MS-DIAL tutorial:

Universal program for untargeted metabolomics

14/04/2020

Justine Chervin and Guillaume Marti

justine.chervin@lrsv.ups-tlse.fr

guillaume.marti@univ-tlse3.fr

https://www.lrsv.ups-tlse.fr/metatoul-en/

1. Prerequisite:

Software installation:

MS-DIAL last version: http://prime.psc.riken.jp/compms/msdial/main.html

Abf (Analysis Base File) Converter: https://www.reifycs.com/AbfConverter/

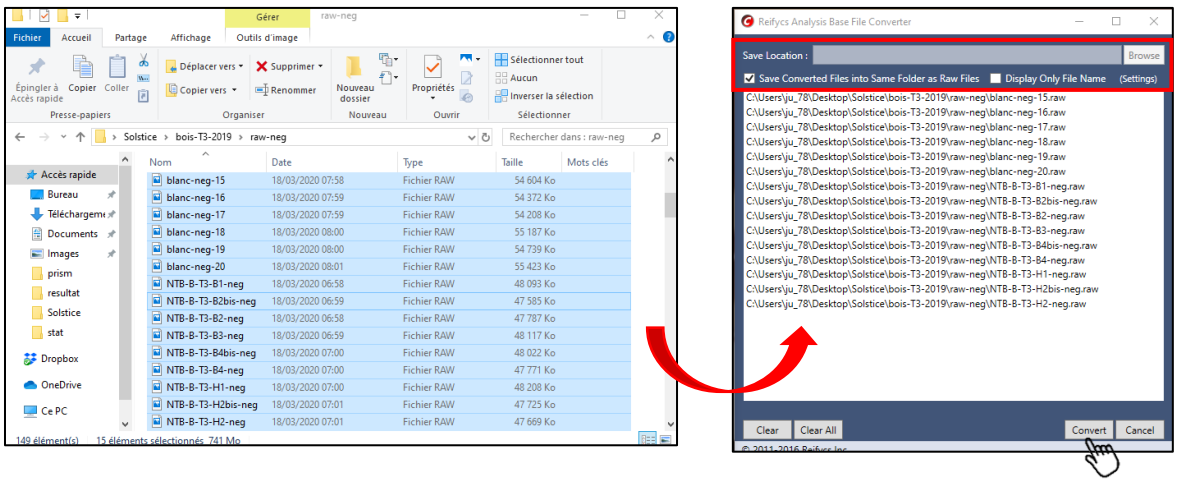
MSFileReader for Thermo data: see the following link which explains how to process http://fields.scripps.edu/rawconv/

1. File conversion:

a. Start the application AnalysisBaseFilConverter
b. **Drag and drop** .raw files into the application
c. Choose the folder of arrivals (cf. red rectangle)

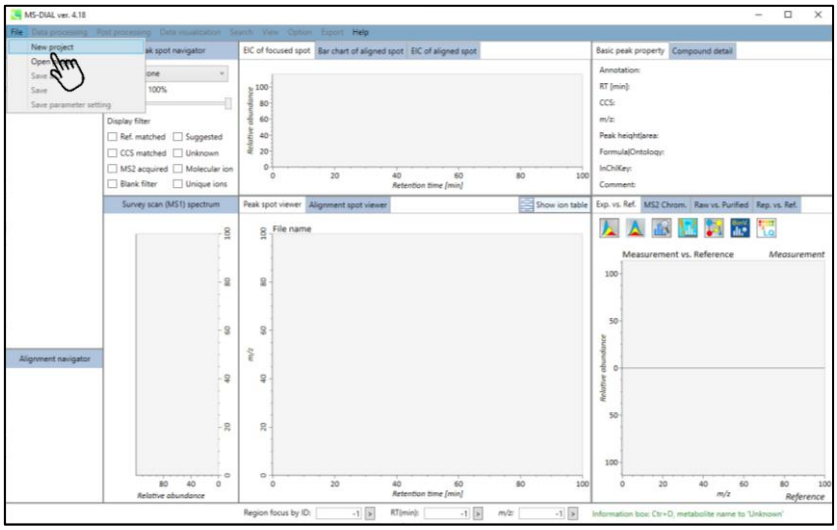
By default, the .abf files will be created in the same folder as raw data since « **Save Converted Files into the Same Folder as Raw Files** » is selected. If you want to generate them into another folder, uncheck this and select the new folder by clicking « **Browse** » on the top right.

d. Click on « **Convert** »

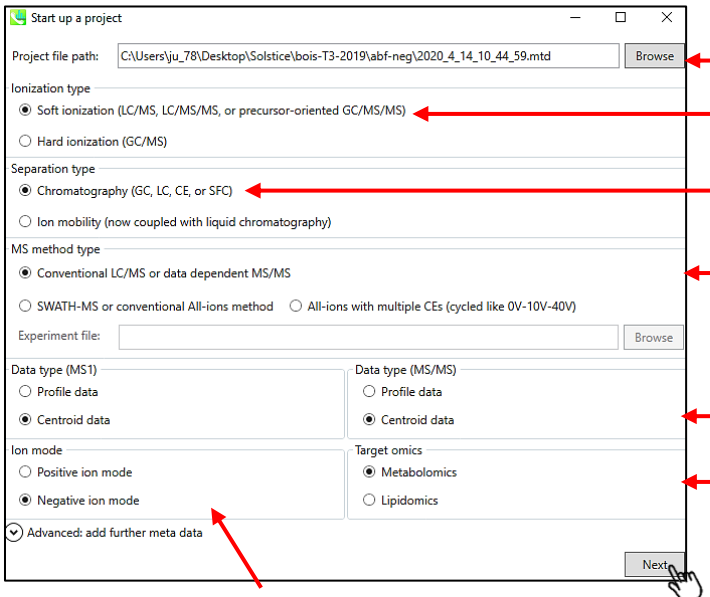


2. Create a project on MS-DIAL and Import data

a. Open the application MS-DIAL and on the « **File** » menu select « **New project** »



The following window appears to define the type of experiments you are about to process:



Select the folder containing the .abf files

Select « Soft ionisation » for ESI experiments
For EI GC/MS select “Hard ionization”

Select the type of chromatography used

Select « Conventional LC/MS » if you want to process LC/MS or LC/MS/MS data in DDA mode
Select SWATCH-MS* if you want to process LC/MS/MS data in DIA mode

Select Profile or Centroid data

Select « Metabolomics » to use your own MSP file database otherwise « Lipidomics » to use the internal lipid database of MS-DIAL

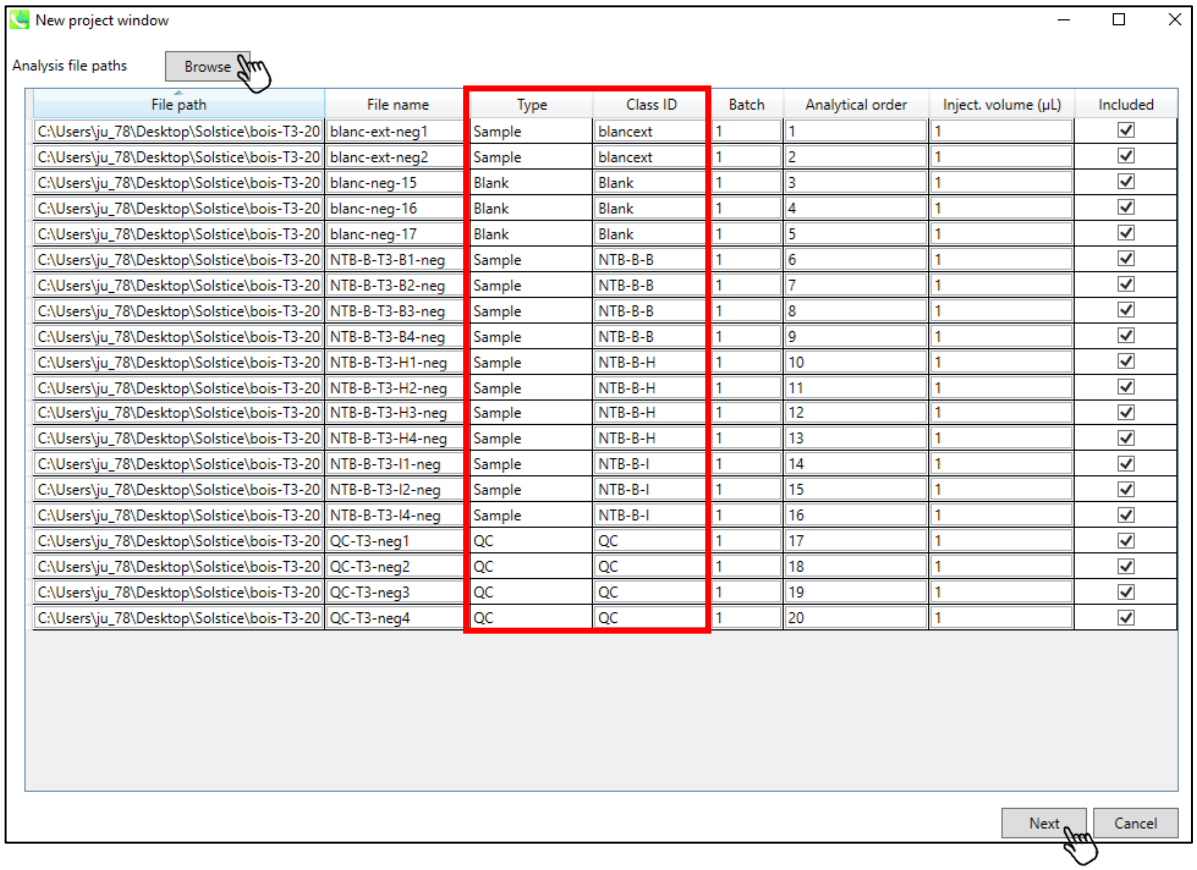
Select the ion mode (you have to process pos and neg separately)

** by selecting SWATCH-MS, you have to indicate a “dictionary file” as tab delimited text file with MS1 scan range and precursor window as follows:*

http://prime.psc.riken.jp/compms/msdial/main.html#Templates

b. Import your data

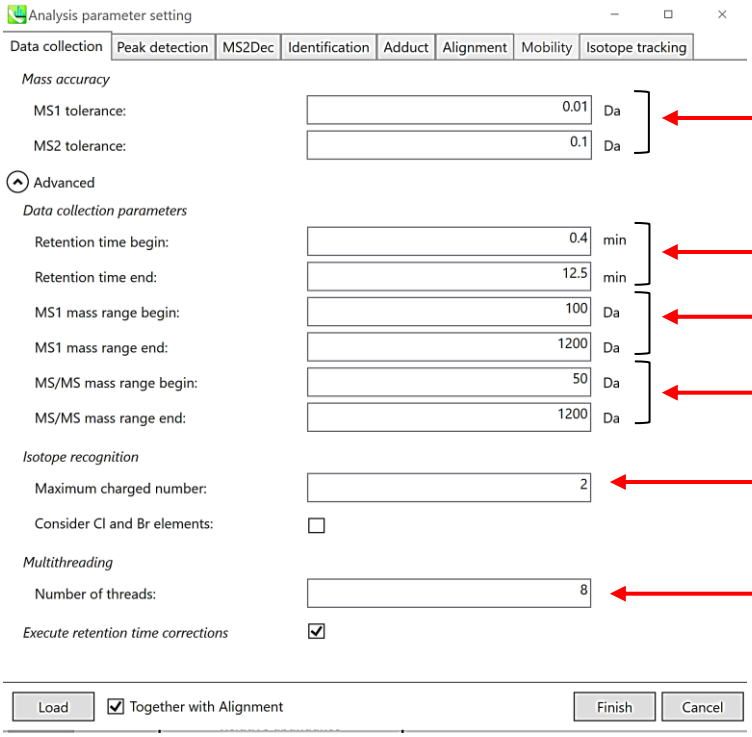
The following window appears and click on “**Browse**” to import the .abf files you want to process. During data importation, it is important to note the type (Blank, QC or Sample) and class of every sample in **Class ID column** (blank, sample class, QC). If both ionization modes have been acquired, be careful to have the **same number of samples** between pos and neg modes and in the **same order.**



File path	File name	Type	Class ID	Batch	Analytical order	Inject. volume (µL)	Included
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	blanc-ext-neg1	Sample	blancext	1	1	1	☑
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	blanc-ext-neg2	Sample	blancext	1	2	1	☑
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	blanc-neg-15	Blank	Blank	1	3	1	☑
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	blanc-neg-16	Blank	Blank	1	4	1	☑
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	blanc-neg-17	Blank	Blank	1	5	1	☑
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	NTB-B-T3-B1-neg	Sample	NTB-B-B	1	6	1	☑
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	NTB-B-T3-B2-neg	Sample	NTB-B-B	1	7	1	☑
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	NTB-B-T3-B3-neg	Sample	NTB-B-B	1	8	1	☑
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	NTB-B-T3-B4-neg	Sample	NTB-B-B	1	9	1	☑
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	NTB-B-T3-H1-neg	Sample	NTB-B-H	1	10	1	☑
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	NTB-B-T3-H2-neg	Sample	NTB-B-H	1	11	1	☑
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	NTB-B-T3-H3-neg	Sample	NTB-B-H	1	12	1	☑
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	NTB-B-T3-H4-neg	Sample	NTB-B-H	1	13	1	☑
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	NTB-B-T3-I1-neg	Sample	NTB-B-I	1	14	1	☑
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	NTB-B-T3-I2-neg	Sample	NTB-B-I	1	15	1	☑
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	NTB-B-T3-I4-neg	Sample	NTB-B-I	1	16	1	☑
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	QC-T3-neg1	QC	QC	1	17	1	☑
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	QC-T3-neg2	QC	QC	1	18	1	☑
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	QC-T3-neg3	QC	QC	1	19	1	☑
C:\Users\ju_78\Desktop\Solstice\bois-T3-20	QC-T3-neg4	QC	QC	1	20	1	☑

3. Data analysis and parameters setting

a. Data collection parameters



Indicate the MS1 and MS2 tolerance Depend of your instrument

Indicate the range of RT

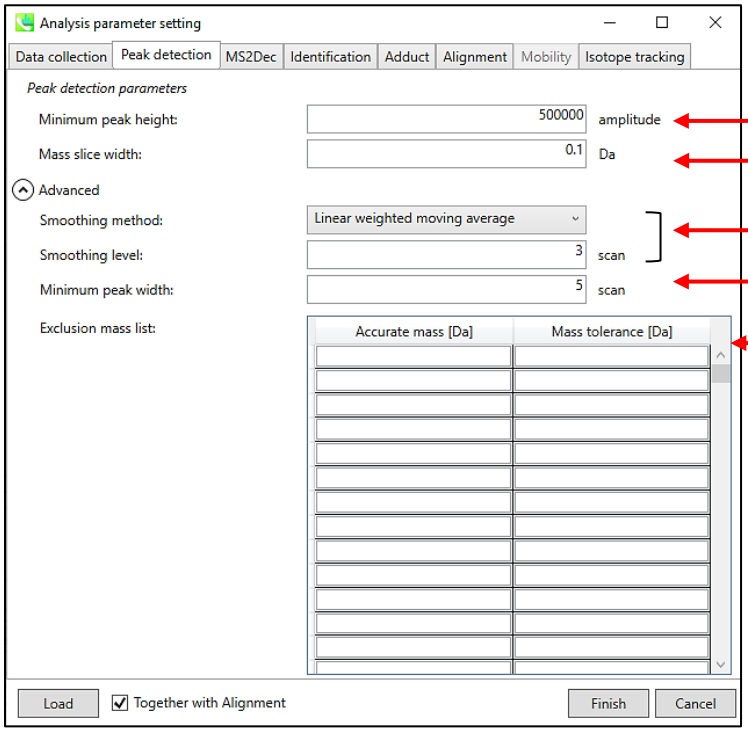
Indicate the range of MS1 mass

Indicate the range of MS2 mass

For small molecules, the maximum charged number can stay at 2 but you can go up to 8 if necessary.

You can fix the count of threads used to process data (see resource monitor in administrative tools (control panel))

b. Peak detection parameters



Fix the peak height threshold*

By default, 0.1 is generally suitable

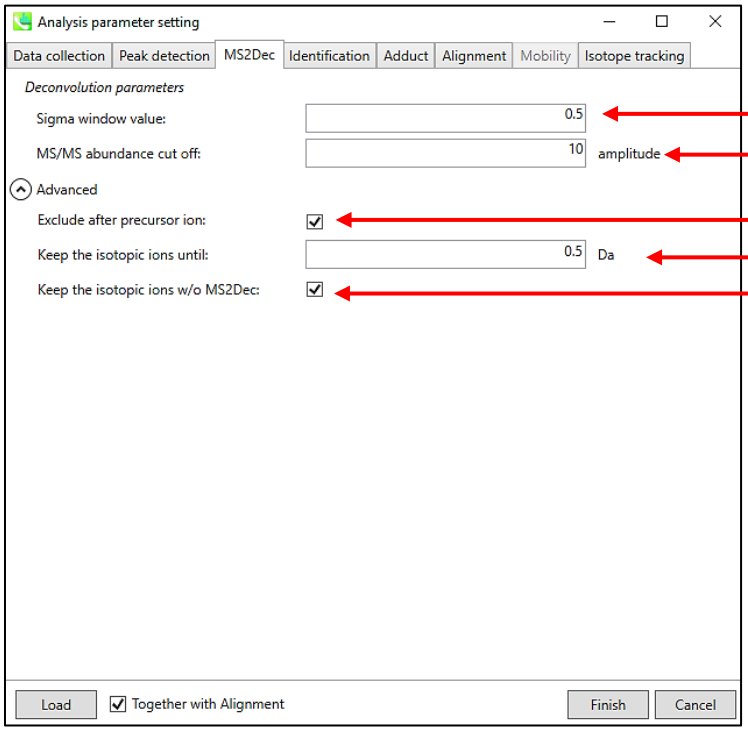
Linear weighted moving average is set by default with recommended smoothing level 1-3

Fix the peak width threshold (by default 5 is ok)

You can specify some unwanted peaks if necessary

****Critical value which greatly influence final results:** *70% under the observed baseline is a good starting point.*

c. MS2Dec parameters



By default, 0.5 is suitable. Under 0.3, you will recognize noise chromatographic peaks. Up to 0.7, you will reduce peak resolution

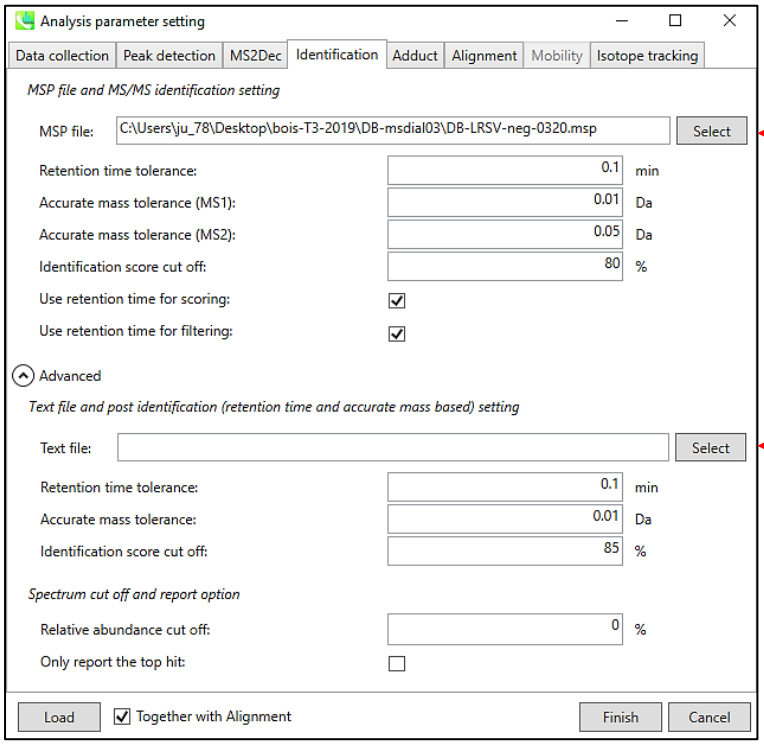
Set a cut off value to reduce the MS/MS noise

Recommended for metabolomics and lipidomics

Set at 0.5 by default.

Select this because the algorithm may erase the precursor's isotopic ions.

d. Identification parameters



You can select a msp format database for identification of level 1. It contains information such as accurate mass, retention time, MS/MS peaks and various identifiers.

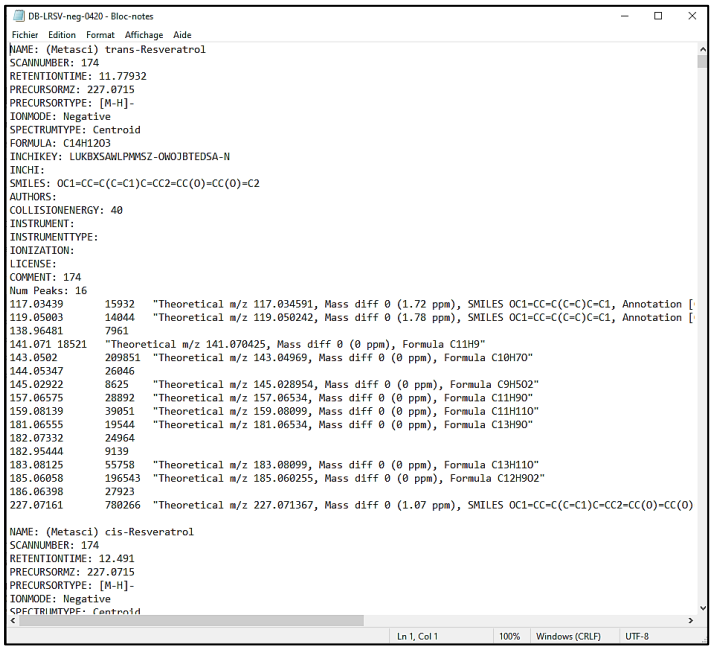
Set the mass and retention time tolerances.
The identification score cut off should be greater than 70%.

You can select a text format database for retention time and accurate mass search (post identification)

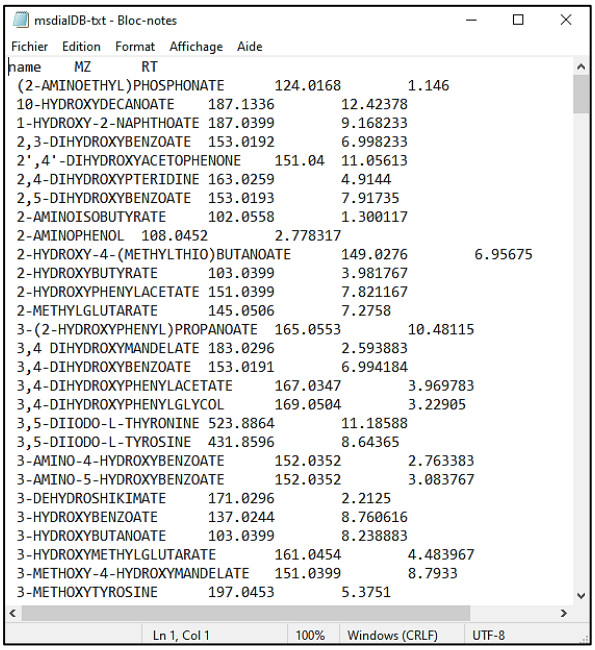
The first two required columns should contain the "metabolite name" and the "accurate mass". You can add additional information such as the retention time in following columns. If so, set a suitable retention time tolerance, otherwise indicate "100".

The identification score cut off should be greater than 70%.

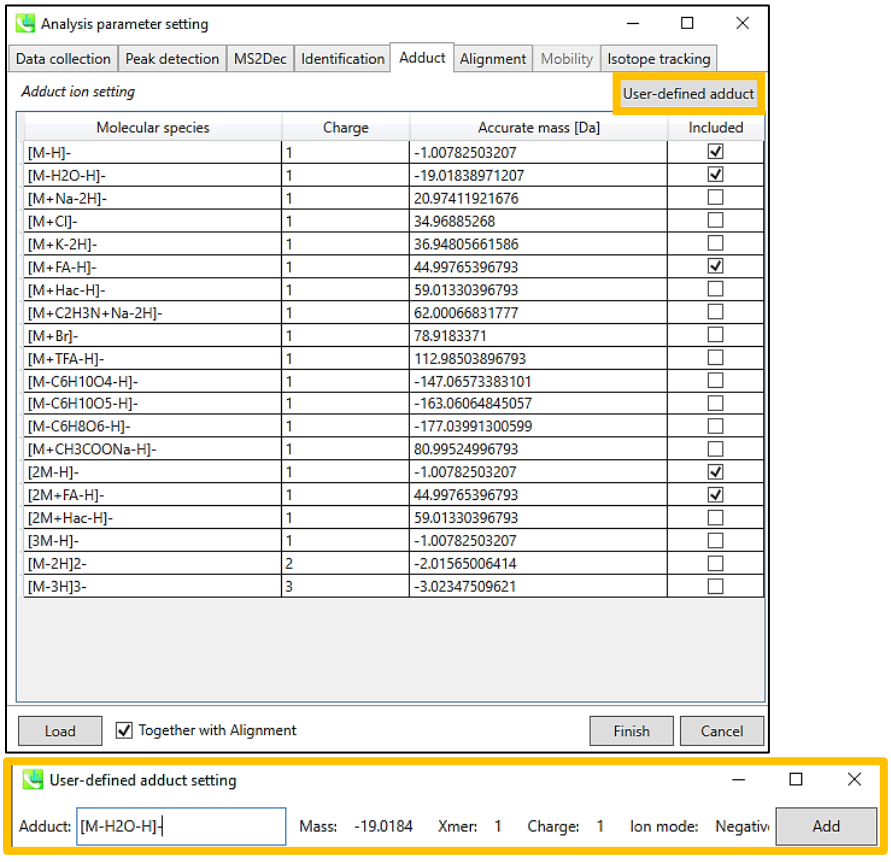
Example of msp format database:



Example of text format database:

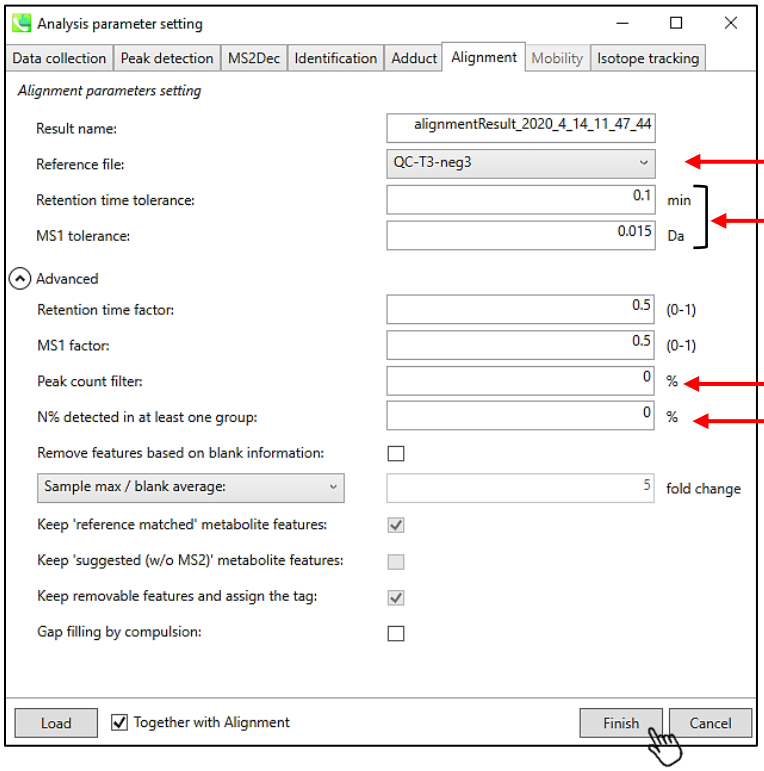


e. Adduct ion setting



You can select the adduct ions suitable. If necessary, you can add your own adduct ion by clicking “**User-defined adduct**”or load a whole adduct list by clicking “**Load**”.

f. Alignment parameters



Specify the QC file. All samples will be aligned on this QC file.

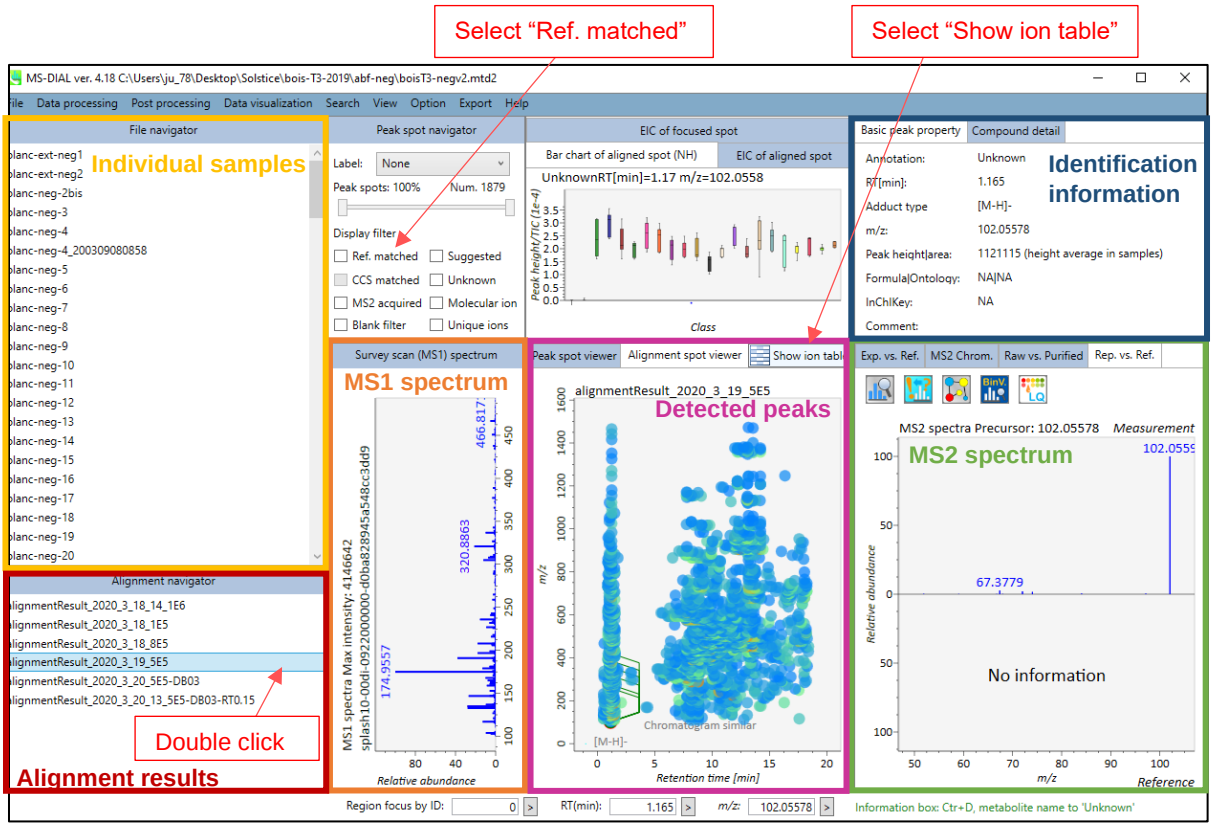
Set the retention time and mass tolerances

To remove peaks with more than N% missing values.

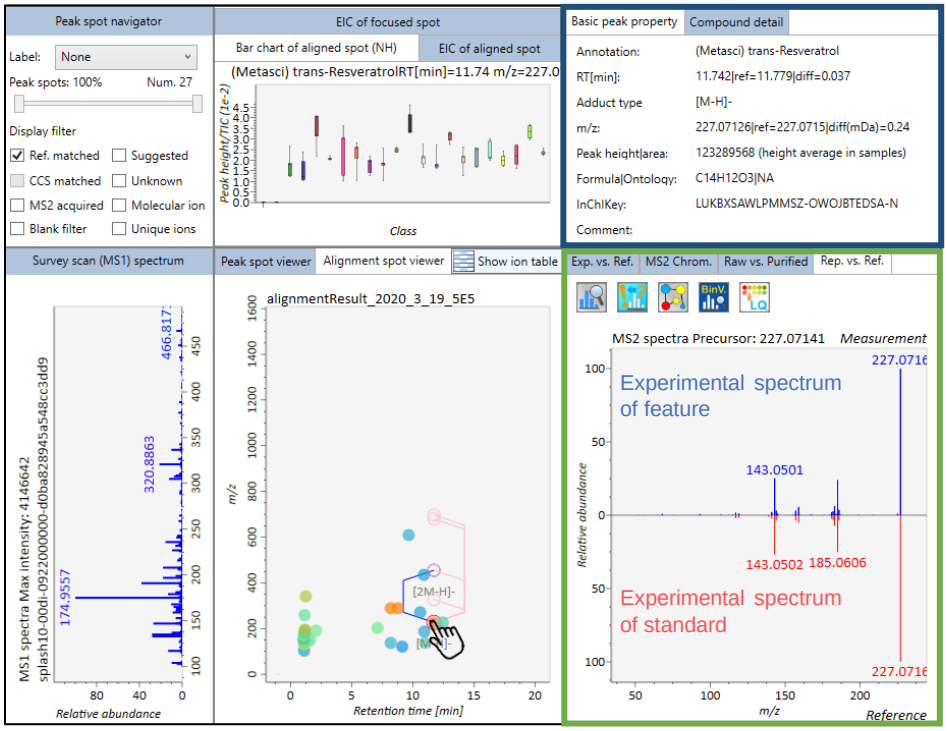
Peaks should be detected in N% of all samples of a class

4. Analysis results

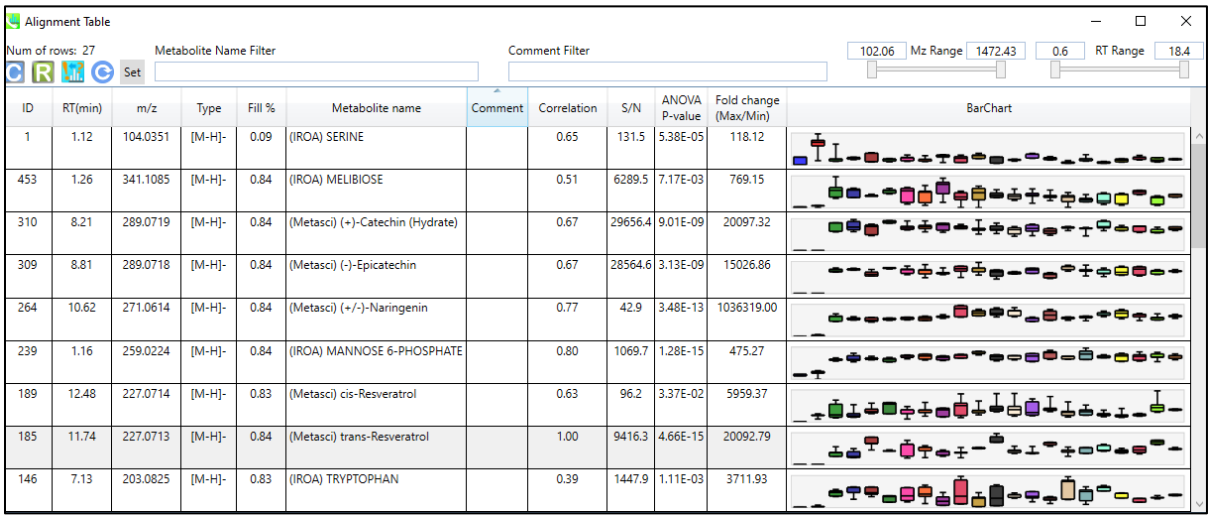
The following window appears after processing:



By clicking on “**Ref. matched**”, you can see the number of matched features based on your imported database. The following example is based on the interrogation of a msp format database based on retention time, accurate mass and MS/MS spectrum for identification of level 1:

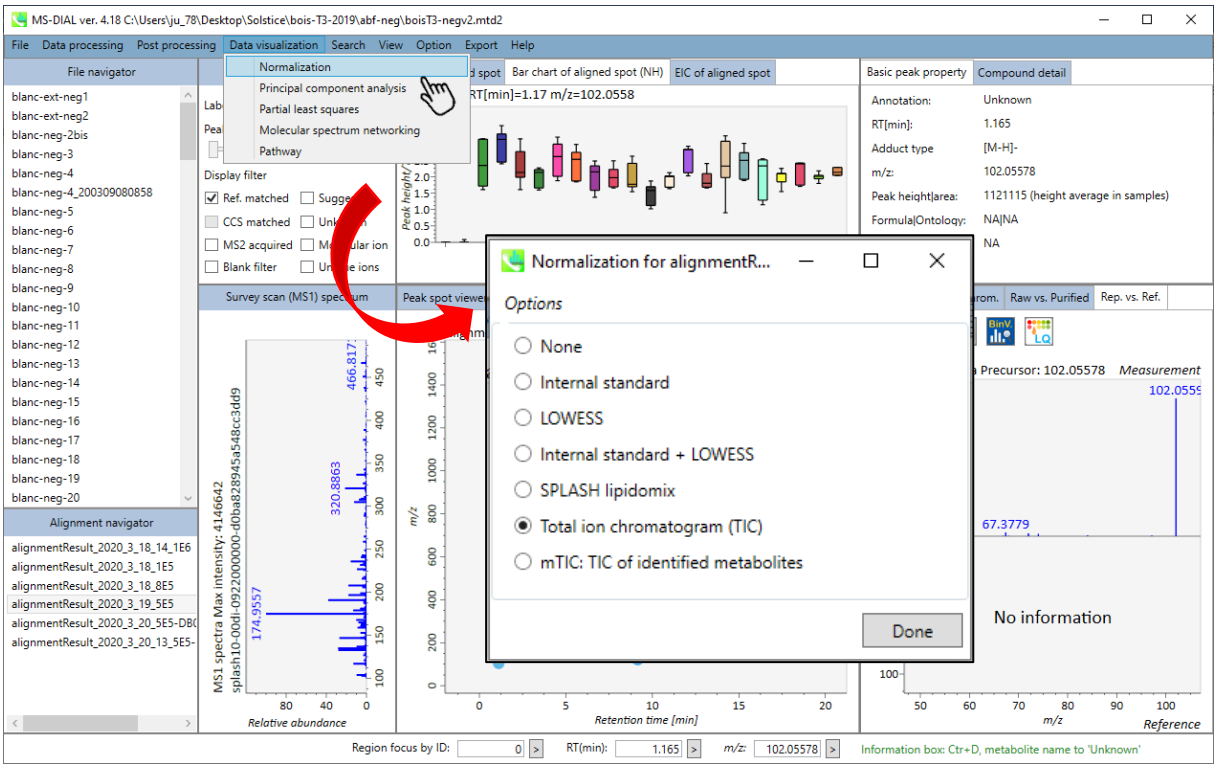


By clicking on “**Show ion table**”, you will easily see all the identified features and their respective intensities in each defined class:



ID	RT(min)	m/z	Type	Fill %	Metabolite name	Comment	Correlation	S/N	ANOVA P-value	Fold change (Max/Min)	BarChart
1	1.12	104.0351	[M-H]-	0.09	(IROA) SERINE		0.65	131.5	5.38E-05	118.12	
453	1.26	341.1085	[M-H]-	0.84	(IROA) MELIBIOSE		0.51	6289.5	7.17E-03	769.15	
310	8.21	289.0719	[M-H]-	0.84	(Metasci) (+)-Catechin (Hydrate)		0.67	29656.4	9.01E-09	20097.32	
309	8.81	289.0718	[M-H]-	0.84	(Metasci) (-)-Epicatechin		0.67	28564.6	3.13E-09	15026.86	
264	10.62	271.0614	[M-H]-	0.84	(Metasci) (+/-)-Naringenin		0.77	42.9	3.48E-13	1036319.00	
239	1.16	259.0224	[M-H]-	0.84	(IROA) MANNOSE 6-PHOSPHATE		0.80	1069.7	1.28E-15	475.27	
189	12.48	227.0714	[M-H]-	0.83	(Metasci) cis-Resveratrol		0.63	96.2	3.37E-02	5959.37	
185	11.74	227.0713	[M-H]-	0.84	(Metasci) trans-Resveratrol		1.00	9416.3	4.66E-15	20092.79	
146	7.13	203.0825	[M-H]-	0.83	(IROA) TRYPTOPHAN		0.39	1447.9	1.11E-03	3711.93	

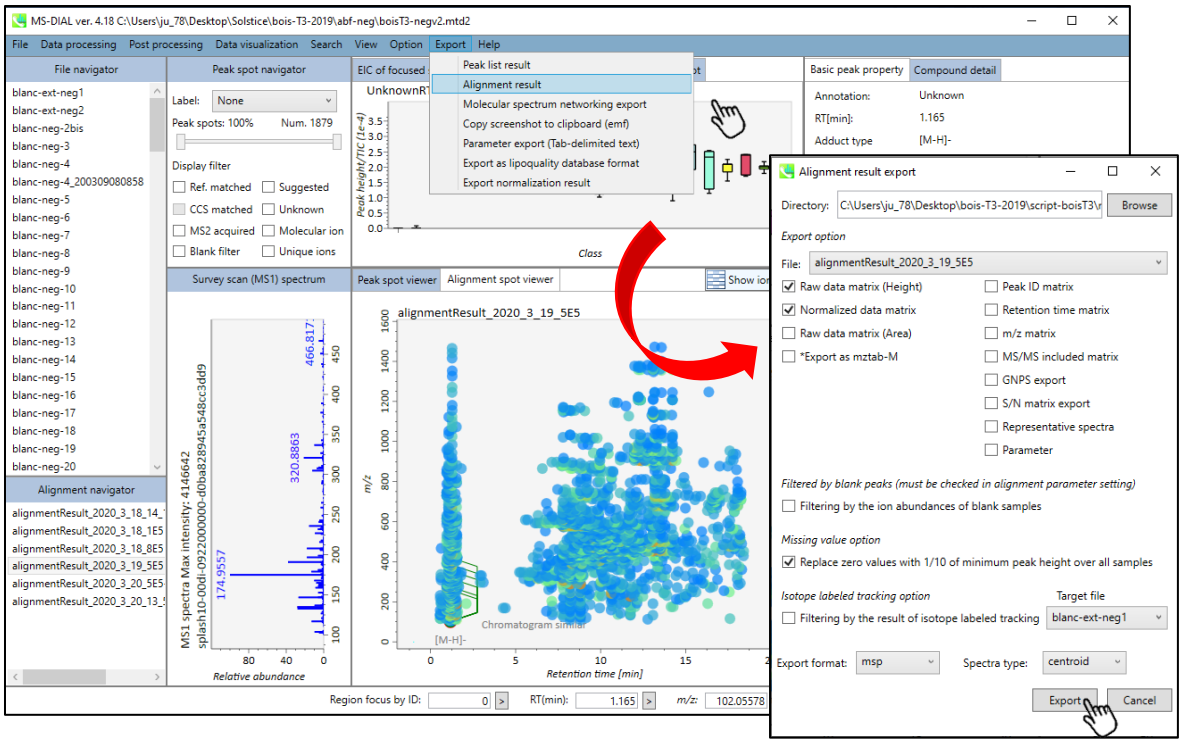
After alignment process, on the “**data visualization**” menu select “**normalization**” and normalized data by Total ion chromatogram (TIC) or another method.



On the “**data visualization**”, after normalization, you will be able to do some statistical analysis such as Principal component analysis.

5. Export peak list

On the “**Export**” menu select “**alignment results**” and export both **Raw data matrix (Height)** and **Normalized data matrix.** You can also select the option **“Replace zero values with 1/10 of minimum peak height over all samples”** for further statistical purpose.



6. Export all peaks

By right-clicking on one point, on the "**search formula and structure**" menu select "**Add component to search list**" and export « **all peaks »** for further identification purpose with MS-FINDER (see MS-FINDER tutorial). On the chosen folder, .MAT files will be generated and directly importable into MS-FINDER. At this point, you should be ready to follow the MS-CleanR and MS-FINDER tutorials.