

EncyclopeDIA Tutorial



Tutorials based on EncyclopeDIA version 4.6.0-SNAPSHOT

Last update on July, 8th 2024

Written by AE Shannon (shannon.225@osu.edu)

During this tutorial, you will learn how to

1. Convert Thermo .RAW files to a universal format (.mzML) using Proteowizard.
2. Search gas-phase fractionated samples against a prosit library to generate a chromatogram library.
3. Searching wide-window sample injections against a chromatogram library to detect peptides in a set of samples.

Additional material includes how to

4. Make a prosit-generated, predicted spectral library using a FASTA through the prosit server.
5. View detections in EncyclopeDIA and other outputs
6. Upload detections to Skyline

For today's tutorial, we will focus on the bolded portions. We will do these together on the screen.

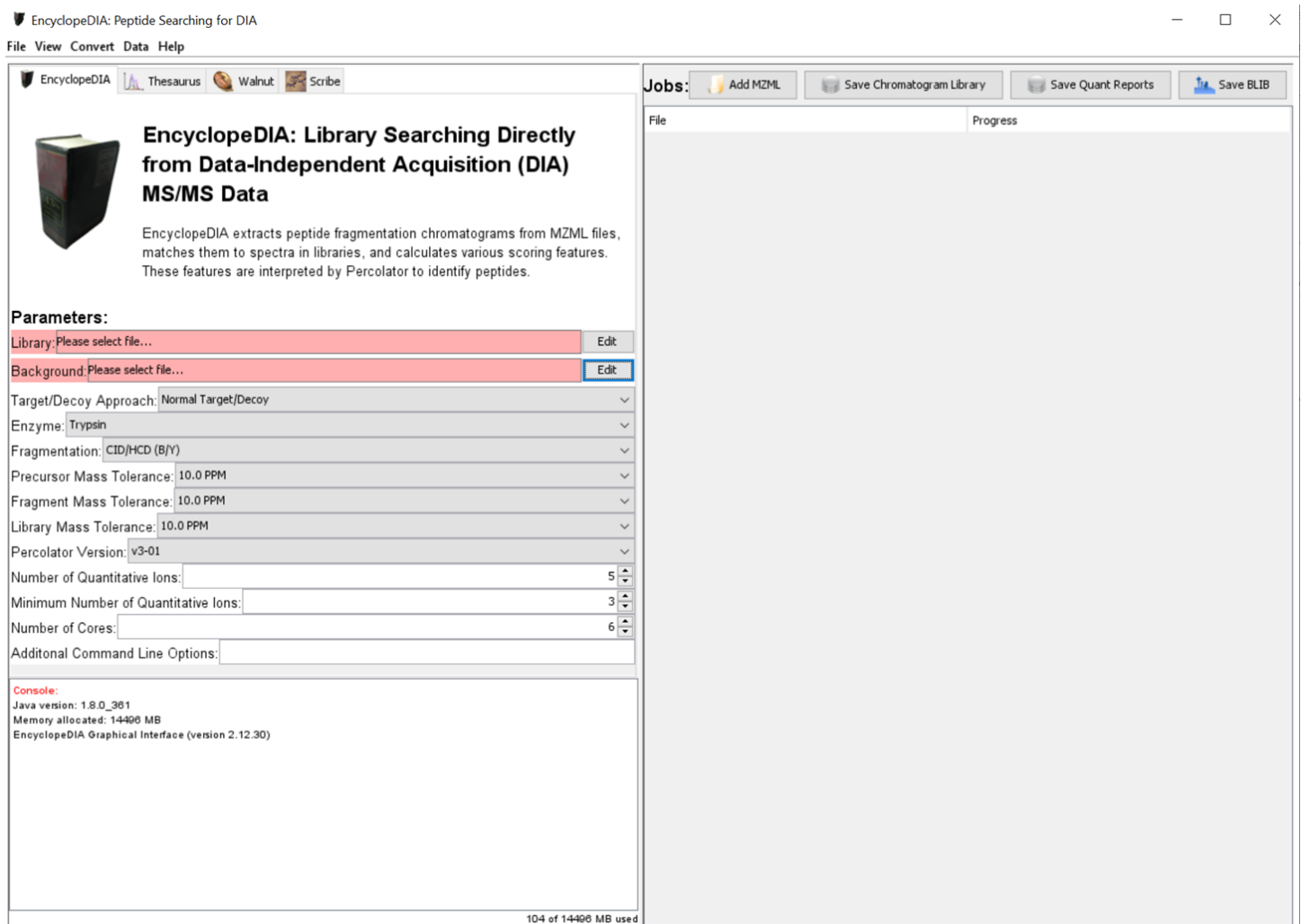
EncyclopeDIA Tutorial.....	1
OVERVIEW.....	2
1. PREREQUISITES AND INSTALLING ENCYCLOPEDIA.....	3
2. GENERATING MZMLS.....	4
3. GENERATING LIBRARIES USING PROSIT AND ENCYCLOPEDIA.....	8
4. ENCYCLOPEDIA SEARCH OPTIONS.....	15
5. GENERATING A CHROMATOGRAM LIBRARY.....	17
6. SEARCHING SAMPLES AGAINST A CHROMATOGRAM LIBRARY.....	18
7. IMPORTING DETECTIONS INTO SKYLINE.....	18

You can find the EncyclopeDIA processed libraries, FASTA file, and backup Skyline documents [here](#).

Raw data can be acquired from the 2022 Wang paper <https://doi.org/10.1038/s41597-022-01845-x>.

OVERVIEW

EncyclopeDIA comes with a user-friendly GUI interface. The upper left pane contains search options, while the right pane contains a process queue. The bottom left contains a console that provides specific information about the processes EncyclopeDIA is running.



This tutorial outlines how to use EncyclopeDIA, what outputs are given, and how to use the outputs to visualize the data.

1. PREREQUISITES AND INSTALLING ENCYCLOPEDIA

Software requirements needed to process data using EncyclopeDIA.

EncyclopeDIA is a cross-platform Java application that has been tested for Windows, Macintosh, and Linux. EncyclopeDIA requires 64-bit Java 1.8. While it is possible to use higher versions of Java, many versions are untested. Using untested versions of Java may result in unknown errors. In particular, Java versions 17 and 18 are known to cause stability issues in the current release.

A. If you don't already have Java, install Java 1.8 on your computer. if you are using Windows, you can download the Windows "x64 Installer" from: <https://www.oracle.com/java/technologies/downloads/#java8>. Other operating system options are available at this URL as well.

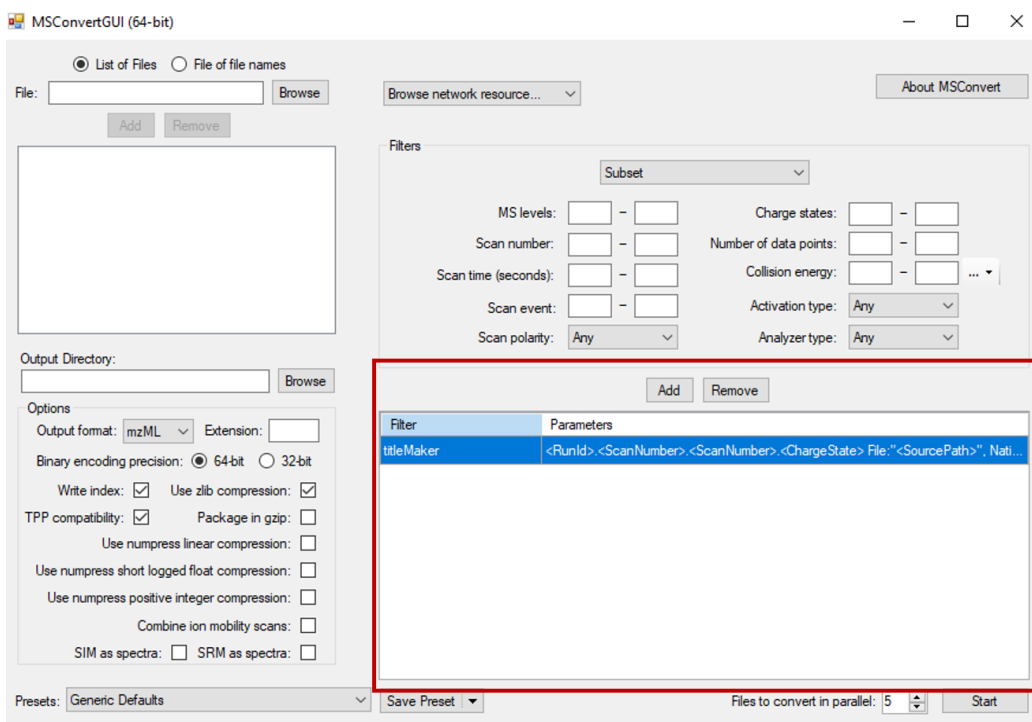
B. Scribe is folded into the EncyclopeDIA software package. After you have 64-bit Java 1.8, go to EncyclopeDIA's bitbucket page (<https://bitbucket.org/searleb/encyclopedia/wiki/Home>) and download the most recent stable version. Once downloaded, double-click on the EncyclopeDIA .JAR file to launch the GUI interface. If you are using a Macintosh, you may need to right-click on the EncyclopeDIA .JAR and select "Open" to execute it for the first time with the proper permissions. Click on the tab named Scribe at the top to search DDA data.

C. We recommend using Proteowizard to create mzML files from your RAW files. You can freely download Proteowizard from here: <https://proteowizard.sourceforge.io/download.html>.

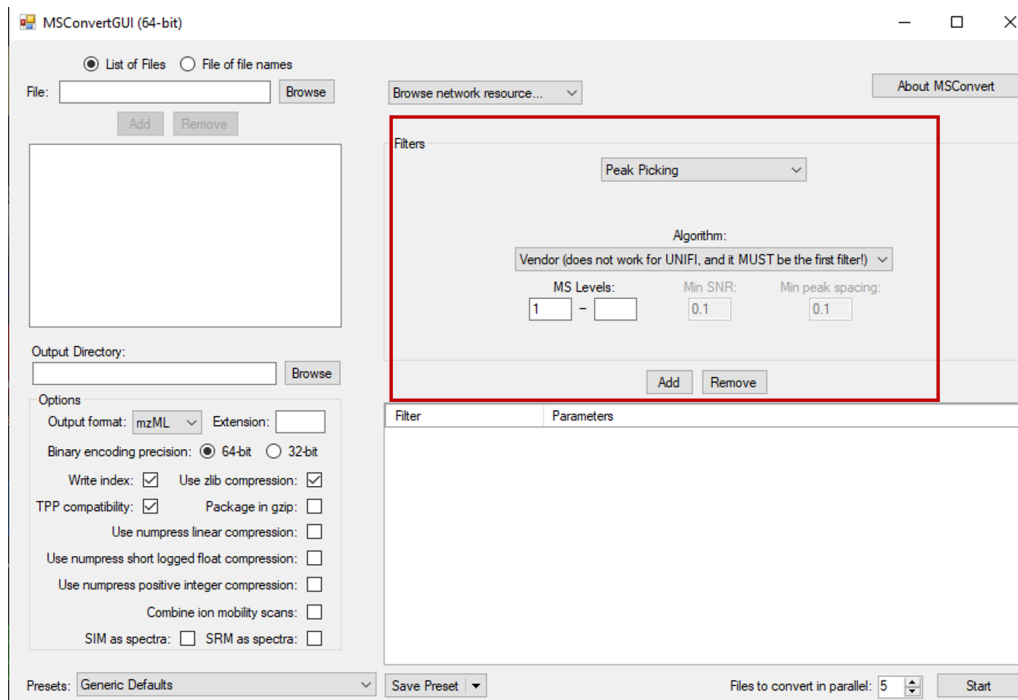
2. GENERATING MZMLS

How to use Proteowizard to generate vendor-neutral .mzML files from vendor-specific RAW files.

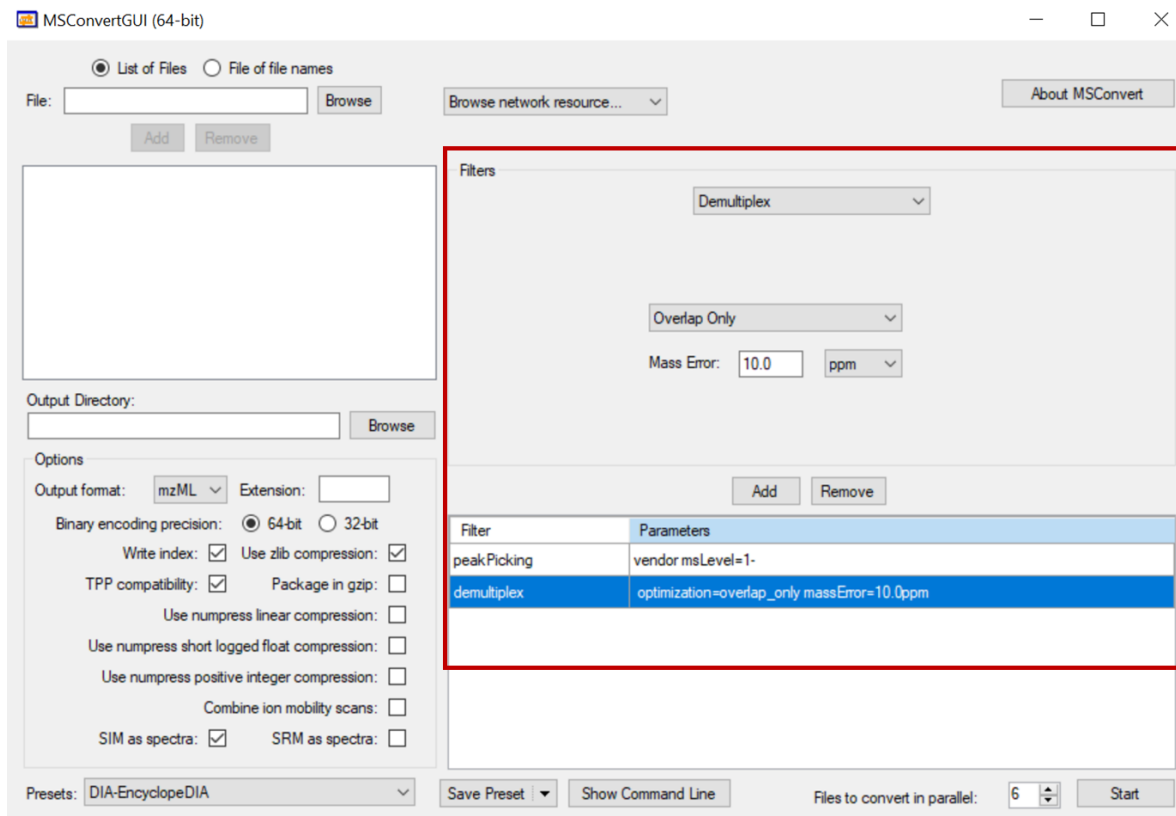
A. Before searching files in Scribe or EncyclopeDIA, RAW files must be converted to .mzML files using Proteowizard. To do so, open Proteowizard. Remove the “titleMakers” filter by selecting the filter in the parameters box, and clicking remove.



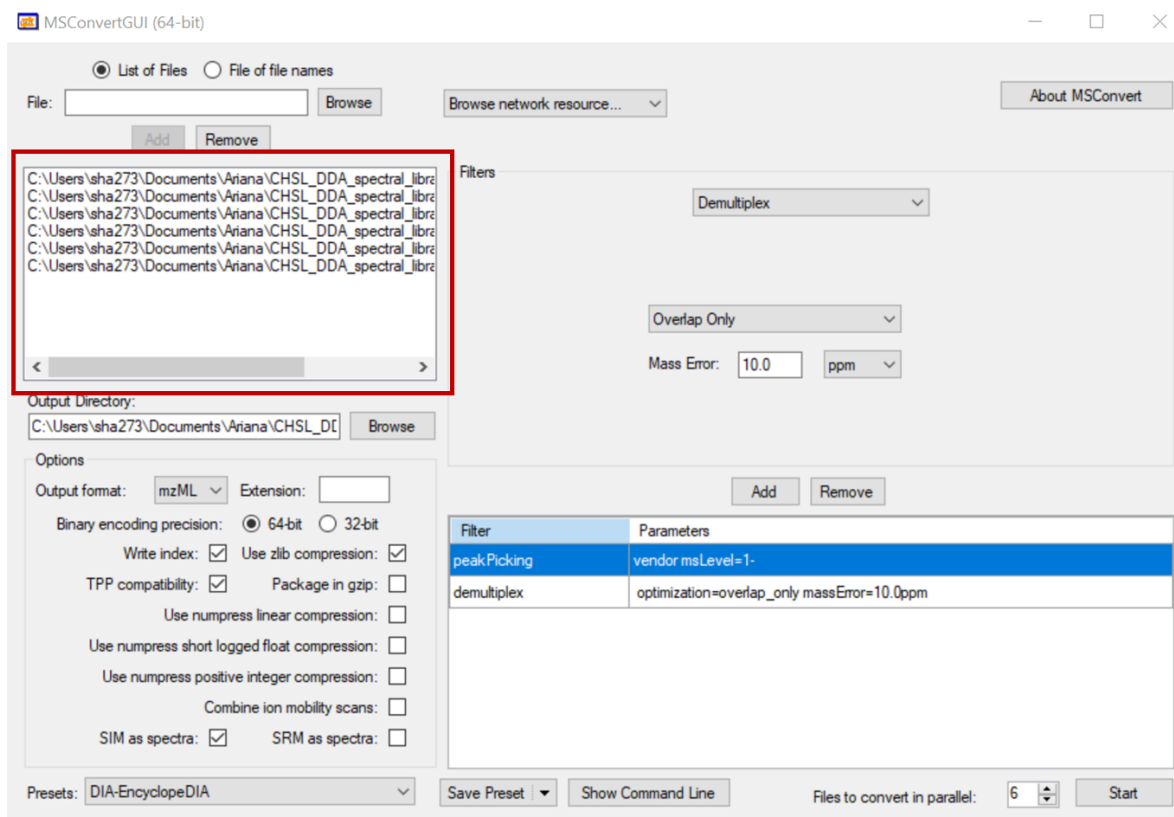
B. For converting DDA injections, or DIA injections that did not use an overlapping isolation windowing scheme, only add “Peak Picking” by selecting the peak picking option under filters, then click add. This should be the only filter in the box.



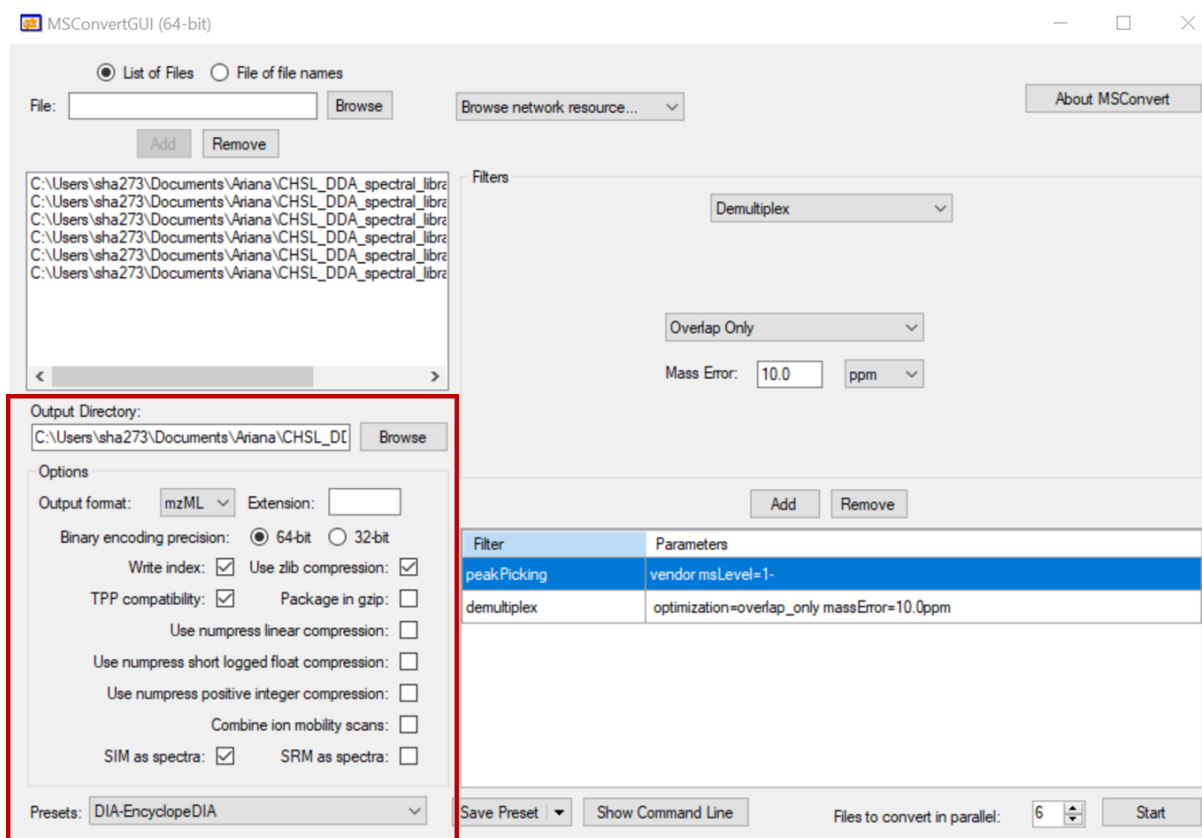
C. ONLY FOR DIA INJECTIONS USING OVERLAPPING WINDOWS: Click the drop down arrow and select demultiplex. Overlapping windows need to be demultiplexed after peak picking. Once you’ve found demultiplexing, you do not need to change anything else. Simply click “Add.” Your window for overlapping DIA data should look at follows:



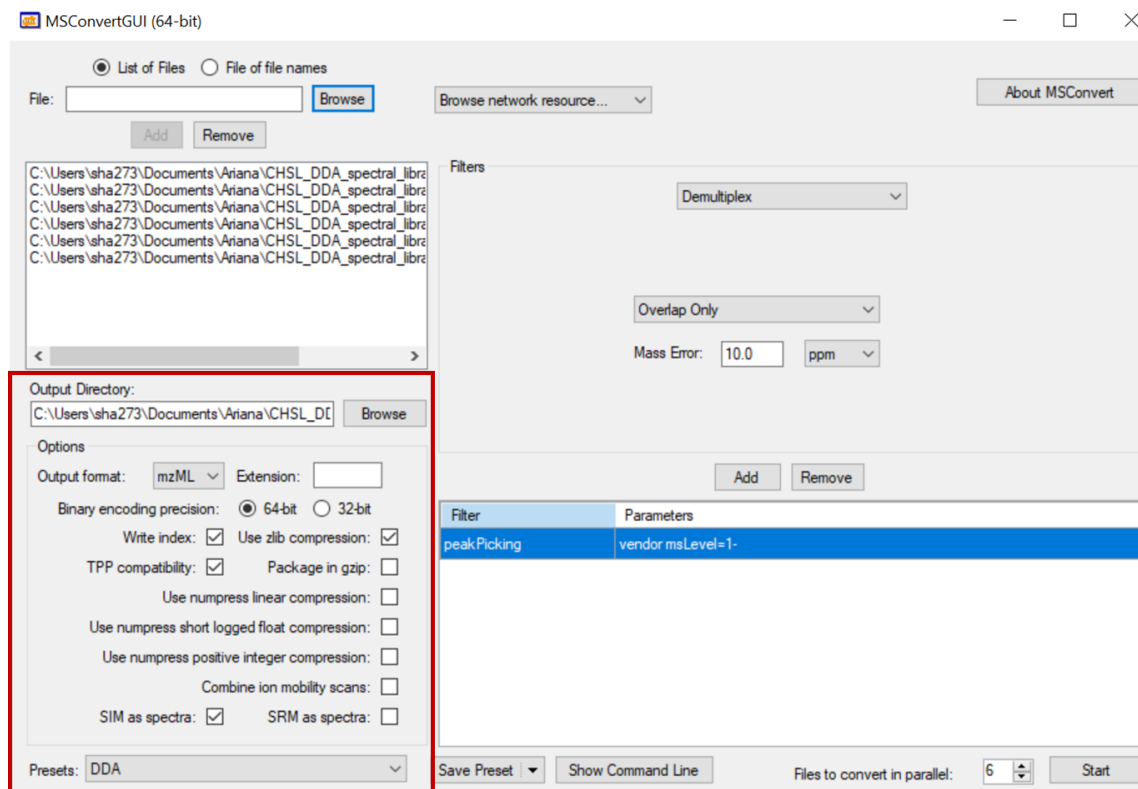
FOR DDA OR DIA WITH OVERLAPPING WINDOWS, your window will look like this:



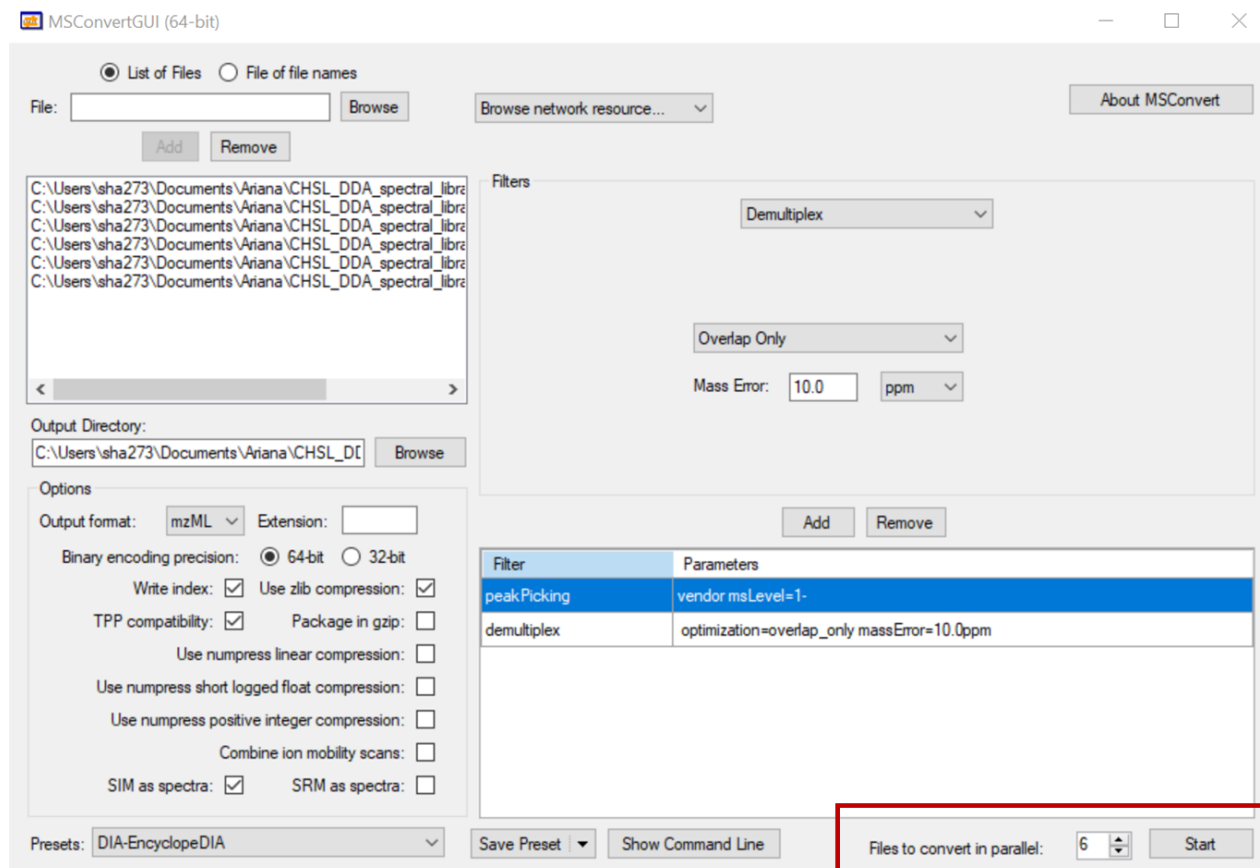
E. Select your files by clicking “browse”, and locate the desired DIA files in your directory. Click “add” on the left-hand side of the screen. The output directory will automatically populate where your files were selected from.

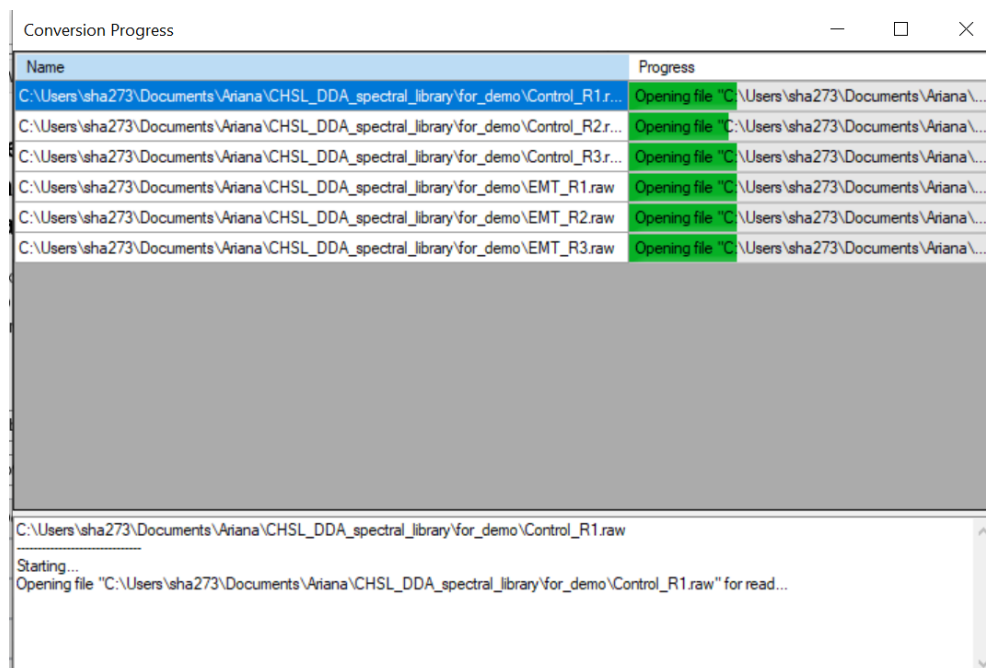


E. Under the options box, the output format should be mzML. You want “write index,” “use zlib compression,” “TPP compatibility, and “SIM as spectra” selected. Your settings should match the screenshot below.

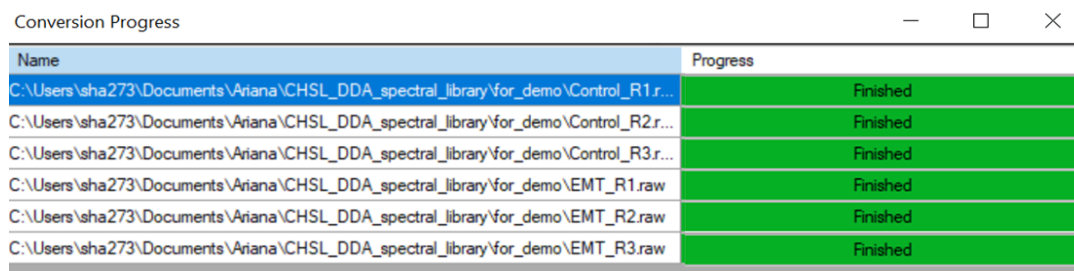


E. Click “Start” in the lower right hand corner to start the file conversion.



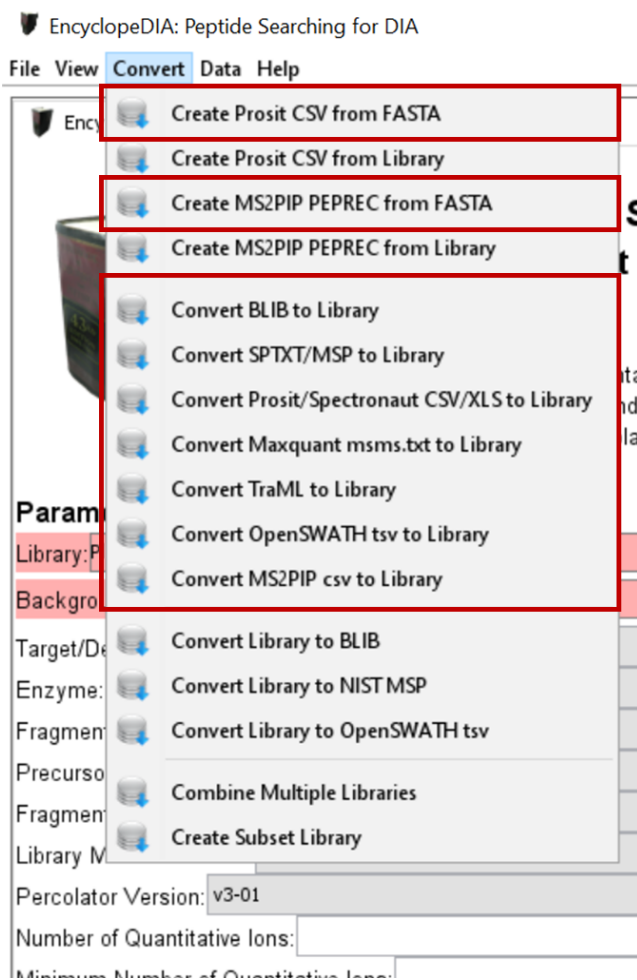


F. Once complete, the window will look like this. You can exit out of MSConvert once the conversion is done. Your files will be in the location previously specified.



3. GENERATING LIBRARIES USING PROSIT AND ENCYCLOPEDIA

How to acquire a FASTA file and then use Prosit and EncyclopeDIA to generate a spectral library file from it.



The primary library format EncyclopeDIA uses is .DLIB and .ELIB. Spectrum libraries from Skyline, NIST, TraML and other formats can be converted to .DLIB using the “Convert” menu. In particular, EncyclopeDIA can search fully predicted libraries generated by either Prosit or MS2PIP, and produce input files for both of those tools. The following tutorial is given on how to obtain a FASTA file and use Prosit to generate a predicted .DLIB. The files generated by EncyclopeDIA are .ELIBs, which are very similar to .DLIBs, but additionally contain retention time information.

A. To generate a Prosit-predicted spectral library (.DLIB) for EncyclopeDIA, start by obtaining a FASTA file for your organism of interest. Given the computational expense of generating libraries, in most cases we recommend using only canonical protein sequences. For example, if you require a FASTA of proteins sequences for Homo Sapiens, go to uniprot.org. Select “Reviewed.”

← → ↻ 🏠 uniprot.org

UniProt

BLASTAlignPeptide searchID mappingSPARQL

Find your protein

UniProtKBAdvanced | ListSearch

Examples: Insulin, APP, Human, P05067, organism_id:9606

UniProt is the world's leading high-quality, comprehensive and freely accessible resource of protein sequence and functional information. Cite UniProt

Proteins
UniProt Knowledgebase

Species
Proteomes

Protein Clusters
UniRef

Sequence Archive
UniParc

Reviewed
(Swiss-Prot)
568,363

Unreviewed
(TrEMBL)
229,928,140

Protein sets for species with sequenced
genomes from across the tree of life

Clusters of protein sequences at 100%,
90% & 50% identity

Non-redundant archive of publicly available
protein sequences seen across different
databases

Specify “Human (20,401).”

Status

Reviewed (Swiss-Prot)
(20,401)

UniProtKB 20,401 results

BLASTAlignMap IDsDownloadAddView: CardsTableCustomize columnsShare

Popular organisms

Human (20,401)

Taxonomy

Filter by taxonomy

Proteins with

3D structure (7,557)

Active site (2,278)

Activity regulation (1,527)

Allergen (6)

Alternative products (isoforms)
(10,634)

More items

Protein existence

Protein level (16,585)

Transcript level (2,310)

Homology (754)

Uncertain (613)

Predicted (139)

Annotation score

5 (14,308)

Entry	Entry Name	Protein Names	Gene Names	Organism	Length
A0A0C5B5G6	MOTSC_HUMAN	Mitochondrial-derived peptide MOTS-c[...]	MT-RNR1	Homo sapiens (Human)	16 AA
A0A1B0GTW7	CIROP_HUMAN	Ciliated left-right organizer metalloproteinase[...]	CIROP, LMLN2	Homo sapiens (Human)	788 AA
A0JNW5	UHRF1_HUMAN	UHRF1-binding protein 1-like[...]	UHRF1BP1L, KIAA0701, SHIP164	Homo sapiens (Human)	1,464 AA
A0JP26	POTB3_HUMAN	POTE ankyrin domain family member B3	POTEB3	Homo sapiens (Human)	581 AA
A0PK11	CLRN2_HUMAN	Clarin-2	CLRN2	Homo sapiens (Human)	232 AA
A1A4S6	RHG10_HUMAN	Rho GTPase-activating protein 10[...]	ARHGAP10, GRAF2	Homo sapiens (Human)	786 AA
A1A519	F170A_HUMAN	Protein FAM170A[...]	FAM170A, ZNFD	Homo sapiens (Human)	330 AA
A1L190	SYCE3_HUMAN	Synaptonemal complex central element protein 3[...]	SYCE3, C22orf41, THEG2	Homo sapiens (Human)	88 AA
A1L3X0	ELOV7_HUMAN	Elongation of very long chain fatty acids protein 7[...]	ELOVL7	Homo sapiens (Human)	281 AA
A1X283	SPD2B_HUMAN	SH3 and PX domain-containing protein 2B[...]	SH3PXD2B, FAD49, KIAA1295, TKS4	Homo sapiens (Human)	911 AA
A2A2Y4	FRMD3_HUMAN	FERM domain-containing protein 3[...]	FRMD3, EPB41L4O	Homo sapiens (Human)	597 AA
A2RU14	TM218_HUMAN	Transmembrane protein 218	TMEM218	Homo sapiens (Human)	115 AA
A2RUB6	CCDC66_HUMAN	Coiled-coil domain-containing protein 66	CCDC66	Homo sapiens (Human)	948 AA
A2RUC4	TYW5_HUMAN	tRNA tryptophan-synthetizing protein 5[...]	TYW5, C2orf60	Homo sapiens (Human)	315 AA
A4D1B5	GSAP_HUMAN	Gamma-secretase-activating protein[...]	GSAP, PION	Homo sapiens (Human)	854 AA
A4GXA9	EME2_HUMAN	Probable crossover junction endonuclease EME2[...]	EME2	Homo sapiens (Human)	379 AA

Feedback

Help

Select “Download.” If you download the compressed version, you will get a .gz and have to uncompress the file. If you download the uncompressed version, the file may take longer, but it will be in a FASTA format for immediate use.

The screenshot shows a web application interface. On the left, a 'Download' dialog box is open, highlighting the 'Download all (20,401)' option. The 'Format' is set to 'FASTA (canonical & isoform)' and 'Compressed' is set to 'Yes'. On the right, a table displays protein data with columns for Gene Names, Organism, and Length. The table lists various proteins from Homo sapiens (Human) with their corresponding gene names and lengths in amino acids (AA).

Gene Names	Organism	Length
MT-RNR1	Homo sapiens (Human)	16 AA
CIROP, LMLN2	Homo sapiens (Human)	788 AA
UHRF1BP1L, KIAA0701, SHIP164	Homo sapiens (Human)	1,464 AA
POTEB3	Homo sapiens (Human)	581 AA
CLRN2	Homo sapiens (Human)	232 AA
ARHGAP10, GRAF2	Homo sapiens (Human)	786 AA
FAM170A, ZNFD	Homo sapiens (Human)	330 AA
SYCE3, C22orf41, THEG2	Homo sapiens (Human)	88 AA
ELOVL7	Homo sapiens (Human)	281 AA
SH3PXD2B, FAD49, KIAA1295, TKS4	Homo sapiens (Human)	911 AA
FRMD3, EPB41L4O	Homo sapiens (Human)	597 AA
TMEM218	Homo sapiens (Human)	115 AA
CCDC66	Homo sapiens (Human)	948 AA
TYW5, C2orf60	Homo sapiens (Human)	315 AA
GSAP, PION	Homo sapiens (Human)	854 AA

B. Open EncyclopeDIA to create a Prosit CSV from a FASTA file. Navigate to the “Convert.” Select “Create Prosit CSV from FASTA.” This will open a dialog window. If you are generating a spectral library to search against DIA data, you should check “Adjust NCE for DIA.”

The screenshot shows the EncyclopeDIA software interface. The 'Convert' menu is open, highlighting 'Create Prosit CSV from FASTA'. The 'Convert FASTA to Prosit CSV' dialog box is also shown, with the 'Adjust NCE for DIA' checkbox checked. The dialog includes fields for FASTA file selection, Enzyme (Trypsin), Charge range (2 to 3), Maximum Missed Cleavage (1), m/z range (396.4 to 1,002.7), Default NCE (33), and Default Charge (3).

EncyclopeDIA: Peptide Searching for DIA

File View **Convert** Data Help

- Create Prosit CSV from FASTA
- Create Prosit CSV from Library
- Create MS2PIP PEPREC from FASTA
- Create MS2PIP PEPREC from Library
- Convert BLIB to Library
- Convert SPTXT/MSP to Library
- Convert Prosit/Spectronaut CSV/XLS to Library
- Convert Maxquant msms.txt to Library
- Convert TraML to Library
- Convert OpenSWATH tsv to Library
- Convert MS2PIP csv to Library
- Convert Library to BLIB
- Convert Library to NIST MSP
- Convert Library to OpenSWATH tsv
- Combine Multiple Libraries
- Create Subset Library

Parameters:

This function will *in silico* digest peptides from your FASTA and create an input file for Prosit. If you use this feature, please cite [Searle et al, 2020](#).

FASTA: Please select file... Edit

Enzyme: Trypsin

Charge range: 2 to 3

Maximum Missed Cleavage: 1

m/z range: 396.4 to 1,002.7

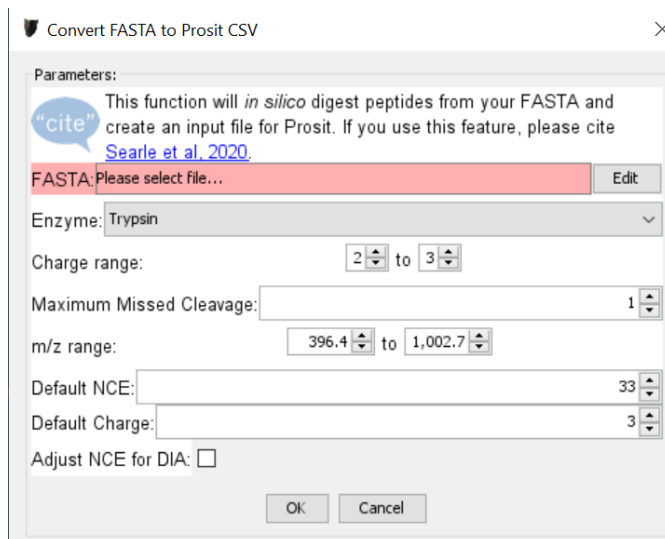
Default NCE: 33

Default Charge: 3

Adjust NCE for DIA: ☒

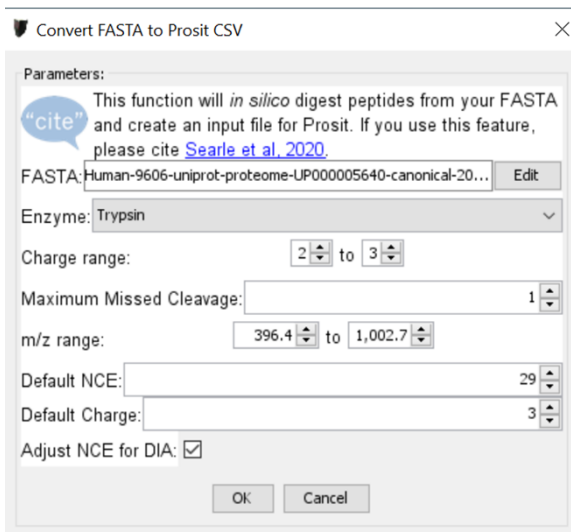
OK Cancel

If you are building a prosit library to search against DDA data, you should leave the “Adjust NCE for DIA” unchecked.



Upload the FASTA you have downloaded. The Prosit CSV will be generated in the same folder where the FASTA is held. If you use the example from the FASTA file provided in the tutorial, you should load [“Human-9606-uniprot-proteome-UP000005640-canonical-2022_04.fasta”](#)

Settings for “human_prosit_generated_abrf.dlib”



This will generate the file
“Human-9606-uniprot-proteome-UP000005640-canonical-2022_04.fasta.fasta.trypsin.z3_nce29,” which is what you will use to put into Prosit.

C. Use Prosit to perform an *in silico* digestion, and obtain an .MSP/NIST format text file. Navigate to the Prosit website, [Prosit](#). When you get to the page, there are three tabs available; “CE CALIBRATION,” “SPECTRAL

LIBRARY,” and “RESCORING.”

! Intensity model works with both CID and HCD fragmentation methods but you need to add fragmentation column to the input. We assume all the sequences are fully labeled and you don't need to add the tmt modification explicitly in your input files.

Prosit 

PREDICT LIBRARIES FAQ STATUS  

Prosit offers high quality MS2 predicted spectra for any organism and protease as well as IRT prediction. Prosit is part of the ProteomeTools (www.proteometools.org/) project and was trained on the project's high quality synthetic dataset. When using Prosit is helpful for your research, please cite "Gessulat, Schmidt et al. 2019" DOI:10.1038/s41592-019-0426-7.

CE CALIBRATION **SPECTRAL LIBRARY** RESCORING

This task estimates the optimal collision energy (CE) based on a given search result. You need to upload a RAW file as well as the MaxQuant's msms.txt for calibration.

Prosit will:

1. Select a random subset of high-scoring PSMS
2. Predict those in for each CE from 18 to 39.
3. Calculate which CE achieves highest correlations with the experimental spectra

Please note: Sequences with amino acid U or O are not supported. Modifications except "M(ox)" are not supported. Each C is treated as Cysteine with carbamidomethylation (fixed modification in MaxQuant).

Also note: Antivirus software may cancel large uploads - turn it off if you experience upload resets.

1 Upload Files
msms.txt and RAW file.

 msms.txt
MaxQuant's msms.txt from a finished search. Note, amino acid U or O are not supported

 RAW
RAW file that was searched (restricted to Thermo Fisher HCD Orbitrap). File size is limited to 2GB.

NEXT >

2 Model
Select intensity and IRT model for prediction

Go to the “SPECTRAL LIBRARY” tab.

! We now offer two new Prosit TMT models that will soon be published. One is for fragment intensities prediction (Prosit_TMT_intensity_2021) and the other is for IRT prediction (Prosit_TMT_irt_2021). The intensity model works with both CID and HCD fragmentation methods but you need to add fragmentation column to the input. We assume all the sequences are fully labeled and you don't need to add the tmt modification explicitly in your input files.

Prosit 

PREDICT LIBRARIES FAQ STATUS  

Prosit offers high quality MS2 predicted spectra for any organism and protease as well as IRT prediction. Prosit is part of the ProteomeTools (www.proteometools.org/) project and was trained on the project's high quality synthetic dataset. When using Prosit is helpful for your research, please cite "Gessulat, Schmidt et al. 2019" DOI:10.1038/s41592-019-0426-7.

CE CALIBRATION **SPECTRAL LIBRARY** RESCORING

This task generates a spectral library either by digesting a given FASTA file, or by predicting a list of peptides given in a CSV file. You need to provide a collision energy (CE) for prediction. To estimate an optimal CE for prediction, please use "CE Calibration".

When a FASTA file is provided, Prosit will:

1. Digest the FASTA, for the given parameters (i.e. protease).
2. Predict all spectra at the given collision energy.

When a CSV with peptides is provided, Prosit will directly predict all spectra.

Please note: Antivirus software may cancel large uploads - turn it off if you experience upload resets.

1 Settings
Indicate collision energy, the maximum number of missed cleavages, and number of oxidized methionines per peptide.

How would you like to provide the list of peptides?

☒ CSV
☐ FASTA (coming soon)

CSV Format

modified_sequence	collision_energy	precursor_charge	fragmentation
M(ox)CSDSDGLAPPQHLIR	15	2	HCD

For the 1st step, CSV is already selected. You can click next to get to the 2nd step. Upload the Prosit CSV created in EncyclopeDIA.

Indicate collision energy, the maximum number of missed cleavages, and number of oxidized methionines per peptide.

2 Upload Files
Fasta or CSV with list of peptides

CSV Format

modified_sequence	collision_energy	precursor_charge		fragmentation
M(ox)CSDSDGLAPPQHLIR	15	2	For TMT models Only	HCD
EMPQSDPSVEPPLSQETFSDLWK	28	2		HCD
TCPVQLWVDSTPPPGTR	35	3		CID
QSQHM(ox)TEVVR	35	5		CID

Please provide all three columns below and use `,` as a separator.

- modified_sequence** Use upper case letters in the column and indicate oxidized Methionine with "M(ox)". Sequence length is restricted to the range of 7 to 30. Each C is treated as Cysteine with carbamidomethylation. Prosit does not support U or O as amino acids.
- collision_energy** Use integer values from 10 and 50.
- precursor_charge** Use integer values from 1 to 6.
- fragmentation** Either HCD or CID, Use upper case letters.*

*Only for TMT model

peptides.csv

containing peptide sequence, collision energy and precursor charge.

< BACK NEXT >

3 Model
Select intensity and IRT model for prediction

4 Isobaric Label
If TMT model is selected

Once the CSV has once loaded, click next.

Settings
Indicate collision energy, the maximum number of missed cleavages, and number of oxidized methionines per peptide.

2 Upload Files
Fasta or CSV with list of peptides

CSV Format

modified_sequence	collision_energy	precursor_charge		fragmentation
M(ox)CSDSDGLAPPQHLIR	15	2	For TMT models Only	HCD
EMPQSDPSVEPPLSQETFSDLWK	28	2		HCD
TCPVQLWVDSTPPPGTR	35	3		CID
QSQHM(ox)TEVVR	35	5		CID

Please provide all three columns below and use `,` as a separator.

- modified_sequence** Use upper case letters in the column and indicate oxidized Methionine with "M(ox)". Sequence length is restricted to the range of 7 to 30. Each C is treated as Cysteine with carbamidomethylation. Prosit does not support U or O as amino acids.
- collision_energy** Use integer values from 10 and 50.
- precursor_charge** Use integer values from 1 to 6.
- fragmentation** Either HCD or CID, Use upper case letters.*

*Only for TMT model

Human-9606-uniprot-proteome-UP000005640-canonical-2022_04.fasta.fasta.trypsin.z3_nce29.csv

containing peptide sequence, collision energy and precursor charge.

< BACK NEXT >

3 Model
Select intensity and IRT model for prediction

Select the desired model. For this example, we want to use the “Prosit_2020_intensity_hcd” for the Intensity prediction model, and the “Prosit_2019_irt” for the iRT prediction model. Click next.

Please note: Antivirus software may cancel large uploads - turn it off if you experience upload resets.

✓ Settings
Indicate collision energy, the maximum number of missed cleavages, and number of oxidized methionines per peptide.

✓ Upload Files
Fasta or CSV with list of peptides

3 Model
Select intensity and iRT model for prediction

Intensity prediction model

☐ Prosit_2019_intensity_hcd

☐ Prosit_2020_intensity_preview

☒ Prosit_2020_intensity_hcd

☐ Prosit_2020_intensity_cid

☐ Prosit_TMT_intensity_2021

iRT prediction model

☒ Prosit_2019_irt

☐ Prosit_TMT_irt_2021

< BACK NEXT >

4 Isobaric Label
If TMT model is selected

5 Task ID
Check if everything is correct and submit the task

For Task ID Output format, select “NIST .MSP Text Format of individual spectra (Skyline and MSPepSearch compatible).”

CE CALIBRATION SPECTRAL LIBRARY RESCORING

This task generates a spectral library either by digesting a given FASTA file, or by predicting a list of peptides given in a CSV file. You need to provide a collision energy (CE) for prediction. To estimate an optimal CE for prediction, please use “CE Calibration”.

When a FASTA file is provided, Prosit will:

1. Digest the FASTA, for the given parameters (i.e. protease).
2. Predict all spectra at the given collision energy.

When a CSV with peptides is provided, Prosit will directly predict all spectra.

Please note: Antivirus software may cancel large uploads - turn it off if you experience upload resets.

✓ Settings
Indicate collision energy, the maximum number of missed cleavages, and number of oxidized methionines per peptide.

✓ Upload Files
Fasta or CSV with list of peptides

✓ Model
Select intensity and iRT model for prediction

✓ Isobaric Label
If TMT model is selected

5 Task ID
Check if everything is correct and submit the task

Output format

☒ NIST .MSP Text Format of individual spectra (Skyline and MSPepSearch compatible)

☐ Generic text (Spectronaut compatible). All fragments are reported.

< BACK SUBMIT ✓

Submit the task. It is helpful to record the task number or bookmark the page shown after you submit the task to come back to once the job is complete.

We now offer two new Prosit TMT models that will soon be published. One is for fragment intensities prediction (Prosit_TMT_intensity_2021) and the other is for IRT prediction (Prosit_TMT_irt_2021). The intensity model works with both CID and HCD fragmentation methods but you need to add fragmentation column to the input. We assume all the sequences are fully labeled and you don't need to add the tmt modification explicitly in your input files.

Prosit 

PREDICT LIBRARIES FAQ STATUS  

Prosit offers high quality MS2 predicted spectra for any organism and protease as well as IRT prediction. Prosit is part of the ProteomeTools (www.proteomertools.org/) project and was trained on the project's high quality synthetic dataset. When using Prosit is helpful for your research, please cite "Gessulat, Schmidt et al. 2019" DOI 10.1038/s41592-019-0426-7.

Task 63E59E6B03CB54EC9E46D2AA185FCB6F

This task is in progress. Tasks may take several hours for full proteomes depending on system load. Please note down your Task ID or save this URL to check back later. You can download the results here upon completion. Resubmitting tasks will not lead to faster results.

Once the task is done, download the .MSP file.

We now offer two new Prosit TMT models that will soon be published. One is for fragment intensities prediction (Prosit_TMT_intensity_2021) and the other is for IRT prediction (Prosit_TMT_irt_2021). The intensity model works with both CID and HCD fragmentation methods but you need to add fragmentation column to the input. We assume all the sequences are fully labeled and you don't need to add the tmt modification explicitly in your input files.

Prosit 

PREDICT LIBRARIES FAQ STATUS  

Prosit offers high quality MS2 predicted spectra for any organism and protease as well as IRT prediction. Prosit is part of the ProteomeTools (www.proteomertools.org/) project and was trained on the project's high quality synthetic dataset. When using Prosit is helpful for your research, please cite "Gessulat, Schmidt et al. 2019" DOI 10.1038/s41592-019-0426-7.

Task 63E59E6B03CB54EC9E46D2AA185FCB6F

Your files are ready.

DOWNLOAD

D. Convert the Prosit output (.MSP/NIST) to a Library (.DLIB) using EncyclopeDIA. At the top of the GUI, go to "Convert," then navigate to "Convert SPTXT/MSP to Library." This window will pop up.

 Convert Prosit/Spectronaut CSV to Library 

Parameters:

Spectronaut CSV/XLS: 

FASTA: 

OK Cancel

Upload the downloaded .MSP file from Prosit, and the FASTA file used to generate the .MSP.

After Uploading Files

 Convert NIST SPTXT/MSP to Library 

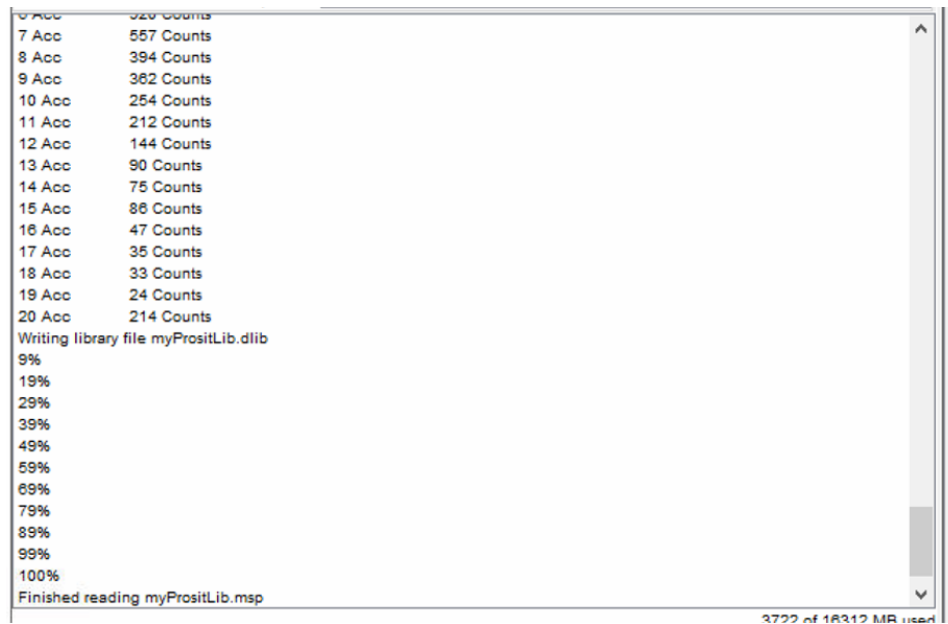
Parameters:

SPTXT/MSP: 

FASTA: 

OK Cancel

Click okay. The dialog box will indicate the .MSP file is being converted to a .DLIB. Once it is complete, the dialog box will tell you.

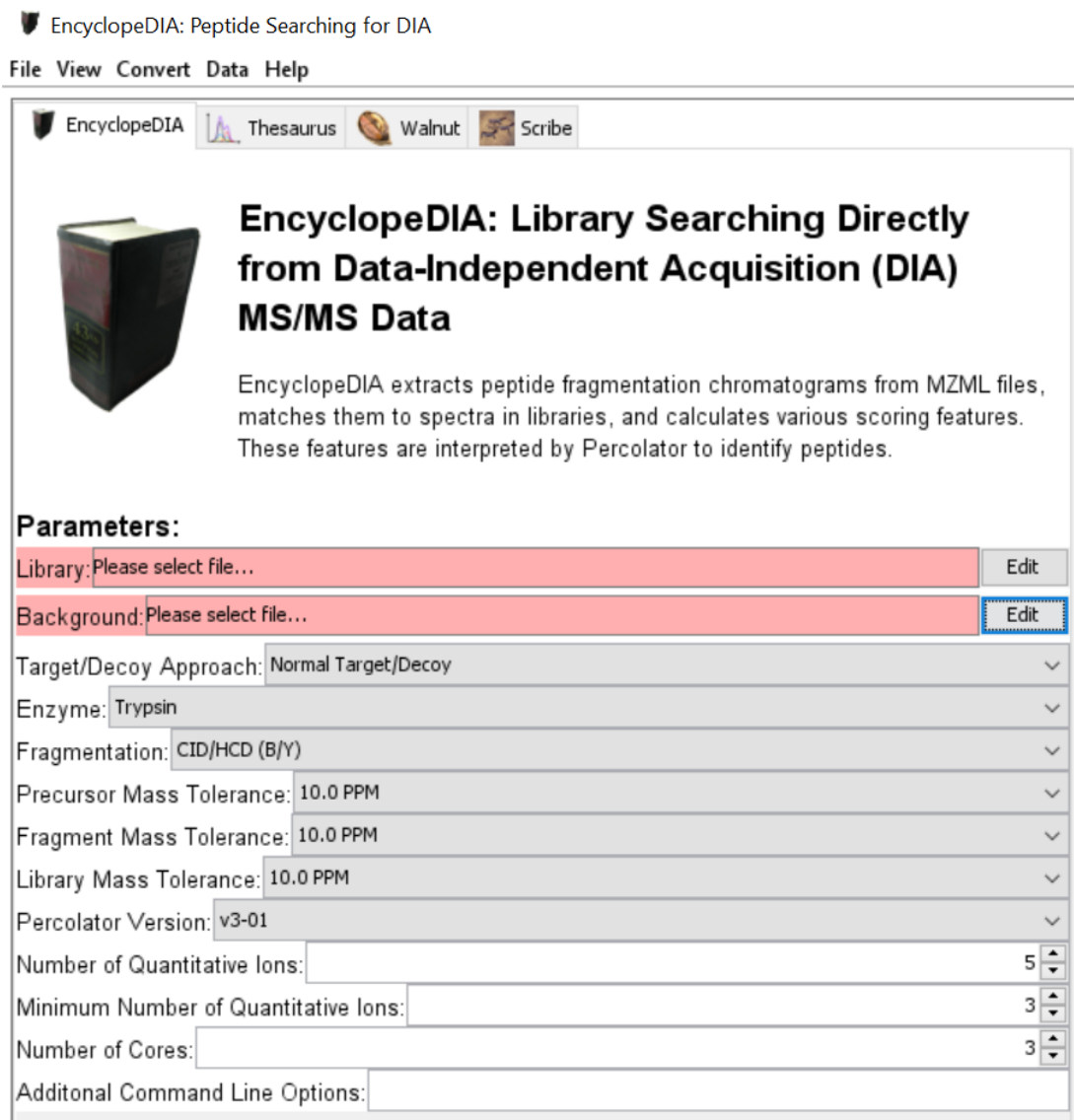


The resulting .DLIB should be analogous to [human_prosit_generated_abrf.dlib](#).

4. ENCYCLOPEDIA SEARCH OPTIONS

Descriptions of the search options available in EncyclopeDIA.

Here is a screenshot of the EncyclopeDIA default search options:



A. EncyclopeDIA has several options for searching files. Before you can start loading data, you need to specify both a .DLIB or .ELIB library to search as well as a background FASTA. These will be shaded red until they are properly specified. Libraries can be either in the .ELIB (chromatogram library) or .DLIB (spectrum library) format.

B. EncyclopeDIA has several other search settings. As a general rule, we recommend using the default search parameters first. Other settings are defined below:

Target/Decoy Approach: In some circumstances, it may be necessary to add additional decoys to improve statistical analysis. However, as a general rule, this should be left at “Normal Target/Decoy”.

Enzyme: Several common digestion enzymes are supported.

Fragmentation: In general, we recommend using CID/HCD (B/Y) fragmentation for most CID or HCD experiments. However, if your library is particularly large or messy you may get improved results with “HCD (y- only)”.

Precursor/Fragment/Library Mass Tolerance: Tolerances can be specified in PPM, AMU, or resolution.

Percolator Version: Percolator 3.1 is recommended for most experiments.

Number of Cores: This is the number of CPU cores you allow EncyclopeDIA to use. The maximum value you should set this to is one less than the number of cores your computer has. You need to leave at least one core for background processes.

C. Additional command line options can be specified in the command line options box. For example:

Additional Command Line Options:

5. GENERATING A CHROMATOGRAM LIBRARY

How to use spectral libraries made from DDA to search against DIA experiments

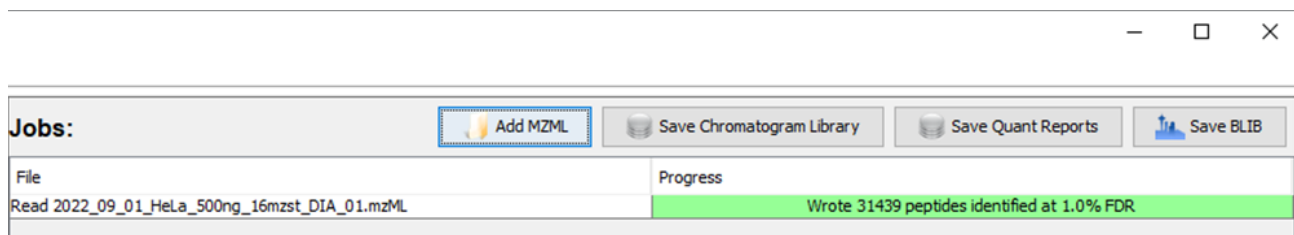
A. A chromatogram library, .DLIB, file is output from EncyclopeDIA. With the latest version of EncyclopeDIA, you can use the acquired spectral library, or prosit-predicted library to search DIA injections. In this example, we are using 6 gas-phase fractionated injections of a pool of the 3 ABRF proteome mixtures. Open EncyclopeDIA, and upload “[combined_abrf_6x_gpf.DLIB](#).” Specify “[LakeTrout-Human-Cow-Contaminants-sPRG2022.fasta](#)” as the background file. The settings should match the screenshot below.

Target/Decoy Approach:	Normal Target/Decoy	▼
Enzyme:	Trypsin	▼
Fragmentation:	CID/HCD (B/Y)	▼
Precursor Mass Tolerance:	10.0 PPM	▼
Fragment Mass Tolerance:	10.0 PPM	▼
Library Mass Tolerance:	10.0 PPM	▼
Percolator Version:	v3-01	▼
Number of Quantitative Ions:	5	▲▼
Minimum Number of Quantitative Ions:	3	▲▼
Number of Cores:	12	▲▼
Additional Command Line Options:		

Click “Add mzML,” then find the 6 DIA injections to run against the spectral library.

Jobs:	 Add MZML	 Save Chromatogram Library	 Save Quant Reports	 Save BLIB
File	Progress			

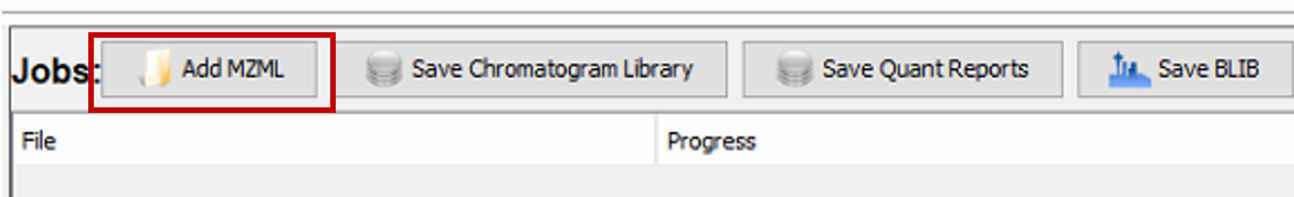
Once the search is complete, all 6 files will say the number of peptides detected. This is not exactly what your screen will look like, but representative



You can then visualize detected peptides as chromatograms using the “ELIB Browser.” The peptide below is an example of a peptide that was detected with little interference and multiple fragment ions.

6. SEARCHING SAMPLES AGAINST A CHROMATOGRAM LIBRARY

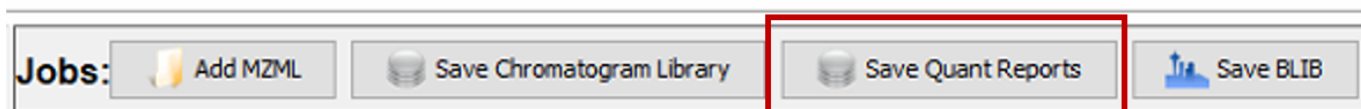
A. Use the “Add MZML” button to add RAW files for library searching. These will be automatically placed on the queue and executed in order using the current settings. If an MZML has been previously analyzed, the GUI will remember where it left off and try to not process it a second time.



Here is a screenshot of an EncyclopeDIA search queue. Again, these are representative, not what your screen will look like:

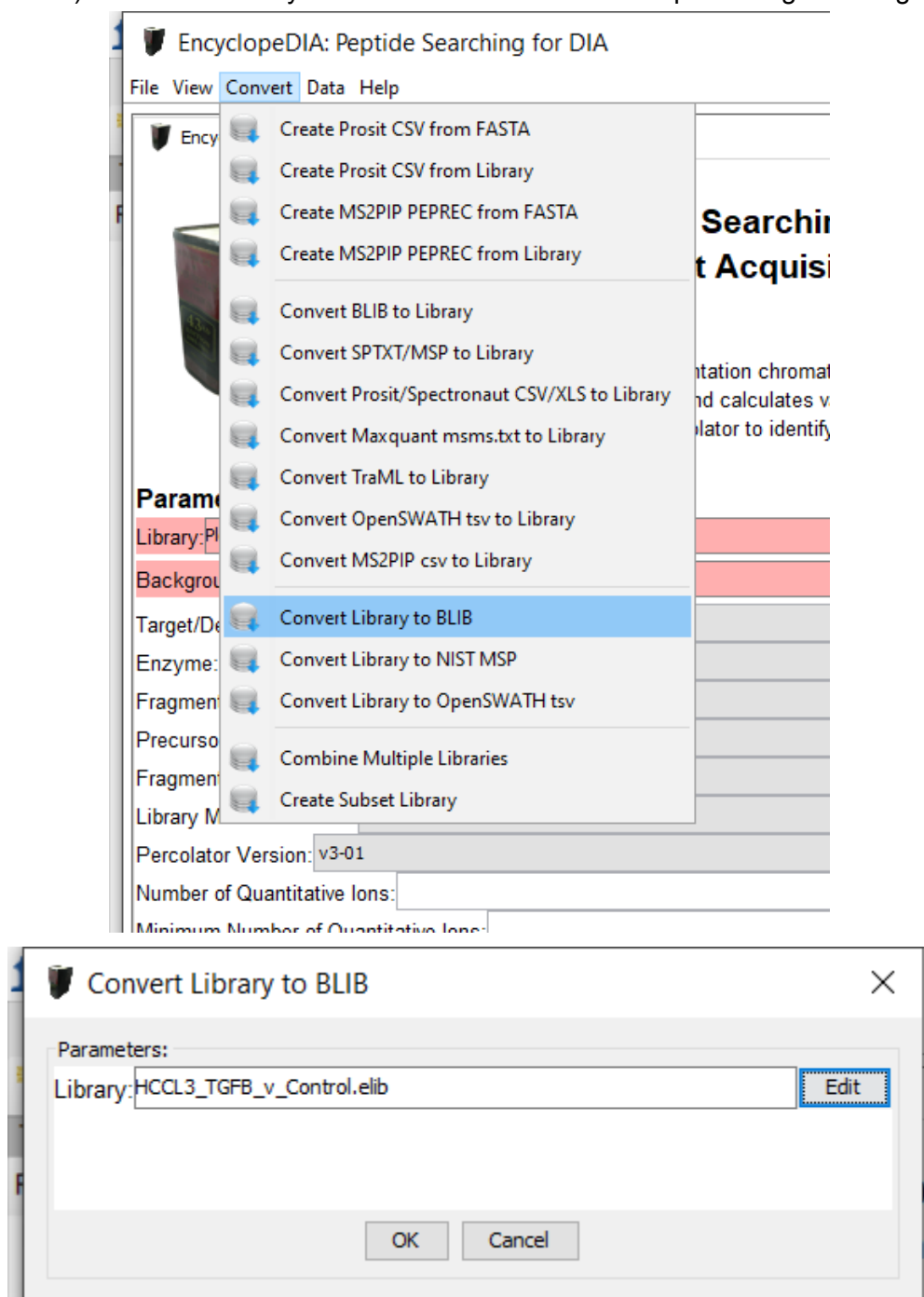
File	Progress
Read 2022_09_01_HeLa_500ng_16mzst_DDA_NCE_22.mzML	Converting files...
Read 2022_09_01_HeLa_500ng_16mzst_DDA_NCE_27.mzML	
Read 2022_09_01_HeLa_500ng_16mzst_DDA_NCE_32.mzML	
Read 2022_09_01_HeLa_500ng_16mzst_DDA_NCE_37.mzML	
Read 2022_09_01_HeLa_500ng_16mzst_DDA_NCE_42.mzML	

B. Search results can also be saved for downstream quantitative assessment using the “Save Quant Reports” button. RAW files between these experiments are expected to contain shared peptides so retention-time alignment is performed and match-between-runs quantification is calculated. This should generate a quant report (peptide.txt and protein.txt),



C. Additionally, save a BLIB version that we can put into Skyline. This should generate an indexed retention time database (.irtdb), and library (.blib) files.

D. (Alternative to C) Convert > Library to BLIB and select the ELIB output from generating a quant report.



This will generate an .irtdb and .blib files.

```

Console:
Java version: 1.8.0_381
Memory allocated: 14548 MB
EncyclopeDIA Graphical Interface (version 2.12.30)
Found 273612 entries from HCCL3_TGFB_v_Control.elib. Writing to
[C:\Users\Admin\Downloads\HCCL3_TGFB_v_Control.blib]...
Writing paired irtdb file [C:\Users\Admin\Downloads\HCCL3_TGFB_v_Control.irtdb]...
Adding VTIAQGGVLPNIQAVLLPK to anchor iRT peptides
Adding AVFVDLEPTVIDEVR to anchor iRT peptides
Adding VFLENVIR to anchor iRT peptides
Adding IGGIGTVPVGR to anchor iRT peptides
Adding AGLQFPVGR to anchor iRT peptides
Adding NTGIIC[+57.021464]TIGPASR to anchor iRT peptides
Adding VAPEEHPVLLTEAPLNPK to anchor iRT peptides
Adding LLLPGELAK to anchor iRT peptides
Adding IIIPEIQK to anchor iRT peptides
Adding EITALAPSTMK to anchor iRT peptides
Finished reading HCCL3_TGFB_v_Control.blib

```

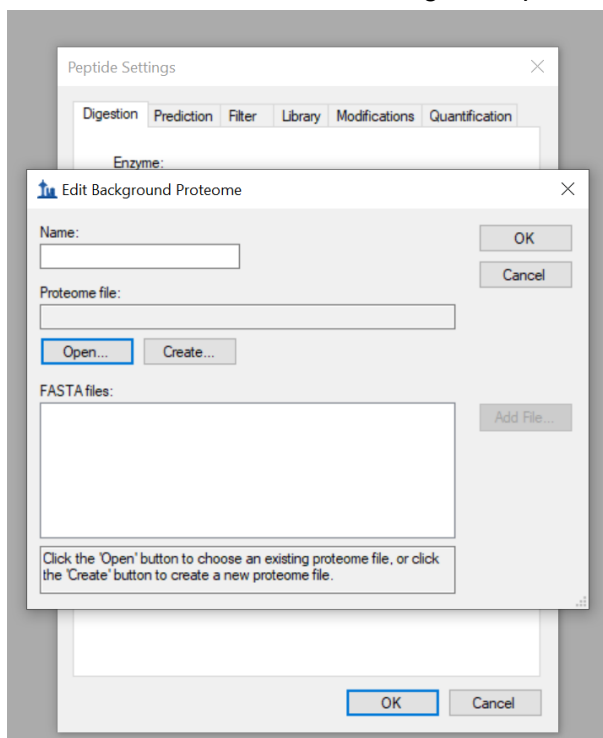
878 of 14548 MB u

The files you will obtain

 HCCL3_TGFB_v_Control.blib	8/14/2023 9:56 AM	BLIB File	79,544 KB
 HCCL3_TGFB_v_Control.irtdb	8/14/2023 9:56 AM	IRTD File	2,916 KB

7. IMPORTING DETECTIONS INTO SKYLINE

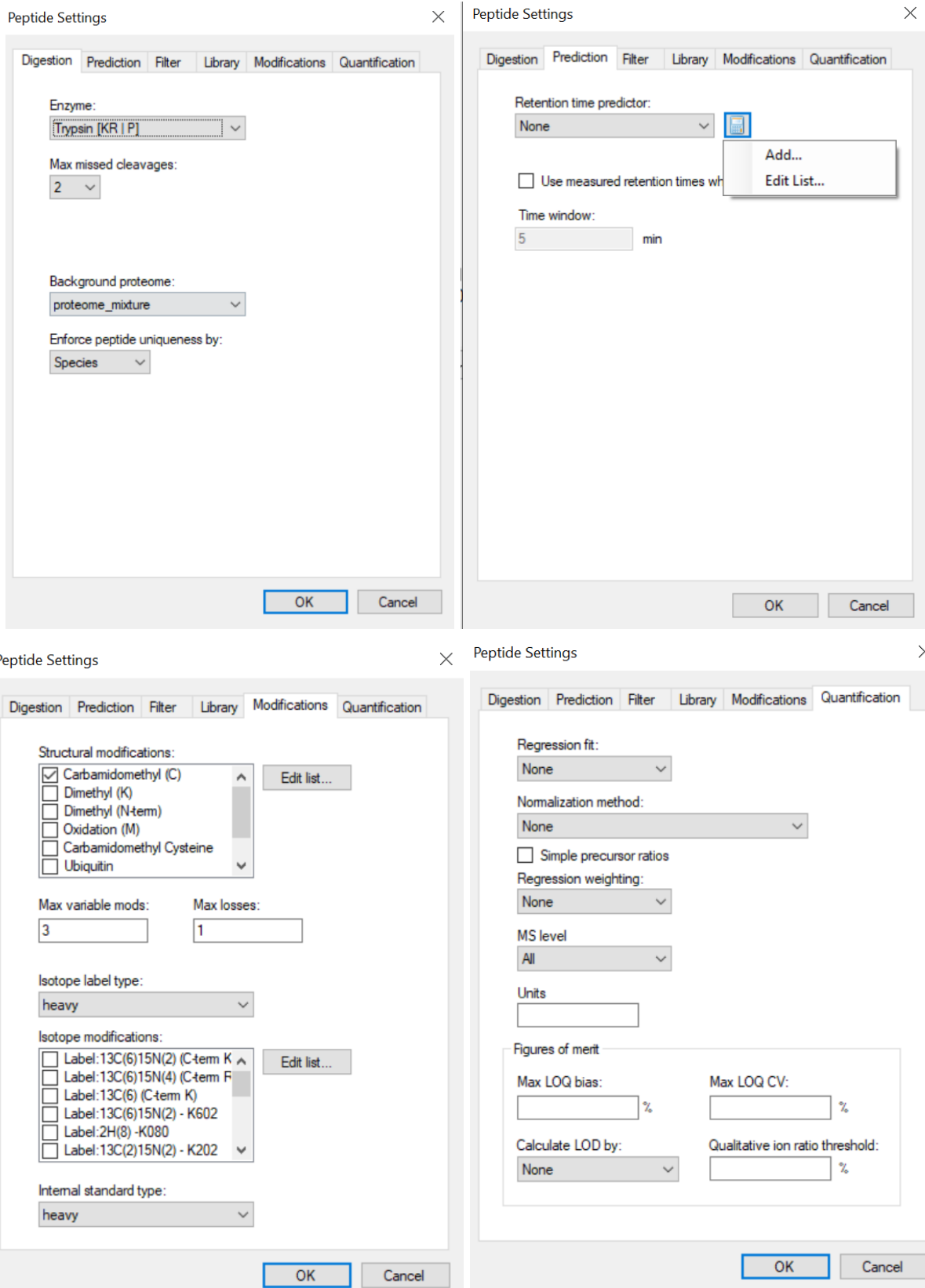
- A. The first window that pops up will be the “Digestion tab.” We will add a background proteome. To do so, click the down arrow and select “add” to make a background proteome. You will see this window:



Click “Open...” and find the FASTA file named “Human-9606-uniprot-proteome-UP000005640-canonical-2022_04.fasta.” You will see proteins importing from the FASTA file.

Once loaded, click “OK.” You will get a warning telling you that there are repeated peptide sequences. This is because the FASTA file was formed from separate FASTAs for each species, concatenated together. Proteins across species may share sequences. Click “OK” to accept the warning.

B. The rest of your Peptide settings will look as follows. Skip the Library tab for now, we will come back to that tab.



B. Set transition settings - set to as follows:

Transition Settings

Prediction

Filter

Library

Instrument

Full-Scan

Ion Mobility

Precursor mass:

Monoisotopic

Product ion mass:

Monoisotopic

Collision energy:

None

Decustering potential:

None

Optimization library:

None

Compensation voltage:

None

☐ Use optimization values when present

OK

Cancel

Transition Settings

Prediction

Filter

Library

Instrument

Full-Scan

Ion Mobility

Peptides

Precursor charges:

2, 3

Ion charges:

1, 2

Ion types:

y, b, p

Product ion selection

From:

ion 3

To:

3 ions

Special ions:

☐ N-terminal to Proline
☐ C-terminal to Glu or Asp
☐ iTRAQ-114
☐ iTRAQ-115
☐ iTRAQ-116
☐ iTRAQ-117

Edit List...

☒ Use DIA precursor window for exclusion
☒ Auto-select all matching transitions

OK

Cancel

Transition Settings

Prediction

Filter

Library

Instrument

Full-Scan

Ion Mobility

Ion match tolerance:

0.05

m/z

☒ If a library spectrum is available, pick its most intense ions

Pick:

9

product ions

6

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

Transition Settings

Prediction

Filter

Library

Instrument

Full-Scan

Ion Mobility

Min m/z:

50

m/z

Max m/z:

1500

m/z

☒ Dynamic min product m/z

Method match tolerance m/z:

0.005

m/z

Firmware transition limit:

Firmware inclusion limit:

Min time:

min

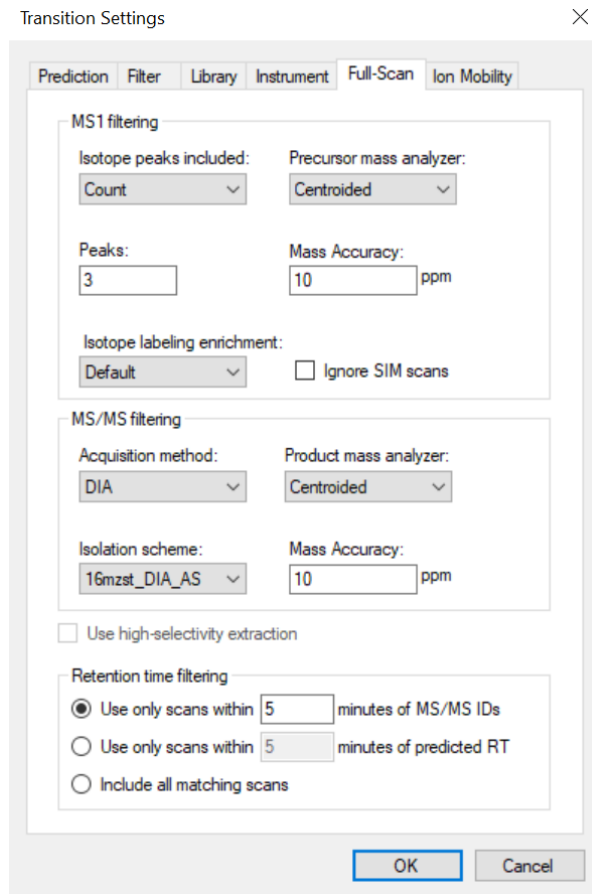
Max time:

min

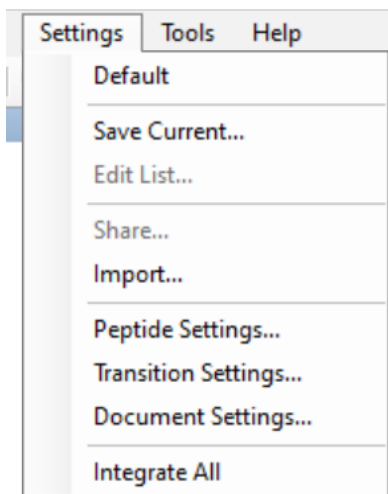
☒ Triggered chromatogram acquisition

OK

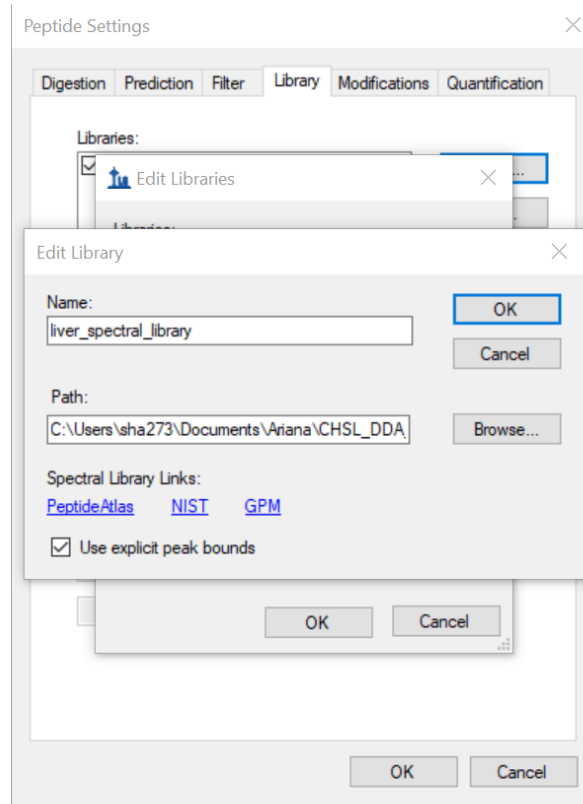
Cancel



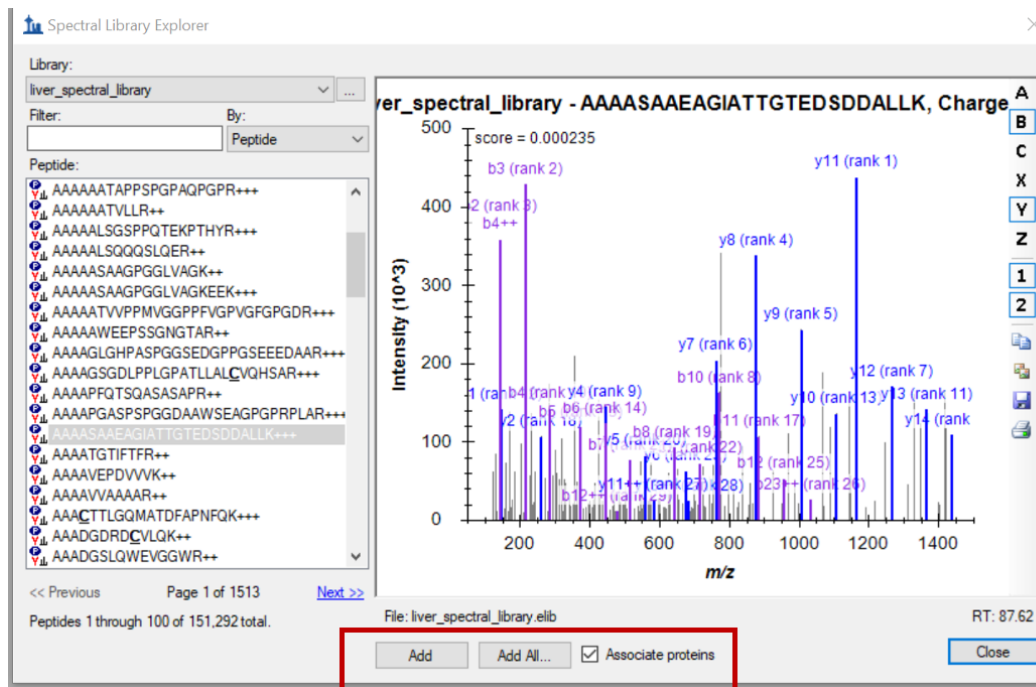
- C. We will ignore the ion mobility window since we did not use any ion mobility with the data acquisition. Click OK to get back to your Skyline document. Next, go to Settings > Integrate all. This will check the integrate all and ensure that all peptides imported into the document will be integrated.



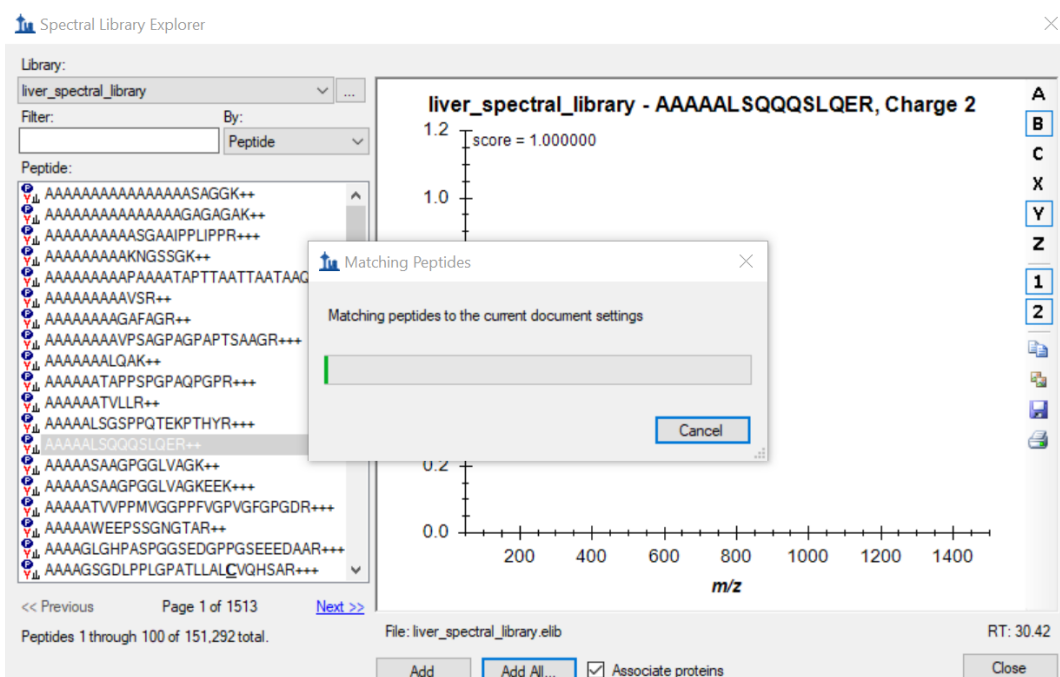
- D. First, we will import the spectral library. Go to Peptide Settings > Library and click Edit List > Add. Then select Browse and find the "liver_spectral_library.blib" file. Name the library file "liver_spectral_library."



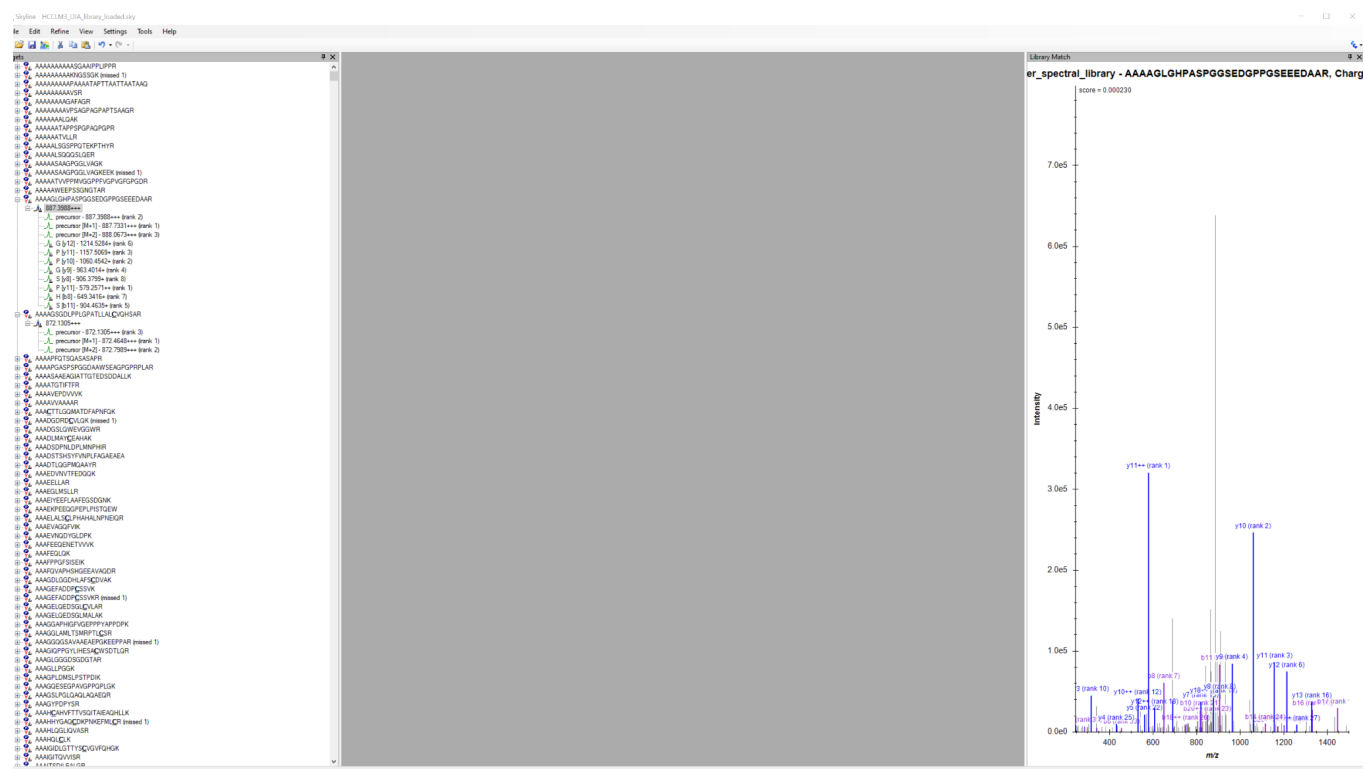
You will be taken back to the Library tab. Check the box next to the library you just imported, and select Explore. This will pop up the spectral library viewer



Check "Associate Proteins" and click the "Add All" button at the bottom of the Spectral Library Explorer window. This will add peptides from the spectral library to Skyline. The screen will look like this while is loading peptides according to the Transition and Peptide settings we have set.



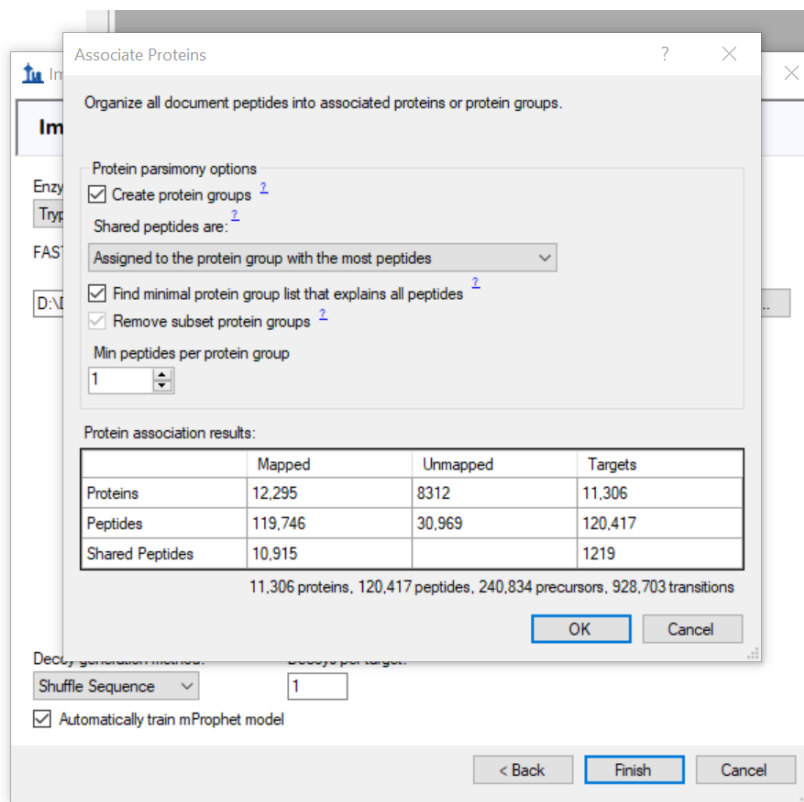
You will get a message that a portion of the peptides do not match the filter settings. We will add these for now, and then perform filtering downstream.



E. Go to File > Import Peptide Search. The first window that pops up will give you the option to generate a library or make a pre-made library. We will select a pre-made library and click on DIA. Find the .BLIB that we printed after the quant report that was generated from EncyclopeDIA.

- F. Next, select the .mzML files used in the search of our individual samples.
 - a. EMT_R1.raw
 - b. EMT_R2.raw
 - c. EMT_R3.raw
 - d. Control_R1.raw
 - e. Control_R2.raw
 - f. Control_R3.raw
- G. Your settings should be set to reflect the following for this search:
- H. Find the Human FASTA we set as the background proteome, and select this for the FASTA.
Ensure the enzyme is set to trypsin, the decoy generation method is Shuffle, and the number of Decoys is set to 1. Automatically train mProphet model should also be checked.

To associate proteins, we will check the box to “Create protein groups,” and click the drop-down menu to set Shared peptides to be “Assigned to the protein group with the most peptides,” and check “Find minimal protein list to explain all peptides”, which will also select the “Remove subsetted proteins” box. In general, we want to filter out duplicated peptides/peptides that match more than one proteins.



- I. You will see your data importing, and once it is in Skyline, an mProphet module will pop up. Train the mProphet according to the decoys generated or second best peak.
- J. Your loaded document should contain the full list of peptides.