Supplementary Note 1: DIA Data Analysis Without Spectral Libraries

This tutorial is a practical guide for how to use the Encyclopedia software suite to build chromatogram libraries for DIA-MS. We have a GUI-based workflow and also a command line workflow. We've included options for visualizing the results of the Encyclopedia analysis in Skyline or in Encyclopedia itself.

Citations

MSconvert (https://www.nature.com/articles/nbt.2377)

A cross-platform toolkit for mass spectrometry and proteomics. Chambers MC et al. *Nat Biotech* 30, 918-920 (2012). doi.org/10.1038/nbt.2377

Encyclopedia (https://www.nature.com/articles/s41467-018-07454-w)

Chromatogram libraries improve peptide detection and quantification by data independent acquisition mass spectrometry. Searle BC et al. *Nat Comm* 9, 5128 (2018). doi.org/10.1038/s41467-018-07454-w

You will need:

- MSConvert from Proteowizard: *Windows only!*
  - http://proteowizard.sourceforge.net/download.html
- EncyclopeDIA suite (*.jar file): *command line and cross-platform GUI*
  - https://bitbucket.org/searleb/encyclopedia/wiki/Home

To exactly replicate the results here, you will also need:

- RAW data files from the tutorial HeLa dataset (MSV000084531)
  - 2018may16_hela_window_size_test_BCS_hela_narrow_1.raw
  - 2018may16_hela_window_size_test_BCS_hela_narrow_2.raw
  - 2018may16_hela_window_size_test_BCS_hela_narrow_3.raw
  - 2018may16_hela_window_size_test_BCS_hela_narrow_4.raw
  - 2018may16_hela_window_size_test_BCS_hela_narrow_5.raw
  - 2018may16_hela_window_size_test_BCS_hela_narrow_6.raw
  - 2018may16_hela_window_size_test_BCS_hela_narrow_7.raw
  - 2018may16_hela_window_size_test_BCS_hela_narrow_8.raw
  - 2018may14_hela_window_size_test_BCS_hela_wide_400_1000_1.raw
  - 2018may14_hela_window_size_test_BCS_hela_wide_400_1000_2.raw
  - 2018may14_hela_window_size_test_BCS_hela_wide_400_1000_3.raw
- FASTA of the Uniprot human reference proteome (reviewed; 20,350 entries)

SUMMARY: Three steps for DIA-MS analysis by chromatogram library

1. Convert .raw files to .mzML using MSConvert
2. Build library using Walnut or XCorDIA in EncyclopeDIA
3. Search wide-window data with library from step 2 using EncyclopeDIA

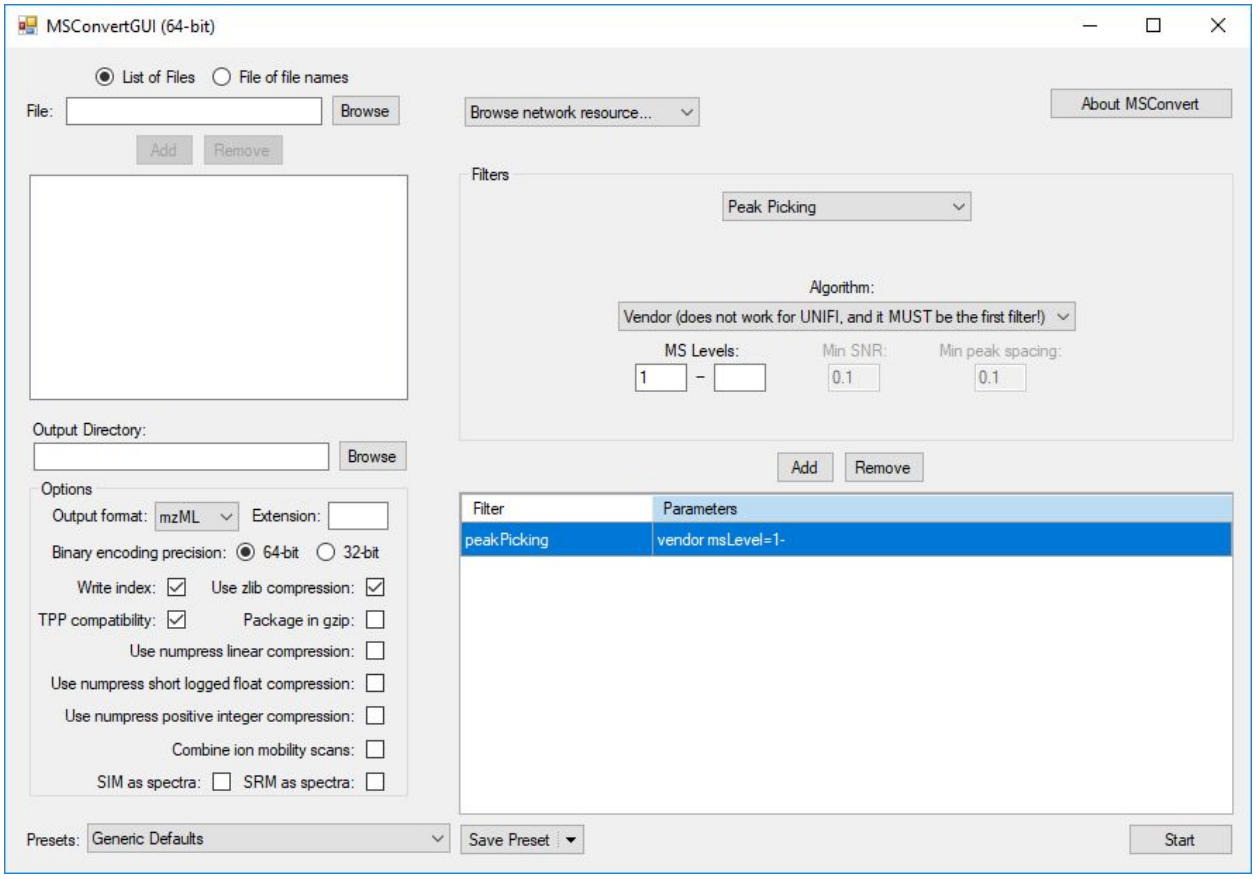
Appendix: Visualization options

A. Skyline
B. EncyclopeDIA viewer
C. Viewing ELIB files with DB Browser for SQLite

GUI-BASED WORKFLOW

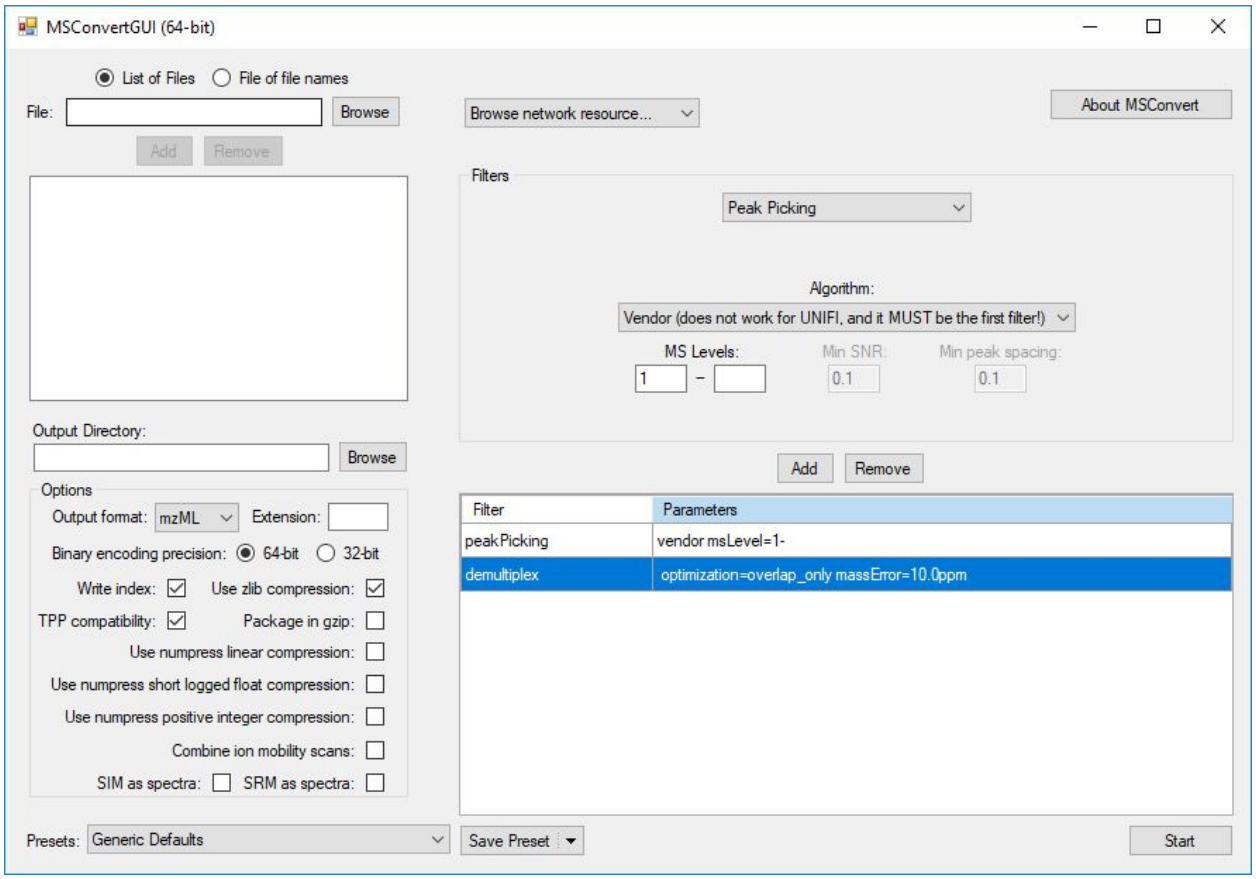
1. Convert .raw files to .mzML using MSConvert

For non-overlapping windows:

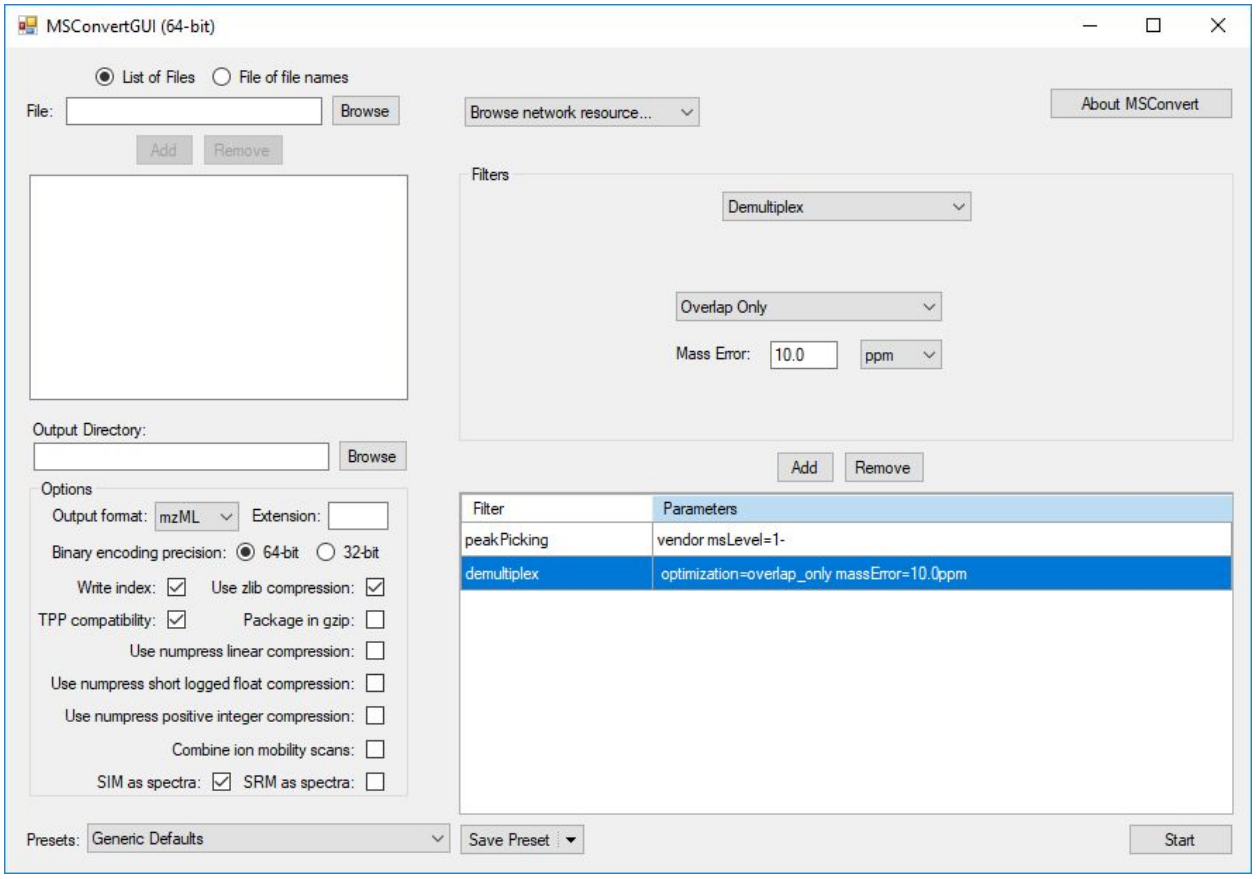


! NOTE: Make sure to have “peakPicking” as the first filter

For overlapping windows:

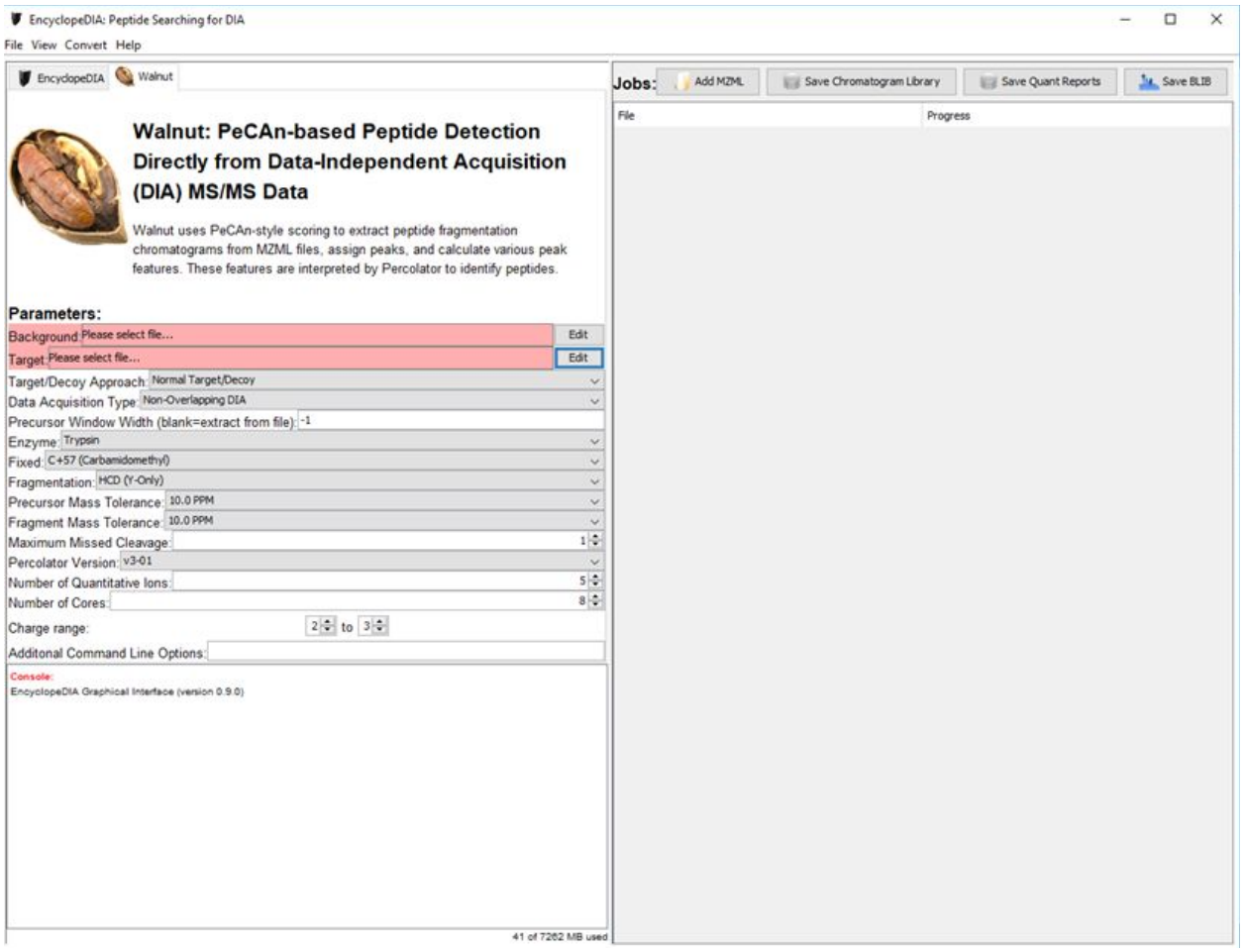


! NOTE: If the MS1 data is acquired as a SIM scan, an acquisition mode generally used for producing chromatograms, you must have another box checked (“SIM as spectra”, bottom left) so the SIM data are written to the mzML file as spectra and not as a wide extracted ion chromatogram. If so, your MSConvert should look like the screenshot below:

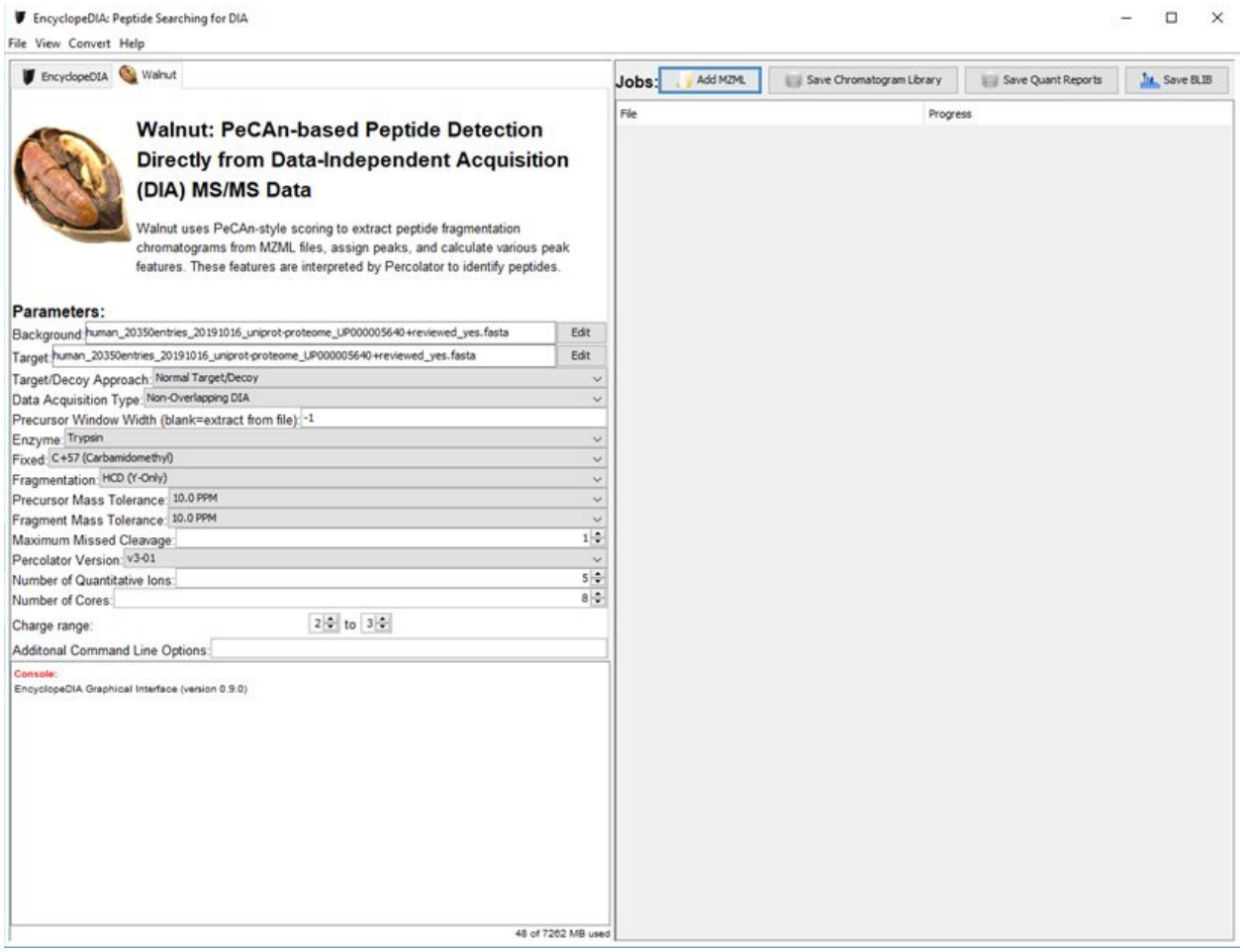


last updated 16 Dec 2019

2. Build chromatogram library using Walnut in EncyclopeDIA



2.1 On the left hand side right underneath the Walnut icon, where it says **Parameters**, find the "**Background**" field and click the corresponding "**Edit**" button to select the background fasta file. The background fasta file should basically be the reference fasta for your model system (*E.coli*, yeast, human, etc). Here, we'll use an example experiment in Hela, so we've downloaded the human reference proteome from Uniprot, and navigated the file explorer to that downloaded fasta file.



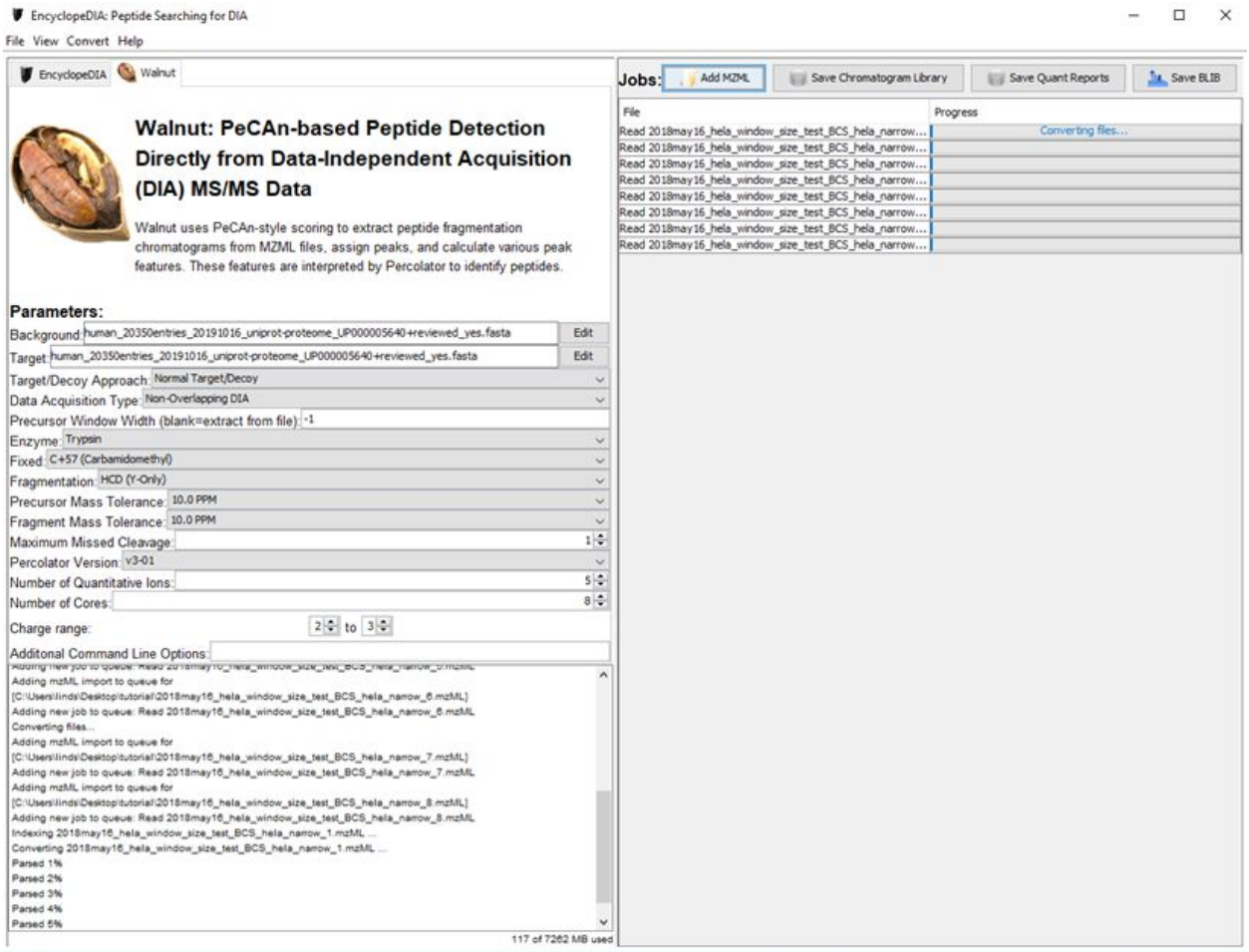
2.2 Again under "**Parameters:**", just underneath the "**Background**" field, which should now contain the filename of that organism fasta, click the "**Target**" field corresponding "**Edit**" button. Navigate to a fasta of your target search proteome. Here, I'm interested in the whole proteome (no specific subcellular fraction like mitochondria) so I'll select the same fasta that I used in the Background field.

! NOTE: *More about the Target/Background fasta: For experiments looking at a "whole proteome" (lysates, for example), both the Target and the Background fasta are the same file (the yeast reference fasta, human reference fasta, etc). For experiments where subcellular fractionation was performed or where you're only interested in some subset of the proteome, use a "Background" fasta of the whole organism and a "Target" fasta just of the proteins you're interested in (for example, a mitochondrial isolation might use the human proteome for a Background fasta, and a MitoCarta fasta that only includes mitochondrial proteins).*

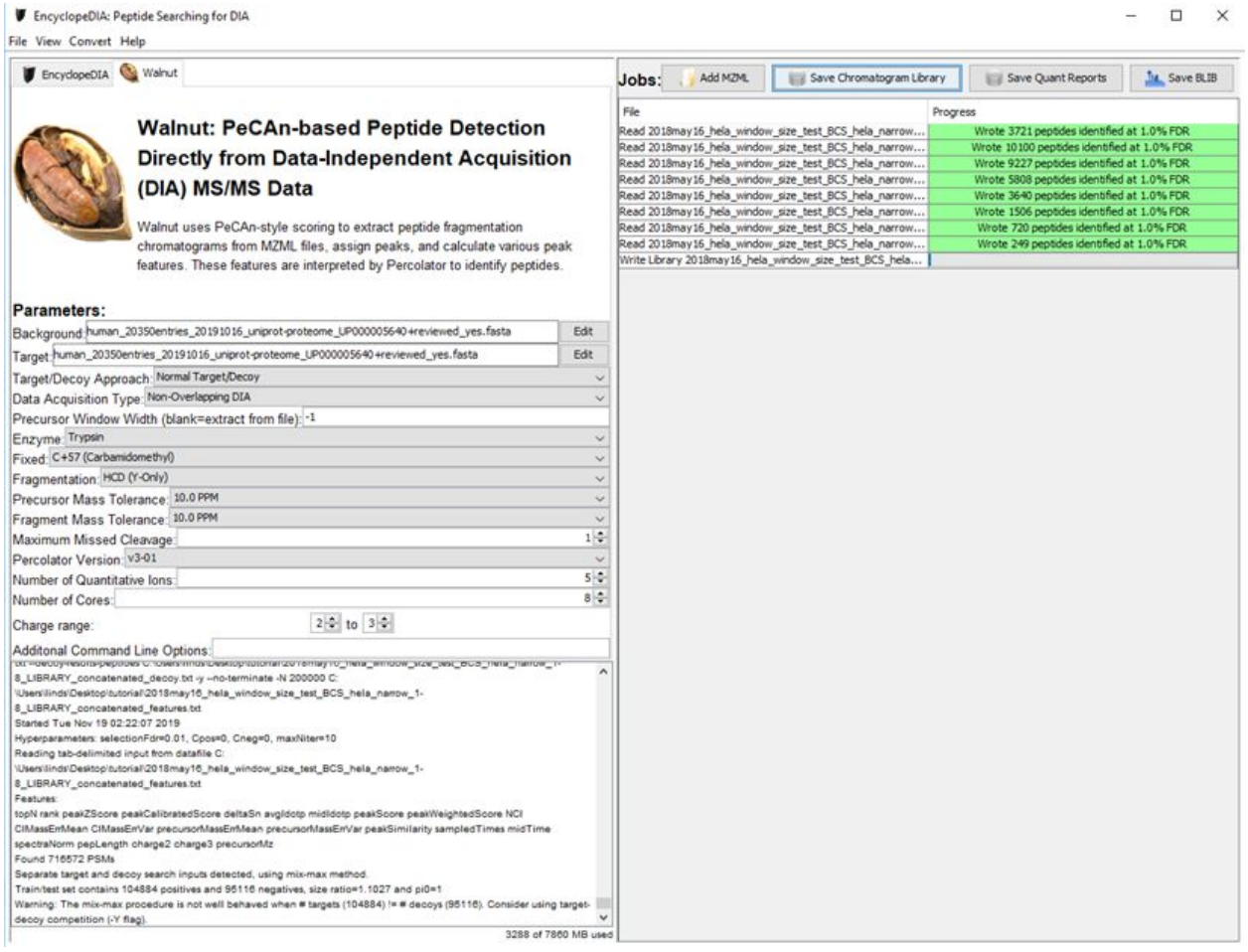
! NOTE *Both Background and Target files are .fasta format. PeCAn users may recall processing a fasta to get a list of peptides for input, but Walnut includes the in silico digest step so you can just give Walnut the .fasta*

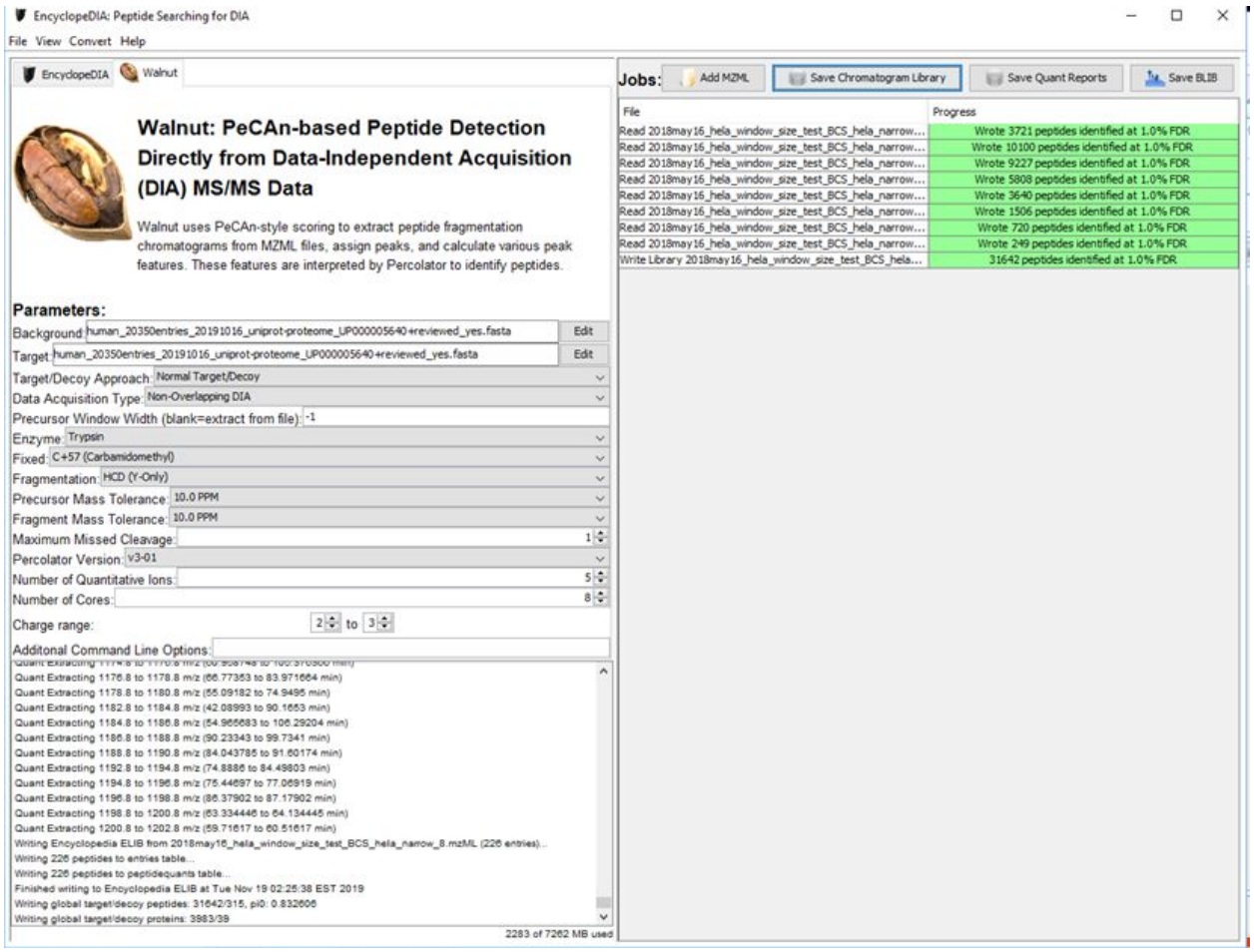
2.3 Set the remaining parameters if you have experiment-specific details that deviate from the defaults (for example, a different digestion enzyme than the default trypsin, or a different fragmentation type than HCD, etc)

2.4 In the top right, where it says "**Jobs:**", click "**Add MZML**". Navigate the file explorer to your converted gas phase fractionated library files from step 1. Select all the gas phase fractionated library MZMLs and click "**Open**". The MZML files you selected should now appear under the "**Jobs**" buttons. You can monitor progress using the GUI.



2.5 When the six gas phase fractionated files have finished running, click "**SAVE CHROMATOGRAM LIBRARY**" and give your library some descriptive filename.

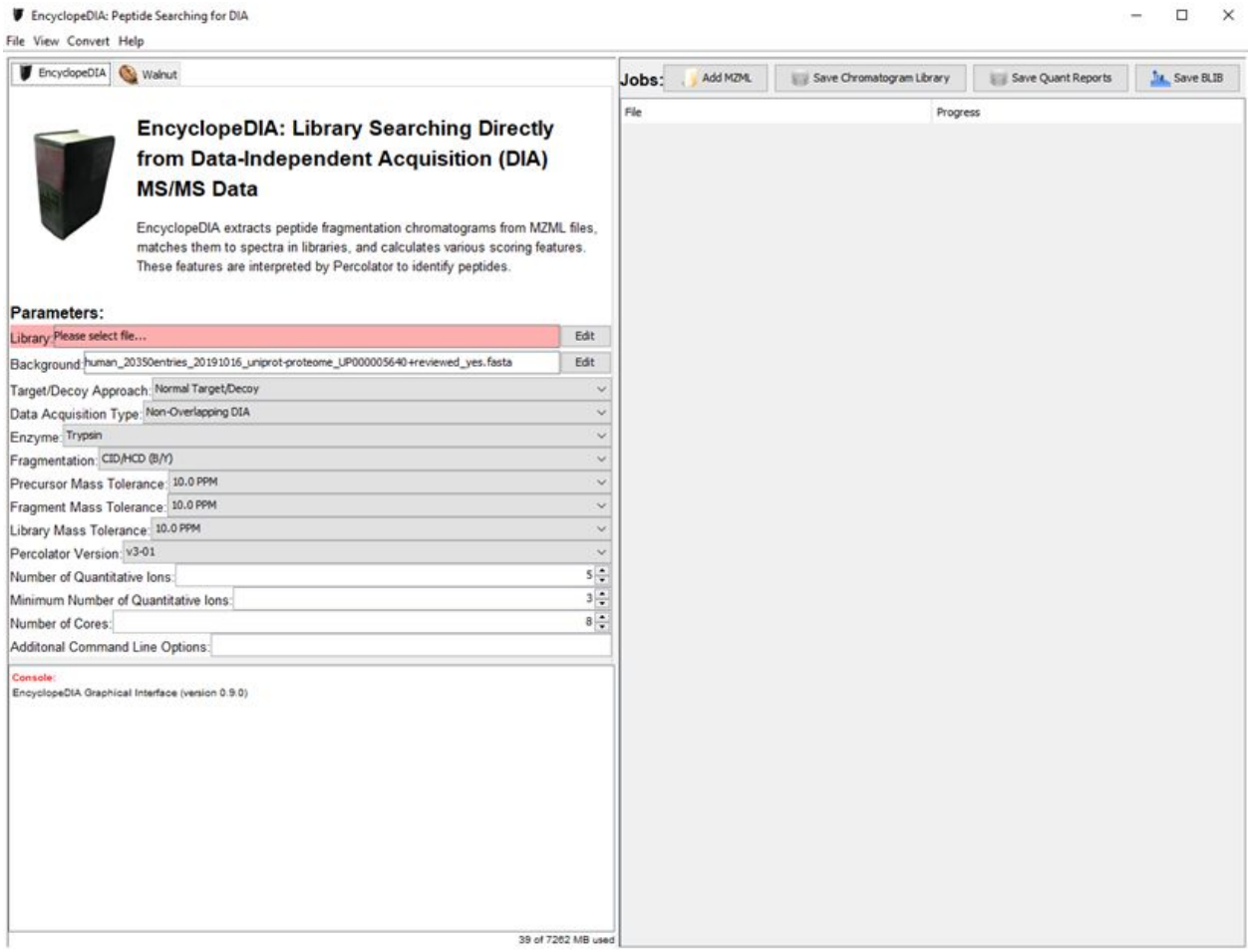




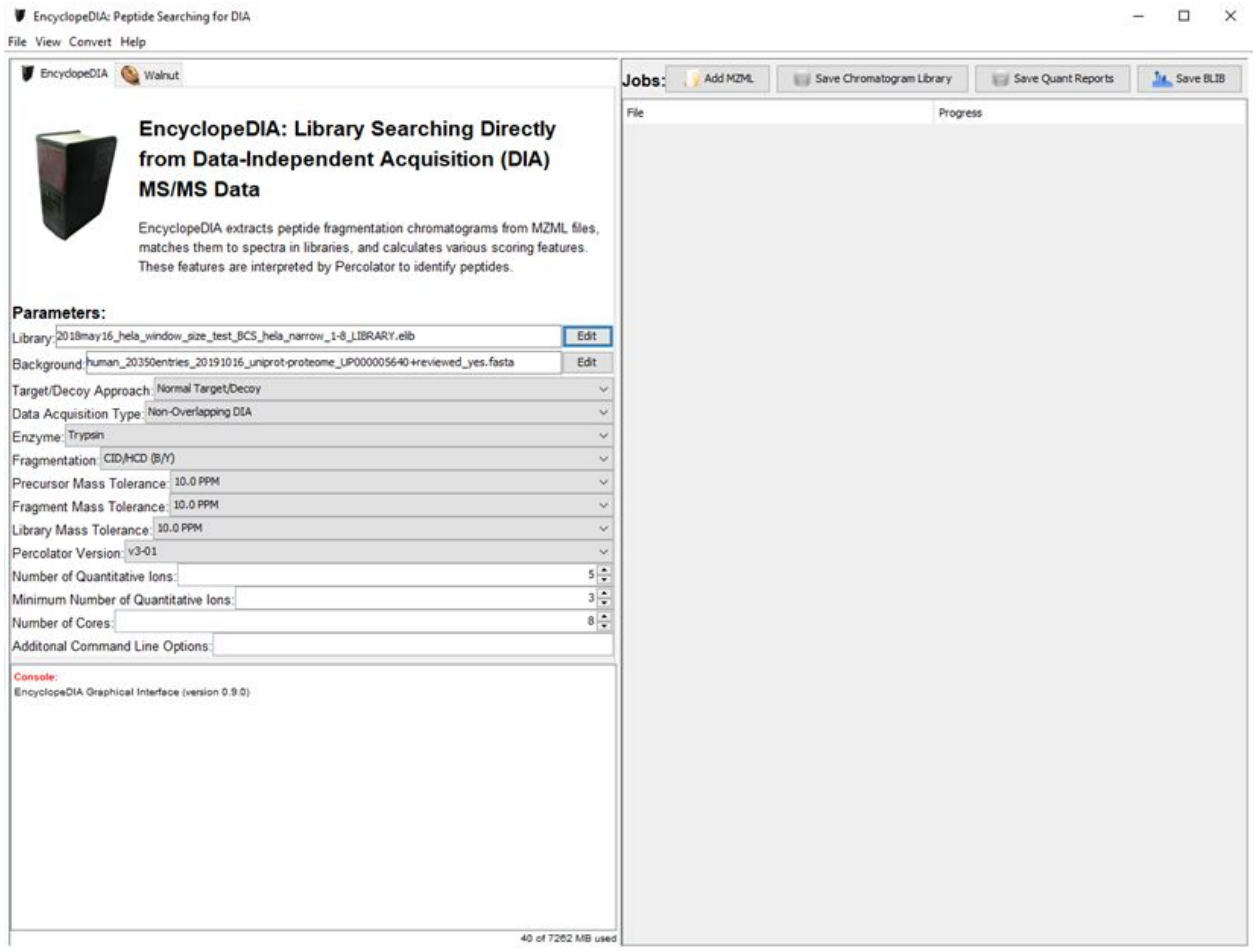
last updated 16 Dec 2019

3. Search wide-window data with chromatogram library from Step 2

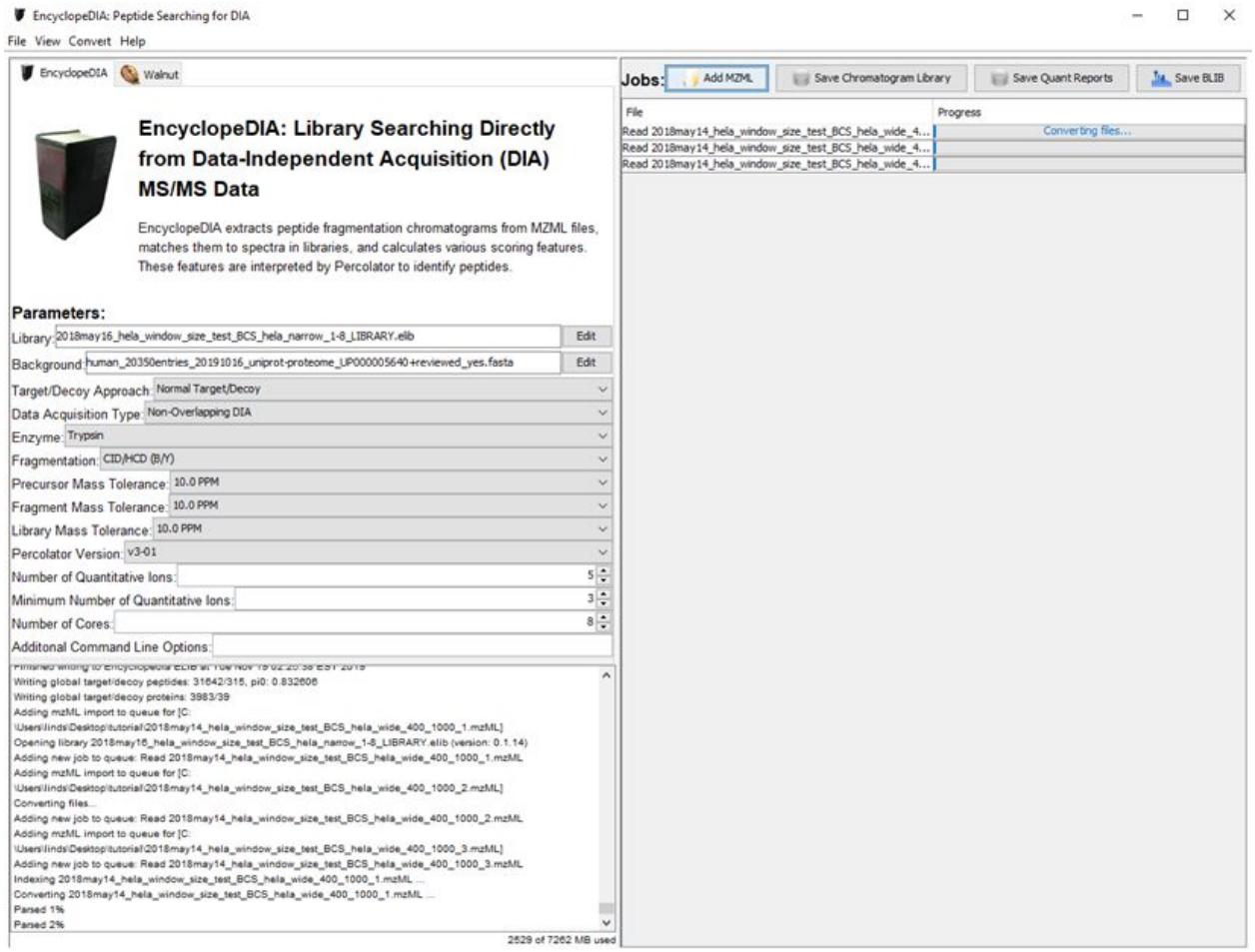
3.1 Close and reopen the EncyclopeDIA GUI to clear EncyclopeDIA's cache/history



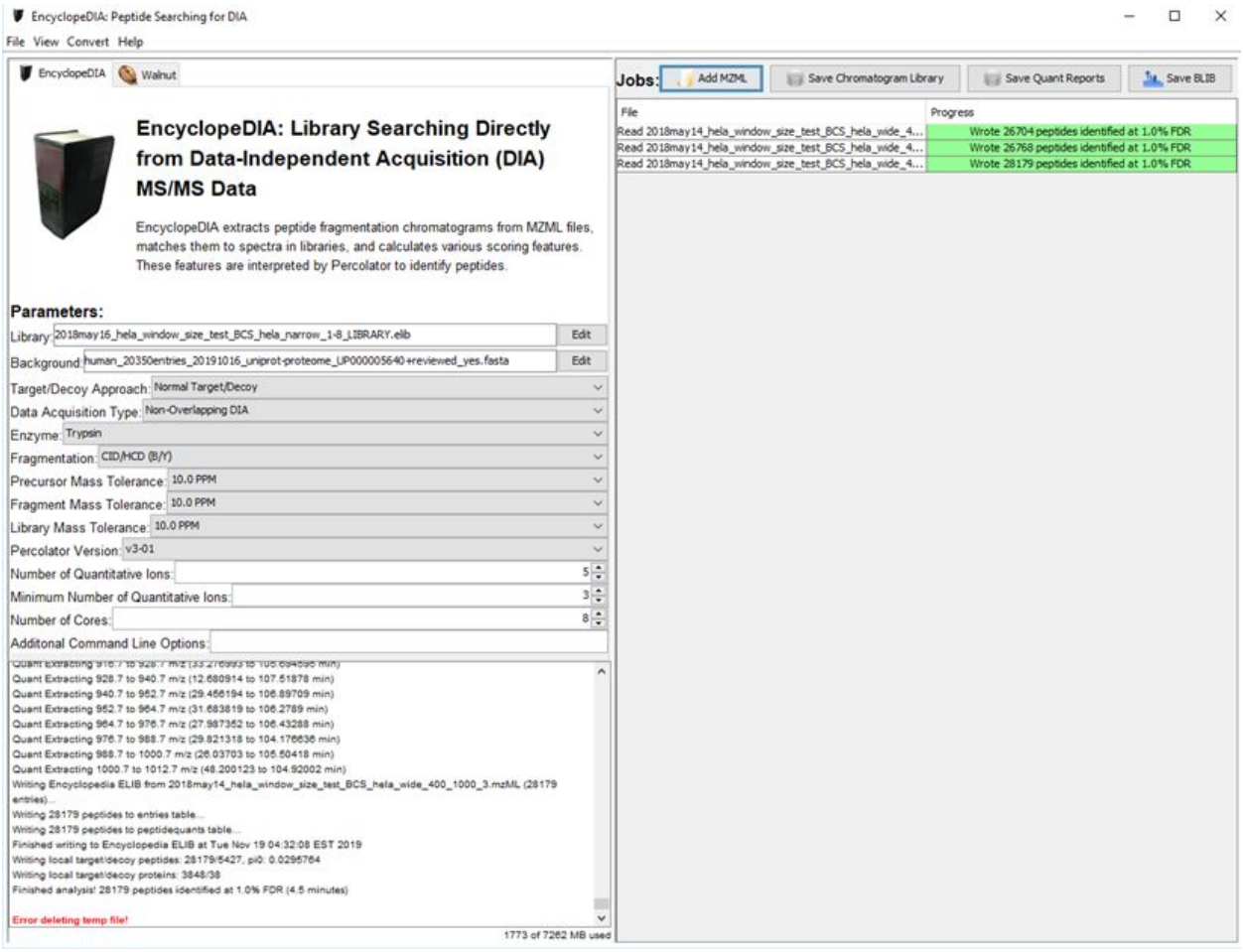
3.2 Within EncyclopeDIA GUI (not Walnut), on the left hand side under “Parameters:” across from the “Library” field, click the “Edit” button. Using the file explorer, select the .elib file you just saved in Step 2.5



3.3 Underneath the “Library” field, across from the “Background” field, click the corresponding “Edit” button and select the appropriate Background file (should be the same fasta you used in 2.1 above!)

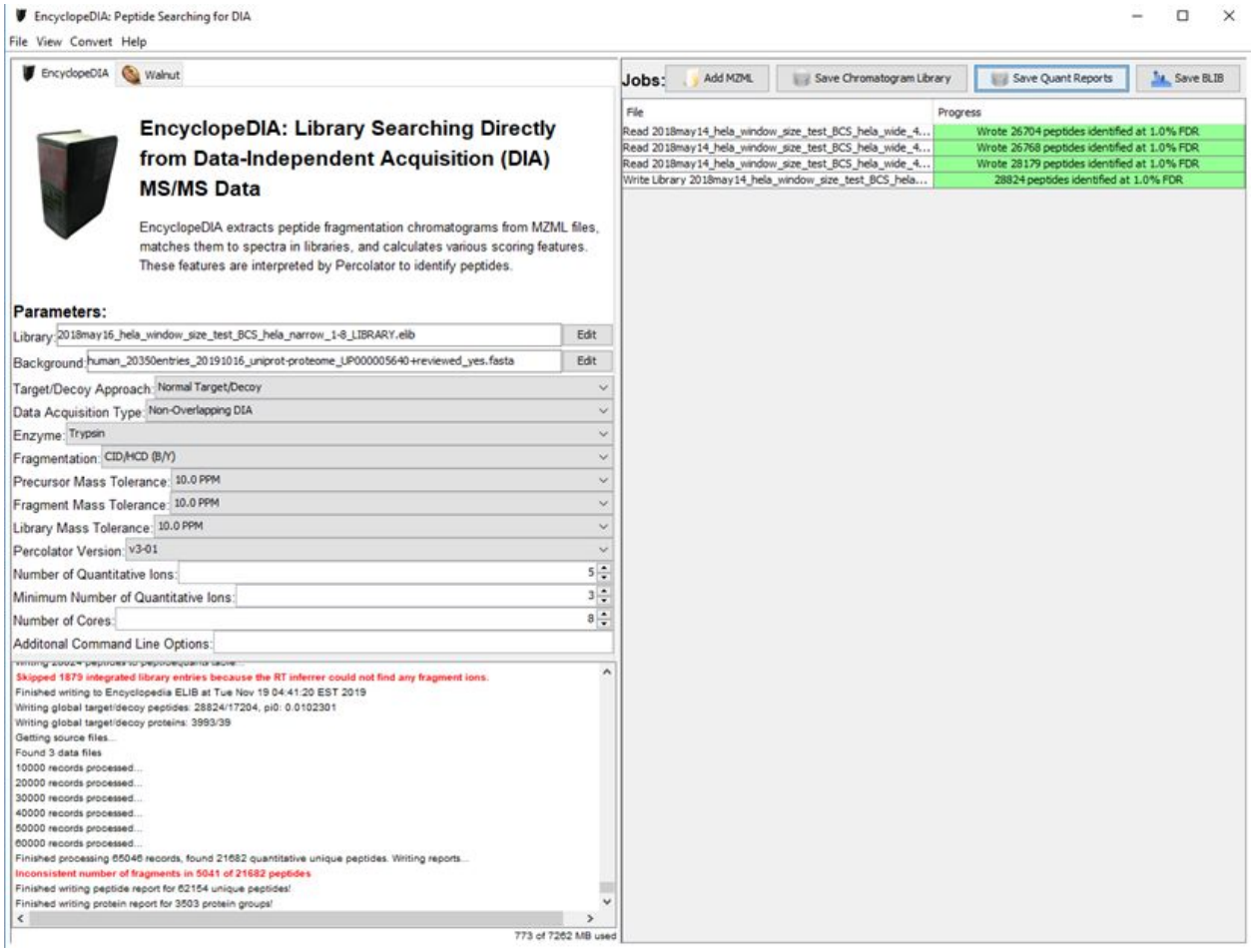


3.4 In the top right, next to “Jobs”, click the “Add MZML” button and select all of the wide-window .mzML files that were acquired using this narrow-window library file.



3.5 Click "**Save Quant Reports"** to perform a final experiment-wide FDR correction and export peptide quant, and protein quant.

- Select "Save Chromatogram Library" to build a file with bonus information like integration boundaries.
    - This is only applicable if you do not want to retention time-align across the MZML files (for example, if your MZMLs are fractionated in a way such that you don't expect to sample the same peptides in each file)
- Select "Save Quant Reports" to get peptide/protein quantitation matrices in the form of a tsv.
    - Pick this if your MZMLs were the experimental samples you want to post-process.
    - If you are following this workflow as-is, this is what you should pick!
- Select "Save BLIB" to build a spectral library file that Skyline can use.
    - This option is effectively depreciated now that Skyline reads ELIB file formats



! NOTE Encyclopedia is determining a lot of information about the DIA experiment. Important details include peptide detections, fragment refinement, and peak boundaries. We'll import that information into Skyline next so that we can visualize the results; however, there's also a visualizer built right into Encyclopedia. See Appendix for details.

! NOTE Encyclopedia's ELIB output can be used in Skyline for visualizing DIA-MS experiments. See Appendix for details.

COMMAND LINE WORKFLOW

1. Convert .raw files into .mzml using MSConvert

For non-overlapping windows:

```
msconvert.exe -v --zlib --64 --mzML --filter "peakPicking true 1-" *.raw
```

For overlapping windows:

```
msconvert.exe -v --zlib --64 --mzML --filter "peakPicking true 1-" --filter "demultiplex optimization=overlap_only" *.raw
```

NOTE: data acquired on Lumos instruments needs an extra flag (`--simAsSpectra`) to convert precursor scans correctly:

```
msconvert.exe --zlib --64 --mzML --filter "peakPicking true 1-" --filter "demultiplex optimization=overlap_only" --simAsSpectra *.raw
```

! NOTE If your file conversion is going slow, you probably aren't using a current version of MSconvert!

2. Build chromatogram library using Walnut (command line)

2.1. Searching DIA data against a FASTA with Walnut: for a full list of options and default values for your version of EncyclopeDIA:

```
java -jar encyclopedia.jar -walnut --help
```

The parameters you may want to change include the enzyme used to prepare the samples,

```
$ java -Xmx<GB_OF_MEM>G -jar encyclopedia.jar -walnut \
  -i <MZML_IN> \
  -f <BACKGROUND_FASTA> \
  -t <TARGET_FASTA> \
  -acquisition DIA \
```

```
  -enzyme <ENZYME> \
  -frag <FRAGMENTATION> \
  -ftol <FRAGMENT_TOLERANCE> \
  -ftolunits <FRAGMENT_TOLERANCE_UNITS> \
  -ptol <PRECURSOR_TOLERANCE> \
  -ptolunits <PRECURSOR_TOLERANCE_UNITS> \
  -minCharge <MIN_CHARGE> \
  -maxCharge <MAX_CHARGE>
```

With a typical DIA setup (trypsin digest, Orbitrap instrument, demultiplexing the RAW file overlapping windows with MSConvert), the command usually looks like this:

```
$ java -Xmx8G -jar encyclopedia.jar -walnut \
  -i DIA_narrow_run_400to500mz.mzML \
  -f human.fasta \
  -t human.fasta \
  -acquisition DIA \
  -enzyme trypsin \
  -frag YONLY \
  -ftol 10.0 \
  -ftolunits ppm \
  -ptol 10.0 \
  -ptolunits ppm \
  -minCharge 2 \
  -maxCharge 3
```

Running Walnut produces several results files, including:

```
<MZML_IN>.dia
<MZML_IN>.mzML.pecan.txt.log
<MZML_IN>.mzML.features.txt
<MZML_IN>.mzML.pecan.txt
<MZML_IN>.mzML.pecan.decoy.txt
```

2.1. Merge Walnut results into a chromatogram library file (.elib): this command must be run within the same directory as the search command from 2a. You have to be in the same directory because EncyclopeDIA will look for the result files in the current

directory in order to compile them into one ELIB, so if the result files aren't there, Walnut has nothing to compile.

For a full list of options and default values for your version of EncyclopeDIA:

```
java -jar encyclopedia.jar -libexport --help
```

`OUTPUT_LIBRARY_NAME:` (filename) Any name of your choice. Entering **`narrow_merged`** here would result in a final output of `narrow_merged.elib` (or `narrow_merged.blib`).
`ALIGN_SPECTRA?:` (**`true`** or **`false`**) You will likely want to set this to **`false`** for this step as you're probably doing a narrow isolation gas phase fractionation. In general, if each of your mzML acquisitions collect an identical precursor range, this should be **`true`** -- otherwise it should be **`false`**.
`USE_BLIB_FLAG:` To export an elib, leave this blank. To export a .blib, set to: **`-blib`**

Typical run:

```
$ java -Xmx<GB_OF_MEM>G -jar encyclopedia.jar -libexport \
  -i <INPUT_DIRECTORY> \
  -o <OUTPUT_LIBRARY_NAME> \
  -a <ALIGN_SPECTRA?> \
  <USE_BLIB_FLAG> \
  -f <BACKGROUND_FASTA> \
  -t <TARGET_FASTA> \
  -ftol <FRAGMENT_TOLERANCE> \
  -ftolunits <FRAGMENT_TOLERANCE_UNITS>
```

Example:

```
$ java -Xmx8G -jar encyclopedia.jar -libexport \
  -i ./ \
  -o narrow_merged \
  -a false \
  -f human.fasta \
  -t human.fasta \
  -ftol 10.0 \
  -ftolunits ppm
```

3. Search wide-window data with chromatogram library from step 2 using EncyclopeDIA (command line)

For a full list of options and default values for your version of EncyclopeDIA:

```
java -jar encyclopedia.jar --help
```

LIBRARY_ELIB_FILE: This should be the result from your narrow library search in step 2.b.ii, e.g. **path/to/your/narrow_merged.elib**

Typical run:

```
$ java -Xmx<GB_OF_MEM>G -jar encyclopedia.jar \
  -l <LIBRARY_ELIB_FILE> \
  -i <MZML_IN> \
  -f <BACKGROUND_FASTA> \
  -t <TARGET_FASTA> \
  -acquisition DIA \
  -enzyme <ENZYME> \
  -frag <FRAGMENTATION> \
  -ftol <FRAGMENT_TOLERANCE> \
  -ftolunits <FRAGMENT_TOLERANCE_UNITS> \
  -ptol <PRECURSOR_TOLERANCE> \
  -ptolunits <PRECURSOR_TOLERANCE_UNITS> \
  -minCharge <MIN_CHARGE> \
  -maxCharge <MAX_CHARGE>
```

Output files:

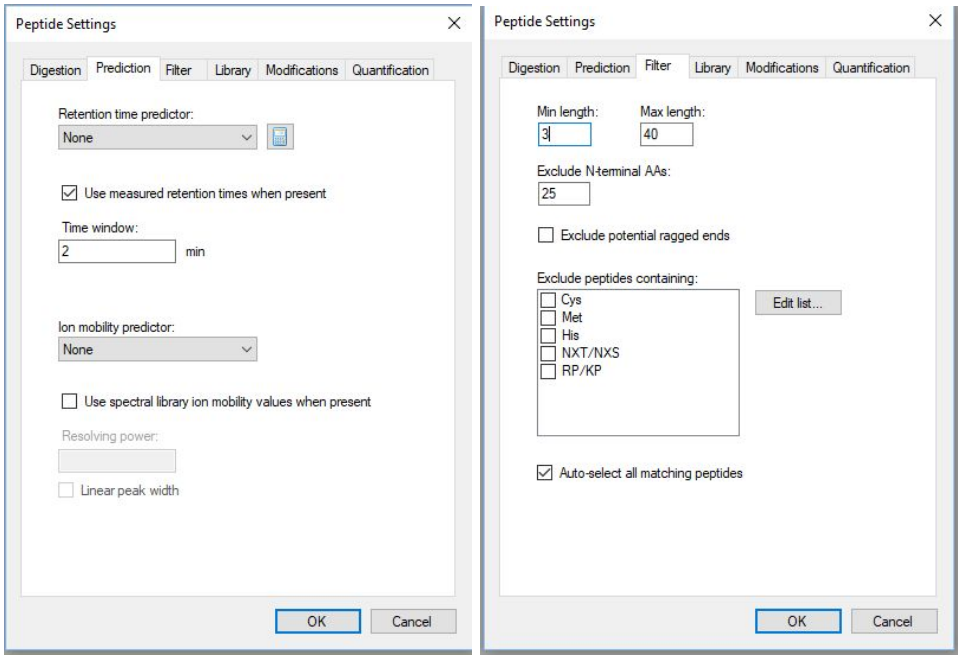
```
<MZML_IN>.dia
<MZML_IN>.mzML.elib
<MZML_IN>.mzML.encyclopedia.txt
<MZML_IN>.mzML.encyclopedia.txt.delta_rt.pdf
<MZML_IN>.mzML.encyclopedia.txt.log
<MZML_IN>.mzML.encyclopedia.txt.rt_fit.pdf
<MZML_IN>.mzML.encyclopedia.txt.rt_fit.txt
```

<MZML_IN>.mzML.features.txt
<MZML_IN>.mzML.first_round.txt

Appendix

Importing peptide detections into Skyline

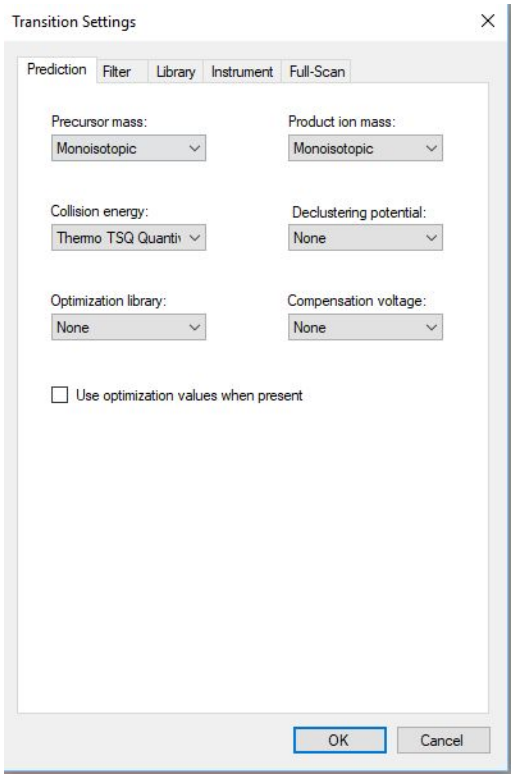
Settings > Peptide Settings.

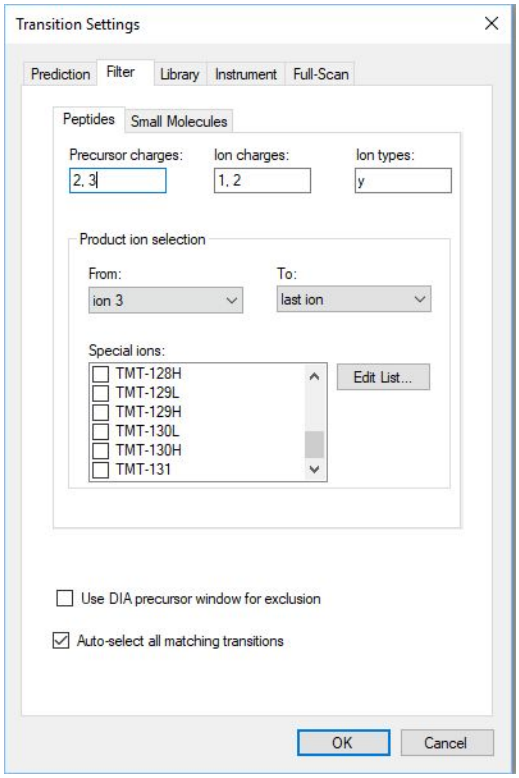


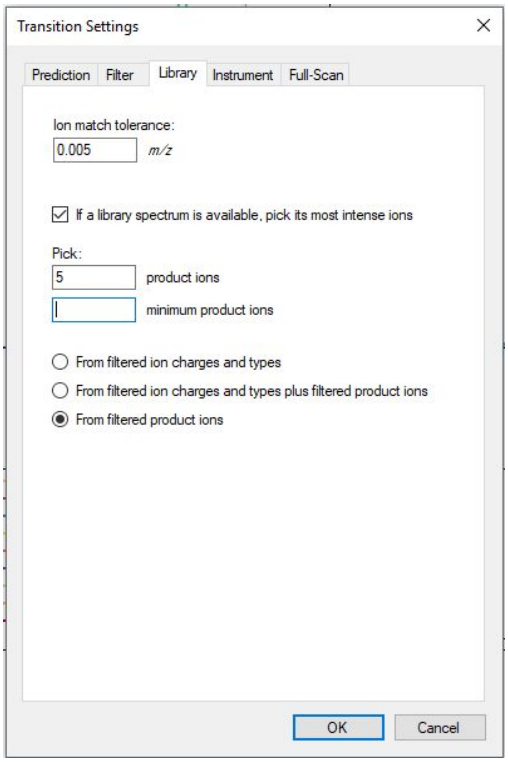
4a. Prediction: Check “Use measured retention times when present”, Time window=2

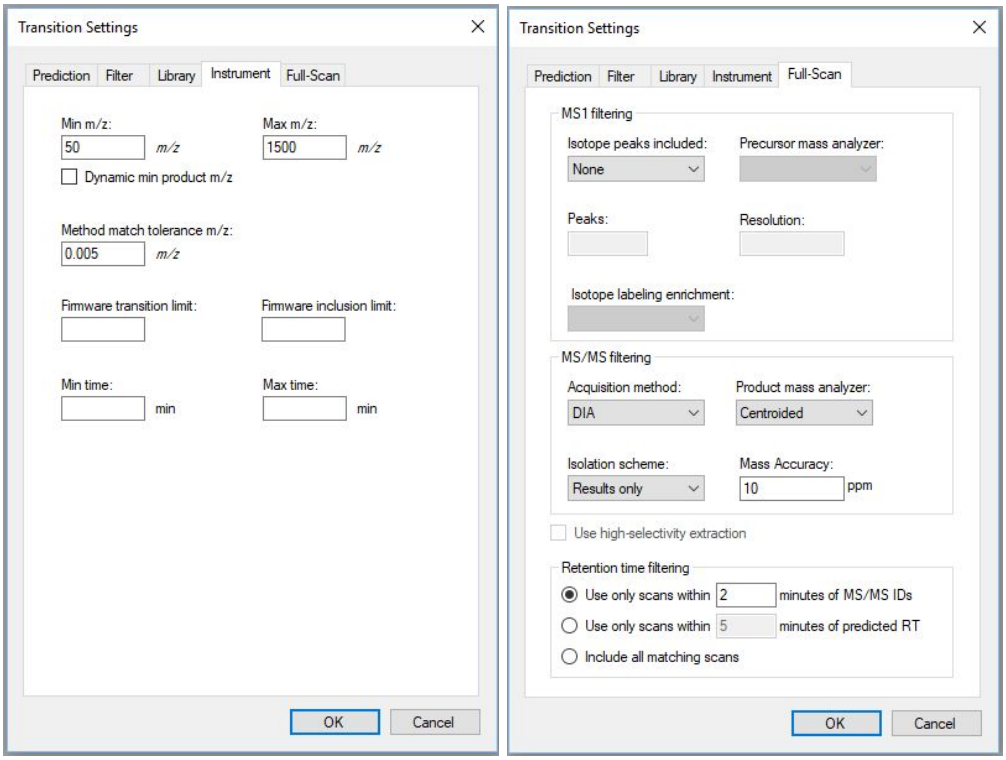
4b. Filter: Min length=3, Max length=40, no excluded amino acids checked

Settings > Transition Settings.


Transition Settings
Prediction
Filter
Library
Instrument
Full-Scan
Precursor mass:
Monoisotopic
Product ion mass:
Monoisotopic
Collision energy:
Thermo TSQ Quantiv
Declustering potential:
None
Optimization library:
None
Compensation voltage:
None
Use optimization values when present
OK
Cancel


Transition Settings
Prediction
Filter
Library
Instrument
Full-Scan
Peptides
Small Molecules
Precursor charges:
2, 3
Ion charges:
1, 2
Ion types:
y
Product ion selection
From:
ion 3
To:
last ion
Special ions:
TMT-128H
TMT-129L
TMT-129H
TMT-130L
TMT-130H
TMT-131
Edit List...
Use DIA precursor window for exclusion
Auto-select all matching transitions
OK
Cancel


Transition Settings
Prediction
Filter
Library
Instrument
Full-Scan
Ion match tolerance:
0.005
m/z
If a library spectrum is available, pick its most intense ions
Pick:
5
product ions
minimum product ions
From filtered ion charges and types
From filtered ion charges and types plus filtered product ions
From filtered product ions
OK
Cancel



4c. Prediction: Precursor/Product ion mass=”Monoisotopic”

4d. Filter: Ion charges=1,2 Ion types=”y”*, From=Ion 3, To=last ion, no special ions

*Set “ion types” to reflect how you searched the data in Encyclopedia!

4e. Library: Ion match tolerance=0.005 m/z, check “If a library spectrum is available, pick its most intense ions”, pick=5 product ions, select “From filtered product ions”

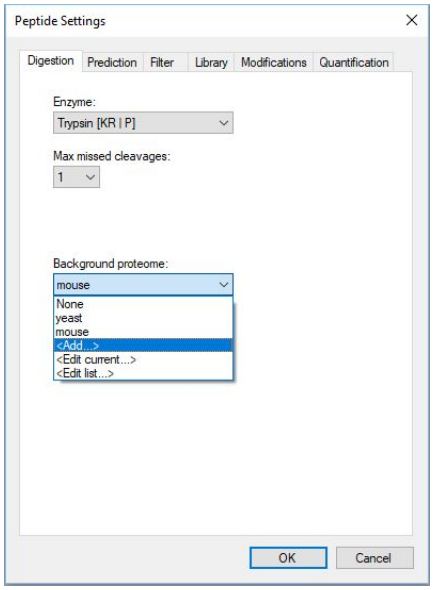
4f. Instrument: “Min m/z=50, Max m/z=1500, Method match tolerance m/z=0.005

4g. Full-Scan (MS1): “Isotope peaks included=None”

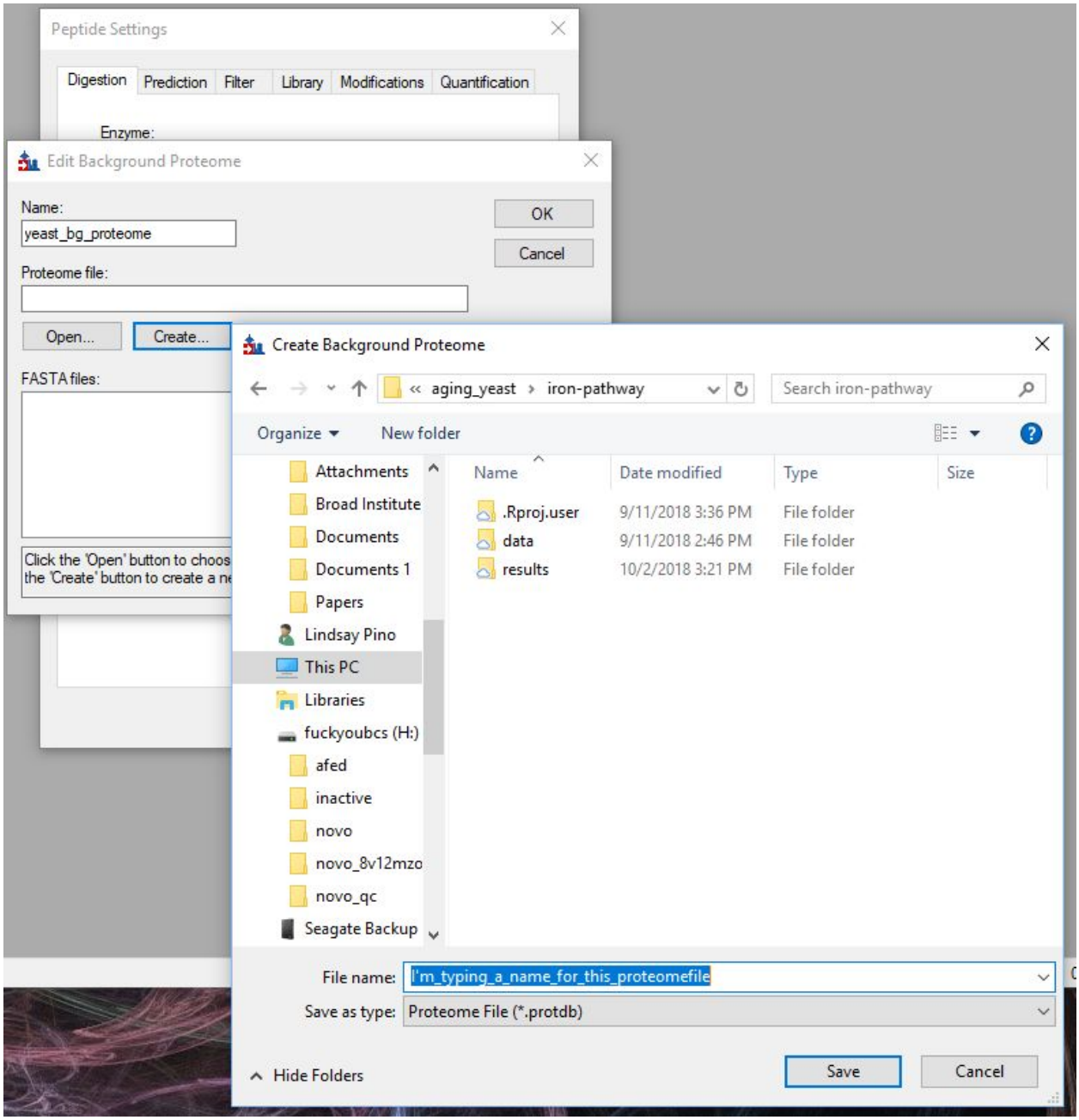
4h. Full-Scan (MS/MS): “Acquisition method=DIA, Product mass analyzer=Centroided, Isolation scheme=Results only, Mass Accuracy=10 ppm, check “Use only scans within “2” minutes of MS/MS IDs

Load the FASTA database and the ELIB library from EncyclopeDIA into Skyline.

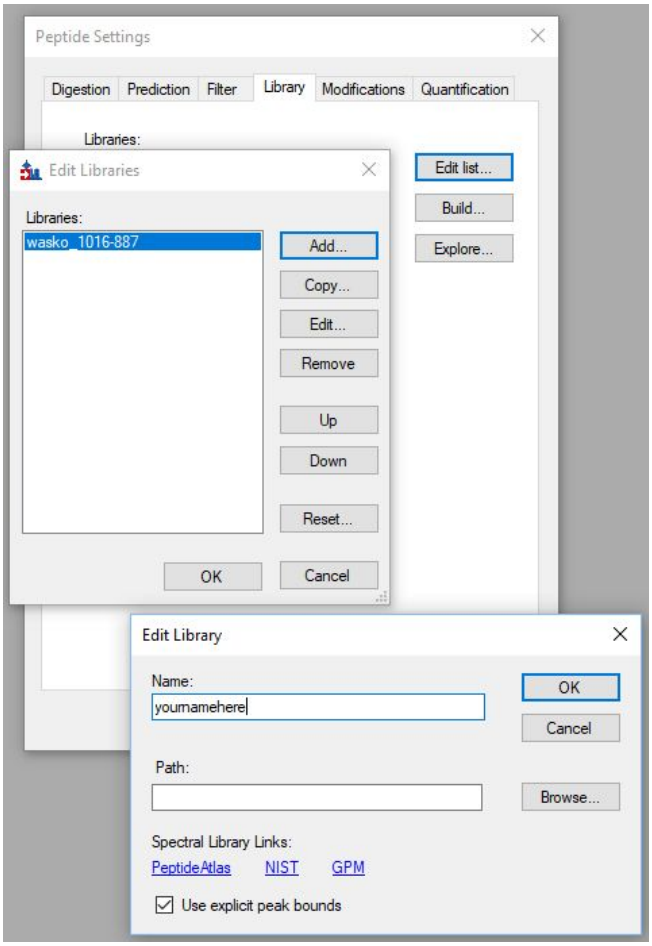
4i. Open Settings/Peptide Settings/Digestion. Select “<Add...>” for Background proteome



4j. Set Name=”[yournamehere]”, then click “Create…” and type in a descriptive name for the soon-to-be-created Proteome File and click “**Save**”. Select “Add File…” to add the background fasta you were just using in the Encyclopedia suite.



4k. Open Settings/Peptide Settings/Library. Select “Edit list…”, then “Add…”. Set the name to be “[yournamehere]” and select “Browse…” to select the .elib that you saved in EncyclopeDIA from the wide-window search (not the gas phase fractionated, narrow window library). “Ok” out to the Peptide Settings.



4l. Check the new “[yournamehere]” library and uncheck all other libraries. Click “Explore…” to view it

> ! If you get a pop up warning that ‘Peptide settings have been changed. Save changes?’ then click “Yes”

4m. Check “Associate proteins” and click “Add All…”. Select “Add to all matching proteins” and “Include all peptides” and hit “OK”.

4n. If you have more than one library (*.elib) for your experiment, you must select each of those libraries from the dropdown menu in "Spectral Library Explorer" and repeat step 4m for each.

4o. Save the Skyline file!

4p. Import .mzML data into Skyline

> ***!*** *Tip: Use the .mzml files, not the .raw files! If you prefer using the .raw files for some reason, you will need to set the windowing scheme parameters in the "MS/MS filtering" box under the "Full Scan" tab in Skyline's "Transition Settings".*

Skyline: Export report for MSstats

To load the MSstats report format, install the MSstats external tool. Alternatively, you can quickly build a custom report using the little black binoculars to find whatever fields you're interested in exporting.

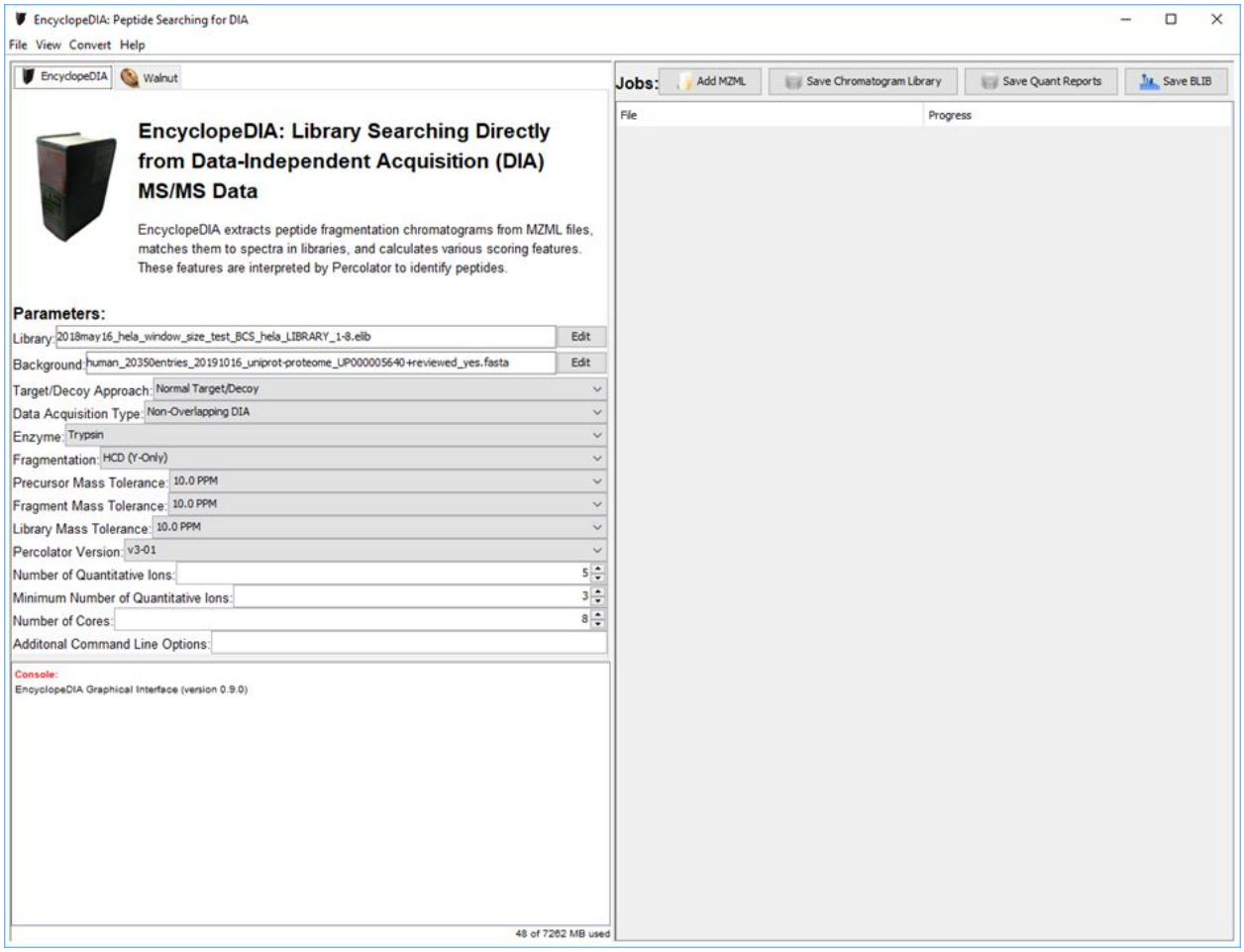
4q. File > Export > Report...

4r. Edit list... > Add...

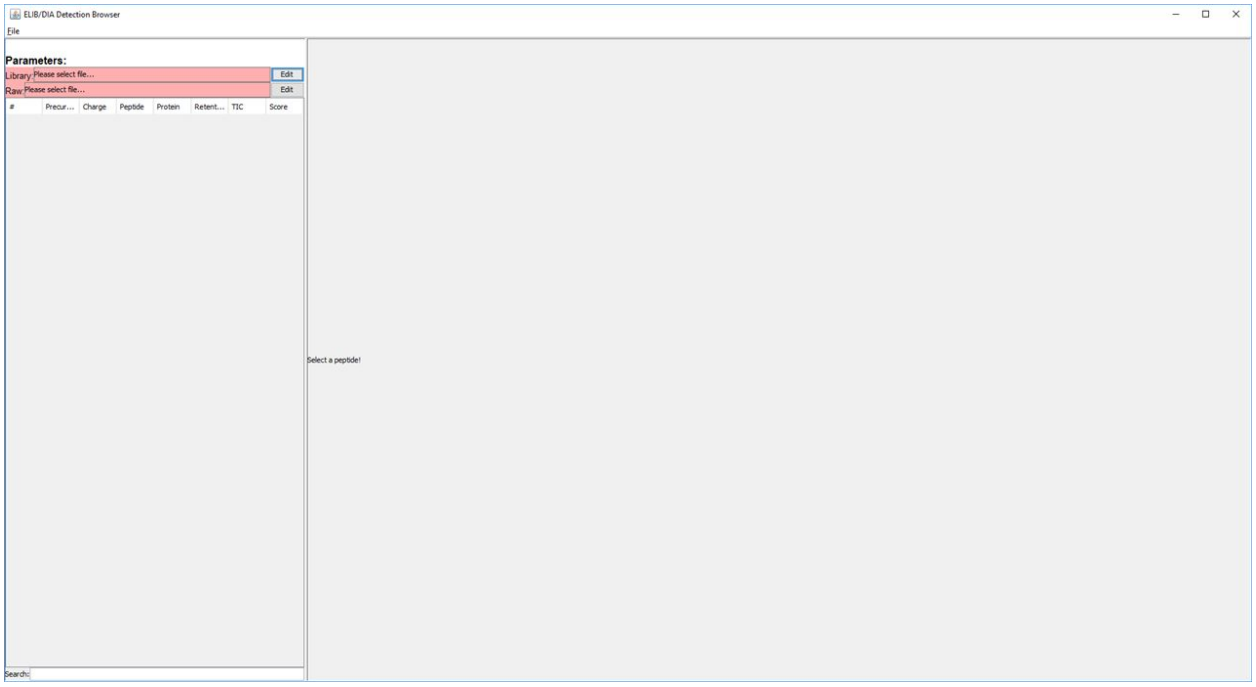
4s. Make sure to name your report ("View Name:"), and check the fields you want the report to include. Binoculars are at the top left, next to the Redo arrow, and will find fields by name.

EncyclopeDIA data visualizers

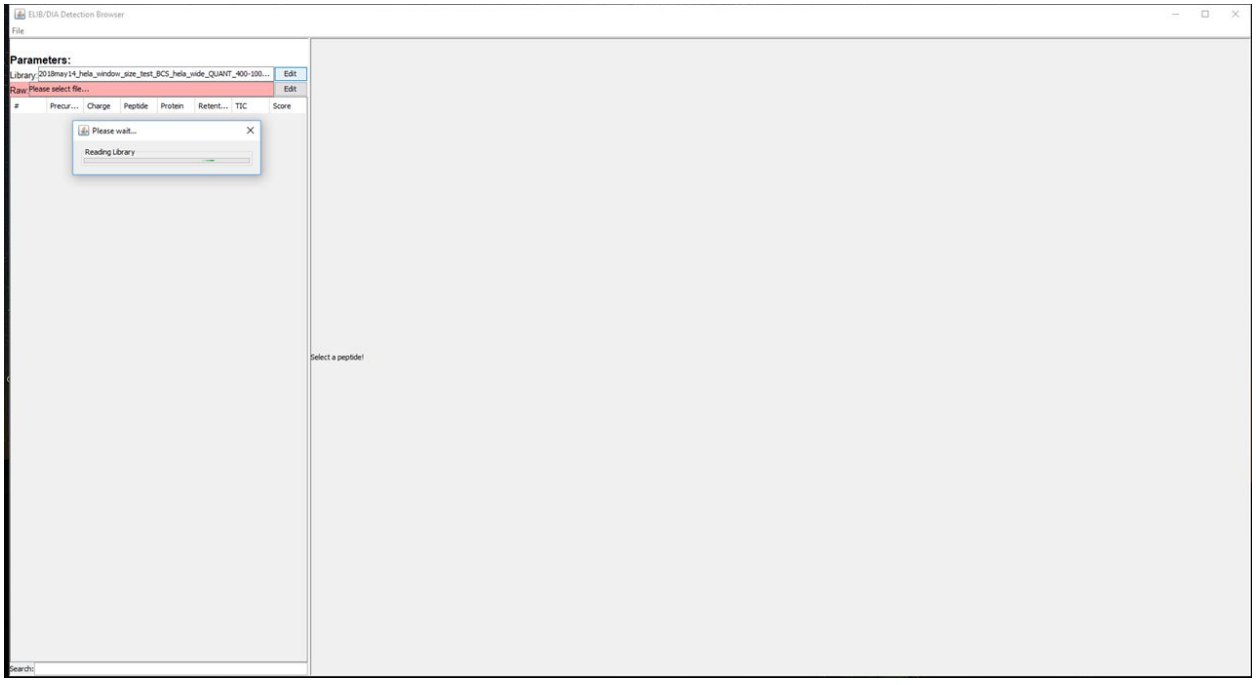
EncyclopeDIA ELIB Browser

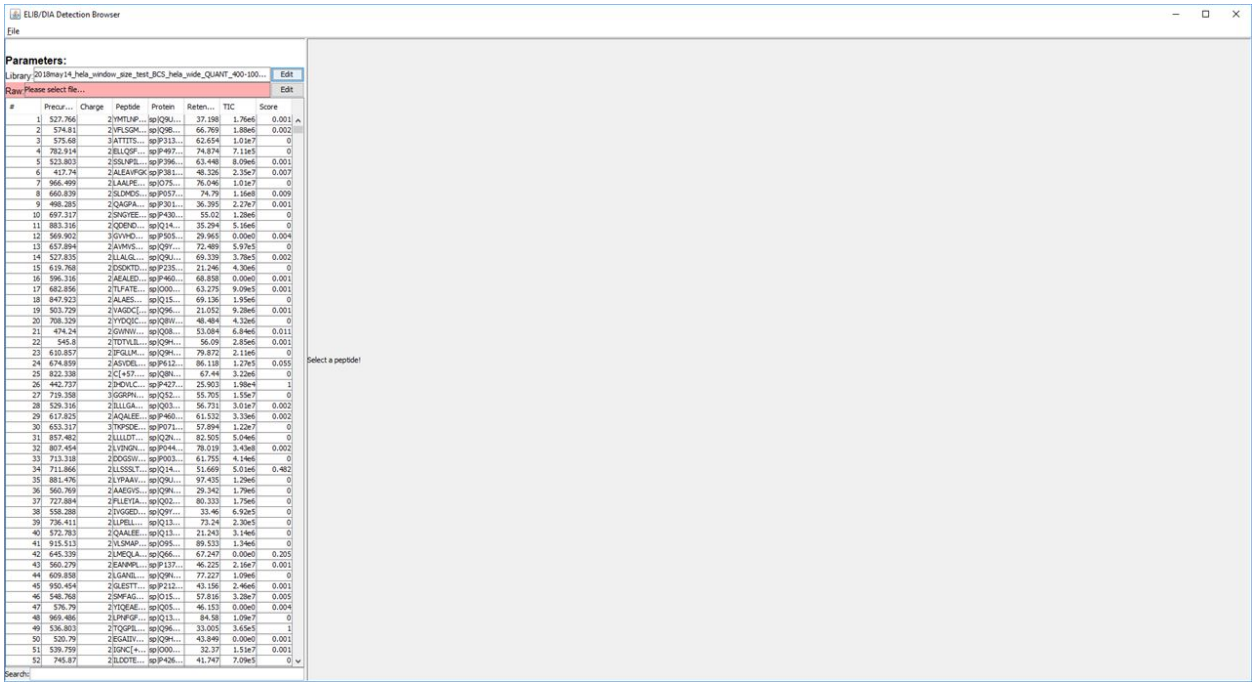


The top left under “View”, select the “Launch ELIB Browser” option

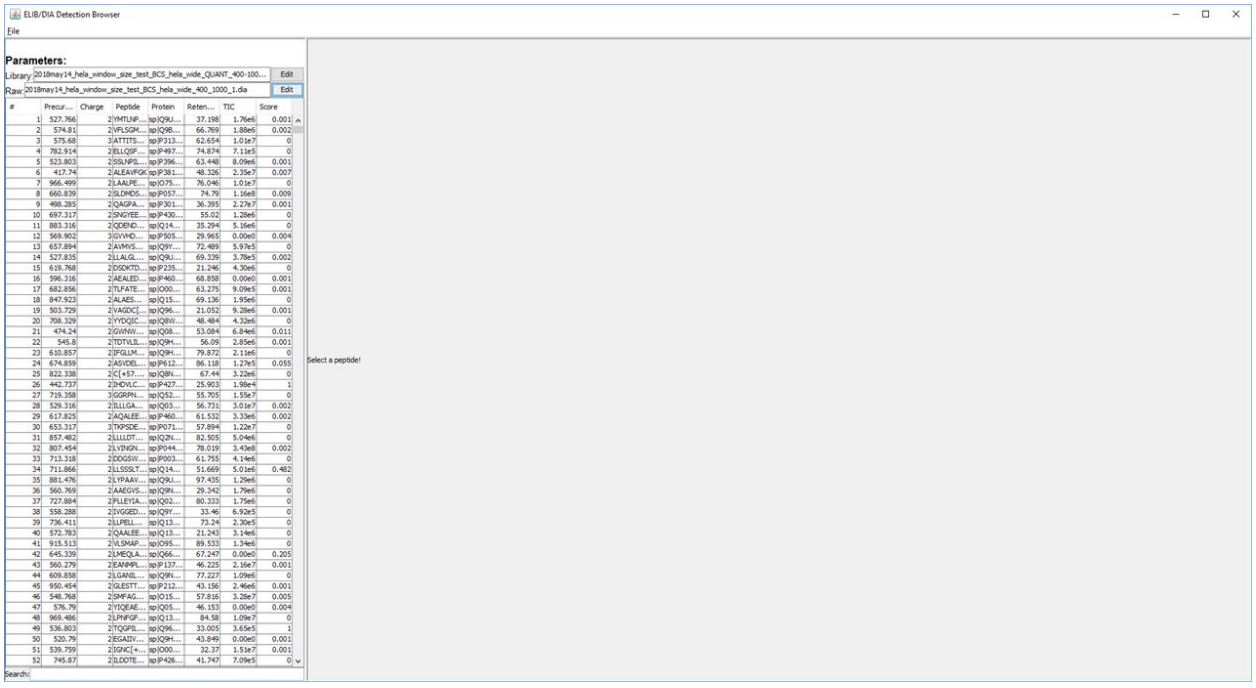


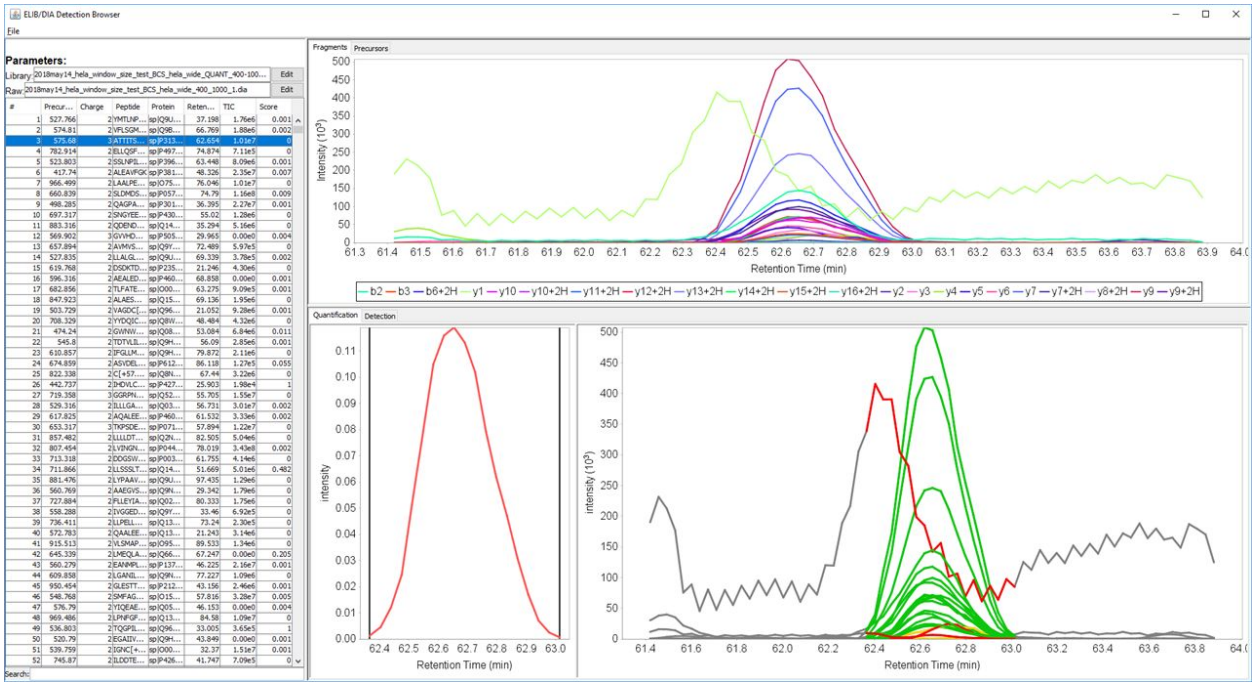
Next to the “Library” field, click the “Edit” button and navigate the explorer to the “Save Chromatogram Library” file you just saved. Once the library has loaded, you should have a populated target list like this:



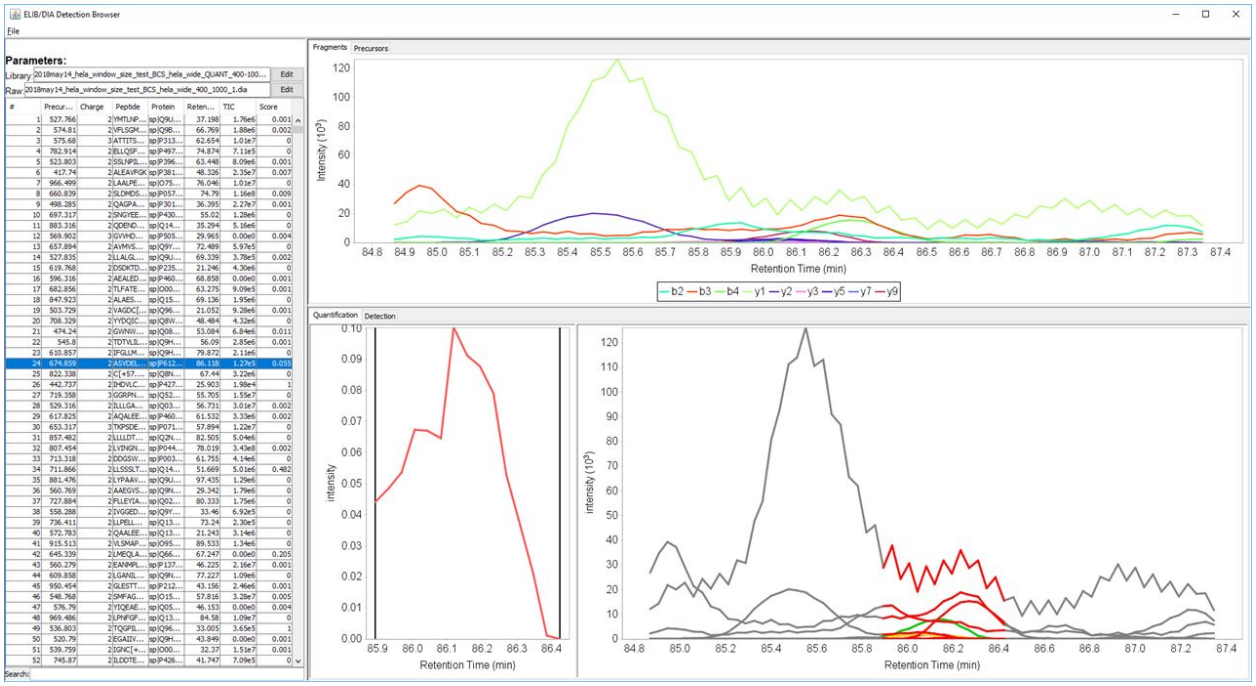


Next to the “Raw” field, click the “Edit” button to select one of the RAW files analyzed in that chromatogram library.





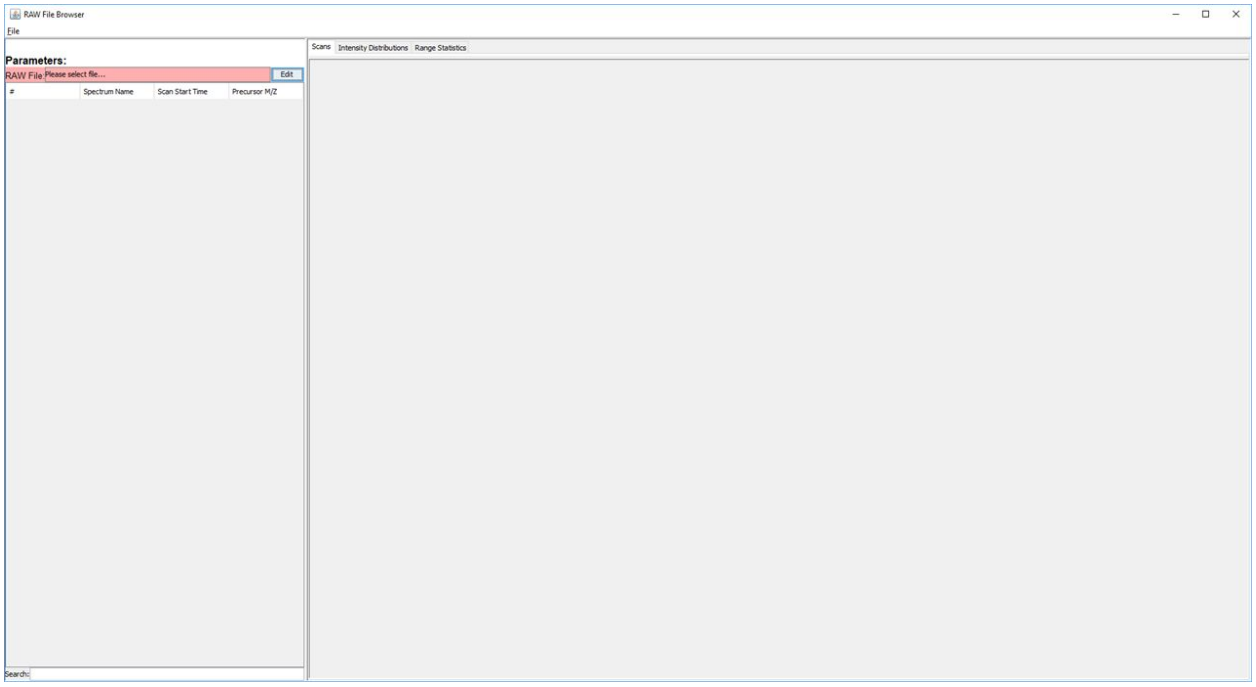
Above is an example of a “good” peptide, where Encylopedia found lots of interference-free fragments (all the fragment traces in the bottom right are green). The fragment ion chromatograms shown in red don’t follow the average profile of the peak group.



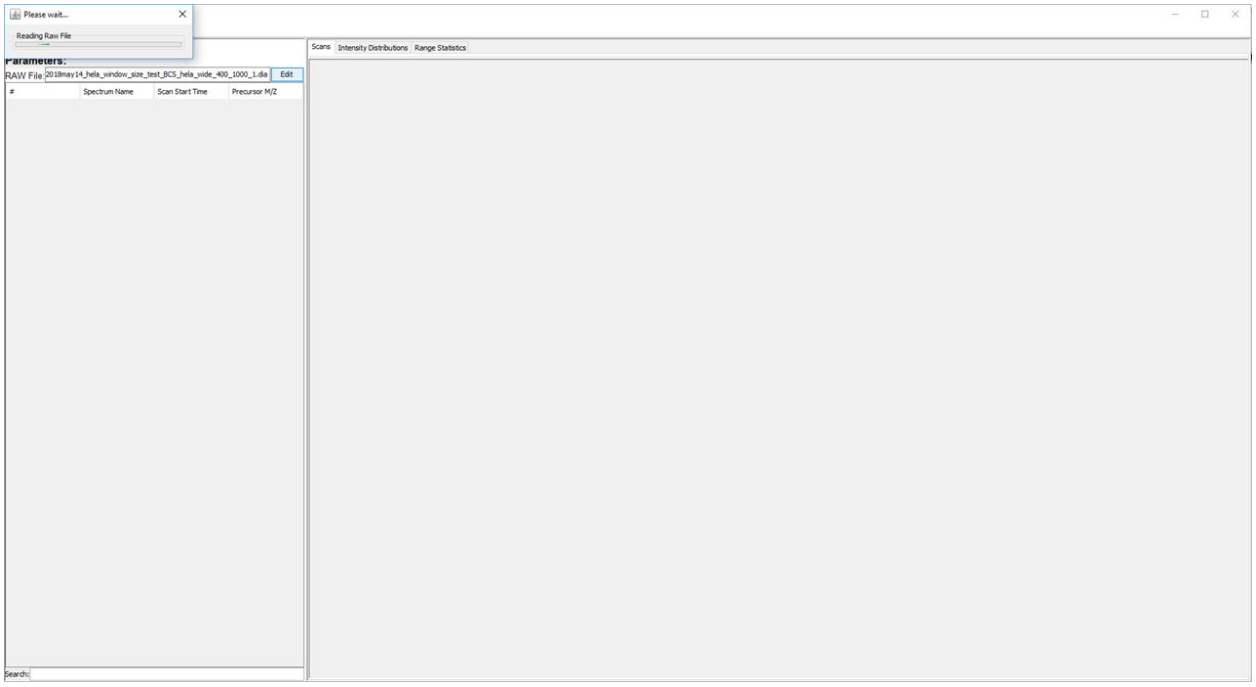
Above is an example of peptide that was detected but doesn’t look quantitative.

EncyclopeDIA RAW File Browser

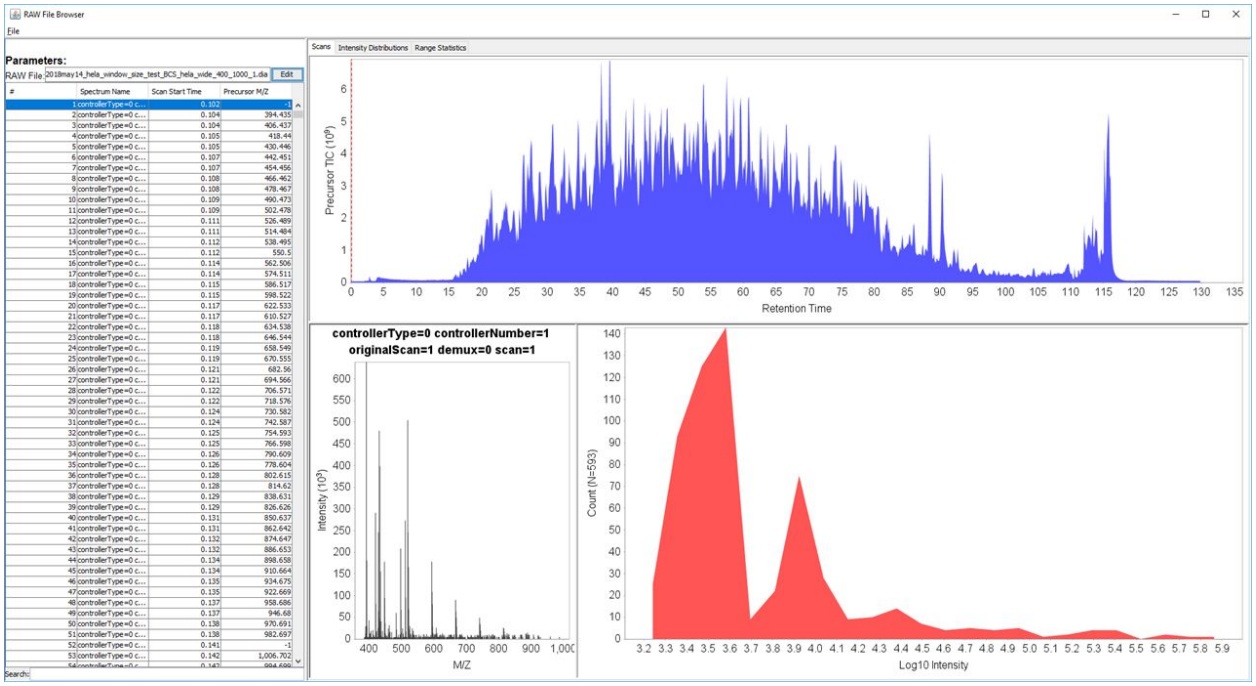
Top left under “View”, select “Launch RAW File Browser”



Select a “RAW” file that has been previously analyzed with the Encyclopedia suite (Walnut or Encyclopedia)

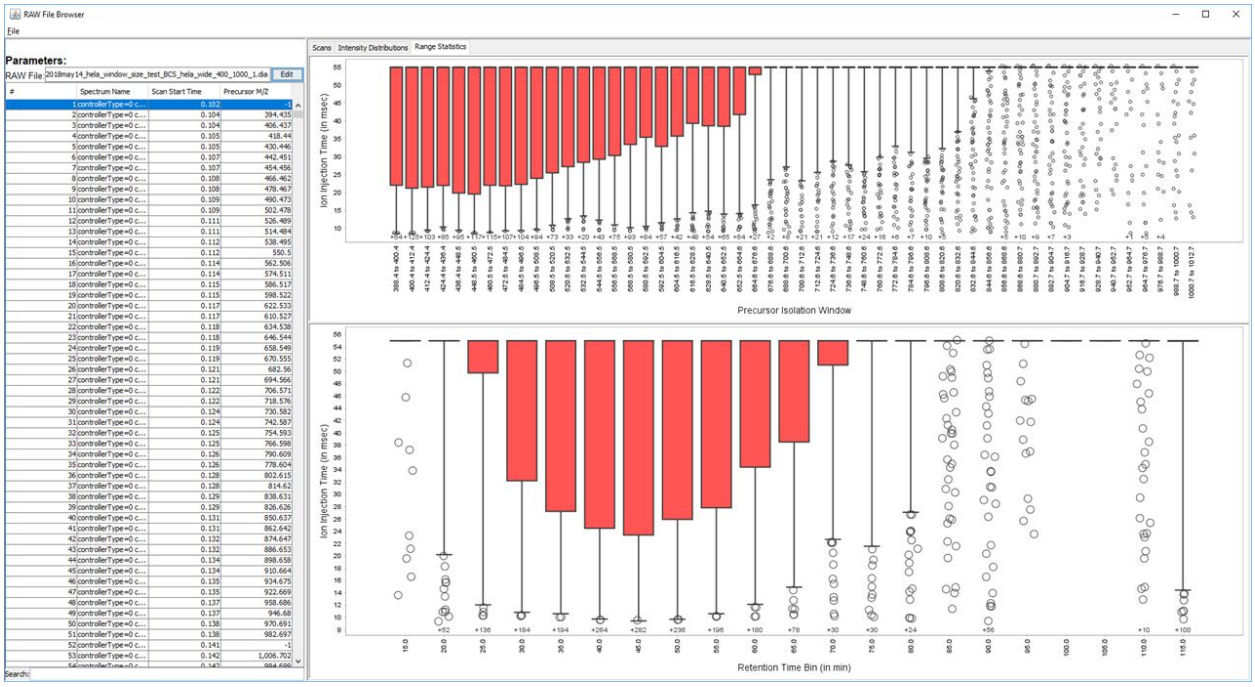


last updated 16 Dec 2019



The target list on the left lists the Scan Number (#), SpectrumName, Scan Start Time, and Precursor M/Z.

On the top above the graphics, there are three tabs. The “Range Statistics” gives some valuable information about the DIA method like the Ion Inject Time across each precursor window and across retention time bin:



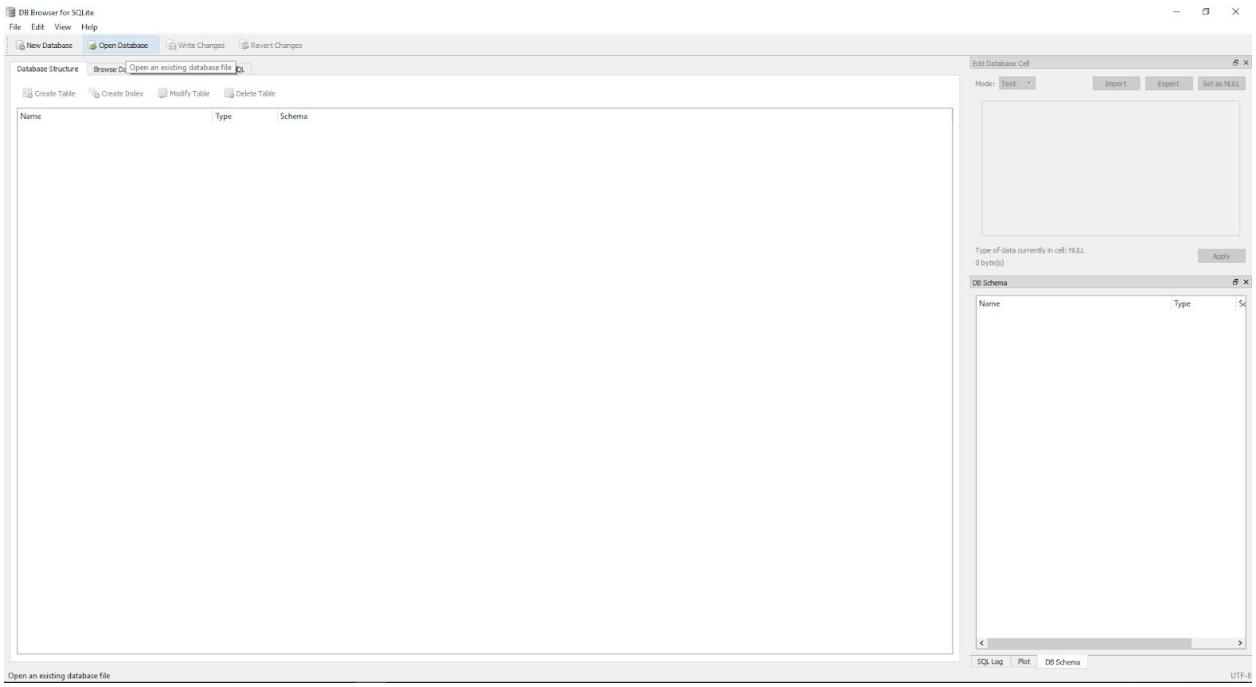
last updated 16 Dec 2019

Viewing ELIB files with DB Browser for SQLite

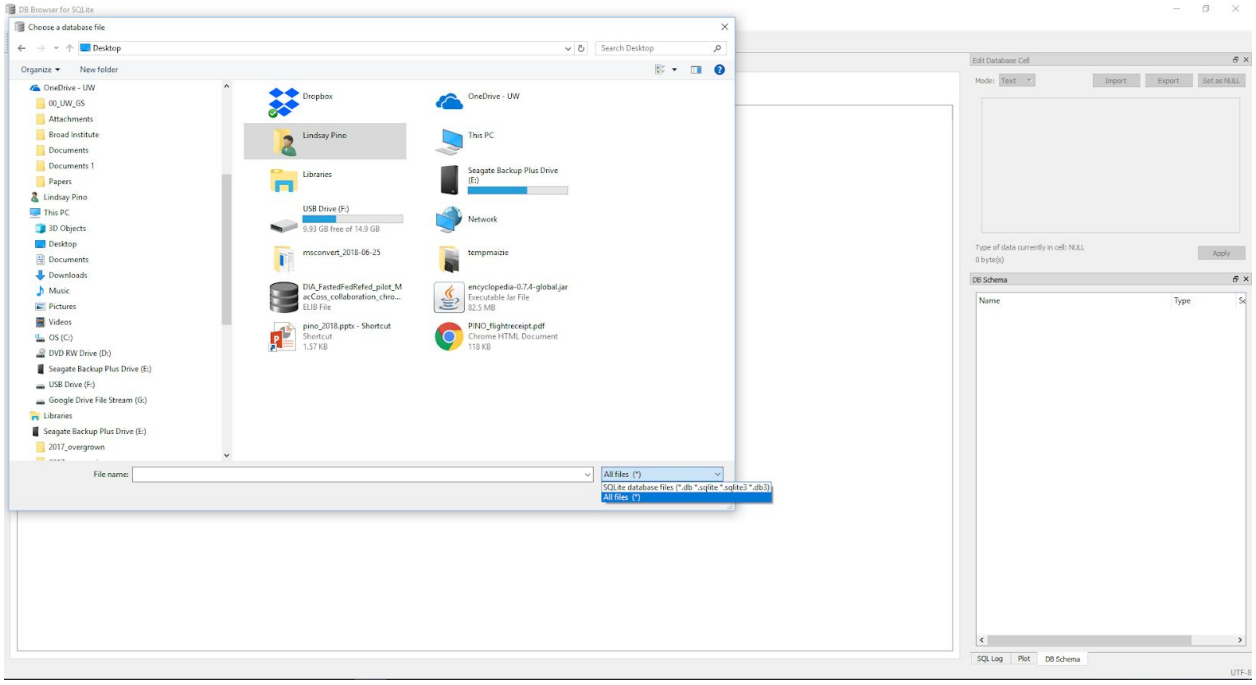
The *.elib files that Encyclopedia builds are SQL databases, which are kind of like multi-tab Excel files but fancier. To view these files and browse the data stored in the elib, follow the steps below:

Download DB Browser for SQLite here: https://sqlitebrowser.org/

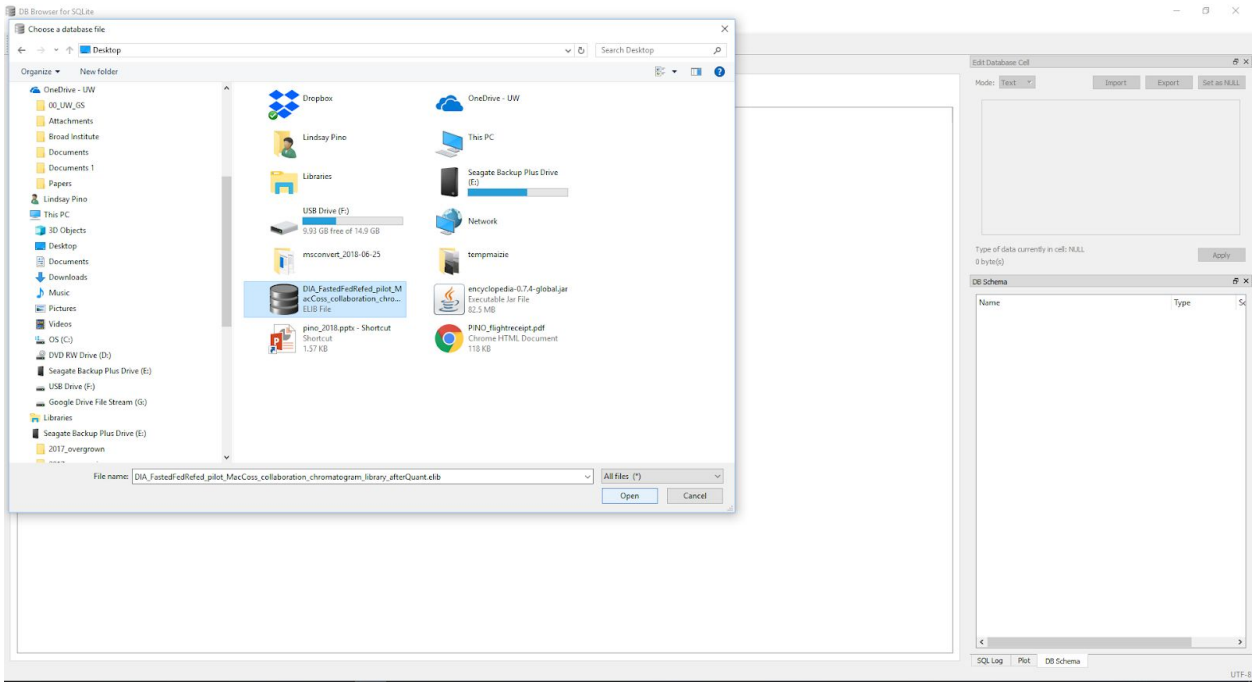
Open DB Browser for SQLite
Navigate to “Open Database” button on top left



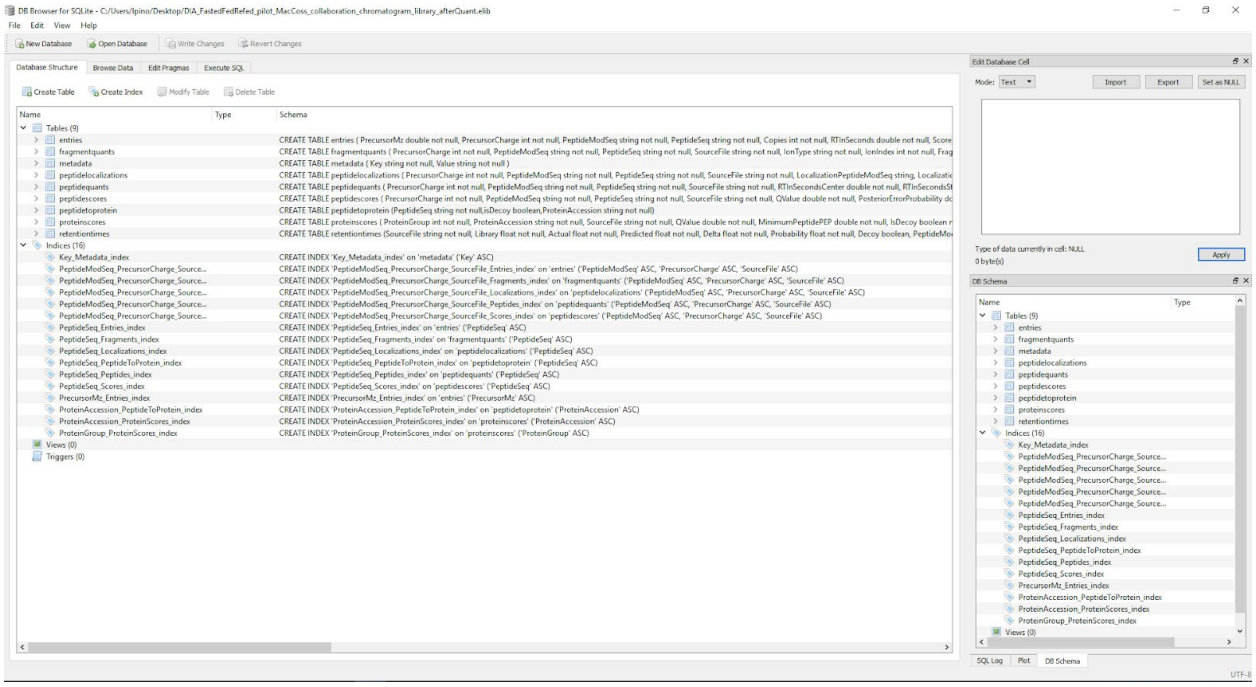
At the bottom of the “Choose a database file” pop-up next to the “File Name” field, select “All files” from the file type dropdown



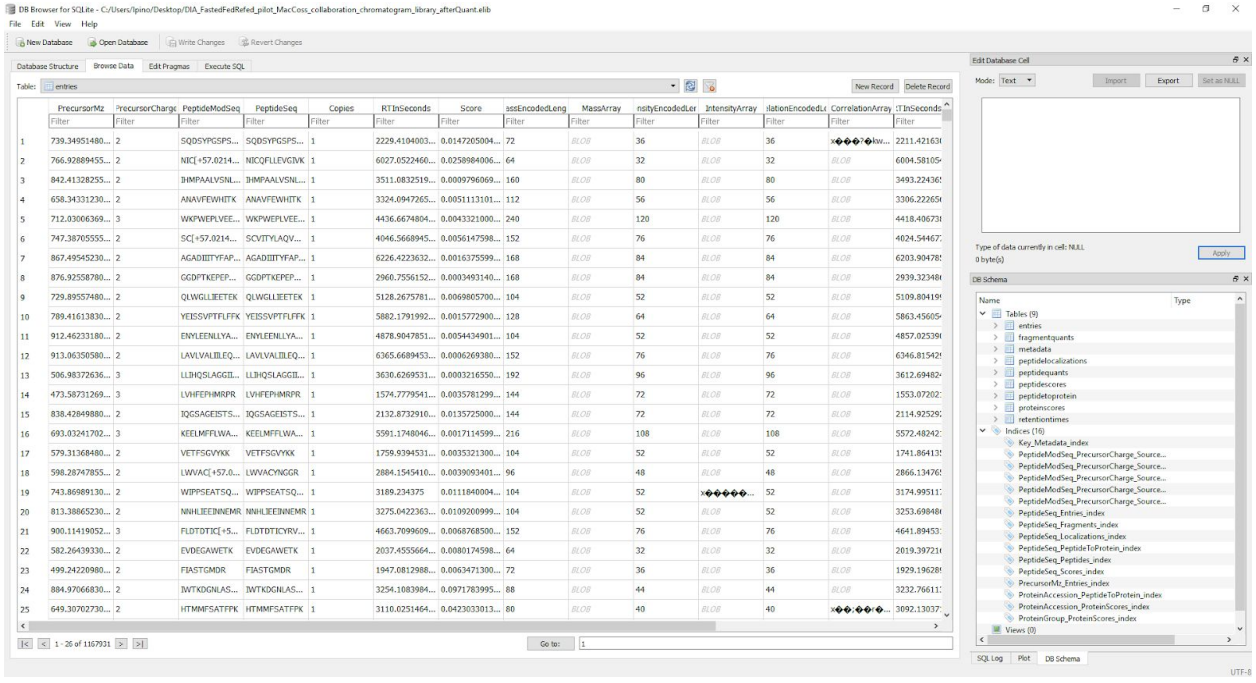
Navigate to the appropriate directory and select the elib file you want to view. Click .Open”



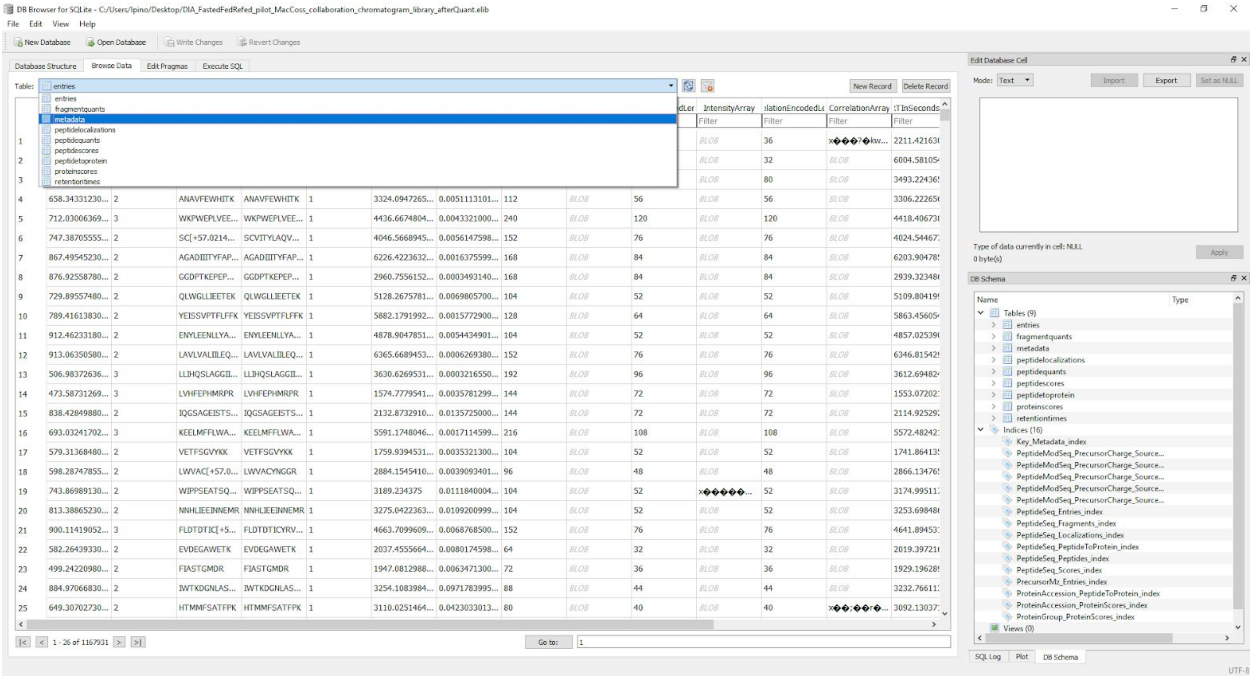
! It might take a minute to load. The window should look like this once it’s loaded:



Click on “Browse Data” tab at top under the “Open Database” button.



Under the “Browse Data” tab, select the table you want to view from the “Table:” dropdown menu. For example, the meta data:



Choosing the “metadata” table will display information such as the parameter settings when Encyclopedia was run, the TIC for each raw file in the chromatogram library/quant report, etc