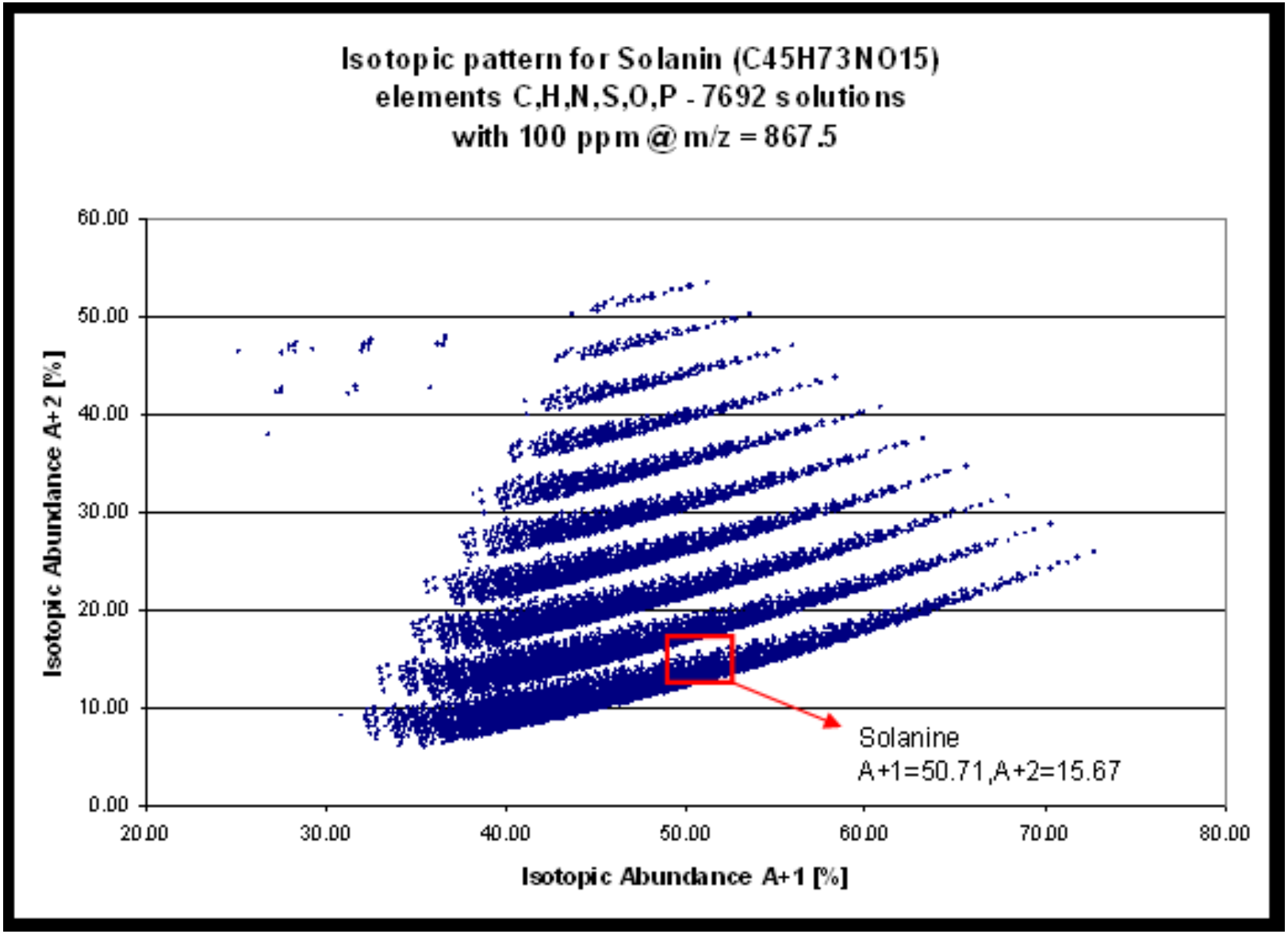
Resolving Power: 1700

Mass Accuracy: 46 ppm

Isotopic Abundance Error: ±1.46%

Example from: http://fiehnlab.ucdavis.edu/projects/Seven_Golden_Rules/Examples/Solanine/

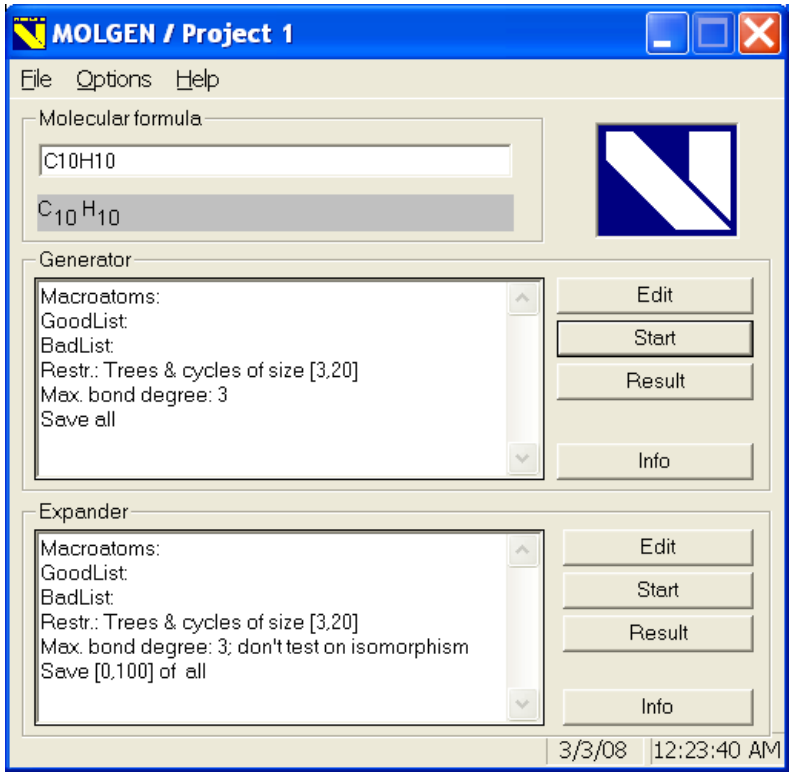
Isotopic abundances as orthogonal filter

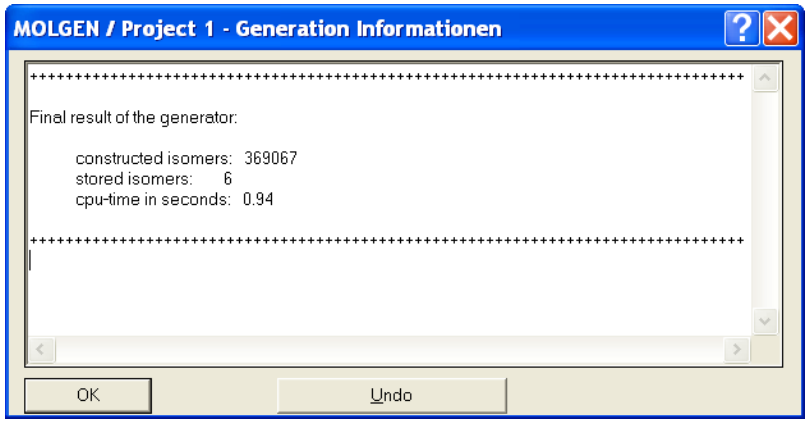


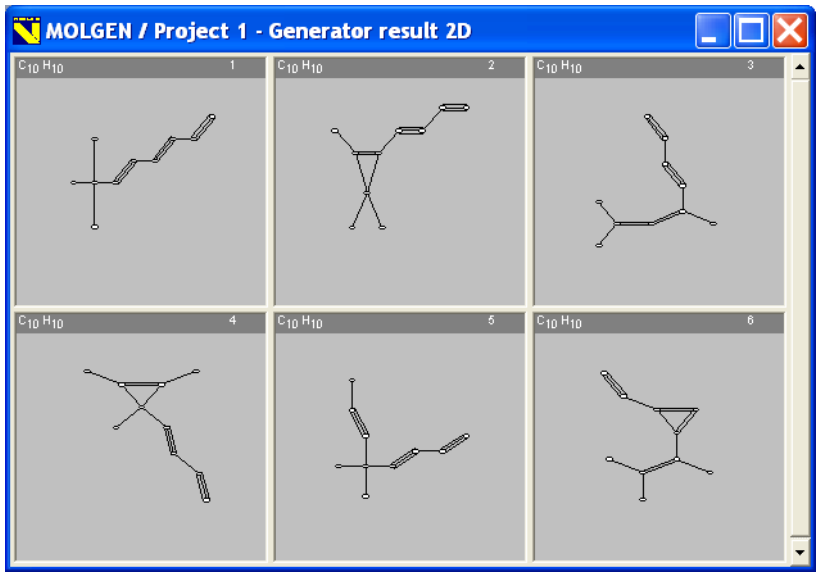
Source: Kind and Fiehn *BMC Bioinformatics* 2006 **7**:234 doi:10.1186/1471-2105-7-234

Molecular Isomer Generators

Isomer generators are used to create all possible **structural isomers** from a given **molecular formula**. **Deterministic** and **stochastic** (random) generators are in use.







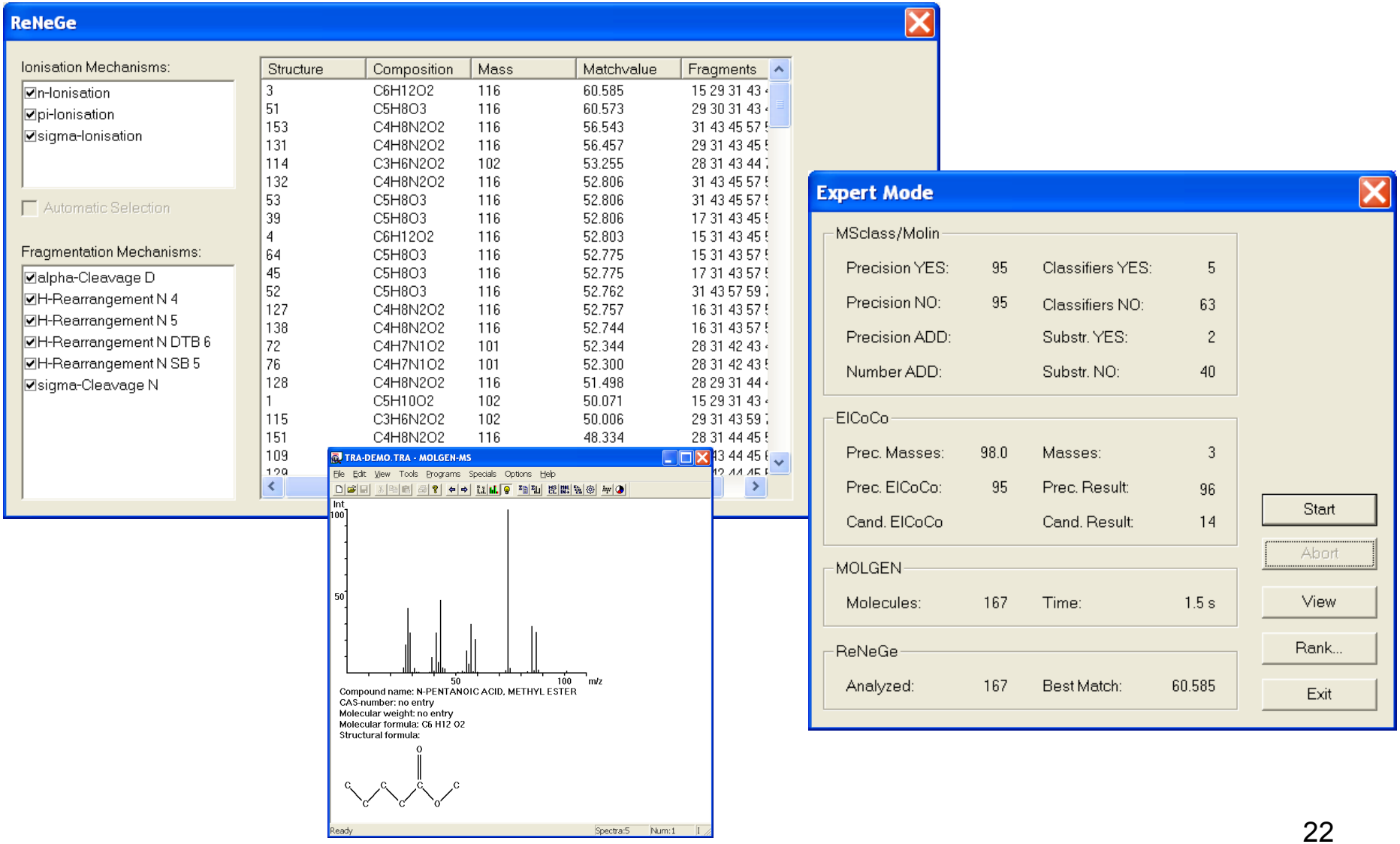
Example: MOLGEN DEMO (Bayreuth)
See also: OMG Open Molecule Generator

The molecular isomer space is unknown

Accurate mass	Formula	Number Isomers	in Beilstein DB
77.99531	CH2O4	6	0
78.04293	CH6N2O2	28	1
78.03169	C2H6O3	10	8
78.02180	C4H2N2	465	2
78.01056	C5H2O	151	2
78.04695	C6H6	217	29
150.04293	C7H6N2O2	100,082,479	153
150.09054	C7H10N4	66,583,863	105
150.03169	C8H6O3	6,717,404	90
150.07931	C8H10N2O	76,307,072	542
150.06808	C9H10O2	6,843,602	667
150.11569	C9H14N2	9,459,132	568
150.02180	C10H2N2	65,563,828	0
150.10446	C10H14O	1,548,361	1938
150.01056	C11H2O	9,414,509	0
150.14084	C11H18	84,051	762
150.04695	C12H6	34,030,905	12

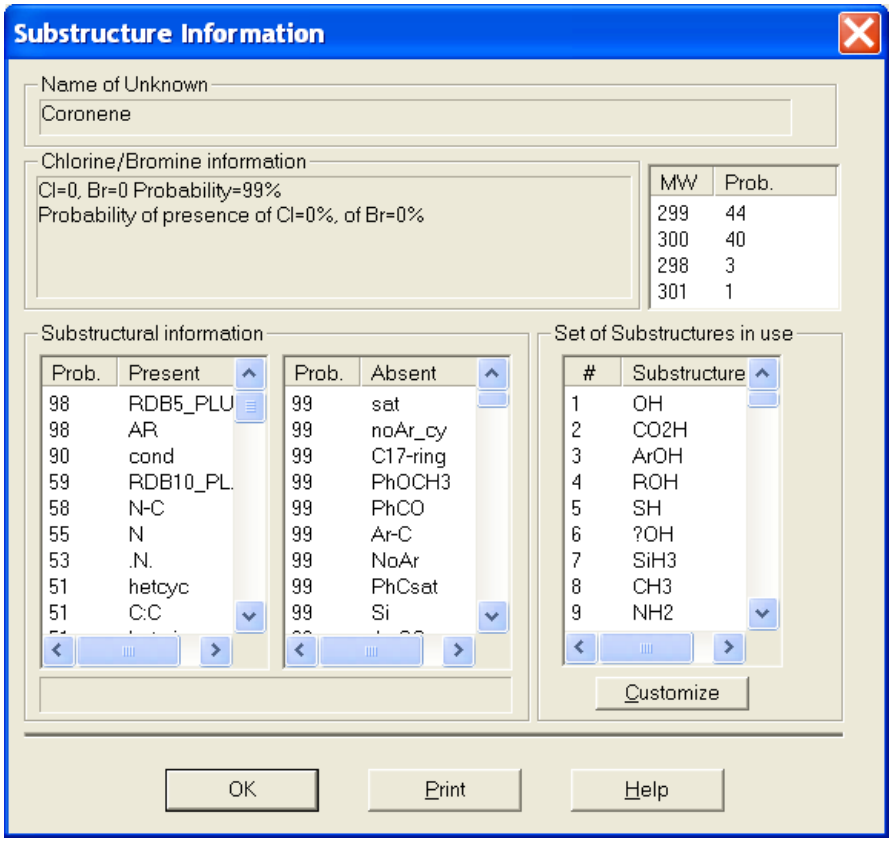
Source: http://fiehnlab.ucdavis.edu/projects/Seven_Golden_Rules/Molecular-Isomer-Generator/
Meringer M: Mathematische Modelle für die kombinatorische Chemie und die molekulare Strukturaufklärung.

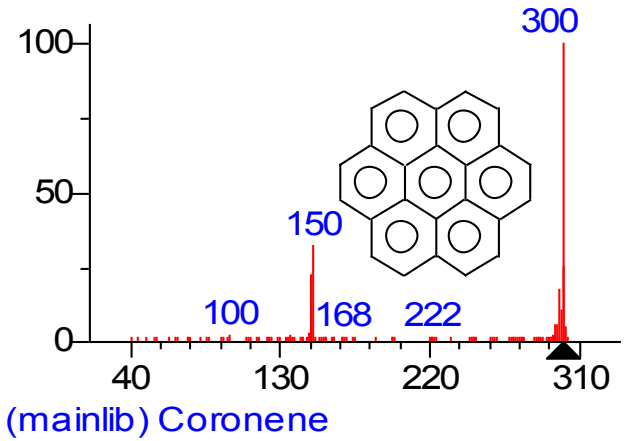
MOLGEN-MS


ReNeGe
Ionisation Mechanisms:
n-Ionisation
pi-Ionisation
sigma-Ionisation
Automatic Selection
Fragmentation Mechanisms:
alpha-Cleavage D
H-Rearrangement N 4
H-Rearrangement N 5
H-Rearrangement N DTB 6
H-Rearrangement N SB 5
sigma-Cleavage N
Structure Composition Mass Matchvalue Fragments
3 C6H12O2 116 60.585 15 29 31 43
51 C5H8O3 116 60.573 29 30 31 43
153 C4H8N2O2 116 56.543 31 43 45 57
131 C4H8N2O2 116 56.457 29 31 43 45
114 C3H6N2O2 102 53.255 28 31 43 44
132 C4H8N2O2 116 52.806 31 43 45 57
53 C5H8O3 116 52.806 31 43 45 57
39 C5H8O3 116 52.806 17 31 43 45
4 C6H12O2 116 52.803 15 31 43 45
64 C5H8O3 116 52.775 15 31 43 57
45 C5H8O3 116 52.775 17 31 43 57
52 C5H8O3 116 52.762 31 43 57 59
127 C4H8N2O2 116 52.757 16 31 43 57
138 C4H8N2O2 116 52.744 16 31 43 57
72 C4H7N1O2 101 52.344 28 31 42 43
76 C4H7N1O2 101 52.300 28 31 42 43
128 C4H8N2O2 116 51.498 28 29 31 44
1 C5H10O2 102 50.071 15 29 31 43
115 C3H6N2O2 102 50.006 29 31 43 59
151 C4H8N2O2 116 48.334 28 31 44 45
109 43 44 45
129 42 44 45
TRA-DEMO.TRA - MOLGEN-MS
File Edit View Tools Programs Specials Options Help
Int
100
50
50 100 m/z
Compound name: N-PENTANOIC ACID, METHYL ESTER
CAS-number: no entry
Molecular weight: no entry
Molecular formula: C6 H12 O2
Structural formula:
Ready Spectra:5 Num:1
Expert Mode
MSclass/Molin
Precision YES: 95 Classifiers YES: 5
Precision NO: 95 Classifiers NO: 63
Precision ADD: Substr. YES: 2
Number ADD: Substr. NO: 40
EICoCo
Prec. Masses: 98.0 Masses: 3
Prec. EICoCo: 95 Prec. Result: 96
Cand. EICoCo Cand. Result: 14
MOLGEN
Molecules: 167 Time: 1.5 s
ReNeGe
Analyzed: 167 Best Match: 60.585
Start
Abort
View
Rank...
Exit

Substructure predictions

Use computer algorithms (machine learning) for automated interpretation of fragments and corresponding substructures. Algorithms creates present/absent list of substructures.



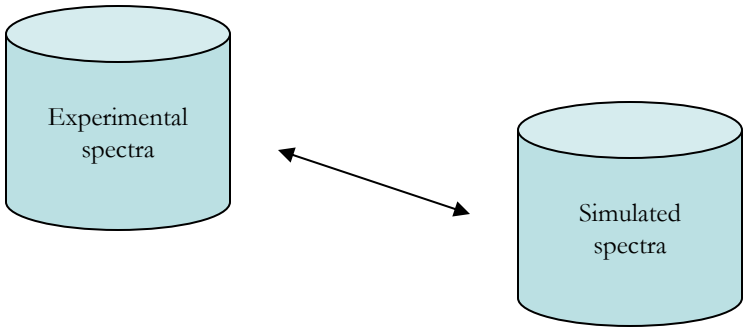


Example from NIST-MS search

Simulation of mass spectra

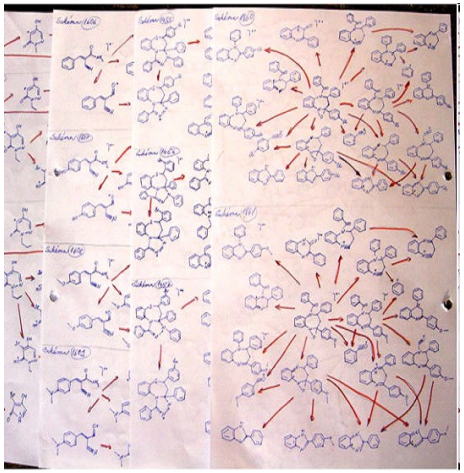
Why is simulation or prediction of mass spectra important?

→ Molecular isomers (structures) can be generated very fast from molecular formulas
→ Only certain mass spectra can be simulated (MS/MS of peptides, oligosaccharides, lipids)
→ Problematic is abundance determination
→ Problematic are all complex rearrangement reactions (gas phase ion chemistry)
→ Simulation of mass spectra from small molecules is new and important research

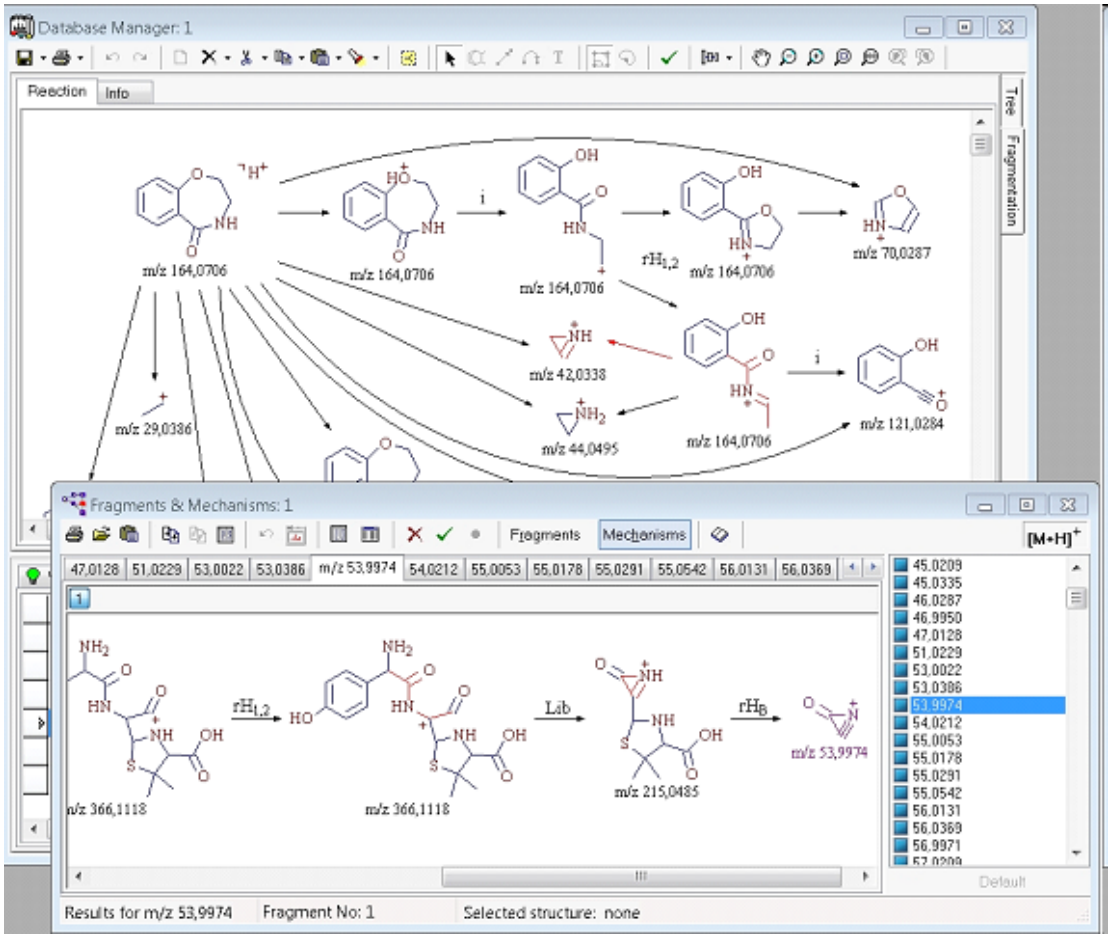


Perform comparison or matching

HighChem MassFrontier contains organic reaction and related mass spectral fragmentation database



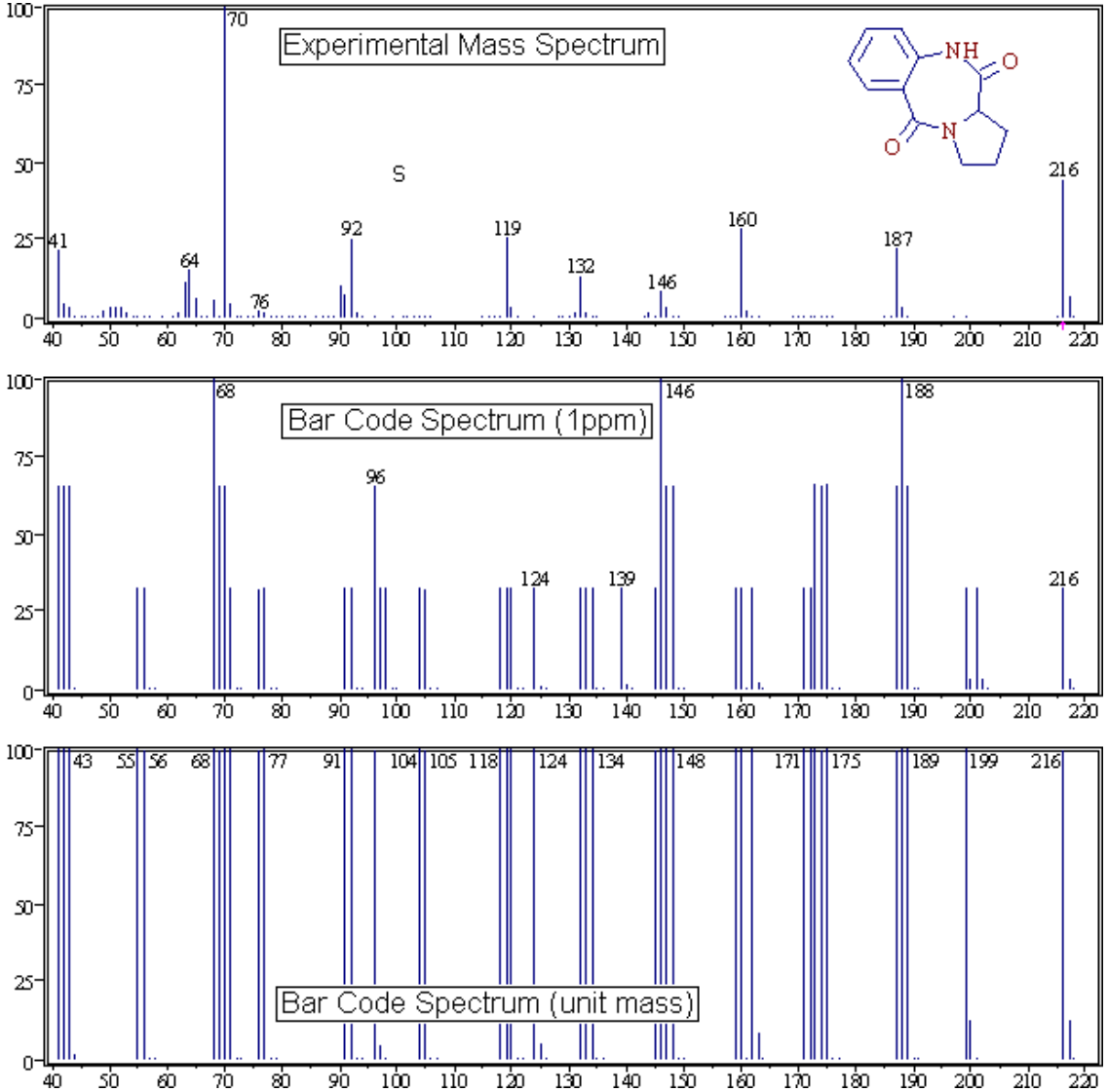
Paper publications



30,936 fragmentations
100,000 individual reaction mechanisms
Can be used to predict fragmentations of unknown molecules

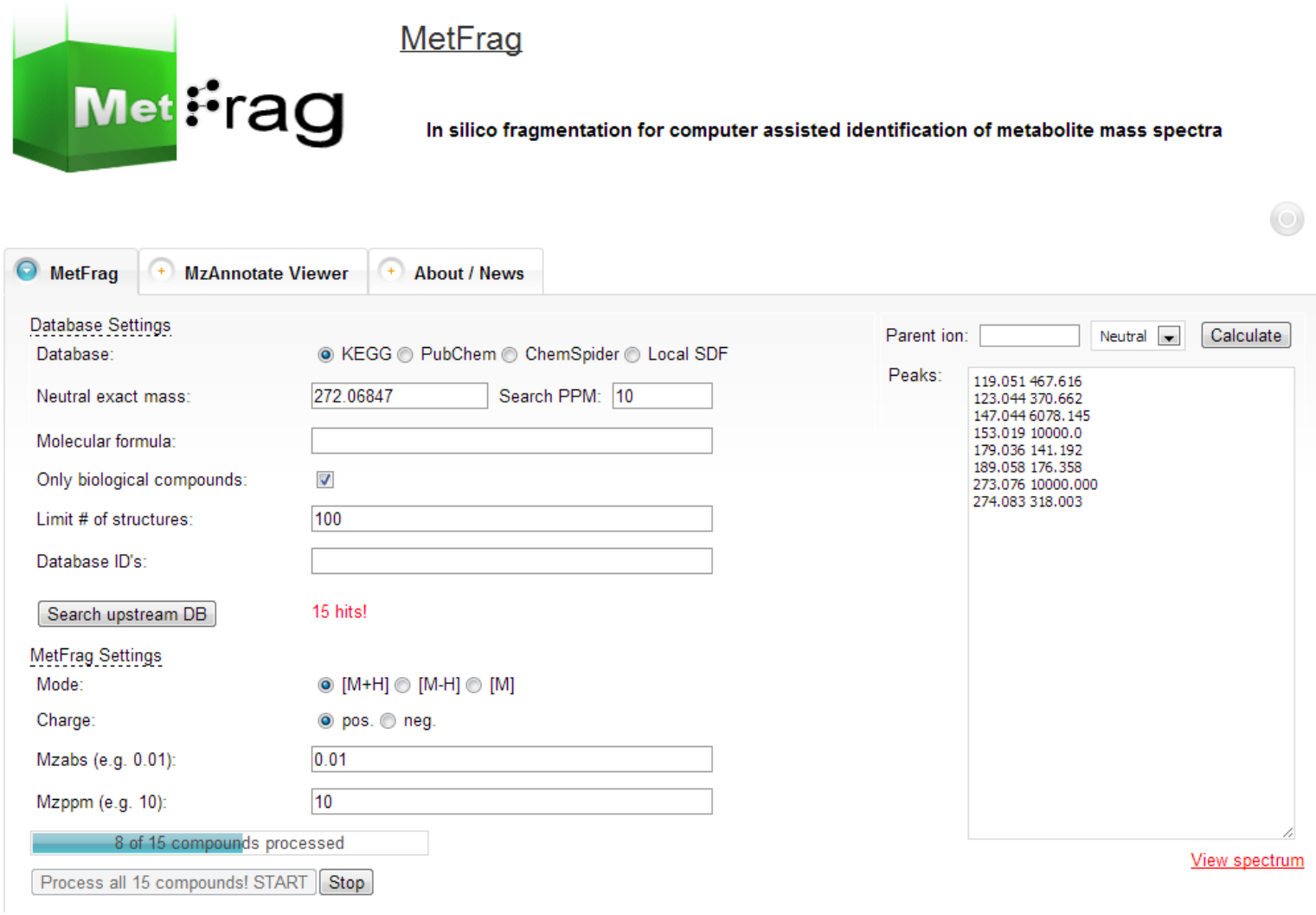
Figures: Robert Mistrik (HighChem)

Barcode spectrum from Mass Frontier



Example: MassFrontier Helpfile

Metfrag – computer based fragmentation of structures from compound DBs



Peptide Sequence Fragmentation Modelling

Peptide Sequence Fragmentation Modelling

File Edit View Tools Help

Sequence: RHPEYAVSVLLR

Match Ions

1 letter notation

Mass Information

MW = 1438.804443

[M+H]1+ 1439.811719 Da

[M+2H]2+ 720.409498 Da

Element Mode: Average / Isotopic

N and C Terminus

N: H (hydrogen)

C: OH (hydroxyl)

Ion Types

A Ions

B Ions

Y Ions

Neutral Losses

A B Y

Loss of H2O

Loss of NH3

Loss of PO4

Charge Options

2+ charged ions

Threshold 500

#	Immon.	a	b	b-H2O	b-NH3	Seq.	y	y-H2O	y-NH3
1	129.11					R	1439.81	1421.80	1422.79
2	110.07	266.17	294.17		277.14	H	1283.71	1265.70	1266.68
3	70.07	363.23	391.22		374.19	P	1146.65	1128.64	1129.62
4	102.06	492.27	520.26	502.25	503.24	E	1049.60	1031.59	1032.57
5	136.08	655.33	683.33	665.32	666.30	Y	920.56	902.55	903.53
6	44.05	726.37	754.36	736.35	737.34	A	757.49	739.48	740.47
7	72.08	825.44	853.43	835.42	836.41	V	686.46	668.44	669.43
8	60.04	912.47	940.46	922.45	923.44	S	587.39	569.38	570.36
9	72.08	1011.54	1039.53	1021.52	1022.51	V	500.35		483.33
10	86.10	1124.62	1152.62	1134.61	1135.59	L	401.29		384.26
11	86.10	1237.71	1265.70	1247.69	1248.67	L	288.20		271.18
12	129.11					R	175.12		158.09

Mass	Intensity	Symbol
136.08	100.00	
139.11	28.81	
159.12	57.63	
175.13	62.71	y1
187.11	62.71	
227.10	13.42	
235.12	8.23	
246.65	29.44	
260.65	64.50	
261.14	16.88	
271.20	7.36	y2-NH3
288.21	27.27	y2
289.21	7.36	
294.19	100.00	b2
295.18	18.18	

Ion Match Options

Remove Precursor Ion

Ion Mass 480.63 Da

Mass Window .5 Da

Ion Matching Window 0.5 Da

Alignment Offset 0 Da

Ion Statistics

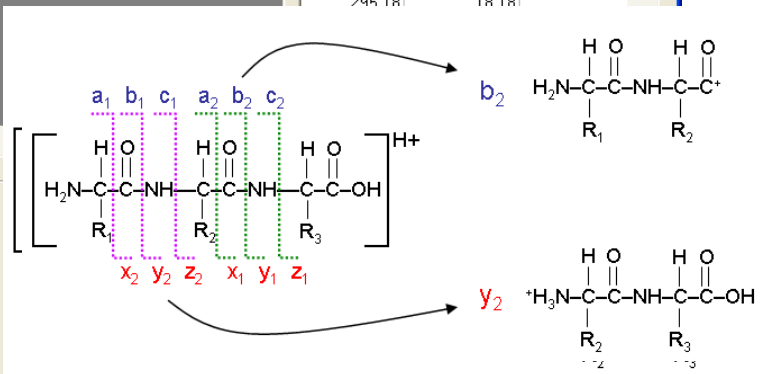
Loaded: 763

Within tolerance: 100 (13.1%)

Precursor removed: 481±0.5

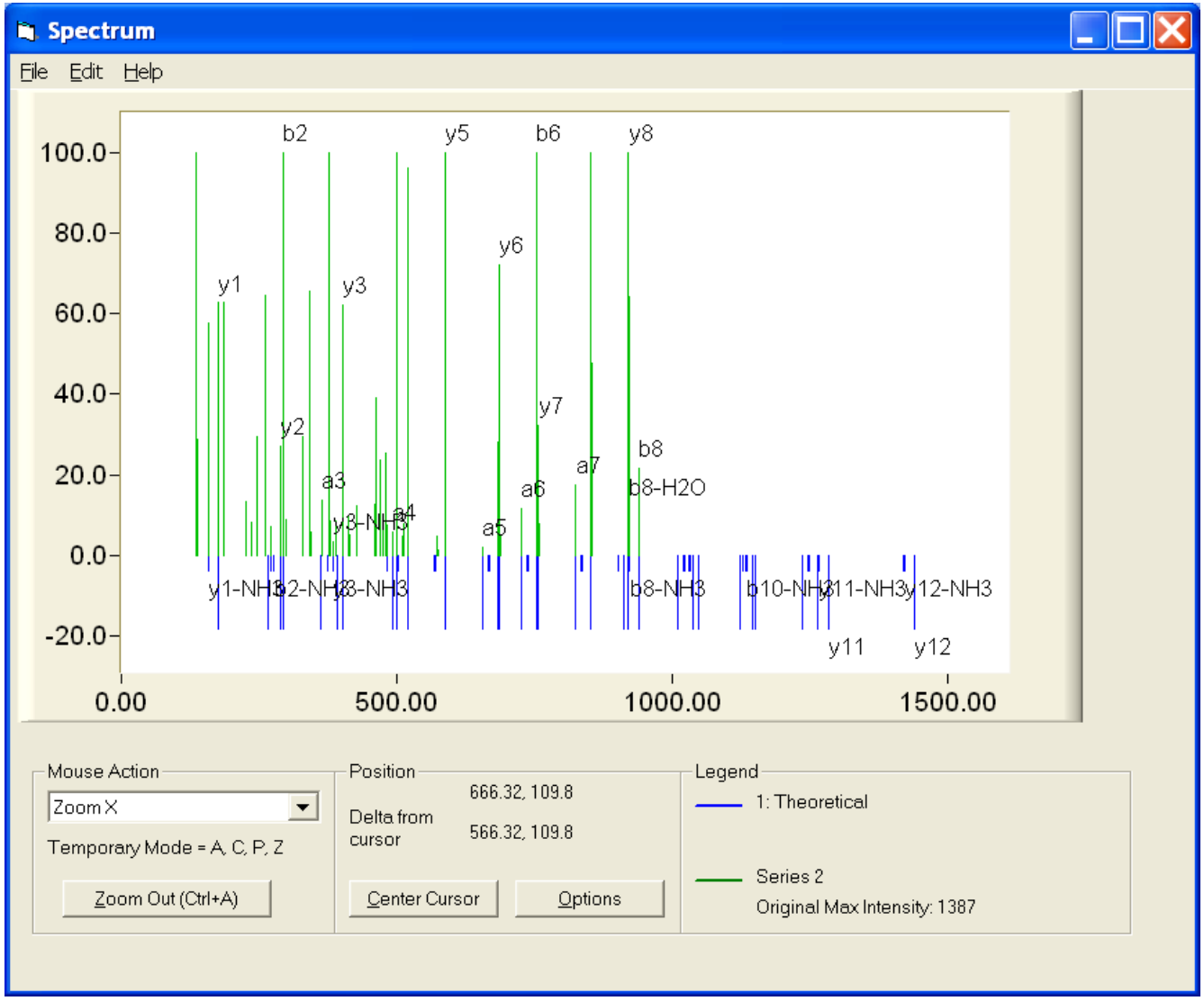
Matches: 23

Score: 235.7



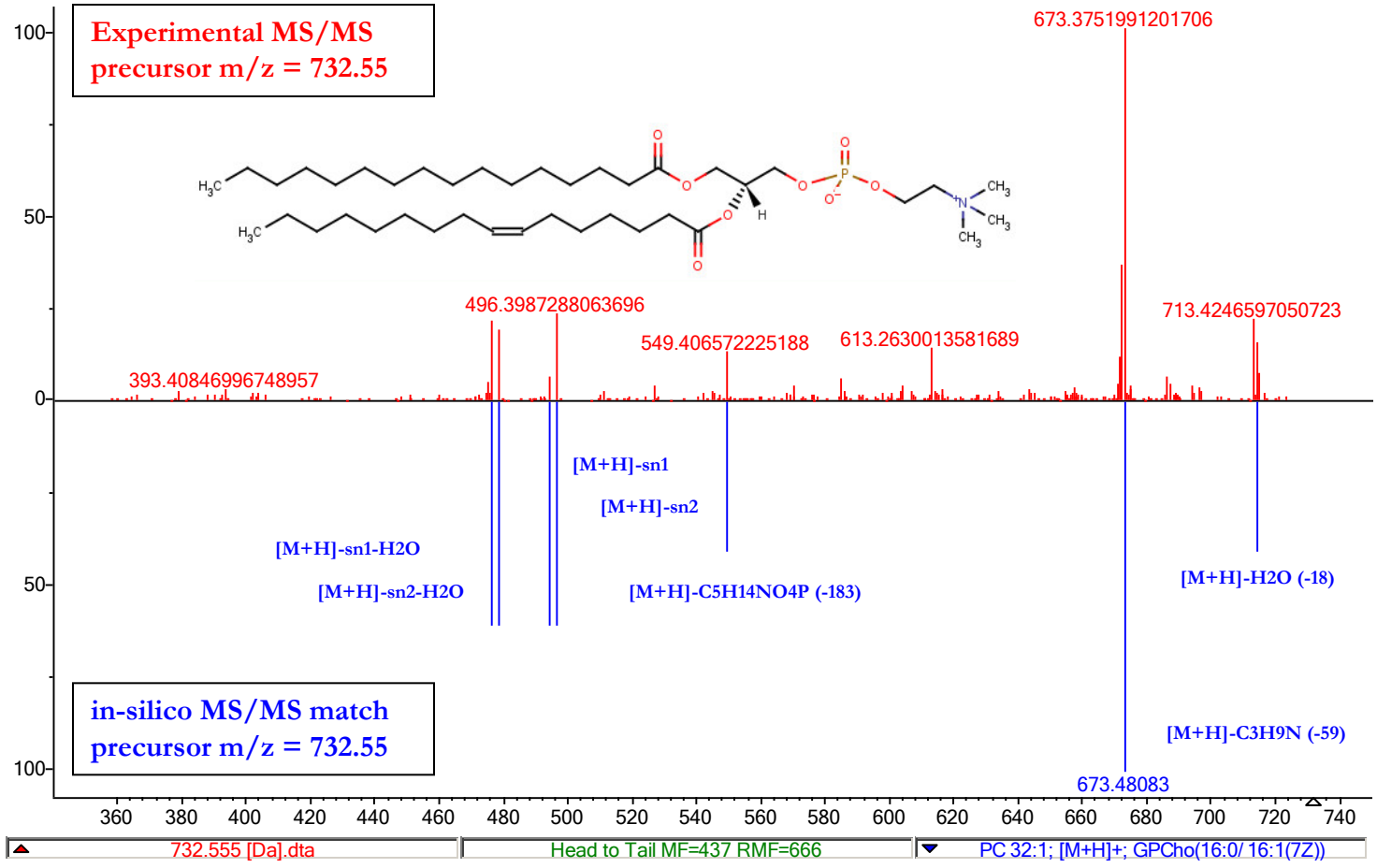
Example: Taken from Molecular Weight Calculator (Matthew Monroe / PNNL)

Matching experimental vs. theoretical sequence



Example: RHPEYAVSVLLR

LipidBlast in-silico MS/MS spectra



Nature Methods. 2013 Aug;10(8):755-8. doi: 10.1038/nmeth.2551.
LipidBlast in silico tandem mass spectrometry database for lipid identification.
Kind T, Liu KH, Lee do Y, Defelice B, Meissen JK, Fiehn O.

Free download: LipidBlast

The Last Page - What is important to remember:

Important performance parameters are accurate mass, resolving power, scan speed and accurate isotopic abundances

Even high resolving power MS can not distinguish between structural isomers

Accurate Mass → Molecular Formula → Structural Isomers → MS/MS

Mass spectra of only some substance classes can be simulated

Only NMR can perform de-novo structure elucidation in an consistent manner

Of general importance for this course:
http://fiehnlab.ucdavis.edu/staff/kind/Metabolomics/Structure_Elucidation/

Advances in structure elucidation of small molecules using mass spectrometry **(Kind & Fiehn 2010)**