

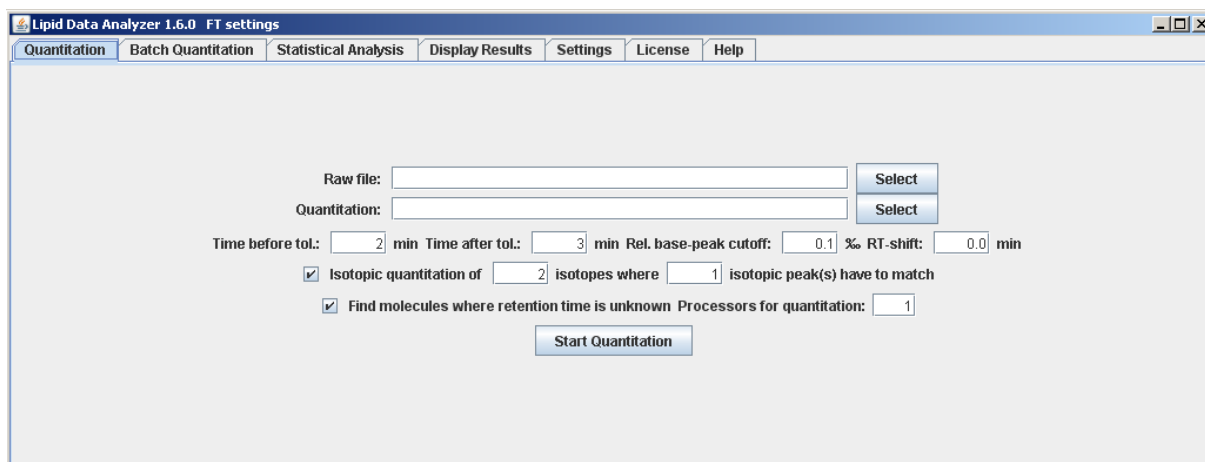
# **Lipid Data Analyzer 2.9**

## User Manual

1	General overview .....	1
2	Settings .....	1
3	Quantitation .....	2
3.1	Polarity switching .....	4
4	Batch quantitation .....	4
4.1	Polarity switching .....	5
5	Manual verification – visualization .....	6
5.1	Visualization – left menu .....	7
5.2	Visualization – 3D viewer .....	10
5.3	Visualization – Chromatogram viewer .....	11
5.4	Visualization – Spectrum viewer .....	12
6	MS <sup>n</sup> spectra interpretation – decision rule sets .....	13
6.1	Organization of decision rule sets .....	13
6.2	Decision rule set file (frag.txt) .....	15
6.3	GUI for MS <sup>n</sup> rule definition/review .....	21
7	Statistical Analysis .....	25
7.1	File selection and settings .....	25
7.2	Heat map and visualization settings .....	27
7.3	Bar charts .....	32
7.4	Overview bar chart .....	34
8	Licensing .....	35
9	Help .....	35
	Appendix A - Preparation of the mass list Excel file .....	37
	Appendix B – Modified lipids .....	38
	Mass list .....	38
	Fatty acid mass list (fattyAcidChains.xlsx) .....	39
	Modification table (modconfig.xml) .....	39
	Appendix C - Configuration settings .....	40
	Performance settings (.settings) .....	40
	Path settings (.settings) .....	40
	Look & feel settings (.settings) .....	41
	Default quantitation settings (.properties) .....	41
	Excel result file settings (.settings) .....	41
	3D viewer default settings (.properties) .....	41
	Standards prefix settings (.settings) .....	41
	mzXML → chrom settings(.properties) .....	41
	Waters raw → chrom settings(.settings) .....	42
	General algorithm settings(.properties) .....	42
	General MS <sup>n</sup> algorithm settings (.properties) .....	42
	Memory settings(.bat, .sh, or. vmoptions) .....	43
	Specific settings for the 3D algorithm (.properties) .....	43
	Appendix D – mzTab settings .....	46
	MS instrument specific settings (.properties) .....	46
	Other mzTab settings (mzTab.properties) .....	47
	Appendix E - Trouble shooting .....	47

# 1 General overview

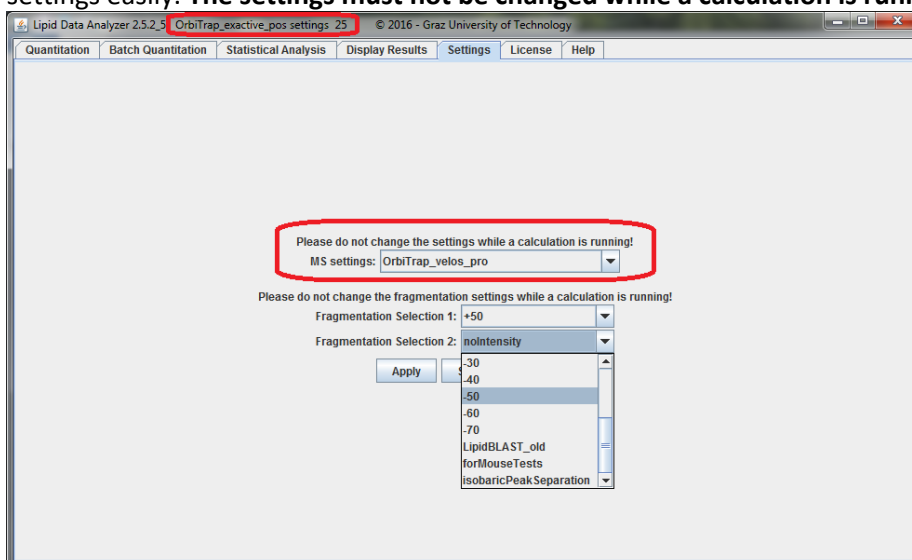
There are two ways to start the program; as plain executable or with a console. If the console is used, a command line window (black one) appears, which contains the logging of the software.



In the LDA window, six tabs are available. For setting up your LDA instance, see chapter 2; for “Quantitation” and “Batch quantitation”, see chapters 3 and 4; for “Statistical analysis” see chapter 7, the “Statistic” section; for “Display Results”, see chapter 5 “Manual verification – visualization”; for licensing information, see chapter 8, and for help section, chapter 9.

## 2 Settings

The LDA can be used for several MS platforms. The Settings tab has been introduced to change these settings easily. **The settings must not be changed while a calculation is running!**



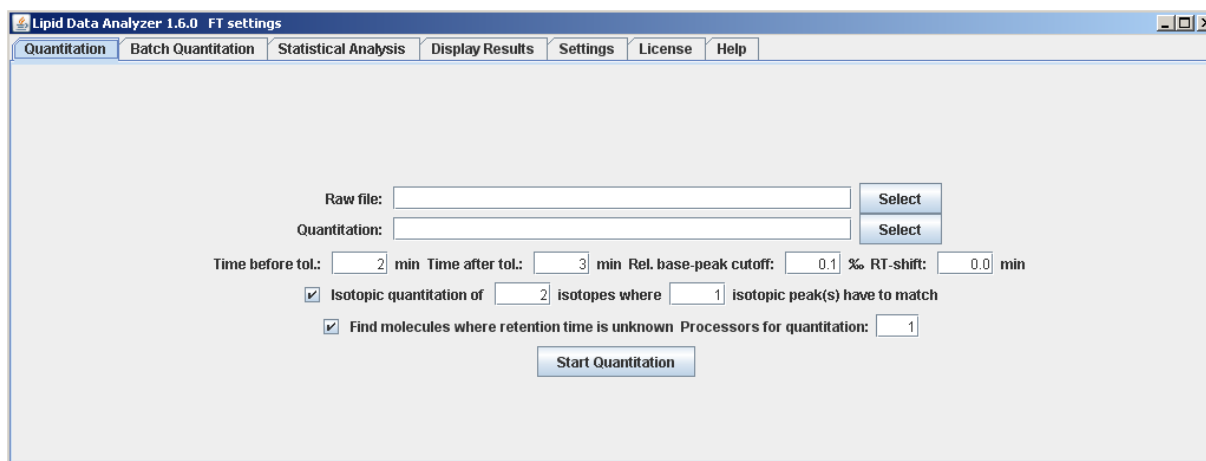
The current settings are displayed in the title of the application. The “MS settings” drop down menu is for selecting the corresponding MS-platform – this affects mainly the performance of the MS<sup>1</sup> quantitation. For the MS<sup>n</sup> interpretation, drop down menus 2 (and 3) are available. These boxes offer for all MS platforms the options “none” (for no MS<sup>n</sup> identification) and “noIntensity” which checks for all possible fragments without considering any intensity information (rather unspecific). Further options are available, if specific MS<sup>n</sup> decision rule sets are available for this MS platform. These

additional options are usually named by the collision settings such as +50 or -30. If another option is selected than “none” or “noIntensity” in “Fragmentation Selection 1”, a third select box appears. These two drop down menus allow for the simultaneous selection of decision rule sets for positive and negative ion mode data. Further details about MS<sup>n</sup> spectra interpretation by decision rule sets, and the development of such rule sets can be found in the chapter 6 “MS<sup>n</sup> spectra interpretation – decision rule sets”.

The “Apply” button applies the settings to the current session. The “Save as default” button applies the selected settings and sets them as default, which will be loaded automatically at the next start of the LDA.

### 3 Quantitation

In order to start a quantitation of a single MS run, a raw file (Raw file) and an Excel file (Quantitation) containing the mass lists are required. The default settings for the input fields can be entered in the LipidDataAnalyzer.properties file (see Appendix C - Configuration settings).



**Raw file:** Files in mzXML or chrom formats are required. However, under Windows operating system, it is possible to use Thermo RAW format, Agilent .d directories, and ABSciex wiff files. In order to use Waters raw directories, the software Mass++ has to be installed and the path to the mspcmd.exe has to be set for the MassPPPath variable of the config file (see Appendix C - Configuration settings).

**Quant. files:** Here, the name of an Excel file has to be entered, containing the mass list. The format of the file is described in Appendix A - Preparation of the mass list Excel file and for MS<sup>n</sup> identifications, at the beginning of chapter 6 “MS<sup>n</sup> spectra interpretation – decision rule sets”. You can find mass list files in the LDA installation folder at the folder exampleMassList. The recommended ones are positive\_noSphingoOH.xlsx and negative\_noSphingoOH.xlsx, except your instrument is the OrbiTrap\_velos\_pro. For this instrument use the positive.xlsx and the negative.xlsx. The reason is that the testing all possible hydroxylations, the fragRules must have certain settings, and the mass list must contain the parameters “OH-Number” and “OH-Range” (see Appendix A). If not this combination is used, an error is thrown. The fragmentation rules of the OrbiTrap\_velos\_pro contain these parameters, and the positive.xlsx and the negative.xlsx do so to. Thus, this combination must be used.

The Q-Exactive has only been partially migrated to the new hydroxylation rules. Thus you have to do the following:

- For the positive ion mode: use the positive\_noSphingoOH.xlsx, but replace the ‘Cer’ tab with the one from the positive.xlsx.
- For the negative ion mode: use the negative.xlsx, but remove the tabs for LSM and HexCer.

Time before tolerance: This field specifies the retention time tolerance in minutes that is applied before the entered ideal retention time in the Excel file. Example is given after the “Time after tolerance” paragraph. This field is used if a retention time (“tR (min)” column) has been entered in the mass-list Excel file.

Time after tolerance: This field specifies the retention time tolerance in minutes that is applied after the entered ideal retention time in the Excel file. An example is given in the next paragraph. This field is just used if a retention time (“tR (min)” column) has been entered in the mass-list Excel file.

Example for time tolerance setting: In the excel file, the retention time for a specific analyte is specified with 24 minutes, in the time before tolerance 2min and in the time after tolerance 3min is entered. Then, the algorithm will look for the analyte in a time range from 22-27 minutes.

Rel. base-peak cutoff: Peaks that are less intense than this per mille value relative to the highest identified peak are discarded by the algorithm. For the determination of the highest peak, all detected analytes are taken into account (irrespective of the analyte class). This threshold can accelerate the quantitation, since intensities that are too small are discarded before the 3D quantitation is performed.

RT-shift: The retention time may shift from batch to batch. The entered value will be added to the retention time defined in the mass list file. Thus, the mass list file does not have to be changed every time a retention time shift occurs.

Isotopic quantitation of ... isotopes where ... isotopic peaks have to match: The checkbox indicates if in addition to the +0 peak other isotopic peaks should be quantified. The first value determines the amount of additional isotopes to be quantified (e.g. 2 would mean that +0, +1, and +2 isotopes will be quantified). The second value determines how many isotopes have to comply with the theoretical isotopic distribution (e.g. 1: just the +1 peak has to comply with the theoretical distribution).

Find molecules where the retention time is unknown: If this checkbox is selected, compounds without retention time entries (“tR (min)” column) in the mass list are quantified, omitted otherwise. However, the quantitation of molecules without retention time requires more time.

Processors for quantitation: The amount of processors that shall be used for the quantitation. LDA detects the amount of processors available, automatically. The default value is always n-1.

#### Start Quantitation:

This button starts the quantitation.

Lipid Data Analyzer 1.6.0 FT settings

Quantitation Batch Quantitation Statistical Analysis Display Results Settings License Help

Raw file: F:\lipidomics\20100210\20100126\_TAG-34.RAW Select

Quantitation: L:\LDA-Paper\quantExcelFiles\masslist-TG\_with\_IS.xls Select

Time before tol.: 2 min Time after tol.: 3 min Rel. base-peak cutoff: 0.0 % RT-shift: 0.0 min

☒ Isotopic quantitation of 2 isotopes where 1 isotopic peak(s) have to match

☒ Find molecules where retention time is unknown Processors for quantitation: 1

Start Quantitation

Quantifying TG 48:2 (3/101) %

After the quantitation has been started, first, data in vendor-specific raw formats are translated in mzXML (if the quantitation is based on the raw file formats). Second, mzXML files are translated in the internal chrom file format. If for the quantitation vendor-specific raw formats are used (not mzXML), the intermediate mzXML is deleted automatically after conversion to chrom format. Based on the chrom files, the quantitation is started. The results of the quantitation will be stored in an Excel file, whereby the name of the results file will be \$path\_of\_the\_chrom\_file\_without\_suffix\$\_\$quant\_file\_name\$ (e.g. raw file: F:\lipidomics\20100210\20100126\_TAG-34.chrom quant file: F:\lipidomics\TG\_NH4\_ACN.xls → result file: F:\lipidomics\20100210\20100126\_TAG-34\_TG\_NH4\_ACN.xls).

### 3.1 Polarity switching

LDA supports polarity switching, but batch quantitation is the recommended mode to quantify such data. When you quantify from vendor file formats or from mzXML, the LDA generates two chrom directories, one ending with “\_positive.chrom”, and the other with “\_negative.chrom”. Since the LDA cannot guess whether your quantitation Excel file contains analytes in positive or negative ion mode, it expects that the name of the Excel file must contain “positive” or “negative”, otherwise an error is reported. For directly quantifying chrom files that originate from data using polarity switching, there are no restrictions regarding the naming of the quantitation file, in the single quantitation file.

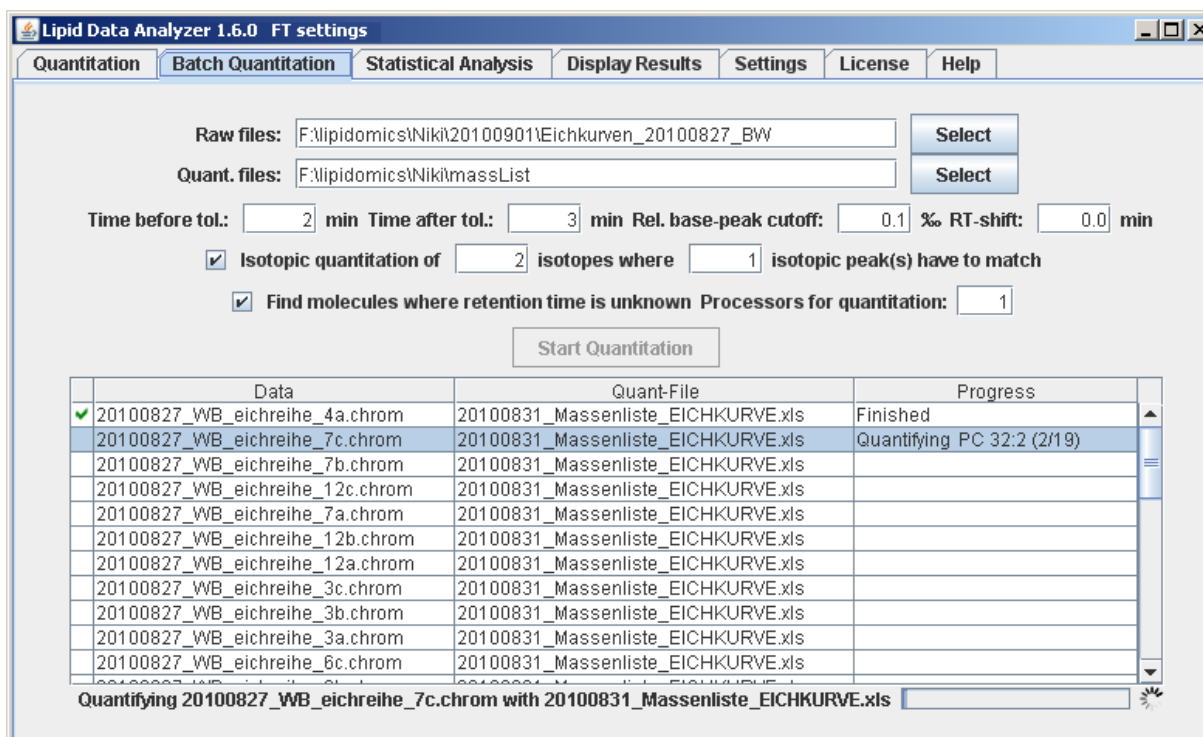
## 4 Batch quantitation

The input masks for “Quantitation” and “Batch Quantitation” look similar. The difference is that for “Quantitation” the file chooser requires a single MS-run (file or directory, depending on the format), whereas in “Batch Quantitation”, a directory containing the files to quantify and a directory containing the Excel mass list files are required.

The screenshot shows the 'Batch Quantitation' tab in the 'Lipid Data Analyzer 1.6.0 FT settings' window. The interface includes the following elements:

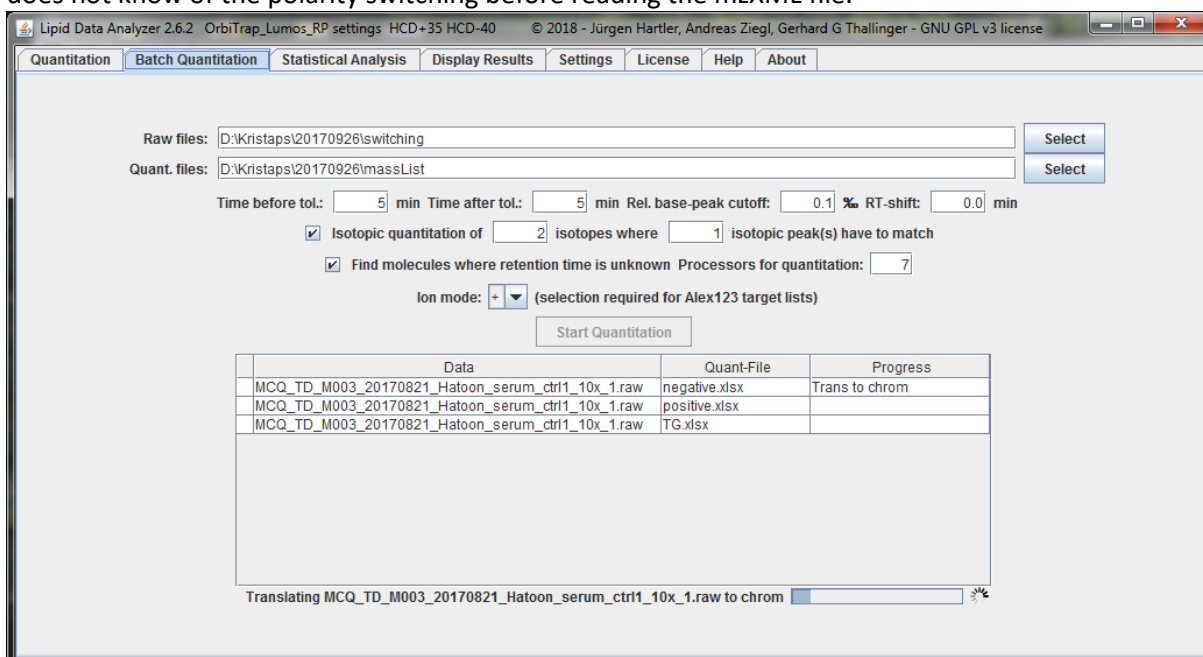
- Raw files:** A text input field with a 'Select' button to its right.
- Quant. files:** A text input field with a 'Select' button to its right.
- Time before tol.:** A numeric input field set to '2' with the unit 'min'.
- Time after tol.:** A numeric input field set to '3' with the unit 'min'.
- Rel. base-peak cutoff:** A numeric input field set to '0.1' with the unit '%'.
- RT-shift:** A numeric input field set to '0.0' with the unit 'min'.
- Isotopic quantitation:** A checked checkbox followed by 'Isotopic quantitation of' and a numeric input field set to '2', followed by 'isotopes where' and a numeric input field set to '1', followed by 'isotopic peak(s) have to match'.
- Find molecules where retention time is unknown:** A checked checkbox followed by this text and 'Processors for quantitation:' and a numeric input field set to '1'.
- Start Quantitation:** A large button at the bottom center of the window.

In contrast to the normal quantitation, in batch mode, a list will appear that shows the progress of the quantitation. The files receive a green check mark, if they were quantified successfully; if not, a red X.

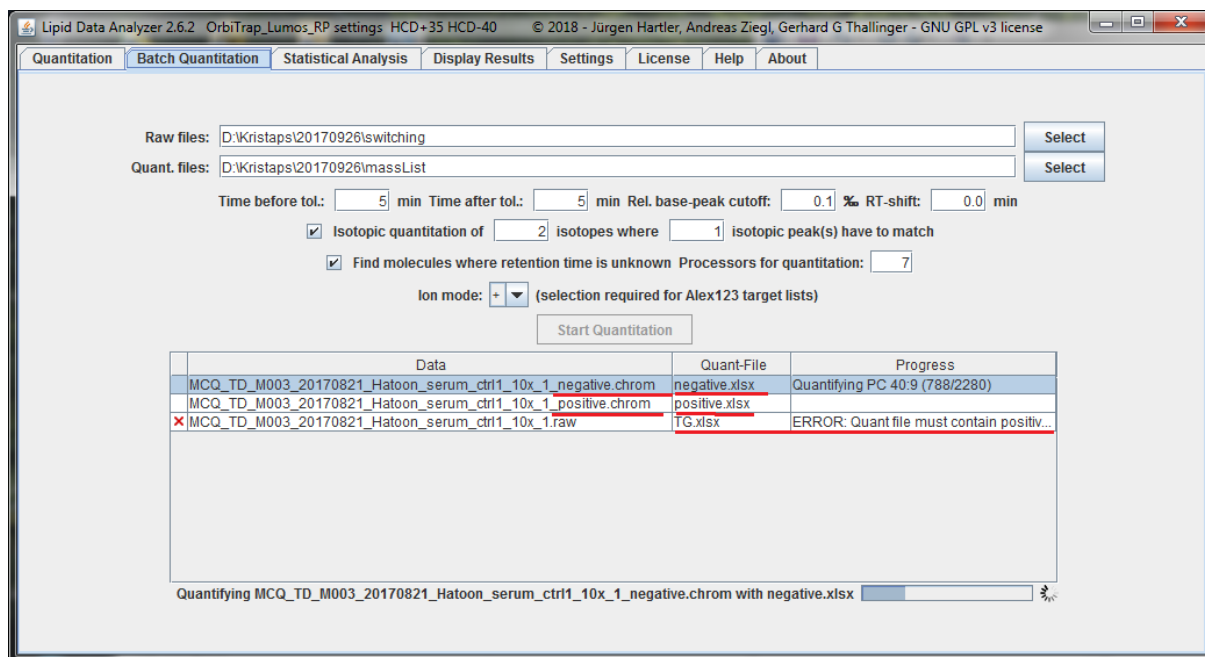


## 4.1 Polarity switching

LDA supports polarity switching, and batch quantitation is the recommended mode to quantify such data. For such data, the raw files and the Excels are first listed as for every other data, since the LDA does not know of the polarity switching before reading the mzXML file.



When you quantify from vendor file formats or from mzXML, the LDA generates two chrom directories, one ending with “\_positive.chrom”, and the other “\_negative.chrom”. After translating to the chroms containing separate polarities, the LDA changes the names of the data files and assigns the Excels files respective their polarity. In order to make the correct assignment, it expects that the name of the Excel file must contain “positive” or “negative”, otherwise an error is reported, as in this example for TG:

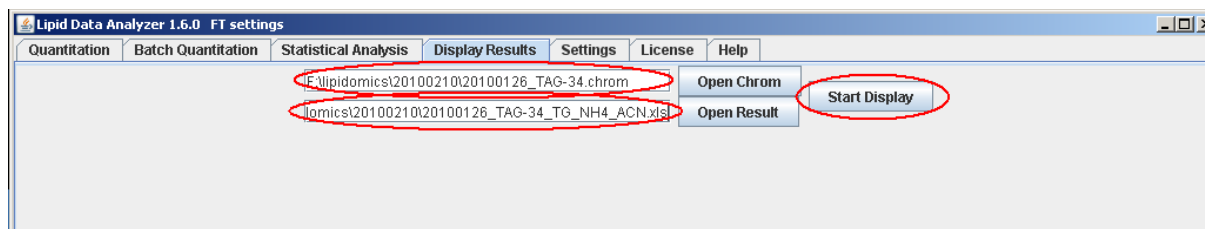


If the Excel file was e.g. named TG\_positive.xlsx, it would have been assigned to the positive chrom file.

For batch quantitation of chrom files from polarity switched data, the Excel files must always contain “positive” or “negative” in their names, otherwise they are ignored.

## 5 Manual verification – visualization

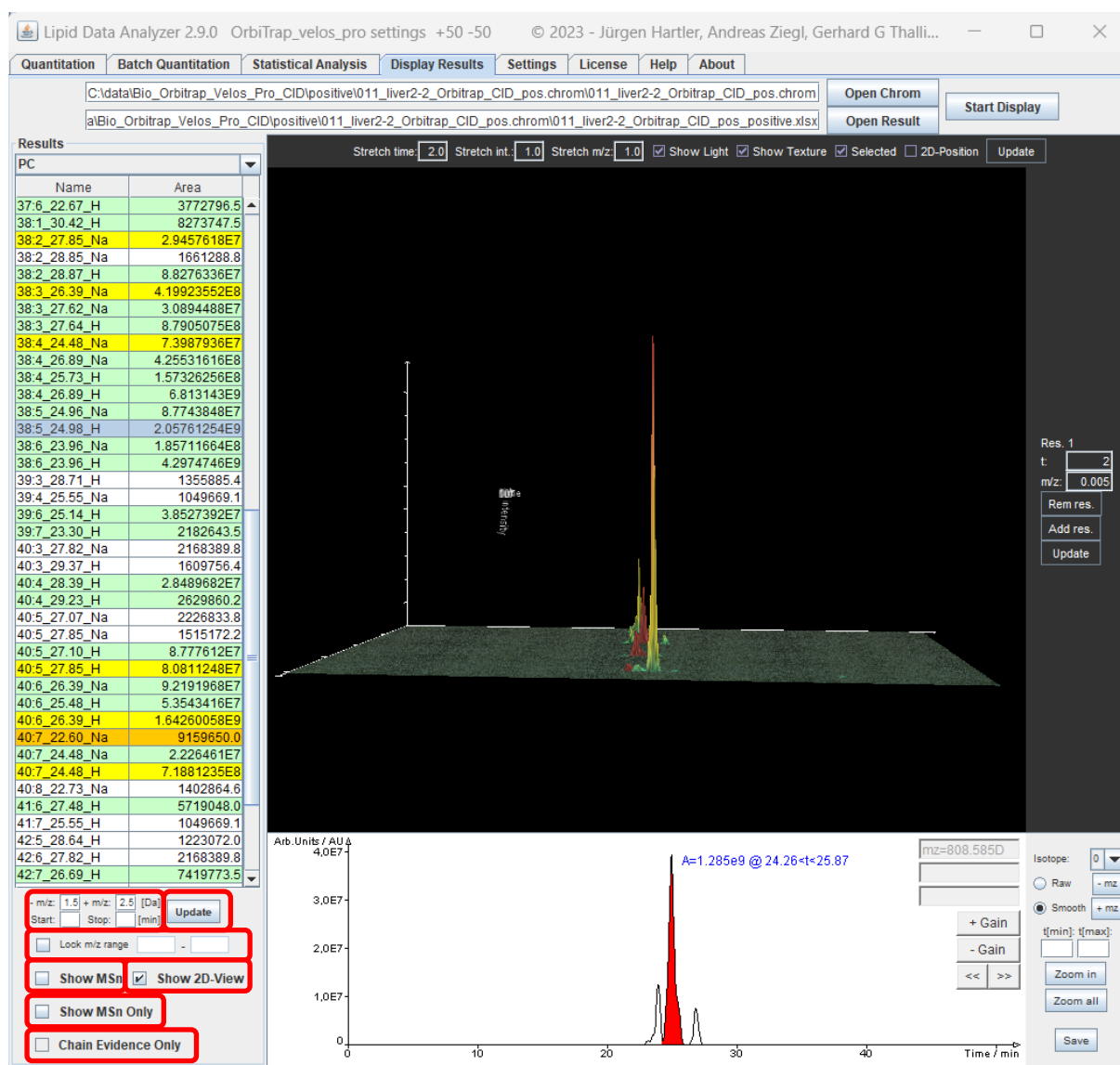
In order to visualize the results, two files are required: the chrom file and the results Excel file:



After the file selection, push the “Start Display” button. The appearing display can be divided into two parts: 1) a menu on the left side (see 5.1); 2) a display frame, whereupon the upper part shows a 3D view (see 5.2) and the lower part displays the 2D chromatogram (see 5.3).



## 5.1 Visualization – left menu



The menu contains a select box at the top, where you can switch between the different lipid subclasses. Then a table follows, whereupon the name of the analyte is in the first column, and the area in the second. The display name consists primarily of “\$name\$.\$double\_bonds\$.\$retention\_time\$” (e.g. 38:5\_24.98). If there are more than one adduct/modification, the modification name is added with \_\$modification\_name\$ (e.g. 38:5\_24.98\_H). The color encoding of the table entries has the following meaning:

- White: MS<sup>1</sup> identification without any MS<sup>n</sup> evidence.
- Green: MS<sup>1</sup> identification verified by MS<sup>n</sup> spectra.
- Yellow: The MS<sup>n</sup> algorithm detected more than one lipid subclass in the MS<sup>1</sup> peak → the peak was split on the MS<sup>1</sup> level in a part for the one and a part for the other lipid subclass. The split is based on information derived from the MS<sup>n</sup> spectra.
- Orange: The MS<sup>n</sup> algorithm detected more than one lipid subclass in this peak, but a split on the MS<sup>1</sup> level is not possible. The intensities of the intact MS<sup>1</sup> peak are split according to the distinct fragments of the MS<sup>n</sup> spectra. This way of splitting is rather unreliable, since usually fragment intensities do not correlate with MS<sup>1</sup> abundances.
- Blue: Current selection for display.

At the bottom of the menu, the range for the 3D viewer can be set. The current center m/z value of the extraction is displayed in the 2D viewer (see chapter 5.3). After the m/z range has been changed,

the “Update” button has to be pressed to change the display. To fix an m/z range (not recommended) irrespective of the selected analyte, click on “Lock m/z range”. The values of the currently active m/z range will be transferred to the text boxes next to this select box. If necessary, change these values according to your needs and press the “Update” Button. This setting is useful when a bigger m/z range needs to be reviewed, since it prevents the time-consuming rendering of the 3D-map whenever another analyte is selected. But be aware in case you select an analyte that falls outside the selected m/z range, it will not be displayed.

Similar to “Lock m/z range”, a fixed chromatographic range can be selected with “Start” and “Stop”. The “Show 2D-View” checkbox can fade out the 2D viewer. The “Show MSn” checkbox switches from the lipid species to the lipid molecular species view of the table where possible:

PC	
Name	Area
20:1_18:0_30.42_H	8273747.5
38:2_27.85_Na	2.9457618E7
38:2_28.85_Na	1661288.8
18:1/20:1_28.87_H	8472244.0
18:0/20:2_28.87_H	7.9804088E7
18:0/20:3_26.39_Na	4.19923552E8
18:0/20:3_27.62_Na	3.0894488E7
21:1_17.2_27.64_H	9.5518848E7
18:0/20:3_27.64_H	7.835319E8
38:4_24.48_Na	7.3987936E7
18:0/20:4_26.89_Na	4.25531616E8
22:4_16:0_25.73_H	1.3572973E7
20:4_18:0_25.73_H	3.7298596E7
18:1/20:3_25.73_H	1.06454688E8
18:0/20:4_26.89_H	6.813143E9
18:1/20:4_24.96_Na	8.7743848E7
16:0/22:5_24.98_H	5.31407136E8
20:0/18:5_24.98_H	4.6825152E8
18:1/20:4_24.98_H	1.05795392E9
16:0/22:6_23.96_Na	1.85711664E8
16:0/22:6_23.96_H	4.2974746E9
38:7_22.03_Na	831522.56
39:3_28.71_H	1355885.4

The table displays now information about the obtained structure according to Liebisch, G. et al. Shorthand notation for lipid structures derived from mass spectrometry. J Lipid Res 54, 1523-1530 (2013). I.e., the sum of carbon atoms and double bonds in the chains if no structural information can be obtained (e.g. 38:5), the fatty acid chains separated by an underscore if fatty acids are detectable (e.g. 20:4\_18:0), or the fatty acid chains separated by a slash if the position of the fatty acid chains is assignable (e.g. 18:1/20:4). Attention: For the manual inspection of MS<sup>n</sup>, there is a difference in the displayed annotation of the spectra, if the lipid species consists of more than one lipid molecular species:

- The lipid species view is selected: The fragments of all detected lipid molecular species are annotated.
- The lipid molecular species view is selected: Only the fragments of the particularly selected lipid molecular species are annotated.

“Show MSn Only” will lead to LDA displaying only identifications, with at least MSn headgroup evidence.

When “Chain Evidence Only” is checked, MSn evidence for chains is needed too, to show the identification.

If a row of the table is clicked, the peaks are displayed in the 3D and the chromatogram viewer. If a row is clicked by the right mouse button, a popup appears:

56:1_27.39	2.5718754E7
56:2_26.28	3.75773344E8
56:3_25.21	3.12709568E8
56:4_24.51	4.19923552E8
56:5_23.82	
56:6_23.08	
56:7_22.64	
56:8_21.92	
56:9_21.23	
56:10_20.55	
58:1_28.62	
58:2_27.30	
58:3_26.21	
58:4_25.21	
58:5_24.82	
58:8_22.53	1.05472192E8
58:9_22.00	1.48642176E8
58:10_21.53	7.6466584E7
58:11_20.80	8695020.0

This popup menu provides options for adding/deleting analytes, sorting options, and options for viewing MS<sup>n</sup> annotations, and developing decision rule sets. To delete several analytes at once, keep Ctrl or Shift pressed and select the analytes; this will prevent refreshing of the 3D viewer and save time. As soon as you release Ctrl or Shift, the 3D viewer will be refreshed by the analyte that was selected last. Before/after just determines the position where the analyte will be added.

Enter new analyte

Add a new analyte before 38:5\_24.98\_H

name:	38:5	formula:	H82 P1 O8 N1 C46
modification:	H	formula:	H1
m/z tolerance:	0.013	exact mass:	
charge:	1	RT:	24.98

Cancel OK

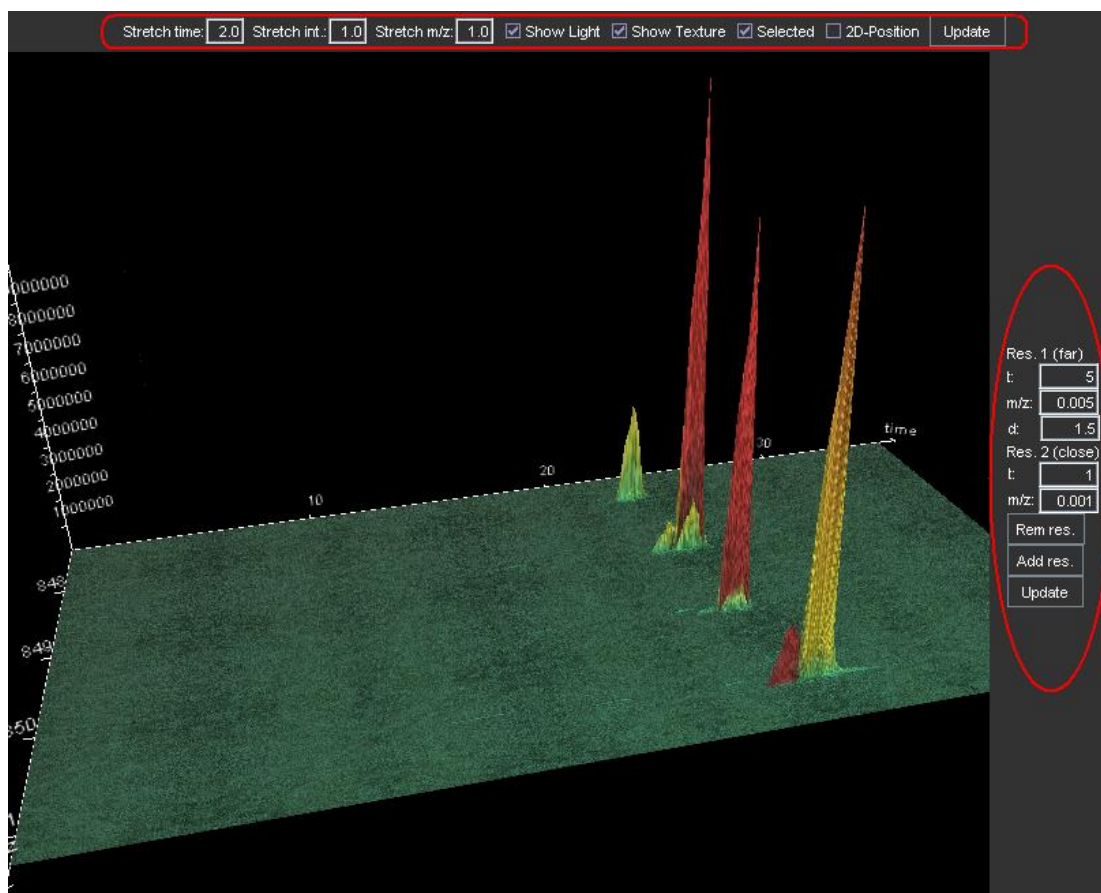
The adding of a new analyte requires the name of the analyte, its chemical formula, the name and the formula of the modification if any, the m/z tolerance of the chromatogram, the exact mass, the charge and the expected retention time.

There are three sorting options available:

- Sort by order: It is the same order as the analytes are stored in the result Excel files(default).
- Sort by intensity: Starting with the most abundant analyte, the analytes are sorted in descending order.
- Sort by mass: Starting with the lowest m/z, the analytes are sorted in ascending order.

“Show MS/MS” displays the spectra as annotated by the used MS<sup>n</sup> decision rules at the time of the quantitation (see chapter 5.4 “Visualization – Spectrum viewer”). Even when other analytes are clicked, LDA will display now MS<sup>n</sup> spectra. For returning to MS1 display, right-click again on the analyte, and select the now available “Show MS1”, or click on an analyte where no spectral information is available (white color). “Edit MSn rule” shows the annotated spectra based on the currently active rules – this option allows for definition/optimization of decision rules and helps in finding reasons for the algorithm’s decision. For further details see chapter 6.3 “GUI for MS<sup>n</sup> rule definition/review”. “Recalculate MSn” repeats the MS<sup>n</sup> identification for the chosen peak based on the currently active decision rule sets. The new identifications will be displayed, and by clicking on “Accept” permanently stored. By clicking on “Edit Rt”, the retention time standing next to the hit can be manually changed.

## 5.2 Visualization – 3D viewer



The 3D viewer provides a menu bar at the top where the appearance of the viewer can be adjusted, and a menu bar on the right side which is responsible for the resolution of the viewer. The view on the plane can be changed when the mouse is clicked inside the drawing canvas.

### Top menu bar:

- Stretch time/int./m/z: The 3D viewer itself treats every dimension equally, so that every dimension should have the same length. However, for the chromatography, it makes more sense to stretch the time axis. Thus, the “stretch time” is automatically set to 2.0. Any of the three dimensions can be stretched or shrunk with these three input fields. The settings are applied after the “Update” button has been pressed.
- Show Light: The 3D viewer has a light source falling on the plane, resulting in bright or shadowy spots. The light effects can be switched off with this check box.
- Show Texture: In order to give the picture more contrast, a texture is rendered over the plane. The texture can be turned off with this check box.
- Selected: If selected, the quantified and stored peaks are colored in red; the ones that are only quantified and not stored are colored in green. If not selected, the coloring depends on the signal intensity (green: low intensity; yellow: medium intensity; red: high intensity).
- 2D-Position: This shows the stripe that has been taken to extract the chromatogram in gold.

### Right menu bar:

- t: Is the time resolution in seconds.
- m/z: Is the m/z resolution in Thompson.
- Add res.: The viewer has the ability to change the resolution level depending on how close the view is to the plane. Several resolution levels can be entered. If an additional resolution

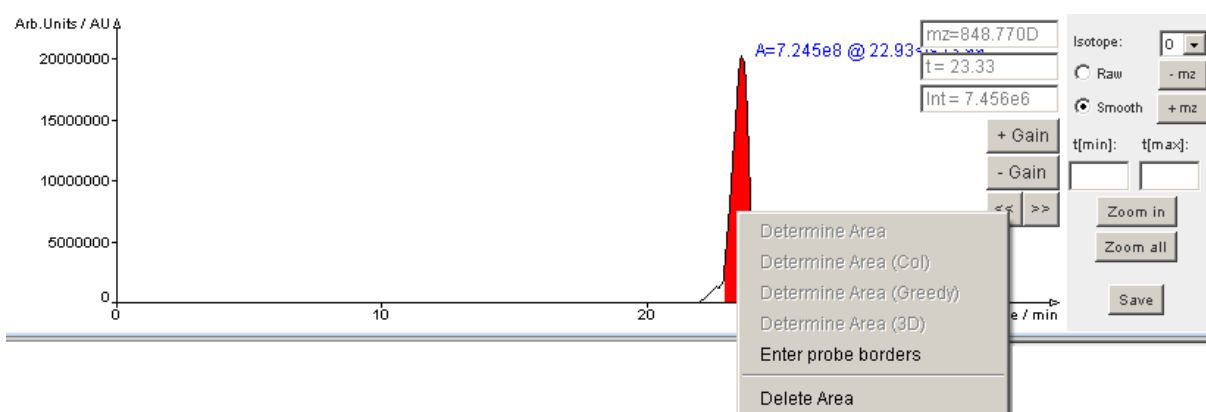
level is required, just press this button. The point where the resolution is changed is defined by the d field.

- d: Is the distance field; it appears if several resolution levels are set. The distance is always measured from the center of the plane. 1 corresponds to a total distance of the m/z axis to the center of the plane. This value is rather empirical, but due to general applicability to different resolutions this approach had been chosen.
- Rem res.: Removes a resolution level
- Update: This button has to be pressed when the resolution settings should take effect for the 3D view.

#### Mouse buttons:

- Left mouse button: Changes the viewing angle on the canvas.
- Middle mouse button: Zooming in and out.
- Right mouse button: Shifting the plane.

### 5.3 Visualization – Chromatogram viewer



If the mouse is hovered over the painting canvas, a crosshairs cursor is rendered, whereupon the current time and intensity is displayed in the two text fields on the right side of the canvas (e.g. “t = 23.33” and “Int = 7.456e6”). The third text field indicates the m/z value of the chromatogram. With +/- Gain, the displayed intensity range can be zoomed, and with “<<” and “>>”, it is possible to shift the currently zoomed time range. If a peak is quantified and stored, it is displayed in red; if it is just quantified and not stored, it is displayed in green. If the right mouse button is clicked under a curve, a popup menu appears with the following options:

- Determine Area (3D): 3D algorithm for more accurate peak border confinement.
- Determine Area: Detects a peak area with the standard ASAPRatio algorithm.
- Determine Area (Col): Detects a peak area with the MASPECTRAS algorithm that integrates over local maxima.
- Determine Area (Greedy): Sets the peak borders at sudden changes in the steepness of the curve.
- Delete Area: Deletes a quantified area.
- Enter probe borders: Manually define the borders of the peak:

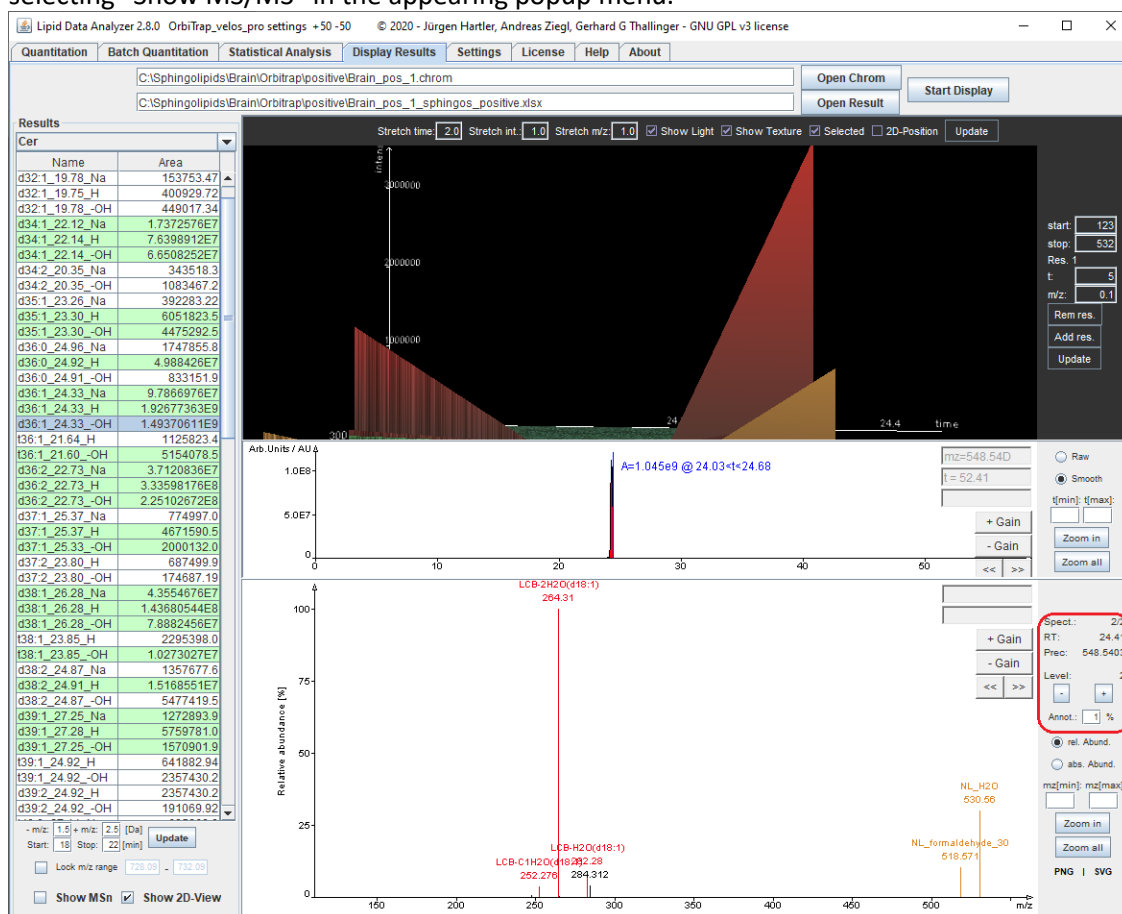
The peak can be defined in 2D or 3D mode. In 2D mode the chromatogram is used and just the start- and stop time has to be entered. For the 3D mode, the m/z values have to be entered, and an ellipse is fitted through the four border values.

In the right menu of the chromatogram viewer, the following options are available:

- Isotope: Switches between the chromatograms of the different isotopes.
- Raw/Smooth: Toggles between the raw and the smoothed version of the chromatogram (smoothed is the default one).
- t[min], t[max] and Zoom in: This is for changing the shown part of time axis. With t[min] and t[max] the display range is defined and with “Zoom in” the settings are used for the viewer. After zooming, the << and >> buttons can be used.
- Zoom all: displays the full time axis again.

## 5.4 Visualization – Spectrum viewer

This view is for the display of MS<sup>n</sup> spectra. It can be reached by a right-click on the lipid species, and selecting “Show MS/MS” in the appearing popup menu.



The topmost view shows the MS/MS fragments in 3D view (the viewer works as the 3D viewer described in 5.2). The center display shows the MS<sup>1</sup> chromatogram (the chromatogram viewer works as described in 5.3, except that peaks cannot be saved), and at the bottom is the spectrum viewer. On the right (encircled in red), the user sees the spectrum sequence number for this precursor, separated by a slash from the number of all available spectra for this precursor. Below is the retention time (RT) of the current spectrum and the exact precursor mass. By clicking on +/-, the user can scroll through the spectra. “Annot.” defines the threshold for annotating unidentified peaks by mass. The rest of the functionality is the same as described in chapter “Visualization – Chromatogram viewer”. In this view, the identified fragments are displayed, independent of the currently used fragmentation settings (only ms2PrecursorTolerance and ms2MzTolerance of the LipidDataAnalyzer.properties of the MS instrument influence the displayed fragments). The



identification in the quantitation is based on all MS<sup>n</sup> spectra inside the range of the MS<sup>1</sup> peak. On the fly identifications based on single spectra can be done in the GUI for MS<sup>n</sup> rule definition (see chapter 6.3).

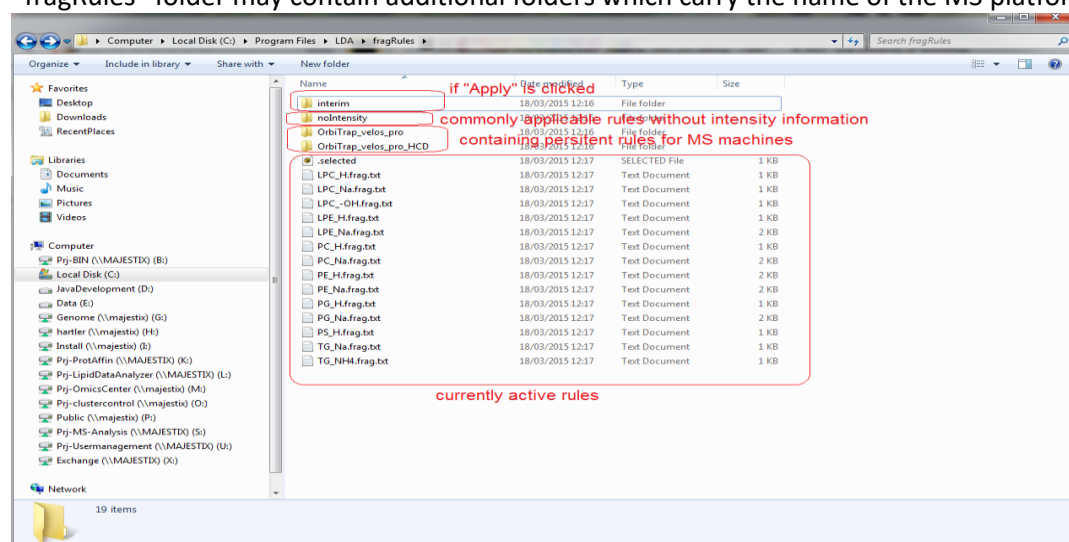
## 6 MS<sup>n</sup> spectra interpretation – decision rule sets

LDA uses Excel files containing the lipid MS<sup>1</sup> species (name, chemical formula, masses) to identify peaks from MS<sup>1</sup> data (see Appendix A). However, due to type of mass spectrometer, to collision energy applied, to adduct ions and charge state, MS/MS spectra of lipids vary heavily. Therefore, in LDA, structural information of MS<sup>n</sup> spectra can be obtained by flexible “decision rule sets” which can be easily adapted to individual MS platforms and setups. These rule sets are based on well-defined fragments and their intensity relationships which are stored in decision rule sets (.frag.txt files). These rule sets are located in the “fragRules” folder of the LDA installation directory. In order to select the correct rule, LDA uses a combination of the lipid subclass name (the name of the Excel tab) and the name of the adduct (the name inside the square brackets “name” in the mass column header). E.g. in the example of Appendix A, the Excel tab is named “TG” (marked by 1. in the picture in Appendix A) and the first mass column header is “mass(form[+NH4] name[NH4])” defining the adduct name as NH4. For MS<sup>n</sup> identification, the software will look for a file named TG\_NH4.frag.txt in the fragRules directory. If this file is not there, the algorithm will do the identification based on MS<sup>1</sup> data only.

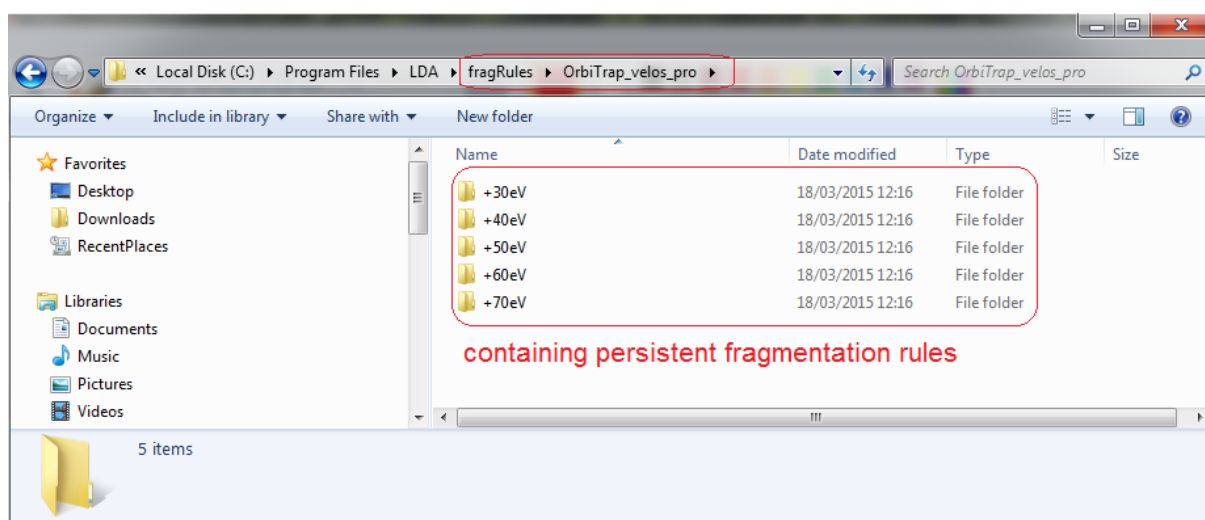
In the following paragraphs, it will be explained how these decision rule sets are organized (see “Organization of decision rule sets”), which parameters are available in the rule sets (see “Decision rule set file (frag.txt)”), and a how the graphical user interface for rule development and visualization works (see “GUI for MS<sup>n</sup> rule definition/review”).

### 6.1 Organization of decision rule sets

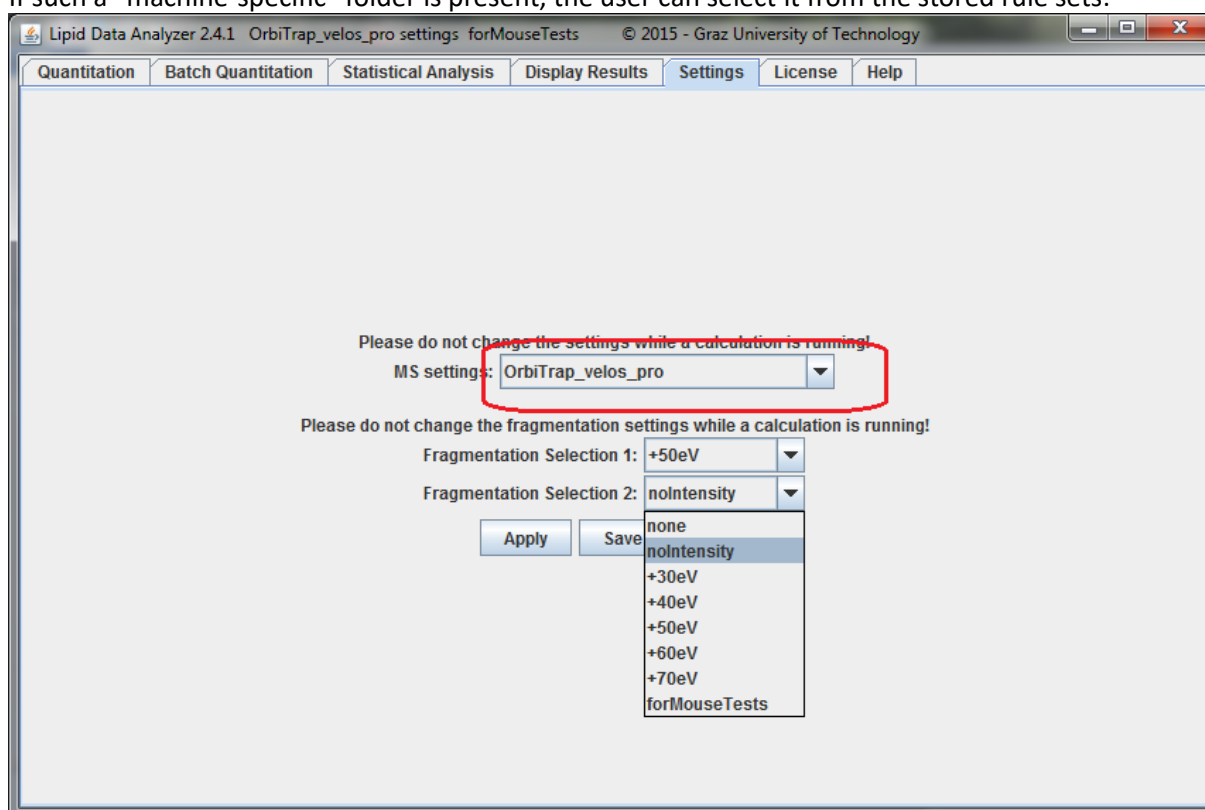
The LDA is applicable to several mass spectrometers and various instrumental setups. However, the fragmentation rules differ from instrument to instrument and depend on settings such as the used collision energies. In order to allow for persistent rule sets which can be selected easily, the “fragRules” folder may contain additional folders which carry the name of the MS platform.



Subfolders within these folders contain fragmentation rule sets for this MS platform. These subfolders can have arbitrary names (e.g. collision energy settings).



If such a “machine-specific” folder is present, the user can select it from the stored rule sets:

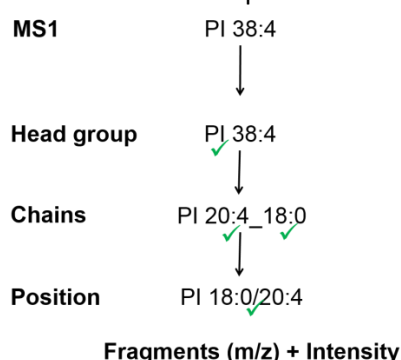


In the present case, the machine Orbitrap\_velos\_pro is selected and the “fragRules” folder contains such a folder containing the subfolders +30eV, +40eV, +50eV, +60eV, +70eV and forMouseTests; these rules are now selectable for the rule sets. If there had been no Orbitrap\_velos\_pro folder, only “none” (for no MS<sup>n</sup> identification) or “noIntensity” (for unspecific MS<sup>n</sup> fragment identification) would have been selectable. By this mechanism, users can easily create rule sets and switch among them. By clicking on “Save as default”, the rules will be transferred to the “fragRules” folder where they are active for quantitation or rule definition. From now onwards, these settings will be automatically loaded whenever LDA is started freshly. “Save as default” should be clicked only if there is no active quantitation running. If LDA is quantifying and the user wants to check data from another machine, define new rule sets, or quantify another data set, he/she can start a new LDA instance, select the appropriate settings and click on the “Apply” button. The rules are now stored in the “fragRules/interim” folder; the user can now work with them without interfering with the running quantitation.



## 6.2 Decision rule set file (frag.txt)

The decision rules for MS<sup>n</sup> interpretation are stored in frag.txt files. Their naming and their storage are explained in the previous paragraphs (see chapters 6 and 6.1). This paragraph gives details about the structure and the parameters of these files.



MS<sup>1</sup> data returns candidates for the intact lipid species, e.g. 38:4. The LDA may obtain the following structural information:

- Verification of the lipid subclass: i.e. a confirmation of detected lipid species which is done by so-called head group fragments; e.g. a verified PI38:4.
- Detection of constituent fatty acid chains: i.e. a detection of one or several fatty acid chain combinations, e.g. PI20:4\_18:0.
- Assignment of sn-positions: i.e. an intensity comparison of detected fatty acid chain fragments; e.g. PI18:0/20:4 where the slash denotes that the positions are found.

The premise of the decision rule sets is to provide users an easy, flexible and straightforward way to configure the spectral interpretation. In fact, decisions are based on well-defined fragments and their intensity relationships which are stored in so-called fragment rules and intensity rules, respectively.

**Fragment rules** are defined by five parameters, and two optional parameters (oh and combiOh):

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

- Name: An arbitrary name for the fragment.
- Formula: The chemical formula, i.e. the instruction for calculating the mass of the fragment. Here, placeholders can be used, such as \$PRECURSOR for the mass of the precursor, and \$CHAIN as

variable for an fatty acyl chain, or \$ALKYLCHAIN, \$ALKENYLCHAIN and \$LCB for fatty alkyl chain, fatty alkenyl chain, and long chain base, respectively. Furthermore, fragments previously defined can be used too.

- Charge: the charge state of the fragment.
- MSLevel: the MS<sup>n</sup> level at which this fragment is observed, e.g., 2 is for MS/MS, 3 is for MS<sup>3</sup>, etc.).
- mandatory: This parameter has three options: i) true, the fragment must be observed; ii) false, the fragment is not always observed in the spectra; iii) other, the fragment is observed in another lipid subclass/adduct; iv) class: the chain fragment is mandatory for the detection of a lipid class (only valid in the [CHAINS] section). The detection of mandatory=true and mandatory=class fragments is essential, while "false" and "other" fragments can be used in intensity rules to find decisions. A peculiarity of "other" fragments is that they are excluded from the calculation of the spectrum coverage to permit the detection of mixed spectra from different lipid subclass/adducts. The exact effects of this parameter are discussed in the following paragraphs of this chapter. Furthermore, in the case it is mandatory to detect one out of two or more fragments, an intensity rule with the '|' (OR) operator can be used (see **Intensity rules**).
- oh (optional): this parameter may be used for classes where different hydroxylation stages are to be expected, such as sphingolipids. This parameter indicates the hydroxylation stage of the fragment. When there is no 'oh' parameter specified, it is assumed that this fragment is detectable at all possible hydroxylation stages. Since the abundance of a fragment often depends on the hydroxylation stage, this parameter allows for overwriting of the default

‘mandatory’ parameter. E.g., ‘mandatory=true’ followed by ‘oh=1=false,2,3,4’ will define that the fragment must be detectable for species containing two, three and four hydroxylation sites, but not for the ones containing one hydroxylation site. Details for the effect of the (overwritten) mandatory settings are given in the description of the ‘mandatory’ parameter, above. ATTENTION: in the [HEAD] section, the hydroxylation stage pertains to the total number of hydroxylation sites of the lipid species, while in the [CHAINS] section, the hydroxylation stage pertains to the particular long chain base (LCB) and fatty acyl moiety, respectively.

- combiOh (optional): this parameter may be used for classes where different hydroxylation stages are to be expected, such as sphingolipids. This parameter indicates the hydroxylation stage of a partnering chain. Correspondingly, this parameter is available only in the [CHAINS] section. Details are given in the description of the [CHAINS] section.

**Intensity rules** reflect the fragmentation behavior of the lipid subclasses/adducts by the definition of intensity relationships. Intensity rules consist of two parameters:

#### 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

- Equation: Previously defined fragments are put in relation to one another. The equations allow for a “greater than” or “smaller than” symbol and multipliers for the fragments to allow for the definition of relations. Furthermore, plus signs and parentheses are available. The placeholder \$BASEPEAK is for the declaration of a minimum intensity of a fragment. Position information can be introduced in square brackets. Furthermore, the OR-operator ‘|’ can be used to define fragments where one out of several fragments must be found, e.g., ‘Equation=NL\_Carboxy|NL\_Ketene’.

- mandatory: If true, the rule has to be fulfilled, false, a violation of the rule is allowed. The exact effects of this parameter are discussed in the following paragraphs of this chapter.

These two types of rules are used to obtain the various levels of structural information.

**A decision rule set file** (frag.txt) is a text-based file which is divided into the four parts [GENERAL], [HEAD], [CHAINS], and [POSITION] according to the obtainable levels of structural information. Beside the [GENERAL] section, all other three sections comprise solely the previously described fragment and intensity rules where the fragment subsection is initiated by the key string “!FRAGMENTS”, and the intensity subsection by the key string “!INTENSITIES”. Furthermore, each of the latter three sections are optional and can be omitted if inappropriate, except for the [CHAINS] section which is required if a [POSITION] section is present.

The first section [GENERAL], contains general information about the current lipid subclass/adduct to define preliminaries for the MS<sup>n</sup> algorithm. It provides the following parameters where “mandatory” denotes parameters which must be declared, while “optional” parameters can be declared:

```
[GENERAL]
AmountOfChains=2
ChainLibrary=fattyAcidChains.xlsx
CATomsFromName=D*(\d+):\d+
DoubleBondsFromName=D*\d+:(\d+)
BasePeakCutoff=0.1%
SpectrumCoverage=10%
RetentionTimePostprocessing=true
```

- AmountOfChains (mandatory): Defines the amount of constituent fatty acid chains; e.g. 2 for PI.
- AlkylChains (optional): How many of the chains are alkylated chains.
- AlkenylChains (optional): How many of the chains are alkenylated chains.

- AmountOfLCBs (optional): How many of the chains are long chain bases.

- ChainLibrary (mandatory): name of library containing possible fatty acid chains (is typically the default one that comes with LDA – dLCB.xlsx: this library contains the dihydroxylated long chain bases; the other hydroxylation stages are calculated by LDA); the libraries are located in the “fattyAcids” directory of the LDA installation.
- LCBLibrary (mandatory when the analyte contains LCBs): name of library containing long chain bases (is typically the default one that comes with LDA – fattyAcidChains.xlsx); the libraries are located in the “fattyAcids” directory of the LDA installation.
- CAtomsFromName (mandatory): Java parsing rule to obtain the total number of carbon atoms in the chains from the lipid species name; this is usually the default one: `\D*(\d+):\d+`.
- DoubleBondsFromName (mandatory): Java parsing rule to obtain the total number of carbon atoms in the chains from the lipid species name: this is usually the default one: `\D*\d+:(\d+)`.
- SingleChainIdentification (optional): Default is false. If true, the chain combination can be fulfilled if only one chain was found. This option has to be used with precaution, since it may introduce many false positives. In such a case, there have to be sufficient chain intensity rules to verify the existence of the identified chain. Nevertheless, this option can be quite useful for checking spectra if single chains were detected in the graphical user interface (see chapter 6.3).
- ChainCutoff: (optional): Default is 0. This is a cutoff relative to the sum of intensities originating from one chain combination. If a chain combination falls below this threshold, it is discarded. This value is usually set in the LipidDataAnalyzer.properties for the whole experiment (usually 1%), however, certain lipid subclasses tend to have noisy spectra; in such a case, it is advised to increase this value.
- BasePeakCutoff (optional): Default is 0. This is an intensity threshold relative to the base peak. This is not the only applied cutoff; LDA automatically calculates a noise level for each spectrum and removes intensities that fall below this threshold.
- SpectrumCoverage (optional): Default is 0. This value assesses a threshold in percent of covered spectral intensities by annotated fragments to declare a hit valid. Typically, peak discard shall be performed by head group rules, however, there are lipid subclasses such as TG and DG which do not show specific head group fragments. In such cases, peak specificity can be reached by the spectrum coverage. For lipid subclasses where head groups are present, it is advised to keep this parameter rather low, because the head rules are usually more robust; particularly in noisy spectra, the spectrum coverage can be quite low, despite the included noise filtering. Thus, head rules are much more specific and allow for reliable peak detection in noisy spectra.
- LcbHydroxylationRange (optional): the range of hydroxylation sites allowed for the long chain base (LCB) – typically, ‘1-3’.
- FaHydroxylationRange (optional): the range of hydroxylation sites allowed for the fatty acyl moiety (FA) – typically, ‘0-1’.
- RetentionTimePostprocessing (optional): Default is false. If true, the detected hits of this subclass/adduct are filtered by a retention time prediction algorithm which works based on found  $MS^n$  hits. The retention time is predicted based on the number of carbon atoms and double bonds, and can remove identifications which are not in the proximity of the prediction. The proximity is dynamically calculated as four times the mean deviation of the  $MS^n$  hits relative to the predicted retention time. At least six verified  $MS^n$  hits are required to calculate the predicted retention times.
- RetentionTimeParallelSeries (optional): Default is false. For some lipid subclasses, there are many false positive peaks at the same  $m/z$  value, e.g., TG+Ethylamine has approximately the same mass of TG+NH<sub>4</sub> with 2 more carbon atoms which can cause false positive identifications. Even though these hits are usually removed by the fragmentation rules, hits where there are no  $MS^n$  spectra detected may remain. These hits can be removed by this option.
- RetentionTimeMaxDeviation (optional): If the dynamic determination of the proximity to the predicted retention time value is not desired, a fixed threshold can be set.
- ValidOnlyWithOtherAdduct (optional): the current adduct is accepted only if other adducts may be identified (OtherAdductValidityTolerance must be set for this parameters). If several other

adducts are involved, they are separated by '|' if the detection of one of them is sufficient, or by comma, if all of them have to be found.

- **OtherAdductValidityTolerance** (can be used only in conjunction with **ValidOnlyWithOtherAdduct**): this gives the maximum retention time tolerance in minutes for the other adducts. The tolerance pertains to the retention time of the peak summit.
- **ForceOtherAdductValidity**: This is helpful when the partnering adduct is overlapped by some other species that produces similar spectra. When true, the partnering adduct is selected to be true, and the overlapping species is removed. This is helpful when there overlaps between species/adducts producing similar fragmentation patterns, such as protonated ceramides ( $\text{Cer [M + H]}^+$ ), and protonated ceramides that lost water ( $\text{Cer [M + H - H}_2\text{O]}^+$ ). An overlap occurs, because the trihydroxylated species of ( $\text{Cer [M + H - H}_2\text{O]}^+$ ) has exactly the same mass as the dihydroxylated protonated species ( $\text{Cer [M + H]}^+$ ). While the sodiated species ( $\text{Cer [M + Na]}^+$ ) is minor compared to the other species, its presence can help to select the correct species. This can be done by setting the **ForceOtherAdductValidity**=true in the fragmentation rules of the sodiated Cer species.
- **ChooseMoreLikelyRtWhenOtherAdductEqual**: This is helpful when the partnering adduct is overlapped by some other species that produces similar spectra, as in the explanation of the "ForceOtherAdductValidity". When both species show exactly the same fragments, this parameters forces to select the presumably correct species based on the retention prediction (see **RetentionTimePostprocessing**): The species where the peak is closer to the predicted retention time is deemed to be correct.
- **MSIdentificationOrder** (optional): Sequence of how MS identifications are performed – default and recommended is "MS1First".
  - **MS1First**:  $\text{MS}^1$  identification is performed first, then, the  $\text{MS}^n$  hits of the found hits are checked.
  - **MSnOnly**:  $\text{MS}^n$  spectra are checked for matching identifications, then,  $\text{MS}^1$  quantitation is performed only at time points where  $\text{MS}^n$  spectra have matched the rules.
  - **MSnFirst**:  $\text{MS}^n$  spectra are checked for matching identifications, then,  $\text{MS}^1$  is performed at time points where  $\text{MS}^n$  spectra matched the rules. Subsequently, a retention time curve is fitted, and  $\text{MS}^1$  identifications are performed at these time points
- **AddChainPositions** (optional): Default is 0. This option defines if there are more positions available than constituent fatty acids; e.g. DG has 2 fatty acid chains which are possible at 3 *sn*-positions; for DG, this value is 1.
- **ClassSpecificMS1Cutoff** (optional): The  $\text{MS}^1$  peak detection algorithm uses two thresholds to eliminate low abundant peaks (**relativeAreaCutoff** and **relativeFarAreaCutoff**). Typically, peaks smaller than 1% of the highest peak in one chromatogram are removed; if it is known that one lipid subclass/adduct is usually detected in the presence of an isobaric lipid subclass/adduct of higher abundance, this threshold can be lowered by this parameter to detect such low abundant species
- **IsobarSCExclusionRatio** (optional): If there are fragments of several isobaric species present, this is a threshold to exclude the smaller lipid subclass/adduct and assign the peak as a whole to the species with the more intense fragments. This value is typically set in the **LipidDataAnalyzer.properties** for the whole experiment (usually 1%), but there are isobaric subclasses requiring higher thresholds such as **PC\_-CH3** and **PE\_H** which are primarily dominated by chain fragments.
- **IsobarSCFarExclusionRatio** (optional): If there are fragments of several isobaric species originating from different subclasses/adducts present, they are usually chromatographically separated. If there are at least two isobaric peaks containing  $\text{MS}^n$  spectra present which are chromatographically separated, and one of them can be assigned to one subclass/adduct, a

higher threshold is used for the removal of the identified subclass/adduct in the other peak, since it must be assumed that two isobars of the same subclass cannot be chromatographically separated by such a high time difference. This preceding is particularly useful for isobaric subclasses such as PC\_-CH3 and PE\_H which are primarily dominated by chain fragments. This value is typically set in the LipidDataAnalyzer.properties for the whole experiment (usually 10%).

- **IsobarRtDiff** (optional): This value defines the retention time difference between two isobaric peaks for the application of the higher isobarSCFarExclusionRatio. This value is typically set in the LipidDataAnalyzer.properties for the whole experiment, since it is rather a chromatographic parameter. However, the overlap might be different between various isobaric subclass/adduct combinations.
- **EnforcePeakUnionTime** (optional): if a lipid subclass smears over the chromatogram, this is an option to reunite several segregated MS<sup>1</sup> peaks returning similar MS<sup>n</sup> evidence. This approach is particularly useful for lipid subclasses such as PS which starts to smear as soon as the column starts to degrade. However, the derived quantities of the peaks have to be treated with caution, since the final stretch of the united peak primarily depends on the presence of MS<sup>n</sup> spectra throughout the smearing peak, and their generation is typically a slightly arbitrary procedure.
- **IgnorePositionForUnion** (optional): This parameter is evaluated only if the "EnforcePeakUnionTime" is set. If true, peaks are still united, even though single spectra return contradicting evidence for the position of the fatty acids. This is usually beneficial, since at low abundant spectra the position detection tends to be unreliable, and such spectra occur typically in smearing peaks.

The second section [HEAD] contains fragment and intensity rules for the verification of the lipid subclasses where the fragments are typically termed as head group fragments.

```
[HEAD]
!FRAGMENTS
Name=PL153          Formula=C3H6O5P      Charge=1      MSLevel=2      mandatory=false
Name=PIhead_223     Formula=C6H8O7P      Charge=1      MSLevel=2      mandatory=false
Name=PIhead_241     Formula=C6H10O8P     Charge=1      MSLevel=2      mandatory=true
Name=PIhead_259     Formula=C6H12O9P     Charge=1      MSLevel=2      mandatory=false
Name=PIhead_297     Formula=C9H14O9P     Charge=1      MSLevel=2      mandatory=false
Name=PIhead_315     Formula=C9H16O10P    Charge=1      MSLevel=2      mandatory=false
Name=NL_Inositol_162 Formula=$PRECURSOR-C6H10O5 Charge=1      MSLevel=2      mandatory=false

!INTENSITIES
Equation=PIhead_241*0.9>PIhead_223 mandatory=false
Equation=PIhead_241*0.9>PIhead_297 mandatory=false
```

Fragments where the mandatory parameter is set to true must be found in the corresponding MS<sup>n</sup> spectra. An absence of such a fragment (in this case PIhead\_241 for PI\_-H) will cause the discard of an MS<sup>1</sup> peak. The detection/absence of mandatory=false or mandatory=other fragments won't cause any discard, but these fragments can be used in intensity rules. The same is true for intensity rules; if mandatory intensity rules are violated, the MS<sup>1</sup> peak will be discarded. Therefore, only mandatory intensity rules impact the outcome of the [HEAD] section, while optional (mandatory=false) intensity rules can be used only for testing if a rule might work. From here onwards, the MS<sup>1</sup> peak is verified and cannot be discarded by the [CHAINS] or [POSITION] section unless the spectrum coverage is not fulfilled, or a mandatory=class fragment is present in the [CHAINS] section. Additionally, for classes with varying hydroxylation stages, the 'oh' parameter can be set to indicate the hydroxylation stage where this particular fragment can be observed. In the [HEAD] section, the given hydroxylation number pertains to the total number of hydroxylation sites of this lipid species, irrespective of its distribution among the long chain base and fatty acyl moiety. Details about the usage of the 'oh' parameter can be found in the description of the **Fragment rules**.

The third section [CHAINS] contains fragment and intensity rules for the detection of the constituent fatty acid chains.

```
[CHAINS]
!FRAGMENTS
Name=Carboxy          Formula=$CHAIN-H          Charge=1          MSLevel=2          mandatory=false
Name=NL_Carboxy       Formula=$PRECURSOR-$CHAIN Charge=1          MSLevel=2          mandatory=true
Name=NL_Ketene        Formula=$PRECURSOR-$CHAIN+H2O Charge=1          MSLevel=2          mandatory=false
Name=NL_Carboxy_Ino   Formula=NL_Carboxy-C6H10O5 Charge=1          MSLevel=2          mandatory=false
Name=NL_Ketene_Ino    Formula=NL_Ketene-C6H10O5 Charge=1          MSLevel=2          mandatory=false

!INTENSITIES
Equation=NL_Carboxy*0.8>NL_Ketene          mandatory=true
Equation=Carboxy+NL_Carboxy_Ino+NL_Carboxy*0>0*$BASEPEAK          mandatory=true
```

The MS<sup>n</sup> algorithm will check all possible fatty acid chains for the presence of their fragments in the spectra. Again, the absence of mandatory fragments will cause the discard of the fatty acid chain (not of the MS<sup>1</sup> peak; except mandatory=class is set and not a single chain fragment of this type is found, then the whole lipid species will be discarded), and again, the violation of mandatory intensity rules will cause the discard of the corresponding fatty acid chain too. The result of the rules check is a list of potential fatty acid chains. Next, all fatty acid combinations will be discarded where no evidence is present for all constituent chains of a combination, except if SingleChainIdentification is set to true in the [GENERAL] section. Finally, minor contributions that fall below the ChainCutoff will be discarded too. The result of the [CHAINS] section is a list of identified fatty acid combinations, such as e.g. 20:4\_18:0, 20:3\_18:1, or if there are no chains detectable, the algorithm returns the confirmed lipid species, e.g. 38:4. Additionally, for classes with varying hydroxylation stages, the 'oh' parameter can be set to indicate the hydroxylation stage where this particular fragment can be observed. In the [CHAINS] section, the given hydroxylation number pertains to the total number of hydroxylation sites on this particular chain, irrespective of the number of the hydroxylation sites on the partnering chain(s). Details about the usage of the 'oh' parameter can be found in the description of the **Fragment rules**. Furthermore, for classes with varying hydroxylation stages where the presence of a particular fragment depends on the partnering chain, the 'combiOh' parameter can be used. In order to designate the partnering chain, the following notation must be used: \$CHAIN\_TYPE\$ \$HYDROXYLATION\_STAGE\$=\$MANDATORY\_VALUE\$; e.g., 'combiOh=\$LCB2=true' means that this fragment must be observed when there is a partnering dihydroxylated long chain base. The values for \$CHAIN\_TYPE\$ are \$CHAIN, \$ALKYLCHAIN, \$ALKENYLCHAIN and \$LCB for fatty acyl chain, fatty alkyl chain, fatty alkenyl chain, and long chain base, respectively.

The fourth and final section [POSITON] will be evaluated only if there are chain combinations detected.

```
[POSITION]
!INTENSITIES
Equation=Carboxy[1]*0.9>Carboxy[2]          mandatory=true
Equation=NL_Carboxy[2]*0.5>NL_Carboxy[1]    mandatory=true
Equation=NL_Ketene[2]*0.8>NL_Ketene[1]       mandatory=false
Equation=NL_Carboxy_Ino[2]*0.8>NL_Carboxy_Ino[1] mandatory=true
Equation=NL_Ketene_Ino[2]*0.8>NL_Ketene_Ino[1] mandatory=false
```

This section consists of intensity rules only. A rule will be evaluated only, if at least one fragment on each side of the relational operator is detected. If there are evaluable mandatory=true rules left, all of them must be fulfilled to assign a position. If there are only mandatory=false rules left, an assignment is made only if a majority of the rules returns the same position. This concept is particularly useful if the position information is reliably present in fragments of low abundance (mandatory=true), and less reliably in the high abundant ones (mandatory=false). If the low abundant fragments are detectable, the positional assignment is based solely on them, while in their absence the less reliable high abundant ones are used.

The next picture shows an example of a whole decision rule set file; i.e. PI\_H for an Orbitrap\_velos\_pro at -50:

```
[GENERAL]
AmountOfChains=2
ChainLibrary=fattyAcidChains.xlsx
CAtomsFromName=\D*(\d+):\d+
DoubleBondsFromName=\D*\d+:(\d+)
BasePeakCutoff=0.1%
SpectrumCoverage=10%
RetentionTimePostprocessing=true

[HEAD]
!FRAGMENTS
Name=PL153          Formula=C3H605P      Charge=1      MSLevel=2      mandatory=false
Name=PIhead_223     Formula=C6H807P      Charge=1      MSLevel=2      mandatory=false
Name=PIhead_241     Formula=C6H1008P     Charge=1      MSLevel=2      mandatory=true
Name=PIhead_259     Formula=C6H1209P     Charge=1      MSLevel=2      mandatory=false
Name=PIhead_297     Formula=C9H1409P     Charge=1      MSLevel=2      mandatory=false
Name=PIhead_315     Formula=C9H16010P    Charge=1      MSLevel=2      mandatory=false
Name=NL_Inositol_162 Formula=$PRECURSOR-C6H1005 Charge=1      MSLevel=2      mandatory=false

!INTENSITIES
Equation=PIhead_241*0.9>PIhead_223      mandatory=false
Equation=PIhead_241*0.9>PIhead_297      mandatory=false

[CHAINS]
!FRAGMENTS
Name=Carboxy          Formula=$CHAIN-H      Charge=1      MSLevel=2      mandatory=false
Name=NL_Carboxy       Formula=$PRECURSOR-$CHAIN Charge=1      MSLevel=2      mandatory=true
Name=NL_Ketene        Formula=$PRECURSOR-$CHAIN+H2O Charge=1      MSLevel=2      mandatory=false
Name=NL_Carboxy_Ino   Formula=NL_Carboxy-C6H1005 Charge=1      MSLevel=2      mandatory=false
Name=NL_Ketene_Ino    Formula=NL_Ketene-C6H1005 Charge=1      MSLevel=2      mandatory=false

!INTENSITIES
Equation=NL_Carboxy*0.8>NL_Ketene      mandatory=true
Equation=Carboxy+NL_Carboxy_Ino+NL_Carboxy*0>0*$BASEPEAK      mandatory=true

[POSITION]
!INTENSITIES
Equation=Carboxy[1]*0.9>Carboxy[2]      mandatory=true
Equation=NL_Carboxy[2]*0.5>NL_Carboxy[1] mandatory=true
Equation=NL_Ketene[2]*0.8>NL_Ketene[1]  mandatory=false
Equation=NL_Carboxy_Ino[2]*0.8>NL_Carboxy_Ino[1] mandatory=true
Equation=NL_Ketene_Ino[2]*0.8>NL_Ketene_Ino[1] mandatory=false
```

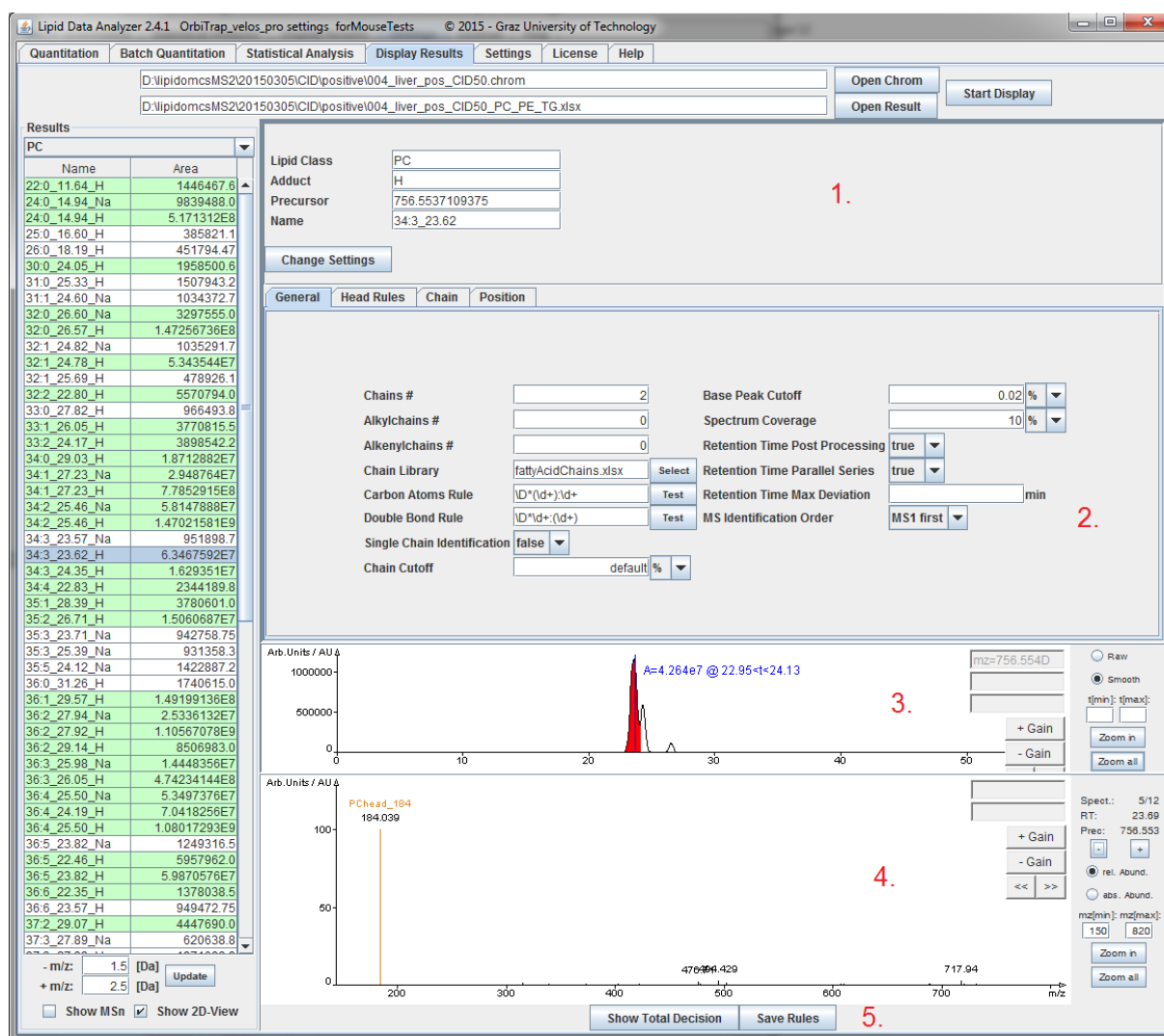
### 6.3 GUI for MS<sup>n</sup> rule definition/review

In order to simplify the development of the text-based decision rule sets (see chapter 6.2), and to obtain direct visual feedback, a graphical user interface is provided. This GUI has two purposes:

- Defining/optimizing fragmentation rules.
- Help finding reasons why an identification does not work as expected.

This GUI can be reached by right-clicking on the analyte and selecting “Edit MS<sup>n</sup> rule” (see chapter 5.1).





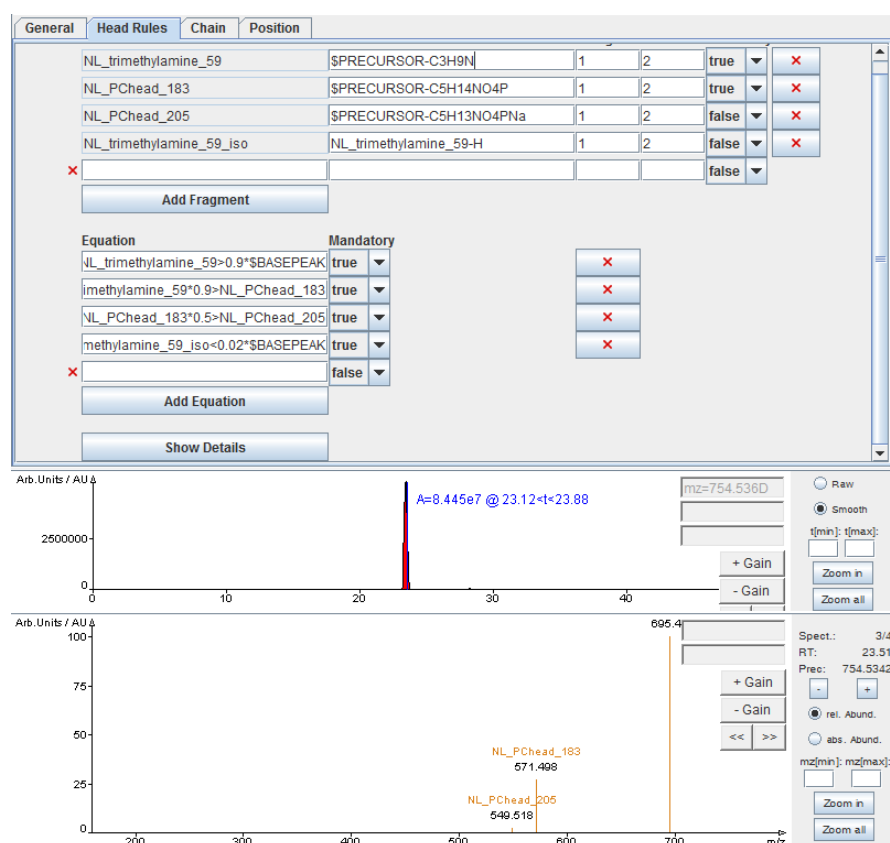
1. General information about precursor selection, lipid subclass, and analyte name – defines the rule name. With this setting, the user can switch to other analytes of this lipid subclass that were not detected, or start with the definition of a completely new rule (new lipid subclass adduct combination).
2. Tabs containing the sections of the rule files (frag.txt): General, Head Rules, Chain and Position. In these tabs, the whole MS<sup>n</sup> identification process can be defined. Changes in these tabs will immediately cause the spectrum viewer to be updated. The only exception is the “General” tab, which updates the spectrum after the user leaves the tab.
3. The chromatogram viewer showing the location of the current spectrum.
4. Spectrum viewer showing the found fragments in the current spectrum.
5. A Button for showing the final identification based on the defined and activated rules, and a button for saving these rules. If “Save Rules” is clicked, the files at “fragRules” will be overwritten, and the rule is active; this should not be clicked while another quantification is running. Only if “Apply” of the settings (see chapter 2) has been used, the rule is written in the interim directory, and does not affect any other running instances of LDA. After “Save Rules” had been clicked, the changed rule is located in the “fragRules” folder. If the settings are changed (see chapter 2), the new rule will be overwritten. For a permanent storage of the rule, copy the corresponding frag.txt file in the appropriate directory for the current collision setup of your MS-platform (see chapter 6.1).



The information about the decision finding is stored in General, Head Rules, Chain and Position tabs. These tabs reproduce the sections of the textual frag.txt file in a graphical environment. Details about the available parameters and their effects can be found in chapter 6.2 “Decision rule set file (frag.txt)”. A short description of the effects of changes on the GUI will be explained in the following few lines.

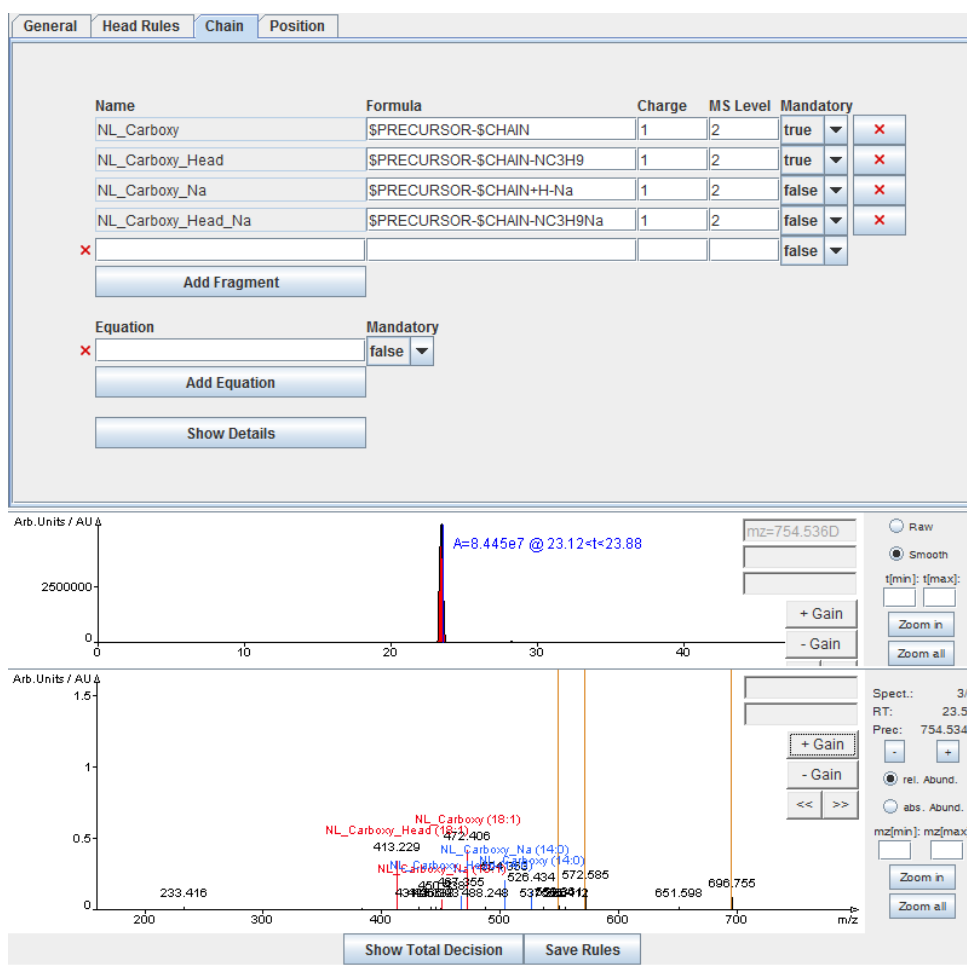
The “General” tab contains general settings for the annotation which has deep impact on the visual output. In order to avoid senseless repainting of the spectra, settings for this tab will have effect as soon as the user activates another tab (is a kind of confirmation that all settings were done).

The “Head Rules” tab is used for fragment definitions and intensity rule definitions of subclass specific fragments (for specificity of MS<sup>1</sup> identification).



If a rule is changed (e.g. the chemical formula), the annotation in the selected spectrum will take effect immediately. If a new fragment or intensity rule is defined, the red X will mutate to a green check box as soon as a fragment is found in the spectrum, or the intensity rule is fulfilled, respectively. “Show Details” gives more information about found fragments and the reasons for discards.

The “Chains” tab is for entering rules for chain identifications. This tab is based on the same concepts as the “Head Rules” tab.



The position tab is for entering rules for position definition. This tab is based on the same concepts as the “Head Rules” and the “Chains” tab, except that only intensity rules can be entered, and the position cannot be shown in the spectrum.

General Head Rules Chain Position

Equation	Mandatory
NL_Carboxy_Head[1]*0.7>NL_Carboxy_Head[2]	true
NL_Carboxy_Head_Na[1]*0.5>NL_Carboxy_Head_Na[2]	true
NL_Carboxy[1]*0.7>NL_Carboxy[2]	false
	false

Add Equation

Show Details

## 7 Statistical Analysis

### 7.1 File selection and settings

The page for the selection of the files can be split into 4 parts whereby changes in this section will effect only the statistics section, after the “Accept” button had been pressed.

Lipid Data Analyzer 2.9.0 OrbiTrap\_velos\_pro settings +50 -50 © 2023 - Jürgen Hartler, Andreas Ziegl, Gerhard G Thalli...

Quantitation Batch Quantitation Statistical Analysis Display Results Settings License Help About

Selection

Add Files Add Dir Remove Remove all Add to group

File Name	Directory
20100126_TAG-34_Massenliste-TAG_mit IS.xlsx	C:\data\lipidomics\20100210
20100126_TAG-35_Massenliste-TAG_mit IS.xlsx	C:\data\lipidomics\20100210
20100126_TAG-36_Massenliste-TAG_mit IS.xlsx	C:\data\lipidomics\20100210
20100126_TAG-37_Massenliste-TAG_mit IS.xlsx	C:\data\lipidomics\20100210
20100126_TAG-38_Massenliste-TAG_mit IS.xlsx	C:\data\lipidomics\20100210
20100126_TAG-39_Massenliste-TAG_mit IS.xlsx	C:\data\lipidomics\20100210
20100126_TAG-40_Massenliste-TAG_mit IS.xlsx	C:\data\lipidomics\20100210
20100126_TAG-41_Massenliste-TAG_mit IS.xlsx	C:\data\lipidomics\20100210
20100126_TAG-42_Massenliste-TAG_mit IS.xlsx	C:\data\lipidomics\20100210

1.

high fat normal fasted

Rename Remove Delete Gro... Rename Remove Delete Gro... Rename Remove Delete Gro...

File Name

20100126\_TAG-34\_Massenliste-TAG\_mit IS.xlsx  
20100126\_TAG-35\_Massenliste-TAG\_mit IS.xlsx  
20100126\_TAG-36\_Massenliste-TAG\_mit IS.xlsx

20100126\_TAG-37\_Massenliste-TAG\_mit IS.xlsx  
20100126\_TAG-38\_Massenliste-TAG\_mit IS.xlsx  
20100126\_TAG-39\_Massenliste-TAG\_mit IS.xlsx

20100126\_TAG-40\_Massenliste-TAG\_mit IS.xlsx  
20100126\_TAG-41\_Massenliste-TAG\_mit IS.xlsx  
20100126\_TAG-42\_Massenliste-TAG\_mit IS.xlsx

2.

TAG

☒ use same settings for all experiments ☒ use same settings for all standards

Dilution factor: 66.67

Ext volume: 20 µL Ext conc.: 1000 µmol/L

Int volume: 50 µL Int conc.: 4 µmol/L

34 35 36 37 38 39 40 41 42

Sample volume: 2 mL End volume: 200 µL

Sample weight: 800 mg

Protein conc.: 178.8 g/L Neutral lipid conc.: 58.64 mg/L

Apply to all Apply to group

3.

Settings for Statistical Analysis

Show hits with different RT separately: ☒ 0.1 min Show MSn Only ☐ Chain Evidence Only ☐ Combine Classes with ox-Classes ☐

4.

Quant file (for correct analyte order): C:\data\lipidomics\massLists\masslist-TG\_with\_IS.xlsx Select

5.

Ion mode: + (selection required for Alex123 target lists)

Internal-standard prefix: IS

6.

External-standard prefix: Ex-IS

Add cutoff settings (Please specify result-files before)

Load settings Save settings Remove absolute settings

Accept

The red box labeled by number 1 is for the selection of files.

- Add Files: This button allows for selection of result files (Excel) for the analysis. The files appear then in the table below.
- Add Dir: Does the same like “Add files” but reads automatically the content of a whole directory and adds all of the Excel files in the directory to the table.

- Remove: Removes one or several results from the table. The files have to be selected before in the table.
- Remove all: Removes all files in the table.
- Add to group: Is for the grouping of results. First, a set of files has to be selected in the table, and then this button has to be pressed. Then, a dialog box requests the name of the group and the group is added to the red box labeled by number 3, the grouping part.

The red box labeled by number 2 displays the groups (they have to be selected at number 1). The grouping is not mandatory, but it can help in the interpretation of results. Each group is displayed in a small table and the following buttons are available:

- Rename: Renames the group.
- Remove: Removes members of one group after they had been selected in the table where the group members are shown.
- Delete Group: Deletes the entire group.

The red box labeled by number 3 is for the standardization on certain values like the protein content or the calculation of absolute values (just valid if there are standards and the ionization efficiencies of the analyte can be assumed to be similar to the ones of the standards). This section is not mandatory. Before this section can be used, the results have to be selected in number 1 first, and the prefixes for the standards have to be defined in number 4. Then, the “Add absolute settings” button has to be pressed. The application reads now the result files and checks for standards.

The upper box of this section is for the assessment of absolute values for the standard. The box has a tab for each lipid subclass, so the settings can be defined specifically for each subclass.

- use same settings for all experiments: If this checkbox is selected, it is assumed that the standards are added to each experiment at the same amounts. Otherwise, specific settings for each experiment are possible.
- use same settings for all standards: If this checkbox is selected, it is assumed that all of the external standards (“Ext volume” and “Ext conc.” have to be entered) and internal standards (“Int volume” and “Int conc.” have to be entered) are added at the same amounts. If not, a list of the standard appears and each standard can be entered separately.

The dilution factor corresponds to the performed dilution between the adding of the external standard until the adding of the internal standards. If the settings are entered for each experiment separately, the buttons “Apply to all” and “Apply to group” appear. These buttons permit the propagation of the entered values to all experiments or to the members of a group.

The lower box is for the specification of sample specific settings like sample volume (before sample preparation) and end volume (after sample preparation). The sample volume and the end volume are mandatory, whereas the “Sample weight”, the “Protein conc.” and the “Neutral lipid conc.” are not. If latter values are entered, standardization on these values is possible. The “Apply to all” and “Apply to group” buttons permit the propagation of the entered values to all experiments or to the members of a group.

The red box labeled by number 4 is for configuring some display settings.

“Show MSn Only” will lead to LDA displaying only identifications, with at least MSn headgroup evidence. When “Chain Evidence Only” is checked, MSn evidence for chains is needed too, to show the identification. “Combine Classes with ox-Classes” will merge all classes with their respective ox - classes, allowing comparison of abundances between the classes.

The red box labeled by number 5 is for the sorting of analytes.

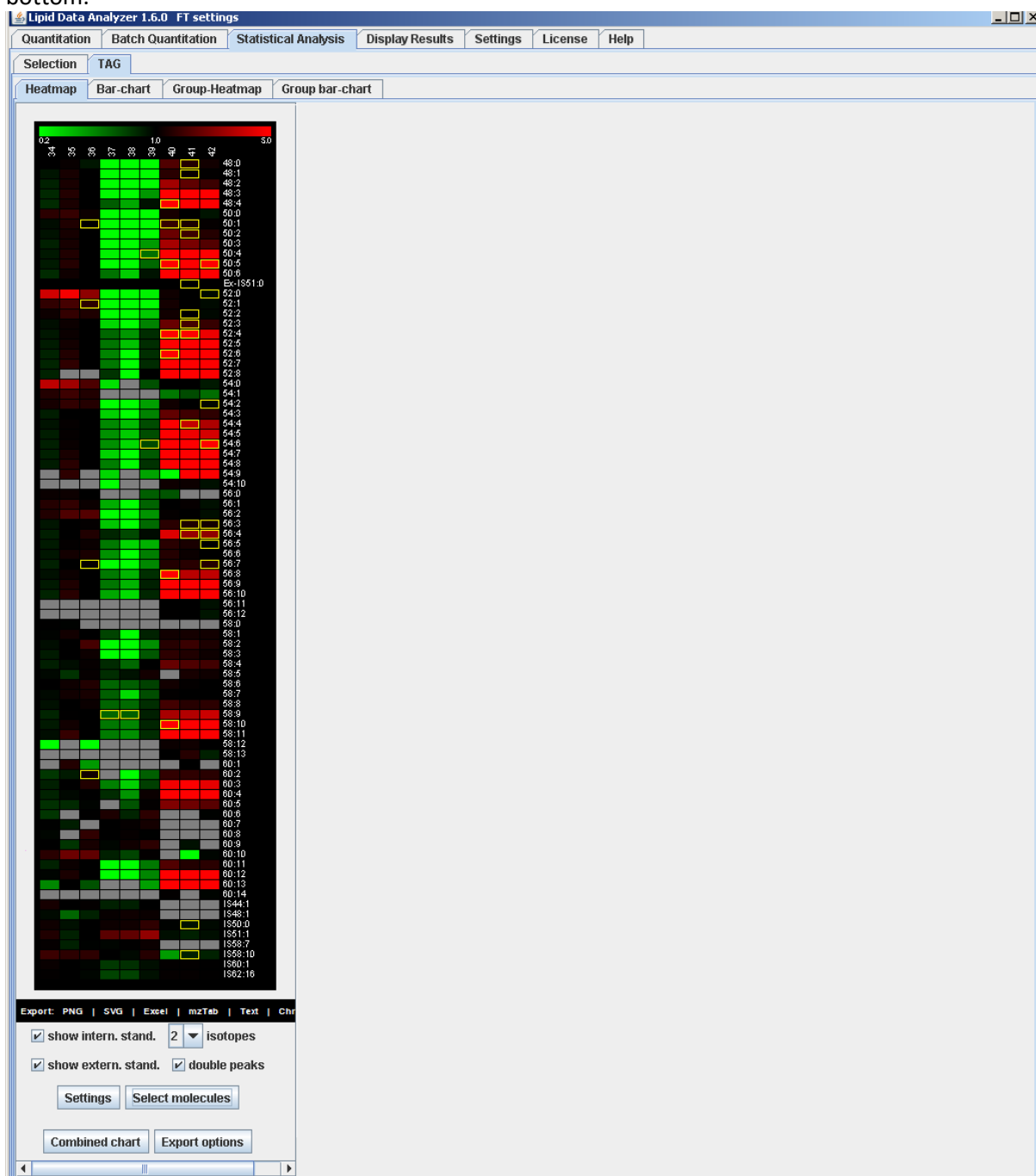
If several analytes cannot be detected in the quantitation step, not all of the analytes appear in the Excel results file. Then, it is hard for the algorithm to specify the correct order. Here, the original quantitation file can be provided. The analytes will be displayed in the order they occur in the Excel

file. The provision of the original Excel file is not mandatory, but can be helpful if several analytes are not detected.

The red box labeled number 6 is for the selection of molecules as reference standard. It is assumed that the standards in the results carry a certain prefix (e.g. IS or Ex-IS). This prefix has to be entered in the two input fields. The usage of standards is not mandatory, so these fields can remain empty. If there are prefixes entered and no compound carries this prefix, the standardization is neglected.

## 7.2 Heat map and visualization settings

After the “Accept” button had been pressed, a tab for each lipid subclass appears. This tab again contains a tabbed pane with the entries “Heatmap” and “Bar-chart”, and if groups are selected “Group-Heatmap” and “Group bar-chart”. The “Heatmap” tab is selected by default. The tab itself contains the heat map at the top, followed by an export bar, and some control elements at the bottom.

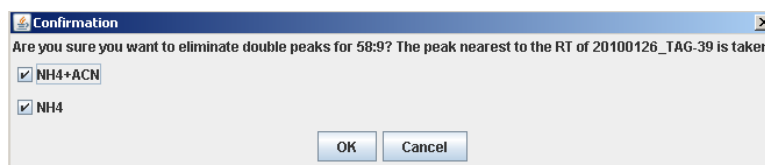


### Heat map:

At the top, the heat map contains a color legend. The values in the map are calculated relative to the median of one molecule over all samples/groups. If the intensity of a molecule of one sample/group is lower than the median, it is colored in green, the ones that are higher are in red, and the ones around the median are in black. If one analyte cannot be quantified, it is in gray. In the heat map, the samples/groups are organized horizontally and the lipids vertically. If a quantitation is based on more than one peak (could be an ambiguous identification; each isotope separately), a yellow rectangle is around the peak. If the mouse is hovered over one heat map cell, a white rectangle is rendered, and at the bottom of the application a line with the name of the lipid, the name of the sample, the value relative to the median, a value depending on the “Settings” option (can be standardized), the original value of the quantitation, and the amount of available isotopes are displayed.

Mouse clicks on cells in the heat map (not valid for the group heat map):

- Left mouse button (does not work for gray fields): The LDA visualizes the quantitation (see chapter 5). If the quantitation is based on several peaks due to adducts/modifications, the LDA jumps to the first one available.
- Right mouse button (does not work for gray fields and the ones with a yellow rectangle): A popup menu appears with 3 options: 1) “Choose just one peak for doubles”; this option is for the automated removal of double peak identifications (yellow rectangle). It automatically selects in the other samples the one peak as correct that is closer to the retention time of the peak in the sample where the right mouse button has been clicked. After this option had been selected, the application asks for which adducts/modifications this operation shall performed.



2) “Quant. anal. at not found”; this option tries to automatically quantify analytes which are grey at the retention time of the selected analyte, whereby it does not take the theoretical isotopic distribution into account. This procedure tries out all of the available quantitation methods, starting with the 3D method, then the “greedy” method followed by the MASPECTRAS and ASAPRatio method (see chapter 5.3). For this option, the application prompts again for which adducts/modifications this operation shall be performed like in “Choose just one peak for doubles”. 3) “Take exact peak for others”; is similar to “Quant. anal. at not found”, but here exactly the same dimensions of the peak in this sample (m/z and time position) are used to define peaks in the other samples. Here, no quantitation methods are used – this is a kind of emergency method, if none of the other methods are able to quantify a peak, and it has to be used with extreme caution, since it can detect the true peak borders only by pure chance.

Mouse clicks on the “sample/group” name:

- Left mouse button: a bar chart for the sample/group is displayed, containing all the available analytes.
- Right mouse button (just for sample): Renaming of sample name.

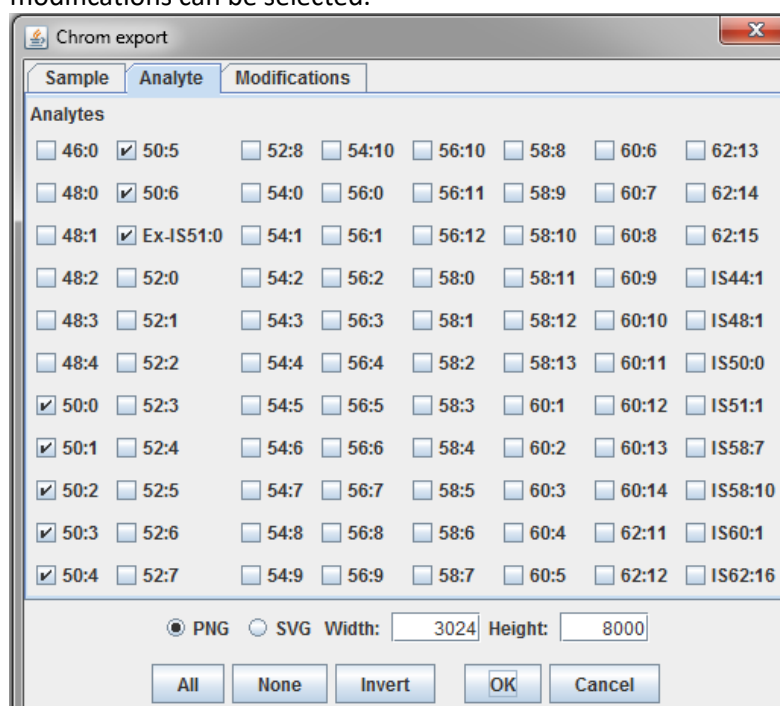
Mouse clicks on the molecule name:

- Left mouse button: a bar chart for the molecule is displayed, containing all available samples/groups.
- Right mouse button (just valid for sample): A popup menu appears containing three options: 1) “Remove analyte in all probes”: deletes one lipid in all result files. 2) “Select analyte”: If there are a lot of lipids and a lot of result files, the removal of a lipid can be quite time consuming. However, if there are several analytes deleted, it consumes the same time. With this option

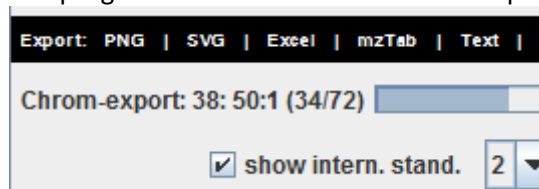
analytes can be selected for the removal; if selected, the name of the analyte is surrounded by a blue rectangle. The removal takes effect on all of the selected analytes as soon as the “Remove analyte in all probes” option is pressed. Here again, a confirmation box appears asking for the adducts/modification on which the removal shall be performed.

#### Export bar:

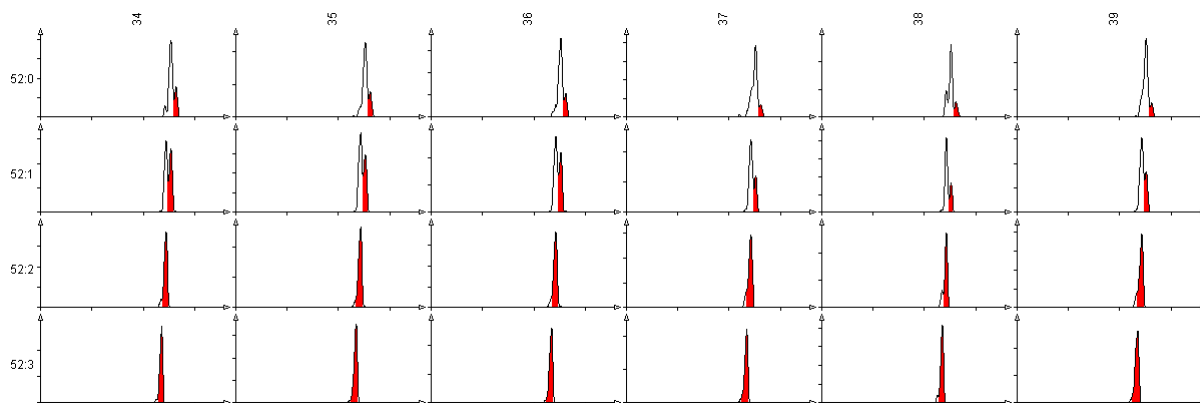
A picture of the heat map or the data respectively can be exported in the following formats: PNG = portable network graphics; SVG = scalable vector graphics; Excel = Microsoft Excel; Text = text based format (tab delimited). Excel and Text export the values in the type that are selected by the user (see next paragraph – Control elements). Furthermore, there is the option to export results in mzTab format, in order to submit them to a public repository. For mzTab, the raw peak areas are exported, these can be affected by standardization and isotope settings only. The mzTab button exports all lipid subclasses at once and generates a file containing all experiments; the data can be exported to chromatograms with the “Chroms” button. Here, the exported analytes, experiments and modifications can be selected.



It is recommended to export just one or two lipid series at once. The chroms export can take some time, depending on the data and the amount of selected hits. The progress of the export is presented in a progress bar underneath the heat map.



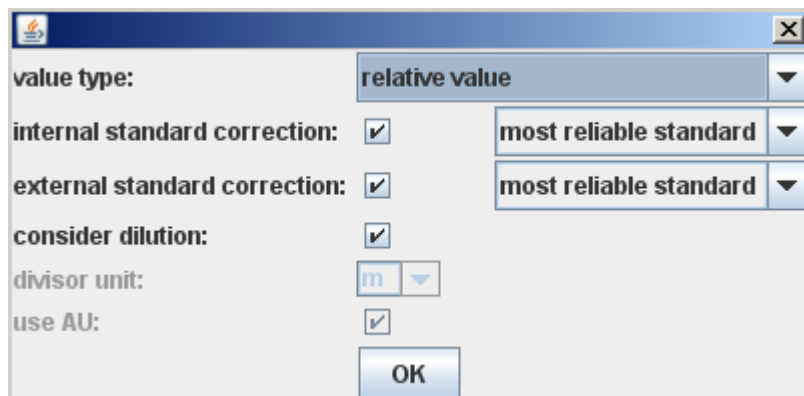
At the end of the chroms export, the user is informed that the export is finished.



This picture shows a chroms export of TG52 of the LDA. It can be easily seen that analytes with more double bonds elute slightly earlier. This is a reliable quality criterion that the algorithm selected the correct hit.

#### Control elements:

- show intern. stand.: Is the internal standards be displayed in the heat map.
- show extern. stand.: Is the external standards be displayed in the heat map.
- isotopes: Defines the highest possible isotope number that should be used for the heat map for all of the analytes. However, the highest isotope number is determined for each analyte separately (e.g. if there are three samples a, b and c, and for a and b the isotopes +0,+1, and+2 are found, whereas in sample c just +0 and +1 are found; just +0 and +1 are taken for the heat map to make the samples comparable).
- double peaks: Shall duplicate peak identifications be flagged with the yellow rectangle.
- Settings button: This is the major settings panel for the display.

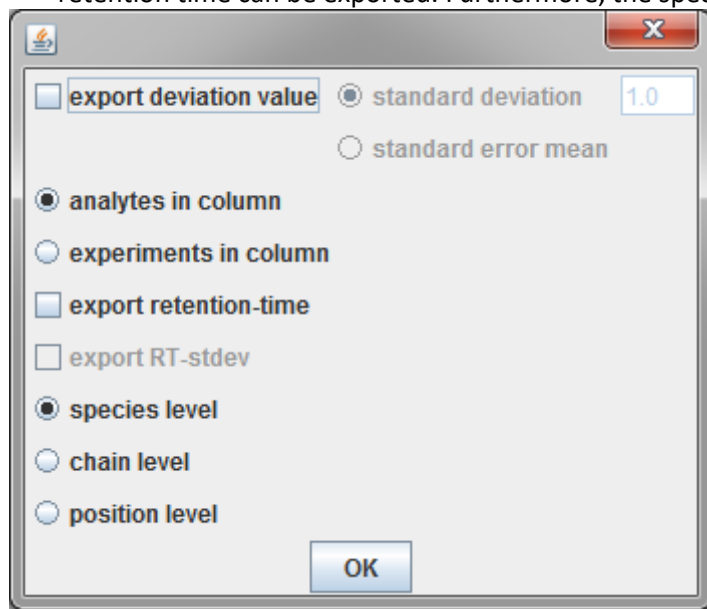


- value type: Which value shall be displayed and used for the heat map or the bar chart.
  - relative value: Just a value in arbitrary units. This value may be the calculated quantity for the analyte itself, or it may be standardized on standards, depending on the other settings.
  - relative to base peak: The values are calculated relative to the highest found peak of this lipid subclass.
  - relative to measured subclass amount: Here values in percent are determined relative to the sum of all analytes for one sample (standards are not considered in this sum) of one lipid subclass.
  - relative to highest total peak: The values are calculated relative to the highest found peak of all available lipid subclasses.
  - relative to total amount: The values are calculated relative to the highest found peak of all available lipid subclasses.



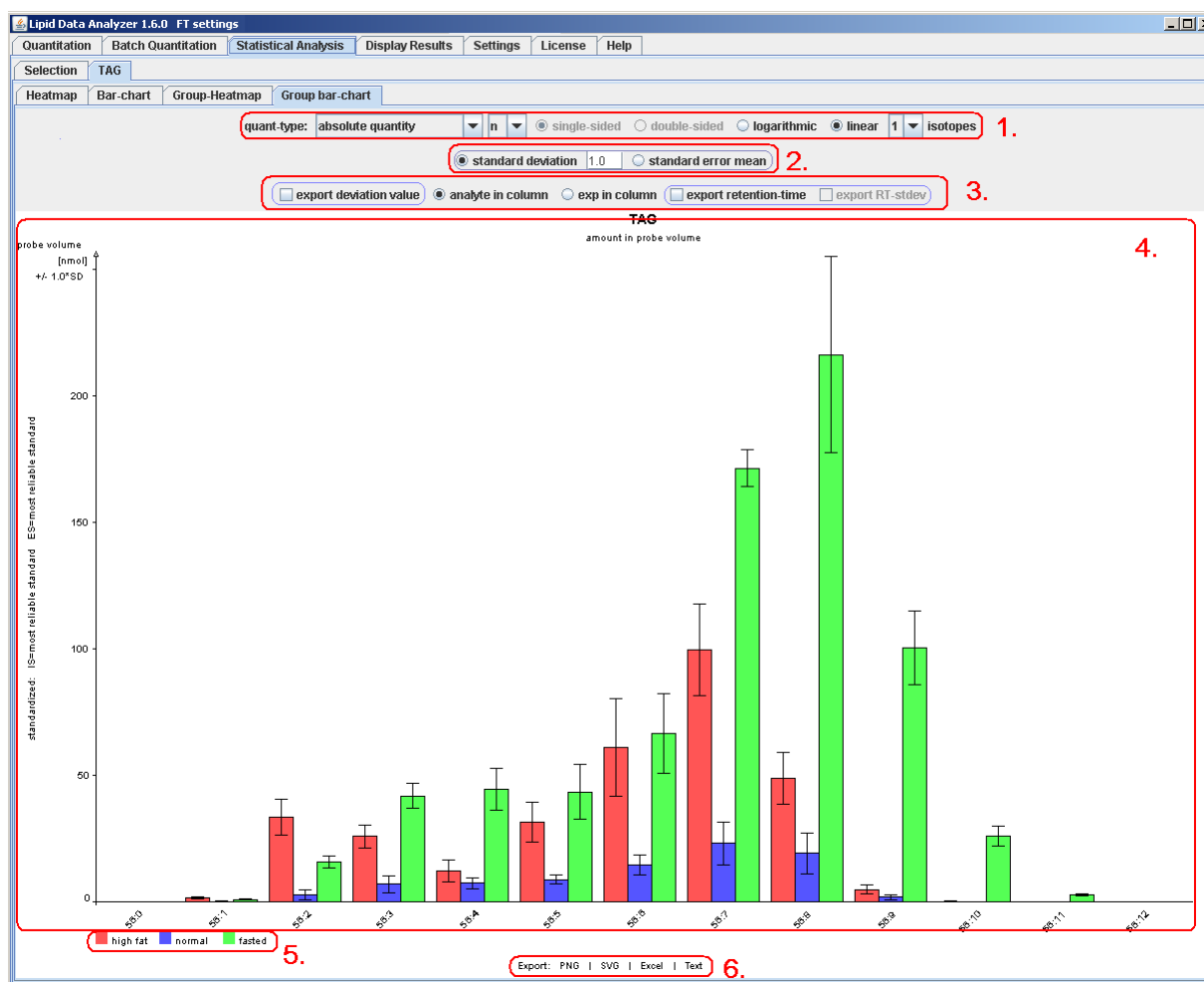
- amount end-volume: Returns the amount [mol] in the end volume, before the MS measurement. For this value, standardization on an internal standard and absolute quantities (see chapter 7.1 number 5) are required.
- conc. end volume: Returns the concentration [mol/L] in the end volume, before the MS measurement. For this value, standardization on an internal standard and absolute quantities (see chapter 7.1 number 5) are required.
- weight end-volume: returns the weight [gram] in the end volume, before the MS measurement. For this value, standardization on an internal standard and absolute quantities (see chapter 7.1 number 5) are required.
- amount sample-volume: Returns the amount [mol] in the sample volume, before the sample preparation steps (dilution). For this value, standardization on an internal and/or external standard and absolute quantities (see chapter 7.1 number 5) are required.
- conc. sample-volume: Returns the concentration [mol/L] in the sample volume, before the sample preparation steps (dilution). For this value, standardization on an internal and/or external standard and absolute quantities (see chapter 7.1 number 5) are required.
- weight sample-volume: Returns the amount [gram] in the sample volume, before the sample preparation steps (dilution). For this value standardization on an internal and/or external standard and absolute quantities (see chapter 7.1 number 5) are required.
- relation to measured neutral lipid: This is for the standardization on the total measured lipid mass[g] of the current lipid subclass measured by MS. For this value, standardization on an internal and/or external standard and absolute quantities (see chapter 7.1 number 5) are required.
- relative to sample weight: This is for the standardization on the total weight of the sample mass[g] of the current lipid subclass measured by MS. For this value, standardization on an internal and/or external standard and absolute quantities (see chapter 7.1 number 5) are required.
- relation to protein content: This is for the standardization on a differently measured protein mass (e.g. measured by a kit). The values for the standardization have to be entered at the absolute quantities (see chapter 7.1 number 5). If there is a standard available, this value is presented in mol/g, else in AU/g.
- relation to neutral lipid content: This is for the standardization on a differently measured lipid mass (e.g. measured by a kit). The values for the standardization have to be entered at absolute quantities (see chapter 7.1 number 5). If there is a standard available, this value is presented in mol/g, else in AU/g.
- internal standard correction: For standardization based on an internal standard desired. If yes, it is possible to choose between 3 standardization methods.
  - most reliable standard: Internally developed method, that detects trustworthy standards and out of these a ratio to the other groups is calculated.
  - median: The median of the standards of one sample is taken as reference value for standardization
  - single standards (several ones are available e.g. IS50:0): A standardization on every entered standard is possible.
- consider dilution: Consider the dilution in the calculation.
- divisor unit: If the value is standardized by some value, this value defines the magnitude of the divisor (e.g. pmol/mL; the m for milli is defined by this setting). This setting is applicable for “conc. end-volume”, “conc. sample-volume”, “relation to measured neutral lipid”, “relative to sample weight”, “relation to protein content”, and “relation to neutral lipid content”.
- use AU: use arbitrary units instead of SI units.
- Select molecules: To fade out specific molecules of the heat map.

- Combined chart: Shows selected molecules in one chart over all samples/groups (see chapter 7.3).
- Export options: Defines the information that is exported to Excel or text format. This dialog box consists of a radio button with the options “analytes in column” and “experiments in column”, and additionally a checkbox for the export of the retention time. Additionally for groups, the standard deviation or the standard error of the values and the standard deviation of the retention time can be exported. Furthermore, the species level of the export can be chosen:



### 7.3 Bar charts

The bar charts are accessible via the heat map or the “Combined chart” button (see chapter 7.2).



The parts encircled by thin violet line are just visible for the grouped view. In general the bar chart painter can be separated into 6 parts:

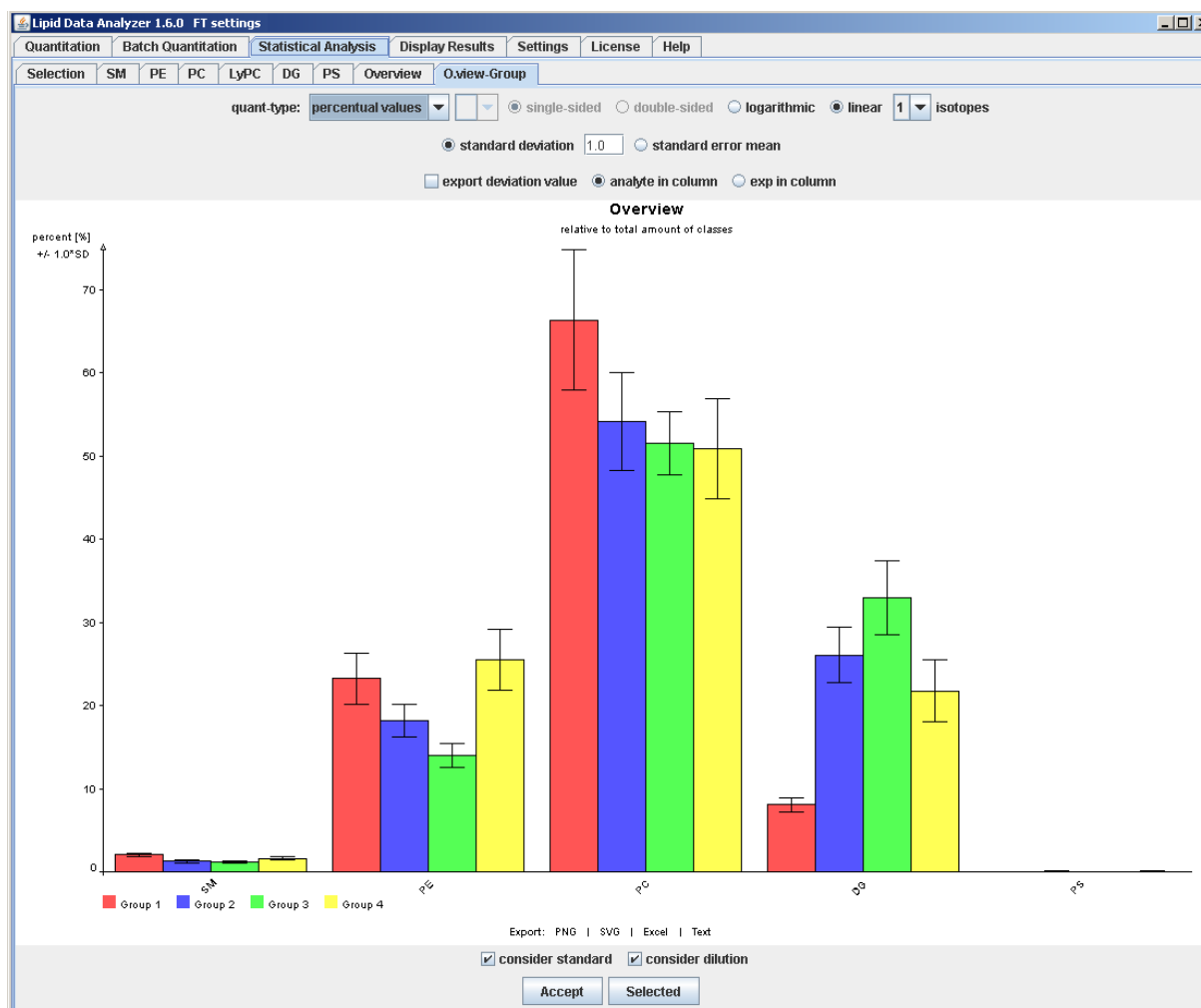
1. Display settings: This is for changing the bar chart display.

- quant-type: Determines the type of values that have to be displayed.
  - a. area absolute: The absolute area values as they are returned from the algorithm, or if there is a standardization on an internal or external standard, these values are the standardized ones.
  - b. percentual values: The selected molecules/group/sample correspond to 100% and the values are displayed relative to this amount.
  - c. area relative to molecule: The values are displayed relative to the median.
  - d. area relative to standard: The values are displayed in relation to the calculated standard (just available if a standardization is selected).
  - e. absolute quantity: The values are displayed in the quantity that is selected in 7.2 "Settings button". The bar chart will always appear in this setting by default, except if "relative value" is selected in the settings.
- a small select box after the quant-type: This box is just visible for the "absolute quantity" settings. The input specifies the magnifier (available are none, m,  $\mu$ , n, p, f, a, %, and %).
- radio button single-sided double-sided: just enabled for "area relative to molecule" and "area relative to standard". For single sided, the values always start with 0. For double sided, the area is relative to the median/standard, and lower values are painted in negative direction and higher values in positive direction.
- radio button logarithmic linear: Available for "area absolute", "percentual values", and "absolute quantity". This option switches between linear and logarithmic (log10) display.

- radio button log10 log2: Available for “area relative to molecule” and “area relative to standard”. Toggles between decadic- and dual logarithm.
  - isotopes: Determines the amount of isotopes that should be used for the quantitative value.
2. Deviation settings: Available just for groups. This value defines how the error bars should be displayed. The deviation can be displayed at any multiplicative value (1.0 is default) and the standard error mean can be displayed.
  3. Export settings: Settings for the export in Excel and Text format. These are the same as in chapter 7.2 “Export options”.
  4. Bar char: The main painting area. At the top in the center, the lipid subclass and the value type is displayed. On the left, written vertically, the used standardization procedure is displayed. At the top of the vertical axis, the current unit is displayed. On the horizontal axis, the molecule, samples or groups are displayed respectively.
  5. Color selection: In this legend, the used color and the corresponding item are displayed. If the mouse is clicked on this legend, the color can be changed.
  6. Export bar: The picture of the heat map or the data respectively can be exported in the following formats: PNG = portable network graphics; SVG = scalable vector graphics; Excel = Microsoft Excel; Text = text based format (tab delimited).

## 7.4 Overview bar chart

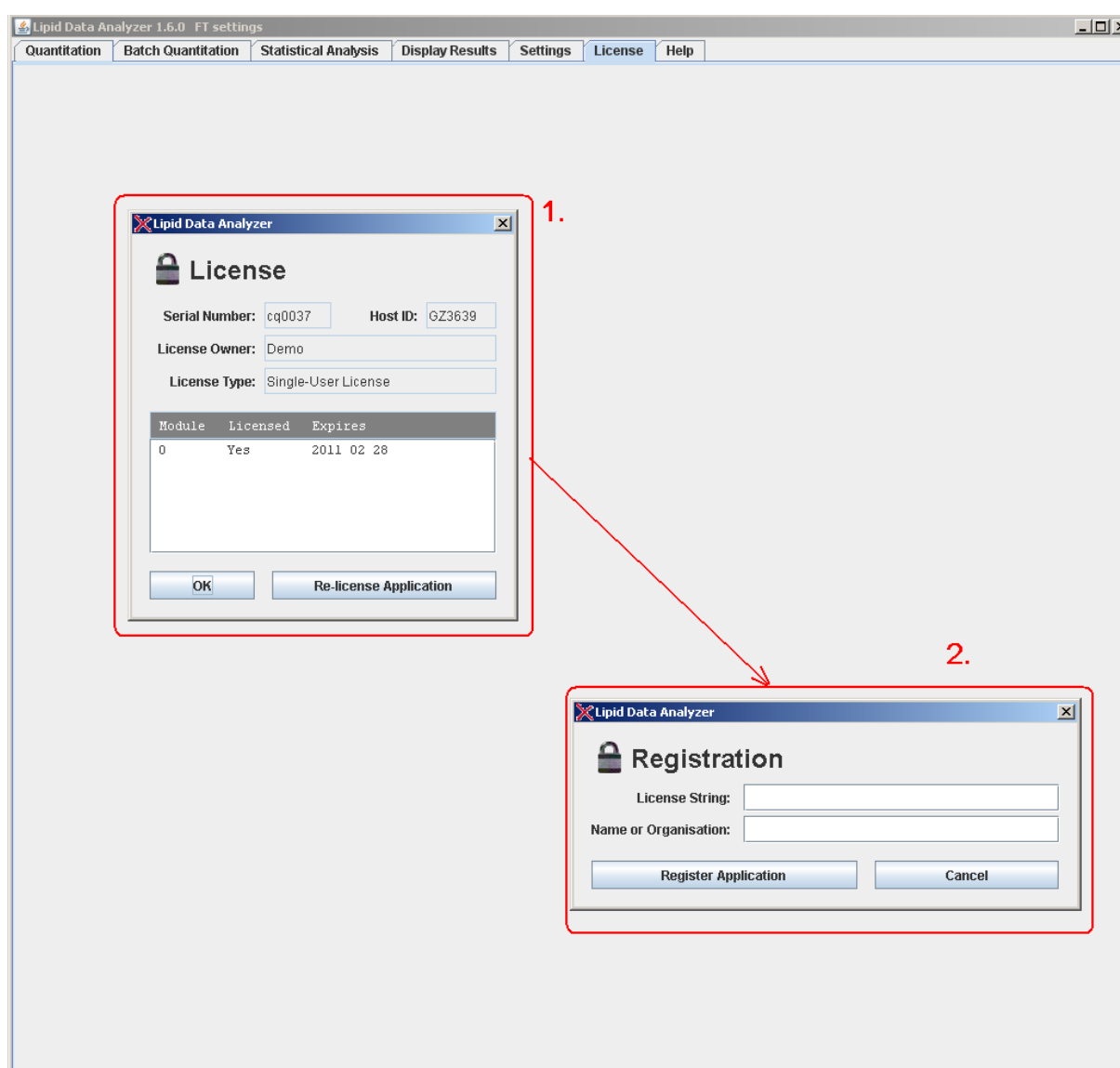
If there are several lipid subclasses available, a bar chart for the overview of the contribution of the separate lipid subclasses is available.



This overview has the option to “consider standard” (the current standard settings are taken) and to “consider dilution”. If the “consider standard” is selected, an overview is displayed containing the standards used. If for one lipid subclass no standard is available, the subclass is neglected (like here LyPC). Without the “consider standard”, these subclasses would have been displayed. Attention: without the “consider standard” it is assumed that all subclasses ionize equally! As quant-type is available only “area absolute” and “percentual values”.

## 8 Licensing

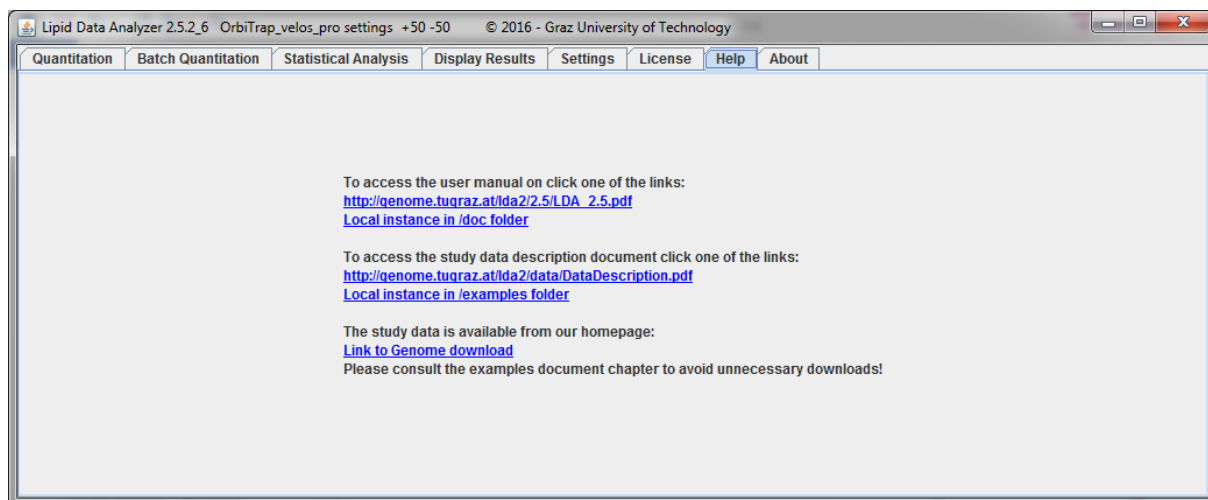
In order to view the license, the tab “License” has to be clicked. Then, a dialog window shows the details about the current license (1. in next figure). If you acquired a new license, click on “Re-license” and a dialog window appears where the license hash plus the name of the institution has to be entered.



## 9 Help

The help page contains links to several resources. The first one is to an online help component (LDA help, explained in detail after the next figure). Furthermore, it provides links to the user manual (this

file), the Examples.pdf (file that shows the functionality of the software with the aid of the published example data), and the example data that can be downloaded from the homepage. The links to PDFs are provided to the home page and to local files that come with the installation package.



## Appendix A - Preparation of the mass list Excel file

The Excel file containing the mass lists must follow some conventions.

The file consists of several sheets, each representing a lipid subclass. If you want to analyze several lipid subclasses, just create several sheets. Second, there must be a header row. The header row must contain the following keywords, otherwise entered analytes won't be accepted: Name, tR (min), at least one mass(...), and at least one element column. Except for the "Name" column, which must be the first column, the order of the columns does not matter for the Excel file. Detailed description of the example in the figure (columns from left to right):

- "Name": The name of the analyte.
- "dbs": This column is optional. If the analyte name contains double bonds, they should be entered in this column. The values of this column must be in integer format. If the name of the analyte does not contain any double bonds, just remove this column. Generally, it is possible to enter the double bonds in the "Name" column directly, however, the sorting of the analytes in the results is improved if this column is used.
- Element columns ("C", "H", "O", "P", "N", "D"): In these columns, the elemental composition (chemical formula) of the analyte has to be entered. The values of these columns must be in integer format. If there is a value in float format (e.g. the M column), the column will be ignored as elemental column (a warning at the beginning of the quantitation appears in the logging – see "Appendix E - Trouble shooting"). A column is an elemental column if it contains just one or two characters, beginning in upper case and ending lower case (e.g. H, Na, He, P, etc.). If new elements are required, just add an additional column and enter the probable occurrences in nature in the elementconfig.xml file (see "Appendix C - Configuration settings" for the location of this file).
- Additional calculation columns (e.g. "M", "MH+"): These columns are not read by the application, and are just a help for the user to calculate the masses. Do not mind that the comma in the float value is a "," and not a ".", the Excel I used for the screen shot is using a German notation.
- Mass columns ("mass(...)"): A mass column contains the mass values to be quantified. There can be several mass columns because of different modifications/adducts. A mass column must contain the keyword "mass" followed by round brackets. Within the round brackets, there must be the keywords "name" and "form" followed by square brackets [...]. In the square brackets after "name", the display name has to be entered; after "form", the chemical formula. The chemical formula of the modification/adduct must be in the notation \$element\_symbol\$element\_amount\$(e.g. H4). If no element amount is entered, it is assumed to be one (e.g. NH4=N1H4). If elements are lost by the modification, enter a "-" in front (e.g.

–NH4 would correspond to a loss of ammonium). **A space is interpreted as +**, that means if you enter “-N H4” this would correspond to –N1+H4. Additionally, the charge state can be entered by charge=\$charge\_state\$. If this information is missing in the brackets, a charge state of 1 is assumed.

- tR (min): This is the retention time column. The input of the retention time is optional. If it has not been entered, the whole chromatogram is searched for a peak (if the checkbox “Find molecules where retention time is unknown” is checked).

Third, it is possible to enter a general start and stop retention time for a lipid subclass. The notation is “Start-RT: ...” and “Stop-RT: ...”. These cells can be in any row before the header row.

Fourth, for classes where different hydroxylation stages are to be expected, such as sphingolipids, the ‘OH-Number: ...’ parameter designates the hydroxylation stage of the provided mass list, e.g. ‘OH-Number: 2’, and the ‘OH-Range: ...’ parameter designates the hydroxylation range the LDA should search for, e.g., OH-Range: 1-3. In the previous example, a dihydroxylated mass list is provided, and the LDA will search for mono-, di-, and trihydroxylated species. The masses for mono-, and trihydroxylated species will be calculated automatically by LDA by using the information provided in the dihydroxylated mass list.

Fifth, it is possible to set some filtering options. Adding of “adductInsensitiveRtFilter” forces the retention time filter to be calculated based on all modifications. Adding of “pickBestMatchBySpectrumCoverage” will lead to picking best matches from duplicates (same lipid class and scan number) by spectrum coverage. In case of same coverage the hit with the lowest number of ox-modifications is picked. In case of same coverage and same number of ox-modifications the hit with the minimum difference between theoretical and evaluated m/z value is picked. These cells can be in any row before the header row.

Pay attention to the naming of the lipid subclasses (sheet name) and naming of the adduct, since the combination of both is used for the selection of the corresponding decision rule set (see chapter 6 “MS<sup>n</sup> spectra interpretation – decision rule sets”). E.g. in this example, the Excel tab is named “TG” (marked by 1) and the first mass column header is “mass(form[+NH4] name[NH4])” defining the adduct name as NH4. For MS<sup>n</sup> identification, the software will look for a file named TG\_NH4.frag.txt in the fragRules directory. If this file is not there, the algorithm will do the identification based on MS<sup>1</sup> data only.

## Appendix B – Modified lipids

In addition to the hydroxylation stages (Appendix A) the current version of the LDA supports analysis of modified lipids (their fatty acyl chains) for custom, user defined modifications.

### ***Mass list***

To this end, a new column called “PSM” was introduced, where cell values enumerate the total amount of possible modifications. Disparate modifications are separated by a semicolon, and if multiple modifications are possible at the same time, they should be separated by a colon. The modifications in the example allow for up to four oxo modifications, a chlorine modification, or an oxo modification in combination with a chlorine modification. *Note the semicolon at the start, which tells the LDA to also allow unmodified chains still.* Since the LDA will automatically infer new masses for modified lipids, the masses in the list should still reflect the unmodified lipid.



Name	db	C	H	O	P	N	mass(form[+HCOO] name[HCOO])	mass(form[-CH3] name[-CH3])	PSM	tr (min)	adductInsensitiveRtFilter
6	0	14	30	7	1	1	400,17419	340,15306	;O1;O2;O3;O4;Cl1;Cl1,O1		
7	0	15	32	7	1	1	414,18984	354,16871	;O1;O2;O3;O4;Cl1;Cl1,O1		
8	0	16	34	7	1	1	428,20549	368,18436	;O1;O2;O3;O4;Cl1;Cl1,O1		
9	0	17	36	7	1	1	442,22114	382,20001	;O1;O2;O3;O4;Cl1;Cl1,O1		
10	0	18	38	7	1	1	456,23679	396,21566	;O1;O2;O3;O4;Cl1;Cl1,O1		
10	1	18	36	7	1	1	454,22114	394,20001	;O1;O2;O3;O4;Cl1;Cl1,O1		
11	0	19	40	7	1	1	470,25244	410,23131	;O1;O2;O3;O4;Cl1;Cl1,O1		
11	1	19	38	7	1	1	468,23679	408,21566	;O1;O2;O3;O4;Cl1;Cl1,O1		
12	0	20	42	7	1	1	484,26809	424,24696	;O1;O2;O3;O4;Cl1;Cl1,O1		
12	1	20	40	7	1	1	482,25244	422,23131	;O1;O2;O3;O4;Cl1;Cl1,O1		
12	2	20	38	7	1	1	480,23679	420,21566	;O1;O2;O3;O4;Cl1;Cl1,O1		
12	3	20	36	7	1	1	478,22114	418,20001	;O1;O2;O3;O4;Cl1;Cl1,O1		
12	4	20	34	7	1	1	476,20549	416,18436	;O1;O2;O3;O4;Cl1;Cl1,O1		
13	0	21	44	7	1	1	498,28374	438,26261	;O1;O2;O3;O4;Cl1;Cl1,O1		
13	1	21	42	7	1	1	496,26809	436,24696	;O1;O2;O3;O4;Cl1;Cl1,O1		
13	2	21	40	7	1	1	494,25244	434,23131	;O1;O2;O3;O4;Cl1;Cl1,O1		
13	3	21	38	7	1	1	492,23679	432,21566	;O1;O2;O3;O4;Cl1;Cl1,O1		
13	4	21	36	7	1	1	490,22114	430,20001	;O1;O2;O3;O4;Cl1;Cl1,O1		
14	0	22	46	7	1	1	512,29939	452,27826	;O1;O2;O3;O4;Cl1;Cl1,O1		
14	1	22	44	7	1	1	510,28374	450,26261	;O1;O2;O3;O4;Cl1;Cl1,O1		
14	2	22	42	7	1	1	508,26809	448,24696	;O1;O2;O3;O4;Cl1;Cl1,O1		
14	3	22	40	7	1	1	506,25244	446,23131	;O1;O2;O3;O4;Cl1;Cl1,O1		
14	4	22	38	7	1	1	504,23679	444,21566	;O1;O2;O3;O4;Cl1;Cl1,O1		
15	0	23	48	7	1	1	526,31504	466,29391	;O1;O2;O3;O4;Cl1;Cl1,O1		
15	1	23	46	7	1	1	524,29939	464,27826	;O1;O2;O3;O4;Cl1;Cl1,O1		
15	2	23	44	7	1	1	522,28374	462,26261	;O1;O2;O3;O4;Cl1;Cl1,O1		
15	3	23	42	7	1	1	520,26809	460,24696	;O1;O2;O3;O4;Cl1;Cl1,O1		
15	4	23	40	7	1	1	518,25244	458,23131	;O1;O2;O3;O4;Cl1;Cl1,O1		
16	0	24	50	7	1	1	540,33069	480,30956	;O1;O2;O3;O4;Cl1;Cl1,O1		

## Fatty acid mass list (fattyAcidChains.xlsx)

The “PSM” column was also integrated into the fatty acid mass list to be able to set and control the modifications per fatty acid chain. In the example, each fatty acid chain can carry one or two oxo modifications, a chlorine modification, an oxo modification in combination with a chlorine modification or no modifications at all.

Name	db	C	H	O	mass	PSM
2	0	2	4	2	60,0211206	;O1;O2;Cl1;Cl1,O1
3	0	3	6	2	74,0367706	;O1;O2;Cl1;Cl1,O1
4	0	4	8	2	88,0524206	;O1;O2;Cl1;Cl1,O1
5	0	5	10	2	102,068071	;O1;O2;Cl1;Cl1,O1
6	0	6	12	2	116,083721	;O1;O2;Cl1;Cl1,O1
7	0	7	14	2	130,099371	;O1;O2;Cl1;Cl1,O1
8	0	8	16	2	144,115021	;O1;O2;Cl1;Cl1,O1
9	0	9	18	2	158,130671	;O1;O2;Cl1;Cl1,O1
10	0	10	20	2	172,146321	;O1;O2;Cl1;Cl1,O1
11	0	11	22	2	186,161971	;O1;O2;Cl1;Cl1,O1
12	0	12	24	2	200,177621	;O1;O2;Cl1;Cl1,O1
13	0	13	26	2	214,193271	;O1;O2;Cl1;Cl1,O1
14	0	14	28	2	228,208921	;O1;O2;Cl1;Cl1,O1
15	0	15	30	2	242,224571	;O1;O2;Cl1;Cl1,O1
16	0	16	32	2	256,240221	;O1;O2;Cl1;Cl1,O1
17	0	17	34	2	270,255871	;O1;O2;Cl1;Cl1,O1
18	0	18	36	2	284,271521	;O1;O2;Cl1;Cl1,O1
19	0	19	38	2	298,287171	;O1;O2;Cl1;Cl1,O1
20	0	20	40	2	312,302821	;O1;O2;Cl1;Cl1,O1
21	0	21	42	2	326,318471	;O1;O2;Cl1;Cl1,O1
22	0	22	44	2	340,334121	;O1;O2;Cl1;Cl1,O1
23	0	23	46	2	354,349771	;O1;O2;Cl1;Cl1,O1
24	0	24	48	2	368,365421	;O1;O2;Cl1;Cl1,O1
25	0	25	50	2	382,381071	;O1;O2;Cl1;Cl1,O1
26	0	26	52	2	396,396721	;O1;O2;Cl1;Cl1,O1
27	0	27	54	2	410,412371	;O1;O2;Cl1;Cl1,O1
28	0	28	56	2	424,428021	;O1;O2;Cl1;Cl1,O1
29	0	29	58	2	438,443672	;O1;O2;Cl1;Cl1,O1
30	0	30	60	2	452,459322	;O1;O2;Cl1;Cl1,O1

## Modification table (modconfig.xml)

The modification table is the last piece of the puzzle, in which the modification symbol (as used in the mass lists) is connected to the actual configuration of the modification. In the example, the symbol

'Cl' is used for chlorinated lipids, where the modification constitutes an addition of a chlorine atom and a subtraction of a hydrogen atom, whereas the symbol 'O' stands for an oxo modification which constitutes the addition of an oxygen atom without subtraction of a hydrogen atom.

```
modconfig.xml
1  <?xml version="1.0" encoding="UTF-8"?>
2
3  <modifications>
4      <modification symbol="O">
5          <element name="O">+1</element>
6      </modification>
7      <modification symbol="Cl">
8          <element name="Cl">+1</element>
9          <element name="H">-1</element>
10     </modification>
11 </modifications>
12
```

## Appendix C - Configuration settings

This chapter covers the manual adaption of LDA setting parameters. General settings can be found in the .settings file, and machine specific ones in the LipidDataAnalyzer.properties file. For each platform such a properties file is present in the properties folder in the installation directory. It is recommended to change the settings for the platform in the corresponding file in the properties folder and, after saving this file, to select the platform and click on "Save as default" (see chapter 2). If the settings are changed directly in the LipidDataAnalyzer.properties, these settings are overwritten the next time somebody clicks on "Save as default". The .settings file can be edited directly.

In the LipidDataAnalyzer.properties file, not only algorithmic parameters are stored, but default settings for the display as well. The changes take effect after the properties file had been stored and the LDA is restarted.

### ***Performance settings (.settings)***

For accelerating the quantification, the time-consuming Savitzkay-Golay smoothing filter can be performed on a CUDA-compatible GPU. To activate this feature set the following parameter to true:

- useCuda: use "true" when the LDA should make some calculations on the CUDA-compatible GPU, "false" otherwise.

### ***Path settings (.settings)***

This part contains path settings to executables or required files.

- MsconvertPath: Path to msconvert executable (msconvert.exe). This parameter is required if Thermo RAW files, ABSciex wiff files, or Agilent .d directories are used directly.
- MassPPPath: Path to Mass++ executable (msppcmd.exe). This parameter is required if Waters raw directories are used directly.
- ElementConfig: Path to the required elementconfig.xml file that contains the used chemical elements and the occurrence of their isotopes in nature.
- ReadWPath (deprecated – use msconvert instead): Path to ReadW executable. This parameter is required if Thermo Finnegan RAW files are used directly.
- MassWolfPath (deprecated – use msconvert instead): Path to the MassWolf executable if the Waters .raw directories shall be used directly.

## ***Look & feel settings (.settings)***

- LookAndFeel: This value is by default “system”, which corresponds to your system specific look & feel (e.g. Windows). If you prefer Java look & feel, change this value to “java”, then you have the same look & feel as in this manual.

## ***Default quantitation settings (.properties)***

This stores default settings for the quantitation input page.

- basePeakCutoff: The default value for the “Rel. base-peak cutoff” (see chapter 2).

## ***Excel result file settings (.settings)***

- OverviewExcelWorkbook: If “true”, an overview tab is created in the Excel file.
- OverviewExcelMass: If “true”, the mass values will be exported in the Excel file.
- OverviewIsotopicValues: If “true”, the quantities of the single isotopes are exported in the Excel file.

## ***3D viewer default settings (.properties)***

For the resolution of the 3D viewer, default settings can be entered.

- threeDViewerDefaultTimeResolution: The default time resolution of the viewer.
- threeDViewerDefaultMZResolution: The default m/z resolution of the viewer.
- MSn3DDisplayTimeWindow (this parameter is in the .settings file): This parameter is for the MS<sup>n</sup> spectrum viewer. It defines the display time [min] to show in 3D view before and after a hit.
- threeDViewerMs2DefaultTimeResolution: The default time resolution of the 3D viewer for MS<sup>n</sup> spectra.
- threeDViewerMs2DefaultMZResolution: The default m/z resolution of the 3D viewer for MS<sup>n</sup> spectra.
- ms2ChromMultiplicationFactorForInt: Defines the highest possible resolution of the 3D viewer for MS<sup>n</sup> spectra. The default value of 10 corresponds to a maximum resolution of 0.1mTh while 100 corresponds to 0.01mTh.

## ***Standards prefix settings (.settings)***

The field for the prefixes of the standards can be set by default (see chapter 7.1).

- ISDefaultInput: Default prefix for the internal standard.
- ESDefaultInput: Default prefix for the external standard.

## ***mzXML → chrom settings(.properties)***

These are settings for the translation of the mzXML to the chrom format.

- maxFileSizeForChromTranslationAtOnce: An approximate value how many MB can be read at once into the memory. The mzXML is translated into the chrom in several iterations. If there is e.g. an mzXML file with 200MB and an m/z range from 600-1400, this file will be translated in two iterations in the case this parameter is 100: first round: masses from 600-1000; second round masses from 1000-1400. Thus this parameter does not directly correspond to the required memory, the required memory is normally more. If this parameter would have been 200, the whole file would have been translated in one round. The reason for this parameter is to permit the translation of huge files on machines with low RAM. If this parameter is increased, the translation time is decreased. If this parameter is too high, a java.lang.OutOfMemoryException will be thrown and the file cannot be translated.
- chromMultiplicationFactorForInt and chromMultiplicationFactorForInt: This parameter defines the highest resolution of the chrom file (e.g. chromMultiplicationFactorForInt=1000 and chromMultiplicationFactorForInt=1 would mean a max resolution of 0.001Th;

chromMultiplicationFactorForInt=1000 and chromMultiplicationFactorForInt=5 would mean a max resolution of 0.005Th; chromMultiplicationFactorForInt=100 and chromMultiplicationFactorForInt=1 would mean a max resolution of 0.01Th).

- MS2: if true, the MS<sup>n</sup> spectra will be translated in the chrom format.

### ***Waters raw → chrom settings(.settings)***

- MergeMultipleMSMSFiles: “true” for DDA mode and “false” for MS<sup>E</sup> mode. For Waters QTOF, several traces of MS/MS are generated, and each one leads to a separate mzXML. If true, the mzXML files are merged to one file which is appropriate for DDA. This setting should be false for MS<sup>E</sup>.
- SkipLastMzXML (applicable only if "MergeMultipleMSMSFiles" is true): The last trace in Waters data is usually a quality trace for detection stability; this should not be merged into the data.

### ***General algorithm settings(.properties)***

This section contains the most general settings of the algorithm. A detailed description of the parameters of the 3D algorithm follows in a later paragraph of this Appendix. The parameters are described in the LipidDataAnalyzer.properties file too.

- machineName: The name of the MS platform. This name must not be changed unless a new MS platform is added. The name must be consistent with the names in the persistently stored properties; e.g. machineName=OrbiTrap\_velos\_pro, the name of the file in the properties directory must be LipidDataAnalyzer\_OrbiTrap\_velos\_pro.properties.
- coarseChromMzTolerance: The +/-m/z range that is used for the extraction of the first chromatogram of the algorithm which is displayed in the 2D viewer (see chapter 5.3).
- chromSmoothRange: The smooth range in seconds for Savitzky-Golay filter of the first chromatogram.
- chromSmoothRepeats: The smooth repeats for Savitzky-Golay filtering of the first chromatogram.
- use3D: If yes, the 3D algorithm is used by default for the quantitation, else the 2D algorithm.
- MS2: If true, the MS<sup>n</sup> spectra will be translated in the chrom format. In this case, the MS<sup>n</sup> functionality of the application can be used.
- neutronMass (optional, default is 1.005): The mass of a neutron for the calculation of the m/z values of the isotopic peaks. This value may be important if very high mass resolution is used.

### ***General MS<sup>n</sup> algorithm settings (.properties)***

These settings are for covering MS platform specific settings such as mass deviation thresholds.

- ms2PrecursorTolerance: Threshold for the precursor m/z deviation in Th.
- ms2MzTolerance: Threshold for m/z deviation in the MS<sup>n</sup> spectra in Th.
- ms2IsobarSCEExclusionRatio: If there are fragments of several isobaric species present, this is a threshold to exclude the smaller lipid subclass/adduct and assign the species totally to the one with the more intense fragments. This default value can be changed for each subclass/adduct in the [GENERAL] section of the fragmentation rules (see IsobarSCEExclusionRatio in chapter 6.2).
- ms2IsobarSCFarExclusionRatio: If there are fragments of several isobaric species present, they are usually chromatographically separated. If there are at least two isobaric peaks containing MS<sup>n</sup> spectra which are chromatographically separated, and one of them can be assigned unambiguously to one subclass/adduct, a higher threshold is used for the removal of the identified subclass/adduct in the other peak, since it must be assumed that two isobars of the same subclass cannot be chromatographically separated by such a high time difference. This preceding is particularly useful for isobaric subclasses such as PC<sub>-</sub>CH3 and

PE\_H which are primarily dominated by chain fragments. This default value can be changed for each subclass/adduct in the [GENERAL] section of the fragmentation rules (see IsobarSCFarExclusionRatio in chapter 6.2).

- ms2IsobaricOtherRtDifference: This parameter defines the retention time difference between two isobaric peaks for the application of the higher isobarSCFarExclusionRatio. This default value can be changed for each subclass/adduct in the [GENERAL] section of the fragmentation rules (see IsobarRtDiff in chapter 6.2).

### ***Memory settings(.bat, .sh, or. vmoptions)***

The memory settings for the executables can be changed in the Lipid Data Analyzer.vmoptions, and for the console in the .bat or .sh, respectively. The -Xmx parameter corresponds to the maximally reservable memory, and -Xms the immediately reserved memory. The -Xms value should be approximately half of the -Xmx value.

### ***Specific settings for the 3D algorithm (.properties)***

- useNoiseCutoff: If "true", the noise level is calculated, and values below this threshold are discarded without attempting a quantitation by the 3D algorithm. Since this value effectively accelerates the quantitation, it is recommended to use "true".
- noiseCutoffDeviationValue (applicable only if "useNoiseCutoff " is true): It is reasonable to use a higher threshold than the noise level. A standard deviation of the presumable noise values is calculated. The final threshold is set to useNoiseCutoff+standardDeviation\*noiseCutoffDeviationValue. A value of 2 is recommended for this setting.
- scanStep: If every scan shall be used for area calculation use 1, if every second one use 2, and so on; this is another option that may be used to accelerate the 3D algorithm. However, the noise cutoff should be used instead, since the probability of missing good identifications is rather high.
- profileMzRangeExtraction: M/z tolerance for the extraction of the profiles. A higher tolerance than the peak width is required to perform proper smoothing.
- profileTimeTolerance: Chromatographic time tolerance around the peak of interest for profile extraction
- profileIntThreshold: It might happen that the chosen peak maximum of the coarse chromatogram (which is used as time for the profile extraction) is exactly between two independent peaks. The result is a profile without or with small intensities. This is an absolute value for declaring values as zero. If all intensities of the profile are smaller than this threshold, a broader profile is extracted.
- broaderProfileTimeTolerance: If all intensities of the profile are smaller than the threshold, declared at profileIntThreshold, a broader profile is extracted. This value is the chromatographic time tolerance for the broader profile.
- profileSmoothRange: Which range is used to smooth the profile.
- profileSmoothRepeats: Number of repeats for smoothing the 2D-profile.
- profileMeanSmoothRepeats: Number of repeats for smoothing the 2D-profile by a mean filter. The typical smoothing filter for the LDA is the Savitzky-Golay le. However, especially for high-resolution data, some data points might be missed by the conventional smoothing range. Hence, a mean filter is applied before the Savitzky-Golay filter to close the parts of the missing spaces. Attention: mean filtering has to be used with caution, since it flattens peaks and shifts the actual true m/z value in the positive direction. Consequently, mean filtered profiles do not return the true m/z value of the peak maximum. This effect is reduced by the parameter profileSmoothingCorrection.
- profileSmoothingCorrection: Smoothing by a mean filter might change the m/z position of the peak summit. This value accounts for this phenomenon.
- profileMzMinRange: Minimum m/z extension of the detected profile peak. Peaks with lower extensions are regarded as noise.

- **profileSteepnessChange1:** The peak borders of the profile peak are detected by a change in the gradient of the peak curve. If the gradient (steepness) of the next data point is **profileSteepnessChange1** times lower than the current one, a peak border is assumed. This threshold (**profileSteepnessChange1**) is used for intensity regions between **profileIntensityCutoff2\*PEAK\_MAXIMUM** to the **PEAK\_MAXIMUM**. E.g. in a region between 20%-100% of the peak maximum.
- **profileSteepnessChange2:** This value has the same purpose as the **profileSteepnessChange1**, but it is used for lower intensity regions. The reason is that intensity changes in lower abundant regions are typically more pronounced, therefore, a higher tolerance is necessary. This threshold (**profileSteepnessChange2**) is used for intensity regions between the **profileIntensityCutoff1\*PEAK\_MAXIMUM** to **profileIntensityCutoff2\*PEAK\_MAXIMUM**. E.g. in a region between 15%-20% of the peak maximum. For regions below **profileIntensityCutoff1**, only a rise in the intensity is regarded as peak border point.
- **profileIntensityCutoff1:** This designates the threshold between the region where no (beneath) or **profileSteepnessChange2** is used to determine the peak border. E.g., if the value is 0.15, below 15% a raise in the peak intensity identifies a peak border, while above 15% already a change in the peak gradient is indicative for a peak border.
- **profileIntensityCutoff2:** Beneath this threshold, **profileSteepnessChange2** is used as threshold in the gradient, above **profileSteepnessChange1**.
- **profileGeneralIntCutoff:** A threshold relative to the peak maximum where a drop in intensity designates a peak border point. E.g. 0.03 corresponds to 3% of the peak maximum.
- **profilePeakAcceptanceRange:** M/z threshold for deviation of the profile peak maximum from the ideal m/z value. If the m/z value of the profile peak maximum is outside this range, it is assumed that the peak comes from another analyte and the peak is discarded.
- **profileMaxRange:** M/z value for the maximally allowed extension of a profile peak. This threshold is for prevention of wrong peak intensities by slowly tapering off peak intensities.
- **smallChromMzRange:** The 3D algorithm extracts two chromatograms based on the values about the profile peak: a small one around the peak maximum, and a broader one with the range of the detected peak borders. This range is an m/z threshold for the extraction of the small chromatogram.
- **smallChromSmoothRepeats:** The smoothing repeats for the small chromatogram (the smoothing range is the same as in the coarse chromatogram).
- **broadChromSmoothRepeats:** The same as **smallChromSmoothRepeats** for the broad chromatogram.
- **smallChromMeanSmoothRepeats:** This parameter indicates the smooth repeats of a mean filter before the conventional Savitzky-Golay filter. This is necessary for very small m/z ranges (missing data points).
- **broadChromMeanSmoothRepeats:** The same as **smallChromMeanSmoothRepeats** for the broad chromatogram.
- **smallChromSmoothRange:** Time range in minutes that is smoothed around the current time point. The reason is that smoothing takes relatively long, so just a small range is smoothed for the small chromatogram.
- **broadChromSmoothRange:** The same as **smallChromSmoothRange** for the broad chromatogram.
- **smallChromIntensityCutoff:** Threshold relative to the peak maximum where a peak border is set (such as **profileGeneralIntCutoff**).
- **broadChromIntensityCutoff:** The same as **smallChromIntensityCutoff** for the broad chromatogram.
- **broadChromSteepnessChangeNoSmall:** Normally the peak-borders are defined by the small chromatogram, however, if there is an overlap expected, the broad chromatogram is taken to define the peak borders with the GreedySteepnessReduction-method which is essentially the same as explained in **profileSteepnessChange1**.

- **broadChromIntensityCutoffNoSmall:** This value is essentially the same as the **profileGeneralIntCutoff** for the broad chromatogram.
- **finalProbeTimeCompTolerance:** Float values can hardly be checked for equality, thus, a tolerance value is introduced to designate equality in the time axis.
- **finalProbeMzCompTolerance:** Float values can hardly be checked for equality, thus, a tolerance value in **Th** is introduced to designate equality in the **m/z** axis.
- **overlapDistanceDeviationFactor:** An overlap is assumed if there is a big difference in the distance between the peak maximum and the border of the peak. If the relation of the **oldDistance/newDistance** is bigger than this value, it is regarded as overlap (old distance is from the coarse chromatogram; new distance is identified by the small and the broad chromatogram). The knowledge about an overlap is used in the verification of the isotopic distribution to apply less stringent criteria.
- **overlapPossibleIntensityThreshold:** If it is not sure that the current peak is an isotopic-peak of a different analyte, it is checked if the time point of the current peak maximum is in a high-intensity region in the **-1Th** chromatogram. If the intensity there is below this settable threshold in relation to the peak maximum (e.g. 0.15 is 15% of the maximum), it must be assumed that the detected peak is not from another isotopic distribution.
- **overlapSureIntensityThreshold:** This is the same relation as explained in **overlapPossibleIntensityThreshold**, but if this value is exceeded (e.g. 0.7 is 70% of the maximum), it must be assumed that this peak is definitely from another isotopic distribution.
- **overlapPeakDistanceDivisor:** If the check by the **overlapPossibleIntensityThreshold** and **overlapSureIntensityThreshold** returned no definite result (e.g. the intensity is between 15% and 70% of the peak maximum), a less reliable check is performed where the algorithm checks if the time point of the current peak is in the inner part of the other peak. For this check, the time distance between the peak center and the border is taken as distance. This parameter designates the ratio for the inner third. In the case of a value of 3, one third is defined as the inner third, and if this time point is within the inner third, it is assumed that this peak is from another isotopic distribution.
- **overlapFullDistanceDivisor:** It is a similar check as before which designates the time point of the current peak to be in the inner part of the overlapping isotope, but here, the time distance is calculated between the peak borders. This is useful if there are small distances from the peak center to the border which occurs in cases of shifted peak centers which are typically observed in case of overlaps. If the full distance is taken, this effect is diminished. Here, two times the **overlapPeakDistanceDivisor** is recommended, since the peak is usually twice as broad as one time the distance from the peak center to the border. If a peak is within the peak borders by any of these two methods, it is assumed that this peak is from another isotopic distribution.
- **peakDiscardingAreaFactor:** In a chromatogram, if peaks are relatively small in relation to the highest found peak they are discarded. A factor of 1000 corresponds to a threshold for peaks that are smaller than 1/1000 of the highest peak, i.e. 0.1%.
- **isotopeInBetweenTime:** Sometimes, small parts of peaks are split by the 3D algorithm, since it is usually very sensitive to variations in peak intensities. If this happens to an adjacent peak from another isotope, the main part of the isotope is removed, but the smaller part remains. In order to remove such peaks, a further check is performed. This parameter designates a maximum chromatographic time distance in seconds between two peaks to trigger a check for an isotope in between.
- **isoInBetweenAreaFactor:** If an isotope is detected in between two peaks, the smaller one is regarded to be a fragment belonging to the other isotope. However, if the area of the smaller one times this factor is higher than the bigger one, it is regarded as independent peak instance and it won't be discarded.
- **isoNearNormalProbeTime:** If there are peaks detected from other isotopic distributions, they are used for the detection of isotopic patterns. If these peaks are not sufficiently close to

correct ones, they are out of interest. If the beginning of an isotopic peak is further away from a correct peak than this threshold in seconds, it is discarded.

- **relativeAreaCutoff:** This value has the same purpose as the **peakDiscardingAreaFactor**, but this higher threshold is applied after the isotopic peak removal.
- **relativeFarAreaCutoff:** Intensities that are close to the highest one of a chromatogram can be members of a twin peak, hence, they should normally not be discarded so quickly. This one is a more stringent cutoff for peaks that are further away.
- **relativeFarAreaTimeSpace:** This is a threshold in seconds to define when a peak is further away and the higher **relativeFarAreaCutoff** has to be applied.
- **relativelsolnBetweenCutoff:** Again, a stricter removal of peaks that may belong to other isotopes is performed, after the isotopic checks are finished.
- **solnBetweenMaxTimeDistance:** If the time distance between the borders of the two peaks is below this threshold in seconds, the check for an isotope in between is performed.
- **closePeakTimeTolerance:** The 3D algorithm is quite sensitive to changes in intensity aberrations to set peak borders. In a last step, it reunites accidentally split peaks. This parameter defines the maximally allowed time distance for the peak borders in seconds to select peaks for a reunion.
- **unionlnBetweenCutoff:** The intensity in between peaks for a peak union must not fall below this threshold which is calculated in relation to the smaller one of the two peak maxima.
- **twinlnBetweenCutoff:** The 3D algorithm is quite sensitive to changes in intensity aberrations to set peak borders which cause peak splits. This union step takes place already after checking if a peak comes from another isotopic distribution for obvious twin peaks. As for the **unionlnBetweenCutoff**, again, the intensity between the two peaks for a peak union must not fall below this threshold which is calculated in relation to the smaller one of the two peak maxima. For this value, there is usually a higher threshold used (95%), since it takes place at an earlier stage where there are more uncertainties about correctness of this peak.
- **twinPeakMzTolerance** (deprecated – not used anymore in LDA 2 versions): Tolerance in m/z direction to allow the union of twin peaks.

## Appendix D – mzTab settings

LDA provides the possibility to export metadata in the mzTab format. None of these information are mandatory for the mzTab export, but helps for the reproducibility of the data. All changes in these settings require a restart of the LDA application.

The metadata are split in two configuration files: i) **.properties** file containing information about the mass spectrometer ii) **mzTab.properties** containing experiment specific information such as affiliations and sample origin information.

### ***MS instrument specific settings (.properties)***

In the current implementation, the following metadata can be exported in mzTab format:

- **mzTabInstrumentName:** The name of the instrumentation.
- **mzTabInstrumentIonSource:** The ion source of the instrumentation (e.g. ESI, or MALDI).
- **mzTabInstrumentAnalyzer:** The analyzer of the instrument (e.g. ion trap, FT-ICR, etc.).
- **mzTabInstrumentDetector:** The ion detector (e.g. electron multiplier).

Examples are provided within the files. Be sure for all parameters that the entries contain always three “,” such as “mzTabInstrumentName=MS,MS:1000287,Time-Of-Flight Mass Spectrometer,”, otherwise these settings will be neglected. If you are using (as recommended) controlled vocabulary, please crosscheck its validity (<http://www.ebi.ac.uk/ontology-lookup/>).



## Other mzTab settings (mzTab.properties)

All the other metadata for mzTab is stored in the mzTab.properties file. The entries must contain three commas such as in the MS instrument specific settings, except for the contact settings which must contain at least two commas. The first comma separates the name, the second one the email address, and the rest of the commas are included in the affiliation. The contacts have to start with one, and MUST have consecutive numbers, e.g.  
contact\_1=\$name\_first\_contact\$, \$email\_first\_contact\$, \$affiliation\_first\_contact\$  
contact\_2=\$name\_second\_contact\$, \$email\_second\_contact\$, \$affiliation\_second\_contact\$.

Other supported metadata settings of this file are:

- species: from which species originates the sample
- tissue: from which tissue originates the sample
- celltype: the cell type of the sample

Other supported metadata where multiplicity is possible (contacts have to start with one, and MUST have consecutive number, such as for contact addresses):

- sampleprocessing\_1: list of separation techniques
- pubmedid\_1: list of PubMed IDs
- doi\_1: list of digital object identifiers

If any of the values are unknown or cannot be assigned, they should be commented or left out. The export to mzTab is possible with missing values.

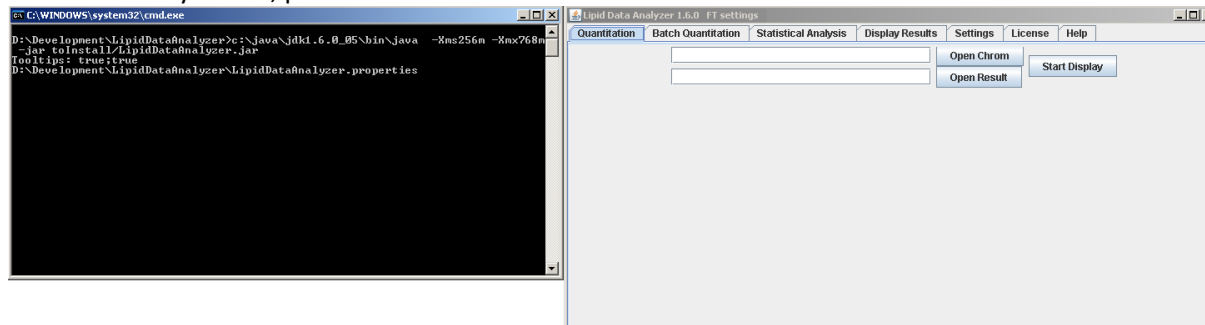
For mzTab export, the generic LDA adduct notation has to be transformed in an IUPAC compatible one. For each adduct type, a line in the following notation has to be entered:

adduct\_\$LDA\_NOTATION\$=\$IUPAC\_NOTATION\$;

e.g. adduct\_HCOO=[M+HCOO]-

## Appendix E - Trouble shooting

If there are any errors, please start the LDA with the console.



Then, please try to reproduce the error! Afterwards, please make a copy of the console output and send it with a detailed description of the preceding user interactions to the developers (lda@genome.tugraz.at).