

Lipid Search Module for



User Manual

Version 1.0

Ansgar Korf
Institute of Inorganic and Analytical Chemistry
University of Münster, Germany
ansgar.korf@wwu.de

Table of Contents

Introduction.....	3
Applicable data sets and data formats.....	3
Data processing.....	3
LC-HRMS data sets.....	3
Direct-infusion data sets.....	9
Using the lipid search module.....	12
Identification of lipids on sum formula level.....	12
Identification of lipid adducts.....	14
Extending the lipid class library.....	16
Identification of lipids on chain level.....	16
Identification of lipids on double bond position level.....	17
Further information.....	24

Introduction

This manual contains information on how to use the lipid search module in MZmine 2. The module is designed to annotate features in feature lists and aligned feature lists as potential lipids. The module supports various identification levels from sum formula, over chain composition up to double bond position level. Furthermore, the library of lipid classes can be extended by modifying the sum formula of a lipid class backbone. The same modification parameter can be used to identify adducts, oxidation products or derivatization products. Additionally, the module can also be applied on selected MS spectra of every MS level, adding annotations directly to the spectra. The module is designed to build custom databases, which are based on user parameter inputs to cover as much lipids as possible.

Applicable data sets and data formats

MZmine 2 is specialized for LC-HRMS data sets. However, shotgun lipidomics data sets (direct infusion experiments) can also be analyzed. All raw data formats supported by MZmine 2 can be used with the lipid search module. Open formats such as mzML are recommended. Not supported raw data formats can potentially be converted to open formats using other specialized software such as MSConvert.

Data processing

This section covers information on how to process raw data to get a suitable feature list for lipid identification.

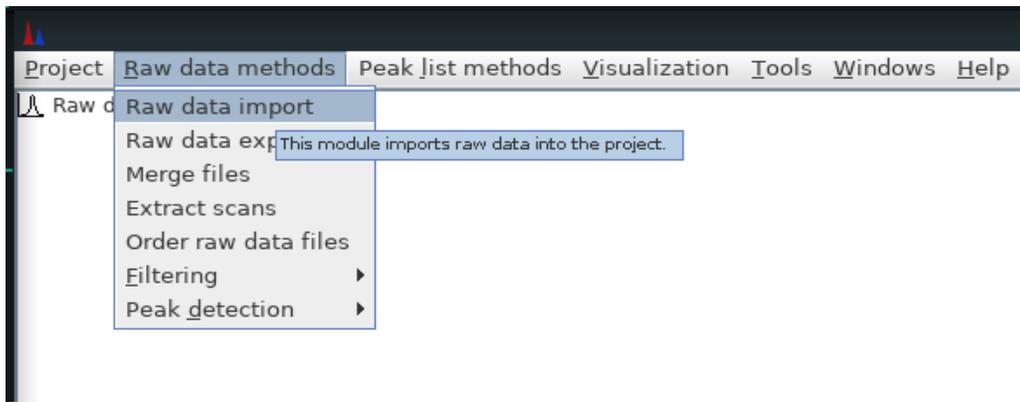
LC-HRMS data sets

Before describing each step to use the Lipid Search module, I like to recommend the Mzmine 2 manual, which covers all basic information from installation to processing, the MZmine 2 tutorial document of Mark Earll, which covers a processing workflow for LC-HRMS metabolomics data sets, and the tutorial video for LC-HRMS data processing by Sandra Castillo. All documents can be accessed online:

<https://mzmine.github.io/documentation.html>

To obtain as much useful information as possible I recommend the following processing steps for lipid identification:

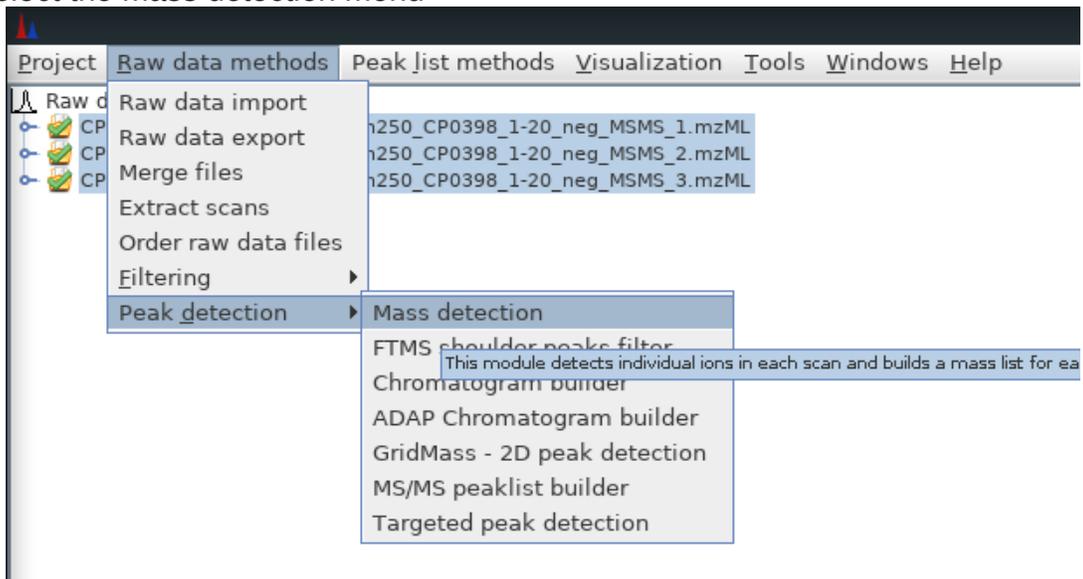
1. Raw data import



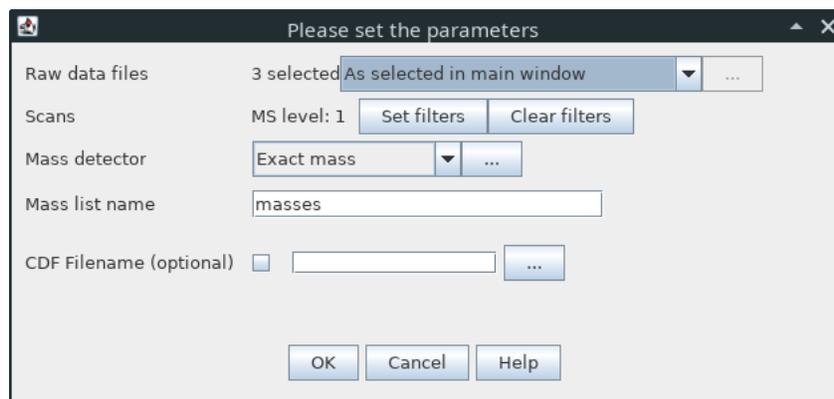
- select all raw data files

2. Mass detection

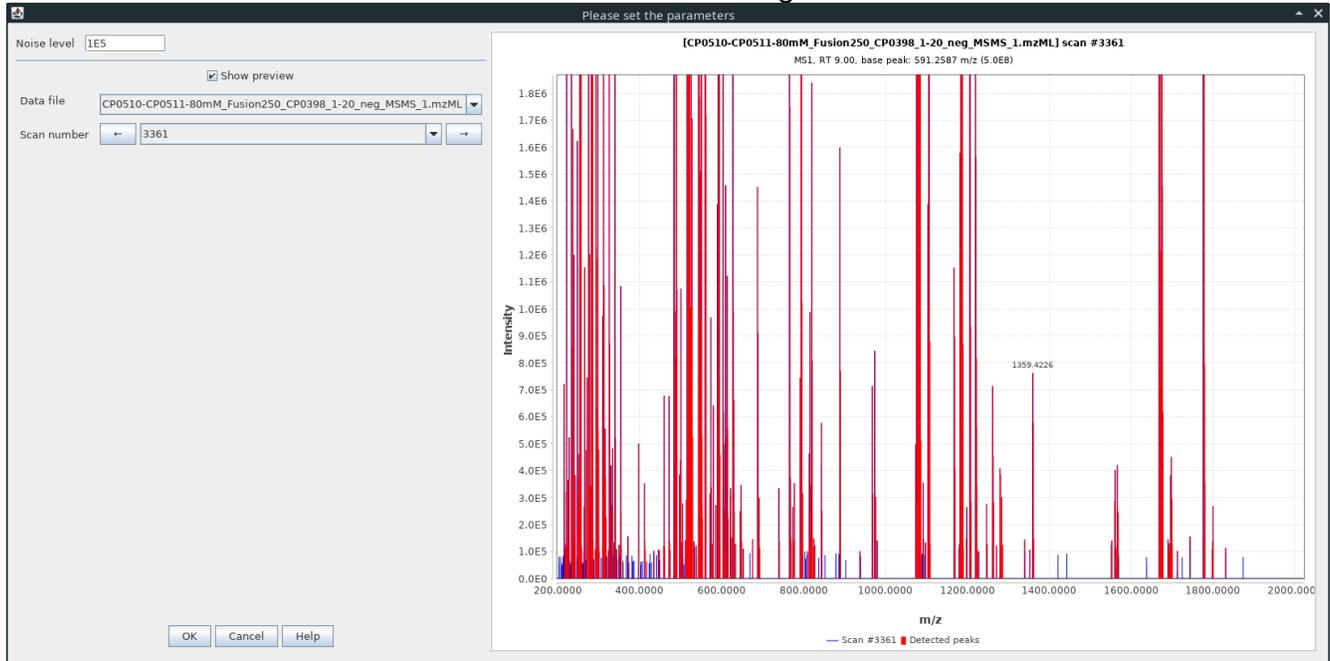
- select the mass detection menu



- select an algorithm (Mass detector) suitable for your data (The “Exact mass” algorithm is suitable for FTMS data)



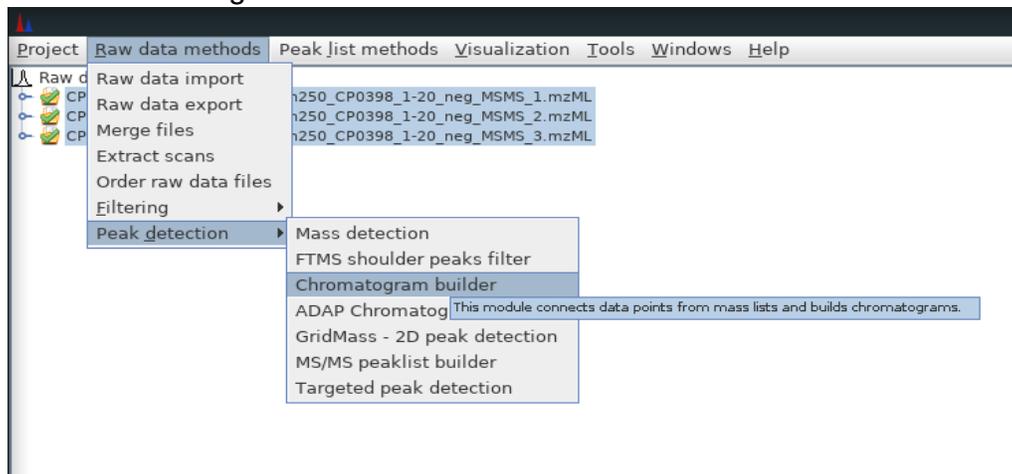
- select the “...” button to enter all required parameters
- select the “Show preview” check box to visualize the results of your settings for an example scan. This will help to determine the noise level parameter. Red signals will be added to the mass list and blue ones will be ignored



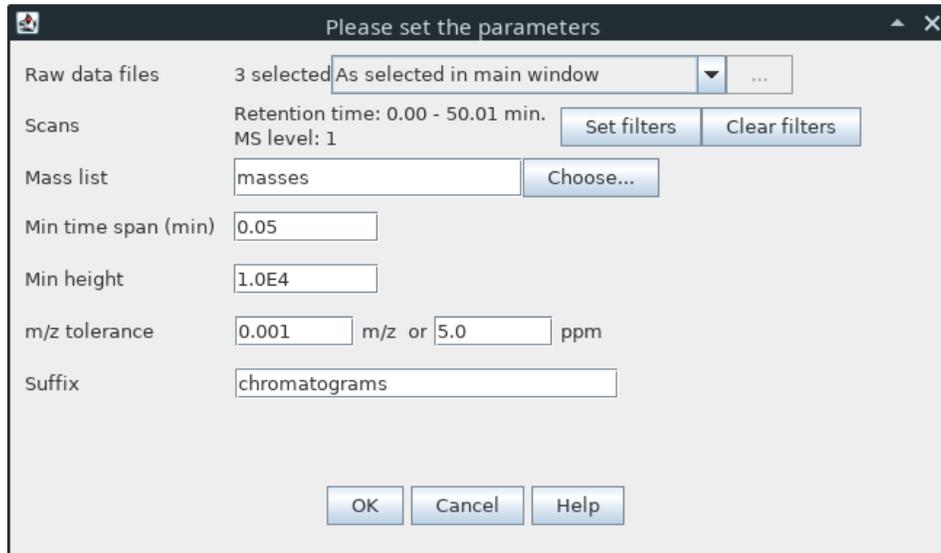
Caution! The noise level may change over the retention time due to different eluent concentrations when using a gradient. In that case you can create multiple mass lists for different sections of the chromatogram.

3. Chromatogram building

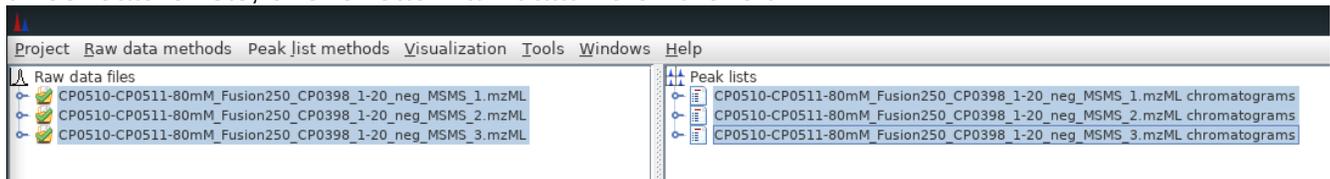
- select the chromatogram builder menu



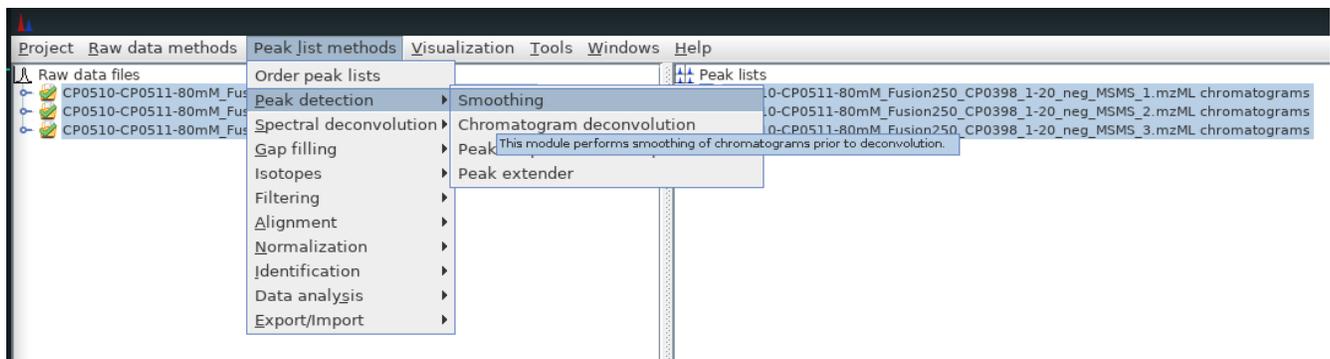
- enter all necessary parameters:
 - select raw data files
 - you can optionally filter the scans used for chromatogram building (e.g MS level, polarity, retention time etc.)
 - select the mass lists you have created in step 2
 - enter the minimum allowed time span of a peak
 - enter minimum peak height
 - enter *m/z* tolerance for scan binning



As a result you will obtain your first feature list on the right hand side. The example shows three feature lists, one for each raw data file on the left:



4. Smoothing (optional)

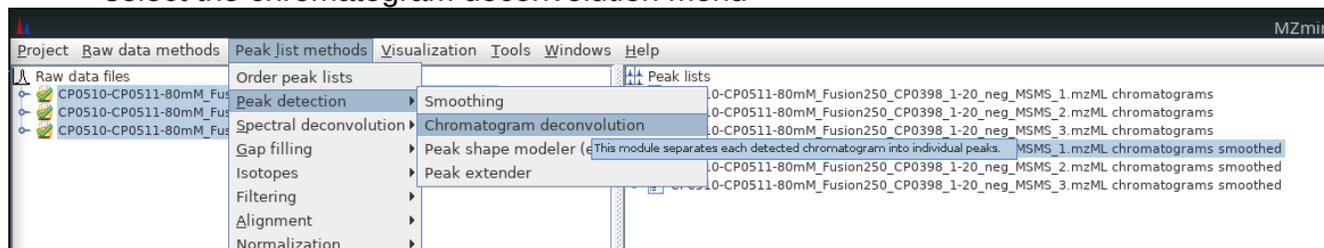


Smoothing can improve the results, especially for step 5 (deconvolution).

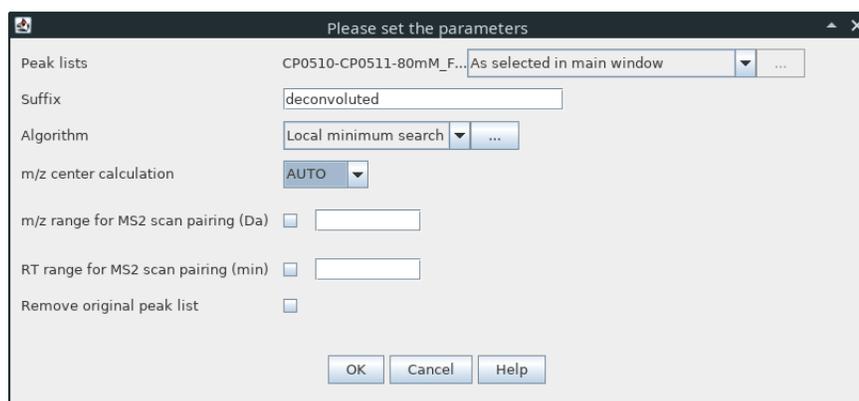
5. Chromatogram deconvolution

Chromatogram deconvolution separates a chromatogram into individual peaks. Also, noise can be removed from the feature list.

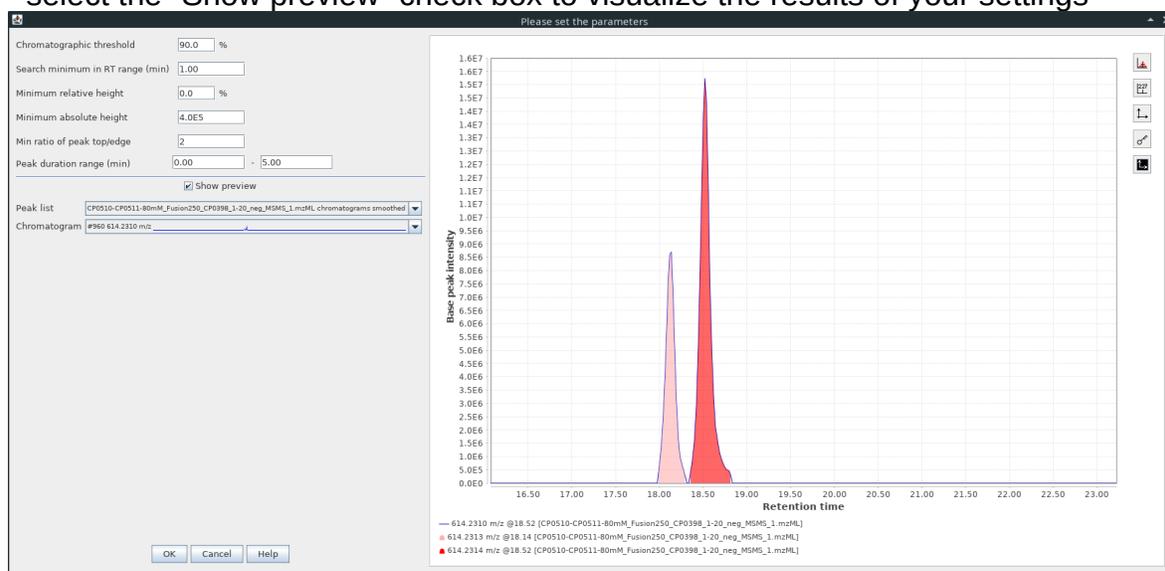
- select the chromatogram deconvolution menu



- select an algorithm and click the “...” button to enter all necessary parameters

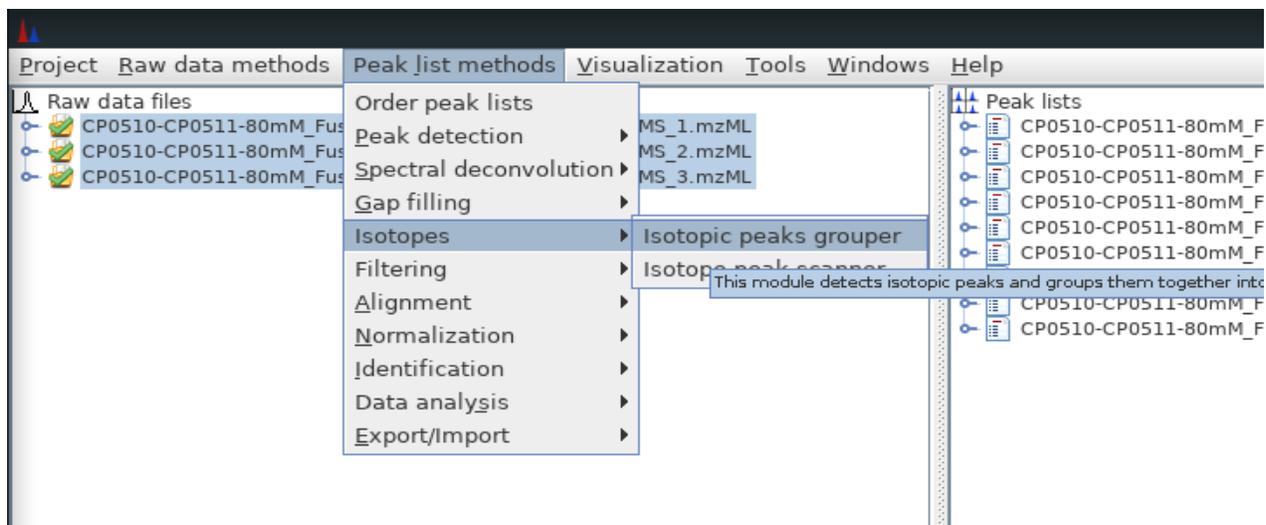


- select the “Show preview” check box to visualize the results of your settings



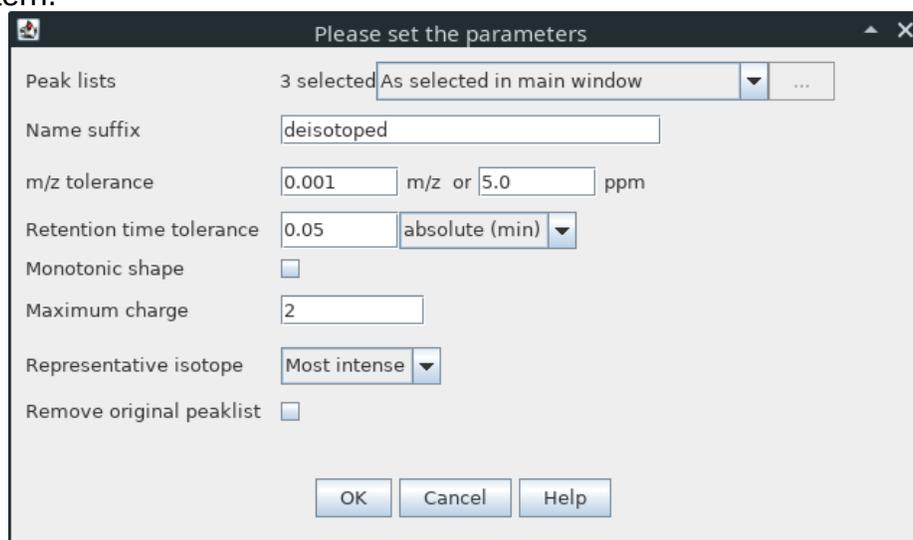
More details on chromatogram deconvolution can be accessed in the integrated MZmine 2 help files.

6. Deisotoping/Isotopic peak grouping



To reduce the number of false positive lipid annotations, isotopic features can be removed by grouping them to one feature. Select the isotopic peaks grouper menu

- enter all necessary parameters:
 - select peak lists
 - enter m/z tolerance window to search for isotopes
 - enter retention time tolerance window to search for isotopes
 - if you check “monotonic shape” then the monotonically decreasing height of an isotopic pattern is required
 - allow a maximum charge to narrow the number of possible isotopes to search
 - select the representative isotope, which is the feature that will be present in the resulting deisotoped feature list. The other isotopes are stored as an isotopic pattern.

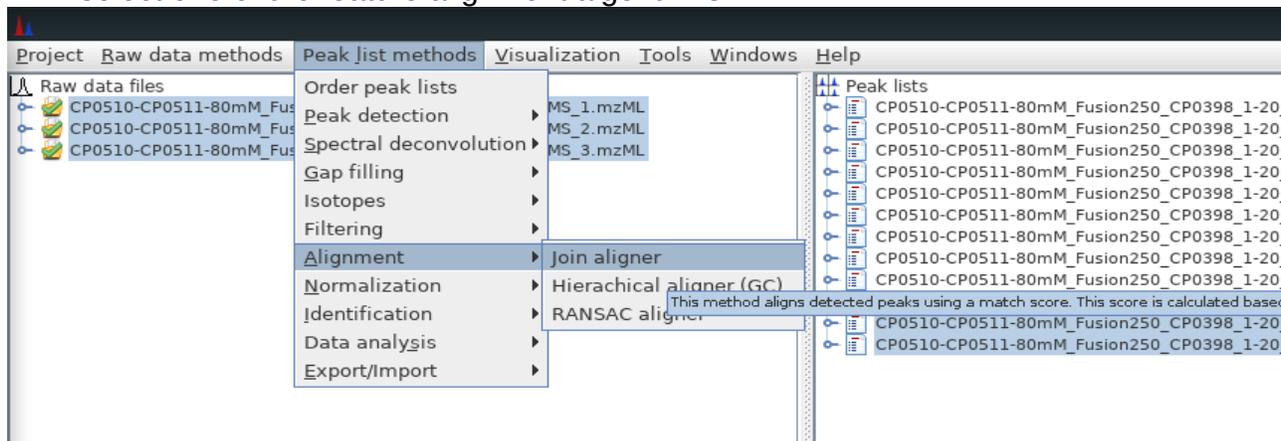


The resulting feature lists are suitable for lipid identification. Additionally, more methods can be applied to further filter the feature list. The lipid search module will technically work with every feature list.

7. Alignment (optional)

Feature alignment can be performed to combine feature lists in one aligned feature list, that compiles all the information in just one list. It makes the comparison between samples much easier.

- select one of the feature alignment algorithms



More details on feature alignment can be accessed in the integrated MZmine 2 help files. If you have decided to proceed with an aligned feature list you should also check the help files for gap filling.

More detailed information for each module used for processing can be accessed in the integrated MZmine 2 help documentation as well.

Direct-infusion data sets

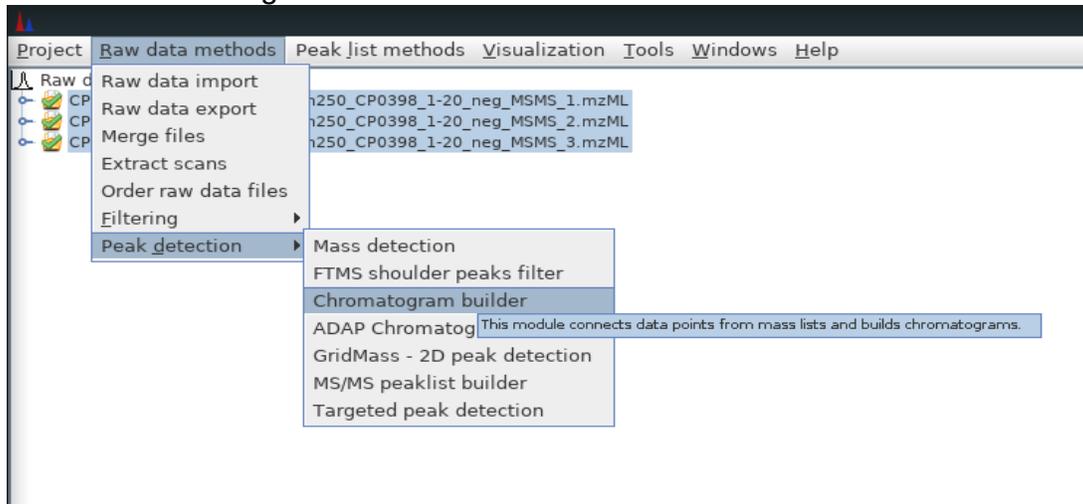
Although MZmine 2 is designed for LC-HRMS data sets, direct infusion data sets can also be effectively analyzed.

Perform steps 1. (raw data import) and 2. (mass detection) as described in the previous section.

3. "Chromatogram" building

The chromatogram builder will be used in this approach to average all recorded scans. You can set the minimum allowed time span to the length of your sample run. This results in filtering out noise signals, which are not present in every scan.

- select the chromatogram builder menu

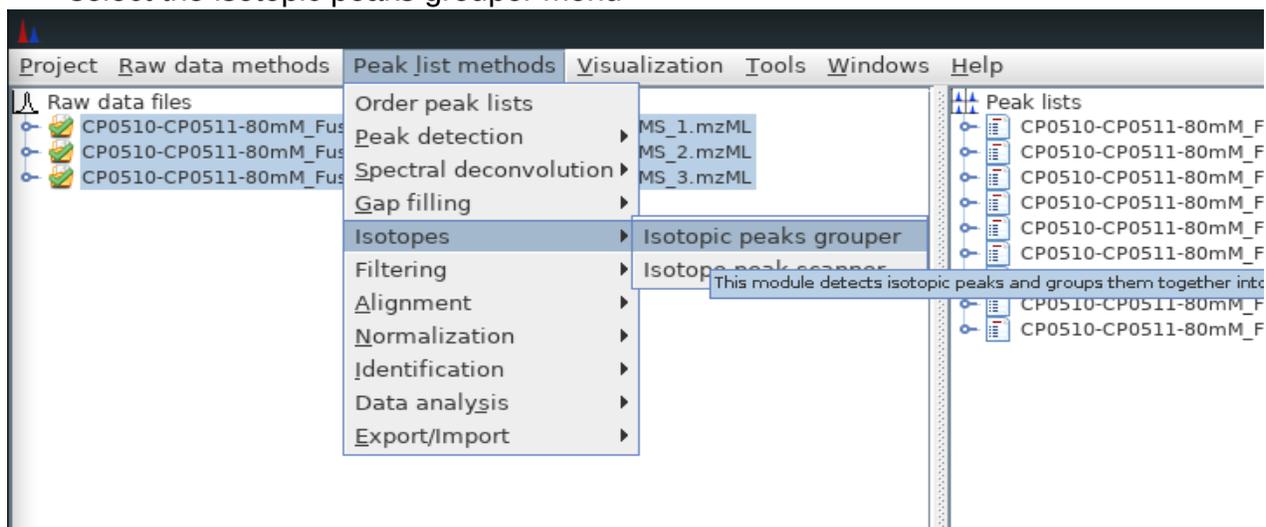


- enter all necessary parameters:
 - select raw data files
 - you can optionally filter the scans used for chromatogram building (e.g MS level, polarity, retention time etc.)
 - select the mass lists you have created in step 2
 - enter the minimum allowed time span. In this approach the time span should be set as long as the entire acquisition time.
 - enter minimum peak height
 - enter m/z tolerance for scan binning

4. Deisotoping/Isotopic peak grouping

To reduce the number of false positive lipid annotations isotopic features can be removed by grouping them to one feature.

- select the isotopic peaks grouper menu



- enter all necessary parameters:
 - select peak lists
 - enter m/z tolerance window to search for isotopes
 - enter retention time tolerance window to search for isotopes. For direct-infusion experiments, the tolerance must be set as long as the entire acquisition time, because the “retention time” will be set to the local maximum.
 - if you check “monotonic shape”, then the monotonically decreasing height of an isotopic pattern is required
 - allow a maximum charge to narrow the number of possible isotopes to search
 - select the representative isotope, which is the feature that will be present in the resulting deisotoped feature list. The other isotopes are stored as an isotopic pattern

The screenshot shows a dialog box titled "Please set the parameters" with the following fields and options:

- Peak lists:** 3 selected, As selected in main window (dropdown menu)
- Name suffix:** deisotoped (text input)
- m/z tolerance:** 0.001 (text input), m/z or 5.0 (text input), ppm (unit)
- Retention time tolerance:** 0.05 (text input), absolute (min) (dropdown menu)
- Monotonic shape:**
- Maximum charge:** 2 (text input)
- Representative isotope:** Most intense (dropdown menu)
- Remove original peaklist:**

Buttons at the bottom: OK, Cancel, Help

The resulting feature lists are suitable for lipid identification.

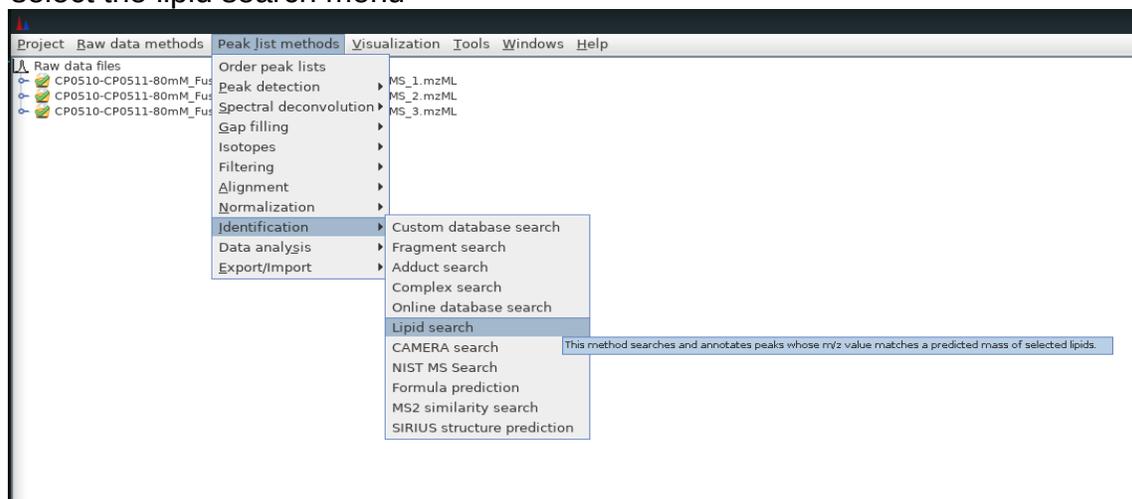
Using the lipid search module

The following sections will cover different approaches and application on how to use the lipid search module.

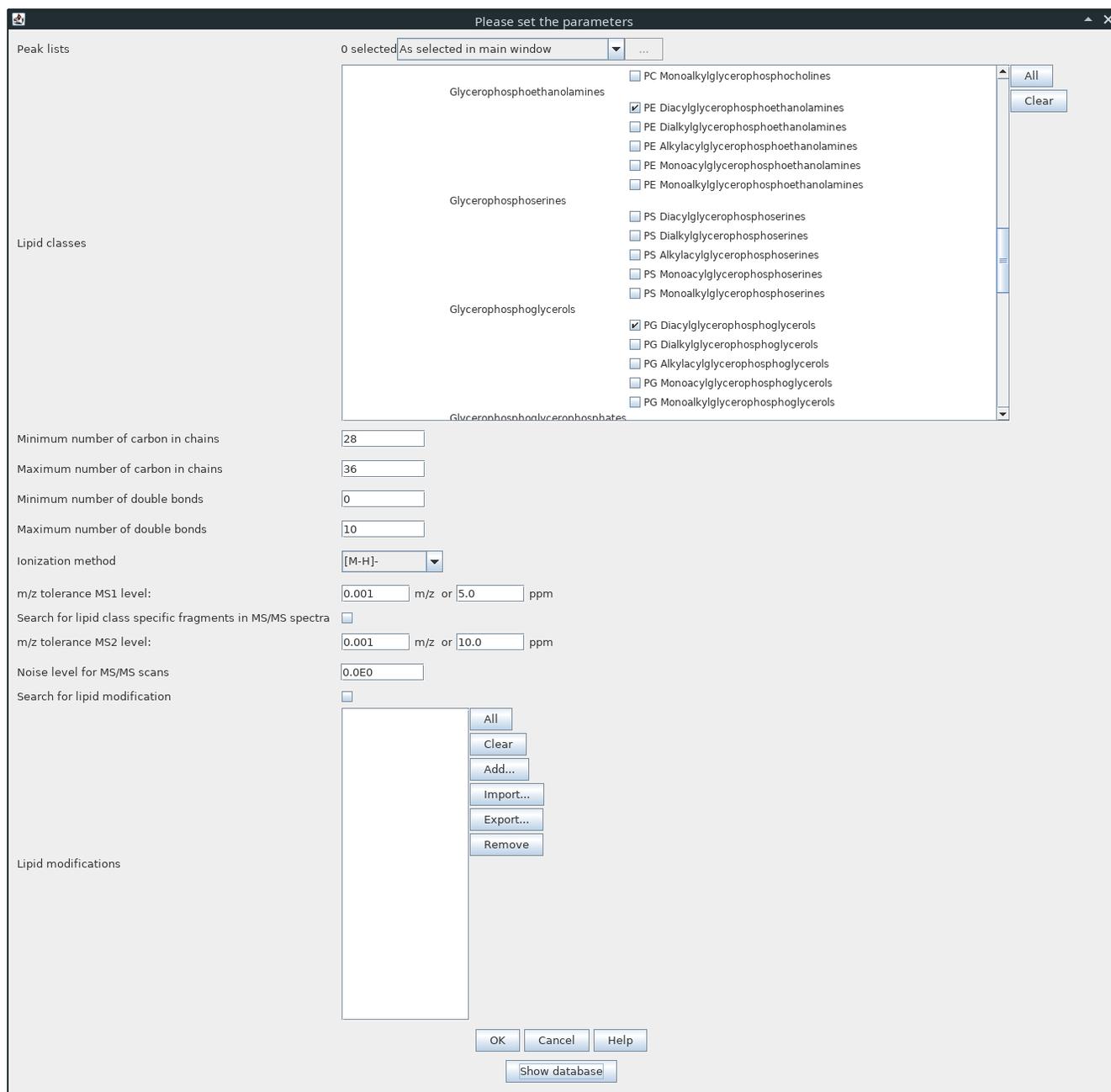
Identification of lipids on sum formula level

Add a lipid identity to a feature based on accurate m/z ratio.

- select the lipid search menu

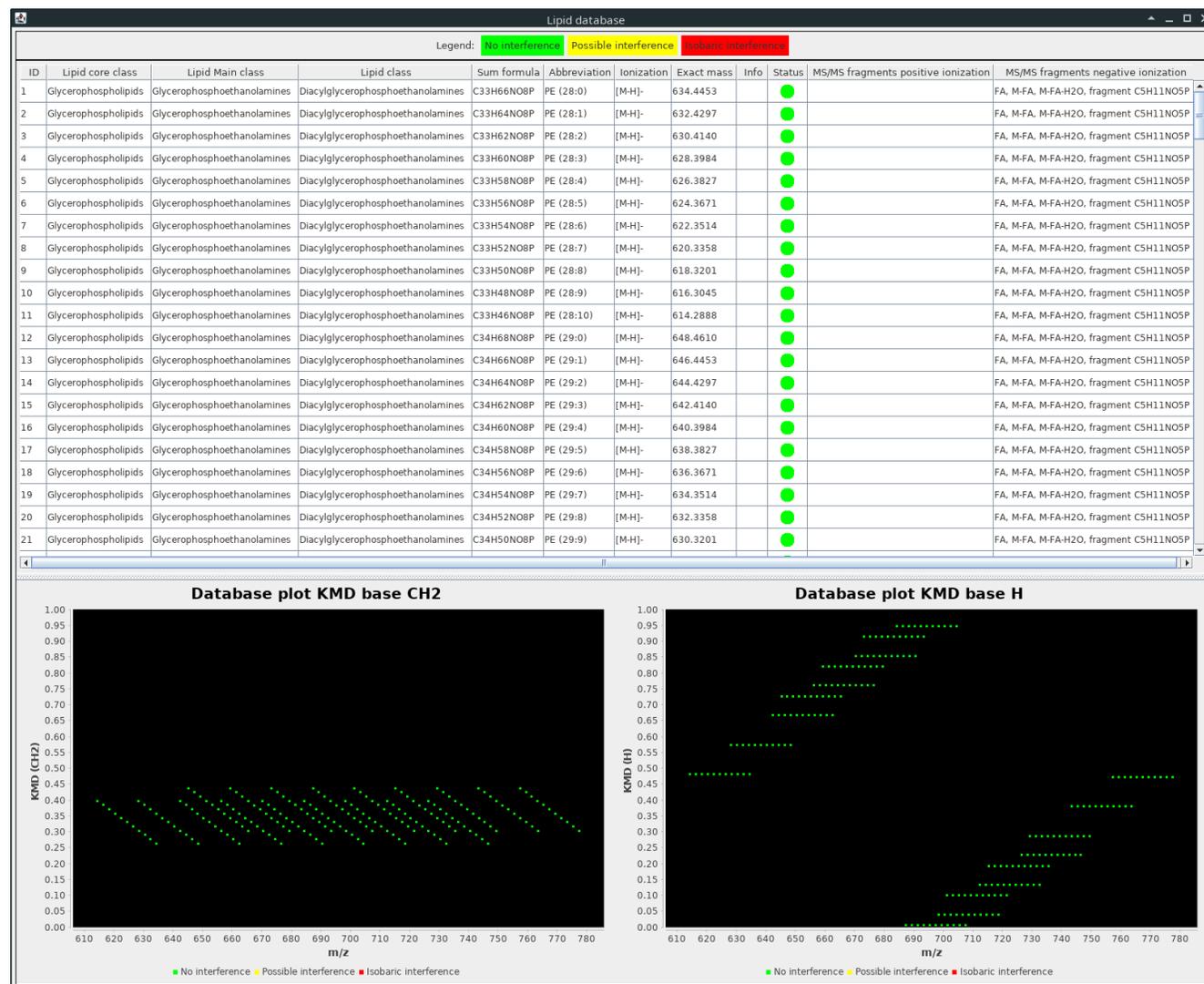


- enter all necessary parameters:
 - select peak lists
 - select all lipid classes you want to search through
 - enter the minimum number of allowed carbon atoms in all lipid chains
 - enter the maximum number of allowed carbon atoms in all lipid chains
 - enter the minimum number of allowed double bonds in all lipid chains
 - enter the maximum number of allowed double bonds in all lipids
 - select the ionization method to calculate the m/z ratios for the selected ion species in the database.
 - enter m/z tolerance on MS1 level to search for database hits



The database search and the annotation of the features will be started when pressing the "OK" button at the bottom of the frame. Prior to starting the search algorithm, the user can examine the generated database by clicking the "Show database" button. A new frame opens containing a table with all information on the generated lipid species. Furthermore, this database table can be copied to an excel sheet to create inclusion lists for data dependent experiments. The table and the two Kendrick mass plots, which haven been recently described in the literature as a data mining approach for LC-HRMS data sets of lipid extracts (Korf et al. 2018 *Three-dimensional Kendrick mass plots as a tool for graphical lipid identification*. *Rapid Communications in Mass Spectrometry*, 32(12), 981-991), below the

table are displaying information on possible interferences with other lipid species in the created database. Green: no interference; Yellow: possible interference in selected m/z tolerance window (e.g. 5 ppm); Red: isomeric interference.



Identification of lipid adducts

It is possible to search for multiple adducts of a lipid at the same time. Therefore, the user can add sum formulas to the lipid modifications parameter to calculate all possible m/z ratios. In the following example the algorithm will search for $[M-H]^-$, $[M+Cl]^-$, $[M+Br]^-$ and $[M+CH_3COO]^-$ ions of PE and PG lipids, ranging from 28 to 36 carbon atoms and 0 to 10 double bonds in all chains. Simply click on the check box "Search for lipid modification". The user can then manually add the ion species of interest using the "Add..." button on the right side of the parameter. Enter the sum formula in the lipid modification text field. Add a "+" **before** all atoms you want to add and the number of atoms directly **after** the atom. The user can also subtract atoms by adding "-" **before** the atoms to subtract. If the user wants to add and subtract atoms in one modification please follow the pattern "+X_nY_m-Z_l". In other words, first

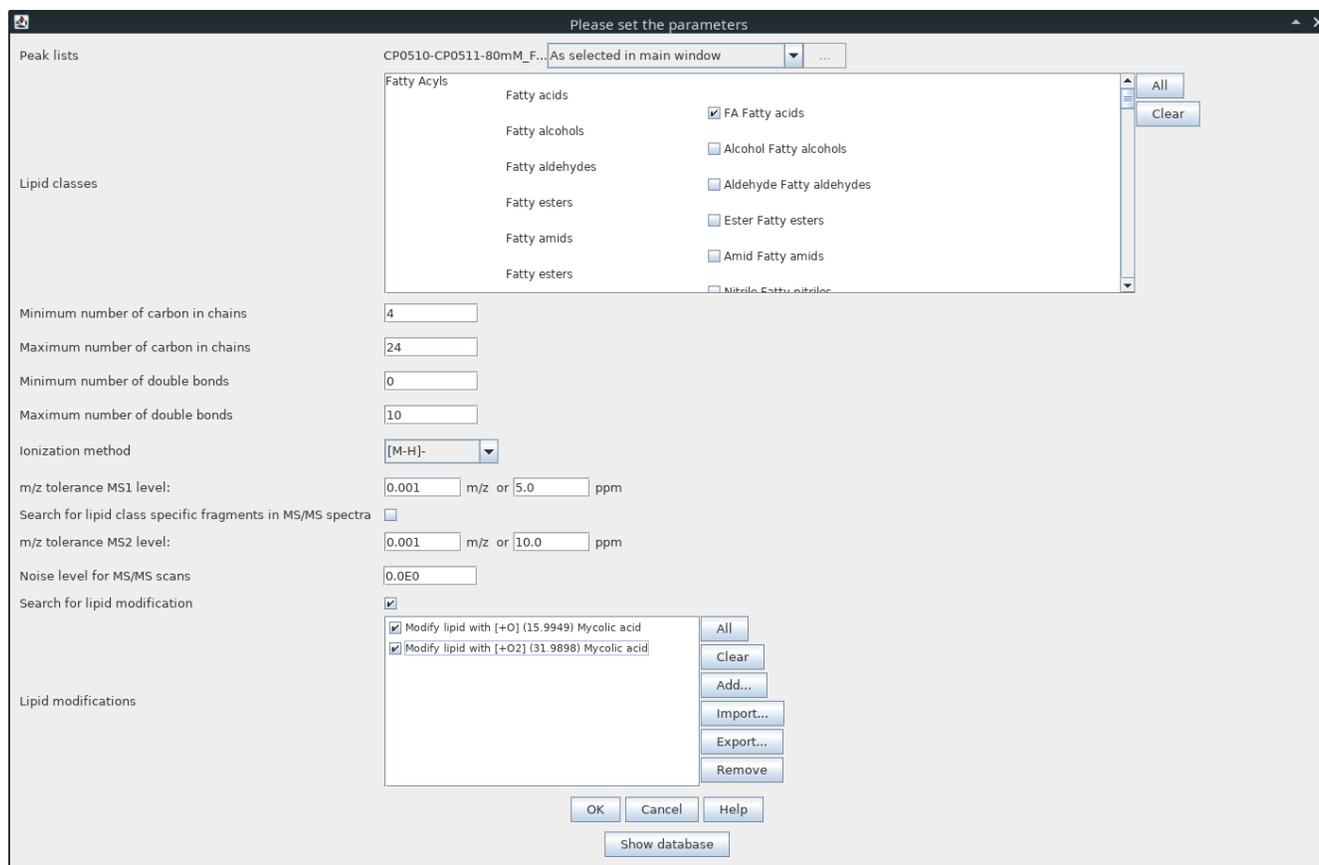
add all atoms, second subtract all atoms, do not mix or subtract first. Once a modification was added by pressing “OK”, the user can review the resulting mass difference of the added modification in the “Lipid modifications” check box list. Furthermore, the user also needs to add a “Lipid modification label”, which allows easier analysis of annotated features.

Important! The algorithm starts calculating a modified lipid from the ionized lipid. Therefore, the lipid modification must add or subtract the atoms of the selected ionization method. In this example [M-H]⁻ was selected as ionization method. To search for [M-Cl]⁻ species, one needs to add “+HCl”.

Added lipid modifications can be exported to a .csv file and re-imported for other analysis tasks.

Extending the lipid class library

The “Lipid modifications” parameter described in detail in section “Identification of lipid adducts”, can also be used to create lipid classes, which are not part of the core library. As an example, the user can systematically search for the highly diverse lipid class Mycolic acids on MS1 level by selecting “Fatty acids” as lipid class and adding various modifications. The user can add e.g. “+O” and/or “+O2” to address hydroxylations as lipid modification.



Further possible side chain modifications can be addressed in a similar fashion.

Identification of lipids on chain level

An automated identification of lipids on chain level is supported for most common lipid classes. The identification is based on in-silico libraries of other projects such as LipidBlast or results derived from the literature. Please note, a data dependent data acquisition is mandatory to use the automated chain level function of this module, because the function is based on MS/MS spectra search. To use the function simply select the check box “Search for

lipid class specific fragments in MS/MS spectra". Furthermore, the algorithm needs an *m/z* tolerance to search for fragments in MS/MS spectra, and a noise level can be set as well. The noise level can also be 0.0, if all signals should be considered as possible fragments.

m/z tolerance MS1 level: m/z or ppm

Search for lipid class specific fragments in MS/MS spectra

m/z tolerance MS2 level: m/z or ppm

Noise level for MS/MS scans

The program searches for a lipid species on MS1 level first. If a feature was annotated as a lipid on MS1 level, the algorithm checks if one or more MS/MS spectra of the precursor have been recorded. If MS/MS scans are available it loops through all MS/MS spectra to search for class specific MS/MS fragments. All identified fragments are then added to the comment section of the feature. Also, information on the scan number is added. If class specific fragments are pointing to an acyl chain, the program checks if the combination of those fragments are in accordance with the lipid species annotation on MS1 level. If everything fits, a fatty acyl chain composition will be added to the comment, again with information on the MS/MS scan number. The following screenshot shows a feature list with annotated PG species on acyl chain level.

Peak list: CP0510-CP0511-80mM_Fusion250_CP0398_1-20_neg_M

ID:	Average		Identity	m/z:	RT:
ID	m/z	RT			Comment
1009	741.4686	9.99	Diacylglycerophosphoglycerols PG(34:4)		Ionization: H-1, Δ 3.573 ppm; FA(16:1)_FA(18:3) MS/MS scan 3765, RT 10.06
1010	741.4687	11.94	Diacylglycerophosphoglycerols PG(34:4)		Ionization: H-1, Δ 3.443 ppm; FA(16:2)_FA(18:2) MS/MS scan 4427, RT 11.81; FA(16:1)_FA(18:3) MS/MS scan 4427, RT 11.81;
1021	743.4847	11.88	Diacylglycerophosphoglycerols PG(34:3)		Ionization: H-1, Δ 2.837 ppm; FA(16:1)_FA(18:2) MS/MS scan 4398, RT 11.74; FA(16:0)_FA(18:3) MS/MS scan 4398, RT 11.74
1029	745.4997	9.96	Diacylglycerophosphoglycerols PG(34:2)		Ionization: H-1, Δ 3.699 ppm; FA(16:1)_FA(18:1) MS/MS scan 3684, RT 9.84; FA(16:0)_FA(18:2) MS/MS scan 3684, RT 9.84
1030	745.4998	11.83	Diacylglycerophosphoglycerols PG(34:2)		Ionization: H-1, Δ 3.574 ppm; FA(16:1)_FA(18:1) MS/MS scan 4374, RT 11.67; FA(16:0)_FA(18:2) MS/MS scan 4418, RT 11.79
971	731.4842	11.91	Diacylglycerophosphoglycerols PG(33:2)		Ionization: H-1, Δ 3.639 ppm; FA(16:1)_FA(17:1) MS/MS scan 4438, RT 11.84
945	717.4687	11.97	Diacylglycerophosphoglycerols PG(32:2)		Ionization: H-1, Δ 3.46 ppm; FA(16:2)_FA(16:0) MS/MS scan 4475, RT 11.94
952	719.4843	10.11	Diacylglycerophosphoglycerols PG(32:1)		Ionization: H-1, Δ 3.596 ppm; FA(16:1)_FA(16:0) MS/MS scan 3774, RT 10.08
953	719.4842	11.97	Diacylglycerophosphoglycerols PG(32:1)		Ionization: H-1, Δ 3.692 ppm; FA(16:1)_FA(16:0) MS/MS scan 4452, RT 11.88
100	278.2193	3.68			
1004	740.5210	16.42			

Identification of lipids on double bond position level

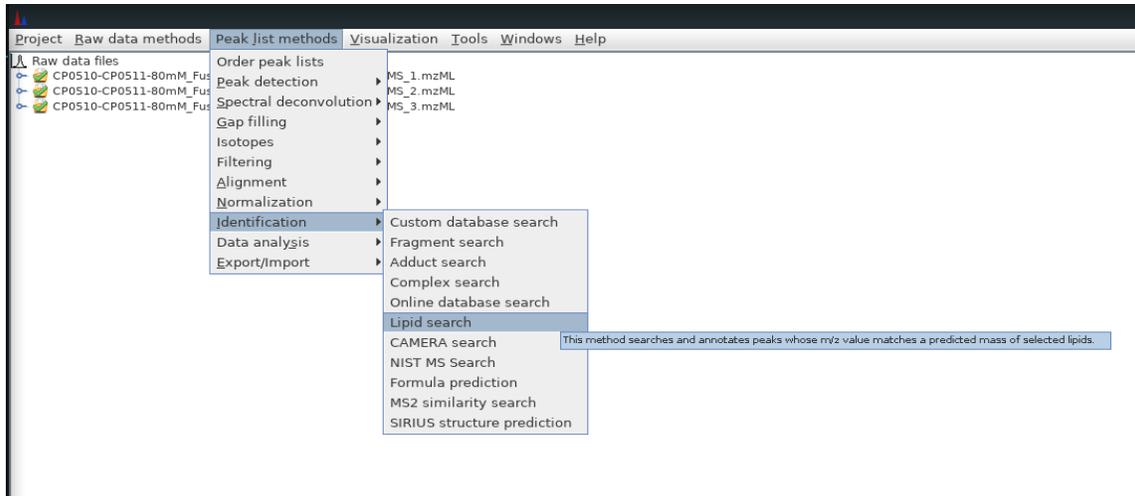
The position of double bonds in unsaturated fatty acids in complex lipids is not accessible by applying conventional low energy collision-induced dissociation (CID) tandem MS. Therefore, derivatization based techniques to obtain diagnostic fragments in MS/MS experiments of lipids that do pinpoint to a double bond position have become of great interest in the literature. One recent approach is the derivatization of double bonds by photochemical Paternò-Büchi (PB) reaction. The PB reaction itself is initiated by UV light activated acetone, which forms an oxetane ring with the double bond of lipid chains. The obtained acetone adducts can be

fragmented in CID-MS/MS experiments to obtain diagnostic fragments, which do pinpoint to a double bonds' position.

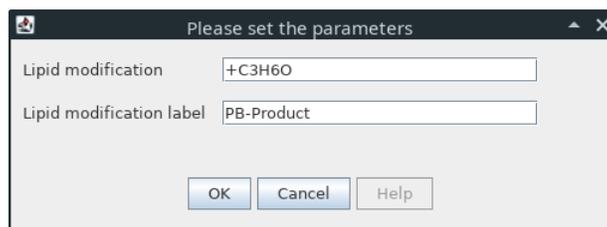
In the following procedure will be described how to identify diacylglyceryltrimethylhomoserine (DGTS) lipid species on double bond position level using the lipid search module in a dataset which was acquired with an online PB reaction as post column derivatization. Please note that the module can be used in a similar fashion for other derivatization reactions.

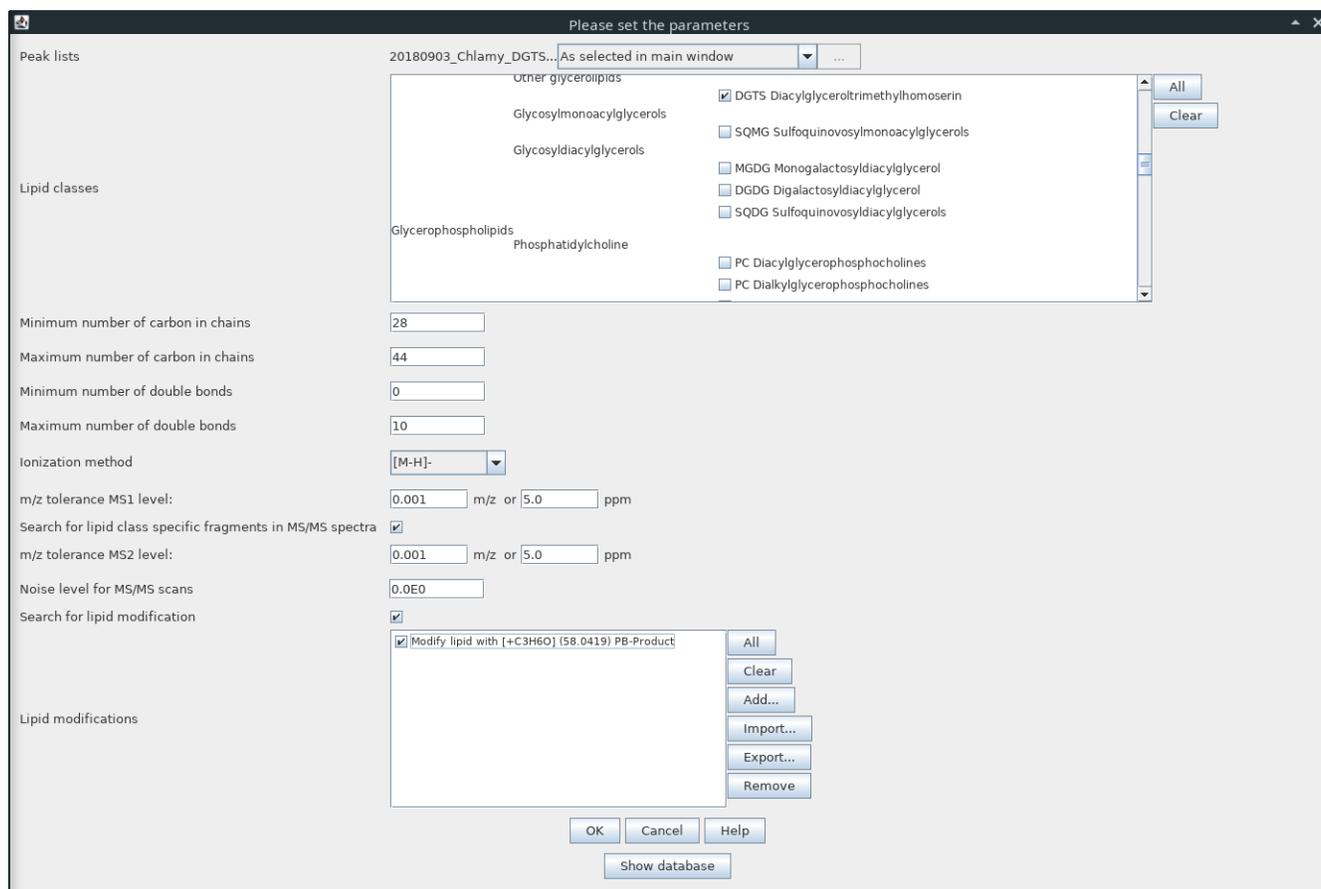
1. Search for lipid species and PB products

- open the lipid search module



- enter all necessary parameters:
 - select peak lists
 - select all lipid classes you want to search through
 - enter the minimum number of allowed carbon atoms in all lipid chains
 - enter the maximum number of allowed carbon atoms in all lipid chains
 - enter the minimum number of allowed double bonds in all lipid chains
 - enter the maximum number of allowed double bonds in all lipids
 - select the ionization method to calculate the m/z ratios for the selected ion species in the database.
 - enter m/z tolerance on MS1 level to search for database hits
 - select the check box "Search for lipid class specific fragments in MS/MS spectra"
 - enter m/z tolerance on MS2 level to search for class specific fragments in MS/MS spectra
 - enter the noise level for MS/MS scans, can be 0.0 to loop through all signals
 - select "Search for lipid modifications" parameters
 - select "Add..." button on the right and add PB-product modification (+C₃H₆O)





- Click “OK” to start the search algorithm

2. Open the analyzed feature list

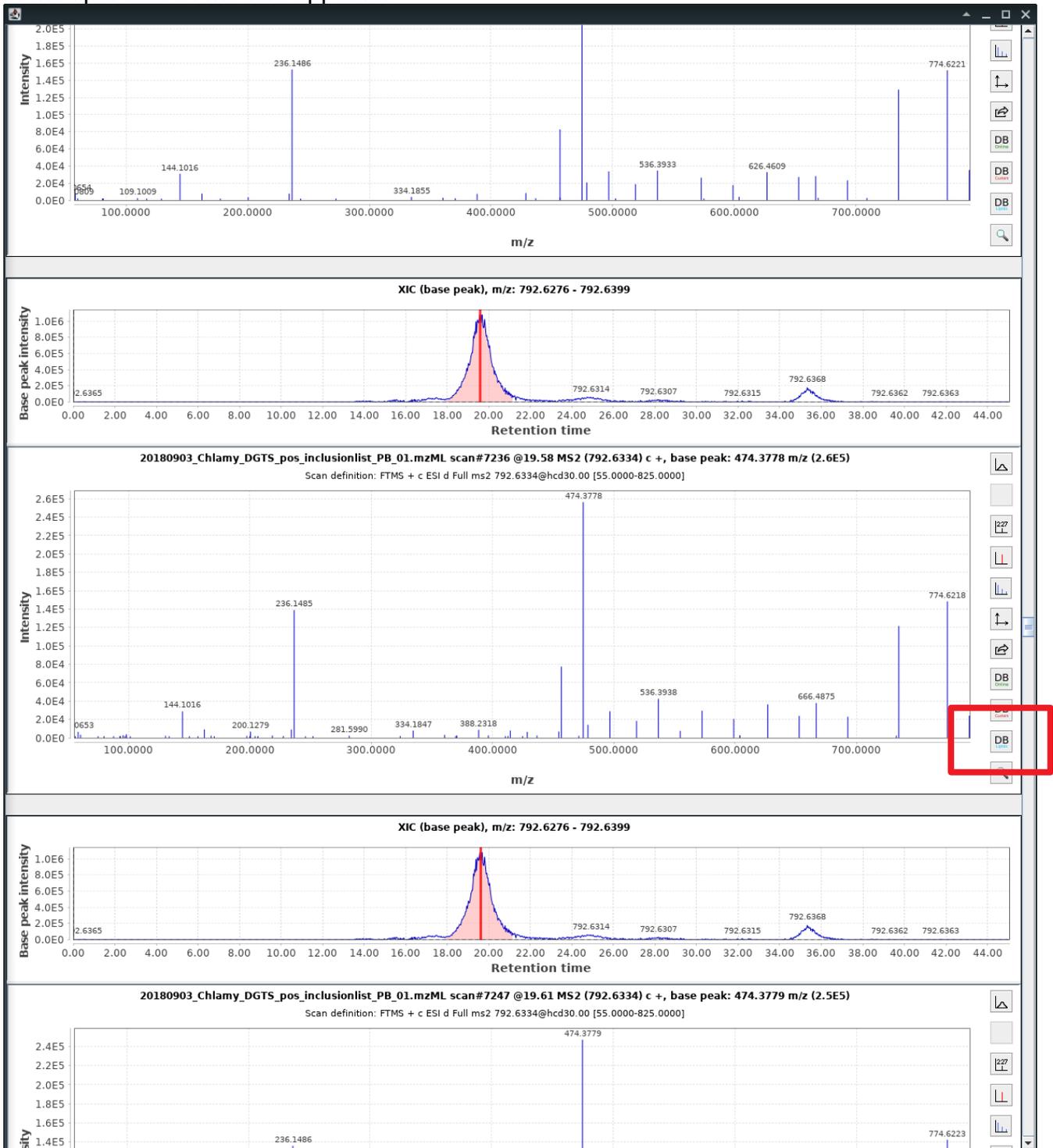
- sort the list by “Identity” (click on identity)

The feature of the lipid species, which includes information on the chain composition, and the corresponding PB-product should appear directly below one another, if present. Please note, post-column derivatizations rarely result in a 100% yield of the derivatization product and it is always a trade-off between yields and peak shape. In this case observing educt and product simultaneously gives information on chain level and double bond position level in a single run.

Peak list: 20180903_Chlamy_DGTS_pos_inclusionlist_PB_01.mzML chromatograms deconvoluted10Min deisotoped filtered							
ID:	Average		Identity	Comment	Peak shape	20180903_Chlamy	
ID	m/z	RT				Status	
1150	736.6067	23.93	Diacylglyceroltrimethylthomoserin DGTS(34:2)	Ionization: H, Δ 2.535 ppm; FA(18:1)_FA(16:1) MS/MS scan 8185, RT 22.04; fragment C7H14NO2 MS/MS scan 8185...		●	
1465	794.6494	24.09	Diacylglyceroltrimethylthomoserin DGTS(34:2) Modify lipid with [+C3H6O] (58.0419) PB-Product	Ionization: H Modify lipid with [+C3H6O] (58.0419) PB-Product. Δ 1.259 ppm		●	
1144	734.5910	19.47	Diacylglyceroltrimethylthomoserin DGTS(34:3)	Ionization: H, Δ 2.57 ppm; FA(18:2)_FA(16:1) MS/MS scan 6077, RT 16.60; fragment C7H14NO2 MS/MS scan 6077, ...		●	
1449	792.6334	19.58	Diacylglyceroltrimethylthomoserin DGTS(34:3) Modify lipid with [+C3H6O] (58.0419) PB-Product	Ionization: H Modify lipid with [+C3H6O] (58.0419) PB-Product. Δ 1.79 ppm		●	
1136	732.5751	15.33	Diacylglyceroltrimethylthomoserin DGTS(34:4)	Ionization: H, Δ 2.977 ppm; FA(18:3)_FA(16:1) MS/MS scan 5337, RT 14.70; FA(18:4)_FA(16:0) MS/MS scan 5337, RT...		●	
1440	790.6178	15.39	Diacylglyceroltrimethylthomoserin DGTS(34:4) Modify lipid with [+C3H6O] (58.0419) PB-Product	Ionization: H Modify lipid with [+C3H6O] (58.0419) PB-Product. Δ 1.727 ppm		●	

3. Observe double bond position diagnostic fragments

A slightly modified version of the parameter setup dialog, specialized for spectra annotation, of the lipid search module appears.

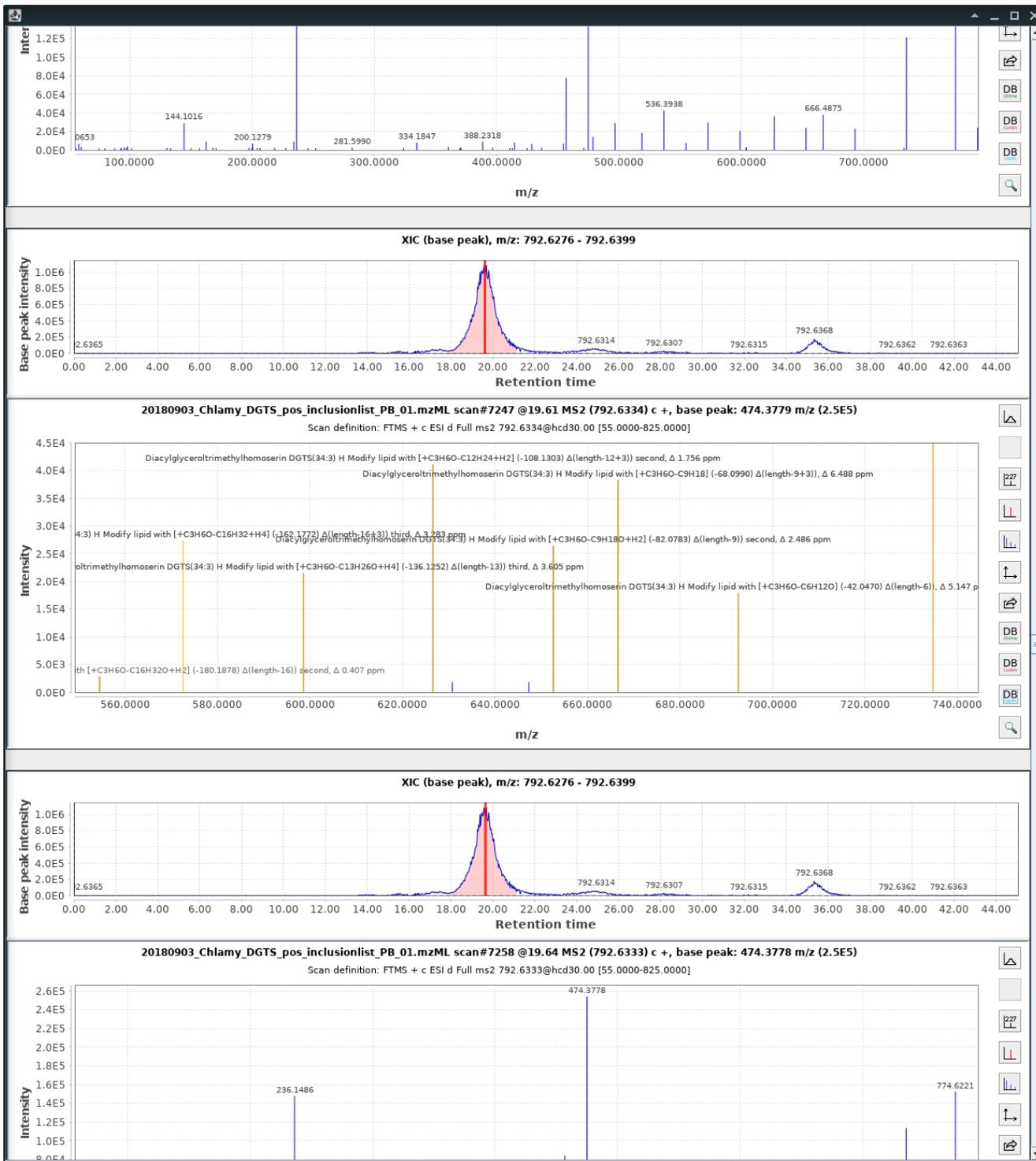


- enter all necessary parameters:
 - select the lipid class of the annotated PB product

- enter the number of carbon atoms of the annotated species as minimum number of allowed carbon atoms in all lipid chains
- enter the number of carbon atoms of the annotated species as maximum number of allowed carbon atoms in all lipid chains
- enter the number of double bonds of the annotated species as minimum number of allowed double bonds in all lipid chains
- enter the number of double bonds of the annotated species as maximum number of allowed double bonds in all lipid chains
- select the ionization method to calculate the m/z ratios for the selected ion species in the database.
- enter m/z tolerance to search for database hits
- enter the noise level, can be 0.0 to loop through all signals
- select “Search for lipid modifications” parameters
- select the “Import...” button on the right and select the .csv file with all possible diagnostic PB fragments, which is in the SI of the corresponding publication of the here presented module.
Please note, you can make your own custom list by adding modifications manually. The list can be saved and re-used for other projects, using the “Import...” button function

Select the “OK” button to start the search algorithm. Signals of identified diagnostic fragments are highlighted in orange and a label is added with information on the double bond position.

The parameters of the lipid search module are stored. This allows the analysis of another scan with only two clicks.



Further information

When you use the Lipid Search module for your lipid identification please cite the following articles:

Korf, Ansgar *et al.* "Lipid species annotation at double bond position level with custom databases by extension of the MZmine 2 open-source software package." *Analytical Chemistry* DOI: 10.1021/acs.analchem.8b05493

and

Pluskal, Tomáš, *et al.* "MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data." *BMC bioinformatics* 11.1 (2010): 395.