Step-by-Step Tutorial for Analyzing DDA and DIA Proteomics Data with Contaminant FASTA and Spectral Libraries

Mass spectrometry-based proteomics is challenged by the presence of contaminant protein background signals. During data analysis, contaminant FASTA libraries allow the search algorithm to distinguish between peptides with similar retention times and *m/z*. In this study, we generated a custom contaminant FASTA and spectral libraries that can be used for both data-dependent acquisition (DDA) and data-independent acquisition (DIA) software. These new contaminant libraries have been shown to reduce false identifications, increase protein IDs, and do not influence protein quantification for both DIA and DDA workflows. We have also modified the contaminant FASTA library to contain a "Cont" prefix before each UniProt identifier, simplifying the process of removing contaminant proteins prior to statistical analysis.

In this tutorial, we describe how to use our contaminant FASTA library with various DDA and DIA software platforms.

Table of Content:

1. Brief Description of Contaminant Libraries

Exogenous contaminant proteins orignated from reagents and sample handling are mostly shared in all bottom-up proteomic experiments. Although widely used for DDA proteomics, the list of common protein contaminants from Maxquant and cRAP list have not been updated for years. These contaminant protein lists contain many incorrect Uniprot IDs, some sample-specific interference proteins that are incorrectly listed as contaminants, and available human protein standards from Sigma-Aldrich which are not contaminant proteins. Therefore, we first built a new contaminant FASTA library by manually merging the available contaminant lists online, updating their Uniprot entry IDs, deleting noncontaminant proteins, searching new contaminant proteins on Uniprot, and combining them into a new FASTA file. Our new contaminant FASTA library contains 381 contaminant proteins including all human keratins and skin-derived proteins, common bovine contaminants from cell culture and affinity columns, various proteolytic enzymes, affinity tags, and other contaminants. When compared to the MaxQuant and cRAP contaminant lists, our new FASTA library is up-to-date for all Uniprot IDs and contains an additional 166 contaminant proteins. This new FASTA library can be used for both DDA and DIA proteomics. We also added a "Cont_" prefix in each contaminant entry in the FASTA library, allowing contaminant proteins to be easily filtered and removed in the result files.

2. Removing Contaminant Proteins from Result Files.

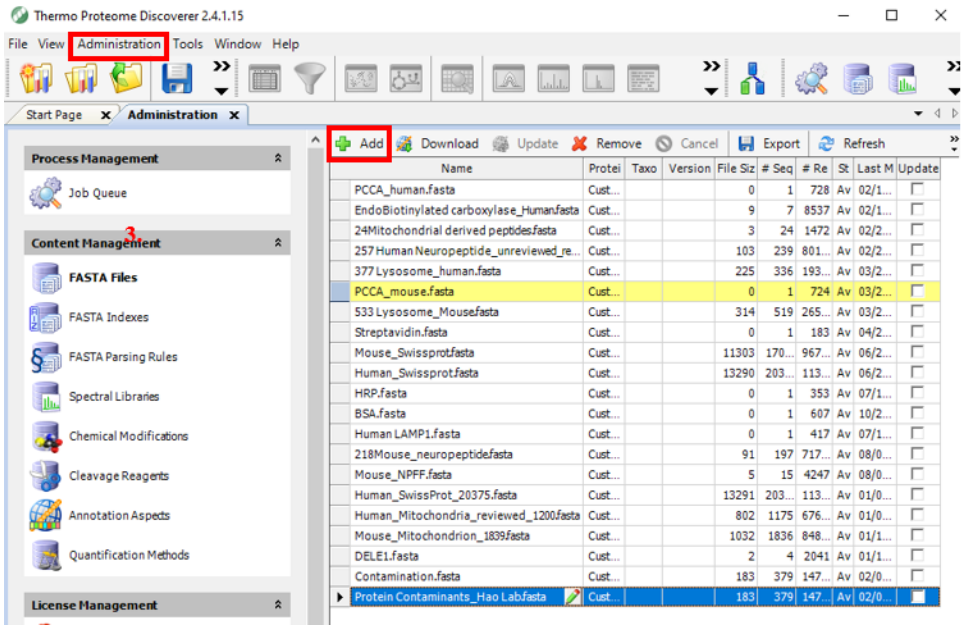
2.1. Launch the results file in Microsoft Excel. In the "Home" tab, click on "Sort & Filter" and then "Filter".

2.2. Navigate to the Protein ID column and type in "Cont".

2.3. This will select all contaminant proteins. All contaminant proteins should be removed prior to statistical analysis.

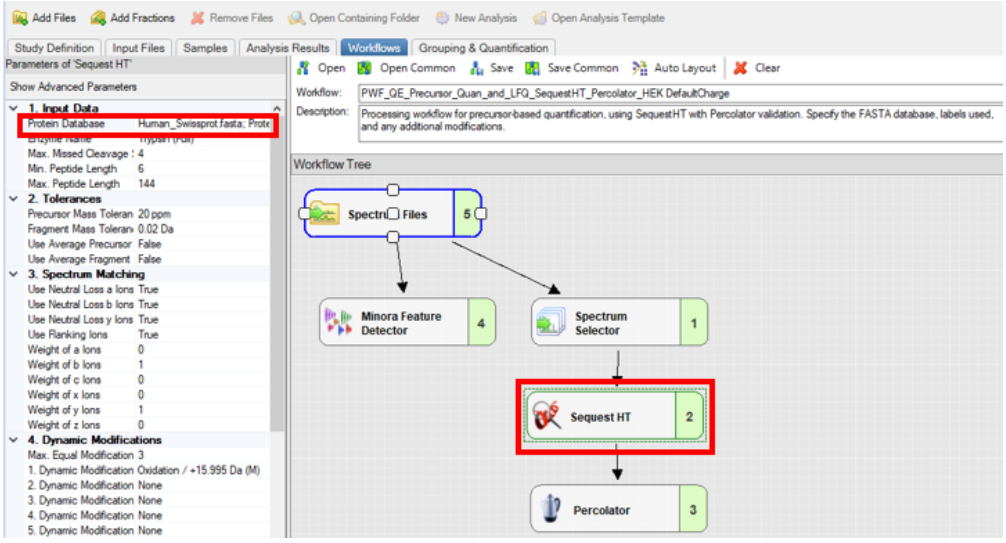
3. Including a Contaminant FASTA library in Proteome Discoverer DDA Workflows

3.1. Click the "Administration" tab and select "Maintain Fasta Files". Click "Add" and then select "Protein Contaminants_Hao Lab.Fasta".

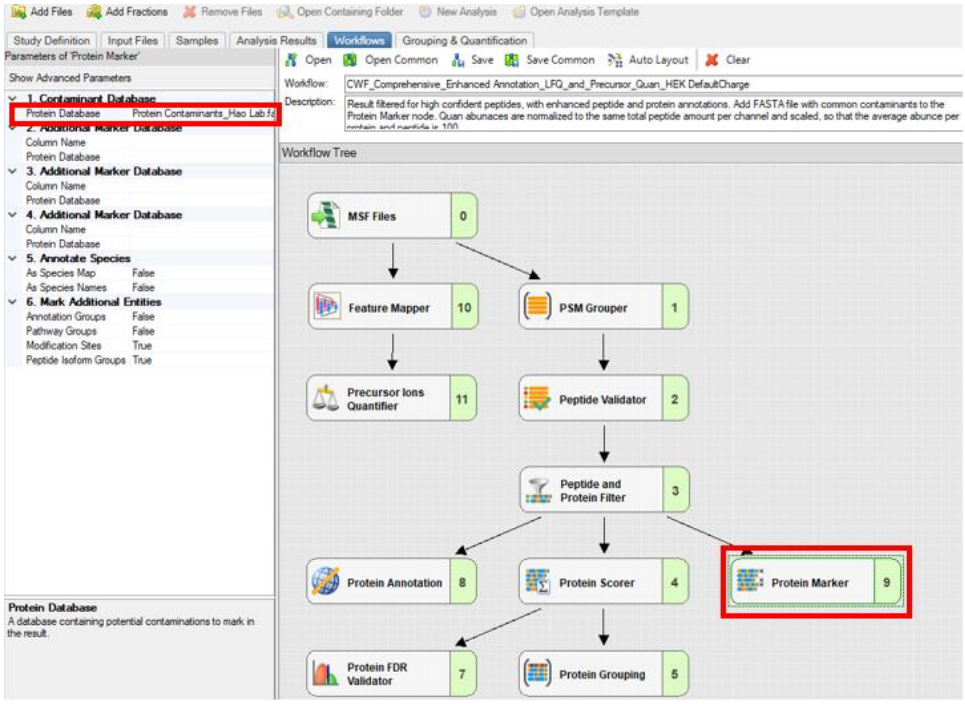


3.2. Open a new study and select a processing step workflow. Click on the "Sequest HT" tab. For protein database, select both the "Protein Contaminants_Hao Lab" and organism FASTA for your sample.

NOTE: The protein contaminant FASTA file must be included to ensure the algorithm does not misassign peptides to the wrong protein.



3.3. Select your consensus step workflow. Under the “Protein Marker” tab, select a contaminant database. This will create a separate column in the result file marking contaminant proteins.



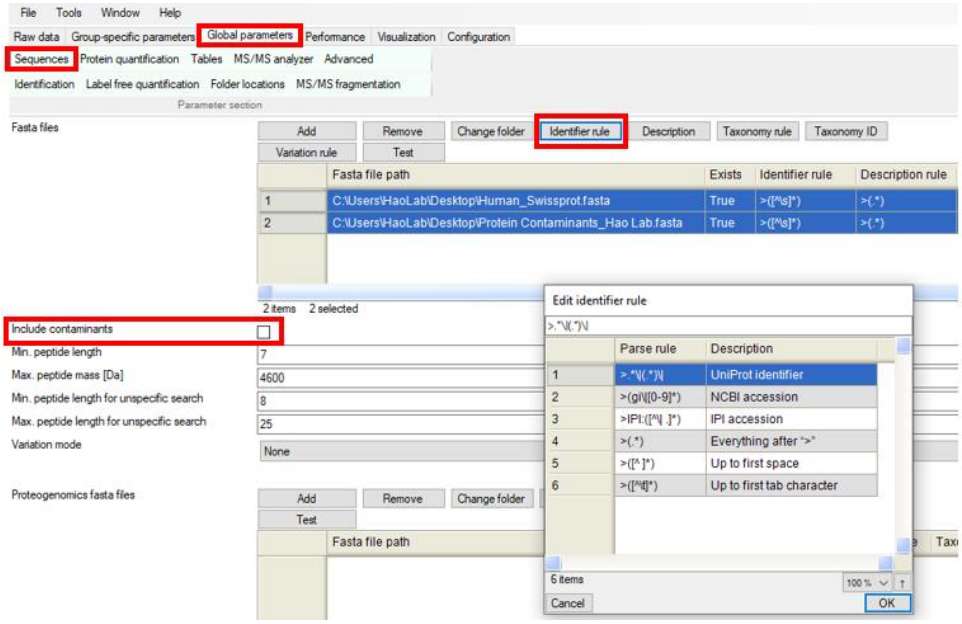
4. Including a Contaminant FASTA library in a DDA MaxQuant Workflow

4.1. Launch MaxQuant. Load *.raw* files. Click the "Global parameters" tab and then select "Sequences".

4.2. Select the "Protein Contaminants_Hao Lab.Fasta" and then click on "Identifier rule".

4.3. Unselect "Include contaminants".

NOTE: Including the MaxQuant contaminant database will not affect results. However, this database includes UniProt IDs that have since been removed or reassigned.



5. Including a Contaminant FASTA library in a MaxDIA Workflow

5.1. For library-based DIA proteomics, you must include the same contaminant and species specific FASTA files used to generate the spectral library. These FASTA files will be included following Steps 2-3.

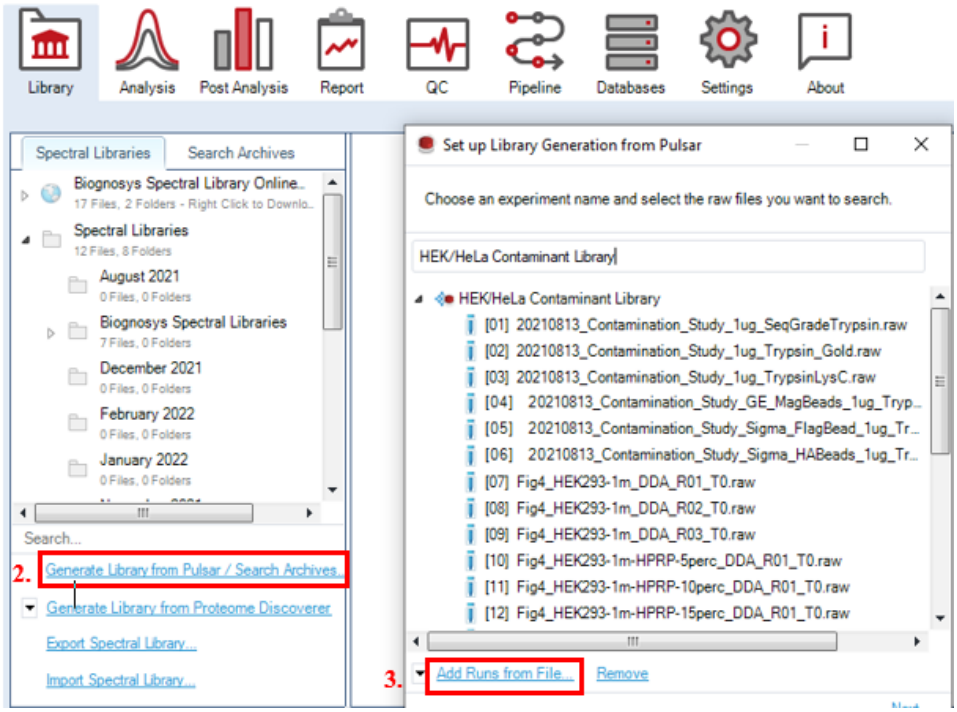
6. Integrating a Contaminant FASTA library into a Library-based Biognosys Spectronaut Workflow

6.1. Launch Biogenesis Spectronaut and select the "Databases" tab. Import the "Protein Contaminants_Hao Lab.Fasta".

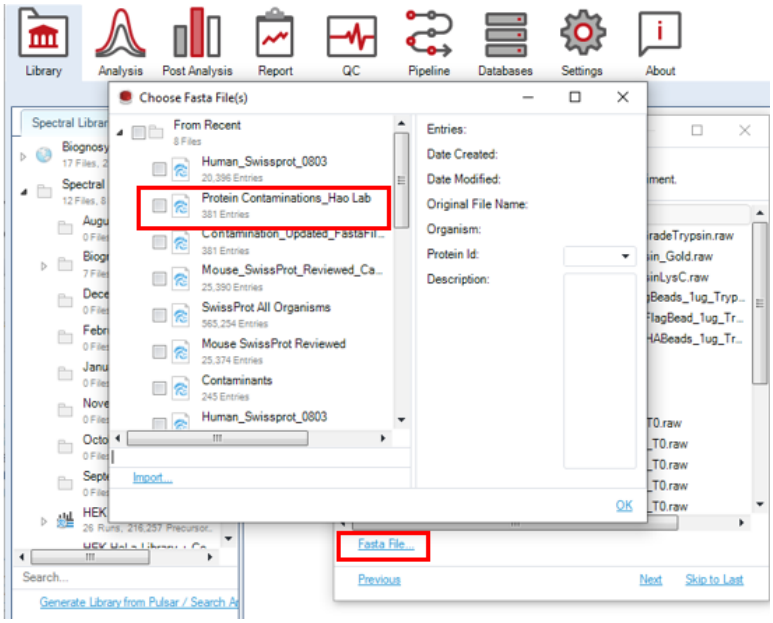
6.2. Select the "Library" tab. Click "Generate Library from Pulsar/Search Archives".

6.3. Select "Add Runs from File" to add *.raw* files.

Note: The *.raw* files from our custom contaminant-only experiment can be included to ensure the accurate detection and inclusion of contaminant spectra within the library.



6.4. Click “Next” and then “Fasta File.” Select the “Protein Contaminantion_Hao Lab FASTA”. Select the remaining settings to build the desired library.

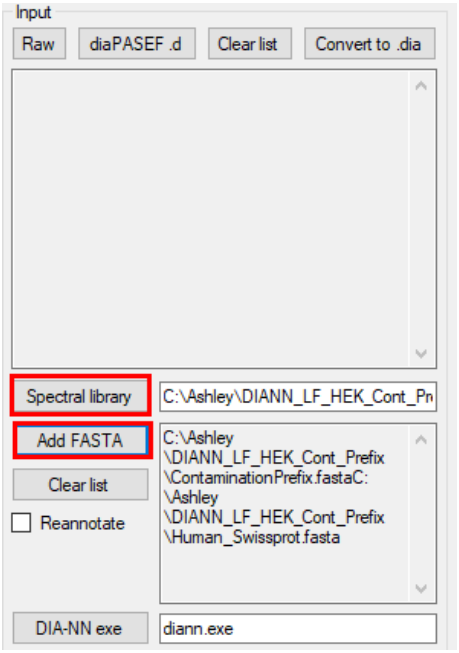


6.5. For library-based DIA proteomics, select the library that was built during data analysis and include the appropriate databases.

7. Integrating a Contaminant FASTA library into a DIA-NN Workflow

7.1. Launch DIA-NN. Click “spectral library” and add the contaminant library that was built using Spectronaut.

7.2. Under “Add FASTA” select the appropriate FASTA libraries that were used to build the spectral library.

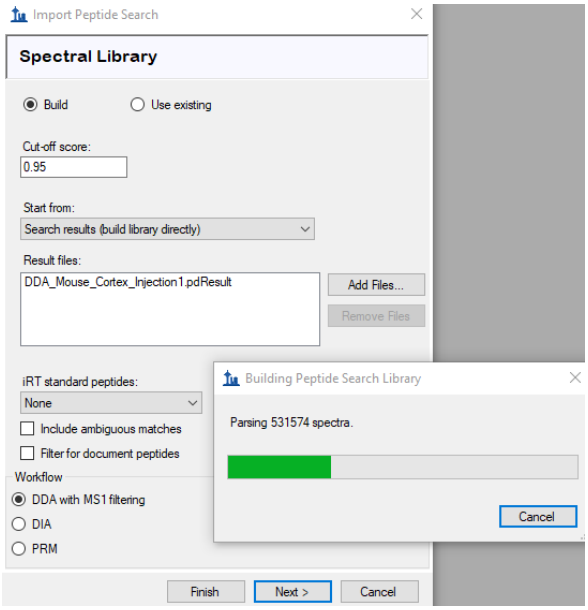


8. Including a Contaminant FASTA library into a Skyline Workflow

8.1. Launch Skyline (version 21.2) and open a "Blank Document".

8.2. A spectral library can be built by selecting "File", "Import" and then "Peptide Search."

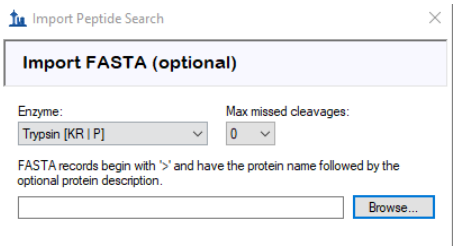
8.3. Import the *.pdResult* file from Proteome Discoverer or *msms.text* file from MaxQuant. Select "Next" to build the process of building the peptide search library.



8.4. Select the appropriate .raw files and click "Next".

8.5. Select the FASTA File and then "Finish".

NOTE: Only a single FASTA library can be imported. The contaminant FASTA file will need to be combined with the organism FASTA.



8.6. Library-based DIA analysis can be conducted using established Skyline workflows. However, the conjoined FASTA file used to build the library should be included during data analysis.

9. Including a Contaminant FASTA library into a PECAN Workflow

9.1. Launch EncylopeDIA (version 1.12.31). Select the Walnut tab.

9.2. Import the contaminant FASTA library to the “Background” and “Target” sections.

NOTE: Only a single FASTA library can be imported into the workflow. The Hao Lab Contaminant library must be combined with your organism FASTA database.