# Converting Raw LC-HRMS/MS Files into mzML files

Benjamin J. Place\*

# 2023-03-15

# Converting To The mzML Format

To use DIMSpec tools, raw data files produced by vendor software must be converted into \*.mzML files. The easiest way to convert data files is to download and install the most recent version of ProteoWizard<sup>1</sup> from https://proteowizard.sourceforge.io/ to use the MSConvert tool (Adusumilli, Ravali and Mallick, Parag 2017). Once installed, follow the next steps to convert the raw file(s) to \*.mzML format.

## 1) Start MSConvert

From the Proteowizard main page, start the MSConvert program with Menu > Proteowizard > MSConvert<sup>2</sup>

🔜 MSConvertGUI (64-bit)			-		×
List of Files					
File: Browse	Browse network resourc	e 🗸	About	MSConve	ert
Add Remove					
	Filters				
		Peak Picking 🗸 🗸			
	_	Algorithm:			
	C	WT (continuous wavelet transform; works for any profile data	a) ~		
		MS Levels: Min SNR: Min peak spar	cing:		
Output Directory:					
Dotions		Add Remove			
Output format: mzML V Extension:	Filter	Parameters			
Binary encoding precision:	titleMaker	<runid>.<scannumber>.<scannumber>.<chargestate:< td=""><td>&gt; File:"<sourc< td=""><td>ePath&gt;", I</td><td>Nati</td></sourc<></td></chargestate:<></scannumber></scannumber></runid>	> File:" <sourc< td=""><td>ePath&gt;", I</td><td>Nati</td></sourc<>	ePath>", I	Nati
Write index: 🔽 Use zlib compression: 🔽	threshold	absolute 1 most-intense			
TPP compatibility: Package in gzip:	peakPicking	cwt snr=0.1 peak Space=0.1 msLevel=1-			_
Use numpress linear compression:					
Use numpress short logged float compression:					
Use numpress positive integer compression:					
Combine ion mobility scans:					
SIM as spectra: SRM as spectra:					
Presets: Generic Defaults	✓ Save Preset ▼	Files to convert in paral	lel: 10 🛓	Star	rt

Figure 1: Initial screen when running MSConvert

<sup>\*</sup>NIST | MML | CSD | benjamin.place@nist.gov

<sup>&</sup>lt;sup>1</sup>Any mention of commercial products is for information only; it does not imply recommendation or endorsement by NIST.

 $<sup>^{2}</sup>MSC$  onvert includes a command line interface for advanced use cases; settings must be the same as described in this guide.

#### 2) Select the files to be converted

The MSConvert software can convert to mzML data files from the the following vendor formats at the time this guide was written: Thermo Fisher Scientific (\*.raw); Waters (\*.raw); SCIEX (\*.wiff2); Agilent (\*.D); Shimadzu (\*.LCD); and Bruker (\*.D)

- Select the files to be converted using the Browse button at the top left (Fig 2A).
   Note: If selecting only one file at a time, instead click the Add button to include the file in the list.
- Select the *Output Directory* to save files by clicking the **Browse** button at the center left (Fig 2B).
   Note: It will default to the same directory as the original file.

After loading the file, the program should look similar to Figure 2.

MSConvertGUI (64-bit)	- 0	×
List of Files      File of file names      File:      Browse      Add Remove	A Browse network resource ~	¢
S_PFAC30PAR_Spectra\PFAC30PAR_PFCA2.raw	Subset       2         MS levels:       -       Charge states:       -          Scan number:       -       Number of data points:       -          Scan time (seconds):       -       Collision energy:       -          Scan event:       -       Activation type:       Any          Scan polarity:       Any       Analyzer type:       Any	C
Data\20 Browse	Add Remove	
Output format: mzML V Extension:	Filter Parameters	
Binary encoding precision:      64-bit      32-bit	titleMaker <runid>.<scannumber>.<scannumber>.<chargestate> File:"<sourcepath>", Na</sourcepath></chargestate></scannumber></scannumber></runid>	ati
Write index: 🗹 Use zlib compression: 🗹	threshold absolute 1 most-intense	_
TPP compatibility: 🗹 Package in gzip: 🗌	peakPicking vendor msLevel=1-2	_
Use numpress linear compression:		
Use numpress short logged float compression:		
Use numpress positive integer compression:		
Combine ion mobility scans:  2 Combine ion mobility scans:  SRM as spectra:  SRM as spectra:		
Presets: Generic Defaults	✓ Save Preset ▼ Files to convert in parallel: 10 ♀ Start	

Figure 2: Interface for MSConvert after a data file has been loaded

#### 3) Adjust the mzML file parameters in the Options box (Fig 2C).

- 1. Select **mzML** from the *Output format* input and make sure the *Extension* input is blank.
- 2. Select **64-bit** under *Binary encoding precision*.
- 3. Ensure boxes are checked next to *Write Index*, *TPP Compatibility*, and *Use zlib compression*; leave all other boxes unchecked.

The proper selections have been made in Figure 2.

## 4) Add conversion filters (Fig 2D)

Apply the following data conversion filters; to adhere to current NIST practice the following settings *must* be used. Select the filter to be applied from the drop down menu at the top of the **Filters** box. Do this sequentially as described here. After setting each parameter click the **Add** button. To remove a parameter (in case of an entry or selection error), select that filter in the table below Fig 2D and click the **Remove** button.

- 1. Select the *Peak Picking* filter and apply the following parameters:
  - Algorithm: Vendor or CWT

Note: Vendor filter does not work for Waters files, you must use CWT

- MS Levels: 1 \_ (leave the second box blank)
- If using CWT:
  - Min. SNR: 0.1
  - Min. peak spacing: 0.1

Parameters should match those in Figure 3.

Click the  $\mathbf{Add}$  Button

Filters	Peak Picking ~
	Algorithm: Vendor (does not work for UNIFI, and it MUST be the first filter!) ~
	MS Levels: Min SNR: Min peak spacing: 1 - 0.1 0.1
	Add Remove

Figure 3: Peak picking filter parameters

- 2. Select the *Threshold Peak Filter* and include the following parameters:
  - Threshold type: Absolute intensity
  - Orientation: Most intense
  - Value: 1

Parameters should match those in Figure 4 (next page) Click the **Add** Button

Filters	old Peak Filter 🗸
Threshold typ Orientatio Valu	e: Absolute intensity ~ m: Most intense ~ e: 1
	Add Remove

Figure 4: Threshold peak filter parameters

- If the raw data is from a Waters Corporation mass spectrometer, apply the Lockmass Refiner filter.
  - Reference m/z: use the mass-to-charge ratio of the lockmass, for Leucine-Enkephalin (most common), the negative ionization m/z is 554.2615 and the positive ionization m/z is 556.2771.
  - m/z Tolerance: 0.1

Parameters should match those in Figure 5 Click the **Add** Button

Lockmass Refiner V
Reference m/z:     554.2615       m/z Tolerance:     0.1
Add Remove

Figure 5: Lockmass refiner parameters

Selected filter parameters should be very similar to those in Figure 2 prior to conversion to \*.mzML.

**Note**: The TitleMaker filter is automatically selected upon opening MSConvert; this can be included and will not affect the data analysis. However, peakPicking must still be the first line in the filters.

Note: Waters Corporation instruments should have the lockmass refiner filter in the as well.

## 5) Convert data files to the \*.mzML format.

Once all settings are as required, click the **Start** button at the bottom right (Fig 2E). This will convert all added files to \*.mzML format and put them in the assigned Output directory. Wait for the conversions to finish.

### 6) (Optional) Save the settings as a preset

Use the **Save Preset** drop down menu to use the same settings in the future (Fig 6) by selecting *Save Preset* As.... To apply a preset profile to future files, select it from this same drop down menu.



Figure 6: MSConvert presets menu

Once all files have been converted to the \*.mzML format they are ready for use in other data tools.

This concludes the Quick Guide to Converting Raw LC-HRMS/MS Files into mzML files.

## References

Adusumilli, Ravali and Mallick, Parag. 2017. "Data Conversion with ProteoWizard msConvert." Methods in Molecular Biology (Clifton, N.J.) 1550: 339–68. https://doi.org/10.1007/978-1-4939-6747-6\_23.