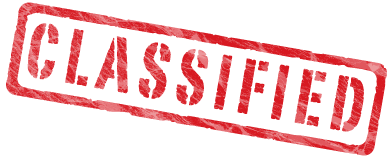
**MetExtract 2.0 β**

User guide



Version 0.4.9   
 Authors Christoph Bueschl   
 Bernhard Kluger   
 Rainer Schuhmacher

Contents

[Scope of MetExtract 4](#_Toc361220084)

[Definitions 4](#_Toc361220085)

[MS signal / Ion signal 4](#_Toc361220086)

[Feature 4](#_Toc361220087)

[Feature Pair 5](#_Toc361220088)

[Feature group 5](#_Toc361220089)

[File conversion 5](#_Toc361220090)

[ReAdW 5](#_Toc361220091)

[Proteo Wizard 5](#_Toc361220092)

[MzXML viewer 5](#_Toc361220093)

[R configuration 6](#_Toc361220094)

[Program start 6](#_Toc361220095)

[Input tab page 6](#_Toc361220096)

[Define group of measurement files 6](#_Toc361220097)

[Inspect or Edit group 7](#_Toc361220098)

[Delete group 7](#_Toc361220099)

[Large/Many groups 7](#_Toc361220100)

[File validation 7](#_Toc361220101)

[Select scan event 7](#_Toc361220102)

[Calculate tab page 7](#_Toc361220103)

[Specify Parameter settings [Individual files processing] 7](#_Toc361220104)

[Selection of Experiment type 8](#_Toc361220105)

[Parameter settings for MS signal processing 10](#_Toc361220106)

[Chromatographic Peak processing 12](#_Toc361220107)

[Save results 14](#_Toc361220108)

[Multiple files annotation 14](#_Toc361220109)

[Group results 14](#_Toc361220110)

[Integrate missing peaks 15](#_Toc361220111)

[Run data processing 15](#_Toc361220112)

[Multi core systems 15](#_Toc361220113)

[Results tab page 15](#_Toc361220114)

[External viewer 15](#_Toc361220115)

[*m/z*s 15](#_Toc361220116)

[*m/z* Bins 16](#_Toc361220117)

[Features 16](#_Toc361220118)

[Feature groups 17](#_Toc361220119)

[Filters 18](#_Toc361220120)

# Scope of MetExtract

The aim of MetExtract is to provide a software tool capable of analysing liquid chromatography high resolution mass spectrometry (LC-HRMS) full scan data derived from mixtures of stable isotope labelled (SIL) and non-labelled (i.e. naturally composed) biological samples (see fig. 1 for a carbon labelling example). The software works with data of samples either derived from full metabolome labelling (cultivation on natural and labelled nutrition source) or samples after isotope labelled tracer / xenobiotic metabolisation (e.g. metabolisation of a mycotoxin).

Figure 1: Illustration of the two distinct isotopic patterns within a mass spectrum of a substance available as   
natural 12C form (teal) and U-13C enriched form (red). The mass difference between the monoisotopic 12C and   
uniformly labelled 13C MS signals is proportional to 1.00335 u (i.e. the mass increment between 12C and 13C isotope)   
multiplied by the number of carbon atoms in the respective substance.

MetExtract is designed to work with centroid data and does not support profile mode data. Further, it requires the natural and labelled compounds to be analysed as mixtures present in the same sample and does not support the evaluation of data obtained from separate, successively measured natural and labelled sample solutions.

MetExtract is not capable of processing MS/MS data. For this purpose please refer to Lehner and Neumann *et al.* (2012) TODO: insert proper citation once the paper is out

## Definitions

### MS signal / Ion signal

In the following, an MS signal corresponds to a centroid signal in a mass spectrum and consists of an m/z value and an ion abundance value. Additionally, for each MS signal the recording MS scan number is known.

### Feature

A feature denotes a chromatographic peak of a particular metabolite ion. Different isotopes as shown in fig. 1 are therefore different features. It can be described as a two-dimensional bounded signal within a defined *m/z* window and a retention time window and each feature has the following properties:

* *m/z* value (i.e. mean of extracted *m/z* window)
* retention time estimate (i.e. center of MassSpecWavelet peak TODO cite DU et al)
* intensity value (i.e. peak area after integration)

### Feature Pair

The MetExtract software offers valuable additional information which can be assigned to each individual monoisotopic feature. For the systematic use of labelling specific isotopic patterns, MetExtract refers to corresponding feature pairs, which are, in case of 13C labelling, originating from the 12C monoisotopic ion (denoted as M) as well as the corresponding fully 13C labelled ion (denoted as M`). Corresponding 12C and 13C features of a feature pair show highly similar chromatographic profiles and directly reveal the number of labelling atoms for the respective feature pair. Additionally the charge state of the extracted feature is deduced and verified by both isotopic distributions.

For metabolisation studies, conjugation products of the applied precursor can be formed which possess additional atoms of the labelling element. The number of atoms of the labelling element per feature pair cannot be deduced directly from the mass difference between non-labelled and corresponding fully labelled ions since only the applied tracer / xenobiotic contains the heavier isotope of the labelling element. For the conjugated moiety, all atoms will show of their natural isotopic composition. Therefore, MetExtract will only annotate the labelled part of the remaining tracer with the number of labelling atoms per measured ion species.

### Feature group

A feature group is derived from combining all features belonging to the same metabolite. Since the individual 12C monoisotopic features of a particular metabolite (e.g. adducts-, in-source fragments) are generated in the ion source of the mass spectrometer (after the chromatographic separation), a retention time window and the Pearson correlation coefficients are used to convolute feature groups. Each feature group represents an individual metabolite.

# File conversion

MetExtract requires the LC-HRMS measurement files to be available as MzXML (TODO: cite Pedrioli) files. Many instrument providers do not support to store the acquired information in the MzXML format. Open source software can be used for the conversion of vendor proprietary data format to MzXML. It is recommended to use the open source software tools ReAdW TODO cite or msConvert from the ProteoWizard package TODO cite.

## ReAdW

ReAdW, a command line tool, is capable of converting Thermo Fischer Raw files (Orbitrap format) into MzXML. It can be downloaded from <http://tools.proteomecenter.org/wiki/index.php?title=Software:ReAdW>.

With the option –c the program converts not only the files but also calculates centroid data from profile mode data using the Thermo Fischer centroidation algorithm.

## Proteo Wizard

msConvert is a part of the ProteoWizard software package and has a graphical user interface, multi core support, and is capable of importing several MS vendor data formats. It also exports to several open data format, most important MzXML and MzML.

It can be downloaded from <http://proteowizard.sourceforge.net/>.

Before the conversion is started, please make sure to add “peakPicking” filter (options 1-), select the “mzXML” output format and discard the “scanHeader” filter. In case MSMS experiments are also recorded in the respective measurement files, please do not use the “msLevel” filter. Compression of mzXML files is not supported by MetExtract as of yet.

# MzXML viewer

To view already converted MzXML files (but also to view result MzXML files), the authors recommend TOPPView from the OpenMS package (Sturm *et al.* (2008)). The viewer allows easy navigation among the files and also provides an easy system to evaluate *m/z* value differences.

# R configuration

Please use R version 2.15.2 (<http://cran.at.r-project.org/bin/windows/base/old/2.15.2/R-2.15.2-win.exe>).

During the setup make sure to select the option “Save version to registry”. For most Windows based operating systems that configuration should work. In case it does not work, please set an environment variable named “R\_HOME” to your R installation folder

# Program start

Start the program with main.exe. Please ensure, that the Microsoft C++ Redistributables 2008 are installed (<http://www.microsoft.com/en-us/download/details.aspx?id=29>)

Once the program has started, it tries to load the default settings stored in the file <PathToExecuteable>/defaultSettings.ini. These settings may be overwritten by the user and may also be specified only partly.

The graphical user interface of MetExtract is divided into the following three sections. For each section separate tab page exists:

* Input tab page  
  In this tab the user can specify measurement groups and the corresponding LC-HRMS measurement files to be processed
* Calculate tab page  
  In the tab calculate all labelling-specific settings and instrument related settings can be specified and data processing is also started from this tab
* Results tab page  
  The Results tab provides a graphical user interface to easily review the extracted information from the input files and also allow their manipulation. At the moment only results from individual files can be viewed and manipulated.

# Input tab page

MetExtract is designed to work with full scan LC-HRMS measurement files. A file is not loaded directly into the program but rather into a group. Each specified group can, for example represent one state of the experiment (e.g. wild type or mutant). Groups can have one or several measurement files.

After any group operation, MetExtract will try to load the input files in order to check if they are valid (see section “File validation”).

## Define group of measurement files

To define a new group click on “Add group”. A new window will appear. In this window the group name has to be specified. With a click on “Add file(s)” the user can select the MzXML files for this group. Files already loaded can be selected and removed with the “Remove selected” button.

During the process of defining a data file group the user can specify how often a particular feature (after having been found at least 1x by MetExtract) has to be found within this group to be accepted for further data processing (Option “Minimum found”). If “omit features” is checked, all features not fulfilling the “Minimum found” option will be rejected for the final data matrix. Therefore, specifying the “Minimum found” option is only necessary if “omit features” is used for a group.

When the user closes this dialog window the newly defined group is displayed in the “Groups” list of the input tab.

## Inspect or Edit group

To edit a group or see which measurement files are contained in the group of interest double click on its name in the “Groups” list in the input tab. Files can be added or deleted from the group. Additionally, Parameter settings such as for “minimum found” (see above) and “Omit features” can be adjusted.

## Delete group

To delete an already defined group, select it and click on “Delete group”. Note that only the specified group will be removed. The measurement files will not be changed with that action.

## Large/Many groups

If an experiment consists of large groups (10 or more measurement files loaded) and/or many different groups (10 or more groups), the file validation step after each group manipulation (creation, deletion or modification of a group) will last for several seconds or even longer. To inactivate this operation step after every group manipulation the “Don’t update” box can be checked. MetExtract will then not update and validate data files within all groups every time. By using the Button “Update” the user can invoke the files verification.

## File validation

The file validation is a functionality, which has to be used before data processing is started, to check if all specified input files are valid according to the mzXML schemata and can be read by MetExtract. If one or several files are incorrectly formatted or cannot be read due to any other error, MetExtract will display an error to the use. Additionally, since MetExtract is designed to work only with one MS scan type (i.e. termed scan event in case of a Thermo instrument) at a time, all scan event types which are available are separately extracted from the input files and displayed under “Scan event”. The user has to choose the appropriate scan event from the list. Only if all measurement files in the different groups share a common scan event, the data processing is possible.

NOTE: MetExtract is not designed for the processing of LC-MS/MS data. However, files containing full scan and MS/MS data can be loaded but only the full scan data events will be available for selection. To perform SIL assisted MS/MS fragmentation experiments please refer to <TODO: Reference Liz Paper>.

## Select scan event

MetExtract is designed to work with only one scan event at a time or with polarity switching. If several different scan events have been recorded, analyse them in sequential order. Only those file can be processed in parallel, which share the same scan events. MSMS or tandem-MS experiments are not supported by MetExtract. Please use FragXtract for this purpose. TODO cite

Polarity switching ionisation is supported by MetExtract. For this, select the same scan event (m/z range, resolution) once in the positive and once in the negative ionisation mode. MetExtract will analyse the two modes independently and merge the individual feature pairs from both modes then.

Do not analyse measurements with polarity switching and such with just one polarity mode concurrently. The polarity switching will halve the total number of scan events per ionisation mode, therefore the number of scans per chromatographic peak will not correspond to the number of scan events per chromatographic peak for measurements with just a single ionisation event. Such measurements need to be processed sequentially.

# Calculate tab page

The calculate tab page lets the user specify all parameters required for data extraction, feature pair comparison among files and optional alignment. The calculate tab page is divided in the following three sections which also correspond to the general workflow of the data processing pipeline:

* Individual files processing  
  This part specifies the settings for MS signal and feature extraction on a per file basis. The listed parameter options do not consider comparison of feature pairs among different measurement files is performed in that section
* Multiple files annotation  
  With these parameter options, feature pairs extracted according to the individual measurement files processing settings will be matched to their pendants in the other processed data files
* Run tasks  
  In this part of the calculate tab page the data processing is activated. The data processing will be displayed by a progress bar once the processing cycle has been started

## Individual files processing

To perform feature extraction for each file separately, the checkbox entitled “Individual files processing” has to be checked. Once started, MetExtract will process each file separately and extract all corresponding 12C and U-13C feature pairs from it. Also, all parameter settings used for this feature pair extraction are located in the “Individual files processing” section of the “Calculate tab page”

Note: For each processed file MetExtract will generate a database file named <FileName>.identified.sqlite. This database contains all extracted data from the respective data file and is used in subsequent grouping of several measurement files as well as results visualisation. Additionally, the file is also used when MetExtract is restarted with the already processed measurement file(s).

### Selection of Experiment type

MetExtract is designed to work either with data derived from global *in vivo* labelling experiments (i.e highly stable isotope enriched metabolomes) or with data originating from metabolisation experiments using (mixtures of) labelled xenobiotics. Depending on the experimental setup, metabolisation studies principally allow several xenobiotics to be tracked in the same measurement files.

After having checked the “individual files processing” box on the “calculation” sheet you can choose between two experiment types: “Full metabolome labelling experiment” or “Tracer experiment”. Both experiment types do share most of the parameter setting options such as *m/z* value picking and *m/z* value clustering as well as those for feature extraction and feature grouping.

#### Full metabolome labelling experiment

Uncheck the box “Tracer experiment” if you want to evaluate data files originating from global *in vivo* labelling experiments (i.e. you expect highly stable isotope enriched metabolomes). In this case you will have the option to specify the natural isotope (isotope A) and the enriched isotope (isotope B) of the element used for labelling during sample cultivation.

The most informative labelling element is carbon, but other labelling elements can also be chosen. Labelling with 13C isotopes will result in the typical SIL associated pattern visualised in fig. 1. Other elements (e.g. N, S, P) will result in very different isotopologue patterns (see fig. 2). For this experimental setup a different extraction method may be used which deduces the feature pairs not based on the isotopic patterns of the labelling element but rather on identical natural carbon isotopic distributions. This option is only available if carbon is not used as the labelling element.

Note: For several different labelling elements predefined relative isotope abundances become available when the user specifies the labelling element. Please ensure that these relative isotope abundances match the experimental setup.

Figure 2: Isotopic pattern derived from labelling with an element other than carbon. The blue distribution shows   
the natural isotopic distribution while the red distribution shows the labelled ion species. The green MS signal bars show   
other isotopes of the labelling element and may or may not be present at all.

#### Tracer metabolisation experiment

If your experiment is to investigate the metabolisation of an isotope labelled tracer, select the “Tracer experiment” box. Using the button “Tracer setup” a new window will open where the used tracers and labelling associated xenobiotics can be specified.

##### Define tracer

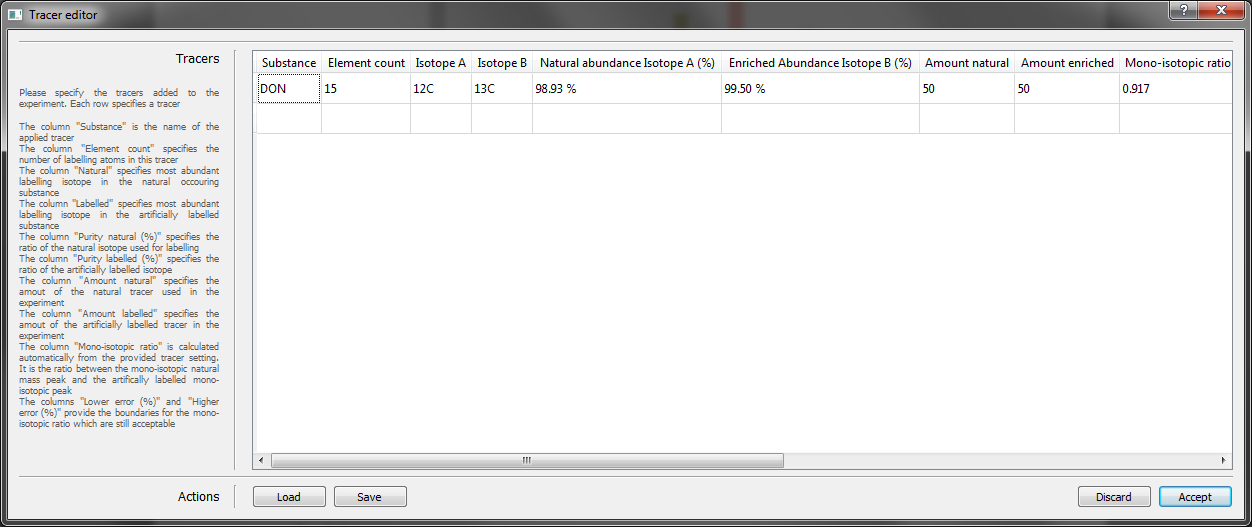


Figure 3: Tracer setup dialog with one tracer configured

One or several tracers can be defined in the Tracer editor dialog (see Figure 2).

Each row in the dialog represents one tracer. For each tracer the following parameters have to be specified:

* Substance  
  The name of the tracer
* Element count  
  The number of atoms of the labelling element per formula unit of the tracer substance
* Isotope A   
  The major natural isotope of the labelling element (e.g., 12C or 14N)
* Isotope B  
  Specify the enriched isotope of the labelling element
* Natural abundance Isotope A (%)  
  Specify the natural abundance of the major natural isotope of the labelling element
* Enriched abundance Isotope B (%)  
  Specify the degree of enrichment of isotope B of the labelling in the precursor (moiety) of interest
* Amount natural  
  The amount of natural tracer relative to the labelled one used in the experiment
* Amount labelled   
  The amount of the artificially enriched tracer relative to the natural one used in the experiment
* Relative isotopic abundance   
  Intensity ratio of the monoisotopic 12C to labelled 13C ion signals of the tracer. Equals the relative amounts of the monoisotopic tracer to the labelled tracer in the mixture used for the metabolisation experiment.   
  This field cannot be specified. Its entry value is calculated automatically from the all previous tracer-specific parameter settings
* Maximum relative negative bias  
  This parameter setting defines the maximum tolerated value of the negative bias of the expected 12C/13C intensity ratio. Thus, with a setting of 0.2, the observed 12C/13C intensity ratio for a putative isotopic pattern has to be less than <relative isotopic abundance>+0.2 (see Figure 3)
* Maximum relative positive bias   
  This parameter setting defines the maximum tolerated relative bias of the expected 12C/13C intensity ratio. Thus, with a setting of 0.2, the observed 12C/13C intensity ratio for a putative isotopic pattern has to be at least <relative isotopic abundance>-0.2 (see Figure 3)

m/z

Allowed ratio  
(e.g. 1:1)

Figure 4: Illustration of a configured tracer. The ratio of the concentrations of the natural and labelled tracer   
compound used for the metabolisation experiment was 1 (1:1; identical signal heights of the monoisotopic (orange)   
and corresponding labelled (blue) ion species signal). Therefore the 12C/13C Intensity ratio is 1 (red line). The two green   
lines indicate the limits of the acceptable intensity interval around the 12C/13C Intensity ratio.

##### use of Several tracers in one experiment

With MetExtract it is possible to evaluate data derived from experiments which use multiple tracers (i.e. precursors to be metabolised). However, the following prerequisites have to be fulfilled:

* The expected 12C/13C monoisotopic intensity ratios of the different tracers must not be identical  
  If several tracers are used concurrently, it has to be ensured that their expected 12C/13C intensity ratios significantly differ from one another. E.g.: Use a 1+1 mixture of the non-labelled and labelled analogue for the first tracer (expected 12C/13C intensity ratio is 1), for the second use e.g. 1.7 (1.7:1), for the third use 2.4 (2.4:1) and so on.  
  Note: always use less from the more expensive tracer – in most cases this is the labelled one
* The expected 12C/13C intensity ratio values ± tolerated bias interval must not overlap.  
  In the above mentioned example the error width should be chosen with 0.3 for both the maximum tolerated positive and negative bias. This would correspond to tolerated 12C/13C intensity ratio intervals of 0.7-1.3 (1±0.3), 1.4-2.0 (1.7±0.3) and 2.1-2.7 (2.4±0.3) parameter settings.

#### Pre-Selection of heavy isotopes X

MetExtract requires specifying a range of labelling elements/isotopes that are expected to be relevant for the respective experiment.

For full metabolome labelling experiments the number can be arbitrary. MetExtract is designed to work with ion pairs for which the two monoisotopic MS signals represent the principal ions of the corresponding ion patterns. Other isotopic distributions may not be identified correctly and the number of labelling atoms may be incorrect.

For tracer metabolisation experiments the maximum number of labelling elements can roughly be chosen as twice the number of labelling elements for the tracer with the highest number of labelling atoms. (e.g. DON has 15 carbon atoms. Therefore, the search for metabolisation products could for example be restricted to a maximum of 32 carbon atoms)

### Parameter settings for MS signal processing

These parameter settings are located in the section MZ picking. These settings are related to the mass accuracy and expected isotopologue ratio accuracy of the used mass spectrometer. These settings are required for the untargeted feature extraction.

The time interval of the chromatogram, in which MetExtract inspects mass spectra and EIC chromatograms is specified with the input fields “Start (min)” and “End (min)”. Only MS spectra and EIC peaks between these time points of the LC-HRMS run are considered for data processing.

An optional intensity threshold can be specified in the input box “Intensity threshold”. All centroided MS signals above this intensity threshold are assumed to be putative signals of monoisotopic ions with natural isotopic composition. The MS signal of the corresponding fully labelled ion may be of lower height than the provided threshold level.

The input field “Max. number of charges” specifies the max. tolerated charge state of an ion to be considered for data processing by MetExtract.

The input box “Max. mass deviation (+/- ppm)” specifies the maximum tolerated mass deviation in ppm, relative to the theoretical *m/z* value (measured *m/z* value of “starting signal” plus n-times 1.00335 u for singly charged ions) of a potential corresponding fully 13C labelled isotopologue. Figure 4 provides a graphical illustration of the tolerated *m/z* interval around the monoisotopic 13C ion signal.



Figure 5: This diagram illustrates the tolerated *m/z* interval resulting from the specification of   
the parameter “Max. mass deviation (+/- ppm)”. The orange leftmost MS ion signal illustrates the monoisotopic 12C peak.   
From the measured *m/z* value of this ion and an assumed number of 34 carbon atoms, the *m/z* value of the   
fully 13C labelled MS signal is calculated. All MS signals within the m/z range   
<calculated m/z>+/-<Max. mass deviation (+/- ppm)> are accepted.

The parameters “Isotopic pattern count (A)” and “Isotopic pattern count (B)” specify how many isotopologues of the natural (A) and the fully 13C labelled isotopologue (B) must be present in the respective isotopic patterns. If “1” is selected, MetExtract will only check for and use the two corresponding monoisotopic MS signals for confirmation. If the parameter is set to 2, MetExtract will search for the monoisotopic signal pair plus the corresponding first isotopic MS signals (M+1.00335 for the non-labelled monoisotopic 12C and M-1.00335 for the fully 13C labelled ion).

When the check-box “Consider isotopologue abundance” is marked, MetExtract evaluates the observed M / M+1 and M` / M`-1 signal intensity ratios and compares them to the theoretically expected ratios (calculated from the number of carbon atoms of that particular MS ion pair and the relative abundance of the respective isotopes).

The parameters “Intensity abundance error” for “Non-labelled ion” and “Labelled ion” specify the maximium relative deviation between the expected and the observed ion intensities of M+1 (in case of the monoisotopic 12C) and M`-1 (for the fully labelled 13C ion). A parameter setting of e.g. 0.2 corresponds to a maximum tolerated relative isotopic abundance error of 20% compared to that theoretically expected.

*m/z* clustering:

Since MetExtract initially finds MS signals by inspecting individual mass spectra, it generates a list with all corresponding *m/z* value pairs and corresponding number of carbon atoms, which fulfil all pre-defined parameter settings. Each of these ion pairs refers to a single mass spectrum in which it was detected. Thus, it is necessary to link MS signal pairs from different MS scans, which show similar m/z values and represent a single metabolite ion species, together. To this end, hierarchical clustering is applied to the *m/z* values. The parameter “Clustering ppm” is used to combine the generated *m/z* signal pairs into different bins. It specifies the mass difference in ppm between two particular *m/z* signals, which is max. tolerated to cluster these ions into the same bin. For a mass difference larger than specified, m/z signals are split into different *m/z* bins. In general this parameter should be a multiple (2 or 2.5) of the parameter “Mass deviation (+/- ppm)” specified above. Additionally, only those *m/z* bins are kept which are found in at least “Min. spectra” MS scans.

### Chromatographic Peak processing

The box “Correct element count” has no functionality yet.

#### Peak Picking [MassSpecWavelet]



Figure 6: Illustration of the chromatographic peak picking process . Within a specified mass tolerance window extracted ion current (EIC) chromatograms of both the natural (e.g. 12C) and its corresponding uniformly labelled (e.g. 13C) MS signals are extracted and inspected for chromatographic peaks (For the example above, in both EICs two chromatographic peaks are found). The Pearson correlation coefficient is then calculated for the intensity values of the corresponding chromatographic peaks and used to verify the presence of co-eluting and similarly shaped EIC peaks. In the illustrated example the two highest peaks have nearly identical profiles. For the other two peaks the EIC chromatograms show only one of the two corresponding (non- and uniformly labelled) ions and are therefore discarded.

As a next step in the data processing workflow, MetExtract extracts chromatographic peaks from the data. For this, each *m/z* bin is processed with the following algorithm:

1. EIC chromatograms of the detected MS signal pairs (monoisotopic non- and uniformly labelled ions) are extracted from the data binwise in a predefined *m/z* interval around the previously deduced *m/z* values using the processing parameter “EIC ppm”.
2. The MassSpecWavelet algorithm described by Du *et al.* (2006) is used for chromatographic peak picking from EIC chromatograms of both non-labelled and the corresponding uniformly labelled *m/z* bins. For this purpose, the setting for peak width is systematically varied to find the wavelet which is best fitting the measured chromatographic peak in the EIC chromatogram. The respective peak width is expressed in scales and specified under “Min scale” and “Max scale” settings. As a consequence the software tries to find best fitting wavelets within the specified range of scale values.
3. A further peak picking criterion is the Signal to noise ratio (SNR). Only peaks with SNR above the preset “SNR threshold” are considered.

Different isotopologues (12C/13C or 14N/15N) of the same metabolite cannot be separated by LC but can easily be distinguished by mass spectrometry. Due to perfect chromatographic co-elution all isotopologues (including labelled and non-labelled ion species) of a particular metabolite can be expected to show identical retention times and chromatographic peak shapes. This verification of co-elution and similar peak shape of corresponding features is performed by the software. The processing parameter “Center error” denotes the maximal tolerated difference (in number of MS scans) between the chromatographic peak centers for the corresponding monoisotopic and fully labelled analogues. The parameter “Min. corr” specifies the minimal value of the Pearson correlation coefficient which must be reached for the inspected peak pair to be accepted as corresponding isotopologues of the same metabolite. See Figure 5 for an illustration.

Note: Each *m/z* bin pair showing two identical chromatographic profiles and fulfilling all previous parameter settings is considered a unique feature pair. Since the same *m/z* bin pairs may correspond to MS signals from different metabolites (i.e. isomers) potentially being separated during the chromatographic step of the LC-HRMS measurement the MS signals of a particular pair of corresponding m/z bins may end up as several distinct feature pairs.

#### Hetero Atom annotation

After feature pairs have been extracted from the data, the software is capable of searching for isotopologue mass peaks originating from other elements than the one used for labelling. Such hetero atoms must consist of at least two different isotopes and their relative intensities should be clearly observable. Chloride, for example, consists of two isotopes namely 35Cl and 37Cl. 35Cl, with an relative abundance of 75%, is clearly the more abundant isotope of chloride and therefore expected to be predominantly incorporated into molecules. However, since the less frequent 37Cl has a quite high abundance of 24% it is recorded in most mass spectra for substances with chloride atoms. On the other hand, oxygen shows a different abundance profile for its 3 stable isotopes. The most abundant oxygen isotope is 16O with approximately 100% abundance. The two other isotopes 17O and 18O are only negligibly present in nature and therefore hardly observed in LC-HRMS data. The following table comprises a short list of the mostly expected hetero atom isotopes observed in metabolomics research:

|  |  |  |
| --- | --- | --- |
| Isotope | Δ m/z to most abundant isotope or respective element | Relative abundance to most abundant isotope of respective element |
| 34S | 1.99579 | 4.43% |
| 37Cl | 1.99705 | 31.98% |
| 41K | 1.99711 | 7.22% |
| 54Fe | -1.99532 | 6.32% |

These hetero isotopologues must not originate from the metabolite itself, but can also originate from common adducts (e.g. [M+K]+, [M+Cl]-).

The search for hetero atom isotopologues is performed after the feature pairs have been extracted. The m/z value M of a feature pair is considered to consist only of the most abundant isotopes of all elements comprising this substance while M’ is considered the same with just one element replaced by the labelling isotope. All scans in the chromatographic region of a feature pair are inspected for mass peaks showing the characteristic m/z offset (restricted by the parameter “Mass deviation”) and relative abundance (parameter “Intensity error”) for the respective isotope. If such a mass peak has been found in at least a predefined number of scans belonging to the respective feature pair, the pair is annotated with the stable isotope and therefore the hetero atom. As more than one hetero atom of the same element may be present in a certain metabolite, the algorithm only reports the most likely number of hetero atoms which is calculated by comparing the observed relative abundance with the theoretical relative abundance of this respective hetero isotopologue.

For isotopes showing a positive m/z offset, the search is performed at M’+isotope\_m/z while for isotopes with a negative m/z offset (e.g. 54Fe) the search for respective mass peaks is performed at M-isotope\_m/z

Several hetero isotopes show similar m/z value offsets and similar relative isotopologue abundances. Depending on the HRMS device, these may or may not be separated. The algorithm will report all possible hetero isotopologues.

Hetero atoms can be configured by clicking on the “Hetero atoms configuration” button. The following information is required for every possible hetero isotope:

* Isotope  
  Name of the isotope (e.g. 34S)
* Δ m/z   
  The m/z ratio offset compared to the elements most abundant isotope if the charge is 1
* Rel. ab [%]  
  The relative abundance of the isotope to the elements most abundant isotope. (e.g. The natural abundances of chloride are 76% for 35Cl and 24% for 37Cl. The ratio for the 37Cl entry is therefore   
  24 % / 76 % = 0.32 ≙ 32 %)
* Min. atoms  
  The minimal number of hetero isotopes to be searched for
* Max. atoms  
  The maximum number of hetero isotopes to be searched for

#### Feature grouping

Ionisation by electrospray may give rise to several ion species for the same substance (e.g. adducts, in-source fragments or dimers). To convolute all ion species of a particular metabolite into a single group, the Pearson correlation coefficient is again used to find features with similar chromatographic peak shapes and retention times. If two feature pairs elute at approximately the same retention time (parameter “Center error”) and the correlation coefficient of their peak shape is higher than “Min. corr”, the respective features are put into one feature group. TODO use higher intensive feature 12C or 13C

After a feature group has been created, MetExtract tries to annotate the chemical relationship between the features of a group by calculating *m/z* value differences between all features within this particular group. For this purpose, adducts or sum formulas of neutral losses can be specified in a dialog window after clicking on the icon “Relationship configuration”. In the upper part of the dialog window frequently observed / expected adducts can be defined. Each line of the table corresponds to a specific adduct with the following setting options:

* Adduct type  
  The common name of the adduct can be specified (e.g. M+Na, M+H…)
* *m/z* offset  
  Specifies the exact (theoretical) mass shift which is obtained by transforming the neutral molecule M to the respective ion species. E.g. for the protonated molecule the *m/z* offset is 1.00728 (i.e. atomic mass of 1H+)
* Polarity  
  Specifies for which ionisation polarity mode the respective adduct is expected to occur. Only “+” and “-“ for the positive and negative mode are allowed.

If no adduct relationship can be detected between two features of a particular feature group, MetExtract considers the specified neutral losses and tries to generate sum formulas for these losses. The allowed elements for the sum formula generation of the neutral losses can be specified in the second table in the Relationship configuration dialog. Each row of this table represents one element with the following properties:

* Element  
  The chemical symbol of the element
* Atomic mass  
  The mass of the elements most abundant isotope
* Valence electrons  
  The number of valence electrons of the respective element TODO check if that can be used to limit possible fragments
* Min count  
  The minimum number of atoms at which this element must occur per molecular formula unit
* Max count  
  The maximum number of atoms at which this element can occur per molecular formula unit

NOTE: If a tracer experiment was performed it is possible that conjugates added during the metabolisation may have several non-labelled atoms of the labelling element. Therefore, a neutral loss may not reduce the number of labelling elements for two features but the neutral loss may still contain the heavier isotope of the labelling element. E.g.: A conjugated metabolite ion is observed having 15 carbon atoms. In the same group the tracer without the conjugated moiety is also observed since moieties (e.g. glucose) may cleave during ionisation. The neutral loss consists then only of the non-labelled conjugate. Thus, the number of labelling atoms for the tracer with the conjugate and only the tracer is identical.

### Save results

Results of the individual files/measurements may be saved in various ways. A tabular format with all features is saved for every measurement in the file <FileName>.tsv

#### Save to PDF

Additionally to the tabular format a PDF file with all features may also be created. Each feature will have a plot of the natural and artificially labelled EIC. For feature groups there will also be a graph where the EICs of all features are scaled and plotted to easily review the grouping process.

#### Save to MZXml

For some MS devices it is possible to save the results to a new MzXML file. MetExtract will try to save the MzXML file but for some devices (e.g.: particularly ones, for which the developers did not have access to the devices) this may fail. In that case, an error will be generated and the MzXML file will not be written.

## Multiple files annotation

### Group results

#### Align chromatograms

### Integrate missing peaks

## Run Tasks

To start the data processing, press the “Start” button. A dialog will appear displaying the overall status and the current operation per file

### Multi core systems

Since many operations of MetExtract can be parallelized, multi core systems are fully supported. The default number of cores MetExtract will use is one less than available on the executing machine to not block the computer for office work. To use all available cores, the box “Keep one core unused” must be unchecked. Additionally, it is not possible to use more CPU cores than available since that only results in threading overhead from the operating system but no gained speed in data processing.

# Results tab page

Results of data processing can be visualized in the third tab named “Results”. The leftmost list (“Processed input files”) holds all processed measurement files. The results of the selected file will be presented in the centred tab (“Extracted information”). The results of each file are divided into the following four categories:

* MZs
* MZ bins
* Features
* Feature groups

## External viewer

To view the measurement MzXML file in an external viewer use the button “open MzXML file externally”.

## *m/z* Plot

The tab “*m/z*s” shows all mass signals from all scans of the original data file that showed the mirror symmetric isotopic patterns and fulfilled all parameter settings for this MS signal processing. The only plot available for this result category is the main plot. It shows a scatter plot of all 12C features found before clustering of those features into *m/z* clusters. If a specific ion signal is selected, it will be coloured red.

Each *m/z* holds the following information:

* *m/z*  
  The 12C feature *m/z* of the extracted ion pair
* xCount  
  The number of labelling atoms deduced for this ion pair
* Scan id  
  The ascending scan number in which this ion pair was found
* Scan time (min)  
  The corresponding retention time of this particular scan
* Charge  
  The ion pairs charge state (determined based on the mass difference between the mono-isotopic and its first carbon isotopologue)
* Intensity  
  The intensity value (ion abundance) of the natural ion in the ion pair

## *m/z* Bin Plot

Individually extracted ion pairs are subsequently clustered into bins of minimal ppm difference. These bins are listed under “*m/z* bins”. Only the main plot is available for a graphical illustration of these bins. A selected bin will be colour highlighted among all extracted bins.

Each *m/z* bin holds the following information:

* *m/z*  
  The average *m/z* ratio of all participating 12C mono-isotopic features
* Delta ppm  
  The ppm difference between the highest and lowest ion pair within the bin
* xCount  
  The number of labelling atoms determined for this bin

## Features

After clustering found ion pairs in *m/z* bins, each bin is inspected for chromatographic peaks. Each bin may have several chromatographic peaks and hence eventuates in several features. A feature is a chromatographic peak having a certain *m/z* ratio and an assigned number of labelling atoms as well as a charge state derived from the ion pairs. For a feature two plots are available. The main plot shows the chromatographic peaks of the natural and artificially enriched isotopologues. The intensities of the labelled chromatographic peaks have negative intensities to better see the natural and labelled peak at once. The second plot for each feature is the mass spectrum (see Figure 6). It shows the scan in the features centre and shows which mass peaks were used.

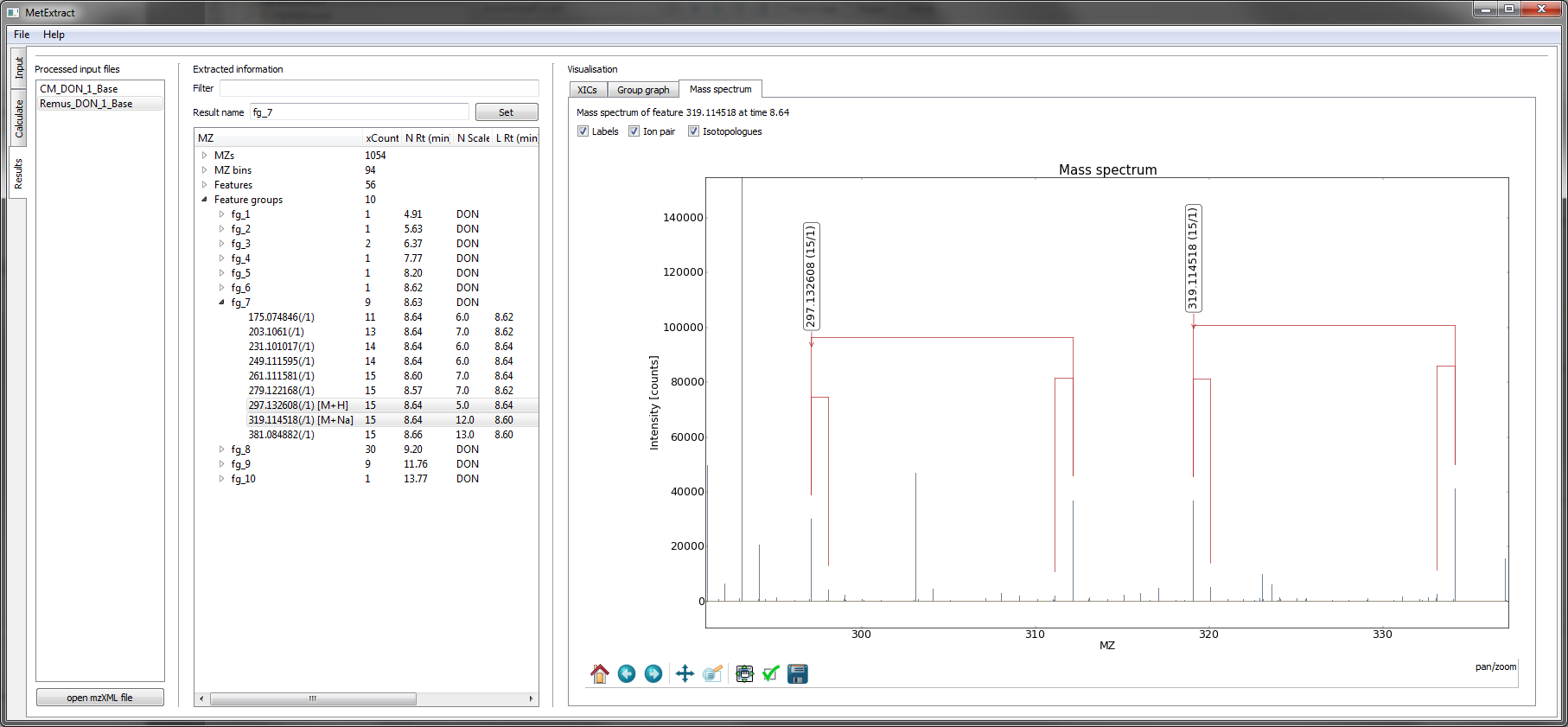


Figure 7: A sample mass spectrum of two features in the same feature group. Both have 15 carbon atoms. The   
balloon indicates the start of the ion pair (natural monoisotopic peak). The artificially enriched isotopologue   
is show on the leftmost end of the red horizontal line. The two first isotopologues are also marked.

Each feature holds the following information:

* *m/z*  
  The average *m/z* ratio derived from extracted ion pairs for this features
* xCount  
  The number of labelling atoms derived from the ion pairs
* N Rt (min)  
  The retention time of the natural chromatographic peak
* N Scale  
  The chromatographic peak width in scales (see Du *et al.* (2006))
* L Rt (min)  
  The retention time of the artificially enriched chromatographic peak
* L Scale  
  The chromatographic peak width in scales
* Corr  
  The Pearson correlation between the natural and artificially labelled chromatographic peak
* Tracer  
  The tracer to which this feature was assigned.

## Feature groups

Feature groups are convolutions of different features most likely originating from the same substance in the sample. In the section Feature groups all groups are listed. Three plots are available for a feature group:

* Main plot  
  This plot shows the chromatographic peaks of all features within a group plotted one above the other
* Group graph  
  This plot shows which features showed high enough correlations to be within one group. Isolated features or grouped islands may indicate false grouping. Thus the graph is an easy method to verify the group and find putative false positive assignments. See Figure 7 for an example
* Mass spectrum  
  This plot shows the mass spectrum in the average centre of the feature group. All ion pairs will be plotted

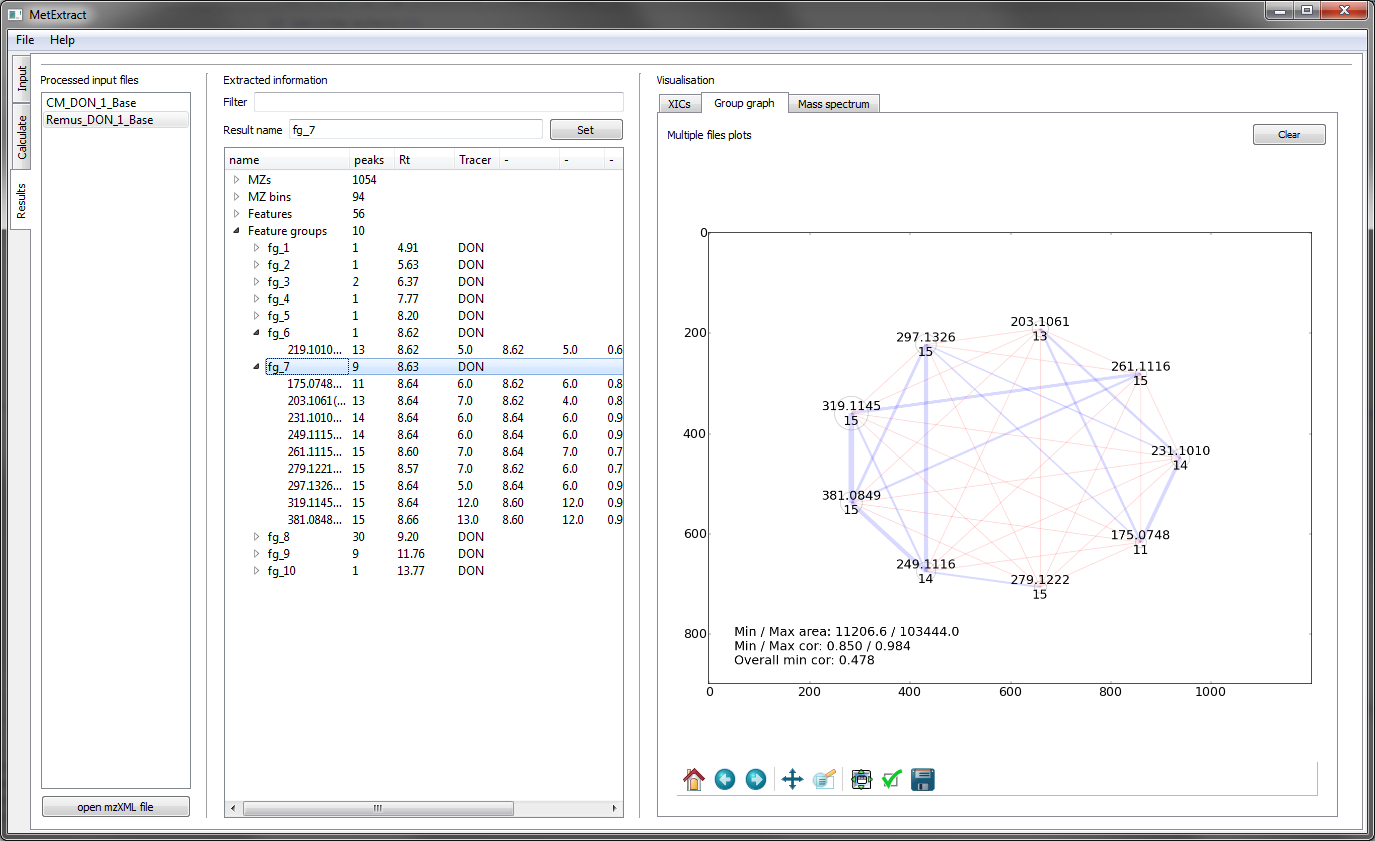


Figure 8: Group graph of a DON. The blue lines (edges) show a correlation between two features (nodes) above the set   
threshold while red edges indicate a lower correlation. This example is not very well connected. The reason is a close by   
co-elution which deranges the chromatographic peaks. A well connected feature group would have no or few red edges.   
The circles around the nodes indicate the features intensity. The bigger such a circle is, the higher the intensity of the   
feature is. In this example the node with *m/z* 319.1145 and 15 carbon atoms has the highest intensity

## Filters

MetExtract offers the possibility to filter all results for certain criteria. The filter can be entered in the text field “Filter” in the “Results” tab. It is applied to each result type. The entered text is first split into chunks using the space character. Each i-th chunk is a filter for the i-th column in the results tab.

EXAMPLE: If the user wants to search for features with a *m/z* ratio of 297.1333, the text “297.1333” is entered in the field “Filter”. MetExtract will then only show results having the text 297.1333 in the first column. If the search shall be further restricted to 15 carbon atoms, the user must enter “297.1333 15”. This will also search in the second column for the text 15 and only show those features, which have 15 or 115 or 1500 carbon atoms.

The direct text search is not very well since the *m/z* ratios may deviate from run to run. It is better to use the “+­-" notation. This allows specifying a base *m/z* value with boundaries.

EXAMPLE: If the user wants to search for 297.1333 +- 1 amu the filter must be set to “297.1333+-1”. This will instruct MetExtract to search for *m/z* values within the range of 296.1333-298.1333 amu. It is also possible to search within a specific ppm window. The filter must then be set to “297.1333+-5ppm”.