Hands-on Automated MS Data Analysis using Lipid Data Analyzer (LDA)

Part I – Introduction

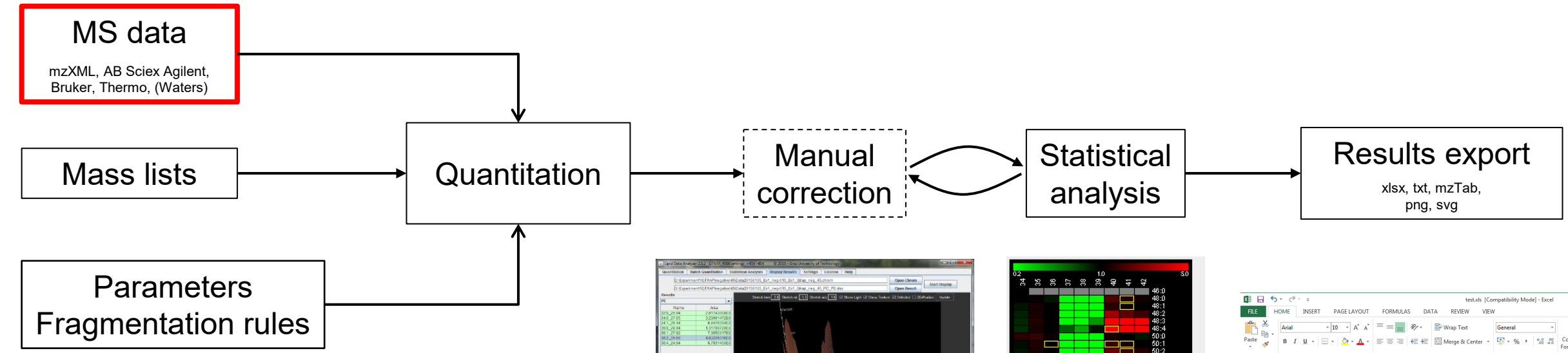
Leonida M. Lamp

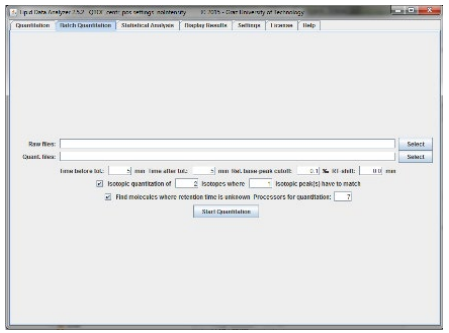
Jürgen Hartler

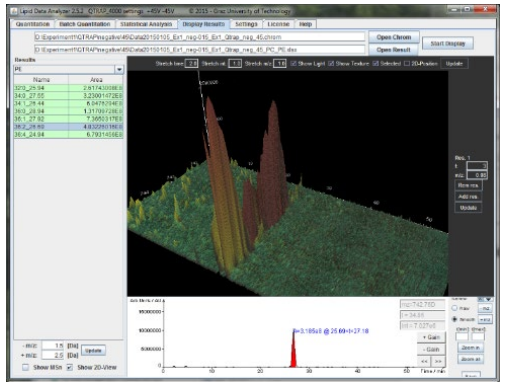
Computational Pharmacology Group

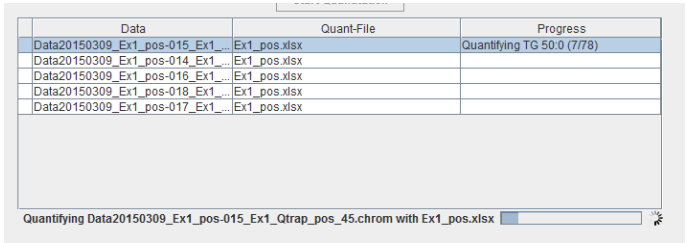
University of Graz

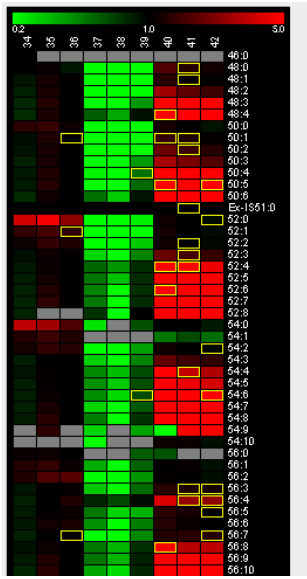
LDA – Data analysis workflow

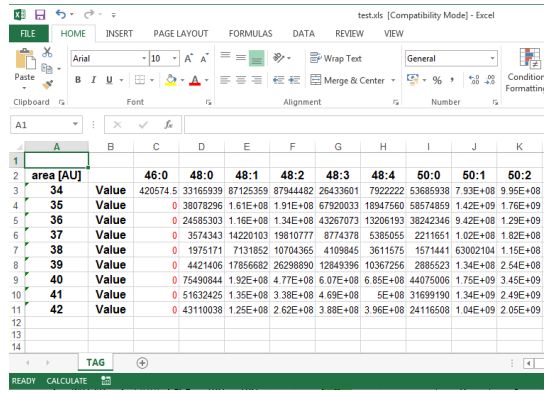






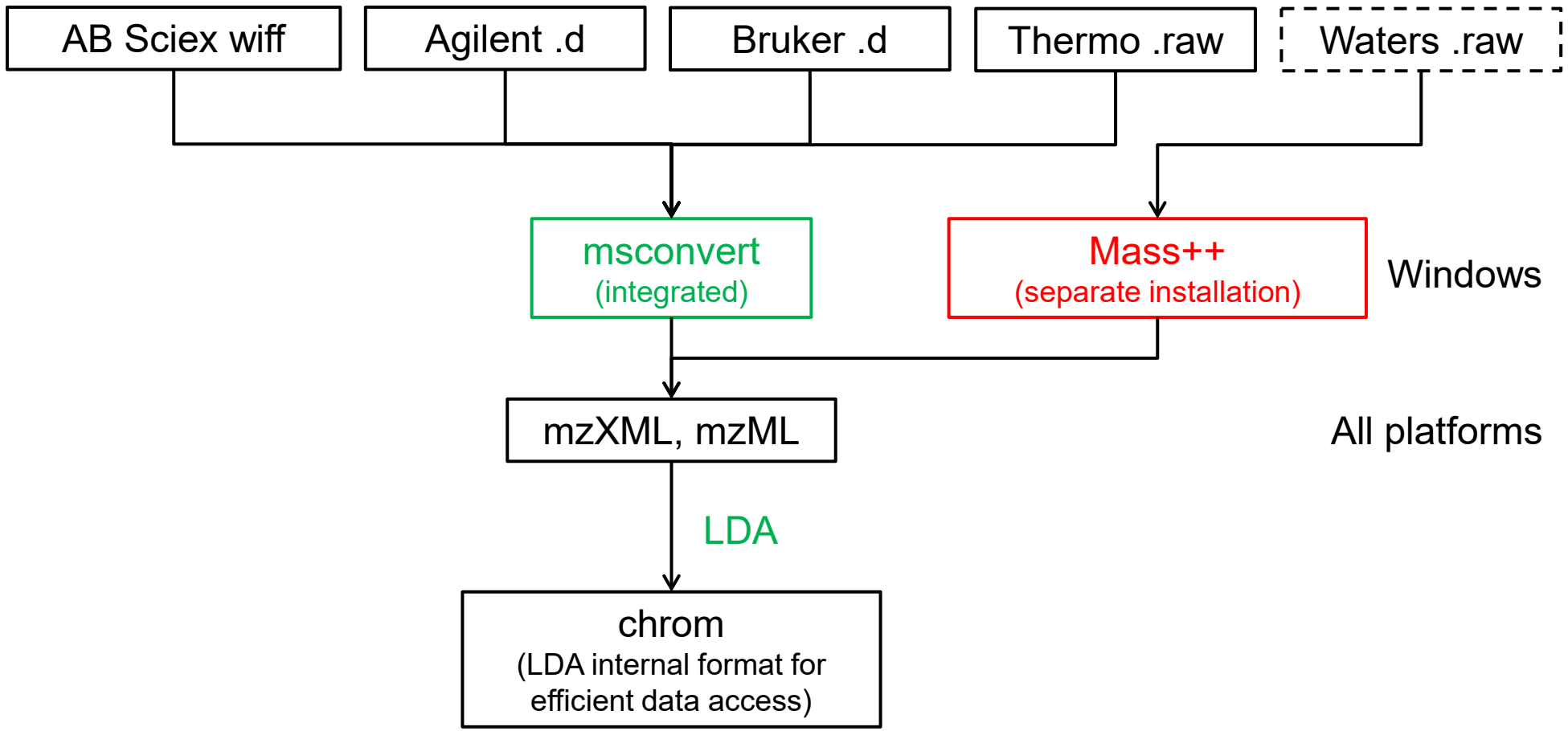




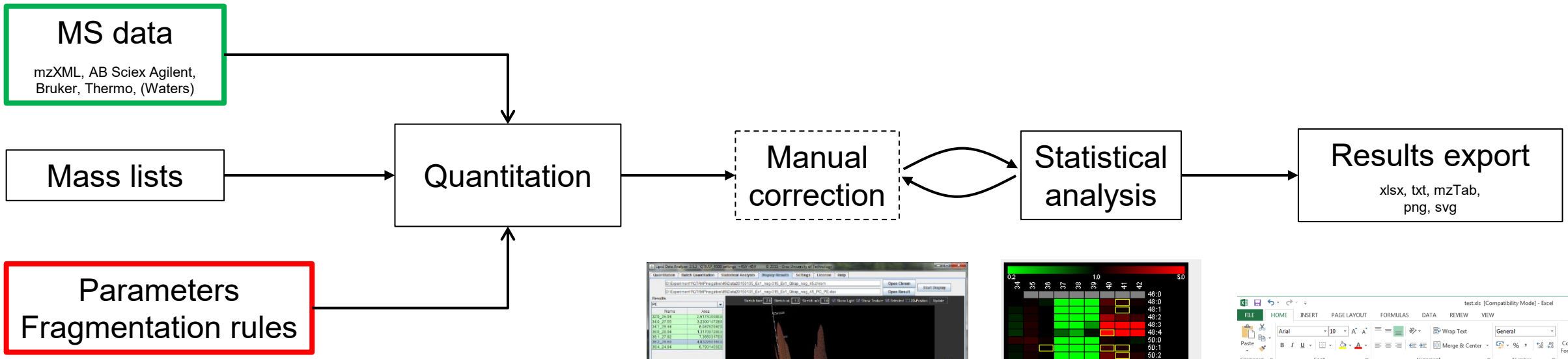


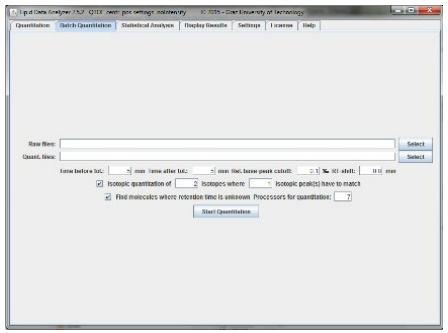
		46:0	48:0	48:1	48:2	48:3	48:4	50:0	50:1	50:2
area [AU]										
34	Value	420574.5	33165939	87125359	87944482	26433601	7922222	53685938	7.93E+08	9.95E+08
35	Value	0	38078296	1.61E+08	1.91E+08	67920033	18947560	58574859	1.42E+09	1.76E+09
36	Value	0	24585303	1.16E+08	1.34E+08	43267073	13206193	38242346	9.42E+08	1.29E+09
37	Value	0	3574343	14220103	19810777	8774378	5385055	2211651	1.02E+08	1.82E+08
38	Value	0	1975171	7131852	10704365	4109845	3611575	1571441	63002104	1.15E+08
39	Value	0	4421406	17856682	26298890	12843996	10367256	2885523	1.34E+08	2.54E+08
40	Value	0	75490844	1.92E+08	4.77E+08	6.07E+08	6.85E+08	44075006	1.75E+09	3.45E+09
41	Value	0	51632425	1.35E+08	3.38E+08	4.69E+08	5E+08	31699190	1.34E+09	2.49E+09
42	Value	0	43110038	1.25E+08	2.62E+08	3.88E+08	3.96E+08	24116508	1.04E+09	2.05E+09

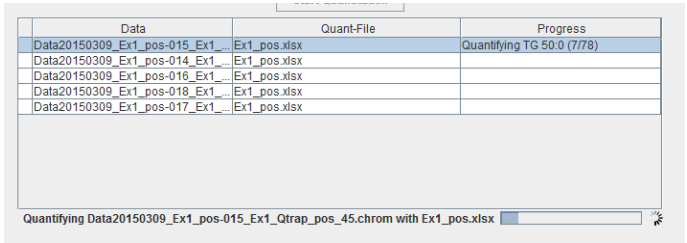
LDA – MS data conversion

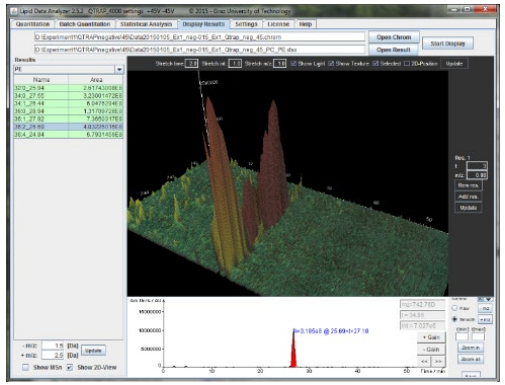


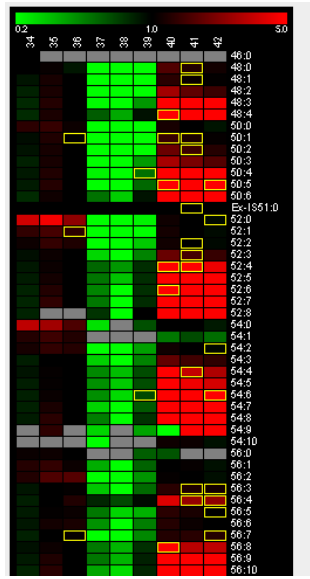
LDA – Data analysis workflow

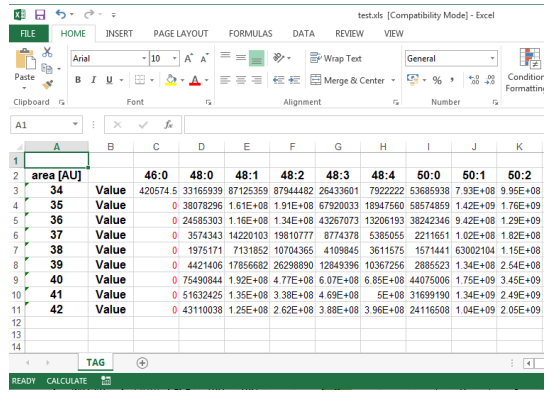






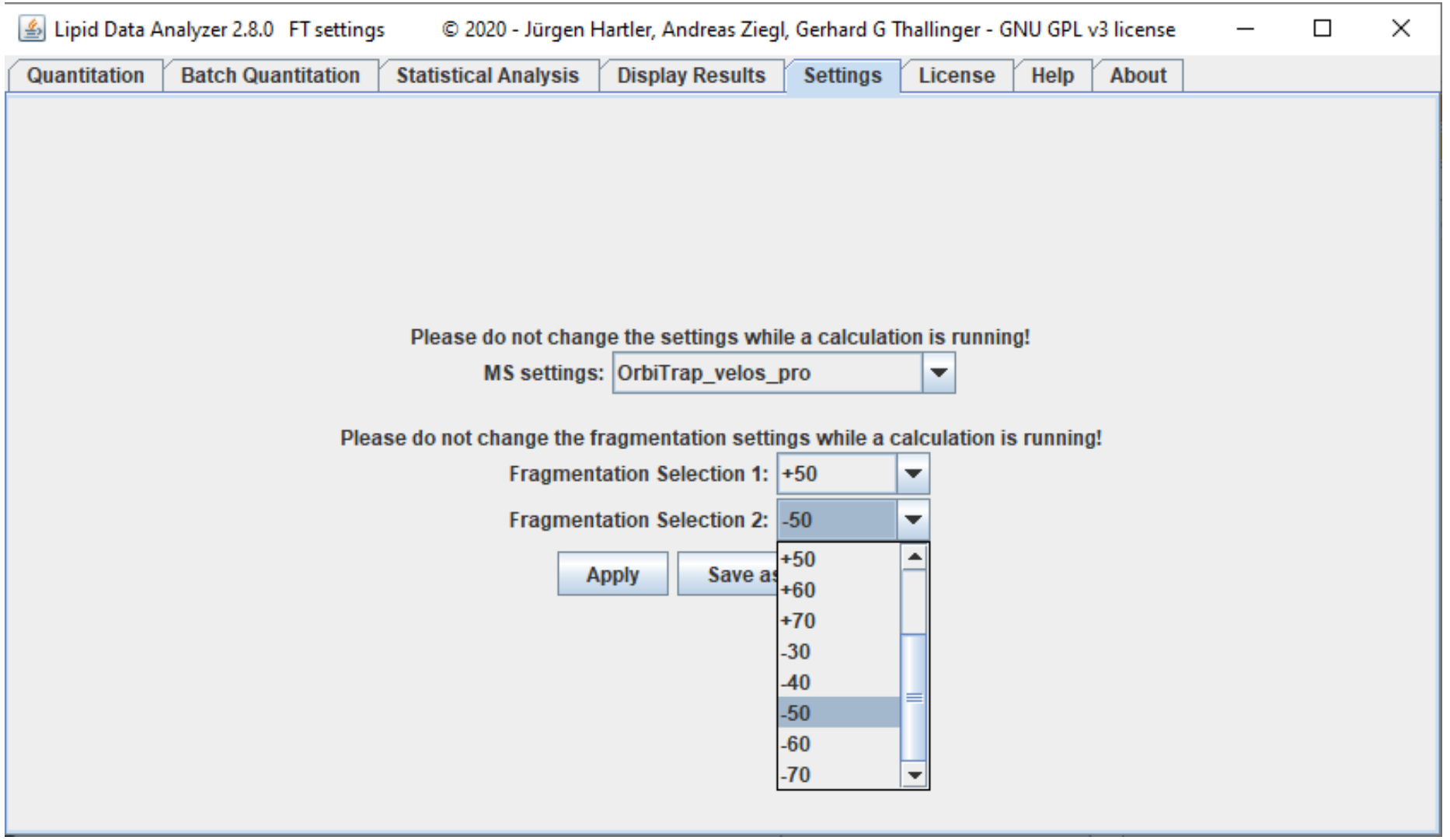






			46:0	48:0	48:1	48:2	48:3	48:4	50:0	50:1	50:2
area [AU]											
34	Value	420574.5	33165939	87125359	87944482	26433601	7922222	53685938	7.93E+08	9.95E+08
35	Value	0	38078296	1.61E+08	1.91E+08	67920033	18947560	58574859	1.42E+09	1.76E+09
36	Value	0	24585303	1.16E+08	1.34E+08	43267073	13206193	38242346	9.42E+08	1.29E+09
37	Value	0	3574343	14220103	19810777	8774378	5385055	2211651	1.02E+08	1.82E+08
38	Value	0	1975171	7131852	10704365	4109845	3611575	1571441	63002104	1.15E+08
39	Value	0	4421406	17856682	26298890	12843996	10367256	2885523	1.34E+08	2.54E+08
40	Value	0	75490844	1.92E+08	4.77E+08	6.07E+08	6.85E+08	44075006	1.75E+09	3.45E+09
41	Value	0	51632425	1.35E+08	3.38E+08	4.69E+08	5E+08	31699190	1.34E+09	2.49E+09
42	Value	0	43110038	1.25E+08	2.62E+08	3.88E+08	3.96E+08	24116508	1.04E+09	2.05E+09

LDA setup


Lipid Data Analyzer 2.8.0 FT settings
© 2020 - Jürgen Hartler, Andreas Ziegl, Gerhard G Thallinger - GNU GPL v3 license
Quantitation
Batch Quantitation
Statistical Analysis
Display Results
Settings
License
Help
About
Please do not change the settings while a calculation is running!
MS settings: OrbiTrap_velos_pro
Please do not change the fragmentation settings while a calculation is running!
Fragmentation Selection 1: +50
Fragmentation Selection 2: -50
Apply
Save as
+50
+60
+70
-30
-40
-50
-60
-70

LDA – Data analysis workflow

MS data

mzXML, AB Sciex Agilent, Bruker, Thermo, (Waters)

Mass lists

Parameters
Fragmentation rules

Quantitation

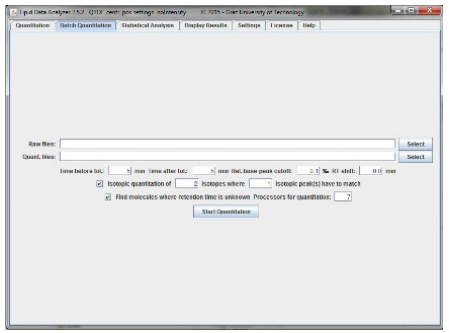
Copy mass list in quant directory

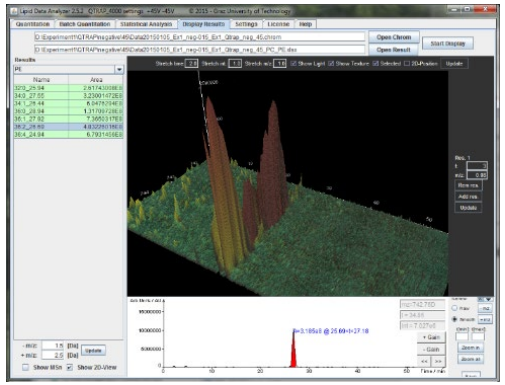
Manual correction

Statistical analysis

Results export

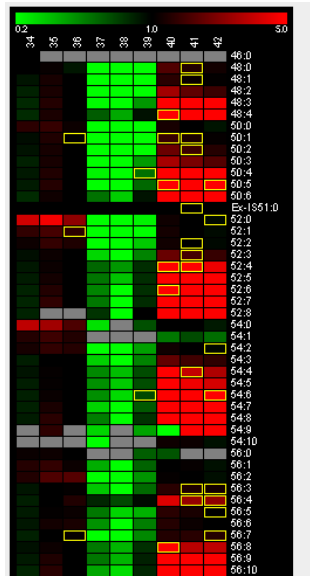
xlsx, txt, mzTab, png, svg

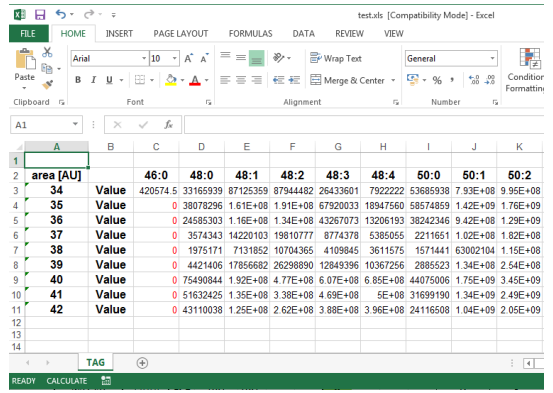




Data	Quant-File	Progress
Data20150309_Ex1_pos-015_Ex1_...	Ex1_pos.xlsx	Quantifying TG 50:0 (7/78)
Data20150309_Ex1_pos-014_Ex1_...	Ex1_pos.xlsx	
Data20150309_Ex1_pos-016_Ex1_...	Ex1_pos.xlsx	
Data20150309_Ex1_pos-018_Ex1_...	Ex1_pos.xlsx	
Data20150309_Ex1_pos-017_Ex1_...	Ex1_pos.xlsx	

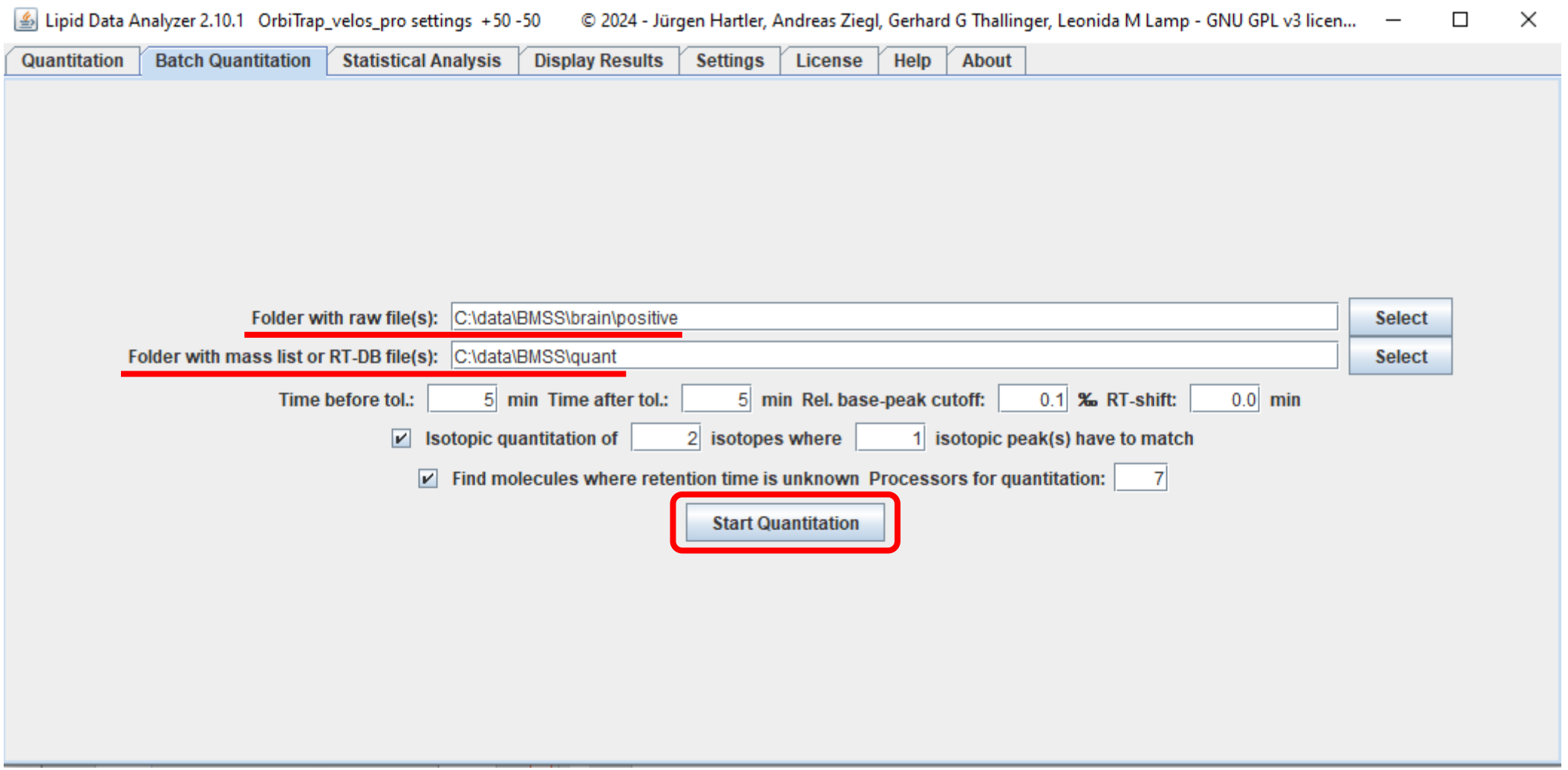
Quantifying Data20150309_Ex1_pos-015_Ex1_Qtrap_pos_45.chrom with Ex1_pos.xlsx



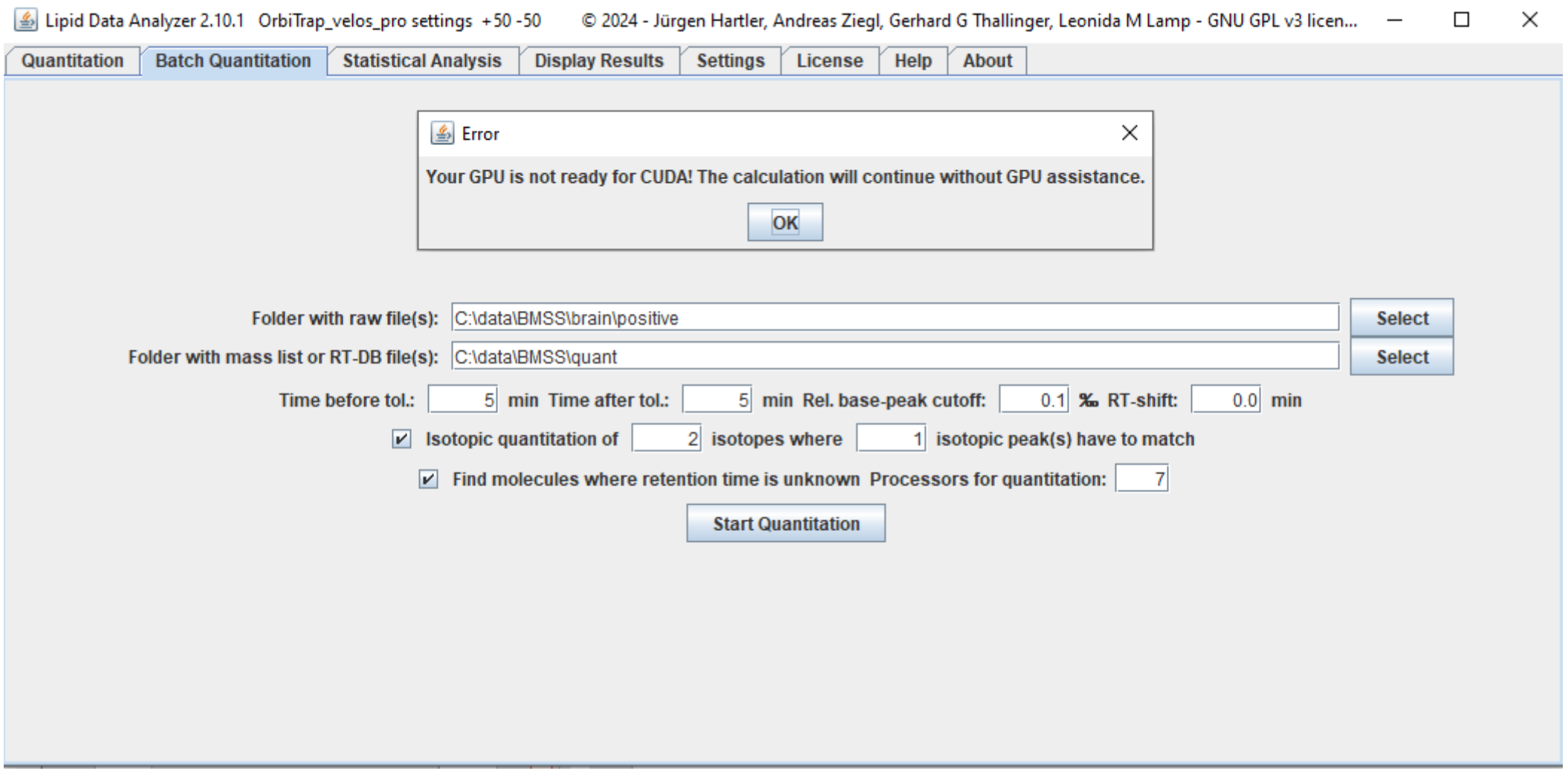


			46:0	48:0	48:1	48:2	48:3	48:4	50:0	50:1	50:2
area [AU]											
34	Value	420574.5	33165939	87125359	87944482	26433601	7922222	53685938	7.93E+08	9.95E+08
35	Value	0	38078296	1.61E+08	1.91E+08	67920033	18947560	58574859	1.42E+09	1.76E+09
36	Value	0	24585303	1.16E+08	1.34E+08	43267073	13206193	38242346	9.42E+08	1.29E+09
37	Value	0	3574343	14220103	19810777	8774378	5385055	2211651	1.02E+08	1.82E+08
38	Value	0	1975171	7131852	10704365	4109845	3611575	1571441	63002104	1.15E+08
39	Value	0	4421406	17856682	26298890	12843996	10367256	2885523	1.34E+08	2.54E+08
40	Value	0	75490844	1.92E+08	4.77E+08	6.07E+08	6.85E+08	44075006	1.75E+09	3.45E+09
41	Value	0	51632425	1.35E+08	3.38E+08	4.69E+08	5E+08	31699190	1.34E+09	2.49E+09
42	Value	0	43110038	1.25E+08	2.62E+08	3.88E+08	3.96E+08	24116508	1.04E+09	2.05E+09

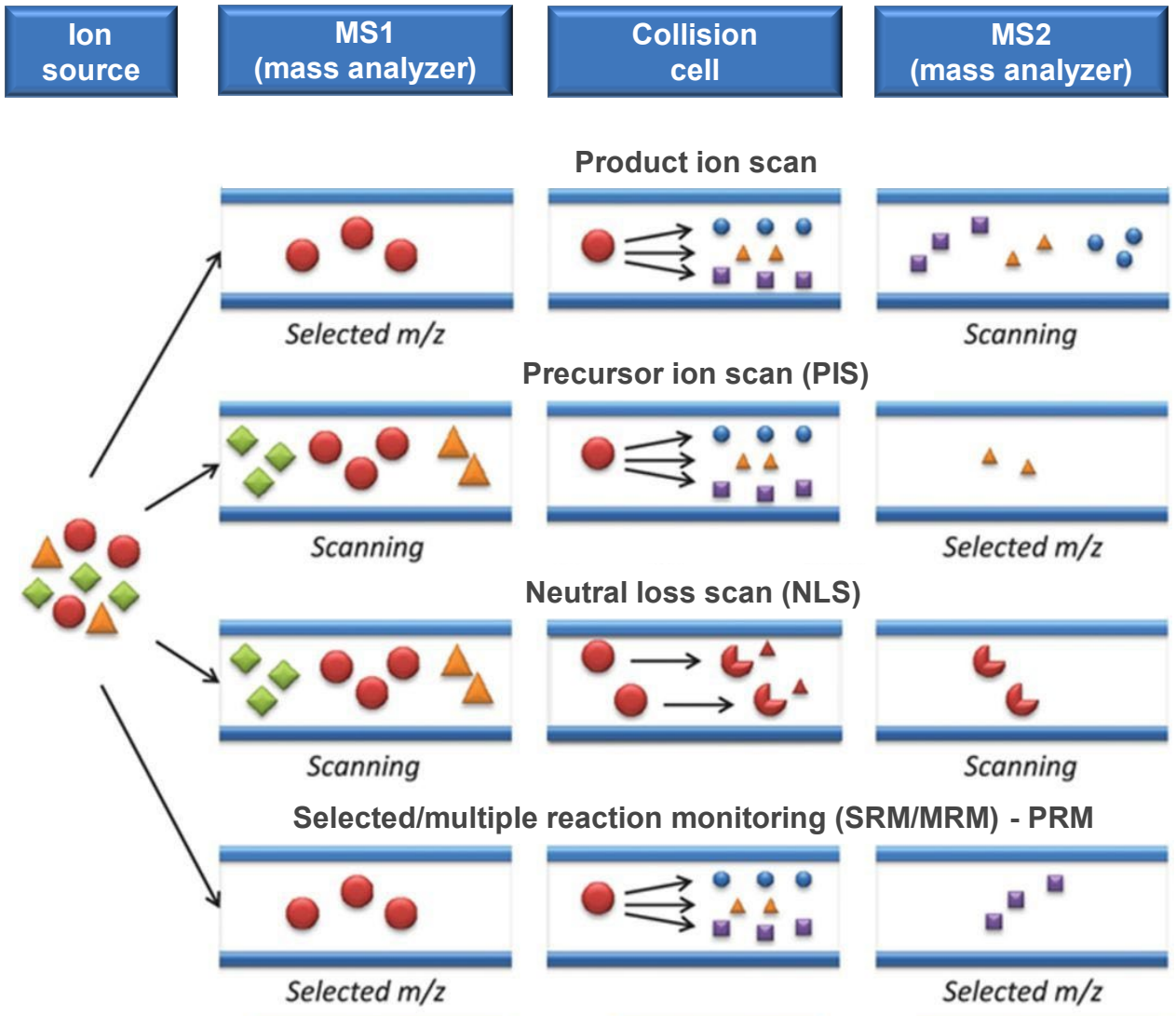
LDA – Start batch quantitation



LDA – Start batch quantitation



Targeted vs untargeted MS data acquisition: differing MS/MS modes



Fragmentation of selected precursor ions

- Data dependent acquisition (DDA): typically, the 4-10 most abundant analytes are selected for fragmentation in one cycle
- Scanning for several precursors takes time → first analyzer is occupied → reduced sampling rates (inaccuracy in quantitation)

Precursor ions with a specified product ion

- Filters analytes of specific types → for investigation of certain analyte classes
- Analytes of low abundance are not missed
- Misses information of additional fragment ions

Precursor ions with a specified neutral-loss fragment

- Filters analytes of specific types → for investigation of certain analyte classes
- Analytes of low abundance are not missed
- Misses information of additional fragment ions

Ions with a specified precursor / fragment m/z pair (transition)

- Highly sensitive filter - for observing specific analytes
- Very low abundant molecules are detectable
- Misses information of additional fragment ions

Helmut H. *et al.* Lipidomics: Quest for Molecular Lipid Biomarkers in Cardiovascular Disease, *Circulation: Cardiovascular Genetics* **7**, 6 (2014). https://doi.org/10.1161/CIRCGENETICS.114.000550

Targeted vs untargeted lipidomics: overview

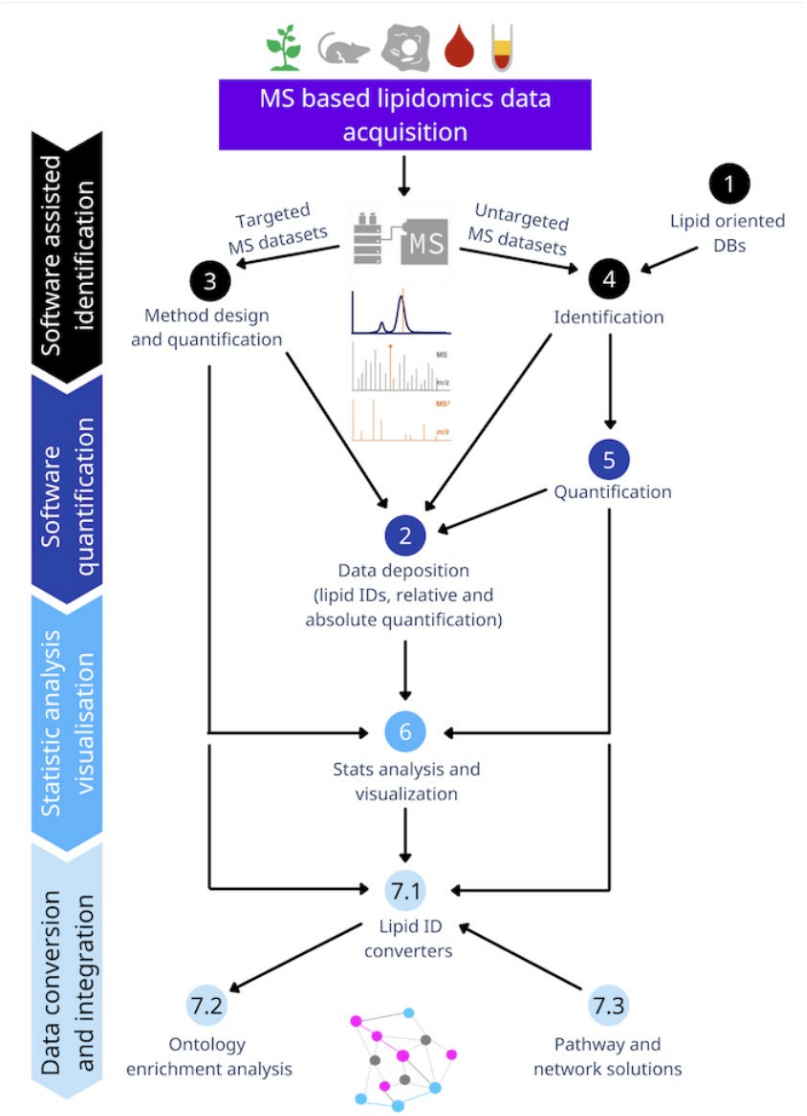
Targeted

- **Scope:** Focus on a specific set of known compounds.
- **Purpose:** Hypothesis-driven, used for validating known metabolic processes and establishing baseline measurements.
- **Quantification**: Provides absolute quantification of selected compounds.
- **Sample preparation**: Requires specific extraction procedures for targeted compounds.
- **MS data acquisition:** MS/MS techniques such as precursor ion scan (PIS), neutral loss scan (NLS) and selected/multiple reaction monitoring (SRM/MRM)
- **Data processing**: Less complex due to the limited search space.

Untargeted

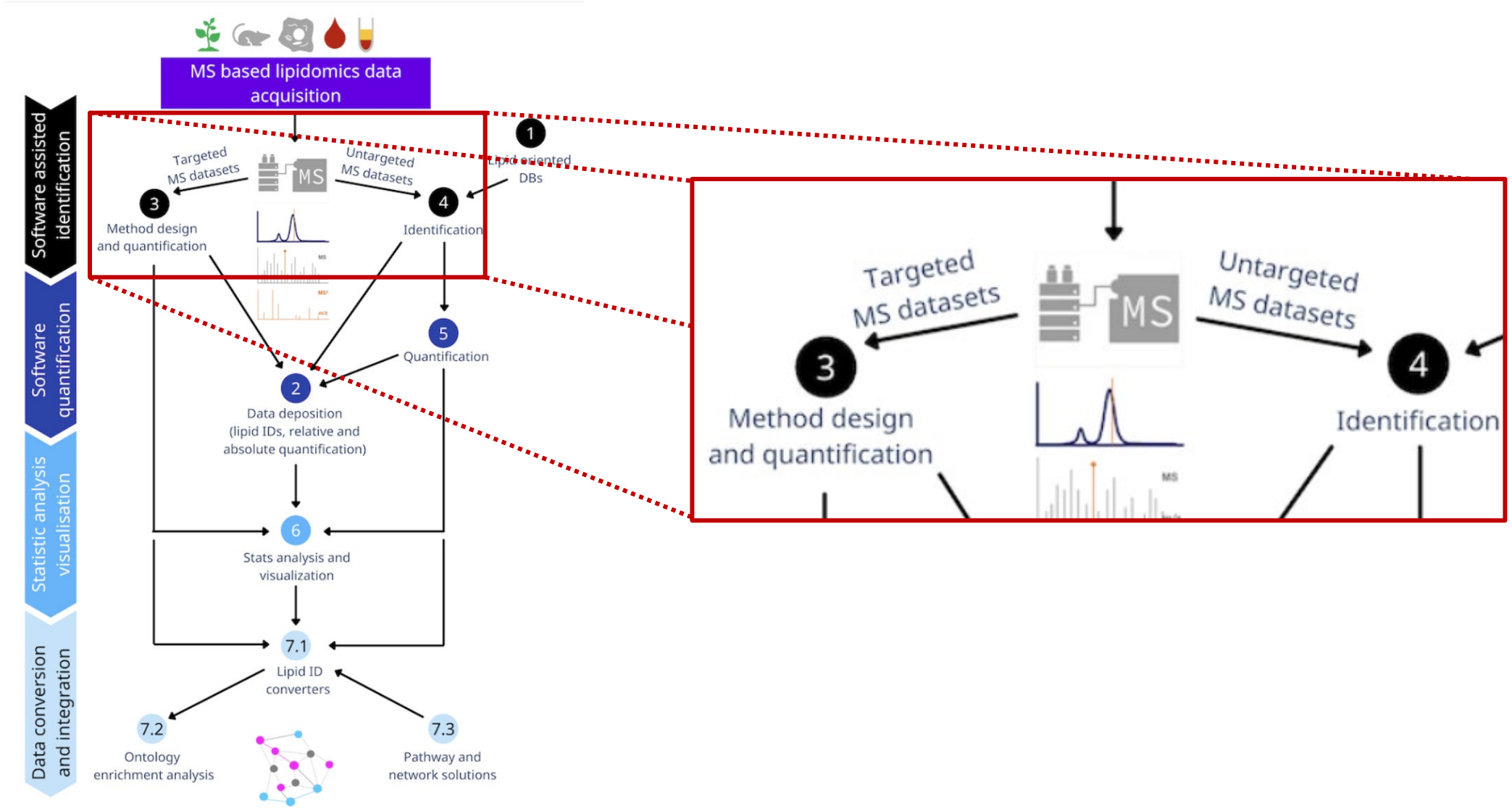
- **Scope:** Global, large scale analysis of both, known and unknown compounds in a sample.
- **Purpose:** Discovery-oriented, used for identifying new compounds, generating hypotheses, and biomarker discovery.
- **Quantification**: Provides qualitative identification and relative quantification of a wide range of compounds.
- **Sample preparation**: Requires global metabolite extraction procedures.
- **MS data acquisition:** DDA approaches (product ion scan)
- **Data processing**: More complex due to the large search space and the unknown complexity of the sample.

Guiding the choice of informatics software and tools for lipidomics research



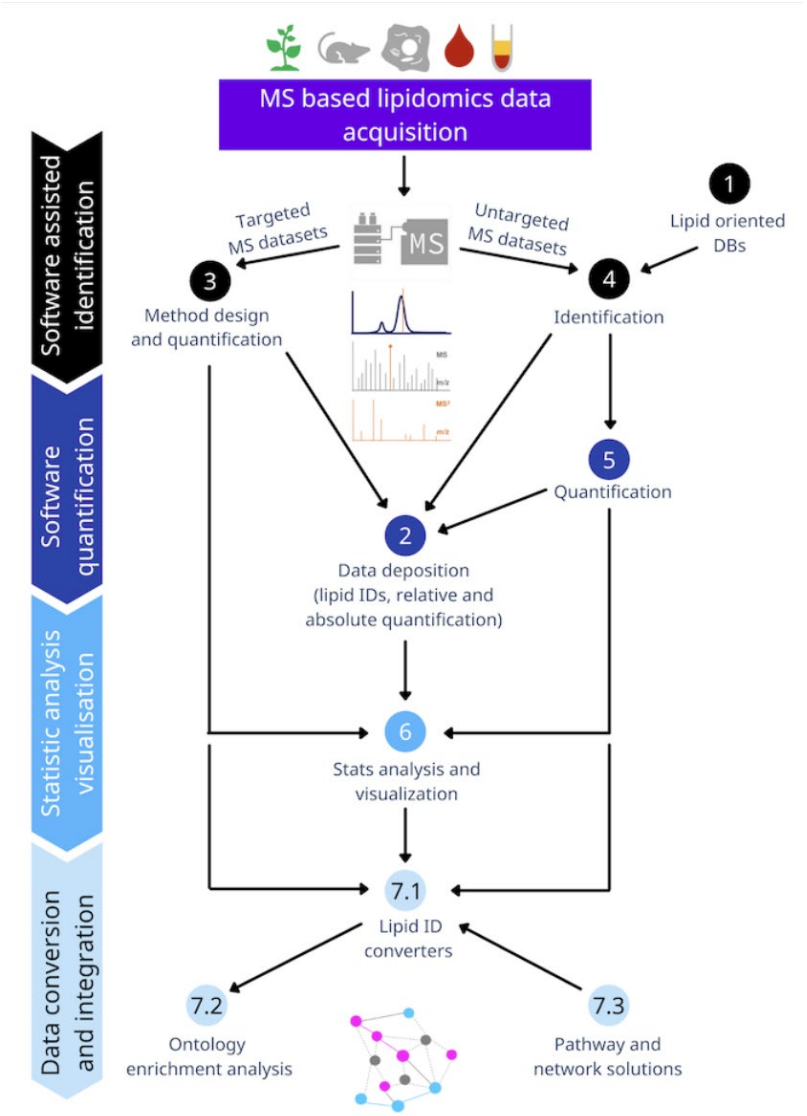
Ni, Z., Wölk, M., Jukes, G. *et al.* Guiding the choice of informatics software and tools for lipidomics research applications. *Nat Methods* **20**, 193–204 (2023). https://doi.org/10.1038/s41592-022-01710-0

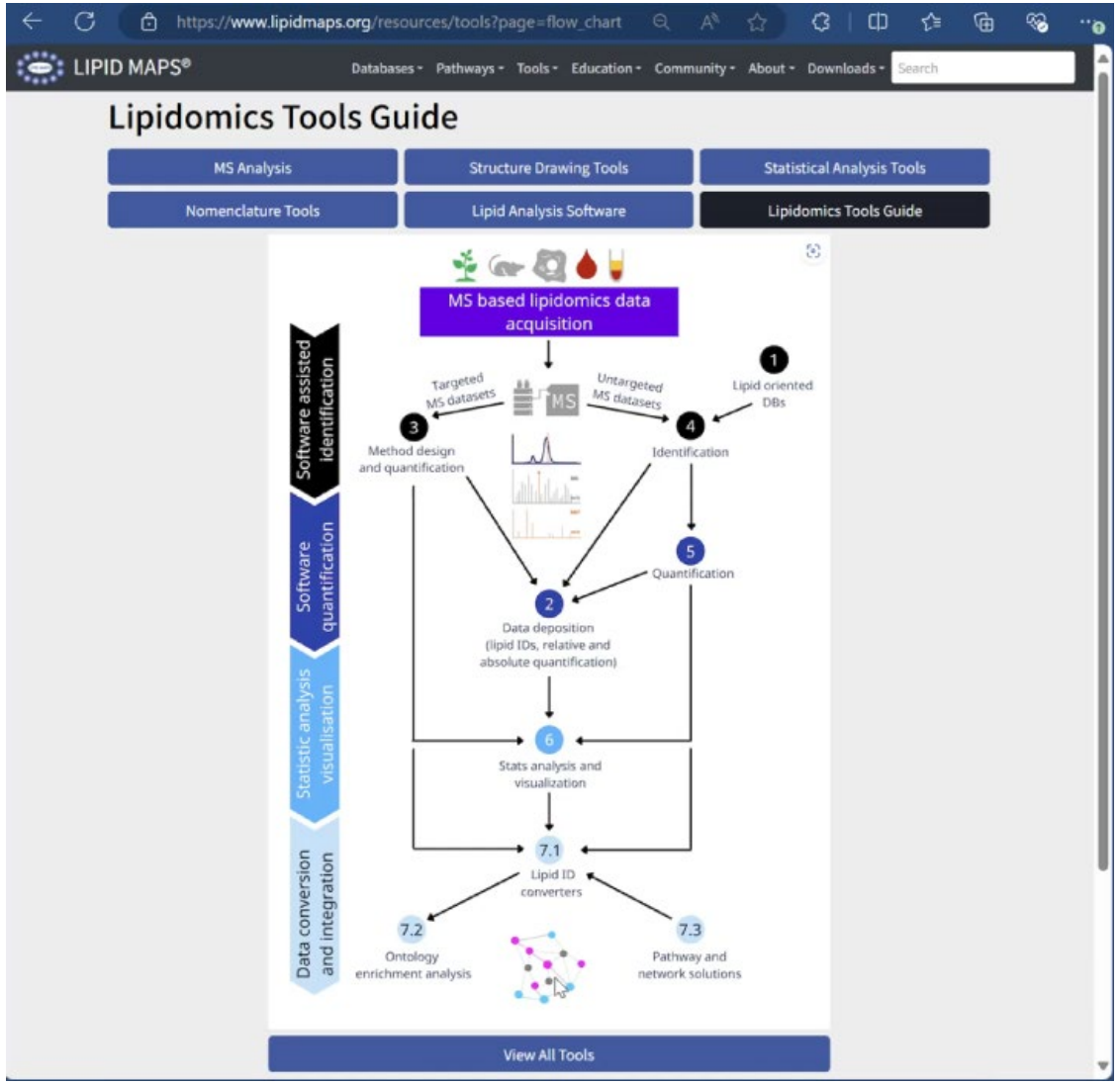
Guiding the choice of informatics software and tools for lipidomics research



Ni, Z., Wölk, M., Jukes, G. *et al.* Guiding the choice of informatics software and tools for lipidomics research applications. *Nat Methods* **20**, 193–204 (2023). https://doi.org/10.1038/s41592-022-01710-0

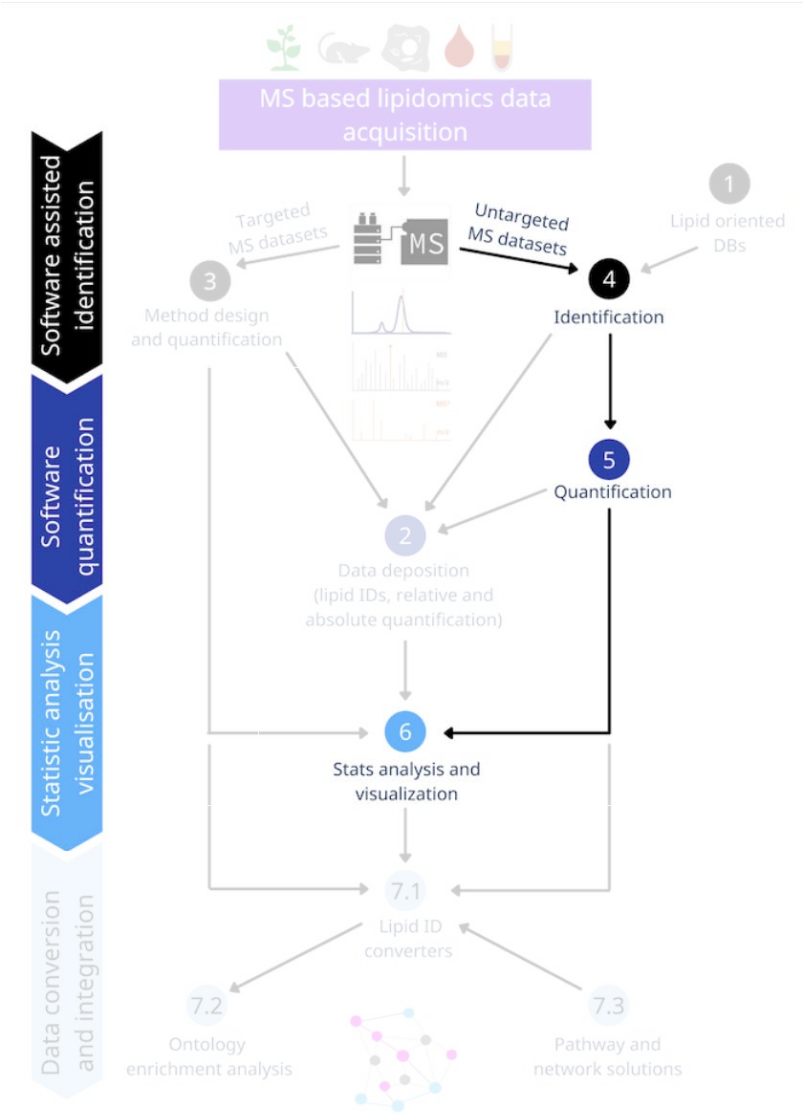
Guiding the choice of informatics software and tools for lipidomics research

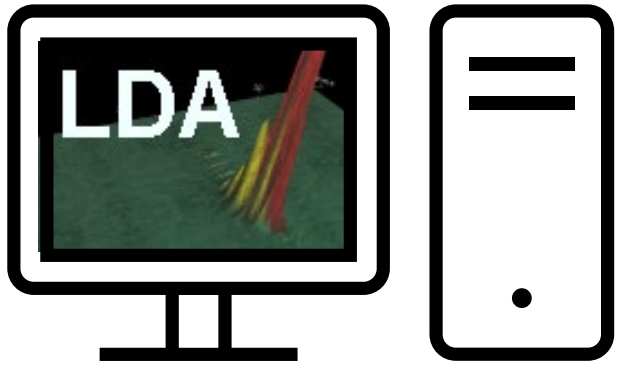




Ni, Z., Wölk, M., Jukes, G. *et al.* Guiding the choice of informatics software and tools for lipidomics research applications. *Nat Methods* **20**, 193–204 (2023). https://doi.org/10.1038/s41592-022-01710-0

Guiding the choice of informatics software and tools for lipidomics research





- **Lipid identification**
- **Quantification**
- **Statistical analysis and visualization**

=> For untargeted MS data!

Ni, Z., Wölk, M., Jukes, G. *et al.* Guiding the choice of informatics software and tools for lipidomics research applications. *Nat Methods* **20**, 193–204 (2023). https://doi.org/10.1038/s41592-022-01710-0

LDA core functions

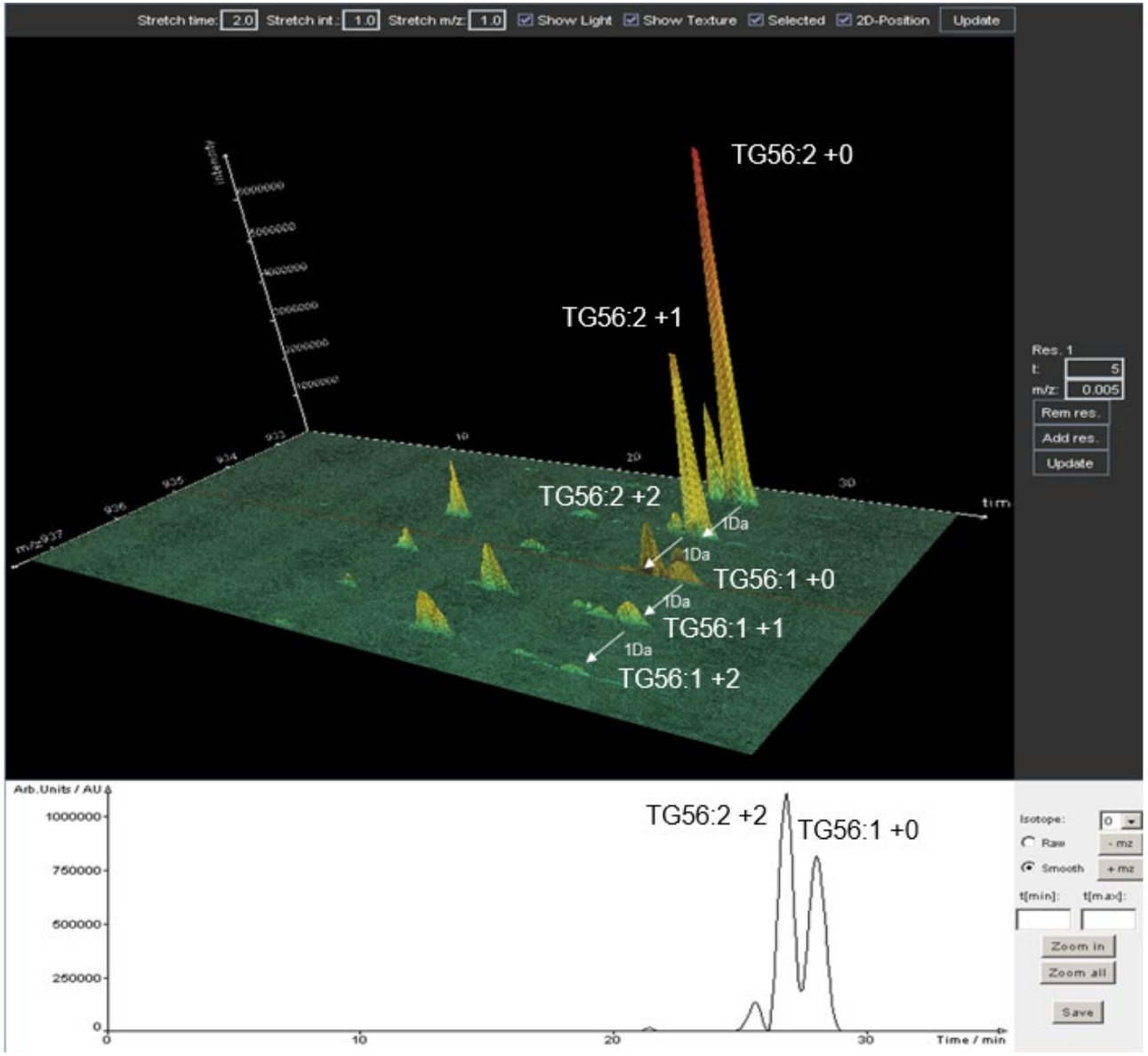
LC-MS^1 data:

Identification and quantitation of intact lipid species

MS^n data:

Verification of lipid class and structural information about the lipid molecular species

Challenges MS¹ – results of a triglyceride analysis



TG 46:0	TG 50:2	TG 52:3	TG 54:2	TG 54:10
TG 48:0	TG 50:3	TG 52:4	TG 54:3	TG 56:0
TG 48:1	TG 50:4	TG 52:5	TG 54:4	TG 56:1
TG 48:2	TG 50:5	TG 52:6	TG 54:5	TG 56:2
TG 48:3	TG 50:6	TG 52:7	TG 54:6	TG 56:3
TG 48:4	TG 52:0	TG 52:8	TG 54:7	TG 56:4
TG 50:0	TG 52:1	TG 54:0	TG 54:8	TG 56:5
TG 50:1	TG 52:2	TG 54:1	TG 54:9	TG 56:6

TG 56:7	TG 58:2	TG 58:10	TG 60:5	TG 60:13
TG 56:8	TG 58:3	TG 58:11	TG 60:6	TG 60:14
TG 56:9	TG 58:4	TG 58:12	TG 60:7	TG 62:11
TG 56:10	TG 58:5	TG 58:13	TG 60:8	TG 62:12
TG 56:11	TG 58:6	TG 60:1	TG 60:9	TG 62:13
TG 56:12	TG 58:7	TG 60:2	TG 60:10	TG 62:14
TG 58:0	TG 58:8	TG 60:3	TG 60:11	TG 62:15
TG 58:1	TG 58:9	TG 60:4	TG 60:12	

Challenges MS^1 – check of isotopic pattern

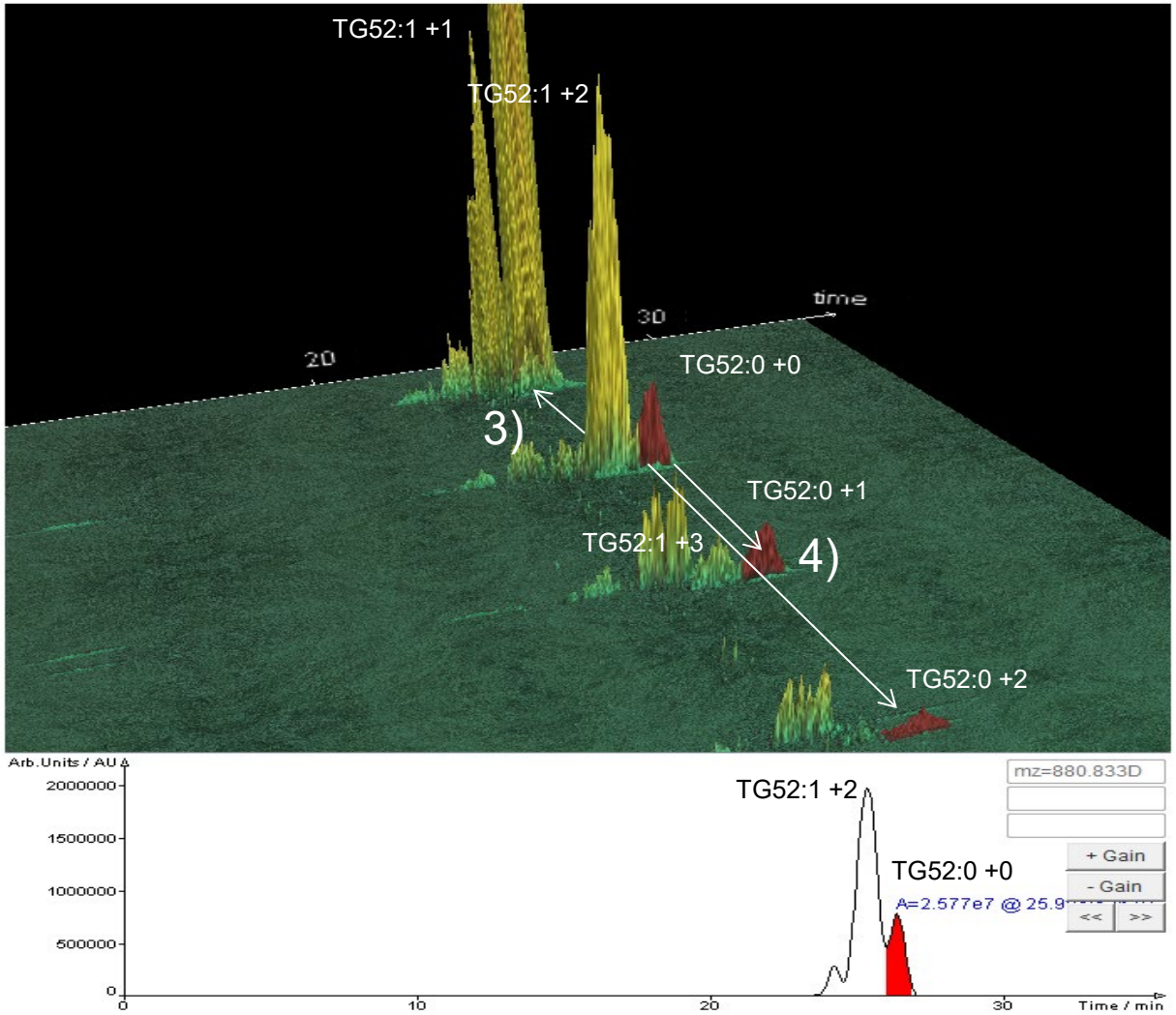
Algorithm:

1) Calculate isotopic distributions

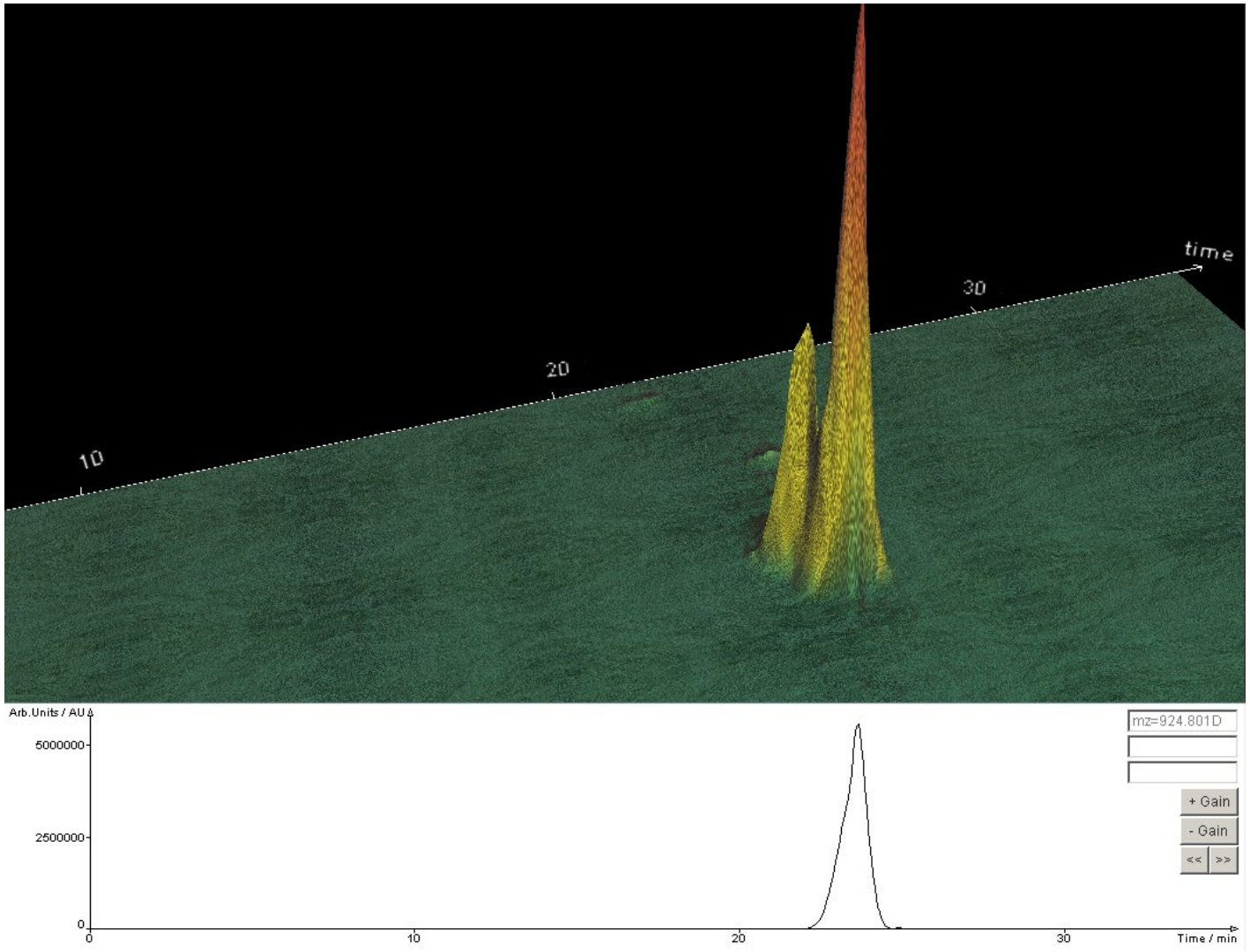
2) Calculate peak intensities

3) Check if peak belongs to another isotope

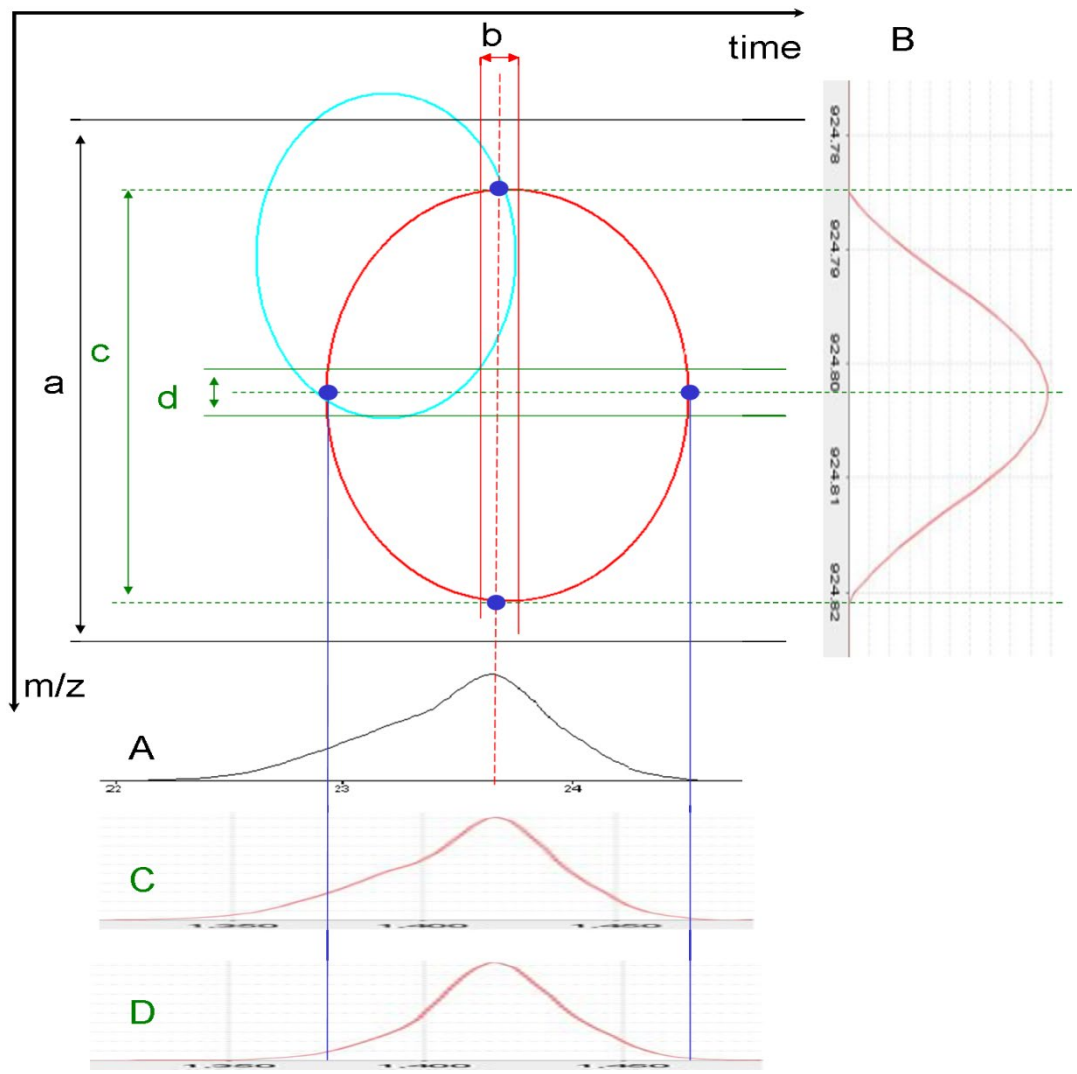
4) Check if isotopic series is correct

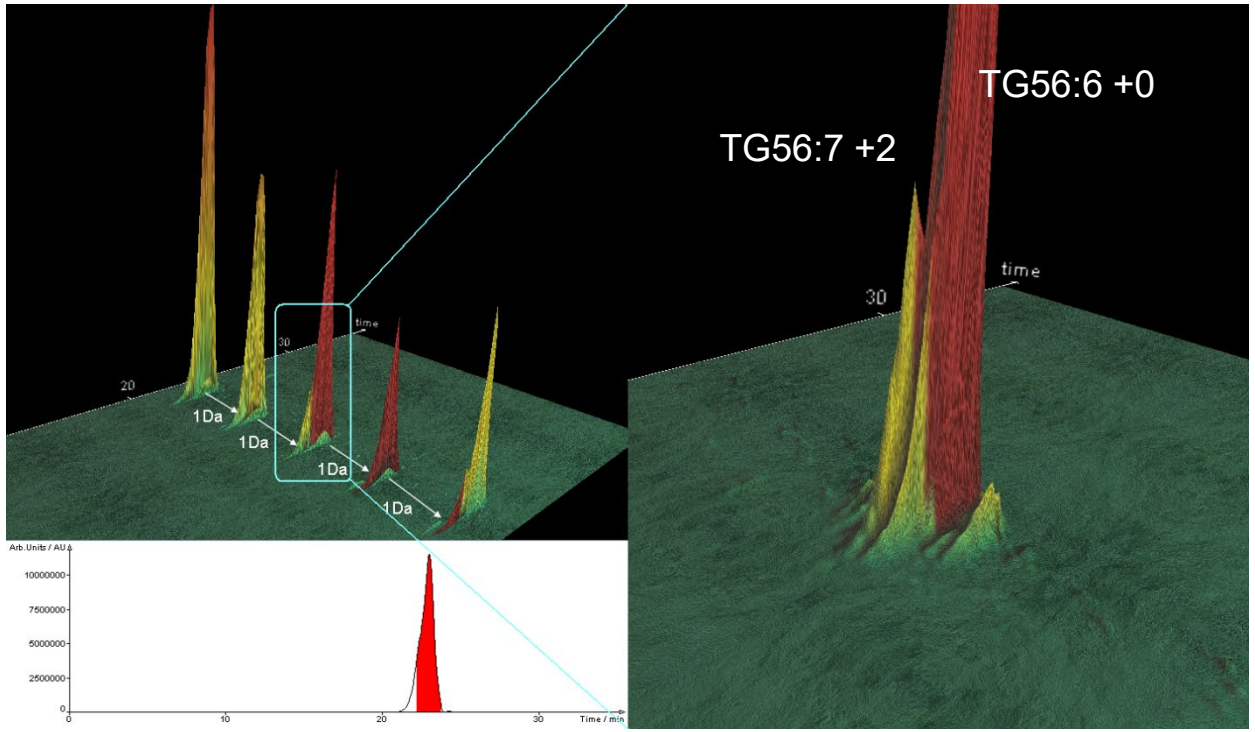


Challenges MS^1 – closely eluting species



Challenges MS^1 – 3D algorithm

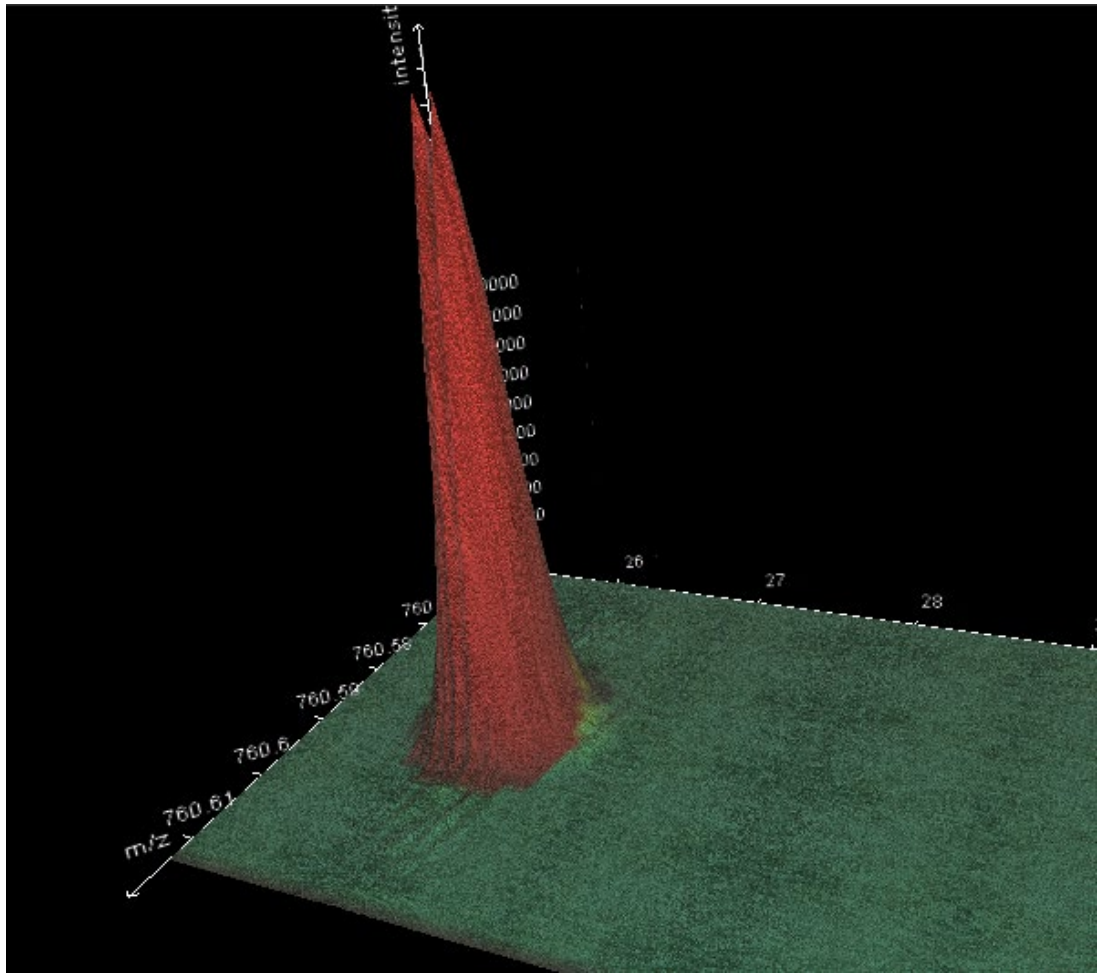


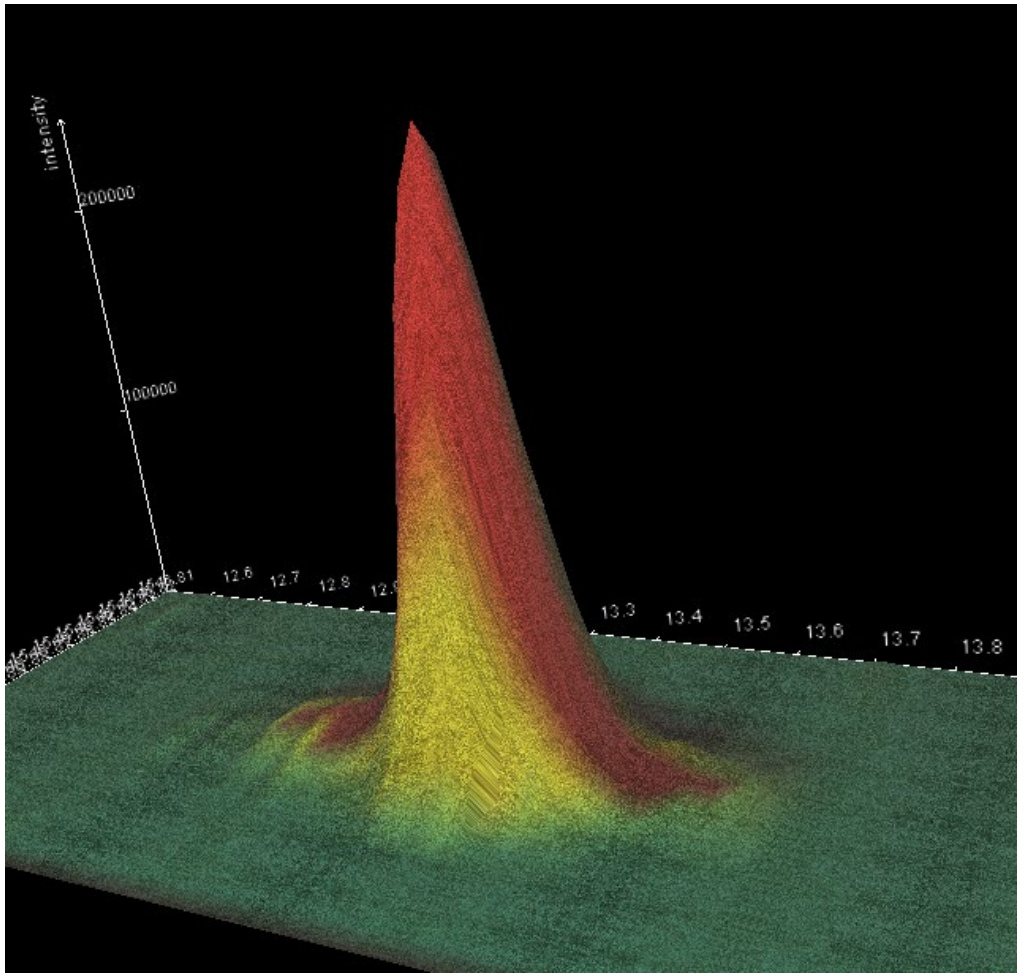


Dynamic range: 10^6	LDA	mzMine2
Sensitivity	93.3%	86.1%
Sensitivity (unambiguous)	90.4%	63.0%
Positive predictive value	99.1%	89.4%

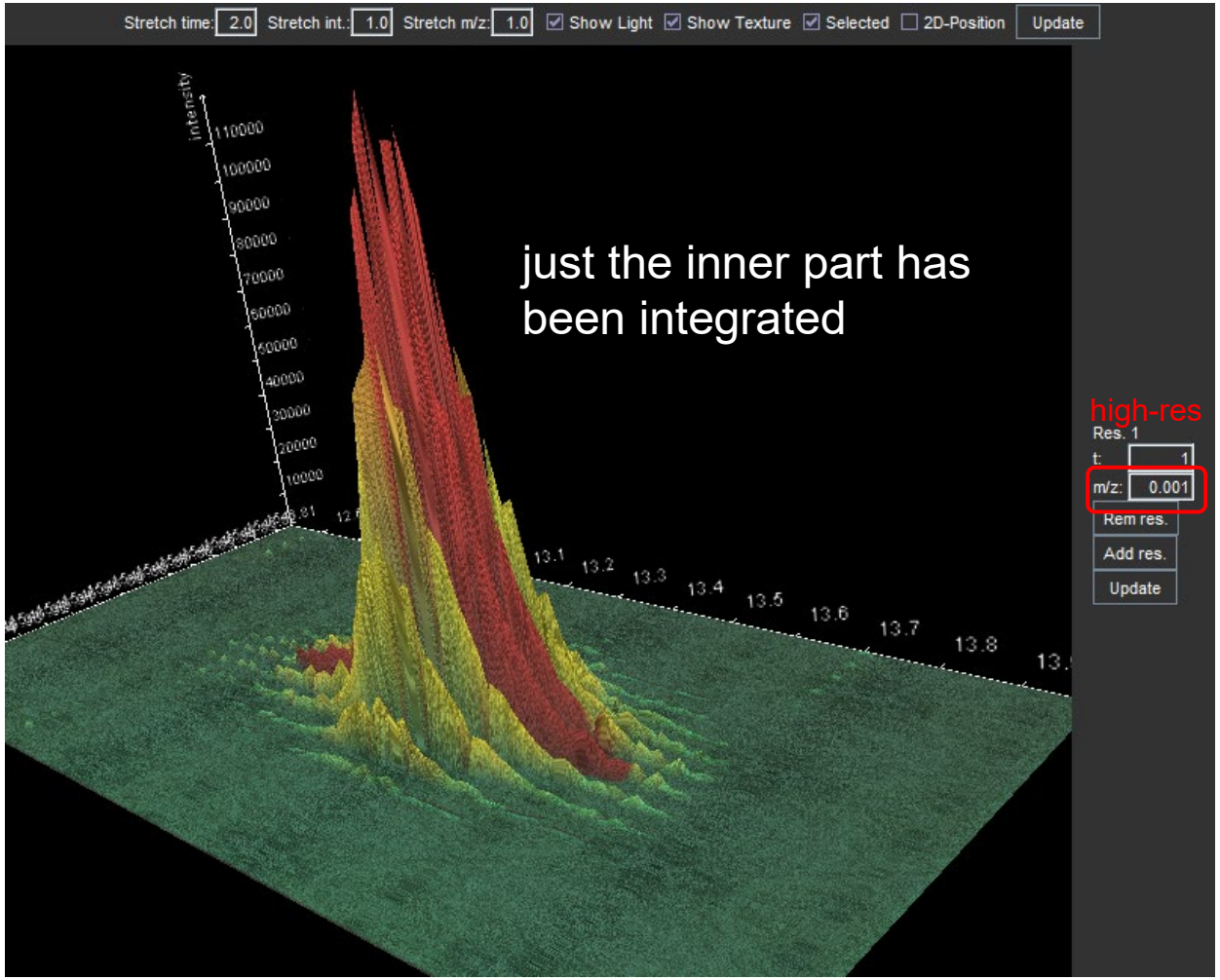
Challenges MS^1 – 3D algorithm – peak integration

not well-integrated





Challenges MS^1 – 3D algorithm – peak integration



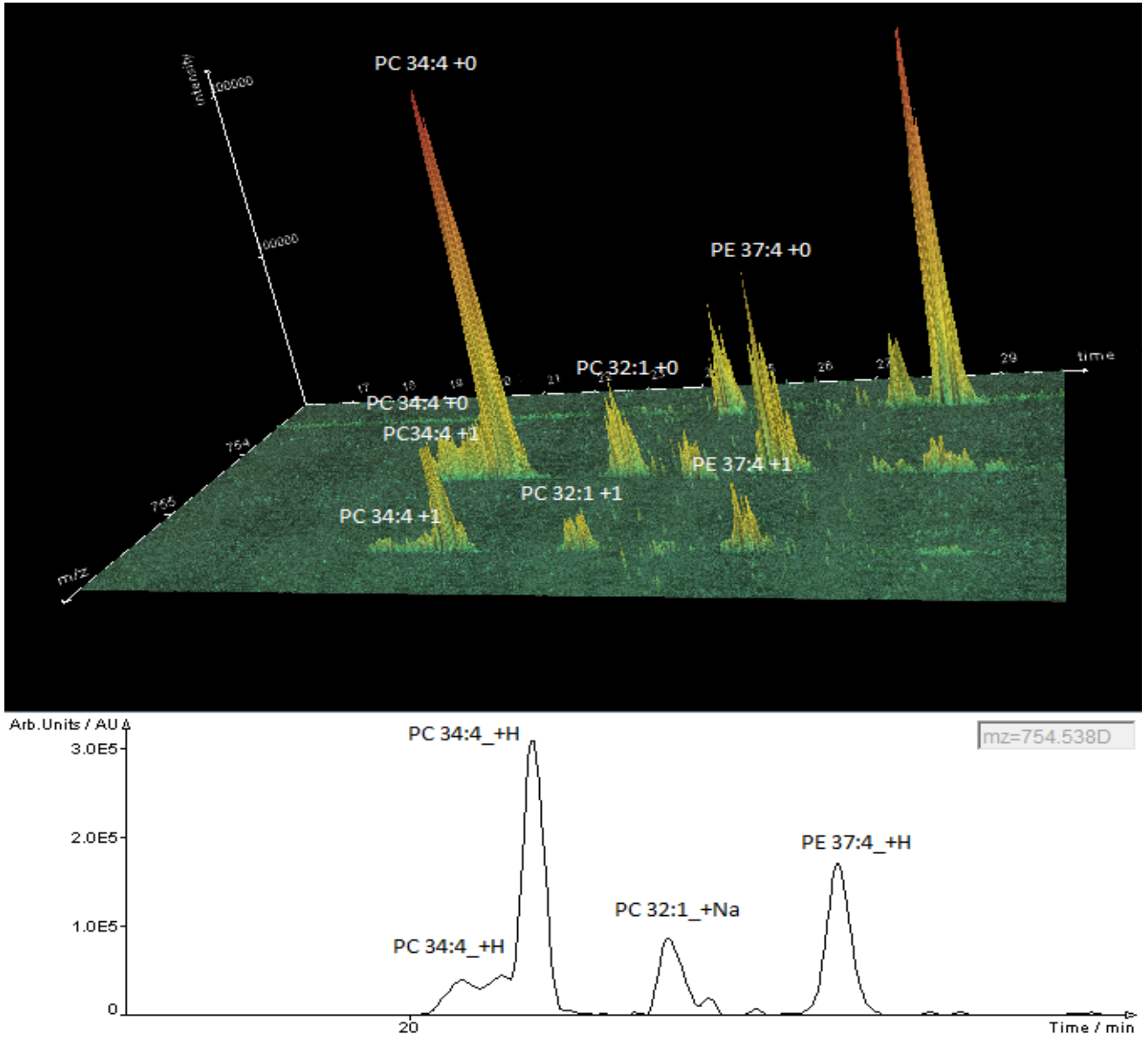
Check this for:

- Strong peak signal
- Weak peak signal
- Low *m/z* value
- High *m/z* value

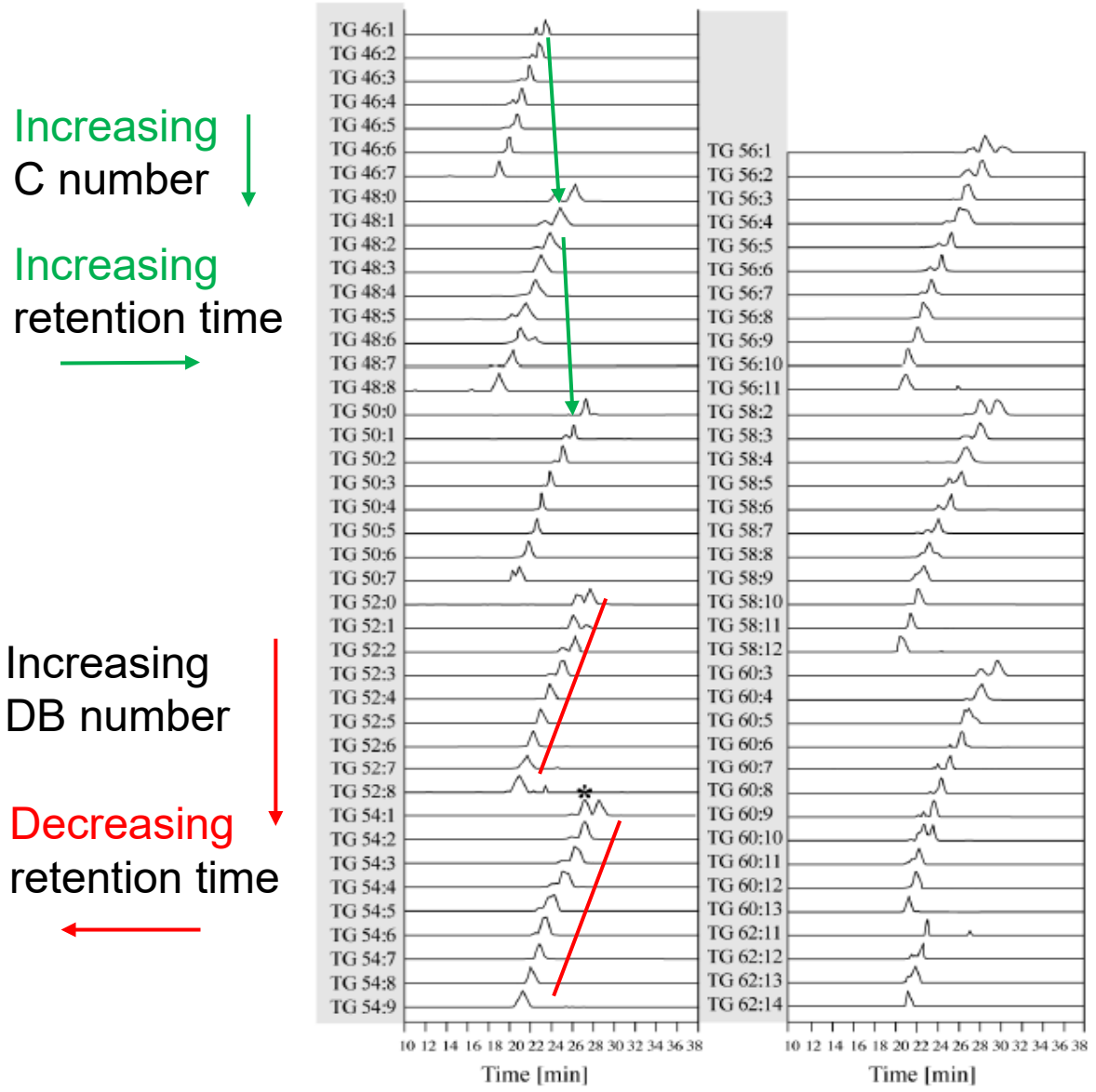
If the integration is not satisfactory:

Get in touch with us!
We will optimize the settings for you

Challenges MS^1 – isomeric/isobaric compounds



Challenges MS^1 – ECN model

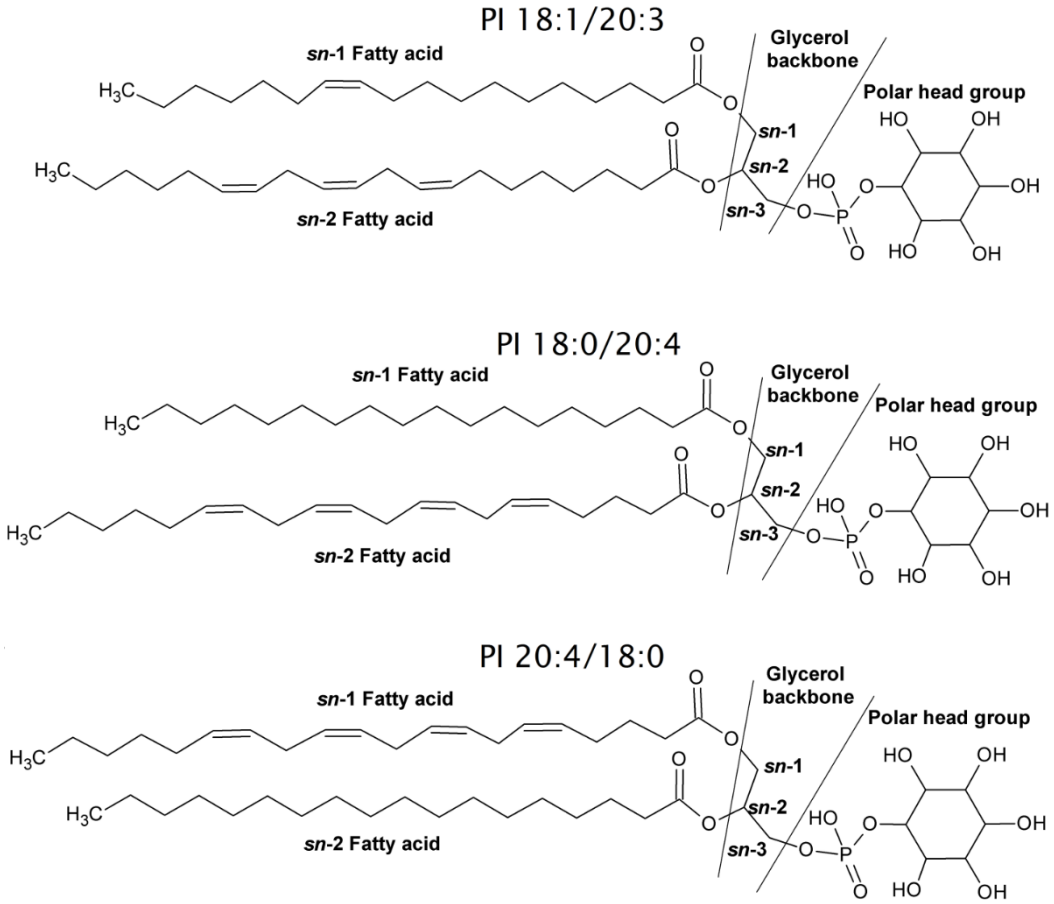


Fauland et al., 2011, *J Lipid Res* 52(12):2314-22

Ovcacikova, et al., 2016, *J Chromatogr A* 1450(12):76-85

Challenges MS^1 – MS^1 misses structural information

PI 38:4

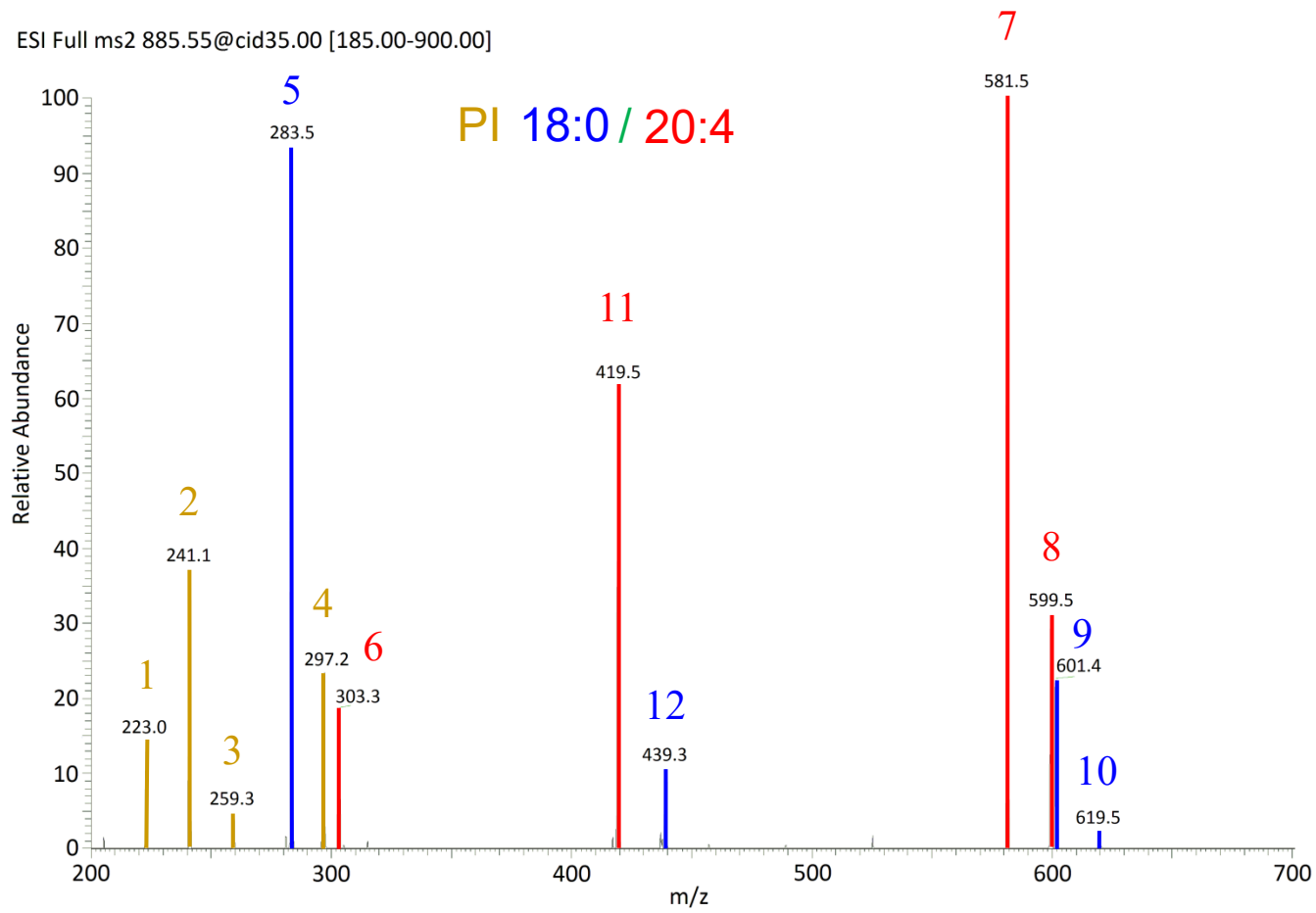


We do not know:

- Fatty acid composition
- Position of FAs
- Position of double bonds

Challenges MS^n – Introduction

PI 38:4 – negative ion mode



1-4: head group

5,6: carboxylates of chains

7,9: neutral loss of chain as fatty acid

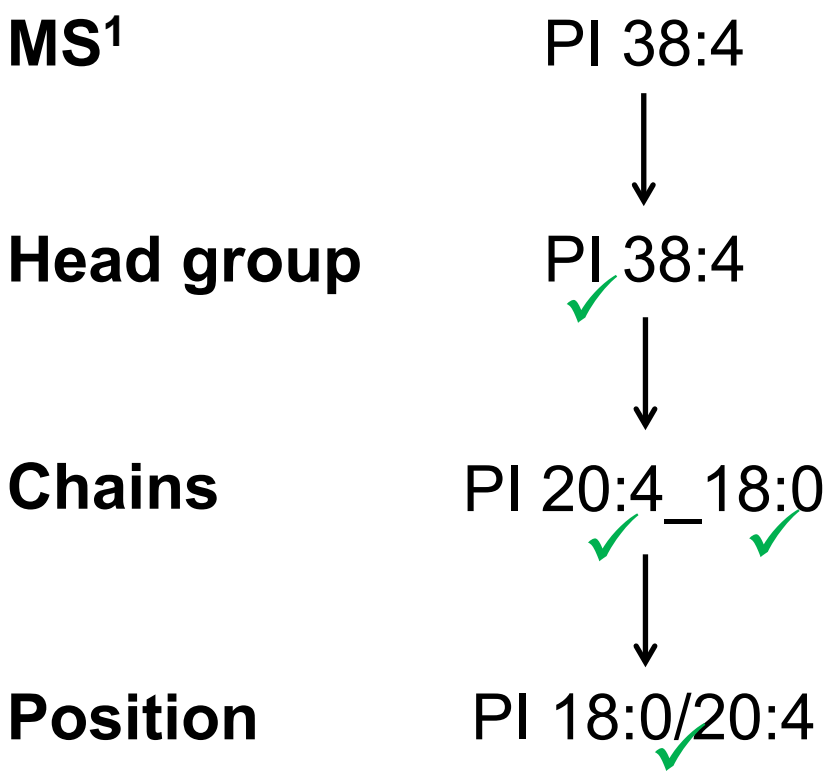
8,10: neutral loss of chain as ketene body

11,12: neutral loss of chain and head group

- charge driven processes favor NL at *sn-2*
- secondary fragmentation processes favor the formation of carboxylates of FA at *sn-1*

Hsu FF and Turk J: **Electrospray Ionization with Low-Energy Collisionally Activated Dissociation Tandem Mass Spectrometry of Complex Lipids: Structural Characterization and Mechanisms of Fragmentation.** In *Modern Methods for Lipid Analysis by Liquid Chromatography-Mass Spectrometry and Related Techniques*. Edited by Byrdwell WC. Champaign, IL: American Oil Chemists Society Press; 2005:61-179.

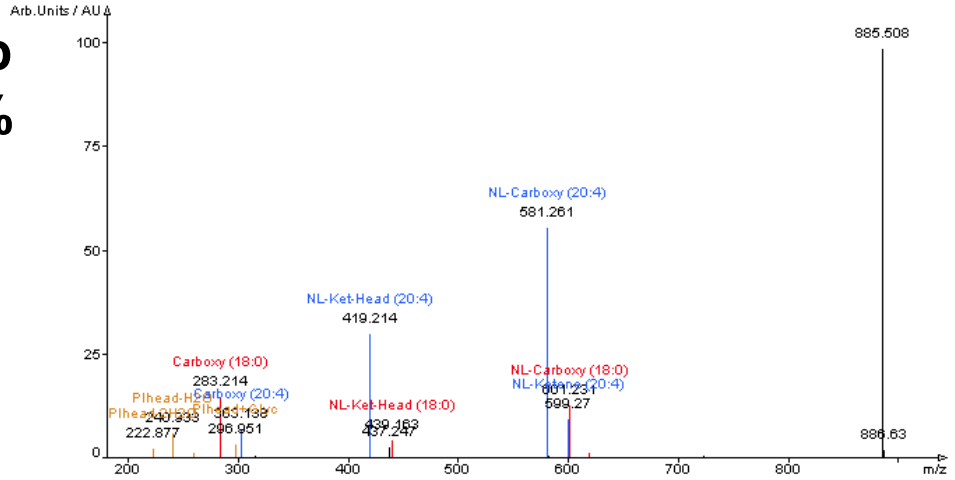
Challenges MS^n – Introduction



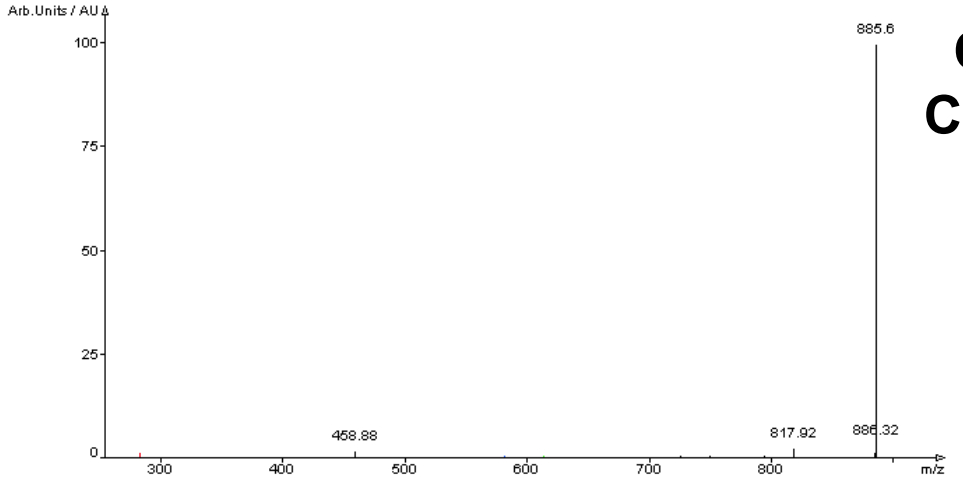
Fragments (m/z) + Intensity

Challenges MS^n – MS^n spectra may vary

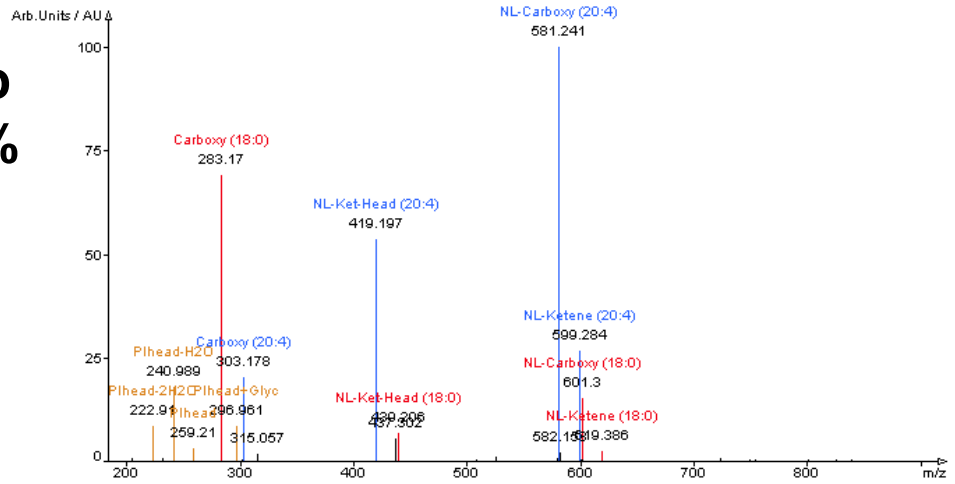
Orbitrap CID 30%



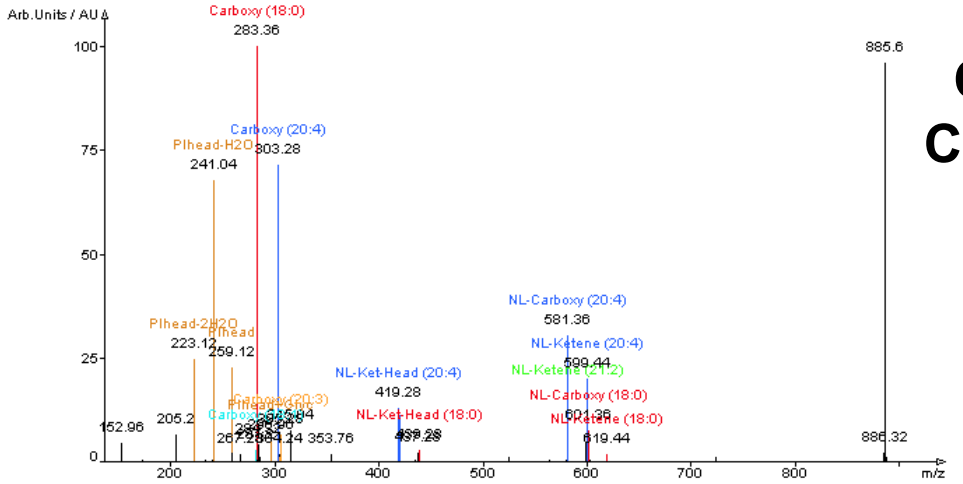
QTRAP CID 30 eV



Orbitrap CID 60%



QTRAP CID 60 eV



A flexible solution is required

Decision rule sets - fragments

Fragment:

Name=PIhead Formula=C6H12O9P Charge=1 MSLevel=2 mandatory=false

Placeholders:

Name=NL-Ketene Formula=$PRECURSOR-$CHAIN+H2O Charge=1 MSLevel=2 mandatory=false

Reuse of definitions:

Name=NL-Ket-Head Formula=NL-Ketene-C6H12O6 Charge=1 MSLevel=2 mandatory=false

Decision rule sets - intensities

Relative intensities:

Equation=PIhead-H2O>2*PIhead mandatory=true

Placeholders:

Equation=PIhead-H2O>0.1*$BASEPEAK mandatory=true

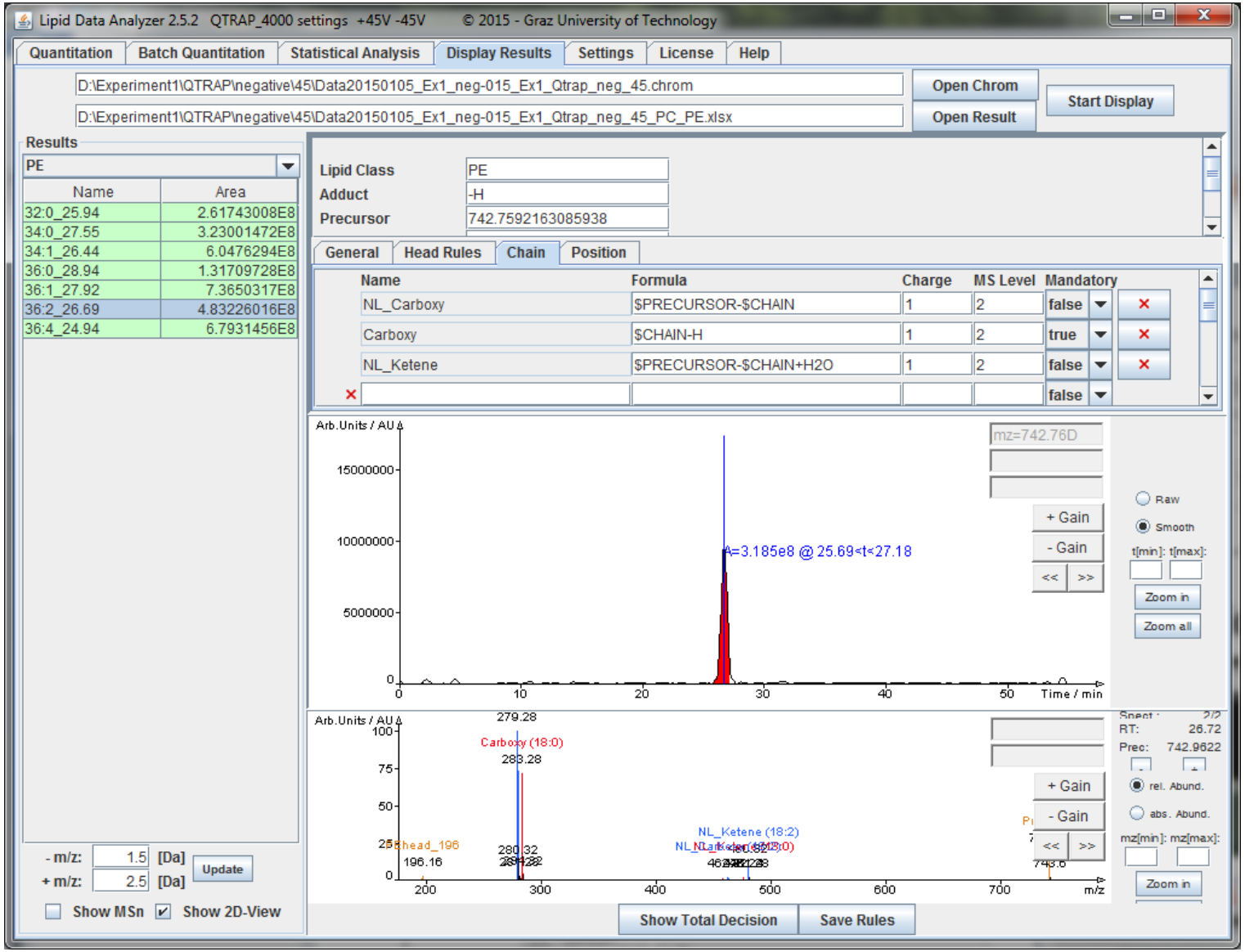
Positions:

Equation=NL-Ketene[2]*0.7>NL-Ketene[1] mandatory=true

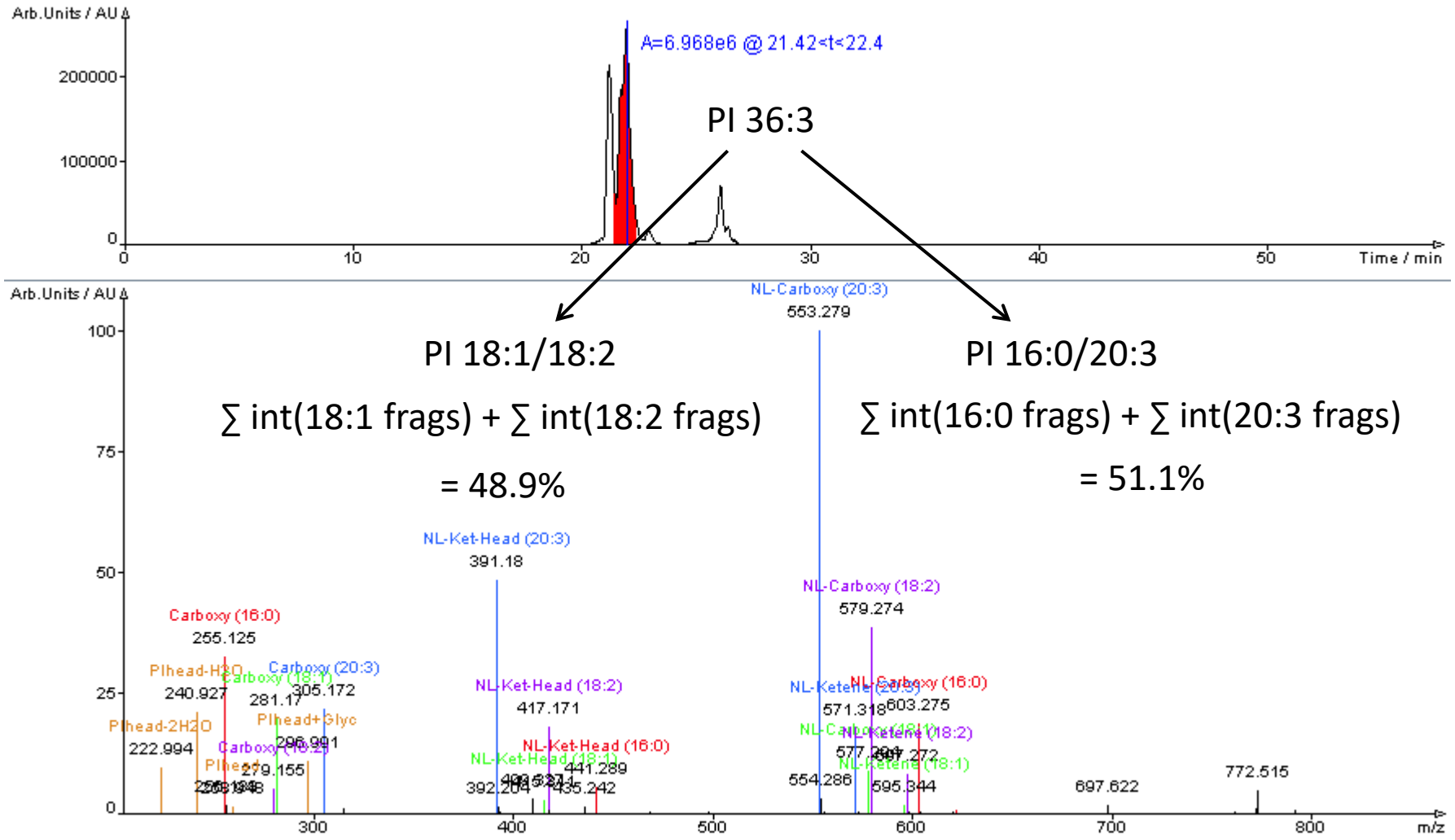
Decision rule sets - example for the lipid class PI

```
[GENERAL]
AmountOfChains=2
ChainLibrary=fattyAcidChains.xlsx
CAtomsFromName=(\d+):\d+
DoubleBondsFromName=\d+:(\d+)
BasePeakCutoff=0.01%

[HEAD]
!FRAGMENTS
Name=PIhead                 Formula=C6H12O9P      Charge=1  MSLevel=2  mandatory=false
Name=PIhead+Glyc Formula=C9H14O9P          Charge=1  MSLevel=2  mandatory=false
Name=PIhead-H2O Formula=C6H10O8P          Charge=1  MSLevel=2  mandatory=true
Name=PIhead-2H2O            Formula=C6H8O7P Charge=1  MSLevel=2  mandatory=false

!INTENSITIES
Equation=PIhead-H2O>2*PIhead                mandatory=true
Equation=PIhead-H2O>0.1*$BASEPEAK mandatory=true
Equation=PIhead-H2O>PIhead-2H2O   mandatory=true
Equation=PIhead+Glyc>2*PIhead               mandatory=false


[CHAINS]
!FRAGMENTS
Name=Carboxy                Formula=$CHAIN-H                  Charge=1  MSLevel=2  mandatory=true
Name=NL-Carboxy  Formula=$PRECURSOR-$CHAIN                    Charge=1  MSLevel=2  mandatory=true
Name=NL-Ketene              Formula=$PRECURSOR-$CHAIN+H2O     Charge=1  MSLevel=2  mandatory=false
Name=NL-Ket-Head            Formula=NL-Ketene-C6H12O6                   Charge=1  MSLevel=2  mandatory=false

!INTENSITIES
Equation=NL-Carboxy>0.6*NL-Ketene   mandatory=true


[POSITION]
!INTENSITIES
Equation=Carboxy[1]>Carboxy[2]                mandatory=false
Equation=NL-Carboxy[2]*0.7>NL-Carboxy[1]      mandatory=true
Equation=NL-Ketene[2]*0.7>NL-Ketene[1]        mandatory=false
Equation=NL-Ket-Head[2]*0.9>NL-Ket-Head[1]    mandatory=true
```

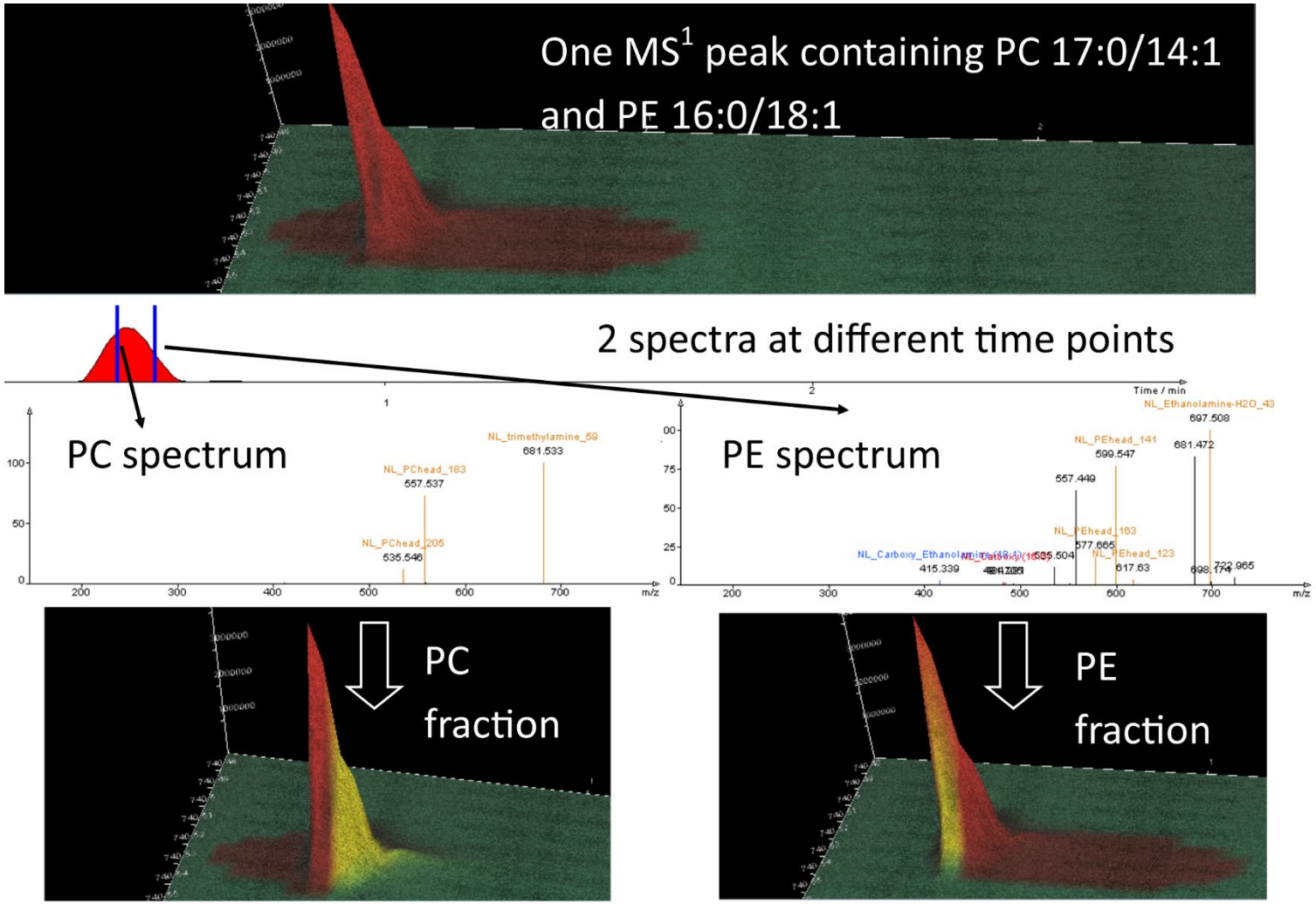
Decision rule sets - graphical user interface



Decision rule sets - multiplexed spectra – co-eluting species



Decision rule sets - mixed MS1 peaks

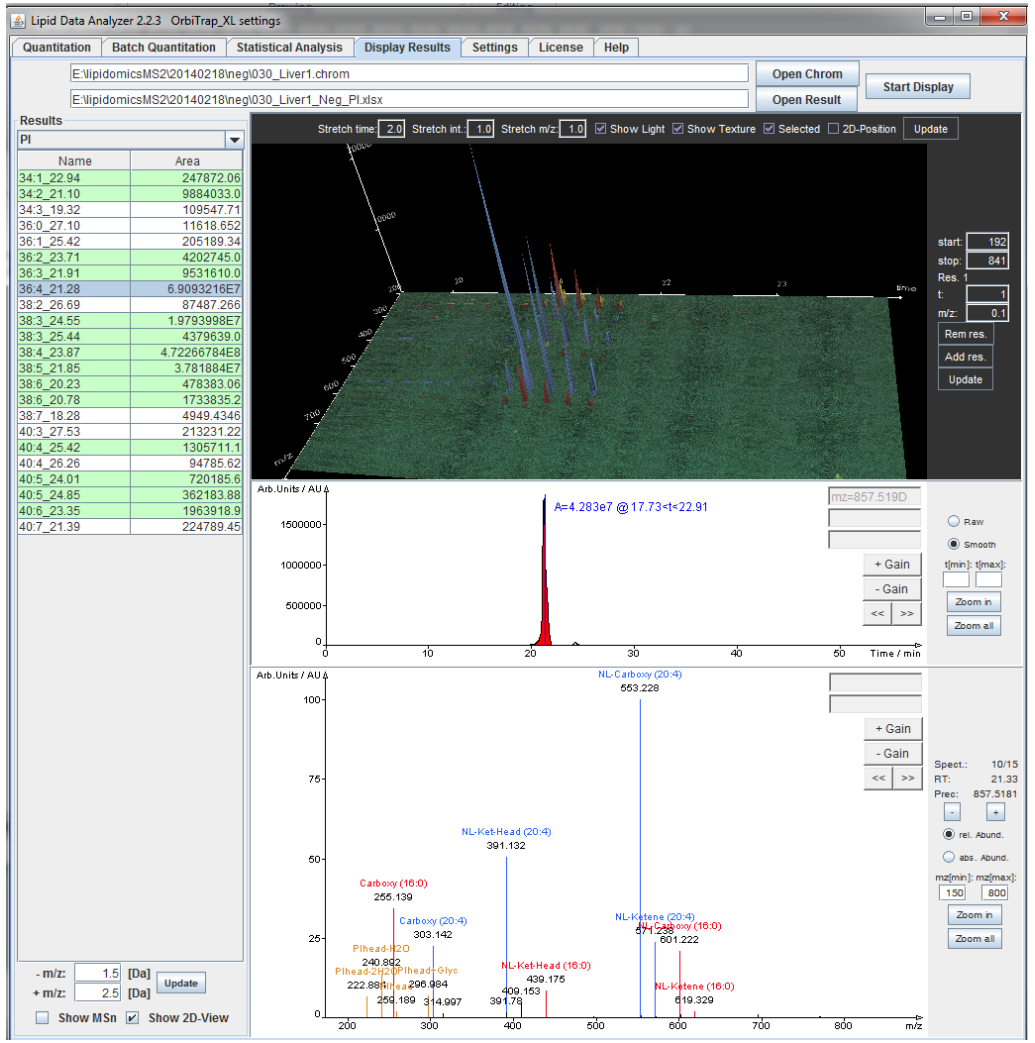


Decision rule sets - retention time filtering

Results

PI

Name	Area
~~34:0_0.64~~	~~3611.3342~~
34:1_22.94	247872.06
34:2_21.10	9884033.0
34:3_19.32	109547.71
~~34:5_0.57~~	~~1490.3436~~
~~34:5_47.81~~	~~507.3144~~
~~34:6_0.69~~	~~3556.525~~
~~36:0_9.21~~	~~2338.4358~~
36:0_27.10	11618.652
36:1_25.42	205189.34
36:2_23.71	4202745.0
36:3_21.91	9531610.0
36:4_21.28	6.9093216E7
~~36:5_19.58~~	~~465799.38~~
~~36:6_8.75~~	~~586.19824~~
~~36:7_50.02~~	~~830.8114~~
38:2_26.69	87487.266
38:3_24.55	1.9793998E7
38:3_25.44	4379639.0
38:4_23.87	4.72266784E8
38:5_21.85	3.781884E7
38:6_20.23	478383.06
38:6_20.78	1733835.2
38:7_18.28	4949.4346
~~38:8_45.00~~	~~514.01514~~
~~40:0_53.85~~	~~9605.83~~
~~40:3_21.46~~	~~50613.062~~
40:3_27.53	213231.22
40:4_25.42	1305711.1
40:4_26.26	94785.62
40:5_24.01	720185.6
40:5_24.85	362183.88
40:6_23.35	1963918.9
40:7_21.39	224789.45
~~40:8_27.39~~	~~428976.2~~



Decision rule sets – performance (high resolution Orbitrap)

positive ion mode

	Total lipid species identified in all quintuplicates: 1077			Total lipid molecular species identified in all quintuplicates: 3567		
	LDA	LB 450	LB 10	LDA	LB 450	LB 10
Sensitivity (%)[a]	97	36	85	80	15	57
PPV (%)[b]	97	91	70	92	91	58

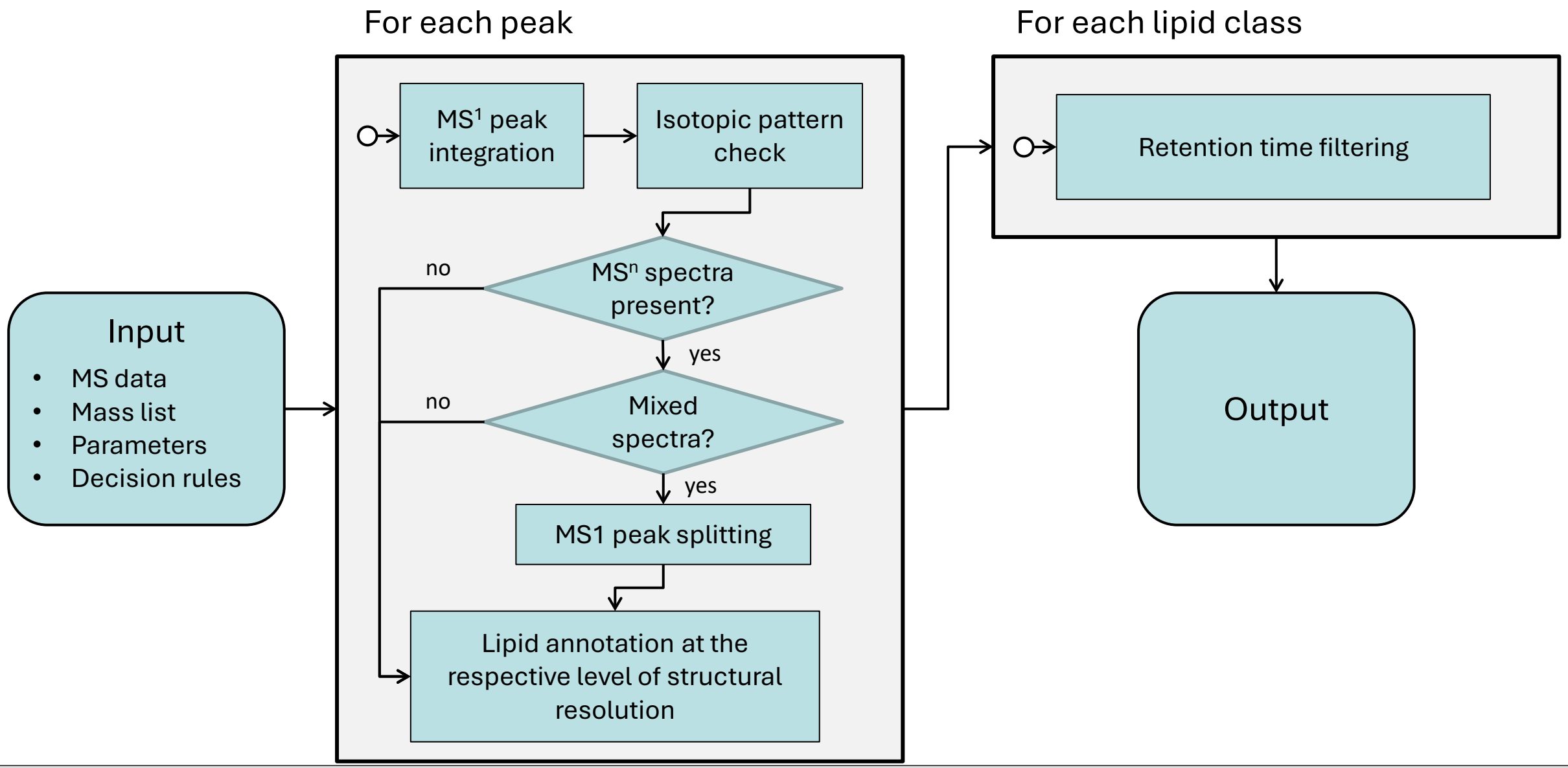
negative ion mode

	Total lipid species identified in all quintuplicates: 573			Total lipid molecular species identified in all quintuplicates: 899		
	LDA	LB 450	LB 10	LDA	LB 450	LB 10
Sensitivity (%)[a]	95	37	96	88	26	83
PPV (%)[b]	99	95	74	96	96	69

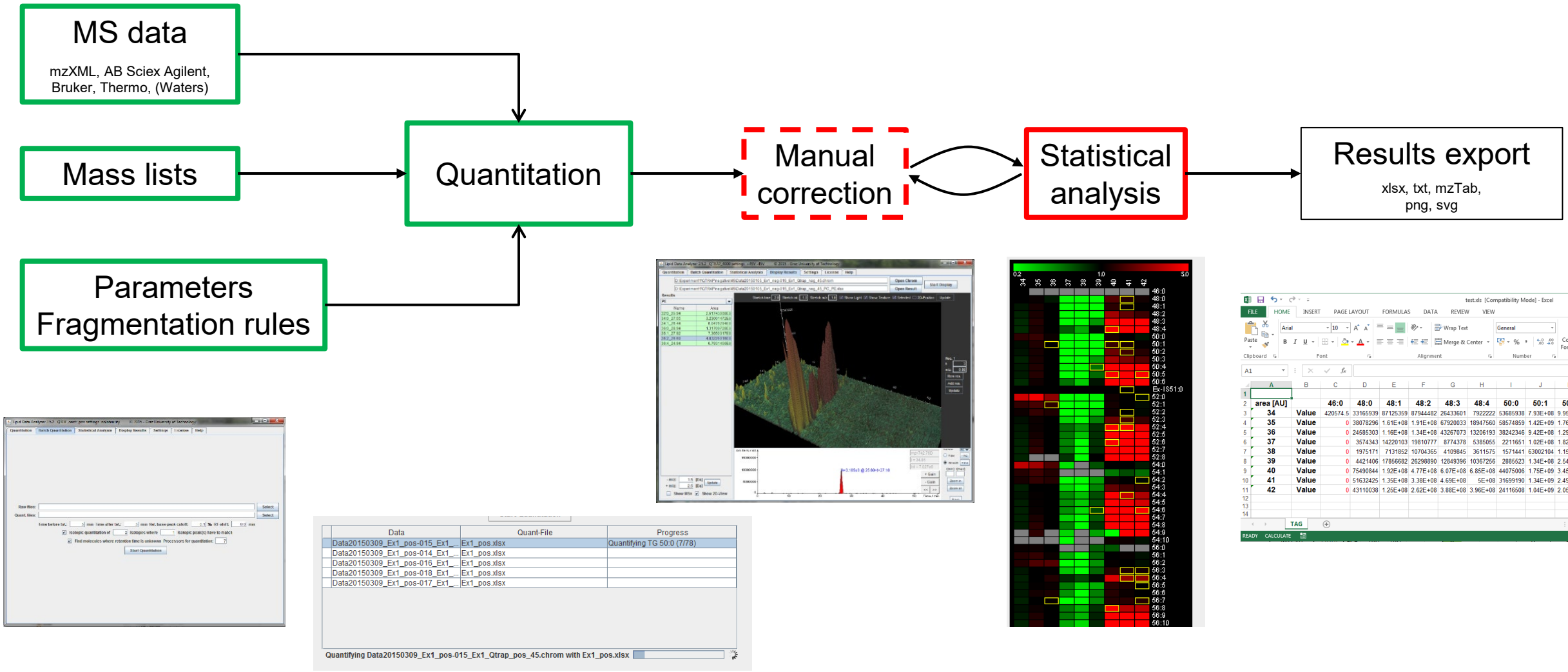
Kind T, *et al.*, Nature Methods, 2013
Hartler J, *et al.*, Nature Methods, 2017

[a] Sensitivity: percent of total species identified;
[b] Positive predictive value: percent of correct identifications.

LDA quantitation - overview



LDA – Data analysis workflow



Exercise 1 – Find the false positives

The LDA is a highly reliable solution

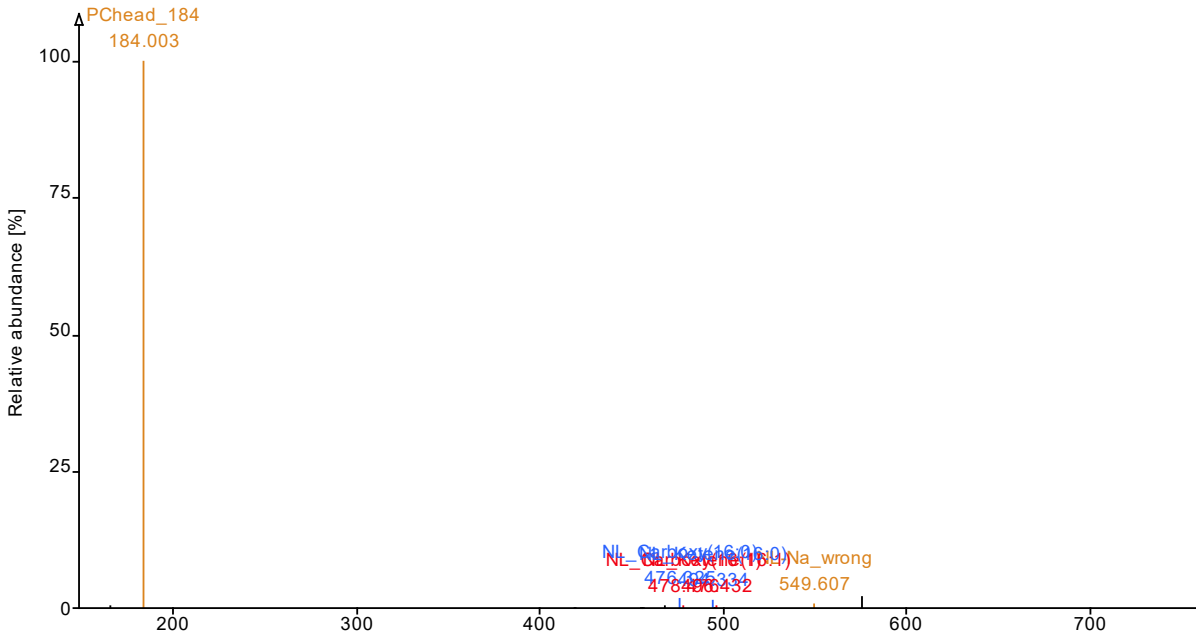
– nonetheless, false positives (FP) are reported with every lipid annotation software

Your task:

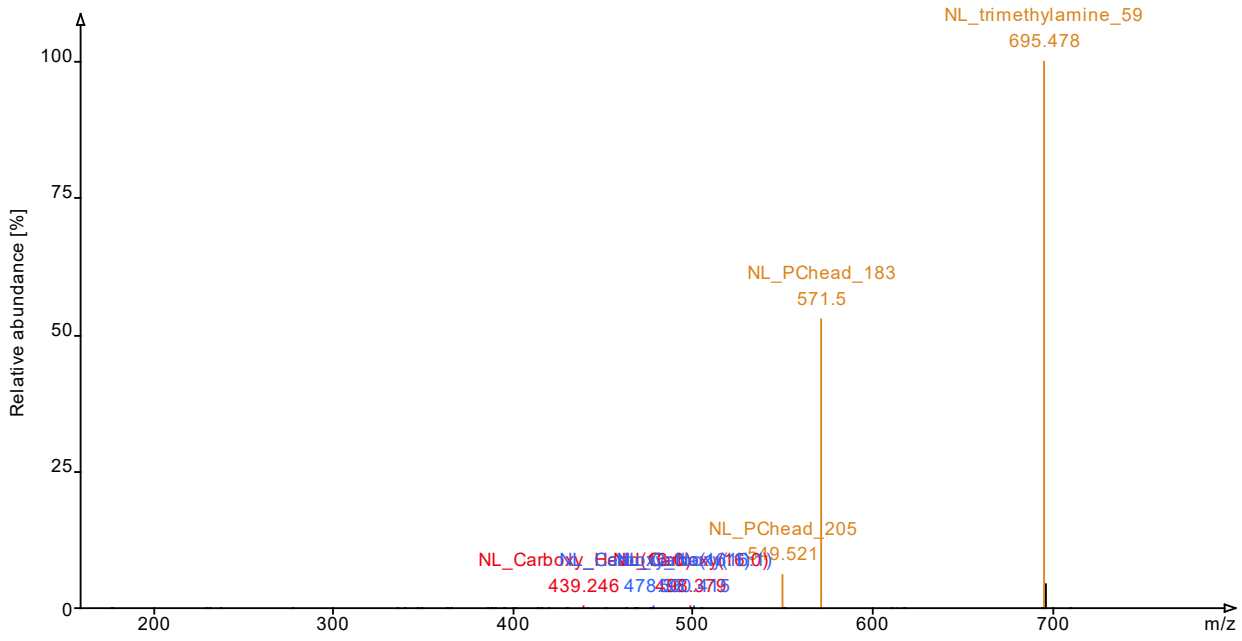
- Find FP annotations in PC, PE and P-PE
- Make use of the ECN model (different adducts should elute at the same time)
- Make use of fragmentation patterns

Exercise 1 – Find the false positives – PC fragmentation

PC $[M+H]^+$



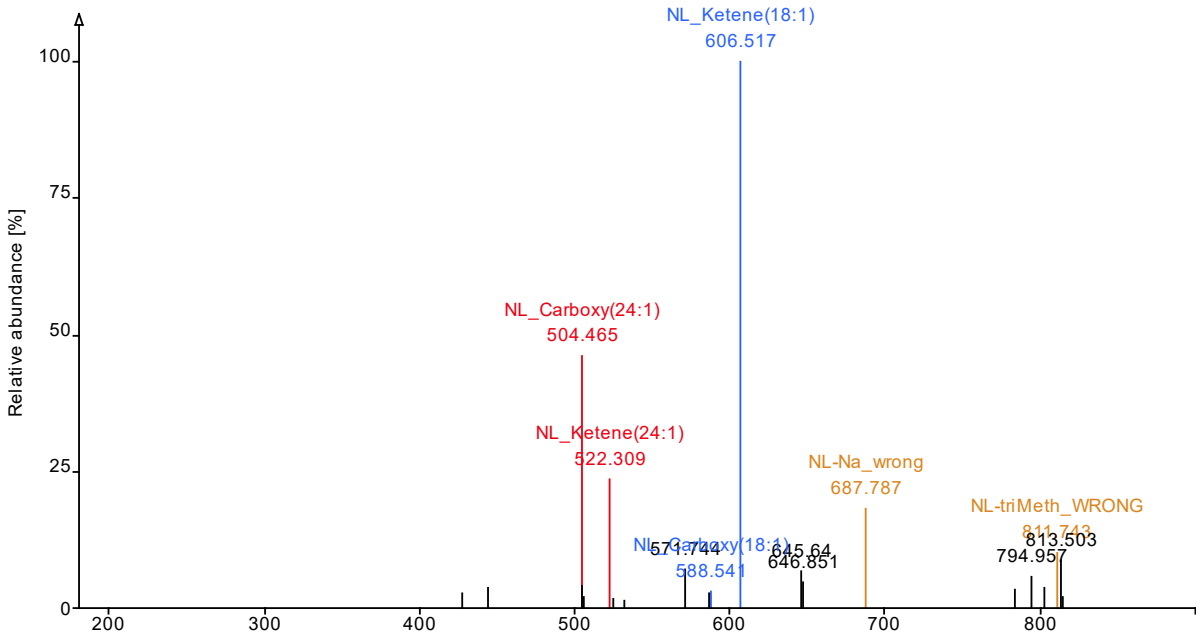
PC $[M+Na]^+$



Exercise 1 – Find the false positives – PC fragmentation

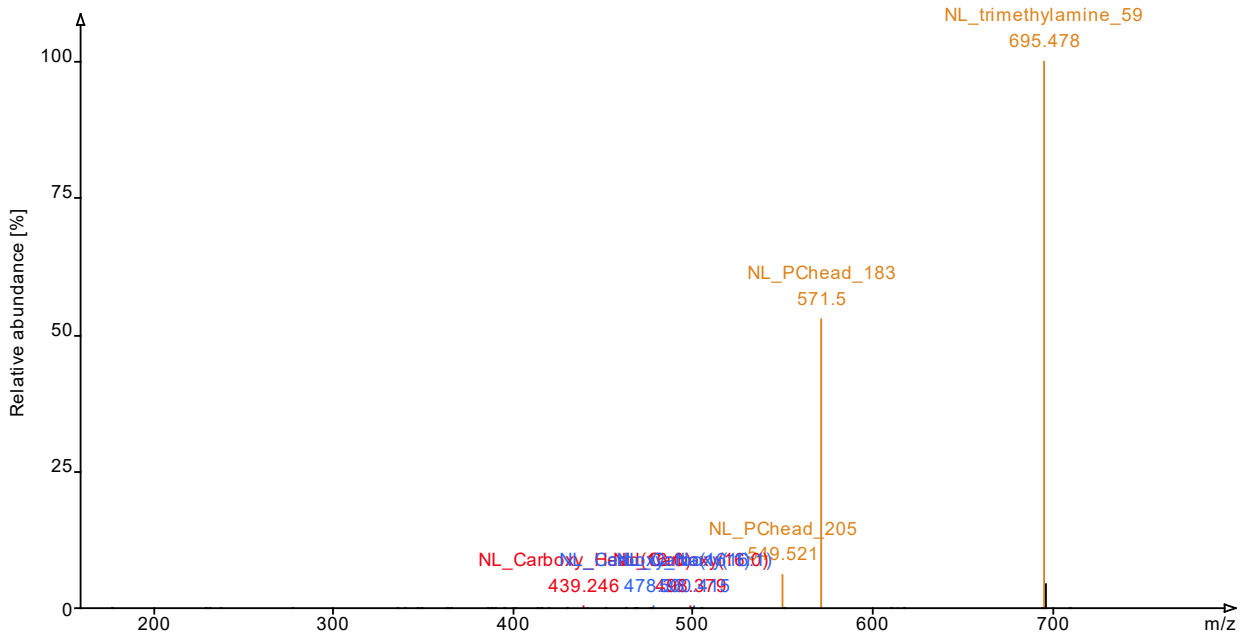
PC [M+H]$^+$

dynamic exclusion of CID cell



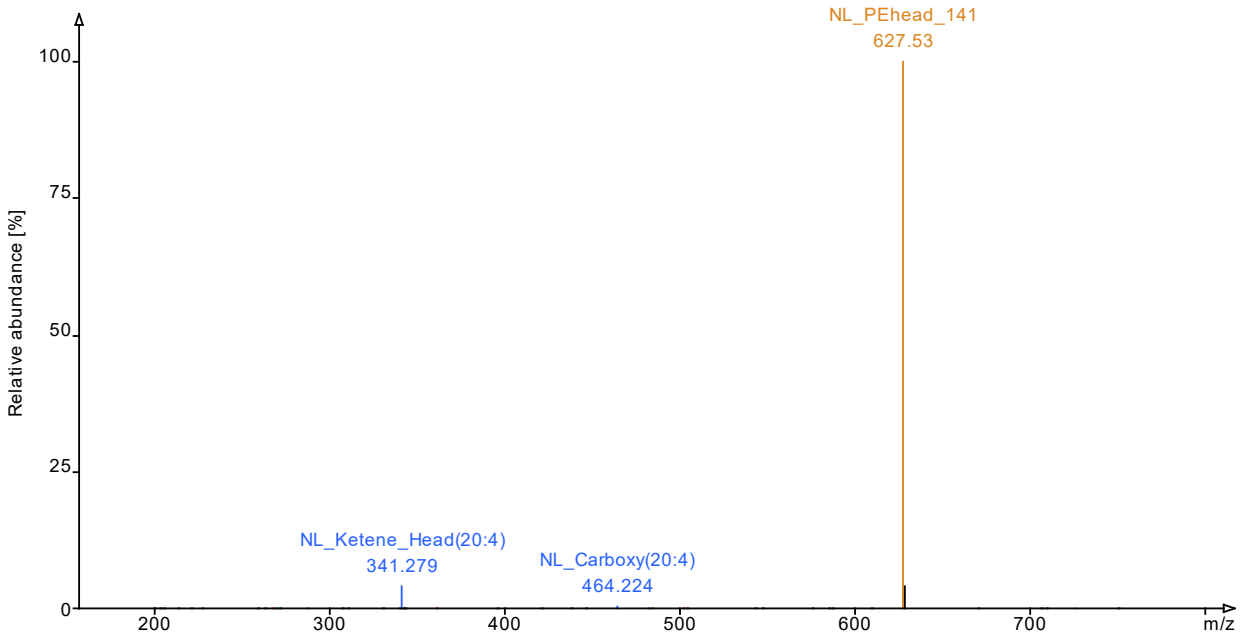
precursor: 870.695 Da

PC [M+Na]$^+$

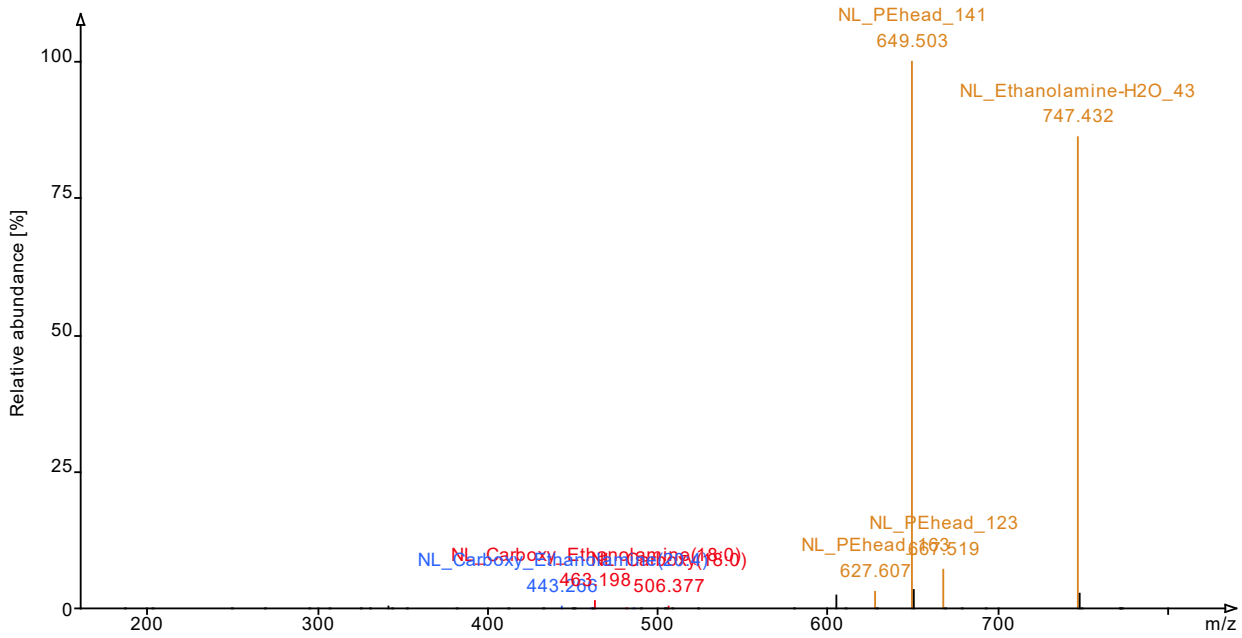


Exercise 1 – Find the false positives – PE fragmentation

PE $[M+H]^+$

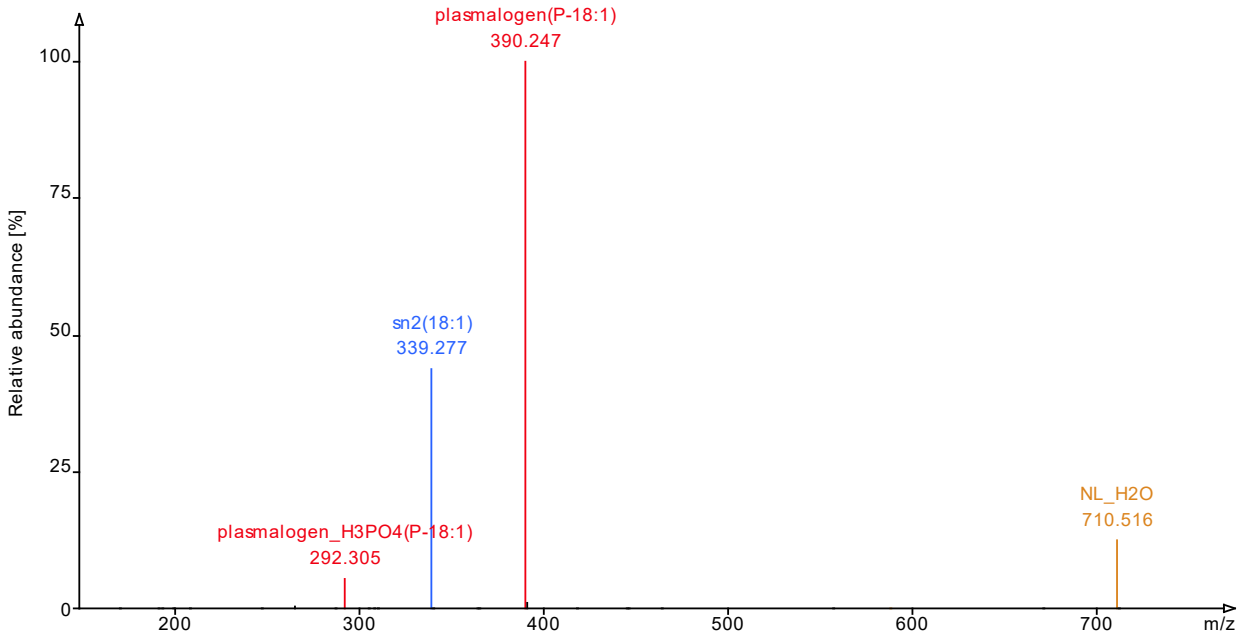


PE $[M+Na]^+$

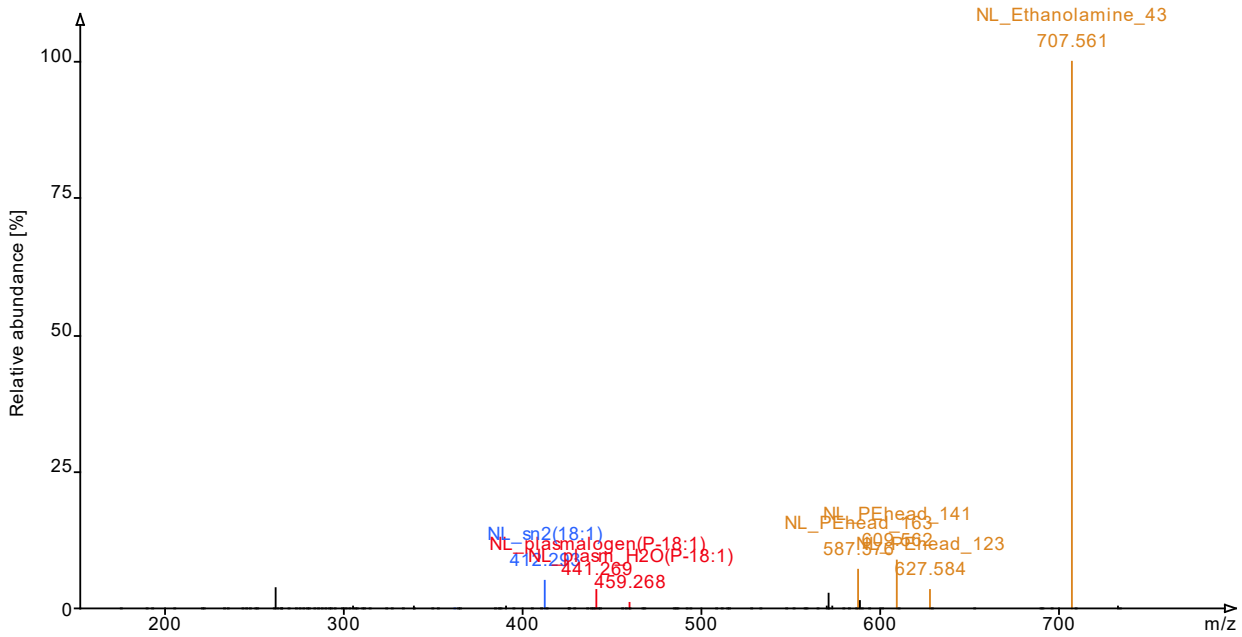


Exercise 1 – Find the false positives – PE fragmentation

P-PE $[M+H]^+$



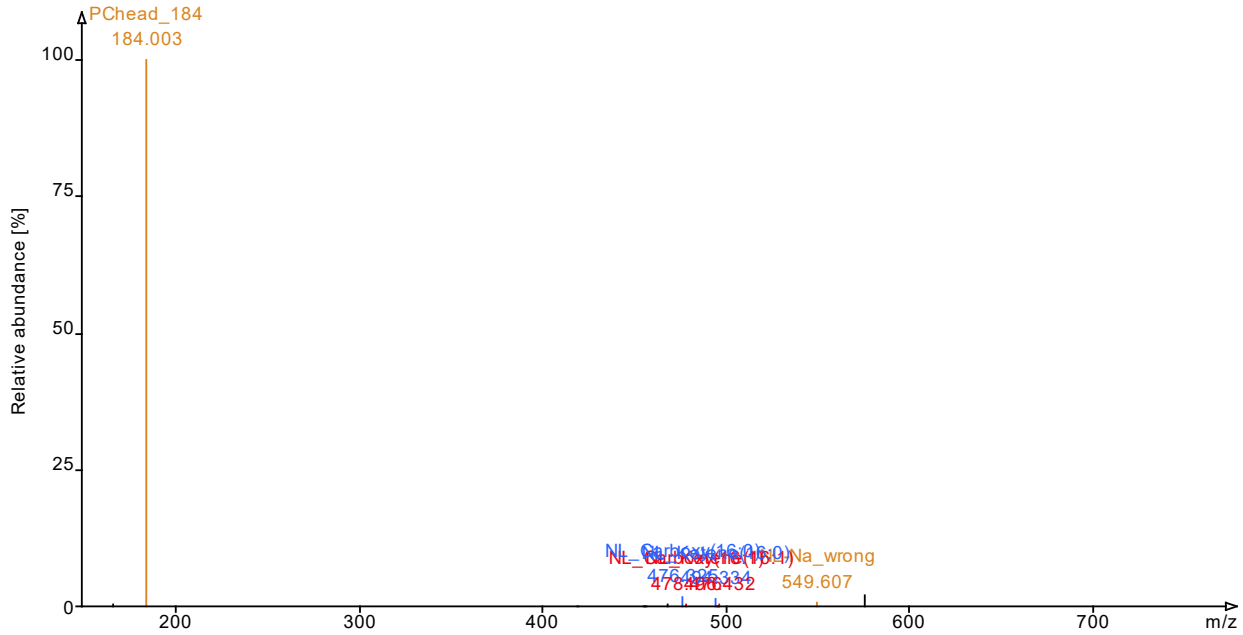
P-PE $[M+Na]^+$



Your task:

- Find FP annotations in PC, PE and P-PE
- Make use of the ECN model (different adducts should elute at the same time)
- Make use of fragmentation patterns

PC $[M+H]^+$



False positives – cheat sheet

PC

30:0_21.86	RT wrong
30:1_20.30	RT wrong
32:2_22.33	MS1
34:4_20.60	MS1
35:3_24.77	PE 38:3_H
36:5_22.00	PE 39:5_H
37:3_26.05	MS1
39:1_30.30	PE 42:1_H
39:2_28.48	PE 42:2_H
39:3_26.01	MS1
39:4_26.15	MS1
39:5_24.94	MS1
39:5_25.44	MS1
39:6_24.19	MS1
39:6_24.92	MS1
39:9_21.60	MS1
39:10_21.17	MS1
44:5_28.17	MS1

PE

33:1_20:30	MS1
37:4_20:60	MS1
38:7_21:23	PE 36:4_Na
40:5_25.07	MS1
40:7_23:44	PE 38:4_Na
40:8_21.67	PE 38:5_Na
42:4_24.15	MS1
42:7_24.74	MS1
42:10_19.48	MS1
44:6_23.63	MS1
45:6_24.62	MS1
46.12_20.57	MS1

PE P-

34:1_22.25_Na	36:4_H
35:1_23.35_Na	37:4_H
35:3_21.89_Na	MS1
36:0_25.69_Na	RT wrong
36:1_24.37_Na	38:4_H
36:4_23.81_Na	MS1
37:4_22.43_Na	MS1
38:1_26.28_Na	RT wrong
38:4_25.76_Na	MS1
38:5_24.25_Na	MS1
38:7_22:25_H	MS1
39:6_24.06_H	MS1
39:7_23.28_Na	MS1
40:0_20.07_Na	MS1
40:1_27.38_Na	MS1
40:2_26.28_Na	RT wrong
40:3_25.07_Na	MS1
40:3_25.80_Na	RT wrong
40:4_23.50_Na	MS1
40:5_26.16_Na	MS1
40:7_20.77_Na	MS1
40:8_22.67_Na	MS1
40:8_22.67_H	MS1
42:3_26.56_H	MS1
42:5_27.61_Na	MS1
44:3_27.41_H	MS1

TG

60:10_20.39	MS1