

Technical guide for acquisition and archivation of high-resolution mass spectrometry (HRMS) chromatograms to NORMAN Digital Sample Freezing Platform (DSFP)

## Contents

<b>1. General recommendations before and during instrumental analysis.....</b>	<b>3</b>
<b>2. Post-data acquisition of Bruker files.....</b>	<b>6</b>
<b>2.1. Conversion of Bruker .d data files to mzML .....</b>	<b>6</b>
<b>2.2. Separation of bbCID collision energy channels .....</b>	<b>7</b>
<b>2.3. Contribution of HRMS chromatograms to DSFP.....</b>	<b>9</b>
<b>3. Post-data acquisition of Agilent files .....</b>	<b>15</b>
<b>4. Post-data acquisition of Waters files .....</b>	<b>18</b>
<b>5. Post-data acquisition of Thermo Fisher Scientific files.....</b>	<b>25</b>
<b>6. Post-data acquisition of AB Sciex.....</b>	<b>26</b>

## 1. General recommendations before and during instrumental analysis

The guide demonstrates how to upload LC-HRMS data to DSFP. The first chapter includes actions to be taken before and during the instrumental analysis (**Figure 1**), whereas the next chapters explain actions to be taken after the instrumental analysis depending on the instrument vendor.

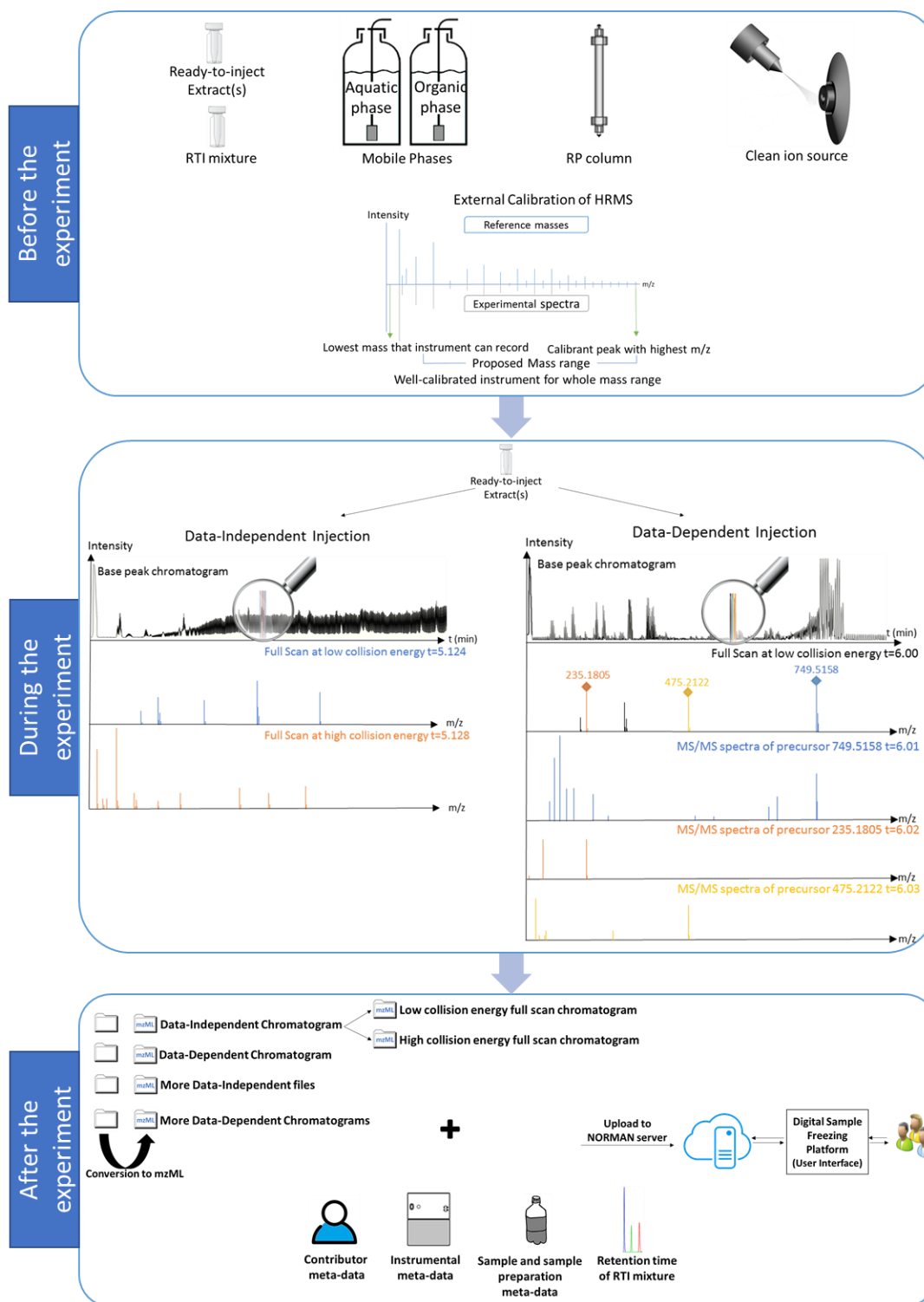
The following actions should be taken before the analysis of the samples:

1. Follow the sample-preparation protocol established by your laboratory, prepare the extracts and reconstitute them according to the selected protocol. Add internal standards to the samples before applying the sample preparation protocol to assure that the extraction was efficient. Prepare field blank and procedural blank samples to capture any unintentional contamination. Spike selected samples at the end of the sample preparation procedure (matrix-matched samples) to evaluate the accuracy of the method and obtain recovery values. Follow all needed quality control and quality assurance measures established by your laboratory. All samples intended to be analysed by LC-HRMS should be reconstituted to the vial, containing an organic phase (most commonly methanol) and some water, which is essential for the chromatography to work properly.
2. Add as much water as needed to the RTI mixture in the vial to achieve the same reconstitution proportion as the other extracts (e.g. 50% H<sub>2</sub>O and 50% MeOH).
3. Prepare sufficient quantity of filtered mobile phases according to the protocols of your laboratory to run all the samples in one batch.
4. Make sure that column is well-equilibrated, and that ion source is clean.
5. Recalibrate the HRMS before starting the sequence according to the vendor instructions. Calibrant peaks should cover the selected mass range.

The following actions should be taken during the analysis of the samples:

1. Use a reversed-phase chromatographic column and gradient program according to your laboratory protocol. The LC method should re-equilibrate the column for the next injection. The first injection should be regarded as chromatographically uncalibrated and should be excluded from subsequent analysis. Make sure that internal standards and spiked compounds are eluted in the expected retention time to verify that the analysis is going as expected.
2. Each extract should be injected in data-independent acquisition (DIA). DIA records all detected masses at low and high collision energy without any prior mass isolation. This, results in complex spectral information suitable for wide-scope suspect screening methods, but less suitable for identification of unknown compounds through non-target screening. Instruments are fast enough and can record one low and one high energy spectra in less than 1 s. Thus, instruments can record low and high collision energy within a single run, either by default or using specific settings. This or any similar approach can be applied as long as the scan rate is not severely affected, i.e. sufficient MS<sup>1</sup> full scan points are required for the DSFP.

- 3.** Depending on the purpose of the experiment, inject the extract in data-dependent acquisition (DDA) as many times as necessary. With this acquisition mode, pre-selected masses are isolated, fragmented and MS/MS spectra are recorded. This mode is ideal for identification of unknown masses of interest. It is recommended to record MS/MS spectra of as many precursors as possible. The DSFP can store these chromatograms as well.
- 4.** Repeat the same injections for the other samples in the sequence
- 5.** In the middle of the sequence, inject the RTI calibrant mixture in full scan mode and record the experimental retention time of the calibrant substances. This will enable you to use retention time index prediction to support the tentatively identified compounds with extra experimental evidence.

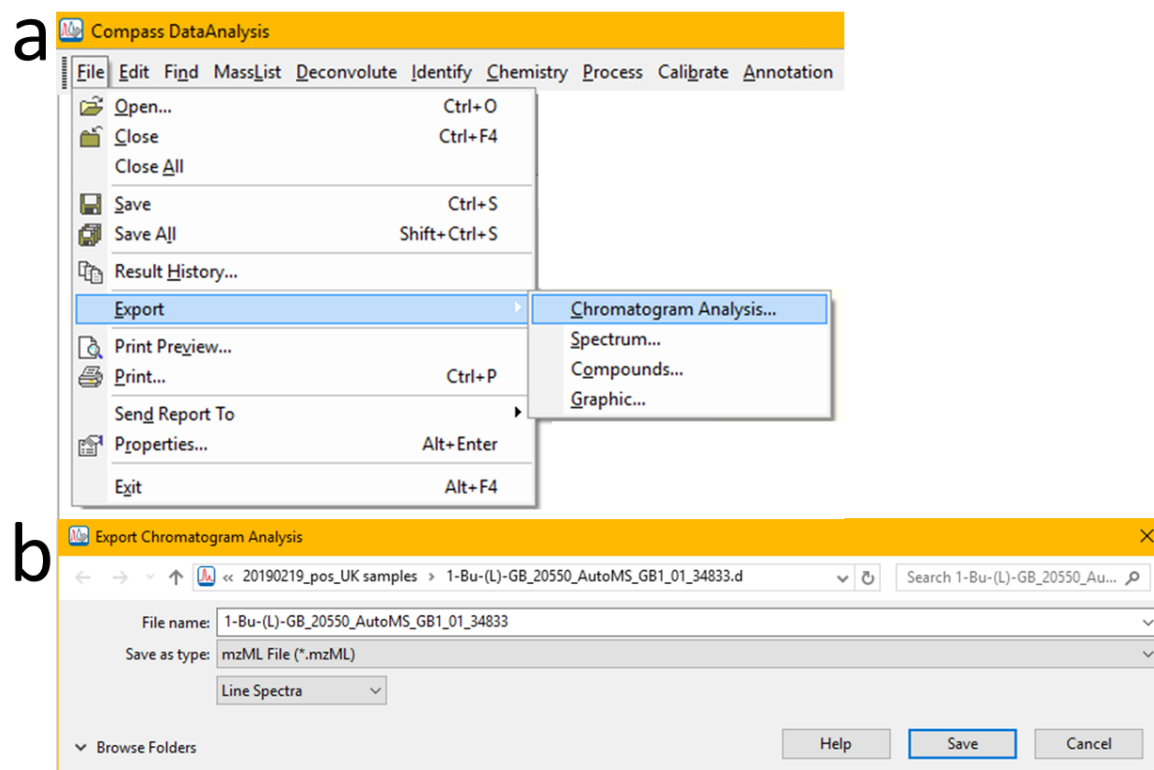


**Figure 1.** Steps to be followed to contribute LC-HRMS data to Digital Sample Freezing Platform (DSFP). Before starting the experiment, the instrument should be in good condition and well-calibrated. During the experiment, inject the samples in data-independent (DIA) and data-dependent acquisition (DDA). Convert the files to mzML and contribute them together with instrumental, sample preparation and RTI information to DSFP

## 2. Post-data acquisition of Bruker files

### 2.1. Conversion of Bruker .d data files to mzML

Bruker files (.d files) can be converted to mzML by Bruker CompassXport (**Figure 2**), which is embedded in DataAnalysis software provided by Bruker. Proteowizard software (**Figure 11**) can also be used for conversion. However, Proteowizard software exports uncalibrated Bruker mzML data files. Therefore, the use of Bruker CompassXport is recommended until the next version of Proteowizard incorporates an update to fix this issue. Most commonly, the calibrant substance is injected in the beginning of each chromatographic run, using a multi-port valve and chromatograms are recalibrated offline based on the experimentally observed and the theoretical  $m/z$ . Once the files are recalibrated, they can be exported by the following option on the Menu of Bruker DataAnalysis software: **File>Export>Chromatogram Analysis**. The same export approach can be used for files acquired in DDA. The disadvantage of the method described above is that files are processed one by one at a time. Extended tests, however, has shown that this way of recalibration and export, assures the lowest possible mass error and allow reliable conversion. The DDA mzML files can be uploaded to DSFP as they are, while DIA files should be separated as described in the next section.



**Figure 2.** Export of Bruker .d files using Bruker CompassXport v. 3.0.9.2 through Bruker DataAnalysis v.4.3.

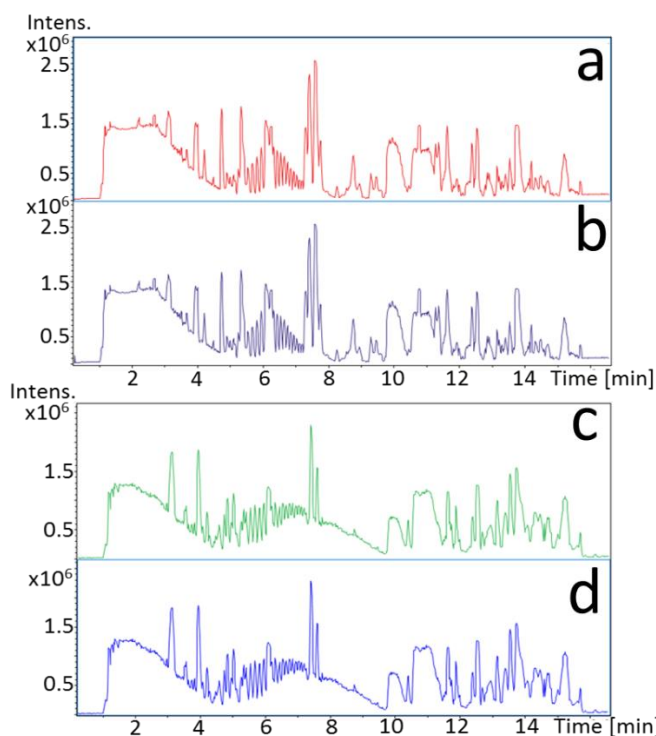
## 2.2. Separation of bbCID collision energy channels

Files acquired in DIA (termed as broadband collision induced dissociation (bbCID)) are further processed to separate the two collision energy channels by a web-tool integrated in DSFP. The tool is available in the website under the choice **More tools>Split Data-Independent data (Figure 3)**. The user can browse the mzML file (**Figure 3b**), insert an intensity cut-off value (**Figure 3d**) and calibrant scan numbers to be removed from the mzML (**Figure 3e**).

The screenshot shows the Norman web tool interface. At the top, the navigation menu includes 'Main Page', 'Batch Mode', 'Contribute', and 'More tools'. The 'More tools' dropdown is open, showing 'Split Data-Independent data' (labeled 'a') and 'NTS Basic Information'. Below the menu, a file upload section (labeled 'b') shows a 'Browse...' button and a text input field containing the file path '1-Bu-(L)-GB\_20550\_GB1\_01\_34801.mzML'. A blue bar indicates 'Upload complete'. To the right, a Windows 'Open' dialog box is shown (labeled 'c'), displaying the file '1-Bu-(L)-GB\_20550\_GB1\_01\_34801.mzML' (430,143 KB) in the 'New Volume (F:)' directory. The 'File name' field contains the same path, and the file type is set to 'Custom Files (\*.mzml;\*.mzml)'. The 'Open' button is highlighted. On the left side of the interface, there are several input fields: 'Intensity cutoff' (labeled 'd') with the value '200', 'Remove lock-mass or calibrant peaks. Enter for example "1,2,3,4" (comma delimited)' (labeled 'e') with the value '1,2,3,4,5,6,7,8,9,10,11', 'How many collision energy channels are contained in the chromatogram' (labeled 'f') with the value '2', 'Which nominal collision energy channels are contained in the chromatogram (i.e. 4eV, 20eV, 40eV)' (labeled 'f') with the value '4', and another 'Which nominal collision energy channels are contained in the chromatogram (i.e. 4eV, 20eV, 40eV)' (labeled 'f') with the value '25'. Below these, there are two large text areas (labeled 'g') containing lists of scan numbers. The first list is for energy channel 4, and the second is for energy channel 25. At the bottom, there are two 'Download channel of energy' buttons (labeled 'h') for '4' and '25'. Below these, a file list (labeled 'i') shows two files: '25\_1-Bu-(L)-GB\_...mzML' and '4\_1-Bu-(L)-GB\_2\_...mzML'. A 'Show all' button is also present.

Figure 3. Split of DIA collision energy layers in Bruker .d files.

“Intensity cut-off” is a numeric value below which spectral peaks are eliminated in all full-scan spectra. Setting an appropriate intensity cut-off value can reduce the size of the produced mzML file drastically and keep intact all the analytical information. The optimum “intensity cut-off” value is the digital noise of the photomultiplier detector. The value is dependent on the instrument vendor and may be different among different models of the same vendor. However, it can be easily determined by the user when the mass spectrometer is “on” and isolated from liquid chromatography. The intensity of the random noise that appears and disappears should be used as the “intensity cut-off”. In case, DSFP does not recognize which full-scans belong to which collision energy layer (depends on the information contained in the mzML file), the user should specify the number of the collision energy channels, the collision energy applied (**Figure 3f**) and which full-scan spectra belong to which collision energy layer (**Figure 3g**). It is advised that the same instrumental method is used for analysis of samples. In this case, the field **g** (**Figure 3g**) will always be the same. DSFP will separate the collision energy channels and appear download buttons, so that the user downloads the mzML files (**Figure 3h** and **Figure 3i**). The converted files (collision layer separated mzML files and the Bruker .d file) contain identical information which is indicated by the base-peak chromatograms for low (**Figure 4a** and **Figure 4b**) and high collision energy channels (**Figure 4c** and **Figure 4d**).



**Figure 4.** a) Full-scan MS1 (4 eV) base-peak chromatogram (.d file), b) Full-scan MS1 (4 eV) base-peak chromatogram (.mzML), c) Full-scan high collision energy (25 eV) layer base-peak chromatogram (.d file), d) Full-scan high collision energy (25 eV) layer base-peak chromatogram (.mzML file)



### 2.3. Contribution of HRMS chromatograms to DSFP

Contribution of LC-HRMS chromatograms to DSFP is possible through the option **Contribute** available in the top bar menu (**Figure 5**). In the tab **Basic Information**, the user should specify the institute name, which will auto-fill many fields (e.g. instrumental meta-data, retention time of calibrant substances, etc.). However, basic information for the contributed samples should be specified (**Figure 5b**; instrument type, short name of the sample, title of the project, location of the sample, date of collection and analysis, and enrichment factor).

The screenshot shows the 'Contribute' module of DSFP at step 1 (Basic Information). The interface is divided into two main sections: 'a Who is contributing' and 'b What is contributed'. Both sections have a red warning: 'All fields are compulsory'. Section 'a' includes a dropdown for 'Institute Name' with a list of institutions, and text input fields for 'Laboratory performing the analysis', 'City of the laboratory performing the analysis', and 'Country of the laboratory performing the analysis'. Section 'b' includes a dropdown for 'Instrument', and text input fields for 'Short name describing the sample', 'Title of Project (Acronym)', 'Country where the sample was taken', 'City where the sample was taken', 'Sampling Date in dd-mm-yyyy', 'Analysis Date in DD-MM-YYYY', 'How many times was sample pre-concentrated (Preconcentration factor)', and 'Unique Identifier'. A 'Proceed to step 2/6' button is located at the bottom of the form.

**Figure 5.** Screenshot from the **Contribute** module of DSFP at step 1 (**Basic Information**).

The location of the sample can be accurately specified in the next step (**Sample Meta-Data; Figure 6**), if the user inputs the exact coordinates of the sample location. In cases in which the exact location of the sample should not be revealed or is not known, the user can remove the latest digits of the decimal coordinates. Afterwards, the user must specify the type of environmental sample collected and input critical matrix-dependent meta-data information. In the next step (**Figure 7**) the ionisation and the instrumental information are specified.

norman

Main Page Batch Mode **Contribute** More tools ▾

Basic-Information **Sample-MetaData** Instrumental-MetaData RTI\_Calibration Spiked-Compounds Upload-Files

### General Sample Information

Longitude in Decimal (Compulsory)  
17.13

Latitude in Decimal (Compulsory)  
48.16

Precision of coordinates (Compulsory)  
Average (range 10-100 m) ▾

Type of data source  
▾

Type of monitoring  
▾

In case of transborder sample specify the 2nd country  
▾

Specify any remarks or important extra information regarding the samples

Type of matrix (Compulsory)

- Choose matrix from the list
- Choose matrix from the list
- Water-Waste water**
  - Municipal
  - Industrial
  - Municipal and Industrial
  - Other
- Water-Surface water**
  - River water
  - Lake water
  - Transitional water
  - Coastal water
  - Territorial (marine) water
  - Reservoirs
  - Other
- Water-Ground water**
  - Water-Ground water
- Sediment**
  - Sediment fom River water
  - Sediment from Lake water

**Figure 6.** Screenshot from the **Contribute** module of DSFP at step 2 (**Sample Meta-Data**).

I have positive ionization for the same samples

I have negative ionization for the same samples

**Positive ionization** All fields are compulsory

Bruker

Model type and number  
maXis Impact

Analytical column  
Thermo Acclaim RSLC C18

Column dimensions [length mm, I.D. mm, Particle size um]  
2.2um, 2.1x100mm

Injection volume  
5

Column temperature [°C]  
25

Composition of the mobile phase  
A 90:10 water:methanol with 0.01% formic acid and 5mM ammonium formate; B methanol with 0.01% formic acid and 5mM ammonium formate

Reconstitution solvent (Use the following format SolventA.SolventB %A.%B)  
Methanol:Water 50:50

**Negative ionization** All fields are compulsory

Bruker

Model type and number  
maXis Impact

Analytical column  
Thermo Acclaim RSLC C18

Column dimensions [length mm, I.D. mm, Particle size um]  
2.2um, 2.1x100mm

Injection volume  
5

Column temperature [°C]  
25

Composition of the mobile phase  
A 90:10 water:methanol with 0.01% formic acid and 5mM ammonium formate; B methanol with 0.01% formic acid and 5mM ammonium formate

Reconstitution solvent (Use the following format SolventA.SolventB %A.%B)  
Methanol:Water 50:50

Proceed to step 4/6

**Figure 7.** Screenshot from the **Contribute** module of DSFP at step 3 (**Instrumental Meta-Data**).

Depending on the selected ionization(s) in previous step (**Figure 7**), the respective RTI calibrant table(s) will appear (**Figure 8**). In this step, the user should specify the retention time of the calibrant substances as indicated in the red box of **Figure 8**. Once the user proceeds to the next step, a list of spiked compounds is requested. More specifically, the spiked concentration level and the observed area or intensity must be specified (**Figure 9**). This optional step allows the semi-quantification of the detected suspects during the batch-mode screening process. It is highly recommended to fill in the table. Overall, it is valid that the more information and meta-data is provided to the system, the better results will be acquired during the batch-mode screening process.

norman

Main Page Batch Mode Contribute More tools ▾

Basic-Information Sample-MetaData Instrumental-MetaData RTI\_Calibration Spiked-Compounds Upload-Files

I have not injected these these calibrants OR I am using GC-HRMS

You have to input the RT of the calibrants. Let the cells empty for compounds without RT.

**Positive ionization**

	Name	RT	CAS	Formula	Ion
1	Guanylurea	1.325	141-83-3	C2H6N4O	103.0614
2	Amitrol	1.392	61-82-5	C2H4N4	85.0509
3	Histamine	1.642	51-45-6	C5H9N3	112.0869
4	Chlormequat	1.875	999-81-5	C5H13ClN	122.0731
5	Methamidophos	2.625	10265-92-6	C2H8NO2PS	142.0086
6	Vancomycin	3.192	1404-90-6	C66H75Cl2N9O24	724.7224
7	Cefoperazone	4.342	62893-19-0	C25H27N9O8S2	646.1497
8	Trichlorfon (Dylox)	5.142	52-68-6	C4H8Cl3O4P	256.9299
9	Butocarboxim	5.992	34681-10-2	C7H14N2O2S	191.0849
10	Dichlorvos	6.908	62-73-7	C4H7Cl2O4P	220.9532
11	Tylosin	7.975	1401-69-0	C46H77NO17	916.5264
12	TCMTB	9.208	21564-17-0	C9H6N2S3	238.9766
13	Rifaximin	10.025	80621-81-4	C43H51N3O11	786.3596
14	Spinosad A (Spinosyn A)	11.525	131929-60-7	C41H65NO10	732.4681
15	Emamectin B1a	12.475	121124-29-6	C49H75NO13	886.5311
16	Avermectin B1a (Abamectin)	13.674	71751-41-2	C48H72O14	890.5260
17	Nigericin	13.908	28380-24-7	C40H68O11	725.4834
18	Ivermectin B1a	14.458	70288-86-7	C48H74O14	892.5436

You have to input the RT of the calibrants. Let the cells empty for compounds without RT.

**Negative ionization**

	Name	RT	CAS	Formula	Ion
1	Amitrole	1.461	61-82-5	C2H4N4	83.0363
2	Benzoic acid	2.760	65-85-0	C7H6O2	121.0295
3	Acephate	2.977	30560-19-1	C4H10NO3PS	182.0046
4	Salicylic acid	3.476	69-72-7	C7H6O3	137.0244
5	Simazine 2-Hydroxy	4.859	2599-11-3	C7H13N5O	182.1047
6	Tepaloxymdim	5.142	149979-41-9	C17H24ClNO4	340.1321
7	Bromoxynil	5.259	1689-84-5	C7H3Br2NO	273.8509
8	MCPA	6.292	94-74-6	C9H9ClO3	199.0167
9	Valproic acid	6.925	99-66-1	C8H16O2	143.1078
10	Phenytoin	7.091	57-41-0	C15H12N2O2	251.0826
11	Flamprop	7.490	58667-63-3	C16H13ClFNO3	320.0495
12	Benodanil	7.891	15310-01-7	C13H10INO	321.9734
13	Dinoterb	8.008	1420-07-1	C10H12N2O5	239.0673
14	Inabenfide	9.207	82211-24-3	C19H15ClN2O2	337.0749
15	Coumaphos	10.989	56-72-4	C14H16ClO5PS	361.0072
16	Triclosan	11.839	3380-34-5	C12H7Cl3O2	286.9439
17	AvermectinB1a (Abamectin)	13.572	65195-55-3	C48H72O14	871.4849
18	Salinomycin	14.488	53003-10-4	C42H70O11	749.4845

Proceed to step 5/6

**Figure 8.** Screenshot from the **Contribute** module of DSFP at step 4 (RTI calibration).

Main Page Batch Mode **Contribute** More tools ▾

Basic-Information Sample-MetaData Instrumental-MetaData RTI\_Calibration **Spiked-Compounds** Upload-Files

You have to input the spiked compounds, the spiked concentration level and the peak area. Let the cells empty for compounds without RT.

	Spiked Compound Name	SMILES	Conc units	Response expression	Conc 1	Response 1	Conc 2	Response 2	Conc 3	Response 3	Conc 4	Response 4	Conc 5
1	Benzotriazole (BTR)	<chem>c1ccc2c(c1)[nH]nn2</chem>	ug/L	Peak Area ▾	25	94,301	50	177,107	75	265,660	100	459,322	150
2	Benzotriazole-5-Me	<chem>Cc1ccc2c(c1)[nH]nn2</chem>	ug/L	Peak Area ▾	25	116,767	50	218,648	75		100	528,229	150
3	Benzotriazole-5,6-di-Me	<chem>Cc1cc2c(cc1C)nn[nH]2</chem>	ug/L	Peak Area ▾	25	85,291	50	159,265	75		100	388,938	150
4	Benzotriazole-1-Methyl	<chem>Cn1c2ccccc2nn1</chem>	ug/L	Peak Area ▾	25	138,969	50	268,138	75		100	702,812	150
5	Benzothiazole-2-Amino	<chem>c1ccc2c(c1)nc(s2)N</chem>	ug/L	Peak Area ▾	25	276,787	50	499,565	75	749,348	100	1,207,429	150
	Positive ionization		ug/L	Peak Area ▾	25		50		75		100		150
		ug/L	Peak Area ▾	25		50		75		100		150	
		ug/L	Peak Area ▾	25		50		75		100		150	
		ug/L	Peak Area ▾	25		50		75		100		150	
		ug/L	Peak Area ▾	25		50		75		100		150	

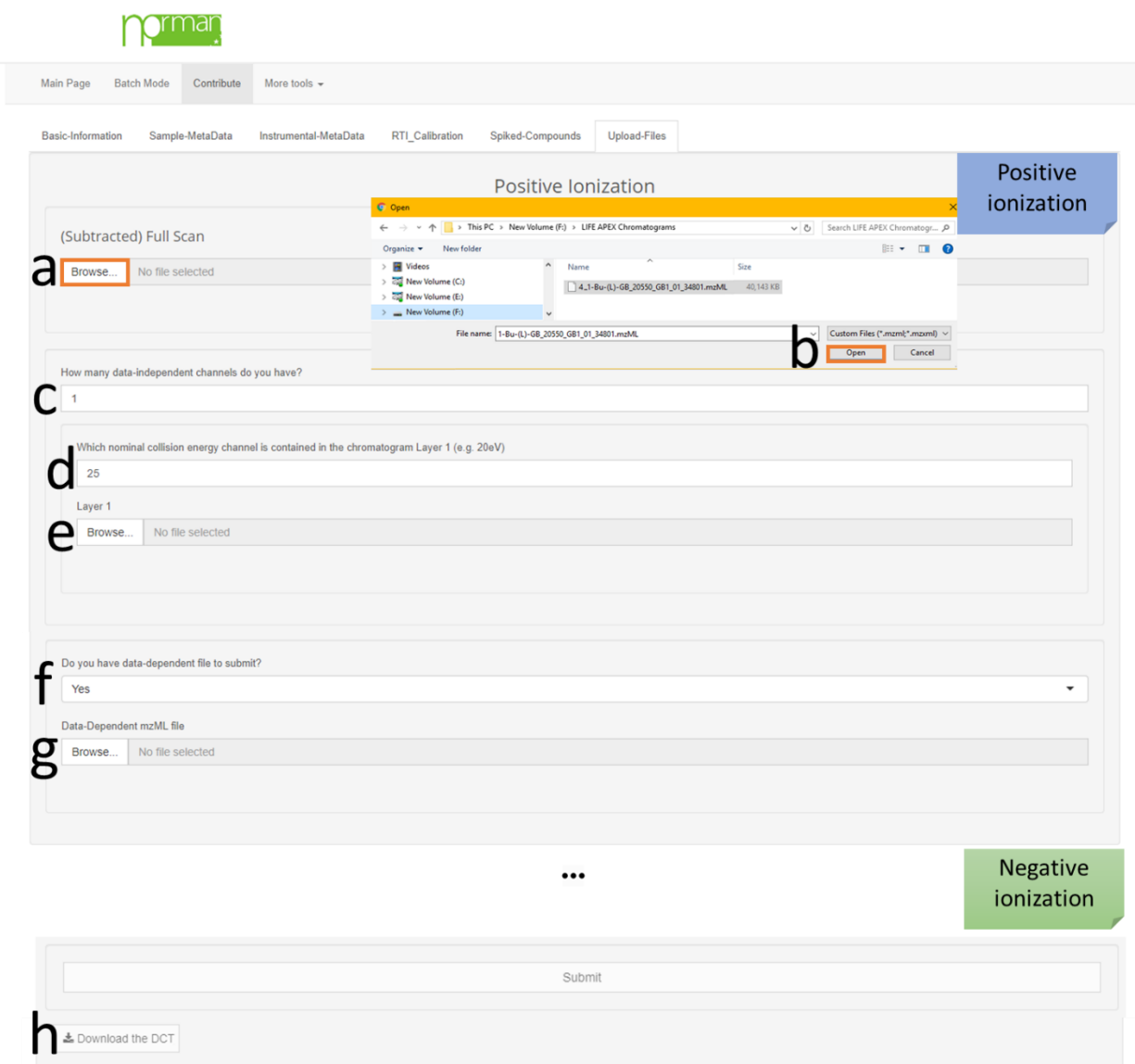
You have to input the spiked compounds, the spiked concentration level and the peak area. Let the cells empty for compounds without RT.

	Spiked Compound Name	SMILES	Conc units	Response expression	Conc 1	Response 1	Conc 2	Response 2	Conc 3	Res
1	Chloramphenicol	<chem>c1cc(ccc1[C@H]([C@@H](CO)NC(=O)C(C)Cl)O)[N+](=O)[O-]</chem>	ug/L	Peak Area ▾	25	9,940	50	19,885	75	2
2	Clofibric acid	<chem>CC(C)(C(=O)O)Oc1ccc(cc1)Cl</chem>	ug/L	Peak Area ▾	25		50		75	
3	Diclofenac	<chem>c1ccc(c(c1)CC(=O)O)Nc2c(cccc2Cl)Cl</chem>	ug/L	Peak Area ▾	25		50		75	
4	Florfenicol	<chem>CS(=O)(=O)c1ccc(cc1)[C@H]([C@@H](CF)NC(=O)C(C)Cl)O</chem>	ug/L	Peak Area ▾	25		50		75	
5	Furosemide	<chem>c1cc(oc1)Cnc2cc(c(cc2C(=O)O)S(=O)(=O)N)Cl</chem>	ug/L	Peak Area ▾	25		50		75	
6			ug/L	Peak Area ▾	25		50		75	
	Negative ionization		ug/L	Peak Area ▾	25		50		75	
		ug/L	Peak Area ▾	25		50		75		
		ug/L	Peak Area ▾	25		50		75		
		ug/L	Peak Area ▾	25		50		75		
		ug/L	Peak Area ▾	25		50		75		

Proceed to step 6/6

**Figure 9.** Screenshot from the **Contribute** module of DSFP at step 5 (**Spiked compounds**).

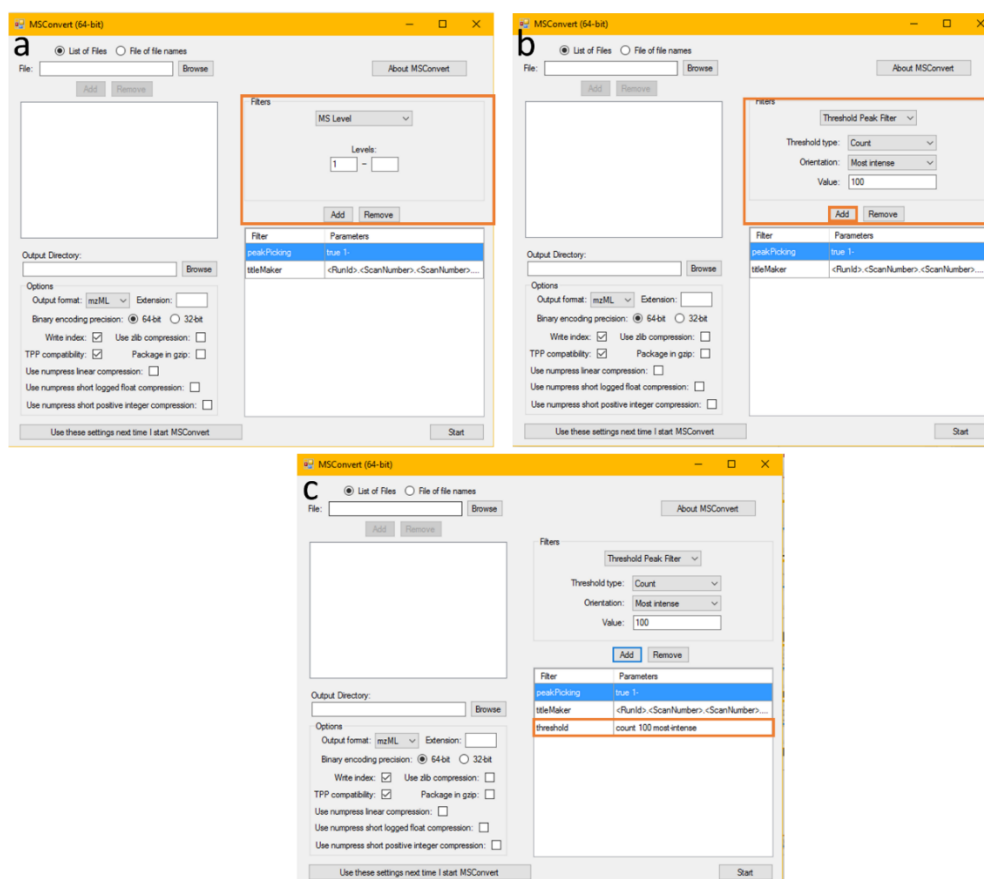
The final step in the procedure is to upload the mzML files (**Figure 10**). The user should select the MS<sup>1</sup> full-scan file (**Figure 10a**). If the user has analysed the sample in DIA, it is requested to specify the number of DIA collision energy channels (**Figure 10b**). Browse buttons (**Figure 10d and Figure 10e**) will appear depending on the number of DIA collision energy channels (input to the field **Figure 10c**). If the sample has been also analysed in DDA method, the user should answer positively (**Figure 10f**) and upload the DDA chromatogram in the respective browse button (**Figure 10g**). If the user has specified in previous steps that chromatograms are available in negative ionisation, then the same fields will also appear for the negative ionisation. Once the user uploads the mzML files, the submit button will be activated. All chromatograms and information provided will be analysed by the DSFP. The progress is indicated by a loading bar. The output of the procedure is reflected in the data collection template (DCT) excel file, which can be downloaded (**Figure 10h**). The next sample can be uploaded by clicking the tab “**Basic Information**”. All previous details remain unchanged, which makes the procedure of uploading of the second sample faster.



**Figure 10.** Screenshot from the **Contribute** module of DSFP at step 6 (**Upload Files**).

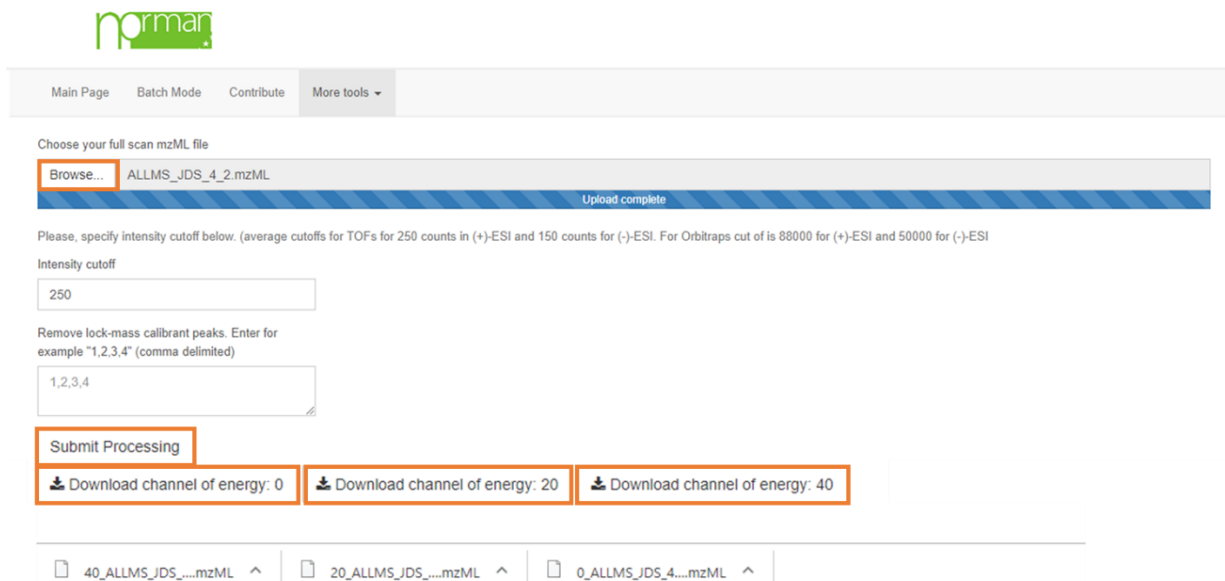
### 3. Post-data acquisition of Agilent files

Agilent files can be converted using Proteowizard software (**Figure 11**), which is available to be downloaded in the following link (<http://proteowizard.sourceforge.net/downloads.shtml>). After installing Proteowizard, use “MSConvert” to convert the files to mzML. Select the files that you want to convert by clicking on browse button. Submitting a conversion request without adding an intensity cut-off may result in mzML files of few Gigabytes. Therefore, it is recommended to enable a filter called “**Threshold Peak Filter**” with “**Count**” as threshold type, “**Most intense**” as orientation and “**Value**” equal to the noise level of the spectral peaks generated by the detector (digital noise). This choice is available under the menu “Filters”. The cut-off value should be equal to the digital noise of the photomultiplier detector. If the data is not recorded in centroid mode, enable the filter “**Peak Picking**” (already enabled in **Figure 11**). Afterwards, the user should press add and click the start button. The files will be converted to mzML by default in the same path of the Agilent .D files unless otherwise specified in the output directory field. DDA mzML files are ready to be uploaded to DSFP.



**Figure 11.** Msconvert.exe from ProteoWizard. The figure demonstrates how to set an intensity cut-off threshold to avoid having large mzML files.

DIA files contain different collision energy layers which can be split using the tool in DSFP under the option **More tools>Split Data-Independent Data (Figure 12)**. The user can browse the mzML file, set an intensity cut-off value and remove any unwanted scans. If the cut-off value is equal or below the cut-off value set in Proteowizard software, no further data reduction and no further spectral peak removal is applied. The scans of mzML files coming from the conversion of Agilent .D files are automatically recognised. Therefore, the split of the different collision energy layers is straight forward; the user must click on the Submit Processing button, a loading bar will appear and once the split is done, the download buttons will appear.



**Figure 12.** Split of DIA collision energy layers in Agilent .D files.

By default, the DIA method of Agilent HRMS, has two collision energy layers (one at 20 eV and one at 40 eV). Thus, the contribution procedure is the same as the one described in section **2.3 Contribution of HRMS chromatograms to DSFP**. The only difference is the number of DIA channels (**Figure 13a**), which should be two. This will add two browse buttons, one for the 20 eV collision energy (**Figure 13b**) channel and another one for the 40 eV collision energy (**Figure 13c**).



**normal**

Main Page Batch Mode **Contribute** More tools ▾

Basic-Information Sample-MetaData Instrumental-MetaData RTI\_Calibration Spiked-Compounds **Upload-Files**

Positive Ionization **Positive ionization**

(Subtracted) Full Scan

Browse... **0\_ALLMS\_JDS\_4\_2.mzML** Upload complete

**a** How many data-independent channels do you have?  
**2**

**b** Which nominal collision energy channel is contained in the chromatogram Layer 1 (e.g. 20eV)  
**20**

Layer 1  
Browse... **20\_ALLMS\_JDS\_4\_2.mzML** Upload complete

**c** Which nominal collision energy channel is contained in the chromatogram Layer 2 (e.g. 20eV)  
**40**

Layer 2  
Browse... **40\_ALLMS\_JDS\_4\_2.mzML** Upload complete

Do you have data-dependent file to submit?  
Yes ▾

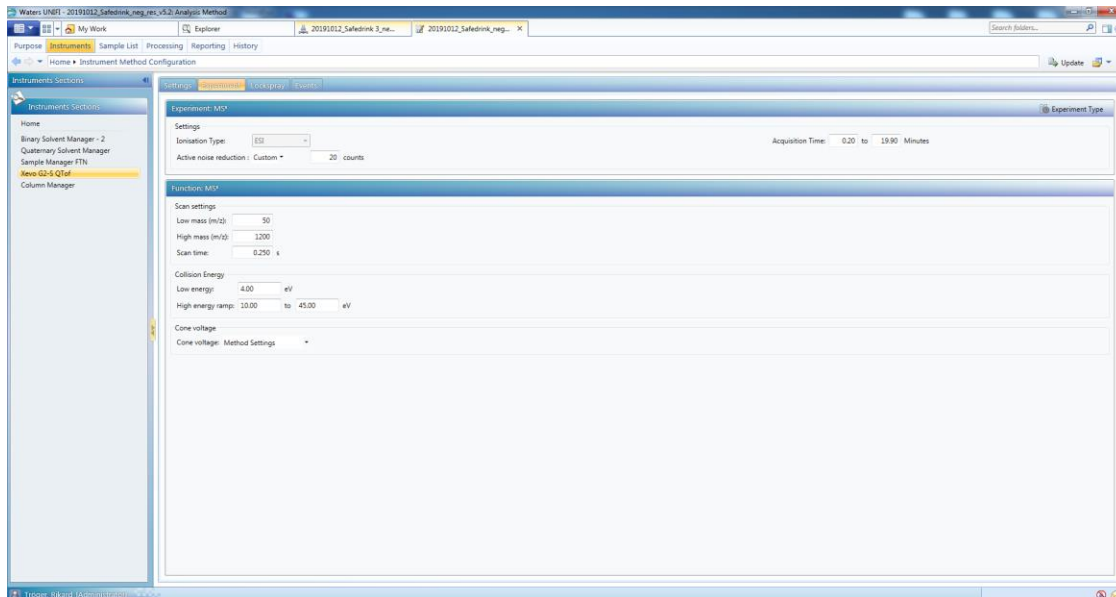
Data-Dependent mzML file  
Browse... **AMSMS\_JDS\_4\_2.mzML** Upload complete

**Figure 13.** Screenshot from the **Contribute** module of DSFP at step 6 (**Upload Files**) for Agilent files, which have two DIA collision energy channels (20 and 40 eV).

## 4. Post-data acquisition of Waters files

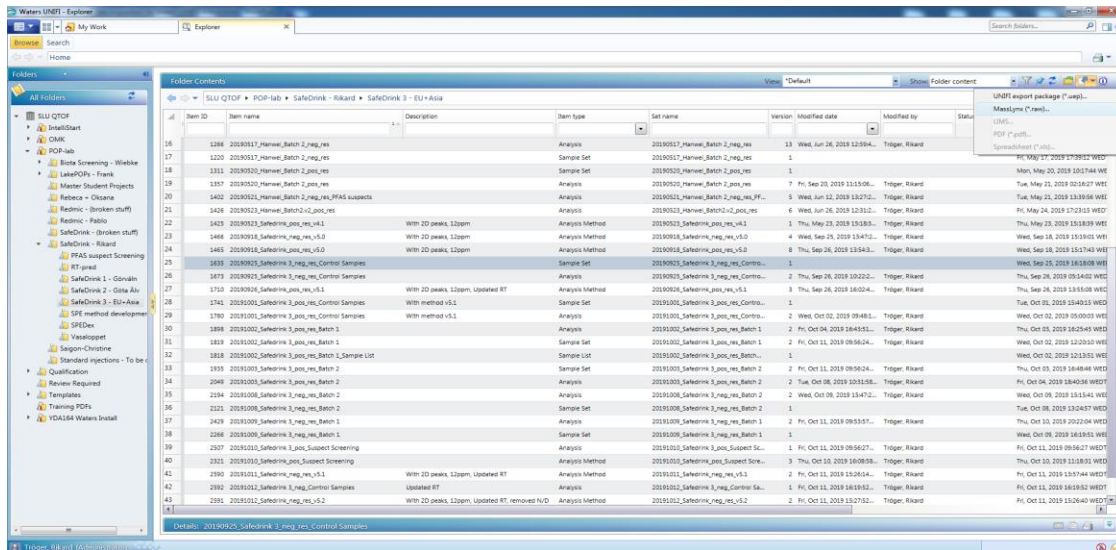
The following steps should be followed for the conversion of Waters files:

1. Make sure that UNIFI (.net 4.5.2), masslynx and ProteoWizard (.net 4.7.2) are installed.
2. To be able to work with the data smoothly, active noise reduction needs to be turned on and set it to value 20 (any value between 10 and 50 can work). Advise **Figure 14** for further details.



**Figure 14.** Screenshot from Waters UNIFI showing how to enable active noise reduction.

3. Go to Browse and export the sample set as MassLynx Raw.



**Figure 15.** Screenshot from Waters UNIFI showing how to export the sample as MassLynx Raw.

4. Open Masslynx and create a new project. Add the files that needs to be converted (**Figure 16**).

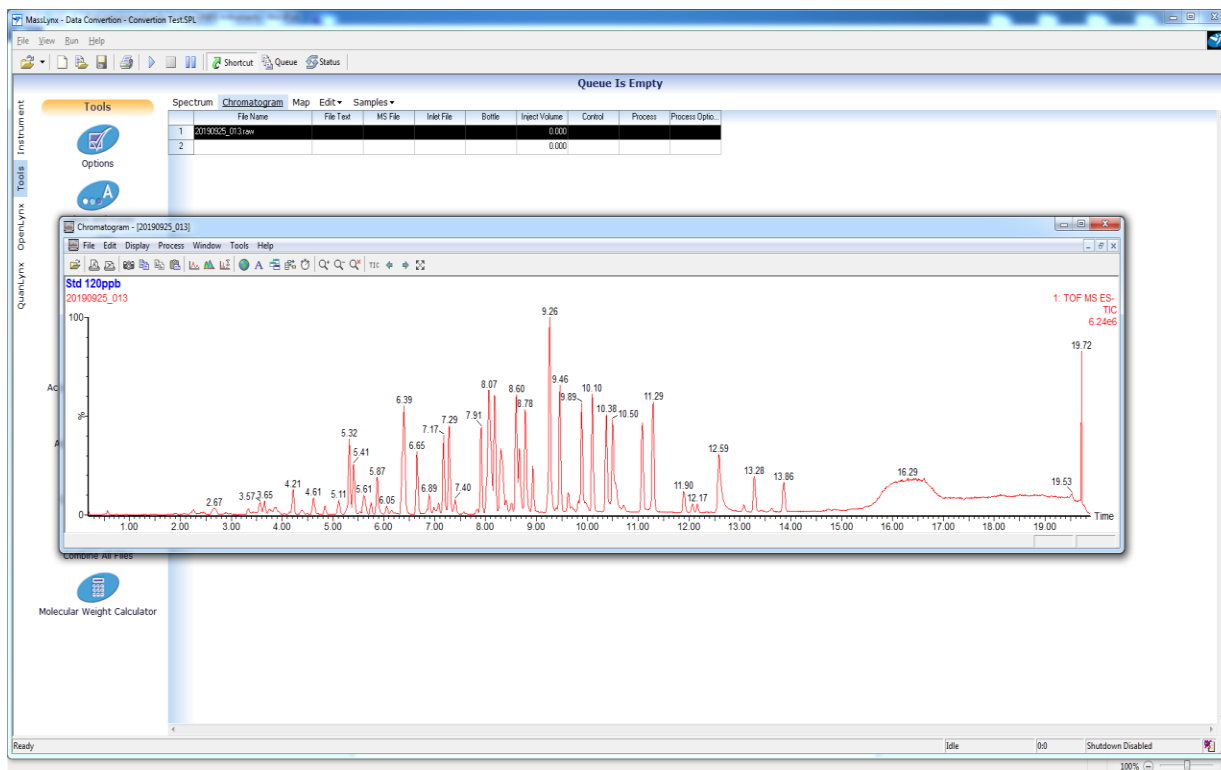


Figure 16. Screenshot from Masslynx.

5. Once files are imported, press Accurate Mass Measure as shown in **Figure 17**.

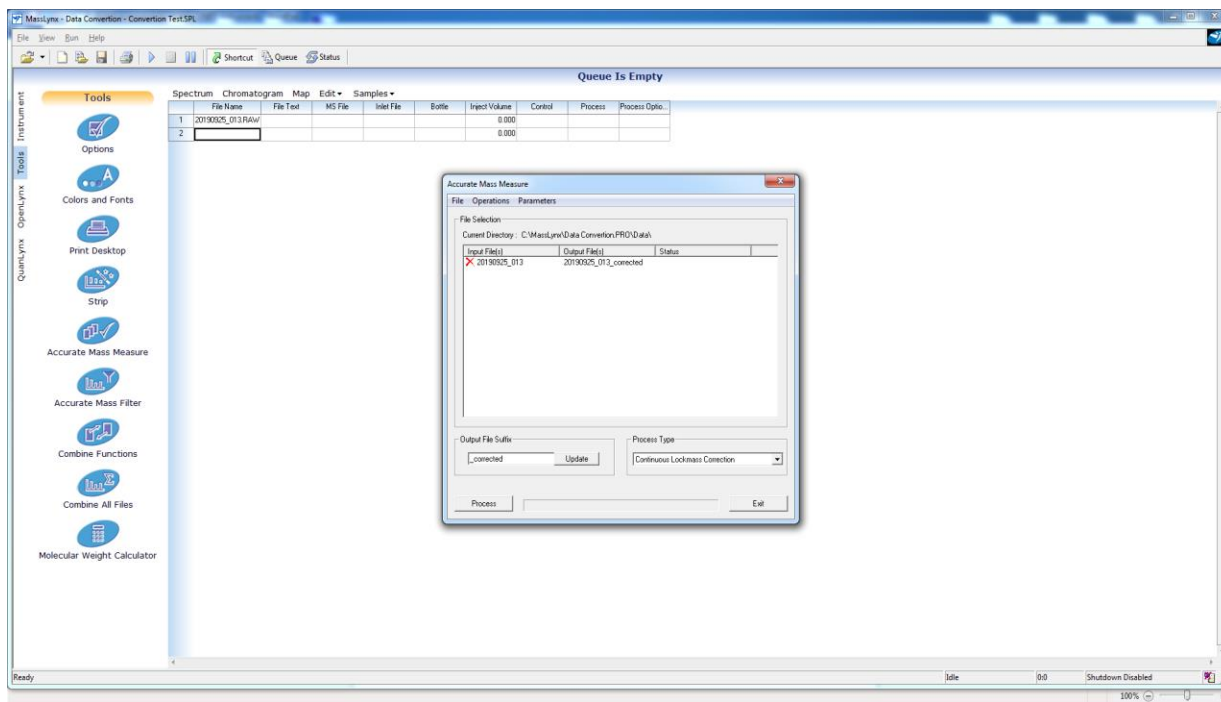


Figure 17. Screenshot from Masslynx with the window for accurate mass measure settings

6. Afterwards, double click on the files that you want to convert and add an output file suffix (Figure 18)

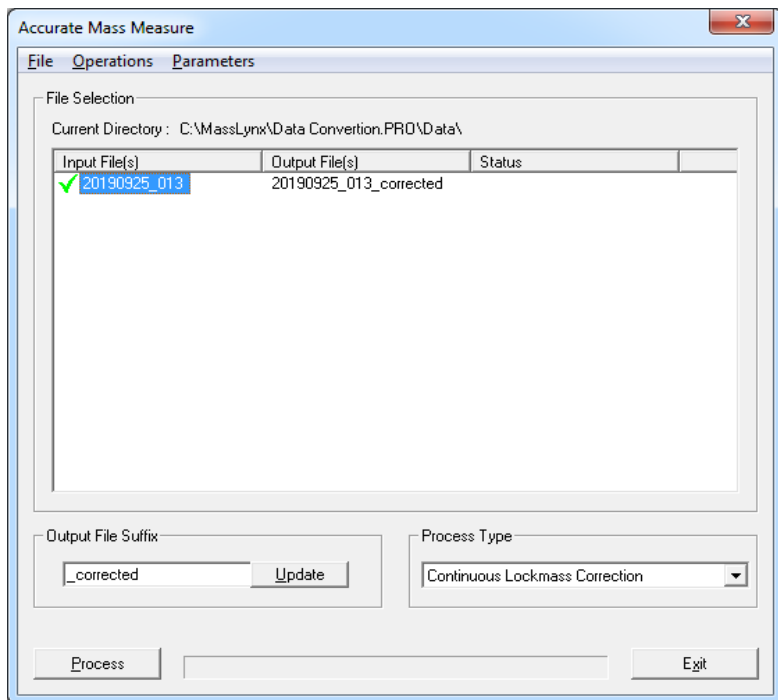


Figure 18. Accurate mass measure settings

7. At that point, press Parameters and Mass Measure Parameters and then the ionization mode that the samples were analysed.

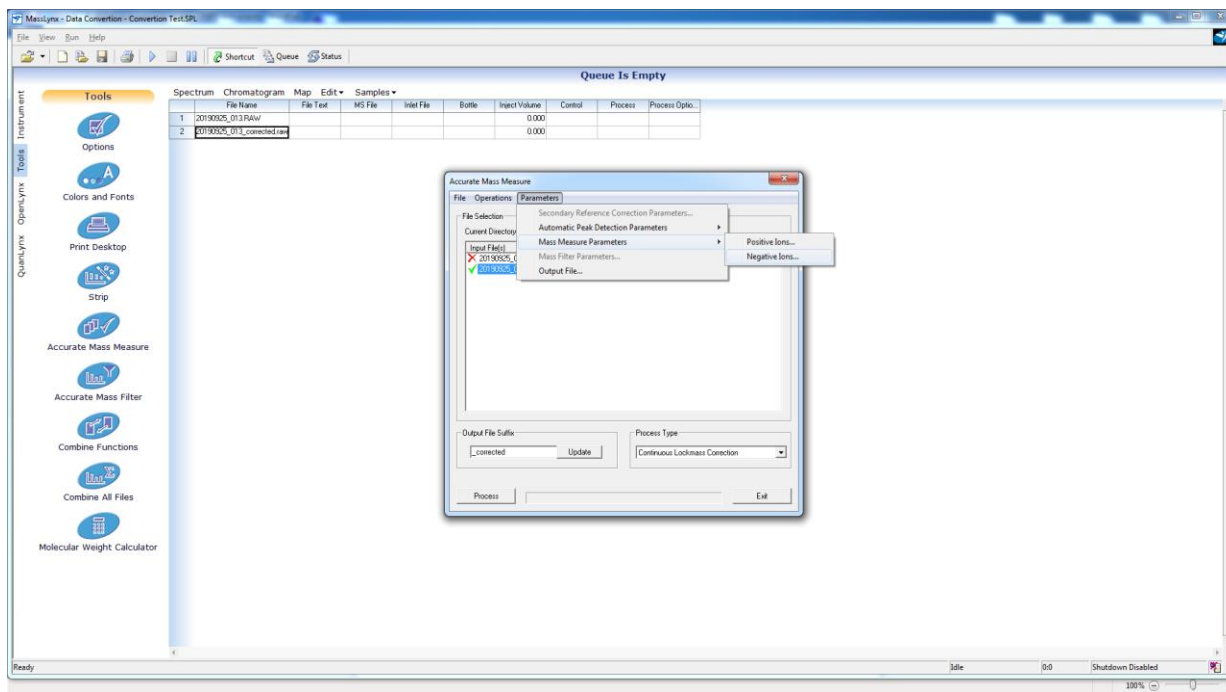


Figure 19. Parameter selection in accurate mass measure settings.

8. Then, press TOF and input the lock mass that you are using

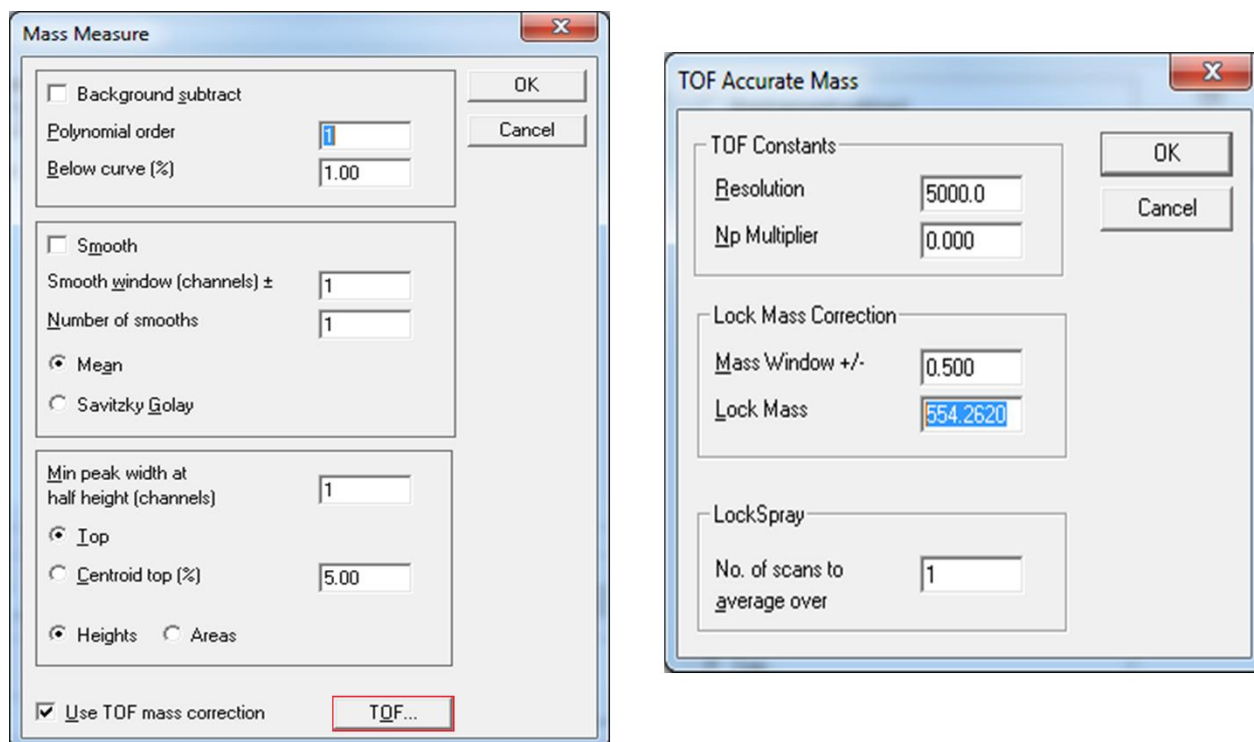


Figure 20. Insertion of lock mass in the TOF accurate mass settings.

9. Then, press process and wait until the lock mass corrected files are generated (Figure 21).

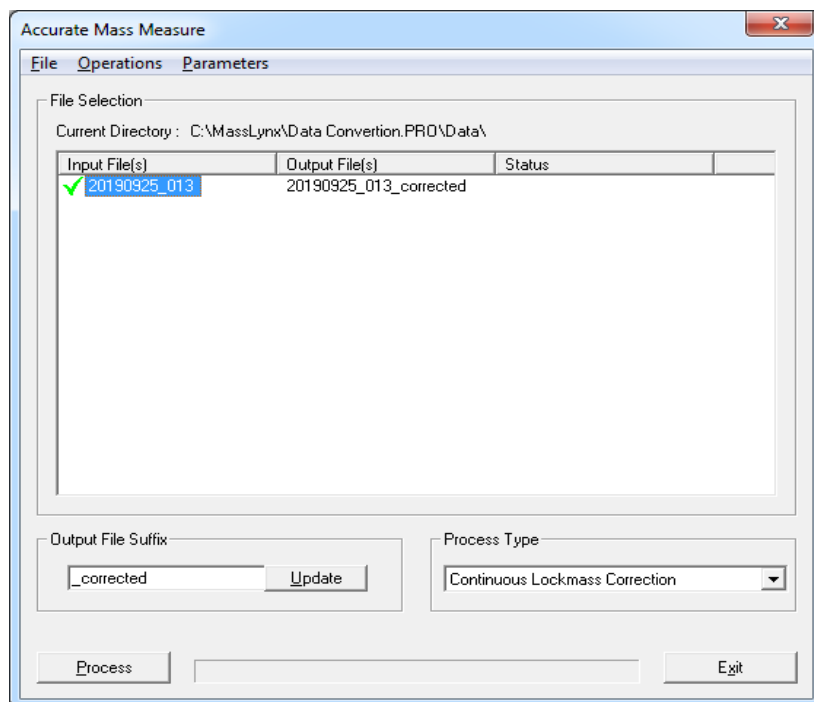
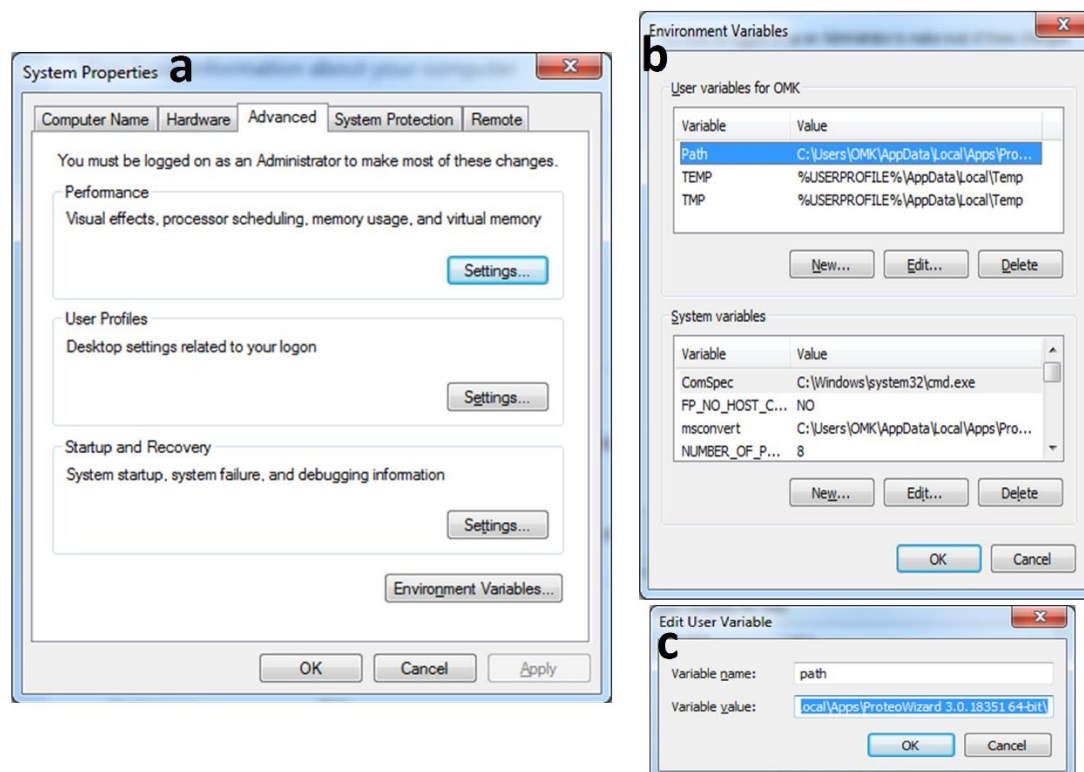


Figure 21. Processing the files based on the lock mass filters set in previous steps.

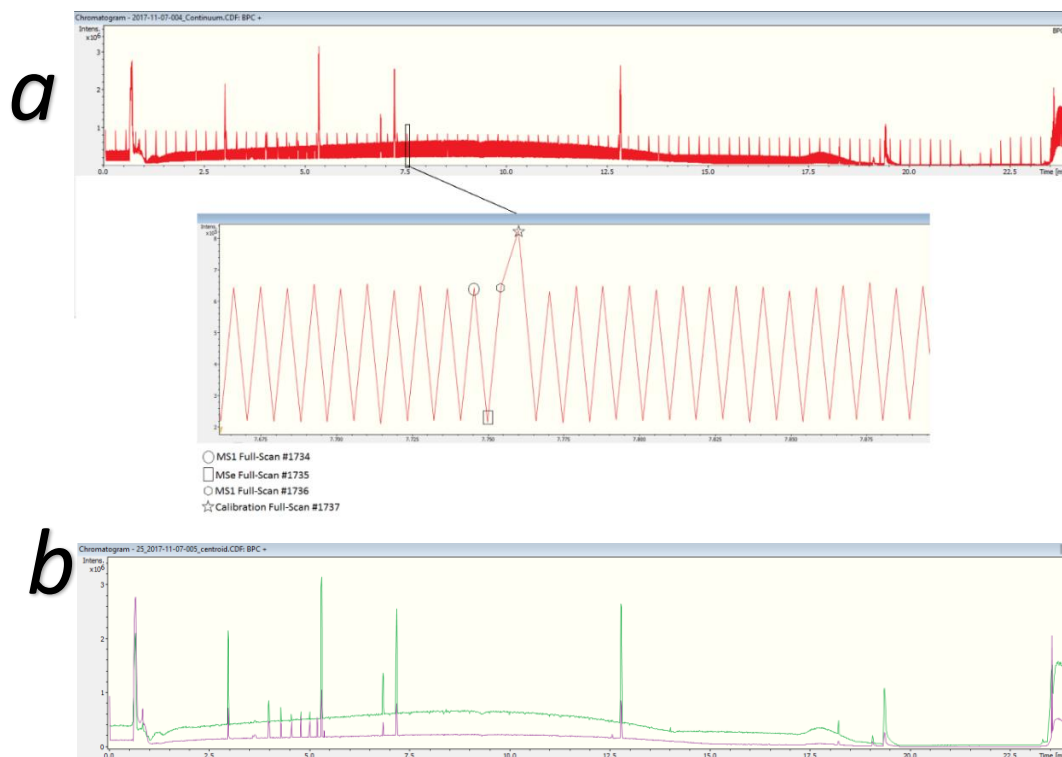
Waters files can be converted to mzML using ProteoWizard (Download link at <http://proteowizard.sourceforge.net/downloads.shtml>). After installing ProteoWizard, use “MSConvert” to convert the files to mzML. Select the files that you want to convert by clicking on browse button. Submitting a conversion request without adding an intensity cut-off may result in mzML files of few Gigabytes. Therefore, it is recommended to enable the filter “**Threshold Peak Filter**” with “**Count**” as threshold type, “**Most intense**” as orientation and “**Value**” equal to the noise level of the spectral peaks generated by the detector (digital noise). For Waters QTOF “**Value**” parameter at around 300 counts is valid for positive ionization. However, the value should be carefully selected, because this value depends on the detector of each instrument. Data should be recorded in centroid mode, otherwise enable filter “**Peak Picking**”.

Conversion can be done using command line if needed. In this case, the installation path of ProteoWizard needs to be set in the system variables as shown in **Figure 22**. After setting the installation path as global system variable, generate a bat file using notepad and add the following command as content `msconvert.exe *raw --mzML --filter "threshold absolute 300 most-intense"`. Afterwards, save the bat file in the directory of the lock-mass calibrated files and run it. You may modify the command of the bat file according to the instructions at <http://proteowizard.sourceforge.net/tools/msconvert.html>.



**Figure 22.** For setting installation path of ProteoWizard to a computer, go to Control Panel>System and Security>System and click on Advanced system settings (part a of the figure). Press settings in “Environment Variables”, edit “Path” variable (part b of the figure) and input the installation path in “Variable value” field.

DDA files can be uploaded to DSFP as they are. On the contrary, the DIA mzML files need further processing so that the collision energy layers are separated (example in **Figure 23**). **Figure 23a** represents the base peak chromatogram of a sample. Waters QTOF record sequentially MS<sup>1</sup> and MS<sup>e</sup> full scans as other QTOF vendors. The difference is that a calibration lock-mass full-scan occurs at a fixed number of full-scans (e.g. every 56 scans as shown in **Figure 23a**). These lock-mass full-scans should be removed. A tool to separate DIA LC-HRMS data to low and high collision energy channels and remove the lock-mass full scan spectra is integrated in DSFP.



**Figure 23** a). Example of Waters DIA chromatogram with lock-mass calibration every 56 scans. B). Chromatogram after the separation of the two collision energy layers (MS<sup>1</sup> layer is green and MS<sup>e</sup> is purple).

On the top bar menu of DSFP, there is an option **More tools>Split Data-Independent Data** to remove the unwanted lock-mass full-scan spectra (**Figure 24a**) and separate the different collision energy layers (**Figure 24b** and **Figure 24c**). Waters vendor software may be used to visualize the scans and specify which scans should be discarded and which not. If the uploaded mzML contains information about the applied collision energies, then the system will request no further information. Otherwise, it will be required to specify the number of collision energy channels, the nominal collision energy of each channel and the scan numbers that correspond to each collision energy channel (**Figure 24**).

norman

Main Page Batch mode Contribute More tools ▾

Choose your full scan mzML file

Browse... NL06\_bbCID\_GE5\_01\_29605.mzML

Upload complete

Unfortunately, your file does not contain collision energy information. You should specify which scans belong to which layer. Follow the instructions below

Please, specify intensity cutoff below. (average cutoffs for TOFs for 250 counts in (+)-ESI and 150 counts for (-)-ESI. For Orbitraps cut of is 88000 for (+)-ESI and 50000 for (-)-ESI)

Intensity cutoff

250

**a** Remove lock-mass or calibrant peaks. Enter for example "1,2,3,4" (comma delimited)

2, 5153, 5208, 5209, 5264, 5265, 5320, 5321, 5376, 5377

Lock-mass full scans to be discarded

How many collision energy channels are contained in the chromatogram

2

Which nominal collision energy channels are contained in the chromatogram (i.e. 4eV, 20eV, 40eV)

4

Which nominal collision energy channels are contained in the chromatogram (i.e. 4eV, 20eV, 40eV)

25

I finished with inputting the information

**b** Write down the scans of channel of collision energy

4

5363, 5365, 5367, 5369, 5371, 5373, 5375, 5378

Full scans of low collision energy

**c** Write down the scans of channel of collision energy

25

5358, 5360, 5362, 5364, 5366, 5368, 5370, 5372, 5374, 5379

Full scans of high collision energy

I finished with inputting the information

2018 © All Rights Reserved.

**Figure 24.** Screenshot from the integrated tool in DSFP, which helps to separate the different layers from DIA LC-HRMS data

Once the separated collision energy channels are saved and the DDA file is saved (if any), select **Contribute** option from the top bar menu and follow the procedure as described in section **2.3 Contribution of HRMS chromatograms to DSFP**. DSFP will guide you step-by-step during the upload.



## 5. Post-data acquisition of Thermo Fisher Scientific files

Users of Thermo fisher scientific HRMS instruments (hybrid ion-traps-orbitraps and quadrupole-orbitraps) employ in most cases the DDA method. However, it is possible that samples are analysed by DIA method if the isolation window is set as wide as the scan range. The DDA files can be converted to mzML with Proteowizard software (<http://proteowizard.sourceforge.net/downloads.shtml>) with the same way as previously described (**Figure 11**).

Briefly, use “MSConvert” included in Proteowizard to convert the .raw files to mzML. Simply, select the files that you want to convert by clicking on browse button (**Figure 11**). You may enable a filter called “**Threshold Peak Filter**” with “**Count**” as threshold type, “**Most intense**” as orientation and “**Value**” equal to the noise level of the spectral peaks generated by the detector (digital noise). This choice is available under the menu “Filters”. The cut-off value should be equal to the digital noise of the photomultiplier detector. Unlike other vendors, the “**Count**” value is few orders of magnitude higher in Thermo fisher scientific HRMS instruments. This happens, because the detectors in Thermo fisher scientific instruments provide signals with number of counts of many orders of magnitude higher than other vendors. A typical “Count” cut-off value is 80,000 for positive ionization and 50,000 for negative ionization. However, these cut-off values are not valid for all Thermo fisher scientific HRMS instruments. In the newest models of Thermo fisher scientific, the resulting mzML files have reasonable sizes (few MB), even if no “**Threshold Peak Filter**” is applied. In these cases, it is not recommended to apply “**Threshold Peak Filter**”, since it is not needed. If the data is not recorded in centroid mode, enable the filter “**Peak Picking**” (already enabled in **Figure 11**). Afterwards, the user should press add and click the start button. The files will be converted to mzML by default in the same path of the .raw files, unless otherwise specified in the output directory field.

DDA mzML files can be uploaded to DSFP without any other action. In case, DIA files are available, conversion of .raw files is performed the same way as DDA. The difference is that the collision energy-layers of DIA files need to be separated. The separation of the collision energy layers is straight forward and possible through the tool in DSFP (**More tools>Split Data-Independent Data**) described in **Figure 12**.

The mzML files together with the meta-data can be uploaded to DSFP through the “Contribute” option on the top bar menu. The contribution procedure is the same as the one described in section **2.3 Contribution of HRMS chromatograms to DSFP**. DSFP will guide you step-by-step during the upload. The field “(Subtracted) Full Scan File” (shown in **Figure 10a**) is mandatory. The DDA mzML file can be uploaded in this field, and also the same file in the field “Data-Dependent mzML file” (shown in **Figure 10g**). If the sample has been analysed in DIA acquisition, then it is recommended to upload the lowest collision energy channel of the DIA run in the field “(Subtracted) Full Scan File” (shown in **Figure 10a**).

## 6. Post-data acquisition of AB Sciex

AB Sciex DDA data can be converted with Proteowizard (<http://proteowizard.sourceforge.net/downloads.shtml>) as data from other vendors (Figure 11). After installing Proteowizard, use "MSConvert" to convert the files to mzML. Select the files that you want to convert by clicking on browse button. It is recommended to enable a filter called "Threshold Peak Filter" with "Count" as threshold type, "Most intense" as orientation and "Value" equal to the noise level of the spectral peaks generated by the detector (digital noise). Most of the times, for AB Sciex data, a low "Value" should be chosen (1 to 10 counts), since the intensity values are low in absolute numbers. If the data is not recorded in centroid mode, enable the filter "Peak Picking" (already enabled in Figure 11). Afterwards, the user should press add and click the start button. The files will be converted to mzML by default in the same path of the AB Sciex .wiff files, unless otherwise specified in the output directory field.

SWATH mzML files can be edited using notepad or any other similar text editor (e.g. notepad++). Replace the following highlighted statements as shown in Figure 25:

- "ms level" value="2" to "ms level" value="1"
- name="collision energy" value="55.0" to ""

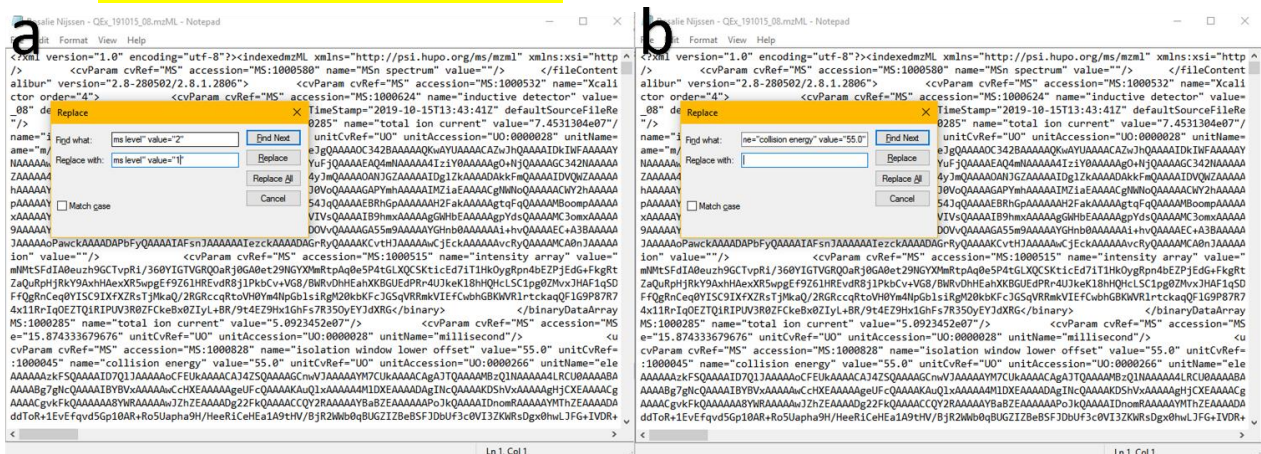


Figure 25. Screenshot from notepad showing the replace of mslevel values to 1 and the removal of collision energy information from the mzML files.

Press File>Save and proceed with the same method for all files. Enter DSFP and go to More tools>Split Data-Independent data. Browse the mzML files to be separated. Set intensity cut-off to a low value (below the Threshold Peak Filter), set the number of collision energy layers (e.g. six in case of MS1 full-scan, and MS2 without prior isolation with scan range <150, <250, <350, <450, <750 Da). Then, input the scans that belong to each channel. The configuration is shown in Figure 26.

Below is an example how to fill in the fields indicating which scans belong to which layer (Figure 26). Since the scan-acquisition rate of instruments is precise, the following number sequences can be easily produced using any programming language. For example, the sequences layers can be produced using the following R-codes:





Choose your full scan mzML file

Browse... Rosalie Nijssen - QEx\_191015\_05.mzML Upload complete

Unfortunately, your file does not contain collision energy information. You should specify which scans belong to which layer. Follow the instructions below  
Please, specify intensity cutoff below. (average cutoffs for TOFs for 250 counts in (+)-ESI and 150 counts for (-)-ESI. For Orbitraps cut is 88000 for (+)-ESI and 50000 for (-)-ESI)

Intensity cutoff

Remove lock-mass or calibrant peaks. Enter for example "1,2,3,4" (comma delimited)

How many collision energy channels are contained in the chromatogram

Which nominal collision energy channels are contained in the chromatogram (i.e. 4eV, 20eV, 40eV)

Which nominal collision energy channels are contained in the chromatogram (i.e. 4eV, 20eV, 40eV)

Which nominal collision energy channels are contained in the chromatogram (i.e. 4eV, 20eV, 40eV)

Which nominal collision energy channels are contained in the chromatogram (i.e. 4eV, 20eV, 40eV)

Which nominal collision energy channels are contained in the chromatogram (i.e. 4eV, 20eV, 40eV)

Which nominal collision energy channels are contained in the chromatogram (i.e. 4eV, 20eV, 40eV)

Write down the scans of channel of collision energy:  
4

Write down the scans of channel of collision energy:  
150

Write down the scans of channel of collision energy:  
250

Write down the scans of channel of collision energy:  
350

Write down the scans of channel of collision energy:  
450

Write down the scans of channel of collision energy:  
750

**Figure 26.** An example of how to define which full scan belongs to which layer in data-independent LC-HRMS data.

Download the separated SWATH files and upload them together with the DDA mzML files (if available), which are ready to be uploaded to DSFP, following the steps described in section **2.3 Contribution of HRMS chromatograms to DSFP**. DSFP will guide you step-by-step during the upload.