

# Peptide detection with EncyclopeDIA

Tutorial based on EncyclopeDIA version 0.6.14, last update on March 2, 2018

Written by Brian Searle ([searleb@uw.edu](mailto:searleb@uw.edu))

Data independent acquisition (DIA) mass spectrometry is a powerful technique that is improving the reproducibility and throughput of proteomics studies. EncyclopeDIA is library search engine comprised of several algorithms for DIA data analysis and can search for peptides using either DDA-based spectrum libraries or DIA-based chromatogram libraries. EncyclopeDIA contains Walnut, an implementation of the PECAN scoring system to enable chromatogram library generation from FASTA protein sequence databases when spectrum libraries are unavailable.

## PREREQUISITES

EncyclopeDIA is a cross-platform Java application that has been tested for Windows, Macintosh, and Linux. EncyclopeDIA requires 64-bit Java 1.8. 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.

## ENCYCLOPEDIA

The screenshot displays the EncyclopeDIA software interface. The main window is titled "EncyclopeDIA: Peptide Searching for DIA". The interface is divided into several sections:

- EncyclopeDIA: Library Searching Directly from Data-Independent Acquisition (DIA) MS/MS Data**: A brief description of the software's function.
- Parameters:** A list of configuration options with input fields and dropdown menus:
  - Library: 2017dec27\_normal\_dia\_6e\_rep1.dia.elib
  - Background: uniprot-9606.fasta
  - Target/Decoy Approach: Normal Target/Decoy
  - Data Acquisition Type: Non-Overlapping DIA
  - 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: 4
- Jobs:** A section for managing search jobs, including buttons for "Add MZML", "Save Chromatogram Library", "Save Quant Reports", and "Save BLIB".
- Progress:** A section showing the current search progress, including a status bar at the bottom indicating "1171 of 3641 MB used".

EncyclopeDIA comes with a user-friendly GUI interface. The left pane contains search options, while the right pane contains a process queue. The bottom left contains a console that provides specific information about what EncyclopeDIA is doing at any given moment.

## ENCYCLOPEDIA SEARCH OPTIONS

### Parameters:

Library:	<input type="text" value="2017dec27_normal_dia_6e_rep1.dia.elib"/>	<input type="button" value="Edit"/>
Background:	<input type="text" value="uniprot-9606.fasta"/>	<input type="button" value="Edit"/>
Target/Decoy Approach:	<input type="text" value="Normal Target/Decoy"/>	<input type="button" value="⌵"/>
Data Acquisition Type:	<input type="text" value="Non-Overlapping DIA"/>	<input type="button" value="⌵"/>
Enzyme:	<input type="text" value="Trypsin"/>	<input type="button" value="⌵"/>
Fragmentation:	<input type="text" value="CID/HCD (B/Y)"/>	<input type="button" value="⌵"/>
Precursor Mass Tolerance:	<input type="text" value="10.0 PPM"/>	<input type="button" value="⌵"/>
Fragment Mass Tolerance:	<input type="text" value="10.0 PPM"/>	<input type="button" value="⌵"/>
Library Mass Tolerance:	<input type="text" value="10.0 PPM"/>	<input type="button" value="⌵"/>
Percolator Version:	<input type="text" value="v3-01"/>	<input type="button" value="⌵"/>
Number of Quantitative Ions:	<input type="text" value="5"/>	<input type="button" value="⌵"/>
Minimum Number of Quantitative Ions:	<input type="text" value="3"/>	<input type="button" value="⌵"/>
Number of Cores:	<input type="text" value="4"/>	<input type="button" value="⌵"/>

EncyclopeDIA has several options for searching files. Before you can start loading data, you need to specify both a library to search as well as a background. These will be shaded red until they are properly specified.

Library:	<input type="text" value="Please select file..."/>	<input type="button" value="Edit"/>
Background:	<input type="text" value="Please select file..."/>	<input type="button" value="Edit"/>

Libraries can be either in the .ELIB (chromatogram library) or .DLIB (spectrum library) format. Spectrum libraries from Skyline, NIST, or in the HUPO PSI format TraML can be converted to .DLIB using the “Convert” menu.

Background files are .FASTA formatted protein databases and provide the necessary peptide-to-protein links not specified in the spectrum library.

Other EncyclopeDIA 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”.

**Data Acquisition Type:** You can specify “Overlapping DIA” or “Non-Overlapping DIA”. In general, we recommend processing overlapping DIA experiments with the PRISM algorithm to deconvolute overlapping windows in MSConvert and specifying “Non-Overlapping DIA” in EncyclopeDIA. However, it is possible to load overlapping window data directly and EncyclopeDIA will use a heuristic-based deconvolution scheme to deconvolute the data.

**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 Quantitative Ions:** This setting is the maximum number of ions integrated for quantitation. In general, we recommend setting this to 5.

**Minimum Number of Quantitative Ions:** This setting requires a minimum number of ions that fit the peak- shape of localizing ions. In general, we recommend setting this to 3.

**Number of Cores:** This is the number of CPU cores you allow EncyclopeDIA to use.

## ENCYCLOPEDIA QUEUE

File	Progress
Read 2017dec27_variable_dia_6b_rep1.dia	Previously found 38220 peptides identified at 1.0...
Read 2017dec27_prm_6b_rep1_180103142426....	Working on 470.9143 to 471.61432 m/z

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, EncyclopeDIA will remember where it left off and try to not process it a second time.

EncyclopeDIA provides three export options. These save options will be placed at the end of the queue and will produce reports on every raw file listed in the queue:

**Save Chromatogram Library:** Narrow-window search results can be saved as a “Chromatogram Library” to enable analyses of other wide-window data acquired at the same time. EncyclopeDIA does not retention-time align between these raw files because in general they do not contain overlapping peptides (e.g. there should be virtually no shared peptides between narrow window acquisitions from 400-500 and 500-600). These results will also contain all detected fragment ions, no matter what setting is made for “Number of Quantitative Ions”.

**Save Quant Reports:** Search results can also be saved for quantitative assessment. These reports are limited to contain only the limited “Number of Quantitative Ions”. These ions are chosen as the most consistent ions across all of the files. In addition, raw files between these experiments are expected to contain shared peptides so retention-time alignment is always performed.

**Save BLIB:** BLIBs can be loaded into Skyline for further visualization.

## WALNUT SEARCH OPTIONS

You can also analyze files with the PECAN algorithm using the Walnut tab.

### Parameters:

Background: UP000005640\_9606.fasta

Edit

Target: uniprot-mtor.fasta

Edit

Target/Decoy Approach: Normal Target/Decoy

Data Acquisition Type: Non-Overlapping DIA

Precursor Window Width (blank=extract from file): -1

Enzyme: Trypsin

Fixed: C+57 (Carbamidomethyl)

Fragmentation: CID/HCD (B/Y)

Precursor Mass Tolerance: 10.0 PPM

Fragment Mass Tolerance: 10.0 PPM

Maximum Missed Cleavage: 1

Percolator Version: v3-01

Number of Quantitative Ions: 5

Number of Cores: 4

Charge range: 2 to 3

Unlike with library searching, here you only need to specify a background and a target .FASTA database. While in general these will be the same .FASTA, in some circumstances it makes sense to analyze a subset of the proteome when asking more targeted questions.

Background: Please select file...	Edit
Target: Please select file...	Edit

Other Walnut-specific options are defined below:

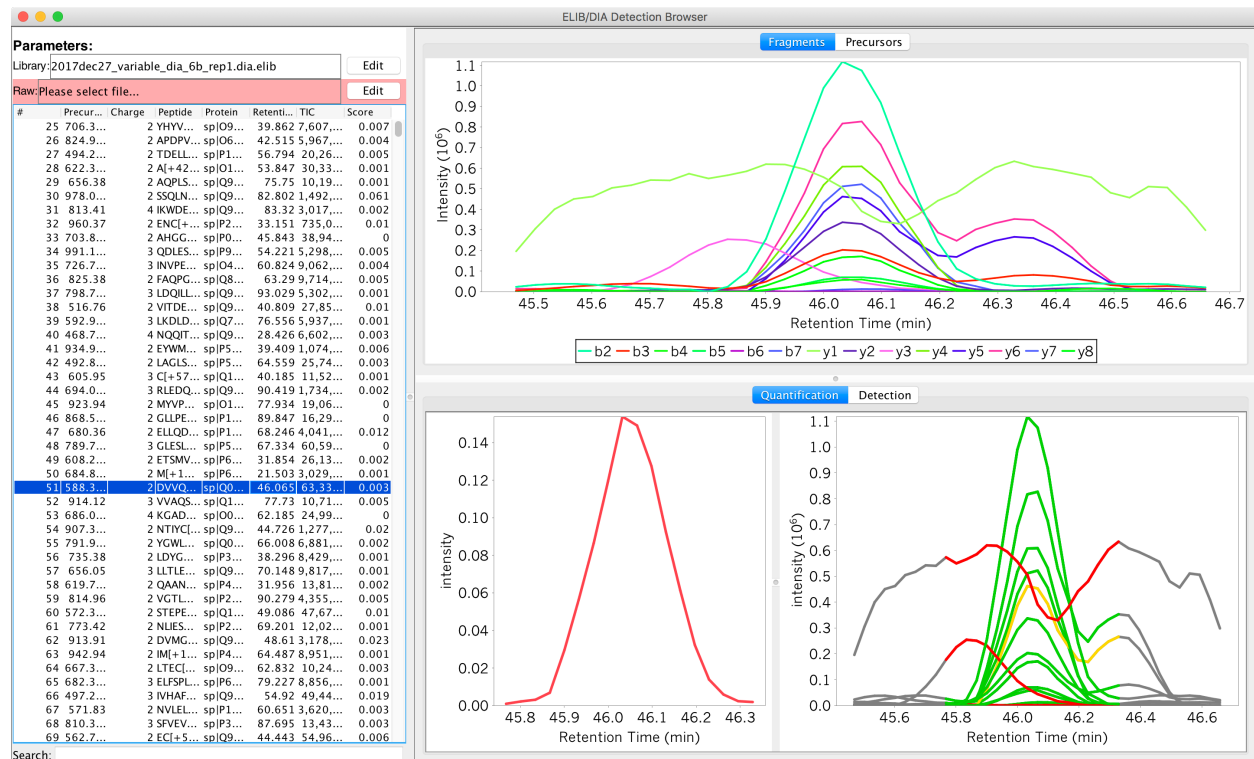
**Fixed:** Several common cysteine alkylation products are supported.

**Fragmentation:** Unlike with EncyclopeDIA, if you are using beam-type fragmentation you will likely get improved results with “HCD (y- only)”.

**Maximum Missed Cleavage:** Often times enzymatic digestion does not go to completion. This setting allows Walnut to consider peptides that contain an internal cleavage site. Searching with more missed cleavages results in lowered discriminatory power, so we recommend that you set this to be no higher than 1.

**Charge range:** This indicates the charge range of peptides to consider. Searching with wider charge ranges also results in lowered discriminatory power, so we recommend that you set this to only +2 to +3, even if you expect +1 or +4 peptides to be present.

## ENCYCLOPEDIA VISUALIZATION



EncyclopeDIA provides several visualization options in the “View” menu. These include:

**ELIB browser:** designed to help analyze acquisition quality using automatically generated .ELIB files for very individual sample.

**Multi-ELIB browser:** designed to help analyze quantitative experiments with multiple samples.

**RAW file browser:** designed to help visualize DIA raw files and troubleshoot data acquisition settings.