

Mar 01, 2024

## 🌐 BAF\_Protocol\_005 Database Search Proteome Discoverer into Scaffold

DOI

[dx.doi.org/10.17504/protocols.io.q26g7p28kgwz/v1](https://dx.doi.org/10.17504/protocols.io.q26g7p28kgwz/v1)

Nicholas Sherman<sup>1</sup>

<sup>1</sup>UVA Biomolecular Analysis Facility Core



Nicholas Sherman

University of Virginia Biomolecular Analysis Facility Core

### Create & collaborate more with a free account

Edit and publish protocols, collaborate in communities, share insights through comments, and track progress with run records.

Create free account

OPEN ACCESS



DOI: <https://dx.doi.org/10.17504/protocols.io.q26g7p28kgwz/v1>

**Protocol Citation:** Nicholas Sherman 2024. BAF\_Protocol\_005 Database Search Proteome Discoverer into Scaffold. protocols.io <https://dx.doi.org/10.17504/protocols.io.q26g7p28kgwz/v1>

**License:** This is an open access protocol distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

**Protocol status:** Working

We use this protocol and it's working

**Created:** February 13, 2024

**Last Modified:** March 01, 2024

**Protocol Integer ID:** 95131

**Keywords:** Database Searching, Proteome Discoverer, Scaffold Software, Data Display, HeLa Standard, standard database search in proteome discoverer, proteome discoverer, search results into scaffold, thermo raw file, hela digest as an instrument benchmark, standard database search, standard search, hela digest, standard search of 200ng, reproducible data, search result

## Abstract

This protocol lays out the basic steps for taking a Thermo RAW file and doing a standard database search in Proteome Discoverer 2.5+ and putting the search results into Scaffold 5.3+ for display. Included is also the standard search of 200ng of HeLa digest as an instrument benchmark for the Exploris 480. The protocol will go through the most important parameters and settings for obtaining quality, reproducible data. Settings can be changed for other, specific experiments as appropriate.

## Guidelines

Benchmarking your specific instrument with a specific standard that is suitable for a particular experiment type is critical for tracking instrument performance over time and for obtaining high quality, reproducible data. In this example we use a specific standard that is purchased and examine data results for a known amount run on a specific program. You then have a weekly check on how your instrument is performing. You also should have some type of internal standard to check LC performance - as most of our samples are tryptic digests, we use the autolysis peaks.

## Materials

Thermo Proteome Discoverer 2.5 - OPTON31040/CPQ00507094 (now 3.1 cloud based)

Proteome Scaffold Q+S 5.3.3 - Q+S (replacing with Scaffold DDA 6.3+)

Thermo HeLa Digest Standard - 88329

## Troubleshooting



## Proteome Discoverer 2.5 Database Searching

- 1 Thermo RAW files are set up to search in PD 2.5 software (Proteome Discoverer) to produce an output MSF file. The RAW files for an individual project are placed in a folder and a sub folder is created with the MSF files produced by PD 2.5 inside. The MSF files will be loaded for display/analysis in Scaffold 5.3+ software.
- 2 Open PD 2.5 and start a new study.
- 3 Choose:  
Study name - will create sub folder with this name for MSF (and associated) files  
Root Directory - where your RAW files are  
Processing Workflow - template with data processing parameters  
Consensus Workflow - template with output display for PD  
  
The workflow templates can be from the Thermo 'stock' but specific ones should be created for the types of analyses most often performed.
- 4 Click 'Add Files' to add your RAW files. They will then be displayed under the input files tab.
- 5 Processing Workflow (to search every spectrum - standard proteomics):
  1. Spectrum Files - No parameters just to get files
  2. Spectrum Selector (set to just take every scan)  
Precursor Selection - Use MS1 Precursor  
Provide Profile Spec - Automatic  
RT, Scan, Charge State - all 0  
Min Precursor Mass - 600 Da  
Max Precursor Mass - 5000 Da  
Total Intensity - 0  
Min Peak Count - 1  
S/N FT - 0
    - Rest parameters just set to your instrument to take all scans
  3. Sequest HT  
Database - FASTA of your species (must be parsed in PD before can choose - restart PD after parsing and before search set up)  
Enzyme - Trypsin (Full)  
Missed Cleavage - 1  
Min length - 5  
Max length - 144  
Precursor Tolerance - 10 ppm



Fragment Tolerance - 0.02 Da  
Averages set to false  
Neutral loss a,b,y and flanking ions - true  
Weight b,y = 1; rest 0  
Max equal modifications = 3  
Dynamic modification oxidation M  
Static modification carbamidomethyl C  
4. Target Decoy PSM Validator  
Target/Decoy - concatenated  
Strict 0.01  
Relaxed 0.05

- 6 Consensus Workflow:  
Just set to defaults as using Scaffold later to display and parse the data. If you want to see specific display in PD 2.5, then you would need to set parameters here.
- 7 Add your files under the Processing workflow. Click 'By File' so each is run as a separate MSF. Click run to start the process. Under Administration the job queue will display the progress.

## Scaffold 5.3+ Data Filtering and Display

- 8 Run Scaffold 5.3.3 and choose new analysis.
- 9 Choose quantitative technique. Spectral counting is standard but you may choose a labeled technique such as SILAC or TMT.
- 10 Add a sample. Should have unique name, category (control, treatment, etc), description. Names and description help to link sample to the experiment while the category will help later with grouping and data analysis. If the MudPit button is checked, the samples will be combined into one analysis output (i.e. added together). You might want to do that if you cut a gel lane into slices and want to see everything in the sample.
- 11 Add each sample to the queue with associated name, category, and description. When done click 'Next'.
- 12 Enter the database which was used in PD for the search. You will have to index in Scaffold just as you did in PD before you can use a database. Use Legacy LFDR, protein cluster analysis, and pre-compute FDR.
- 13 Click 'Load Data' and allow to run.



- 14 Once all data is loaded, filters may be set using a FDR, Peptide/Protein Prophet or XCorr - or some combination. These are choices made depending on instrument and experiment. For our general settings in proteomics we use - min peptide 1, protein prophet 90%, peptide prophet 60%, DeltaCN 0, Xcorr - +1>1.8, +2>2.0, +3>2.2, +4>3.0. For our data run using the nLC1200 on the Exploris 480 with the above search conditions, this produces a FDR of <1% and a balance of showing the most data with the least amount of false positives. We work in a Core lab with very diverse sample types so these settings will change depending on samples, instruments and investigators.

## HeLa Digest Standard

- 15 At least every week, 200ng of Thermo HeLa digest standard is run on the instrument according to Protocol\_004. Using the above PD 2.5 search and Scaffold 5.3 display, we expect to get ~1800 proteins (2+ peptides), ~2200 proteins (1+ peptides), and ~25,000 PSM. In the RAW file we expect to see 3-5E8 basepeak MS and 2-5E9 TIC with smooth, narrow peak shape. This data is tracked from installation to retirement for the instrument and is particularly important when using a new column or buffer mix or to assess if the instrument needs cleaning. Having this standard ensures that the LC and instrument are functioning normally before investigator samples are run.

## Publication Parameters in Short

- 16 10ppm precursor, 0.02Da fragments, full trypsin, carbamidomethyl Cys fixed, oxidized Met variable. Sequest Xcorr score (+1>1.8, +2>2.0, +3>2.2, +4>3.0), delta CN 0, peptide probability >60%, protein probability >90%, 1 unique peptide. The final data FDR of <1%.