Using EncyclopeDIA and Prosit to Make Empirically-Corrected Peptide Libraries

Walkthrough based on EncyclopeDIA version 0.8.3, last update on June 22, 2019
Contact Brian Searle (bsearle@systemsbiology.org) with questions or bugs

Data-independent acquisition (DIA) mass spectrometry is a powerful label-free proteomics technique. DIA methods typically rely on sample-specific spectrum libraries from deeply fractionated data-dependent acquisition (DDA) experiments. This tutorial will walk you through how to create equal- or higher-quality libraries using only DIA data.

Prerequisites

Prosit is a web application that is accessible from any operating system without installation. EncyclopeDIA is a cross-platform Java application that has been tested for Windows, Macintosh, and Linux. EncyclopeDIA requires 64-bit Java 1.8 or higher. If you don't already have it, you can download "Windows x64 Offline" from:
http://www.oracle.com/technetwork/java/javase/downloads/jre8-downloads-2133155.html

After you have 64-bit Java 1.8, 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.

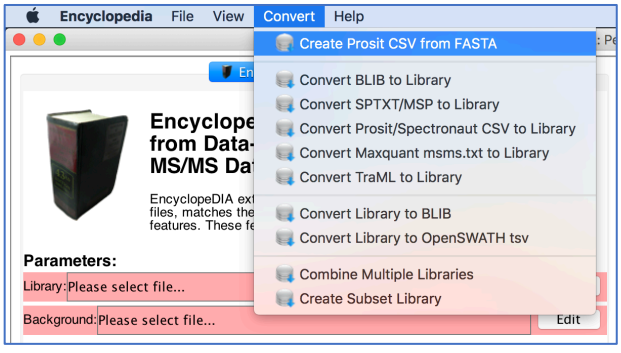
Collecting DIA Data

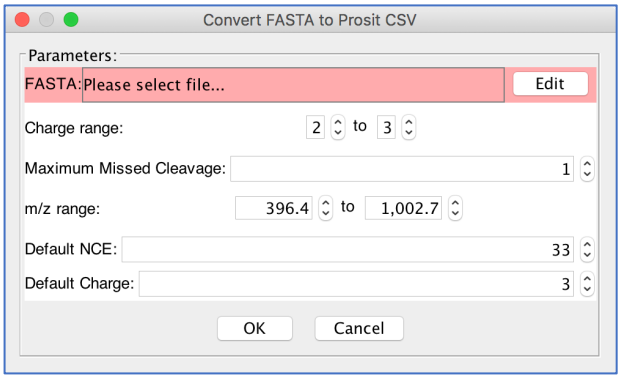
We recommend consulting this general-purpose overview for collecting DIA data:
https://bitbucket.org/searleb/encyclopedia/downloads/dia%20methods%20setup.pdf

We recommend these parameters for collecting single-injection and gas-phase fractionated DIA:
https://docs.google.com/spreadsheets/d/1A8AQlmLroAkQcAcsiGTNvnGBE2lGpkMwhh0YLTBHXKA

Building Prosit CSV Input with EncyclopeDIA

The "Convert/Create Prosit CSV from FASTA" menu option launches a dialog for building a Prosit CSV. Select the input FASTA, charge range, the number of maximum missed cleavages, and the m/z range of interest. In general, because FDR is calculated based on the total number of peptides searched, we recommend using small FASTA databases (fewer than 25,000 entries) and narrow charge ranges to shrink the necessary search space:





Based on how you acquired your DIA experiment, set the target NCE and default charge state for the prediction. With DIA all peptides are fragmented assuming they are the same charge. For peptides that are not the default charge state, EncyclopeDIA back calculates what the Prosit NCE should be. NCE settings are based on the Thermo Fusion Lumos mass spectrometer. The default (33 NCE) should work in most cases, but if you'd like to fine-tune this setting then the following conversions are a good starting point:

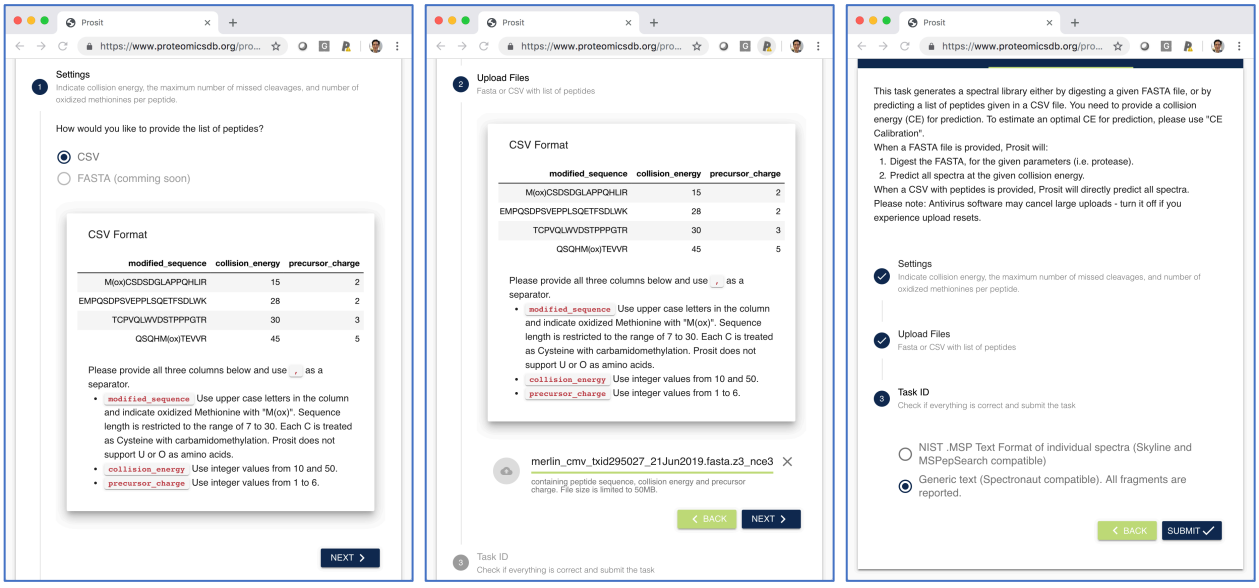
- Fusion-class orbitrap instruments: Actual NCE setting
- Q-Exactive-class orbitrap instruments: Actual NCE setting + 6
- ToF instruments: EV/2.5

This will create a CSV file in the same directory as your FASTA.

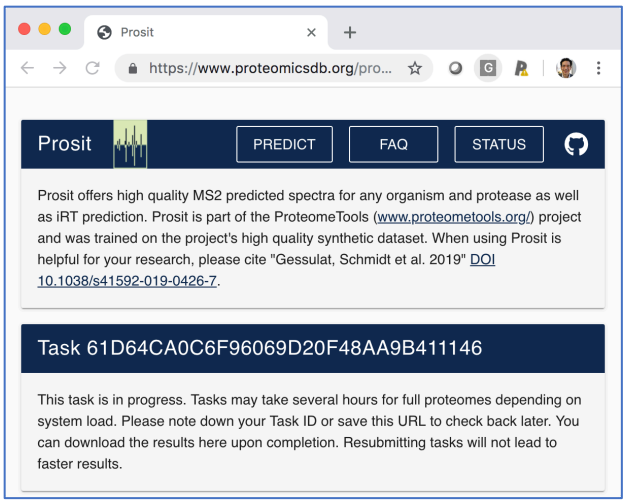
Generating Prosit Predictions

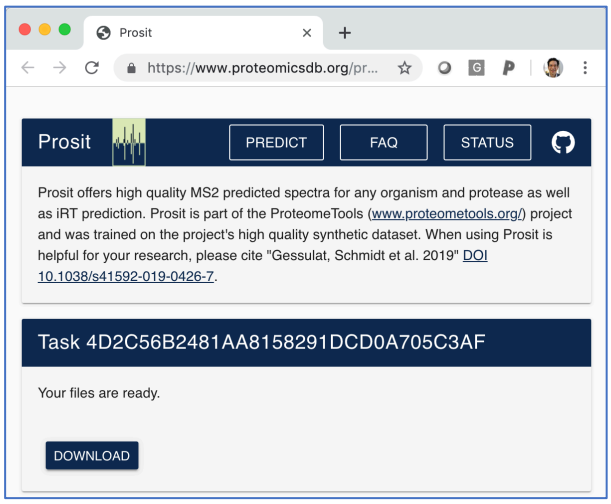
First, go to the Prosit website at https://www.proteomicsdb.org/prosit/. Using the "Spectral Library" tab and follow the three steps:

1. "Settings": select "CSV" to provide the list of peptides, then hit "next"
2. "Upload Files": click the cloud icon to upload the CSV from EncyclopeDIA, then hit "next"
3. "Task ID": select "Generic text" for the return format and hit "submit"



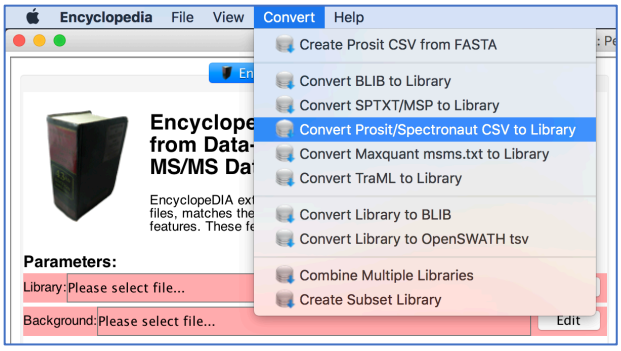
After the job has been submitted, record your Task ID. Depending on the size of the FASTA, your task may take hours to days to process. You can refresh the URL to check if your job has finished, or save the URL to check back at a later time. Once your task is complete, download the resulting files:

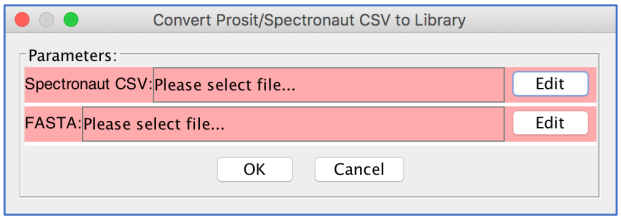




Creating an EncyclopeDIA Library from Prosit CSV Output

The “Convert/Create Prosit CSV from FASTA” menu option launches a dialog for building a predicted library. Upload the Prosit output CSV and your original FASTA database:

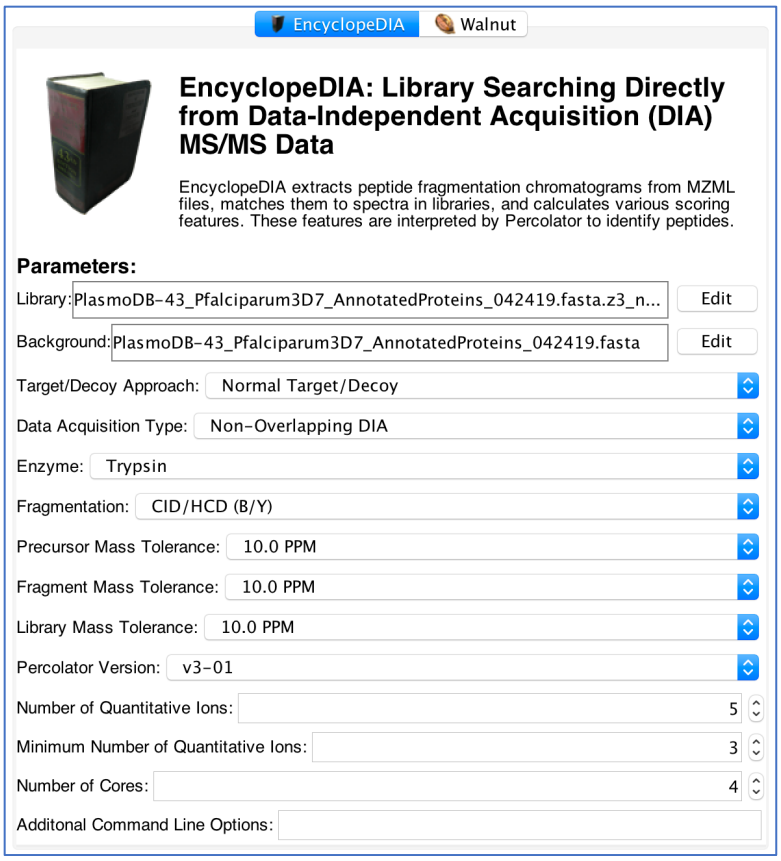




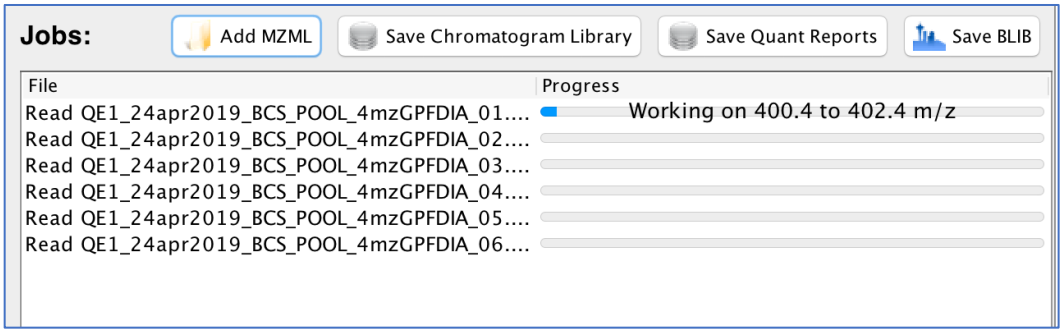
This will create a DLIB library in the same directory as your CSV.

Making an Empirically-Corrected Library from Gas-Phase Fractionated Runs

Using EncyclopeDIA parameter tab, specify the DLIB library and your FASTA database. Then set up your EncyclopeDIA search using settings appropriate for your experiment. For example, the following settings are appropriate for most orbitrap mass spectrometers:



Then, queue up mzML files from your gas-phase fractionated runs. We recommend using MSConvert (Proteowizard) for building vendor-neutral mzML files from your vendor-specific raw files:



Once your files have finished processing, press the button “Save Chromatogram Library” to create your empirically-corrected library in an ELIB format. You can use this ELIB library in EncyclopeDIA, Skyline, or Scaffold DIA to analyze your single-injection DIA experiments.