

Proteoform Suite Manual 0.2.0

Lloyd M. Smith Research Group

ps_support@chem.wisc.edu

October 27, 2020

Project Overview

Proteoform Suite is a user interface for the analysis of proteoform MS data. The source-code is openly available at <https://github.com/smith-chem-wisc/proteoform-suite>. This software identifies proteoforms by comparing the intact mass for each observed proteoform to theoretical proteoforms generated from known protein sequences and annotated post-translational modifications (PTMs) as well as to other coeluting observed proteoforms. These comparisons reveal both exact-mass matches and mass differences characteristic of known PTMs. Proteoforms are grouped together based on these mass differences into proteoform families. A proteoform family is the set of proteoforms derived from a given gene. Bottom-up peptide identifications and top-down proteoform identifications can be integrated into the analysis to improve proteoform identifications. The program also quantifies relative proteoform abundances between two conditions by calculating intensity ratios for each identified proteoform. Finally, Proteoform Suite streamlines the visualization of proteoform families as networks in the program Cytoscape.

Key Publications

- Shortreed, M. R. et al. Elucidating Proteoform Families from Proteoform Intact-Mass and Lysine-Count Measurements. *J Proteome Res* **2016**, *15*, 1213–21
- Cesnik, A. J. et al. Proteoform Suite: Software for Constructing, Quantifying, and Visualizing Proteoform Families. *J Proteome Res* **2018**, *17*, 568–578
- Dai, Y. et al. Elucidating Escherichia coli Proteoform Families Using Intact-Mass Proteomics and a Global PTM Discovery Database. *J Proteome Res* **2017**, *16*, 4156–4165
- Schaffer, L. V. et al. Expanding Proteoform Identifications in Top-Down Proteomic Analyses by Constructing Proteoform Families. *Anal Chem* **2018**, *90*, 1325–1333
- Schaffer, L. V. et al. Identification and Quantification of Murine Mitochondrial Proteoforms Using an Integrated Top-Down and Intact-Mass Strategy. *J Proteome Res* **2018**, *17*, 3526–3536
- Dai, Y. et al. Constructing Human Proteoform Families Using Intact-Mass and Top-Down Proteomics with a Multi-Protease Global Post-Translational Modification Discovery Database. *J Proteome Res* **2019**, *18*, 3671–3680
- Schaffer, L. V. et al. Intact-Mass Analysis Facilitating the Identification of Large Human Heart Proteoforms. *Anal Chem* **2019**, *91*, 10937–10942
- Schaffer, L. V. et al. Improving Proteoform Identifications in Complex Systems Through Integration of Bottom-Up and Top-Down Data. *Journal of Proteome Research* **2020**, DOI: 10.1021/acs.jproteome.0c00332

Contents

Overview	2
Table of Contents	3
1 Getting Started	6
1.1 Installation	6
1.2 System Requirements	6
1.3 Basic Workflow	6
2 Load Results: Standard	8
2.1 Choose Analysis	8
2.2 Set Parameters	9
2.3 Load Data Using Drop Down Menu	9
2.4 Deconvolution Results for Identification	10
2.5 Deconvolution Results for Quantification	10
2.6 Protein Databases	11
2.7 Top-Down Hit Results	11
2.8 MetaMorpheus Bottom-Up Unique Peptides	11
2.9 Start Analysis	12
3 Deconvolution	13
3.1 Thermo Deconvolution 4.0	13
3.2 FLASHDeconv	13
3.3 Other	14
4 Top-down Search	15
4.1 TDPortal	15
4.2 MetaMorpheus	15
5 Calibration	17
5.1 Overview	17
5.2 Load Files	17
5.3 Set Parameters	18
6 Theoretical Database	19
6.1 Overview	19
6.2 Run Page	19
6.3 Set Parameters	19
6.4 Results	21
7 Top-Down	24
7.1 Overview	24
7.2 Run Page	24
7.3 Set Parameters	24
7.4 Results	25

8 Raw Experimental Components	30
8.1 Overview	30
8.2 Run Page	30
8.3 Set Parameters	30
8.4 Results	31
9 NeuCode Pairs	34
9.1 Overview	34
9.2 Run Page	34
9.3 Set Parameters	34
9.4 Results	35
10 Aggregated Proteoforms	38
10.1 Overview	38
10.2 Run Page	38
10.3 Set Parameters	38
10.4 Results	39
11 Experiment Theoretical Comparison	43
11.1 Overview	43
11.2 Run Page	43
11.3 Set Parameters	43
11.4 Results	44
12 Experiment Experiment Comparison	49
12.1 Overview	49
12.2 Run Page	49
12.3 Set Parameters	49
12.4 Results	50
13 Proteoform Families	54
13.1 Overview	54
13.2 Run Page	54
13.3 Set Parameters	54
13.4 Results	56
14 Visualizing Proteoform Families in Cytoscape	59
14.1 Overview	59
14.2 Visualize Families in Cytoscape	59
14.3 Proteoform Families Key	60
14.4 Exporting Family Visualizations into Adobe Illustrator	60
15 Identified Proteoforms	61
15.1 Overview	61
15.2 Results	61

16 Quantification	63
16.1 Overview	63
16.2 Run Page	63
16.3 Set Parameters	63
16.4 Results	65
17 Results Summary	70
17.1 Overview	70
17.2 Results	70
References	76

1 Getting Started

1.1 Installation

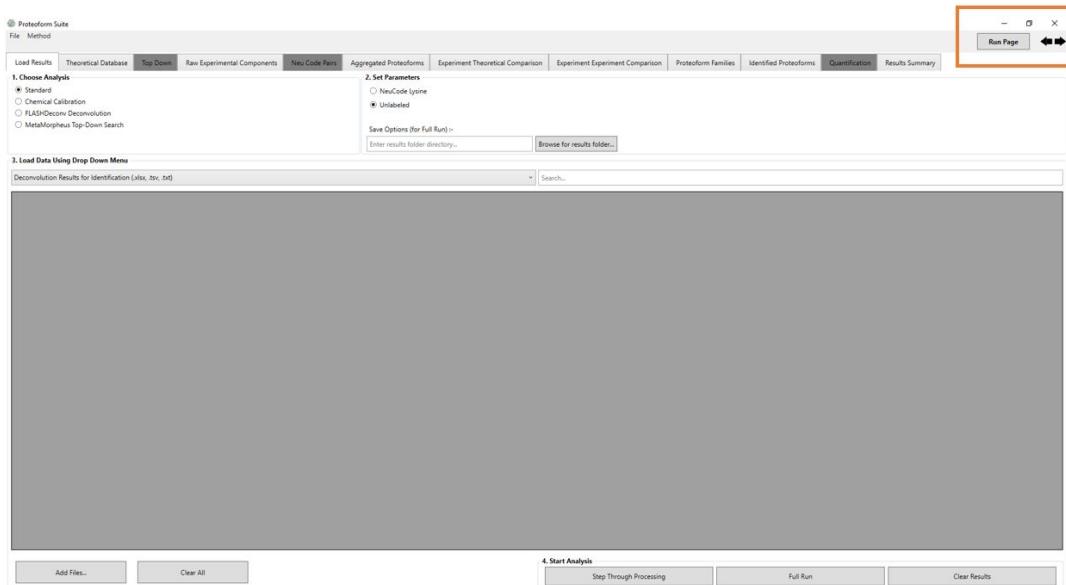
Download a release from <https://github.com/smith-chem-wisc/proteoform-suite>.

1.2 System Requirements

- 8 GB of RAM is recommended for yeast; more RAM is required for larger databases.
- 64-bit operating system
- .NET Core 3.1:
<https://dotnet.microsoft.com/download/dotnet-core/thank-you/runtime-desktop-3.1.3-windows-x64-installer>
- For visualization of proteoform families: Cytoscape^{9,10} version 3.5.0:
<https://cytoscape.org/download.html>
- For visualization of quantitative proteoform families: Need to install enhancedGraphics in Cytoscape using the App Manager under the Tools menu.

1.3 Basic Workflow

- Load results on the Load Results page under Standard analysis
- Click through each tab at the top or use the arrows (top right) to navigate through the pages
- On each page, set parameters and then click the Run Page button (top right)

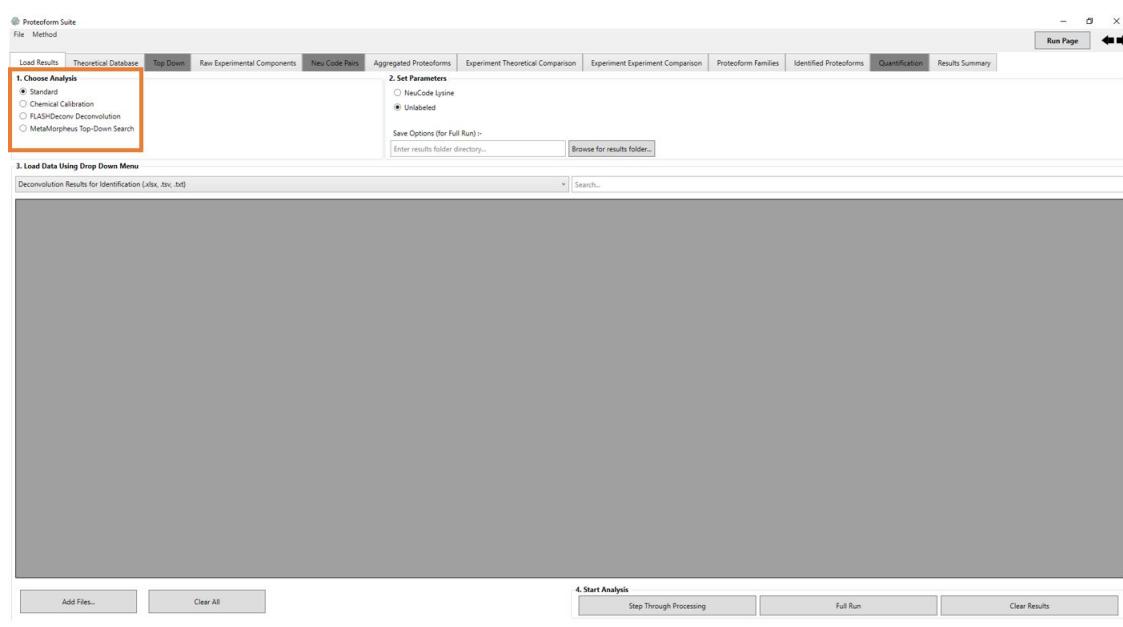


- To save the tables on a specific page: File > Export Table
- To save all tables from all pages: File > Export All Tables
- To save a method .xml (including all parameters, files, file labels): Methods > Save Method
- To load a method .xml (including all parameters, files, file labels): Methods > Load Method

2 Load Results: Standard

This page is where the user loads in deconvolution results, top-down results, bottom-up results, and protein databases for the analysis on subsequent pages. On this page, chemical calibration can be performed on the deconvolution and top-down results (see **Calibration** section). Deconvolution and a top-down search can also be performed to generate deconvolution and top-down results (see **Deconvolution** and **Top-Down Search** sections, respectively). This section will describe the Load Results page for Standard analysis.

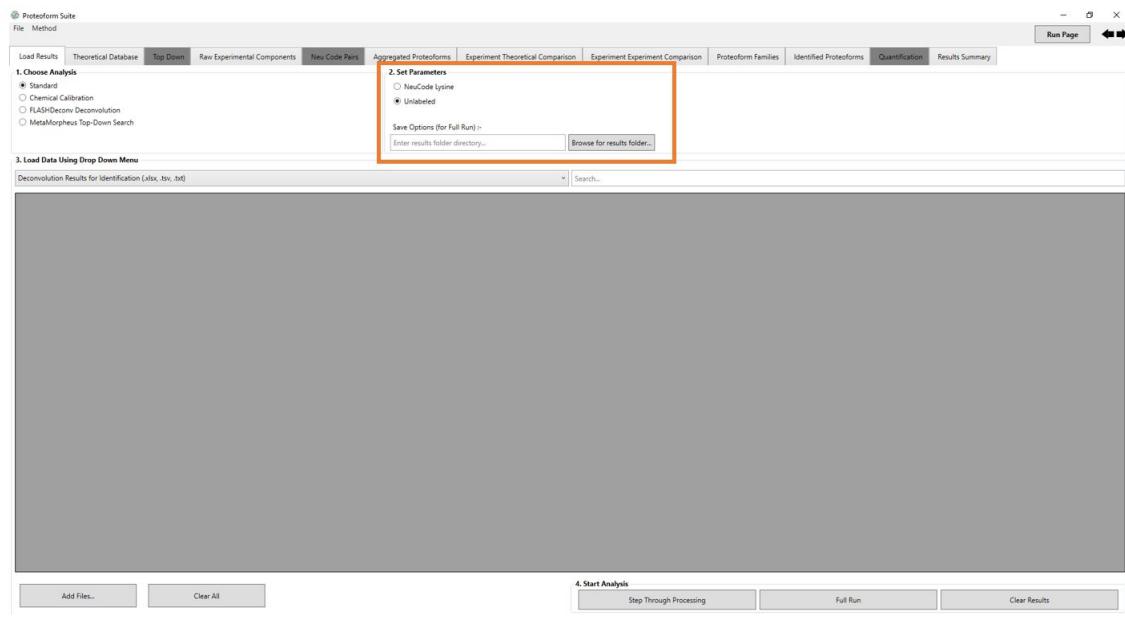
2.1 Choose Analysis



This option selects the type of analysis to perform.

- Standard: load in results under standard before navigating through the different pages
- Chemical Calibration: calibrate mass and retention time of deconvolution and top-down results (see **Calibration** section)
- FLASHDeconv Deconvolution: deconvolute .mzML files (see **Deconvolution** section)
- MetaMorpheus Top-Down Search: search .mzML or .raw files for list of MS/MS identified proteoforms (see **Top-Down Search** section)

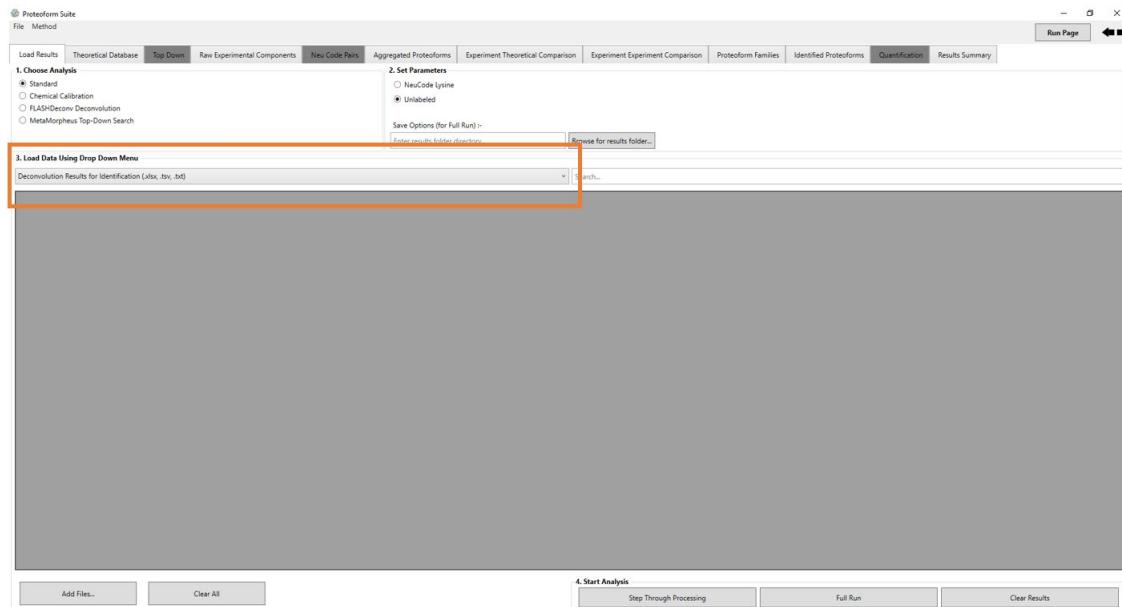
2.2 Set Parameters



- NeuCode Lysine: select if cell culture was performed with heavy and light NeuCode lysine tags
- Unlabeled : select if no labeling was utilized (typical)
- Save Options: for full-run of a method .xml file (see below), select a file path to output results

2.3 Load Data Using Drop Down Menu

- Drop down menu: Change file type to be added by selecting an item in the drop down box. Each file type is described below.
- Add Files...: add files of the type selected in the drop down box
- Clear All: clear files of the type selected in the drop down box



2.4 Deconvolution Results for Identification

Results from deconvolution of MS1 spectra, *i.e.*, observed proteoform masses. These deconvolution results will be used to identify proteoforms by intact-mass analysis. There are three options for deconvolution results (see **Deconvolution** section):

- Results from Thermo Deconvolution 4.0 (.xlsx)
- Results from FLASHDeconv (.tsv)
- A three column tab-separated .tsv or .txt file with columns mass, intensity, retention time

There is the option to label the Biological Replicate, Fraction, Technical Replicate, and Condition for each file. To change one of these labels for a single file, click the appropriate cell in the table. To change the label for more than one file or cell, select the cells you would like to change the label for, right click your mouse, enter a label, click Okay.

2.5 Deconvolution Results for Quantification

It is only necessary to enter deconvolution results for quantification if you plan to perform a quantitative analysis of proteoform abundance changes between two conditions (see **Quantification** section). Results from deconvolution of MS1 spectra, *i.e.*, observed proteoform masses. The proteoform intensity values in these deconvolution results will be used to quantify proteoforms.^{2,5} These results files can be the same or different than the deconvolution results for identification. There are three options for deconvolution results (see **Deconvolution** section):

- Results from Thermo Deconvolution 4.0 (.xlsx)
- Results from FLASHDeconv (.tsv)

- A three column tab-separated .tsv or .txt file with columns mass, intensity, retention time

To perform a quantification analysis, it is necessary to label the Biological Replicate, Fraction, Technical Replicate, and Condition for each file. The change one of these labels for a single file, click the appropriate cell in the table. To change the label for more than one file or cell, select the cells you would like to change the label for, right click your mouse, enter a label, click Okay.

2.6 Protein Databases

Download a protein database from UniProt (<https://www.uniprot.org/proteomes/>). It is recommended to use the Reviewed entries only. A database from MetaMorpheus generated from a bottom-up search and the global post-translational modification discovery strategy (G-PTM-D) can also be utilized.^{3,6} Check the contaminant column if the database is a contaminant database. There are two options for protein databases:

- .xml or .xml.gz: contains annotated PTM information and subsequences
- .fasta: option to include isoforms when downloaded

2.7 Top-Down Hit Results

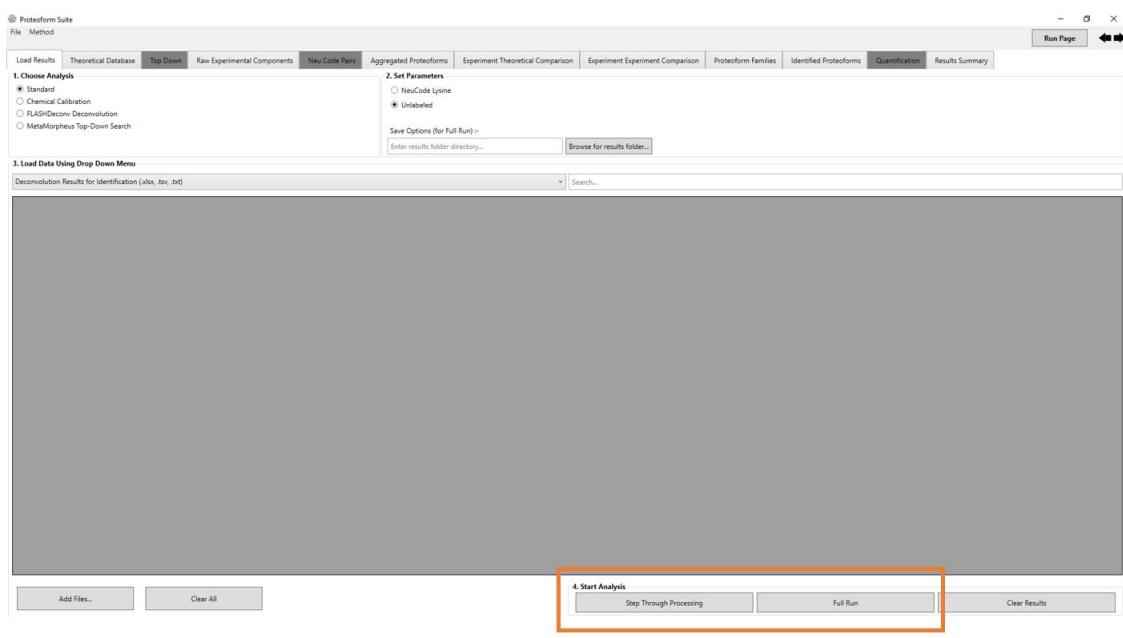
Results from a top-down search of MS/MS spectra, *i.e.*, proteoform identifications. There are two options for top-down results (see **Top-Down Search** section):

- Results from TDPortal (.xlsx)
- Results from MetaMorpheus (.psmtsv)

2.8 MetaMorpheus Bottom-Up Unique Peptides

Results from a bottom-up search of MS/MS spectra, *i.e.*, peptide identifications.⁸ Download a release of MetaMorpheus and run a bottom-up search (<https://github.com/smith-chem-wisc/MetaMorpheus/releases>). In the results folder of the search, load the AllPeptides.psmtsv file.

2.9 Start Analysis



- Step Through Processing button: instructions display on how to step through different pages
- Full Run button: load in a method .xml or use preset defaults to automatically perform a full run through analysis
- Clear Results button: clear all files in the file table

3 Deconvolution

Results from deconvolution of MS1 spectra are used to identify proteoforms by intact-mass analysis and to construct proteoform families of observed proteoforms. Deconvolution results are loaded on the Load Results page (see **Load Results: Standard**) under Deconvolution Results for Identification and Deconvolution Results for Quantification. This section describes how to obtain deconvolution results to import into Proteoform Suite.

3.1 Thermo Deconvolution 4.0

- See Thermo Fisher website for quote and user guide
- Run Xtract algorithm for high resolution data
- For each .raw file:
 - Select Open Results in the Run Queue
 - Right-click the Results table and choose Export All
 - Open the .xls file exported and save as a .xlsx file
- The saved .xlsx file is used for Deconvolution Results for Identification and Deconvolution Results for Quantification in the Standard analysis on the Load Results page

3.2 FLASHDeconv

FLASHDeconv is an ultra-fast deconvolution algorithm for high resolution mass spectrometry data developed by the OpenMS team.¹¹

- On the Load Results page, select FLASHDeconv Deconvolution under Choose Analysis (top left)
- Input .mzML files into the table (Load Data drop down menu will be set to Spectra Files)
- FLASHDeconv requires .mzML files. Use MSConvert to convert other file types (<http://proteowizard.sourceforge.net/>). Peak picking is NOT recommended.
- Set Parameters:
 - Min Charge: minimum charge state allowed for deconvolution
 - Max Charge: maximum charge state allowed for deconvolution
- To begin deconvolution, hit the Deconvolute button under Start Analysis (bottom right)
- For more advanced parameter options, you can also run the command line version of FLASHDeconv, available at <https://www.openms.de/comp/flashdeconv/>
- The resulting .tsv file is used for Deconvolution Results for Identification and Deconvolution Results for Quantification in the Standard analysis on the Load Results page

3.3 Other

- If you have another deconvolution algorithm of choice, simply create a three column tab-separated file

Monoisotopic mass \t intensity \t retention time

- This .tsv or .txt file can be used for Deconvolution Results for Identification and Deconvolution Results for Quantification in the Standard analysis on the Load Results page

4 Top-down Search

Results from a top-down MS/MS search can be used in Proteoform Suite. These top-down proteoform identifications improve intact-mass analysis and proteoform family construction. Additionally, top-down results can be integrated with bottom-up peptide results. Top-down results are loaded on the Load Results page (see **Load Results: Standard**) under Top-Down Hit Results. This section describes how to obtain top-down results to import into Proteoform Suite.

4.1 TDPortal

TDPortal is a high-throughput global proteome analysis software for top-down data^{12–14} available through the National Resource for Translational and Developmental Proteomics.

- Request for access to TDPortal: <http://nrtdp.northwestern.edu/tdportal-request/>
- Download TDViewer to access results: <http://topdownviewer.northwestern.edu/>
- Under Reports tab in TDViewer, export Hit Report
- The resulting .xlsx file is used for Top-Down Hit Results in the Standard analysis on the Load Results page

4.2 MetaMorpheus

MetaMorpheus is an MS/MS search software program for both bottom-up and top-down analysis.^{15,16}

- On the Load Results page, select MetaMorpheus Top-Down Search under Choose Analysis (top left)
- Load .raw or .mzML files in the table with the Load Data drop down menu set to Spectra Files
- Load an .xml or .fasta database in the table with the Load Data drop down menu set to Protein Databases
- Set Parameters:
 - Precursor Mass Tolerance: the difference in mass between the observed precursor and the theoretical proteoform
 - Product Mass Tolerance: the difference in mass between the product ions generated by fragmentation and the theoretical proteoform's theoretical fragmentation spectra
 - Fixed Carbamidomethyl Mod: fixed modifications are applied to EVERY amino acid in the database specified in the list. Check this box for protein samples that have been reduced and alkylated with iodoacetamide
 - Dissociation Type: the dissociation type used to fragment intact proteoforms and produce product ions for the tandem mass spectra
- To begin the search, hit the MetaMorpheus Top-Down Search button under Start Analysis (bottom right)

- For more advanced parameter options, you can also run the GUI or command line version of MetaMorpheus, available at <https://github.com/smith-chem-wisc/MetaMorpheus/releases>
- The resulting AllPSMs.psmtsv file is used for Top-Down Hit Results in the Standard analysis on the Load Results page

5 Calibration

Calibration is an optional pre-processing step to improve the mass accuracy of the deconvolution and the top-down search results.^{4,15} High-scoring top-down identifications are used as calibration points. Retention time across files can also be calibrated to correct run-to-run variation. Calibrated deconvolution results are loaded on the Load Results page (see **Load Results: Standard**) under Deconvolution Results for Identification and Deconvolution Results for Quantification, and calibrated top-down results are loaded on the Load Results page (see **Load Results: Standard**) under Top-Down Hit Results. This section describes how to perform calibration.

5.1 Overview

- On the Load Results page, select Chemical Calibration under Choose Analysis (top left)
- Load and label files (see below)

To change one of these labels for a single file, click the appropriate cell in the table. To change the label for more than one file or cell, select the cells you would like to change the label for, right click your mouse, enter a label, click Okay.

- Set parameters (see below)
- To begin calibration, hit the Calibrate button under Start Analysis (bottom right)

5.2 Load Files

- Set the Load Data drop down menu to Spectra Files
 - Add all .raw or .mzML files using the Add Files button or with drag-and-drop
 - Any raw files deconvoluted to generate Deconvolution Results AND searched to generate Top-Down Hit results must be added
- Set the Load Data drop down menu to Uncalibrated Deconvolution Results
- Add all deconvolution results files using the Add Files button or with drag-and-drop
- Set the Load Data drop down menu to Uncalibrated Top-Down Hit Results
- Add all top-down results files using the Add Files button or with drag-and-drop
- Label the Biological Replicate, Fraction, Technical Replicate, and Condition for each file.
 - Each spectra file and deconvolution result file must have a different label
 - Each spectra file label should exactly match the corresponding deconvolution result label
 - Calibration is performed across technical replicates for the same biological replicate, fraction, and condition. Therefore, if you wish to have a top-down file calibrate an intact-mass file, it is necessary to have the biological replicate, fraction, and condition match while the technical replicate label varies (ex: 1, 2, 3, etc.)

5.3 Set Parameters

- NeuCode Lysine: select if cell culture was performed with heavy and light NeuCode lysine tags
- Unlabeled : select if no labeling was utilized (typical)
- Write Calibrate Raw Files: if checked, calibrated .mzML files will be exported in the same file location as the original spectra files
- Calibrate Top-Down Files: if checked, a calibrated top-down results file will be exported in the same file location as the original top-down results file
- Calibrate Masses: if checked, calibration will be performed on proteoform masses
- Mass Tol. (ppm): mass tolerance used for mass calibration if Calibrate Masses is checked
- Calibrate Retention Times: if checked, calibration will be performed on proteoform retention times
- RT Tol. (min): retention time tolerance used for retention time calibration if Calibrate Retention Times is checked

6 Theoretical Database

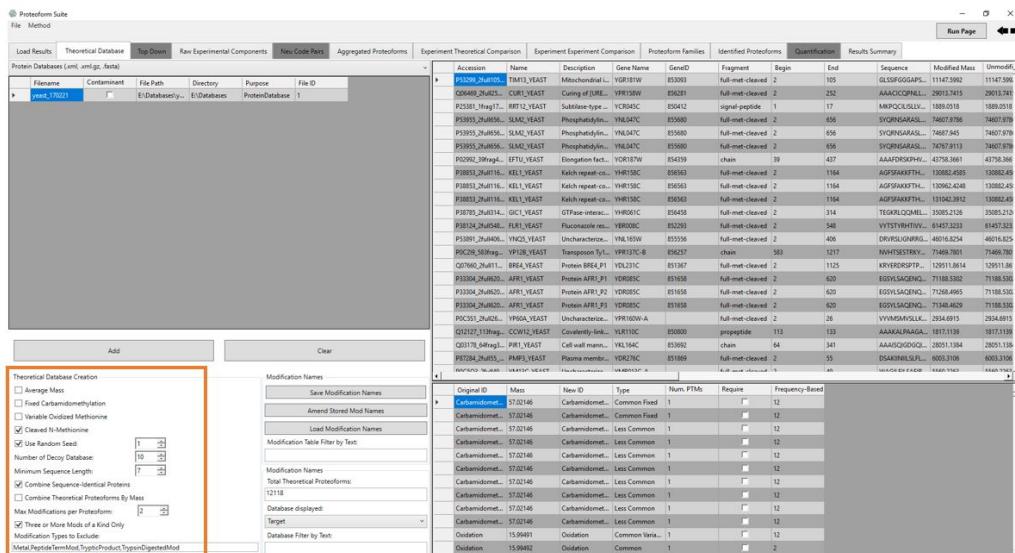
6.1 Overview

On this page, theoretical proteoforms are created using the file(s) loaded under Protein Databases on the Load Results page. This theoretical proteoform database includes theoretical proteoforms with combinations of annotated post-translational modifications (PTMs) and subsequences. Theoretical proteoforms are used in intact-mass analysis to identify proteoforms (see **Experiment-Theoretical Comparison** section). Decoy databases are also generated to compute a false discovery rate for intact-mass proteoform identifications. The modifications table (bottom right) enables a user to edit modifications. The bottom-up result file(s) loaded under MetaMorpheus Bottom-Up Unique Peptides on the Load Results Page is used to create a list of bottom-up peptides, which are integrated with theoretical proteoforms.

6.2 Run Page

- Load database file(s) on the Load Results page under Protein Databases (see **Load Results** section)
 - Set all parameters as desired for current analysis (see below)
 - Click Run Page button (top right)

6.3 Set Parameters



- Average Mass: if checked, the average mass of each theoretical proteoform will be used instead of the default monoisotopic mass
 - Fixed Carbamidomethylation: if checked, each cysteine on each theoretical proteoform will have a carbamidomethylation group added

- Variable Oxidized Methionine: if checked, theoretical proteoforms will be added with oxidation modifications up to the number of methionine residues/the Max Modifications per Proteoform parameter setting
- Cleaved N-Methionine: if checked, methionine will be cleaved off of each full sequence (sub-sequences UniProt containing the N-methionine will still be added)
- Use Random Seed: a random seed will be used in the random number generator creating decoy databases, resulting in the same decoy database each time (with the same given parameters)
- Random Seed: this number will be used for the random seed if the Use Random Seed box is checked
- Number of Decoy Databases: the number of decoy databases generated. The average across each decoy database result is used to compute the false discovery rate for intact-mass identifications
- Minimum Sequence Length: the minimum length of a theoretical proteoform in the database
- Combine Sequence-Identical Proteins: if checked, sequences that are the same from different genes/proteins will be combined into a single theoretical proteoform entry
- Combine Theoretical Proteoforms by Mass: if checked, different proteoforms with the same mass (up to 4 decimal places) will be combined into a single theoretical proteoform entry
- Max Modifications per Proteoform: theoretical proteoforms will be generated for each UniProt protein entry containing annotated modifications in different combinations up to this number
- Three or More Mods of a Kind Only: if checked, modification combinations with different modifications will only go up to 2 modifications per theoretical proteoform. If Max Modifications per Proteoform is set to greater than 2, only modification combinations containing annotated modifications of the same type (ex: triphospho) will go from three modifications up to this number
- Modification Types to Exclude: comma-separated list of types of modifications to exclude for annotated modifications in the database

- Modification Names table: the bottom right table displays each input modification. Several columns can be edited in this table:

The screenshot shows the Proteoform Suite interface with several tabs at the top: Load Results, Theoretical Database, Tag Library, Raw Experimental Components, New Code Panel, Aggregated Proteoforms, Experiments Theoretical Comparison, Experiment Experiment Comparison, Proteoform Families, Identified Proteoforms, Quantification, Results Summary, and Run Page.

The main area contains two tables. The top table, "Proteoform Families", lists various modifications with columns for Accession, Name, Description, Gene Name, Gerofit, Frequency, Begin, End, Sequence, Modified Mass, and Unmodified. The bottom table, "Modification Names", lists modifications with columns for Original ID, Name ID, Type, Num. PTMs, Requires, and Frequency-Based. A red box highlights the "Modification Names" table.

On the left, there is a sidebar for "Theoretical Database Creation" with checkboxes for "Create Mod", "Find Carbamidomethylation", "Denature Oxidized Methionine", "Use Random Seed" (set to 1), "Number of Decay Database" (set to 10), "Minimum Sequence Length" (set to 7), "Combine Sequence Identical Proteins", and "Combine Theoretical Proteoforms By Mass". It also includes fields for "Max Modifications per Proteoform" (set to 2) and "Modification Types to Exclude" (set to "Metal/PeptideTermMod, TrypticProduct, TrypsinDigestedMod").

- New ID: the unlocalized version of the modification name, used in Proteoform Suite intact-mass analysis
- Num. PTMs Represented: the number of PTMs represented by this PTM entry (ex: Dimethylation - 2 PTMs)
- Require Proteoform Without This Modification: if checked, a proteoform is only identified containing this modification if a proteoform without this modification is also identified. Potentially useful for modifications such as adducts
- Frequency-Based Rank of PTM Mass: this rank is used in intact-mass analysis to favor more common modifications (lower number is prioritized).
- Save Modification Names: saves a new .modnames file with edits in Modification Names table
- Amend Stored Mod Names: saves and overwrites .modnames file used in Proteoform Suite
- Load Modification Names: load a new .modnames file for use in current analysis
- Modification Table Filter by Text: filter the Modification Names table (bottom right) by any entered text

6.4 Results

- Theoretical Proteoforms table: the top right table displays all theoretical proteoforms generated
 - Accession: accession given by Proteoform Suite for this specific theoretical proteoform
 - Name: protein name from UniProt

- Description: protein description from UniProt
- Gene Name: gene name from UniProt
- GeneID: gene ID from UniProt
- Fragment: sequence description
- Begin: theoretical proteoform begin residue in protein full sequence in UniProt
- End: theoretical proteoform end residue in protein full sequence in UniProt
- Sequence: theoretical proteoform sequence
- Modified Mass: monoisotopic (or average) mass of theoretical proteoform with any modifications
- Unmodified Mass: monoisotopic (or average) mass of theoretical proteoform without modifications
- PTM Mass: mass of PTMs on theoretical proteoform
- Contaminant: checked if theoretical proteoform is from contaminant database
- Lysine Count: number of lysines in theoretical proteoform sequence
- PTM Description: PTMs on theoretical proteoform
- GO Term IDs: Gene Ontology term IDs from UniProt
- Grouped Accessions: accession grouped if Combine Sequence-Identical Proteins and/or Combine Theoretical Proteoforms By Mass are checked
- Top-Down Theoretical: theoretical proteoform identified by top-down analysis (must run Top-Down page first)
- Not in Original Database: theoretical proteoform added because of top-down identification, not in original database pre-top-down-analysis (must run Top-Down page first)
- Bottom-Up PSMs Count: number of bottom-up PSMs derived from this theoretical proteoform

- Peptide-Specific Modified Bottom-Up PSMs: modified residues confirmed ID'd by bottom-up peptides derived from this theoretical proteoform, keeping unique peptidoforms separate
- Modified Bottom-Up PSMs: modified residues confirmed by ID'd bottom-up peptides derived from this theoretical proteoform
- Bottom-Up Evidence for All PTMs: checked if all PTMs on this theoretical proteoform are confirmed by at least one modified bottom-up peptide
- Bottom-Up Evidence for Begin: checked if bottom-up peptide identified with begin residue at this theoretical proteoform's begin residue
- Bottom-Up Evidence for End: checked if bottom-up peptide identified with end residue at this theoretical proteoform's end residue
- Total Theoretical Proteoforms: number of theoretical proteoforms generated
- Database displayed: which database is displayed (target or decoy database)
- Database filter by Text: filter the Theoretical Proteoforms table (top right) by any entered text

The screenshot shows the ProteinScape software interface with the 'Theoretical Proteoforms' table highlighted. The table lists various peptides with their Accession numbers, Names, Descriptions, Gene Names, and various IDs. Columns include Accession, Name, Description, Gene Name, GeotID, Fragment, Begin, End, Sequence, Modified Mass, Unmodified Mass, and Frequency. A 'Quaternification' tab is selected, showing a summary of the number of peptides found in different states (Full-met, half-met, etc.). Below the table, there are several filter and search panels:

- Theoretical Database Creation:** Includes checkboxes for Average Mass, Combine Theoretical Proteoforms By Mass, Max Modifications per Proteoform, Use More Mods of a Kind Only, and Modification Types to Exclude.
- Modification Names:** Panels for Save Modification Names, Amend Stored Mod Names, Load Modification Names, and Modification Table Filter by Text.
- Total Theoretical Proteoforms:** Displays the count (12118) and a list of modifications used.
- Database displayed:** Set to 'Target'.
- Database Filter by Text:** An input field for filtering the proteoforms table.

7 Top-Down

7.1 Overview

On this page, the top-down hit results are read in from the file(s) loaded under Top-Down Hit Results on the Load Results page. A top-down hit is a proteoform spectral match. Top-down hits are then aggregated into top-down proteoforms by identification and retention time. The theoretical database is supplemented with top-down proteoform identifications not already present in the database. Bottom-up peptide results are integrated with the top-down proteoform results.

7.2 Run Page

- Load top-down results file(s) on the Load Results page under Top-Down Hit Results (see **Load Results** section)
- Set all parameters as desired for current analysis (see below)
- Click Run Page button (top right)

7.3 Set Parameters

The screenshot shows the 'Run Page' configuration window in the Proteoform Suite. The window has several tabs at the top: Load Results, Theoretical Database, Top Down, Raw Experimental Components, New Code Pairs, Aggregated Proteoforms, Experiment Theoretical Comparison, Experiment Experiment Comparison, Proteoform Families, Identified Proteoforms, Quantification, and Results Summary. The 'Top Down' tab is selected. In the 'Top Down' tab, there are input fields for 'Min. C-Score' (set to 3.0), 'Ret. Time Tolerance (min)' (set to 5.00), and checkboxes for 'TightAbsoluteMass', 'Biomarker', and 'Show Proteoform-Specific Peptides'. Below these are two tables: 'Top Down Hits' and 'Top Down Proteoforms'. The 'Top Down Hits' table lists 6092 unique PPRs with IDs Q9Y256_2_30_S... through Q9Y256_21_30_S... and retention times ranging from 40.25 to 26.67 minutes. The 'Top Down Proteoforms' table lists 1681 unique PPs with IDs Q9Y256_31_39_S... through Q9Y256_21_30_S... and retention times ranging from 124.5232 to 120.3231 minutes. At the bottom of the window, there is a decorative footer with the text 'EMDEEDKAFK'.

- Min. C-Score: the minimum C-score¹⁷ for TDPortal results required for a top-down hit to be included. C-scores of 3 and higher correspond to identified proteoforms and C-scores of 40 and higher correspond to well-characterized proteoforms
- Ret. Time Tolerance (min): retention time tolerance used for aggregated top-down hits of the same proteoform identifications. Top-down hits of the same ID that elute outside of this tolerance will be aggregated into separate top-down proteoforms

- Tight Absolute Mass: if checked, TDPortal hits from the Tight Absolute Mass search will be included
- Biomarker: if checked, TDPortal hits from the Biomarker search will be included

7.4 Results

- Top-Down Hits: total number of top-down hits (proteoform spectral matches)
- Unique PFRs: unique proteoform identifications
- Top-Down Proteoforms: number of aggregated top-down proteoforms. May be greater than the number of unique PFRs if some hits of the same ID fall outside of the retention time tolerance
- Table Filter: filter the Top-Down Proteoforms table (left) by any entered text

Proteoform Suite
File Method

Load Results Theoretical Database Top Down Raw Experimental Components Neu Code Pairs Aggregated Proteoforms Experiment Theoretical Comparison Experiment Experiment Comparison Proteoform Families Identified Proteoforms Quantification Results Summary

Min. C-Score: 3.0 Ret. Time Tolerance (min): 5.00

TightAbsolute Mass Biomarker Show Proteoform-Specific Peptides

Top-Down Hits: 6892 Unique PFRs: 1681 Top-Down Proteoforms: 1621

Table Filter:

PFR Accession	Input Filename	Scan	Reported Mass	Theoretical	Retention Time	Uniprot ID	Sequence
Q9Y258_1_36...	20101210_Velos...	5760	1240.5275	1240.5282	40.25	Q9Y258	EMDEEDKAFK
Q9Y258_51_59...	20100616_Velos...	5281	812.4755	812.4756	26.30	Q9Y258	GPLATGKIK
Q9Y258_1_30...	20101210_Velos...	3301	1256.5233	1256.5231	26.67	Q9Y258	EMDEEDKAFK

E M[D[E E[D[K[A]F[K

Accession Modified Mass Retention Time Observations PFR Accession Original Description Gene Name

Q9Y258_2_36... 6972.9015 43.19 13 Q9Y258_2_64... Common Biol. Translation mat. TM47

Q9577_2_70... 10307.2893 72.42 7 Q9577_2_36... Common Biol. Un srRNA-ssco... LSM9

P6212_2_60..._1... 9032.7333 62.95 8 P6212_2_80..._1... Common Biol. Un srRNA-ssco... LSM9

P05114_2_100... 10521.5334 41.22 33 P05114_2_100... Common Biol. Non-histone ch. HMDN1

Q7964_2_70..._10... 11132.2278 69.52 14 Q7964_2_10... Common Biol. ATP synthase - ATP5L

Q7280_1_70..._4... 8619.7000 78.67 11 Q7280_1_74... MRDPAKAVAYV... Small integral - SMM15

P0304_2_60..._2... 9256.0591 37.33 138 P0304_2_90..._P... PRKRAEGAO4... Non-histone ch. HMDN2

Q9958_2_70... 11375.0948 68.07 23 Q9958_2_90... Common Biol. Protein S100-A13 S100A13

P6220_2_70..._1... 8401.4233 64.72 8 P6220_2_76..._5... Common Biol. Small nuclear n... SNRNP

P5681_2_70..._1... 5645.0700 51.04 9 P5681_2_51... VAVYRQLQGLSV... ATP synthase - ATP5L

P6223_2_70..._2... 4960.4938 44.76 6 P6223_2_44..._5... Common Biol. Thymosin beta TMSB4X / TM

P6108_2_70..._15... 17703.0084 66.39 6 P6108_2_15... Common Biol. Ubiquitin-cong UBE2N

Q15847_2_70..._7... 7760.9693 64.65 1 Q15847_1_76... Common Biol. Adipogenesis r... ADIRF

P6317_2_70..._7... 8081.7000 47.78 8 P6317_2_70... PRKIEKEKFLPY... 60S ribosomal RPL38

P6311_2_70..._1... 4933.5199 45.07 1 P6311_2_44..._1... Common Biol. Thymosin beta TMSB10

P05204_2_70..._3... 9255.9900 21.40 46 P05204_2_90..._P... PRKRAEGAO4... Non-histone ch. HMDN2

Q15841_1_76..._7... 8554.6783 53.39 6 Q15841_1_76..._M... MLLKVKTITKEL... NEDD8 NEDD8

P31949_2_70..._2... 11543.8190 67.54 27 P31949_2_10..._1... Uniprot P-ace... Protein S100-A... S100A11 | S19

Q9P059_1_70..._2... 11599.0212 76.90 5 Q9P059_1_112... Common Biol. Transmembrane TMEM14C

Q7550_2_70..._6... 8449.4881 74.13 1 Q7550_2_76... Common Biol. Heat shock fact. HSP90

Q3ZAC7_1_70..._10... 11388.8921 87.90 3 Q3ZAC7_1_10... Common Biol. Vacuolar ATPase VMA21

P24311_2_70..._2... 6358.2054 64.08 2 P24311_2_80..._P... SHQKRTGFHD... Cytochrome c COX7B

P05114_2_70..._1... 89014.4974 41.22 4 P05114_2_100..._P... PRKRAEGAO4... Non-histone ch. HMDN1

- Top-down Proteoforms table: the left table displays all top-down proteoforms. For MetaMorpheus results, ambiguous identifications in a single top-down proteoform have information separated by a "—".

Proteoform Suite
File Method

Load Results Theoretical Database Top Down Raw Experimental Components New Code Prefs Aggregated Proteoforms Experiment Theoretical Comparison Experiment Experiment Comparison Proteoform Families Identified Proteoforms Quantification Results Summary

Min. C-Score: 3.0 Top-Down Hits: 6892
Ret. Time Tolerance (min): 5.00 Unique PFRs: 1681
 TightAbsolute Mass Top-Down Proteoforms: 1621
 Biomerator
 Show Downstream Cross-Database Table Filter

Accession	Modified Mass	Retention Time	Observations	PFR Accession	Original Description	Gene Name
Q9Y258_260L...	6972.9015	43.19	13	Q9Y258_2_34_S...	[Common Biol., Translation m...]	TMAT
P05777_260L...	1030.2893	72.42	7	Q95777_2_36_T...	[Common Biol., US snRNA-ass...]	L3MB
P62112_260L...	9032.7533	62.95	8	P62112_2_36_S...	[Common Biol., US snRNA-ass...]	L5MB
P05114_260L...	10521.5394	41.22	33	P05114_2_100_P...	[PDRXYSAGA...]	Non-histone ch., HMGN1
Q7H644_260L...	11382.3278	69.52	14	Q7H644_2_103...	[Common Biol., ATP synthase ...]	ATP5B
Q7ZB80_260L...	8619.7931	78.67	11	Q7ZB80_2_74...	[MDKIAWAEV...]	Small integral ...
P05104_260L...	9256.0051	37.33	138	P05104_2_90_P...	[PKRALEGAG...	HMGN15
Q99584_260L...	11370.8448	68.07	23	Q99584_2_38_A...	[Common Biol., Protein S100-A11]	S100A13
P62108_260L...	8401.4233	64.72	8	P62108_2_76_S...	[Common Biol., Small nuclear r...]	SNRPG
P56181_260L...	5645.0740	51.04	9	P56181_2_31_V...	[VAVWPAQASLY...]	ATP synthase ...
P62128_260L...	4960.4938	44.76	6	P62128_2_44_S...	[Common Biol., Thymosin beta...]	TM584X/TM
P06102_260L...	17038.0084	66.39	6	P06102_2_152...	[Common Biol., Ubiquitin-con...]	UBE2N
C15447_260L...	7760.9603	64.65	1	C15447_2_76_A...	[Common Biol., Adipogenesis r...]	ADIPF
P61173_260L...	8081.7048	47.78	8	P61173_2_76_P...	[PRKEEKFPLT...]	65S ribosomal ...
P63111_260L...	4931.5169	45.07	1	P63111_2_44_A...	[Common Biol., Thymosin beta...]	TM5810
P05104_260L...	9255.9995	21.40	46	P05104_2_90_P...	[PKRALEGAG...	Non-histone ch., HMGN42
C151843_167L...	8554.6763	53.39	6	C151843_1_76_M...	[MKVYVLTIGKL...]	NEDD8
P31196_260L...	11643.8199	67.54	27	P31196_2_105...	[Uniprot&Prot...]	Protein S100-A...
Q9P599_161L...	11590.0121	76.90	5	Q9P599_1_112...	[Common Biol., Transmembrane...]	TMEN14C
P75596_260L...	8448.1481	74.13	1	P75596_2_32_A...	[Common Biol., Heat shock fact...]	HSPB1
Q7ZAO7_161L...	11380.8921	87.90	3	Q7ZAO7_1_101...	[Common Biol., Vacuole ATPas...]	VMA21
P24311_260L...	6358.2054	64.08	2	P24311_2_50_S...	[SHQKTFPFD...]	Cytochrome c ...
D01114_260L...	10691.4874	41.59	9	D01114_2_100_D...	[DPRKVVGEG...	Non-histone ch.

E M D E E D K A F K

■ G R E G G K J K J K P L L K Q P P K J K Q A J K E J M D E E D K A F K Q K Q K E E E Q K K L E E L K A K A A G K G P F L A T G G I J K F S G K K

- Accession: accession given by Proteoform Suite for this specific top-down proteoform
- Modified Mass: monoisotopic mass of top-down proteoform (average of aggregated top-down hits)
- Retention Time: retention time of top-down proteoform (average of aggregated top-down hits)
- Observations: number of aggregated top-down hits in this top-down proteoform
- PFR Accession: protein accession_begin residue_end residue_full sequence with PTMs
- Original PFR Accession/full-sequence: PFR accession in TDPortal, full sequence reported in MetaMorpheus
- Description: protein description from UniProt
- Gene Name: gene name from UniProt
- UniProt ID: protein UniProt ID
- Accessions: UniProt protein accession
- PTM Description: PTMs on top-down proteoform
- Begin and End: top-down proteoform begin and end residue in protein full sequence in UniProt
- Sequence: top-down proteoform sequence
- UniProt-Annotated Modifications: all UniProt annotated residues for PTMs on this top-down proteoform in UniProt database provided
- Potentially Novel Mods: checked if this top-down proteoform contains PTMs not annotated in UniProt database provided

- Best Hit Score: best C-score (TDPortal) or MetaMorpheus score (MetaMorpheus) out of aggregated hits for this top-down proteoform
 - Best Hit Delta Score: best delta score (score difference between next best scoring identification) out of aggregated hits for this top-down proteoform (MetaMorpheus results only)
 - Best Hit Q-Value: best q-value out of aggregated hits for this top-down proteoform
 - Level: proteoform identification level based on five-level scheme¹⁸
 - Level Description: description of proteoform level assignment (sources of ambiguity)
 - Mass Error: difference in mass between experimental and theoretical proteoform monoiso-topic mass
 - Best Hit Info: filename for highest scoring hit aggregated into this top-down proteoform
 - Family: proteoform family number (must have run through full Proteoform Suite analysis)
 - Linked Proteoform References: proteoforms in family network path of identification to the nearest theoretical proteoform (must have run through full Proteoform Suite analysis)
 - Bottom-Up PSMs Count: number of bottom-up PSMs derived from this top-down proteoform
 - Different Ambiguity in Bottom-Up PSMs: checked for ambiguous top-down identifications where the different IDs have a different number of bottom-up PSMs
 - Modified Bottom-Up PSMs: modified residues confirmed by ID'd bottom-up peptides derived from this top-down proteoform
 - All Modified Bottom-Up PSMs from Protein: modified residues confirmed by ID'd bottom-up peptides derived from this top-down proteoform's protein
 - Bottom-Up PSMs Separate Peptides: modified residues confirmed ID'd by bottom-up peptides derived from this top-down proteoform, keeping unique peptidoforms separate
 - Bottom-Up Evidence for Begin: checked if bottom-up peptide identified with begin residue at this top-down proteoform's begin residue
 - Bottom-Up Evidence for End: checked if bottom-up peptide identified with end residue at this top-down proteoform's end residue
 - Bottom-Up Evidence for All PTMs: checked if all PTMs on this top-down proteoform are confirmed by at least one modified bottom-up peptide
 - Sequence Specific: description of difference in PTMs between bottom-up peptides from this top-down proteoform sequence and this top-down proteoform's PTMs
 - All Peptides from Protein: description of difference in PTMs between bottom-up peptides from this top-down proteoform's protein and this top-down proteoform's PTMs
 - Fragments: if MetaMorpheus results, MS/MS fragments identified
- Top-down Proteoform Sequence box: bottom box on page where sequence is displayed with PTMs for the top-down proteoform selected in the Top-Down Proteoforms table (left). If MetaMorpheus results, sequence is annotated based on identified MS/MS fragments

Proteoform Suite

File Method

Load Results Theoretical Database Top Down Raw Experimental Components Neu Code Pairs Aggregated Proteoforms Experiment Theoretical Comparison Experiment Experiment Comparison Proteoform Families Identified Proteoforms Quantification Results Summary

Min. C-Score
Ret. Time Tolerance (min)
TightAbsolute Mass
Biomarker

Show Proteoform-Specific Peptides

Accession	Modified Mass	Retention Time	Observations	PFR Accession	Original	Description	Gene Name	
Q9Y506_2n64...	6972.9015	43.19	13	Q9Y506_2A_50...	[Common Biol., Translation m...]	TMA7		
Q95777_2n65...	10307.2893	72.42	7	Q95777_2A_50...	[Common Biol., Urs s...]	LSPM		
P62312_2n610...	9032.7533	62.95	8	P62312_2A_50...	[Common Biol., Urs s...]	LSPM		
P05114_2n610...	10521.5334	41.22	33	P05114_2A_100...	[Common Biol., PRKXSSAGA...	HMGN1		
Q75964_2n610...	11322.2378	69.52	14	Q75964_2_100...	[Common Biol., ATP synthase ...]	ATPS1		
Q72380_2n610...	8619.7931	78.67	11	Q72380_1_70...	[Common Biol., MIFOKAWLEV...	SMIM15		
P05104_2n610...	9256.0051	37.33	138	P05104_2A_50...	[Common Biol., PRKXAGAG...	Non-histone ch...	HMGN2	
Q95584_2n610...	11375.0548	68.07	23	Q95584_2A_50...	[Common Biol., PRKXAGAG...	Non-histone ch...	HMGN2	
P62308_2n701...	8401.4238	64.72	8	P62308_2_75...	[Common Biol., Small nuclear ...]	SNRPG		
P56031_2n611...	5645.0740	51.04	9	P56031_3_55...	[Common Biol., VAVVHQDQGLSY...	ATP synthase ...]	ATPS2	
P62328_2n612...	4960.4993	44.76	6	P62328_2_44...	[Common Biol., Thymosin beta...	TMSK1		
P01688_2n612...	17038.0084	66.39	6	P01688_2_15...	[Common Biol., Ubiquitin-con...	UBZN1		
Q15457_2n612...	7786.9663	64.65	1	Q15457_2_70...	[Common Biol., Adipogenesis r...	ADIRF		
P61733_2n612...	8811.7046	47.78	8	P61733_2_70...	[Common Biol., PRKXENDFLT...	65S ribosomal ...]	RPL28	
P60313_2n604...	4933.5168	45.07	1	P60313_2_44...	[Common Biol., Thymosin beta...	TMSB10		
P06034_2n605...	9253.9993	21.40	46	P06034_2_50...	[Common Biol., PRKXAGAG...	Non-histone ch...	HMGN2	
Q15451_2n616...	8554.6763	53.59	6	Q15451_1_76...	[Common Biol., MIFOKAWLEV...	NEDD8		
P31949_2n617...	11643.8190	67.54	27	P31949_2_100...	[UniprotKB-exact, Protein S100A11 S10...			
Q9W595_2n617...	11599.0721	76.90	5	Q9W595_1_112...	[Common Biol., Transmembrane...	TMMB4C		
Q75964_2n617...	7766.9663	74.13	1	Q75964_2_70...	[Common Biol., Heat shock fact...	HSPB1		
P61733_2n617...	11332.2378	69.52	11	P61733_2_70...	[Common Biol., Vacuolar ATPas...	VMA21		
P24311_2n617...	6336.2054	64.08	2	P24311_2_50...	[Common Biol., SHQKTPFH...	Cytochrome c	CYTB	
B0H114_2n618...	10601.4874	41.33	6	B0H114_2_100...	[UniprotKB-exact, Non-histone ch...	NEDD8		

Top-Down Hits: 6892
Unique PFRs: 1681
Top-Down Proteoforms: 1821

Table Filter

E M D E E D K A F K

■ G R E J G G K J K K P L J K Q P K K Q A K J E M D E D K A F K Q K Q E E Q K K L E E L J K A K J A A G K J G P L A T T G G I J K K S G K K

- Show Proteoform-Specific Peptides: if checked, only peptides specific to the selected top-down proteoform in the Top-Down Proteoform table (left) are displayed in the Bottom-Up Peptide table (right). If unchecked, all bottom-up peptides from the top-down proteoform's protein are displayed.
- Bottom-Up Peptides table: the right table displays bottom-up peptides from the top-down proteoform selected in the Top-Down Proteoforms table (left) if Show Proteoform-Specific Peptides is checked. The right table displays all bottom-up peptides from the top-down proteoform's protein selected in the Top-Down Proteoforms table (left) if Show Proteoform-Specific Peptides is unchecked.

Proteoform Suite

File Method

Load Results Theoretical Database Top Down Raw Experimental Components Neu Code Pairs Aggregated Proteoforms Experiment Theoretical Comparison Experiment Experiment Comparison Proteoform Families Identified Proteoforms Quantification Results Summary

Min. C-Score
Ret. Time Tolerance (min)
TightAbsolute Mass
Biomarker

Show Proteoform-Specific Peptides

Accession	Modified Mass	Retention Time	Observations	PFR Accession	Original	Description	Gene Name	
Q9Y506_2n64...	6972.9015	43.19	13	Q9Y506_2A_50...	[Common Biol., Translation m...]	TMA7		
Q95777_2n65...	10307.2893	72.42	7	Q95777_2A_50...	[Common Biol., Urs s...]	LSPM		
P62312_2n610...	9032.7533	62.95	8	P62312_2A_50...	[Common Biol., Urs s...]	LSPM		
P05114_2n610...	10521.5334	41.22	33	P05114_2A_100...	[Common Biol., PRKXSSAGA...	HMGN1		
Q75964_2n610...	11322.2378	69.52	14	Q75964_2_100...	[Common Biol., ATP synthase ...]	ATPS1		
Q72380_2n610...	8619.7931	78.67	11	Q72380_1_70...	[Common Biol., MIFOKAWLEV...	SMIM15		
P05104_2n610...	9256.0051	37.33	138	P05104_2A_50...	[Common Biol., PRKXAGAG...	Non-histone ch...	HMGN2	
Q95584_2n610...	11375.0548	68.07	23	Q95584_2A_50...	[Common Biol., PRKXAGAG...	Non-histone ch...	HMGN2	
P62308_2n701...	8401.4233	64.72	8	P62308_2_75...	[Common Biol., Small nuclear ...]	SNRPG		
P56031_2n611...	5645.0740	51.04	9	P56031_3_55...	[Common Biol., VAVVHQDQGLSY...	ATP synthase ...]	ATPS2	
P62328_2n612...	4960.4993	44.76	6	P62328_2_45...	[Common Biol., Thymosin beta...	TMSK1		
P01688_2n612...	17038.0084	66.39	6	P01688_2_15...	[Common Biol., Ubiquitin-con...	UBZN1		
Q15457_2n612...	7786.9663	64.65	1	Q15457_2_70...	[Common Biol., Adipogenesis r...	ADIRF		
P61733_2n612...	8811.7046	47.78	8	P61733_2_70...	[Common Biol., PRKXENDFLT...	65S ribosomal ...]	RPL28	
P60313_2n604...	4933.5168	45.07	1	P60313_2_44...	[Common Biol., Thymosin beta...	TMSB10		
P06034_2n605...	9253.9993	21.40	46	P06034_2_50...	[Common Biol., PRKXAGAG...	Non-histone ch...	HMGN2	
Q15451_2n616...	8554.6763	53.59	6	Q15451_1_76...	[Common Biol., MIFOKAWLEV...	NEDD8		
P31949_2n617...	11643.8190	67.54	27	P31949_2_100...	[UniprotKB-exact, Protein S100A11 S10...			
Q9W595_2n617...	11599.0721	76.90	5	Q9W595_1_112...	[Common Biol., Transmembrane...	TMMB4C		
Q75964_2n617...	7766.9663	74.13	1	Q75964_2_70...	[Common Biol., Heat shock fact...	HSPB1		
P61733_2n617...	11332.2378	69.52	11	P61733_2_70...	[Common Biol., Vacuolar ATPas...	VMA21		
P24311_2n617...	6336.2054	64.08	2	P24311_2_50...	[Common Biol., SHQKTPFH...	Cytochrome c	CYTB	
B0H114_2n618...	10601.4874	41.33	6	B0H114_2_100...	[UniprotKB-exact, Non-histone ch...	NEDD8		

Top-Down Hits: 6892
Unique PFRs: 1681
Top-Down Proteoforms: 1821

Table Filter

E M D E E D K A F K

■ G R E J G G K J K K P L J K Q P K K Q A K J E M D E D K A F K Q K Q E E Q K K L E E L J K A K J A A G K J G P L A T T G G I J K K S G K K

- PFR Accession: protein accession_begin residue_end residue_full sequence with PTMs
 - Input Filename: filename for this bottom-up peptide identification
 - Scan: scan number for this bottom-up peptide identification
 - Reported Mass: observed monoisotopic mass for this bottom-up peptide identification
 - Theoretical Mass: theoretical mass for this bottom-up peptide identification
 - Retention Time: retention time for this bottom-up peptide identification
 - UniProt ID: protein UniProt ID
 - Sequence: peptide sequence
 - Begin: peptide begin residue in protein full sequence in UniProt
 - End: peptide end residue in protein full sequence in UniProt
 - PTM Description: PTMs on peptide
 - Accession: UniProt protein accession
 - Name: name of protein from UniProt
 - Q-value: PEP q-value for this bottom-up peptide
 - Score: MetaMorpheus score for this peptide
 - Shared: checked if this peptide is a shared peptide between multiple proteins
- Bottom-Up Sequence box: middle right box on page where sequence is displayed with PTMs for the peptide selected in the Bottom-Up Peptides table (right). If MetaMorpheus results, sequence is annotated based on identified MS/MS fragments
-
- The screenshot shows the Proteiform Suite software interface. On the left, there is a table titled "Bottom-Up Peptides" with columns: Accession, Modified Mass, Retention Time, Observations, PFR Accession, Original, Description, Gene Name, and PTM Description. The table contains several rows of peptide data. On the right, there is a larger window titled "Sequence" which displays the peptide sequence "R[E][G][K][P][L][K][Q][A][K][E][M][D][E][E][D][K][A][F][K][Q][K][Q][E][E][Q][K][J][L][E][E][L][K][A][K][A][G][K][G][P][I][A][T][G][G][I][K][K][S][G][K][K]" with various PTMs highlighted in red. The software has a standard Windows-style interface with tabs at the top and a "Run Page" button.

8 Raw Experimental Components

8.1 Overview

On this page, raw experimental components are read in from the file(s) loaded under Deconvolution Results for Identification and Deconvolution Results for Quantification on the Load Results page. A raw experimental component is an individual proteoform observation as reported in the deconvolution results file(s).

8.2 Run Page

- Load the deconvolution results file(s) on the Load Results page under Deconvolution Results for Identification. If desired, label the biological replicate, fraction, technical replicate, and condition for each file (see **Load Results** section)
- If performing a quantitative analysis, load the deconvolution results file(s) on the Load Results page under Deconvolution Results for Quantification. Label the biological replicate, fraction, technical replicate, and condition for each file (see **Load Results** section)
- Set all parameters as desired for current analysis (see below)
- Click Run Page button (top right)

8.3 Set Parameters

- Mass tolerance for Merging Artifacts (ppm): mass tolerance used to merge deconvolution artifacts including missed monoisotopic mass differences and charge state harmonics
- Cosine threshold between per-charge intensities and fitted gaussian distribution: minimum value as described for FLASHDeconv deconvolution results

- Cosine threshold between averagine and observed isotope pattern: minimum value as described for FLASHDeconv deconvolution results

8.4 Results

- Unprocessed Raw Experimental Components: the number of raw experimental components for identification before merging artifacts
- Raw Experimental Components: the number of raw experimental components for identification after merging artifacts
- Missed Monoisotopic Raw Experimental Components Merged: the number of raw experimental components for identification artifacts merged due to being missed monoisotopic errors within the set mass tolerance
- Harmonic Raw Experimental Components Merged: the number of raw experimental components for identification artifacts merged due to being charge state harmonic errors within the set mass tolerance
- Unprocessed Raw Quantitative Components: the number of raw experimental components for quantification before merging artifacts
- Raw Quantitative Components: the number of raw experimental components for quantification after merging artifacts
- Missed Monoisotopic Raw Quantitative Components Merged: the number of raw experimental components for quantification artifacts merged due to being missed monoisotopic errors within the set mass tolerance
- Harmonic Raw Quantitative Components Merged: the number of raw experimental components for quantification artifacts merged due to being charge state harmonic errors within the set mass tolerance

The screenshot shows the Proteform Suite software interface. The top menu bar includes File, Method, Run Page, and a back/forward button. Below the menu is a navigation bar with tabs: Load Results, Theoretical Database, Top Down, Raw Experimental Components, New Code Pairs, Aggregated Proteforms, Experiment Theoretical Comparison, Experiment Experiment Comparison, Proteform Families, Identified Proteforms, Quantification, and Results Summary. The 'Aggregated Proteforms' tab is selected. On the left, there's a tree view of raw experimental components: LCA_UVS_1082... (1), LCA_DB_UVS_12... (1), and a collapsed section for Unprocessed Raw Experimental Components. The main area contains three input fields: Mass Tolerance for Merging Artifacts (ppm) set to 0.70, Cosine threshold between per-charge-intensities and fitted gaussian distribution set to 0.70, and Cosine threshold between average and observed isotopic pattern set to 0.70. To the right of these fields is a table titled 'Identified Proteforms'. The table has columns: Input File Unique ID, Input Filename, Input File Purpose, Scan Range, Component ID, Accepted, Weighted Monoisotopic Mass, No. Charge States, and Charge States. The table lists 12 entries, each corresponding to a raw experimental component from the tree view.

- Raw Experimental Components table: the top right table displays raw experimental components for either Identification or Quantification (depending on selection of Components Displayed to the left of the table)

This screenshot is nearly identical to the previous one, showing the same software interface and configuration. The main difference is the content of the raw experimental components tree view on the left. It now shows: LCA_UVS_1082... (1), LCA_DB_UVS_12... (1), and a collapsed section for Unprocessed Raw Experimental Components. The 'Identified Proteforms' table on the right also contains 12 entries, matching the structure and data of the previous screenshot.

- Input File Unique ID: file ID number for filename of Deconvolution Results for Identification or Quantification for this raw experimental component
- Input Filename: filename of Deconvolution Results for Identification or Quantification for this raw experimental component
- Input File Purpose: either Identification or Quantification
- Scan Range: MS scan range for this raw experimental component

- Component ID: Proteoform Suite given ID for this raw experimental component; file ID_component #
 - Accepted: checked if this raw experimental component is accepted for analysis
 - Weighted Monoisotopic Mass: monoisotopic mass of raw experimental component weighted by intensity of each charge state (more intense charge states are higher weighted)
 - No. Charges: number of charge states for this raw experimental component
 - Charge States: comma-separated list of charge states for this raw experimental component
 - Intensity Sum: intensity of this raw experimental component, charge state normalized
 - RT Range: retention time range
 - Apex RT: apex retention time as reported
 - Reported Monoisotopic Mass: monoisotopic mass reported by deconvolution input
 - Reported Intensity: intensity reported by deconvolution input
- Charge States table: the bottom right table displays all charge states from the raw experimental component selected in the Raw Experimental Components table (top right)

The screenshot shows the Proteoform Suite software interface. At the top, there is a navigation bar with tabs: Load Results, Theoretical Database, Top Down, Raw Experimental Components, New Code Park, Aggregated Proteoforms, Experiment Theoretical Comparison, Experiment Experiment Comparison, Proteoform Families, Identified Proteoforms, Quantification, and Results Summary. The 'Raw Experimental Components' tab is currently selected.

The main area contains two tables:

- Raw Experimental Components Table:** Shows a list of raw experimental components with columns: Filename, Biological, Fraction, Technical, Condition, and Components Displayed. One row is highlighted with a blue background: LCA_UVS_12082.
- Quantification Components Table:** Shows a list of quantification components with columns: Input File Unique ID, Input File Name, Input File Purpose, Scan Range, Component ID, Accepted, Weighted Monoisotopic Mass, No. Charge States, and Charge States. A specific row is highlighted with a red border: LCA_UVS_12082_Identification_285-292. This row corresponds to the highlighted row in the first table.

Below the tables, there are several input fields and dropdown menus:

- Mass Tolerance for Merging Artifacts (ppm): 5
- Cosine threshold between per-charge-intensities and fitted gaussian distribution: 0.70
- Cosine threshold between average and observed isotope pattern: 0.70
- Filter options: Unprocessed Raw Experimental Components, Raw Experimental Components, Missed Monoisotopic Raw Experimental Components Merged, Harmonic Raw Quantitative Components Merged, Unprocessed Raw Quantitative Components, Raw Quantitative Components, Missed Monoisotopic Raw Quantitative Components Merged, and Harmonic Raw Quantitative Components Merged.

- Calculated Mass: monoisotopic mass of this charge state
- Centroid m/z: monoisotopic m/z value of this charge state
- Intensity: charge normalized intensity of this charge state
- Charge count: number of this charge state

9 NeuCode Pairs

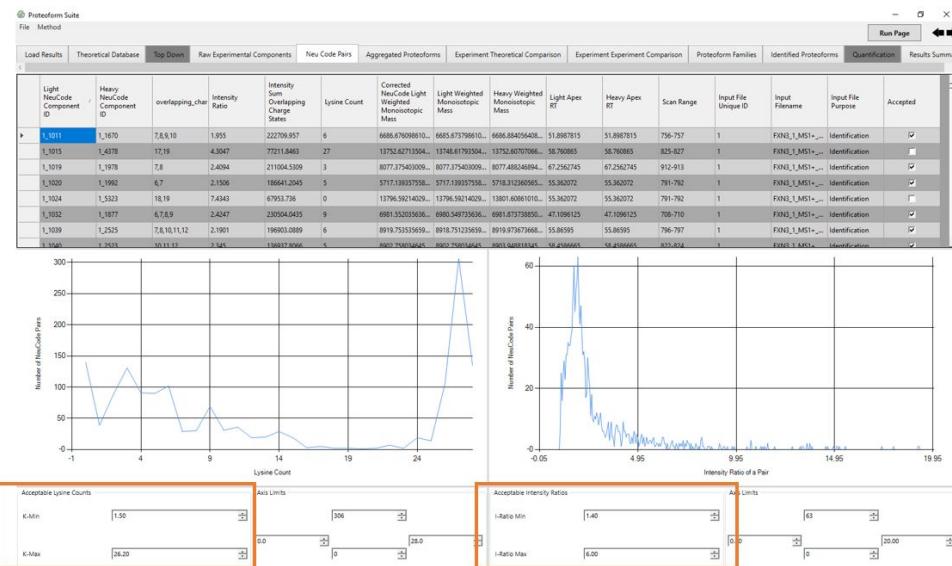
9.1 Overview

On this page, NeuCode pairs are generated from the raw experimental components. Each NeuCode pair consists of a light raw experimental component and a heavy raw experimental component; the lysine count is determined based upon the delta mass difference between the light and heavy raw experimental components of the NeuCode pair. For more information on generating NeuCode labeled data for proteoform identification in Proteoform Suite, see the methods sections of previously published analyses.^{1-3,6}

9.2 Run Page

- The Raw Experimental Components page must be run before running this page
- Click Run Page button (top right)
- Set all parameters as desired for current analysis (see below). The results automatically refresh; you do not need to re-run the page after adjusting the parameters. Only accepted NeuCode pairs will be included in further analysis in Proteoform Suite.

9.3 Set Parameters

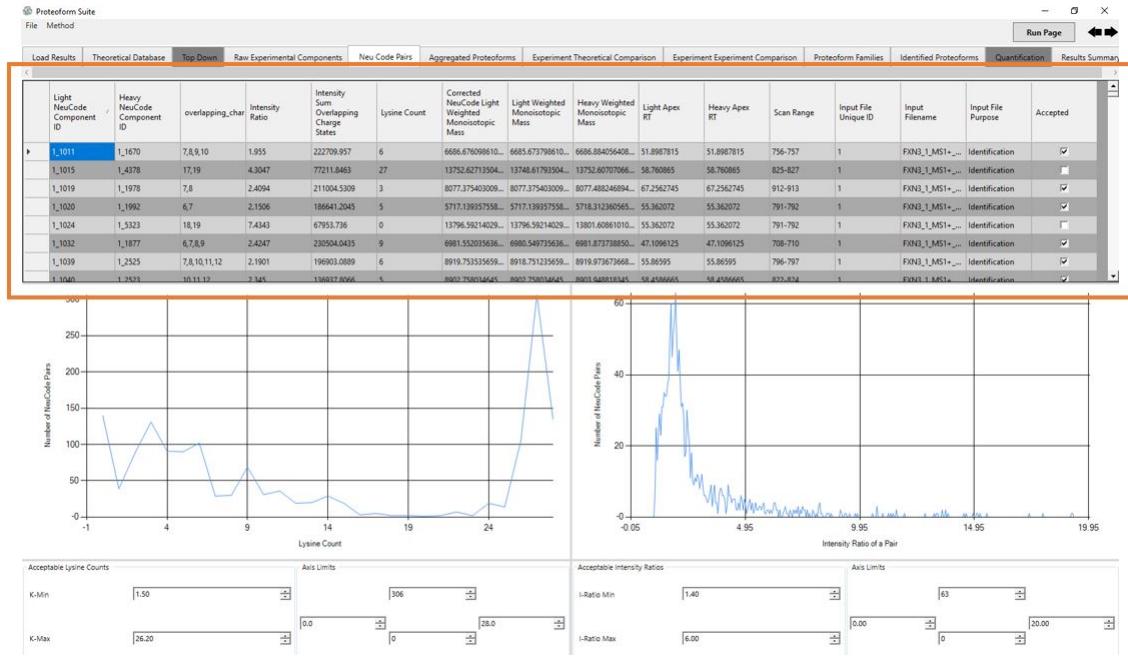


- Acceptable Lysine Counts: only NeuCode pairs with an acceptable number of lysines will be accepted
 - K-Min: minimum acceptable lysine count for a NeuCode pair to be accepted for further analysis
 - K-Max: maximum acceptable lysine count for a NeuCode pair to be accepted for further analysis

- Acceptable Intensity Ratios: only NeuCode pairs with an acceptable intensity ratio between light/heavy raw experimental components will be accepted
 - I-Ratio Min: minimum acceptable intensity ratio for a NeuCode pair to be accepted for further analysis
 - I-Ratio Max: maximum acceptable intensity ratio for a NeuCode pair to be accepted for further analysis

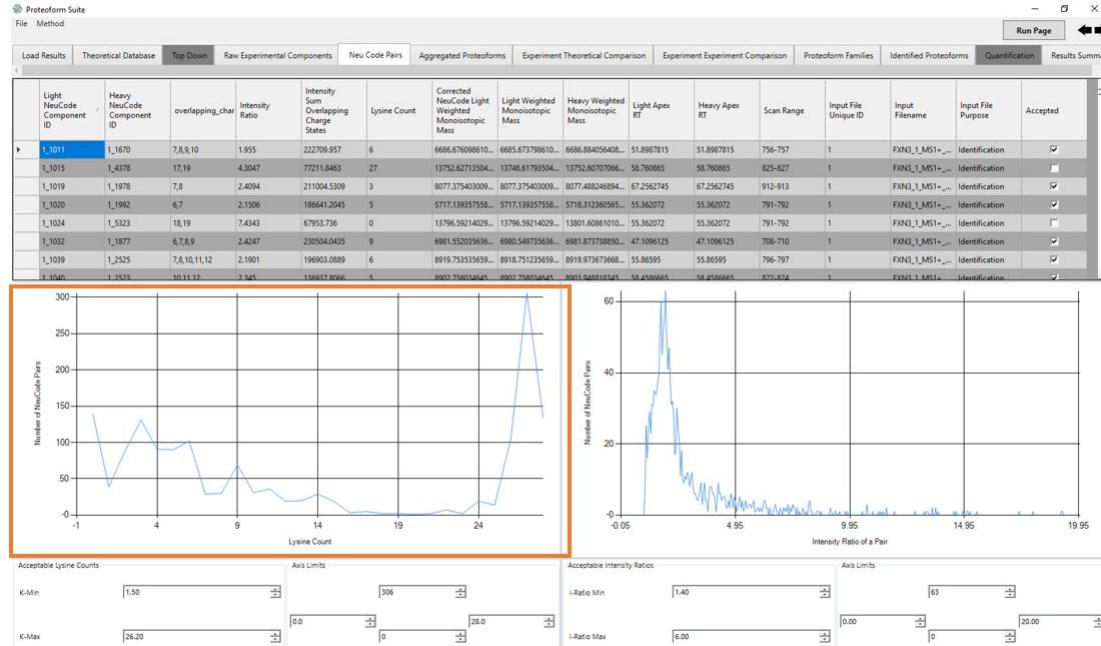
9.4 Results

- NeuCode pairs table: the top table displays all NeuCode pairs generated

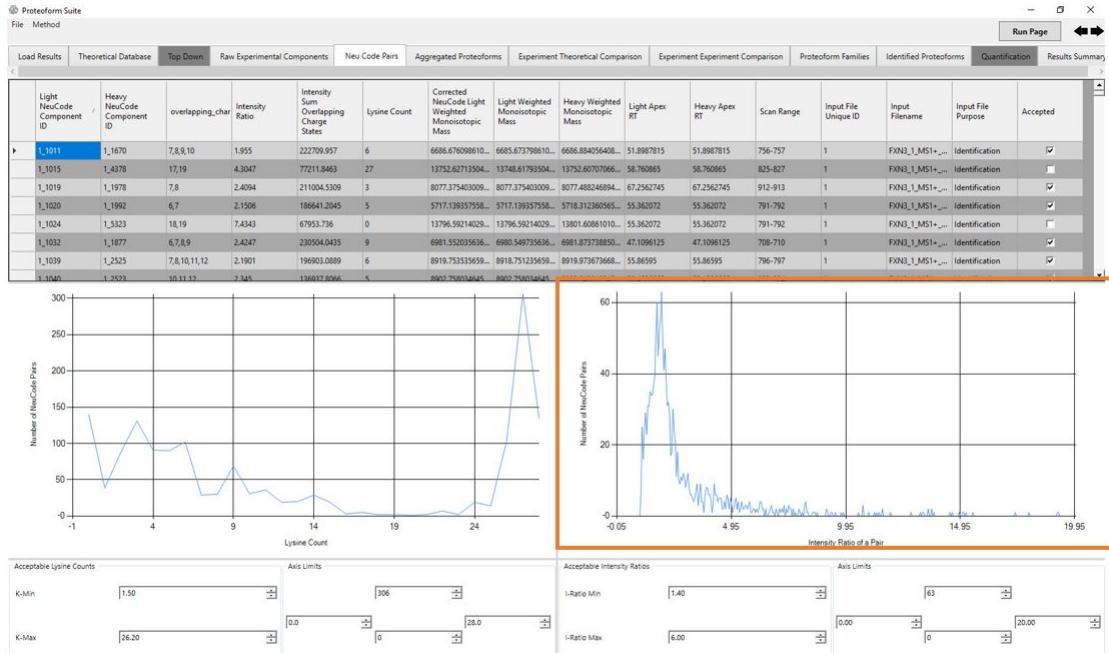


- Light NeuCode Component ID: Proteoform Suite given ID for the light raw experimental component; file ID_component #
- Heavy NeuCode Component ID: Proteoform Suite given ID for the heavy raw experimental component; file ID_component #
- Overlapping Charge States: comma-separated list of charge states observed for both the light and heavy raw experimental components components in this NeuCode pair
- Intensity Ratio: intensity ratio of light raw experimental component intensity : heavy raw experimental component intensity
- Intensity Sum Overlapping Charge States: summed intensity for all charge states observed for both the light and heavy raw experimental components in this NeuCode pair
- Lysine Count: lysine count for this NeuCode pair; determined by mass difference between light and heavy raw experimental components

- Corrected NeuCode Light Weighted Monoisotopic Mass: monoisotopic mass of light raw experimental component used for this NeuCode pair, monoisotopic mass errors corrected
- Light Weighted Monoisotopic Mass: monoisotopic mass of light raw experimental component
- Heavy Weighted Monoisotopic Mass: monoisotopic mass of heavy raw experimental component
- Light Apex RT: apex retention time of light raw experimental component
- Scan Range: MS scan range for light and heavy raw experimental components in this NeuCode pair
- Input File Unique ID: file ID number for filename of Deconvolution Results for Identification or Quantification for the light and heavy raw experimental components in this NeuCode pair
- Input Filename: filename of Deconvolution Results for Identification or Quantification for light and heavy raw experimental components in this NeuCode pair
- Input File Purpose: either Identification or Quantification
- Accepted: checked if this NeuCode pair has a lysine count and intensity ratio within the min and max range allowed (see Set Parameters); if accepted, this NeuCode pair will be utilized in subsequent Proteoform Suite analysis
- Lysine Count histogram: the left graph shows a histogram of lysine counts for all NeuCode pairs. The Axis Limits box below this graph adjusts the x- and y-axes.



- Intensity Ratio histogram: the right graph shows a histogram of light:heavy intensity ratios for all NeuCode pairs. The peak of this histogram should fall close to the experimentally performed mixing ratio for light:heavy protein samples (ex: 2:1 light:heavy). The Axis Limits box below this graph adjusts the x- and y-axes.



10 Aggregated Proteoforms

10.1 Overview

On this page, experimental proteoforms are created by aggregating either raw experimental components (unlabeled analysis) or NeuCode pairs (NeuCode labeled analysis). If Deconvolution Results for Quantification were provided on the Load Results page, these raw experimental components for quantification will be binned with experimental proteoforms based upon the set parameters. The experimental proteoforms are used in intact-mass analysis to identify proteoforms and construct proteoform families.

10.2 Run Page

- The Raw Experimental Components page must be run before running this page. If applicable, the Top-Down and the NeuCode Pairs pages must also be run before running this page.
- Set all parameters as desired for current analysis (see below)
- Click Run Page button (top right)

10.3 Set Parameters

The screenshot shows the 'Aggregated Proteoforms' tab selected in the top navigation bar. The left sidebar contains various filter options and checkboxes for 'Add Top-Down Proteoforms' and 'Validate Proteoforms'. The main area displays a table of aggregated proteoforms with columns for experimental proteoform ID, mass tolerance, retention time tolerance, and other quantitative and qualitative parameters. A detailed view of one row is shown in a modal window at the bottom, displaying intensity ratios, scan ranges, and input file details for each component.

Experimental Proteoform ID	Mass Tolerance (ppm)	Retention Time Tolerance (min)	Intensity Sum Overlapping Charge States	Lysine Count	Connected NeuCode Light Monoisotopic Mass	Light Weighted Monoisotopic Mass	Heavy Weighted Monoisotopic Mass	Light Ape RT	Heavy Ape RT	Scan Range	Input File Unique ID	Input File Name	Input File Purpose	Accepted
E_29	8, 72	5, 6, 7, 8	2,5491	2183532,5985	7966,08374039..., 7962,07514539...	7668,02655951..., 7668,02655951...	96,866498	96,866498	1176-1177	8	F004_2_MS1_C...	Identification	<input checked="" type="checkbox"/>	
E_34	8, 102	5, 6, 7, 8	2,8314	10902,021,1022	7966,08454696..., 7962,07514596...	7668,02324563..., 7668,02324563...	96,763608	96,763608	1175-1176	8	F004_2_MS1_C...	Identification	<input checked="" type="checkbox"/>	
E_36	8, 100	5, 6, 7, 8	2,6622	10969,732,4543	7966,08731519..., 7962,07815619...	7668,03635473..., 7668,03635473...	96,947435	96,947435	1177-1178	8	F004_2_MS1_C...	Identification	<input checked="" type="checkbox"/>	
E_43	8, 126	5, 6, 7, 8	2,7513	15325,696,622	7966,08699563..., 7962,07849956...	7668,03722697..., 7668,03722697...	97,071945	97,071945	1177-1179	8	F004_2_MS1_C...	Identification	<input checked="" type="checkbox"/>	
E_45	8, 144	5, 6, 7, 8	3,1166	15051,714,8965	7966,09110707..., 7962,07981070...	7668,03516502..., 7668,03516502...	96,751995	96,751995	1174-1176	8	F004_2_MS1_C...	Identification	<input checked="" type="checkbox"/>	
E_51	8, 180	5, 6, 7, 8	2,9667	10596476,3035	7966,05205935..., 7962,05205935...	7668,04194826..., 7668,04194826...	97,119905	97,119905	1178-1180	8	F004_2_MS1_C...	Identification	<input checked="" type="checkbox"/>	
E_74	8, 316	5, 6, 7, 8	4,3243	834002,8343	7966,08731247..., 7962,07811247...	7668,03987905..., 7668,03987905...	96,841452	96,841452	1177-1178	8	F004_2_MS1_C...	Identification	<input checked="" type="checkbox"/>	
E_103	8, 312	5, 6, 7	3,3911	6539500,4462	7966,09119993..., 7962,08999993...	7668,04471890..., 7668,04471890...	97,219858	97,219858	1176-1181	8	F004_2_MS1_C...	Identification	<input checked="" type="checkbox"/>	
E_132	8, 421	5, 6, 7	3,4456	5194996,4772	7966,09589526..., 7962,08699526...	7668,04731146..., 7668,04731146...	97,270343	97,270343	1180-1181	8	F004_2_MS1_C...	Identification	<input checked="" type="checkbox"/>	
E_185	8, 364	5, 6, 7	3,8289	4189536,0681	7966,08791811..., 7962,08891811...	7668,05148494..., 7668,05148494...	97,372265	97,372265	1181-1182	8	F004_2_MS1_C...	Identification	<input checked="" type="checkbox"/>	
E_424	8, 1181	5, 6, 7	4,5105	1665470,1109	7966,10090502..., 7962,0917736052...	7668,056728765...	97,977145	97,977145	1187-1188	8	F004_2_MS1_C...	Identification	<input checked="" type="checkbox"/>	

- Mass Tolerance (ppm): mass tolerance used to aggregate raw experimental components (unlabeled) or NeuCode pairs (NeuCode labeled)
- Ret. Time Tolerance (min): retention time tolerance used to aggregate raw experimental components (unlabeled) or NeuCode pairs (NeuCode labeled)

- Missed Monoisotopics (num): number of missed monoisotopic units to aggregate raw experimental components (unlabeled) or NeuCode pairs (NeuCode labeled)
- Missed Lysine Counts (num): the number of missed lysine counts to aggregate NeuCode pairs (accounts for $\pm 100\%$ labeling efficiency)
- Min. # Consecutive Charge States: minimum number of charge states for a raw experimental component to be considered for aggregation
- Minimum Required Observations: set the # (left) and the requirement (right drop down box) to require observations of an experimental proteoform in more than one file type. Must have labeled biological replicates, technical replicates, and conditions for Deconvolution Results for Identification on the Load Results page
- Add Top-Down Proteoforms: if checked, top-down proteoforms from the Top Down page will be aggregated with experimental proteoforms created from raw experimental components (unlabeled) or NeuCode pairs (NeuCode labeled). Top-down proteoforms will replace experimental proteoforms using the set parameters above
- Validate Proteoforms: if checked and if NeuCode labeled data, will verify that aggregated NeuCode pairs are in the tolerances determined with set parameters

10.4 Results

- Total Accepted Aggregated Proteoforms: the number of accepted experimental proteoforms that will be included in further analysis
- Aggregated Proteoform Table Filter: filter the Aggregated Proteoforms table (top right) by any entered text
- Components Displayed Upon Selecting an Experimental Proteoform: determines which components will be displayed in the Components table (bottom) when an aggregated proteoform is selected in the Aggregated Proteoform Table (right)

Proteoform Suite

File Method

Load Results Theoretical Database Top Down Raw Experimental Components Neu Code Pairs Aggregated Proteoforms Experiment Theoretical Comparison Experiment Experiment Comparison Proteoform Families Identified Proteoforms Quantification Results Summary

Mass Tolerance (ppm) 5
Ret. Time Tolerance (min) 5.00
Missed Monoisotopes (num) 3
Missed Lysine Counts (num) 2
Min. # Consecutive Charge Sta 1
Minimum Required Observations 1 Minimum Biopsies with Observations From Any Single Condition
Add Top-Down Proteoforms Validate Proteoforms

Total Accepted Aggregated Proteoforms 221

Aggregated Proteoform Table Filter Components Displayed upon Selecting an Experimental Proteoform
 Identification Components
 Light Quantification Components
 Heavy Quantification Components

Experimental Proteoform ID	Aggregated RT	Aggregated Mass	Aggregated Decconvolution Intensity	Top-Down Proteoform	Lysine Count	Manually Shifted Mass	Aggregated Component Count for Identification	Description
E1	7666.0800	96.94	125312851.8974		26		11	
E2	6791.1951	51.00	2042350.1199		14		5	
E1	6001.3020	49.09	2821022.0338		26		2	
E7	13763.1648	54.67	16881623.3999		12		3	
E5	10937.9847	55.13	11630992.2957		24		3	
E9	5546.1483	52.76	15006392.1835		4		12	
E11	6745.2085	51.05	1337806.5033		26		5	
E3	14033.8251	58.99	12290567.1853		5		3	
E8	6824.1956	50.88	9600853.1711		14		2	
E4	13987.8415	58.28	8595019.3960		3		3	
E6	6890.2437	48.83	6196097.7287		26		1	
E9	5735.1450	55.05	6081402.7211		5		8	
E16	6734.1573	50.85	1067143.1515		15		4	
E14	6887.2803	49.87	9104862.7027		26		2	
E13	6795.1935	51.62	8820857.1521		7		1	
E41	6470.4696	51.94	5078342.7468		6		9	
E20	13543.7230	52.80	5124645.4334		17		3	

Light NeuCode Component ID	Heavy NeuCode Component ID	overlapping_char	Intensity Ratio	Intensity Sum Overlapping Charge States	Lysine Count	Connected NeuCode Light Weighted Monoisotopic Mass	Light Weighted Monoisotopic Mass	Heavy Weighted Monoisotopic Mass	Light Aces RT	Heavy Aces RT	Scan Range	Input File Unique ID	Input File Name	Input File Purpose	Accepted
E_29	8_72	5,6,7,8	2,5491	21835322.5865	26	7666.08374039..	7662.07454039..	7666.03265591...	96.866468	97.715195	1176-1177	0	FN4A_2_MS1_C...	Identification	<input checked="" type="checkbox"/>
E_34	8_102	5,6,7,8	2,8314	1900201.1022	26	7666.08451495..	7662.07314956..	7666.03245603...	96.795608	97.715008	1175-1176	8	FN4A_2_MS1_C...	Identification	<input checked="" type="checkbox"/>
E_36	8_100	5,6,7,8	2,6622	18069732.4543	26	7666.08756196..	7662.07815616..	7666.03815473...	96.967425	96.967425	1177-1178	8	FN4A_2_MS1_C...	Identification	<input checked="" type="checkbox"/>
E_43	8_126	5,6,7,8	2,7513	15325696.623	26	7666.08899603..	7662.07949603..	7666.03726971...	97.071945	97.071945	1177-1179	8	FN4A_2_MS1_C...	Identification	<input checked="" type="checkbox"/>
E_45	8_144	5,6,7,8	3,1166	1503174.8965	26	7666.08818070..	7662.07801807..	7666.03316502...	96.751595	96.751595	1176-1177	0	FN4A_2_MS1_C...	Identification	<input checked="" type="checkbox"/>
E_51	8_188	5,6,7,8	2,9607	10596478.3035	26	7666.08263863..	7662.08238633..	7666.04148216..	97.119508	97.119508	1176-1180	8	FN4A_2_MS1_C...	Identification	<input checked="" type="checkbox"/>
E_74	8_316	5,6,7,8	4,3243	8340023.543	26	7666.08712473..	7662.07812473..	7666.03987908..	96.841482	96.841482	1177-1175	8	FN4A_2_MS1_C...	Identification	<input checked="" type="checkbox"/>
E_103	8_312	5,6,7	3,3911	659500.442	26	7666.09159963..	7662.08899526..	7666.04471890..	97.219585	97.219585	1176-1181	8	FN4A_2_MS1_C...	Identification	<input checked="" type="checkbox"/>
E_132	8_421	5,6,7	3,4456	519496.472	26	7666.09595226..	7662.08695226..	7666.04721146..	97.27043	97.27043	1180-1181	8	FN4A_2_MS1_C...	Identification	<input checked="" type="checkbox"/>
E_165	8_564	5,6,7	3,8289	4189536.0881	26	7666.09798113..	7662.08898113..	7666.05148694..	97.371205	97.371205	1181-1182	8	FN4A_2_MS1_C...	Identification	<input checked="" type="checkbox"/>
E_424	8_1181	5,6,7	4,5105	1665470.1109	26	7666.10093602..	7662.09179052..	7666.05627675..	97.877145	97.877145	1187-1188	8	FN4A_2_MS1_C...	Identification	<input checked="" type="checkbox"/>

- Aggregated Proteoforms table: the top right table displays the aggregated experimental proteoforms

Proteoform Suite

File Method

Load Results Theoretical Database Top Down Raw Experimental Components Neu Code Pairs Aggregated Proteoforms Experiment Theoretical Comparison Experiment Experiment Comparison Proteoform Families Identified Proteoforms Quantification Results Summary

Mass Tolerance (ppm) 5
Ret. Time Tolerance (min) 5.00
Missed Monoisotopes (num) 3
Missed Lysine Counts (num) 2
Min. # Consecutive Charge Sta 1
Minimum Required Observations 1 Minimum Biopsies with Observations From Any Single Condition
Add Top-Down Proteoforms Validate Proteoforms

Total Accepted Aggregated Proteoforms 221

Aggregated Proteoform Table Filter Components Displayed upon Selecting an Experimental Proteoform
 Identification Components
 Light Quantification Components
 Heavy Quantification Components

Experimental Proteoform ID	Aggregated RT	Aggregated Mass	Aggregated Decconvolution Intensity	Top-Down Proteoform	Lysine Count	Manually Shifted Mass	Aggregated Component Count for Identification	Description
E1	7666.0800	96.94	125312851.8974		26		11	
E2	6791.1951	51.00	2042350.1199		14		5	
E1	6001.3020	49.09	2821022.0338		26		2	
E7	13763.1648	54.67	16881623.3999		12		3	
E5	10937.9847	55.13	11630992.2957		24		3	
E9	5546.1483	52.76	15006392.1835		4		12	
E11	6745.2085	51.05	1337806.5033		26		5	
E3	14033.8251	58.99	12290567.1853		5		3	
E8	6824.1956	50.88	9600853.1711		14		2	
E4	13987.8415	58.28	8595019.3960		3		3	
E6	6890.2437	48.83	6196097.7287		26		1	
E5	5735.1450	55.05	6081402.7211		5		8	
E16	6734.1573	50.85	1067143.1515		15		4	
E14	6887.2803	49.87	9104862.7027		26		2	
E13	6795.1935	51.62	8820857.1521		7		1	
E41	6470.4696	51.94	5078342.7468		6		9	
E20	13543.7230	52.80	5124645.4334		17		3	

Light NeuCode Component ID	Heavy NeuCode Component ID	overlapping_char	Intensity Ratio	Intensity Sum Overlapping Charge States	Lysine Count	Connected NeuCode Light Weighted Monoisotopic Mass	Light Weighted Monoisotopic Mass	Heavy Weighted Monoisotopic Mass	Light Aces RT	Heavy Aces RT	Scan Range	Input File Unique ID	Input File Name	Input File Purpose	Accepted
E_29	8_72	5,6,7,8	2,5491	21835322.5865	26	7666.08374039..	7662.07454039..	7666.03265591...	96.866468	97.715195	1176-1177	0	FN4A_2_MS1_C...	Identification	<input checked="" type="checkbox"/>
E_34	8_102	5,6,7,8	2,8314	1900201.1022	26	7666.08451495..	7662.07314956..	7666.03245603...	96.795608	97.715008	1175-1176	8	FN4A_2_MS1_C...	Identification	<input checked="" type="checkbox"/>
E_36	8_100	5,6,7,8	2,6622	18069732.4543	26	7666.08756196..	7662.07815616..	7666.03815473...	96.967425	96.967425	1177-1178	8	FN4A_2_MS1_C...	Identification	<input checked="" type="checkbox"/>
E_43	8_126	5,6,7,8	2,7513	15325696.623	26	7666.08899603..	7662.07949603..	7666.03726971...	97.071945	97.071945	1177-1179	8	FN4A_2_MS1_C...	Identification	<input checked="" type="checkbox"/>
E_45	8_144	5,6,7,8	3,1166	1503174.8965	26	7666.08818070..	7662.07801807..	7666.03316502...	96.751595	96.751595	1176-1177	0	FN4A_2_MS1_C...	Identification	<input checked="" type="checkbox"/>
E_51	8_188	5,6,7,8	2,9607	10596478.3035	26	7666.08263863..	7662.08238633..	7666.04148216..	97.119508	97.119508	1176-1180	8	FN4A_2_MS1_C...	Identification	<input checked="" type="checkbox"/>
E_74	8_316	5,6,7,8	4,3243	8340023.543	26	7666.08712473..	7662.07812473..	7666.03987908..	96.841482	96.841482	1177-1175	8	FN4A_2_MS1_C...	Identification	<input checked="" type="checkbox"/>
E_103	8_312	5,6,7	3,3911	659500.442	26	7666.09159963..	7662.08899526..	7666.04471890..	97.219585	97.219585	1176-1181	8	FN4A_2_MS1_C...	Identification	<input checked="" type="checkbox"/>
E_132	8_421	5,6,7	3,4456	519496.472	26	7666.09595226..	7662.08695226..	7666.04721146..	97.27043	97.27043	1180-1181	8	FN4A_2_MS1_C...	Identification	<input checked="" type="checkbox"/>
E_165	8_564	5,6,7	3,8289	4189536.0881	26	7666.09798113..	7662.08898113..	7666.05148694..	97.371205	97.371205	1181-1182	8	FN4A_2_MS1_C...	Identification	<input checked="" type="checkbox"/>
E_424	8_1181	5,6,7	4,5105	1665470.1109	26	7666.10093602..	7662.09179052..	7666.05627675..	97.877145	97.877145	1187-1188	8	FN4A_2_MS1_C...	Identification	<input checked="" type="checkbox"/>

- Experimental Proteoform ID: unique ID given by Proteoform Suite for this experimental proteoform
- Aggregated Mass: monoisotopic mass of experimental proteoform, weighted average of (light) raw experimental components by intensity
- Aggregated RT: retention time of experimental proteoform, average of raw experimental components (unlabeled) or NeuCode pairs (NeuCode labeled)

- Aggregated Deconvolution Intensity: sum of intensity of aggregated raw experimental components
- Top-Down Proteoform: checked if this experimental proteoform is a top-down identified proteoform
- Lysine Count: number of lysines (NeuCode labeled data)
- Manually Shifted Mass: checked if monoisotopic mass was shifted by an isotopic mass unit (see **Experiment-Theoretical Comparison** section)
- Aggregated Component Count for Identification: number of aggregated raw experimental components (unlabeled) or NeuCode pairs (NeuCode labeled). If top-down proteoform, the number of top-down hits
- Description: if identified, the protein description from UniProt
- Gene Name: if identified, the gene name
- GeneID: if identified, the gene ID
- Grouped Accessions: if identified, the protein accessions from UniProt
- PTM Description: if identified, the post-translational modifications on this experimental proteoform
- Begin and End: if identified, the experimental proteoform begin and end residue in protein full sequence in UniProt
- Sequence: if identified, proteoform sequence
- UniProt-Annotated Modifications: if identified, all UniProt annotated residues for PTMs on this experimental proteoform in UniProt database provided
- Potentially Novel Mods: if identified, checked if this experimental proteoform contains PTMs not annotated in UniProt database provided
- Bottom-Up PSMs Count: if identified, the number of bottom-up PSMs derived from this experimental proteoform
- Modified Bottom-Up PSMs: if identified, modified residues confirmed by ID'd bottom-up peptides derived from this experimental proteoform
- Bottom-Up Evidence for All PTMs: if identified, checked if all PTMs on this experimental
- Level: proteoform identification level based on five-level scheme¹⁸
- Level Description: description of proteoform level assignment (sources of ambiguity)
- New Intact-Mass ID: if identified, this intact-mass identification was not identified by top-down
- Ambiguous: checked if ambiguous intact-mass identification
- Adduct: checked if identification is due to presence of an adduct (oxidation, SDS adduct, sulfate adduct)
- Contaminant: checked if identification is from a theoretical proteoform in a contaminant database
- Mass Error: mass difference between observed mass and theoretical mass of this identification

- Family ID: proteoform family number (must have run through full Proteoform Suite analysis)
- Family: identified, ambiguous, or unidentified proteoform family
- Linked Proteoform References: proteoforms in family network path of identification to the nearest theoretical proteoform (must have run through full Proteoform Suite analysis)
- M/z values: m/z values observed for this experimental proteoform
- Charges: charge state numbers observed for this experimental proteoform
- Abundant Component for Manual Validation of Identification: file information for the most abundant raw experimental component aggregated into this experimental proteoform
- Components table: the bottom table displays all raw experimental components (unlabeled) or NeuCode pairs (NeuCode labeled) aggregated into the experimental proteoform selected in the Aggregated Proteoforms table (top right)

Proteoform Suite
File Method

Load Results	Theoretical Database	Top Down	Raw Experimental Components	Neu Code Pairs	Aggregated Proteoforms	Experiment Theoretical Comparison	Experiment Experiment Comparison	Proteoform Families	Identified Proteoforms	Quantification	Results Summary
Mass Tolerance (ppm)	5										
Ret. Time Tolerance (min)	5.00										
Missed Monoisotopes (num)	3										
Missed Lysine Counts (num)	2										
Min. # Consecutive Charge Sta	1										
Minimum Required Observations	1	<input type="checkbox"/> Minimum Bumps with Observations From Any Single Condition									
Add Top-Down Proteoforms	<input checked="" type="checkbox"/>	Validate Proteoforms									
Total Accepted Aggregated Proteoforms	221										
Aggregated Proteoform Table Filter											
Components Displayed upon Selecting an Experimental Proteoform	<input checked="" type="checkbox"/> Identification Components	<input type="checkbox"/> Light Quantification Components	<input type="checkbox"/> Heavy Quantification Components								

Experimental Proteoform ID: 6946_0008
Aggregated Mass: 96.94
Aggregated RT: 125312051.8974
Aggregated Decconvolution Intensity: 26
Top-Down Proteoform: Lysine Count: 11
Manually Shifted Mass: 52.76
Aggregated Component ID: 11
Description: 11

Experimental Proteoform ID: 6970_1951
Aggregated Mass: 51.00
Aggregated RT: 2642380.0199
Aggregated Decconvolution Intensity: 14
Top-Down Proteoform: Lysine Count: 5
Manually Shifted Mass: 52.76
Aggregated Component ID: 5
Description: 5

Experimental Proteoform ID: 6901_3023
Aggregated Mass: 49.09
Aggregated RT: 2821802.0038
Aggregated Decconvolution Intensity: 26
Top-Down Proteoform: Lysine Count: 2
Manually Shifted Mass: 52.76
Aggregated Component ID: 2
Description: 2

Experimental Proteoform ID: 13783_3648
Aggregated Mass: 54.67
Aggregated RT: 16681623.3999
Aggregated Decconvolution Intensity: 12
Top-Down Proteoform: Lysine Count: 3
Manually Shifted Mass: 52.76
Aggregated Component ID: 3
Description: 3

Experimental Proteoform ID: 10937_0647
Aggregated Mass: 55.13
Aggregated RT: 14639909.2957
Aggregated Decconvolution Intensity: 24
Top-Down Proteoform: Lysine Count: 3
Manually Shifted Mass: 52.76
Aggregated Component ID: 3
Description: 3

Experimental Proteoform ID: 5564_1483
Aggregated Mass: 52.76
Aggregated RT: 15090630.3035
Aggregated Decconvolution Intensity: 4
Top-Down Proteoform: Lysine Count: 12
Manually Shifted Mass: 52.76
Aggregated Component ID: 12
Description: 12

Experimental Proteoform ID: 4754_3095
Aggregated Mass: 51.05
Aggregated RT: 13179095.0033
Aggregated Decconvolution Intensity: 5
Top-Down Proteoform: Lysine Count: 5
Manually Shifted Mass: 52.76
Aggregated Component ID: 5
Description: 5

Experimental Proteoform ID: 14033_8251
Aggregated Mass: 58.99
Aggregated RT: 32290567.0353
Aggregated Decconvolution Intensity: 5
Top-Down Proteoform: Lysine Count: 3
Manually Shifted Mass: 52.76
Aggregated Component ID: 3
Description: 3

Experimental Proteoform ID: 6824_1955
Aggregated Mass: 50.88
Aggregated RT: 9608551.1711
Aggregated Decconvolution Intensity: 14
Top-Down Proteoform: Lysine Count: 2
Manually Shifted Mass: 52.76
Aggregated Component ID: 2
Description: 2

Experimental Proteoform ID: 13987_0415
Aggregated Mass: 58.29
Aggregated RT: 8395010.3960
Aggregated Decconvolution Intensity: 3
Top-Down Proteoform: Lysine Count: 3
Manually Shifted Mass: 52.76
Aggregated Component ID: 3
Description: 3

Experimental Proteoform ID: 6879_2407
Aggregated Mass: 48.82
Aggregated RT: 6196097.7287
Aggregated Decconvolution Intensity: 26
Top-Down Proteoform: Lysine Count: 1
Manually Shifted Mass: 52.76
Aggregated Component ID: 1
Description: 1

Experimental Proteoform ID: 3753_1450
Aggregated Mass: 55.05
Aggregated RT: 6081340.7121
Aggregated Decconvolution Intensity: 5
Top-Down Proteoform: Lysine Count: 6
Manually Shifted Mass: 52.76
Aggregated Component ID: 6
Description: 6

Experimental Proteoform ID: 6724_1673
Aggregated Mass: 50.85
Aggregated RT: 10676143.1515
Aggregated Decconvolution Intensity: 15
Top-Down Proteoform: Lysine Count: 4
Manually Shifted Mass: 52.76
Aggregated Component ID: 4
Description: 4

Experimental Proteoform ID: 6897_2853
Aggregated Mass: 49.67
Aggregated RT: 915465.7027
Aggregated Decconvolution Intensity: 26
Top-Down Proteoform: Lysine Count: 2
Manually Shifted Mass: 52.76
Aggregated Component ID: 2
Description: 2

Experimental Proteoform ID: 6792_1935
Aggregated Mass: 51.42
Aggregated RT: 8620051.1521
Aggregated Decconvolution Intensity: 7
Top-Down Proteoform: Lysine Count: 1
Manually Shifted Mass: 52.76
Aggregated Component ID: 1
Description: 1

Experimental Proteoform ID: 6470_4096
Aggregated Mass: 51.94
Aggregated RT: 317842.7488
Aggregated Decconvolution Intensity: 6
Top-Down Proteoform: Lysine Count: 9
Manually Shifted Mass: 52.76
Aggregated Component ID: 9
Description: 9

Experimental Proteoform ID: 12543_7230
Aggregated Mass: 52.80
Aggregated RT: 5124645.4034
Aggregated Decconvolution Intensity: 17
Top-Down Proteoform: Lysine Count: 3
Manually Shifted Mass: 52.76
Aggregated Component ID: 3
Description: 3

Light NeuCode Component ID: 8_39
Heavy NeuCode Component ID: 9_72
overlapping_charge: 2.5491
Intensity Ratio: 21635532.8965
Intensity Sum Overlapping Charge States: 26
Lysine Count: 7956.02740339
Corrected NeuCode: 7942.07454039
Light Weighted Monoisotopic Mass: 96.865498
Heavy Weighted Monoisotopic Mass: 96.865498
Light Apex RT: 1176-1177
Heavy Apex RT: 1176-1176
Scan Range: 0
Input File Unique ID: F1XN4_2_M51_C...
Input Filename: F1XN4_2_M51_C...
Input Purpose: Identification
Accepted:

Light NeuCode Component ID: 8_34
Heavy NeuCode Component ID: 8_102
overlapping_charge: 2.8314
Intensity Ratio: 19002201.1022
Intensity Sum Overlapping Charge States: 26
Lysine Count: 7966.08431496
Corrected NeuCode: 7952.07514966
Light Weighted Monoisotopic Mass: 96.865608
Heavy Weighted Monoisotopic Mass: 96.865608
Light Apex RT: 1175-1176
Heavy Apex RT: 1175-1176
Scan Range: 8
Input File Unique ID: F1XN4_2_M51_C...
Input Filename: F1XN4_2_M51_C...
Input Purpose: Identification
Accepted:

Light NeuCode Component ID: 8_36
Heavy NeuCode Component ID: 8_100
overlapping_charge: 2.6462
Intensity Ratio: 1809732.4243
Intensity Sum Overlapping Charge States: 26
Lysine Count: 7966.07375196
Corrected NeuCode: 7952.07514966
Light Weighted Monoisotopic Mass: 96.864735
Heavy Weighted Monoisotopic Mass: 96.864735
Light Apex RT: 1177-1178
Heavy Apex RT: 1177-1178
Scan Range: 8
Input File Unique ID: F1XN4_2_M51_C...
Input Filename: F1XN4_2_M51_C...
Input Purpose: Identification
Accepted:

Light NeuCode Component ID: 8_43
Heavy NeuCode Component ID: 8_126
overlapping_charge: 2.7513
Intensity Ratio: 15129986.42
Intensity Sum Overlapping Charge States: 26
Lysine Count: 7966.08999683
Corrected NeuCode: 7952.07499683
Light Weighted Monoisotopic Mass: 96.8651729971
Heavy Weighted Monoisotopic Mass: 96.8651729971
Light Apex RT: 97017945
Heavy Apex RT: 97017945
Scan Range: 8
Input File Unique ID: F1XN4_2_M51_C...
Input Filename: F1XN4_2_M51_C...
Input Purpose: Identification
Accepted:

Light NeuCode Component ID: 8_45
Heavy NeuCode Component ID: 8_144
overlapping_charge: 3.1166
Intensity Ratio: 1531714.0495
Intensity Sum Overlapping Charge States: 26
Lysine Count: 7966.08181070
Corrected NeuCode: 7952.07691070
Light Weighted Monoisotopic Mass: 96.8651729971
Heavy Weighted Monoisotopic Mass: 96.8651729971
Light Apex RT: 7688.03351852
Heavy Apex RT: 97151585
Scan Range: 8
Input File Unique ID: F1XN4_2_M51_C...
Input Filename: F1XN4_2_M51_C...
Input Purpose: Identification
Accepted:

Light NeuCode Component ID: 8_41
Heavy NeuCode Component ID: 8_188
overlapping_charge: 2.9607
Intensity Ratio: 10209478.8035
Intensity Sum Overlapping Charge States: 26
Lysine Count: 7966.02926383
Corrected NeuCode: 7952.02930383
Light Weighted Monoisotopic Mass: 96.8641942164
Heavy Weighted Monoisotopic Mass: 96.8641942164
Light Apex RT: 97119086
Heavy Apex RT: 97119086
Scan Range: 8
Input File Unique ID: F1XN4_2_M51_C...
Input Filename: F1XN4_2_M51_C...
Input Purpose: Identification
Accepted:

Light NeuCode Component ID: 8_74
Heavy NeuCode Component ID: 8_318
overlapping_charge: 4.3243
Intensity Ratio: 8340027.8344
Intensity Sum Overlapping Charge States: 26
Lysine Count: 7966.07512475
Corrected NeuCode: 7952.07812475
Light Weighted Monoisotopic Mass: 96.864423
Heavy Weighted Monoisotopic Mass: 96.864423
Light Apex RT: 1173-1175
Heavy Apex RT: 1173-1175
Scan Range: 8
Input File Unique ID: F1XN4_2_M51_C...
Input Filename: F1XN4_2_M51_C...
Input Purpose: Identification
Accepted:

Light NeuCode Component ID: 8_169
Heavy NeuCode Component ID: 8_312
overlapping_charge: 3.3991
Intensity Ratio: 659506.442
Intensity Sum Overlapping Charge States: 26
Lysine Count: 7966.05199961
Corrected NeuCode: 7952.05299961
Light Weighted Monoisotopic Mass: 97.219586
Heavy Weighted Monoisotopic Mass: 97.219586
Light Apex RT: 1179-1181
Heavy Apex RT: 1179-1181
Scan Range: 8
Input File Unique ID: F1XN4_2_M51_C...
Input Filename: F1XN4_2_M51_C...
Input Purpose: Identification
Accepted:

Light NeuCode Component ID: 8_132
Heavy NeuCode Component ID: 8_421
overlapping_charge: 3.4406
Intensity Ratio: 519496.4772
Intensity Sum Overlapping Charge States: 26
Lysine Count: 7966.09560326
Corrected NeuCode: 7952.08666526
Light Weighted Monoisotopic Mass: 97.217041
Heavy Weighted Monoisotopic Mass: 97.217041
Light Apex RT: 1180-1181
Heavy Apex RT: 1180-1181
Scan Range: 8
Input File Unique ID: F1XN4_2_M51_C...
Input Filename: F1XN4_2_M51_C...
Input Purpose: Identification
Accepted:

Light NeuCode Component ID: 8_165
Heavy NeuCode Component ID: 8_364
overlapping_charge: 3.8239
Intensity Ratio: 4189516.0601
Intensity Sum Overlapping Charge States: 26
Lysine Count: 7966.07891313
Corrected NeuCode: 7952.0866901313
Light Weighted Monoisotopic Mass: 97.371285
Heavy Weighted Monoisotopic Mass: 97.371285
Light Apex RT: 1181-1182
Heavy Apex RT: 1181-1182
Scan Range: 8
Input File Unique ID: F1XN4_2_M51_C...
Input Filename: F1XN4_2_M51_C...
Input Purpose: Identification
Accepted:

Light NeuCode Component ID: 8_424
Heavy NeuCode Component ID: 8_181
overlapping_charge: 4.5105
Intensity Ratio: 1665470.1109
Intensity Sum Overlapping Charge States: 26
Lysine Count: 7966.10093052
Corrected NeuCode: 7952.09173052
Light Weighted Monoisotopic Mass: 97.5771415
Heavy Weighted Monoisotopic Mass: 97.5771415
Light Apex RT: 1187-1188
Heavy Apex RT: 1187-1188
Scan Range: 8
Input File Unique ID: F1XN4_2_M51_C...
Input Filename: F1XN4_2_M51_C...
Input Purpose: Identification
Accepted:

- See Raw Experimental Components table in **Raw Experimental Components** section (unlabeled) or NeuCode Pairs table in **NeuCode Pairs** section (NeuCode labeled) for a description of each column

11 Experiment Theoretical Comparison

11.1 Overview

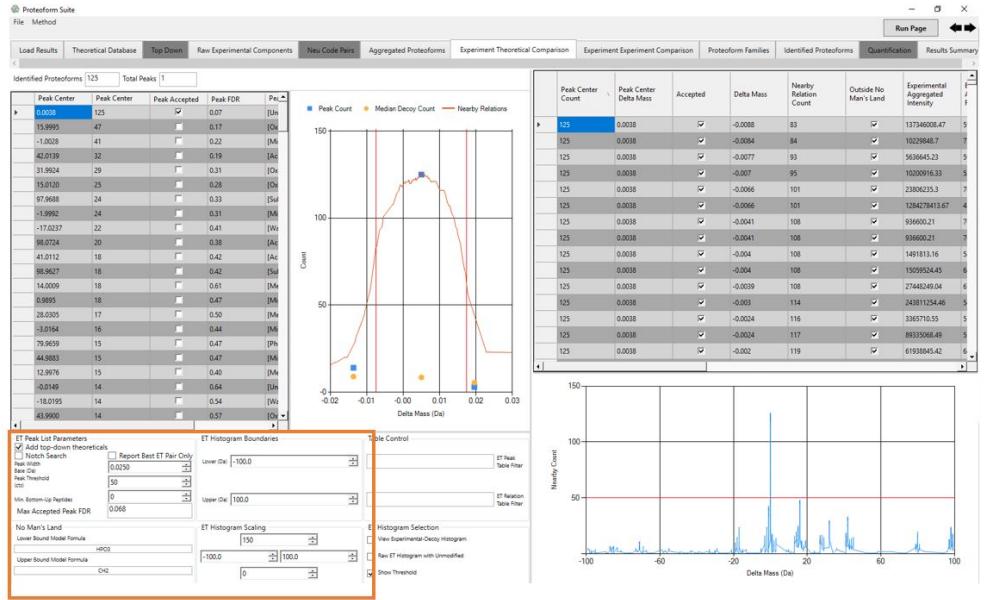
On this page, experimental proteoforms are compared to theoretical proteoforms from the theoretical database, generating a list of experiment-theoretical pairs. Each experiment-theoretical pair has a mass difference between the experimental proteoform and the theoretical proteoform in the pair; pairs are generated for mass differences that correspond to a known set of modifications. Each experimental proteoform can be in a pair with one theoretical proteoform per protein; heuristics are used to determine the most likely pair based on the delta mass differences. A histogram is generated of the mass differences for all experiment-theoretical pairs; experiment-theoretical pairs in accepted delta mass peaks are used to construct proteoform families. Experiment-decoy pairs are generated by comparing the experimental proteoforms to the decoy databases; these pairs are used to estimate a false discovery rate for each delta mass peak.

11.2 Run Page

- The Theoretical Proteoforms and Aggregated Proteoforms pages must be run before running this page.
- Set all parameters as desired for current analysis (see below)
- Click Run Page button (top right)
- Browse the list of delta mass peaks from the histogram of experiment-theoretical pairs delta masses (top left table). Accept peaks that have an acceptable false discovery rate and correspond to common/likely modifications. For unlabeled analyses, typically only the delta mass peak closest to 0 (exact matches) is accepted.

11.3 Set Parameters

- Add top-down theoreticals: if checked, theoretical proteoforms that were supplemented to the database due to the presence of a top-down proteoforms will be included in the experiment-theoretical comparison
- Notch Search: if checked, a notch search will be performed. For each modification, experiment-theoretical pairs will be generated if the delta mass is within the set tolerance from the modification's delta mass
- Report Best ET Pair Only: if checked, only the closest matching experiment-theoretical pair for each experimental proteoform will be included (each experimental proteoform will only be able to belong to one experiment-theoretical pair)
- Peak Width Base (Da): if notch search is unchecked, this is the size of bins used for generating the delta mass histogram from experiment-theoretical pair delta masses
- Peak Threshold (cts): the minimum number of experiment-theoretical pairs that must belong to a delta mass peak for the peak to be accepted

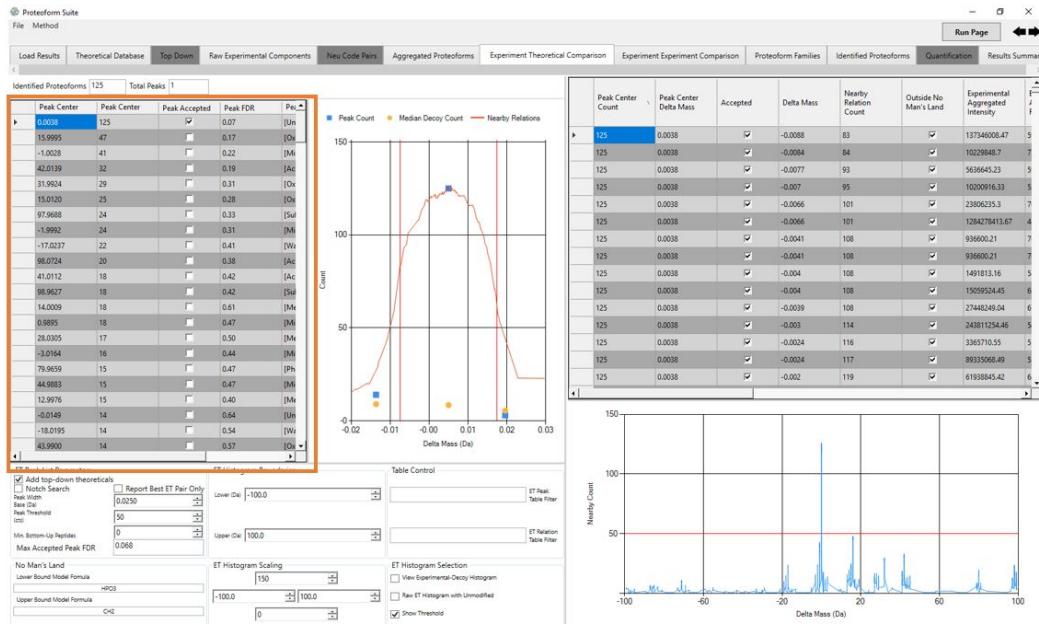


- Min. Bottom-Up Peptides: this is the minimum number of bottom-up peptides that a theoretical proteoform must have in order to be included in the experiment-theoretical comparison. Top-down identified proteoforms are also included in the database.
- Notch Tolerance: if a notch search is performed, this tolerance will be used to generate experiment-theoretical pairs at each modification delta mass
- ET Histogram Boundaries: Lower (Da) and Upper (Da) delta masses to be considered for an experiment-theoretical pair to be generated

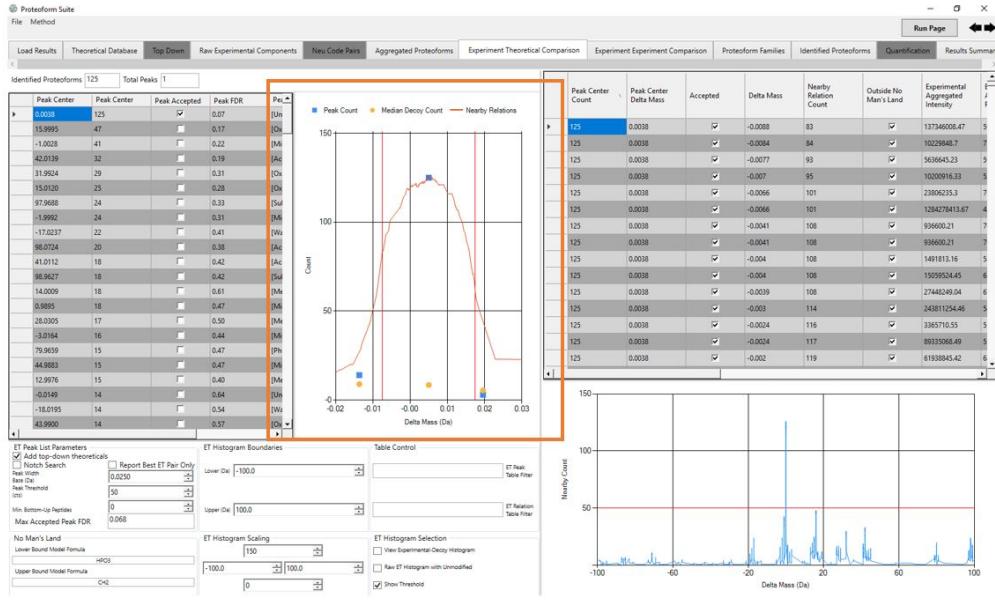
11.4 Results

- Identified Proteoforms: the number of accepted experiment-theoretical pairs
- Total Peaks: the number of accepted experiment-theoretical delta mass peaks

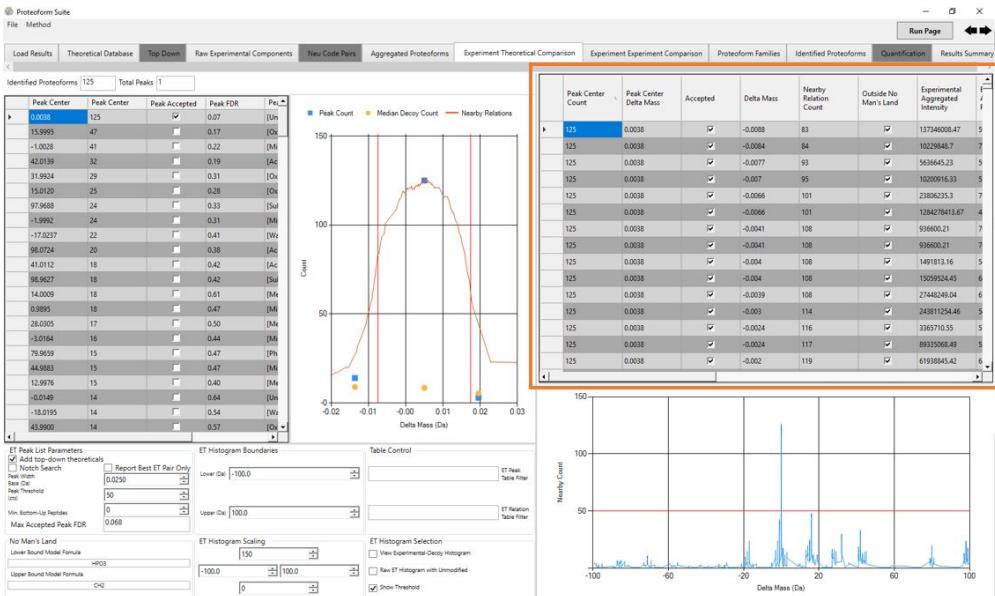
- Experiment-Theoretical Delta Mass Peaks table: the top left table displays the delta mass peaks from the histogram of experiment-theoretical pair delta masses. If a notch search is performed, each peak is a different modification delta mass/notch



- Peak Center Delta Mass: delta mass at the center of this delta mass peak
- Peak Center Count: the number of experiment-theoretical comparisons delta masses that are part of this peak
- Peak Accepted: checked if peak is accepted (peak center count is above peak threshold in set parameters or manually changed by user). This check box can be checked or unchecked to accept or unaccept a delta mass peak
- Peak FDR: the false discovery rate for this delta mass peak; determined based on the average number of experiment-decoy pairs that fall within this peak delta mass plus/minus half of the peak width base
- Peak Assignment Possibilities: modifications/combinations of modifications that could correspond to the delta mass of this delta mass peak
- Mass Shifter: set this number to a positive or negative integer and rerun this page; the monoisotopic mass of experimental proteoforms in experiment-theoretical pairs in this peak will be shifted by the number of monoisotopic mass units of this column; useful if a peak is at a missed monoisotopic value from 0 or a modification delta mass (ex: -1 or +1 Da)
- Experiment-Theoretical Delta Mass Peak Zoom-in Graph: when a peak is selected in the Experiment-Theoretical Delta Mass Peak table, this top left graph displays a zoom-in of the peak from the Experiment-Theoretical Delta Mass Histogram (bottom right graph). Blue square point is the number of experiment-theoretical pairs in the peak, and yellow circle point is median number of experiment-decoy pairs in this peak. The red line plots the nearby relations histogram count for each delta mass value



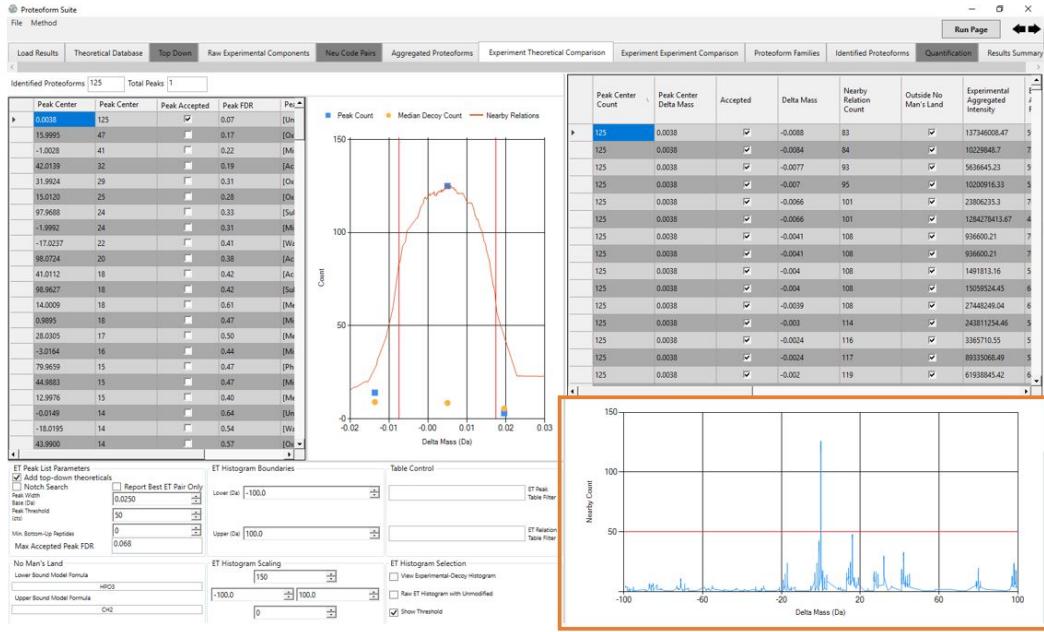
- Experiment-Theoretical Pairs: this right table displays all experiment-theoretical pairs, consisting of a theoretical proteoform, an experimental proteoform, and their mass difference.



- Peak Center Count: if this pair is in a delta mass peak, the number of pairs in the peak
- Peak Center Delta Mass: if this pair is in a delta mass peak, the delta mass at the center of the peak
- Accepted: checked if this pair is in a delta mass peak that is accepted in the Experiment-Theoretical Delta Mass Peaks table

- Delta Mass: mass difference between the experimental and theoretical proteoform in this pair
- Nearby Relation Count: number of pairs with a delta mass close to this pair’s delta mass; this value is used to plot the delta mass histogram
- Outside No Man’s Land: checked if this pair is an acceptable delta mass regarding the numbers after the decimal point. Pairs with a delta mass in no man’s land are not joined into delta mass peaks
- Experimental Aggregated Intensity: sum of intensity of aggregated raw experimental components for the experimental proteoform in this pair
- Experimental Aggregated RT: retention time of experimental proteoform in this pair
- Number Experimental Observations: number of aggregated raw experimental components (unlabeled) or NeuCode pairs (NeuCode labeled) for the experimental proteoform in this pair. If top-down proteoform, the number of top-down hits
- Experimental Aggregated Proteoform Mass: mass of experimental proteoform in this pair
- Experimental Accession: unique ID given by Proteoform Suite for experimental proteoform in this pair
- Abundant Exp. Component for Manual Validation: file information for the most abundant raw experimental component aggregated into the experimental proteoform of this pair
- Theoretical Proteoform Mass: mass of theoretical proteoform in this pair
- Accession: accession given by Proteoform Suite for the theoretical proteoform in this pair
- Name: protein name from UniProt for the theoretical proteoform in this pair
- Fragment: sequence description for the theoretical proteoform in this pair
- Description: protein description from UniProt for the theoretical proteoform in this pair
- PTM Description: PTMs on theoretical proteoform in this pair

- Experiment-Theoretical Delta Mass Histogram graph: this bottom right graph shows a histogram of the delta masses for the experiment-theoretical pairs



- View Experiment-Decoy Histogram: if checked, delta mass histogram for experiment-decoy pairs will be plotted
- Raw ET Histogram with Unmodified: if checked, a histogram with only unmodified theoretical proteoforms will be plotted using all delta mass pairs
- Show Threshold: if checked, a red line on the graph shows the Peak Threshold set

12 Experiment Experiment Comparison

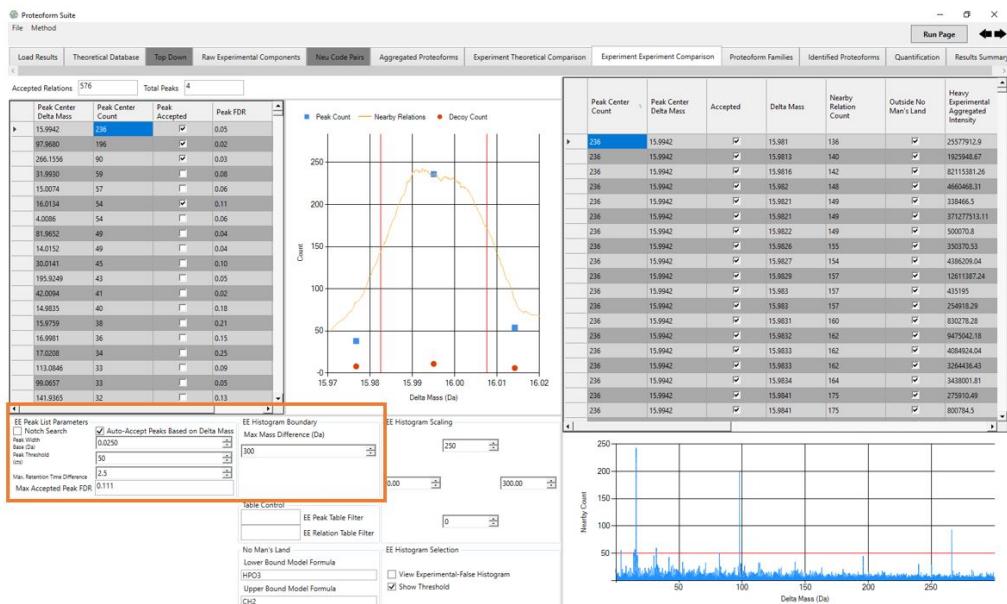
12.1 Overview

On this page, experimental proteoforms are compared to one another, generating a list of experiment-experiment pairs. Each experiment-experiment pair has a mass difference between the two experimental proteoforms in the pair. A histogram is generated of the mass differences for all experiment-experiment pairs; experiment-experiment pairs in accepted delta mass peaks are used to construct proteoform families. Experiment-false pairs are generated from experimental proteoforms with a different number of lysines (NeuCode labeled) or from proteoforms eluting at a different retention times (unlabeled); these pairs are used to estimate a false discovery rate for each delta mass peak.

12.2 Run Page

- The Theoretical Proteoforms and Aggregated Proteoforms pages must be run before running this page.
- Set all parameters as desired for current analysis (see below)
- Click Run Page button (top right)
- Browse the list of delta mass peaks from the histogram of experiment-experiment pairs delta masses (top left table). Accept peaks that have an acceptable false discovery rate and correspond to common/likely modifications.

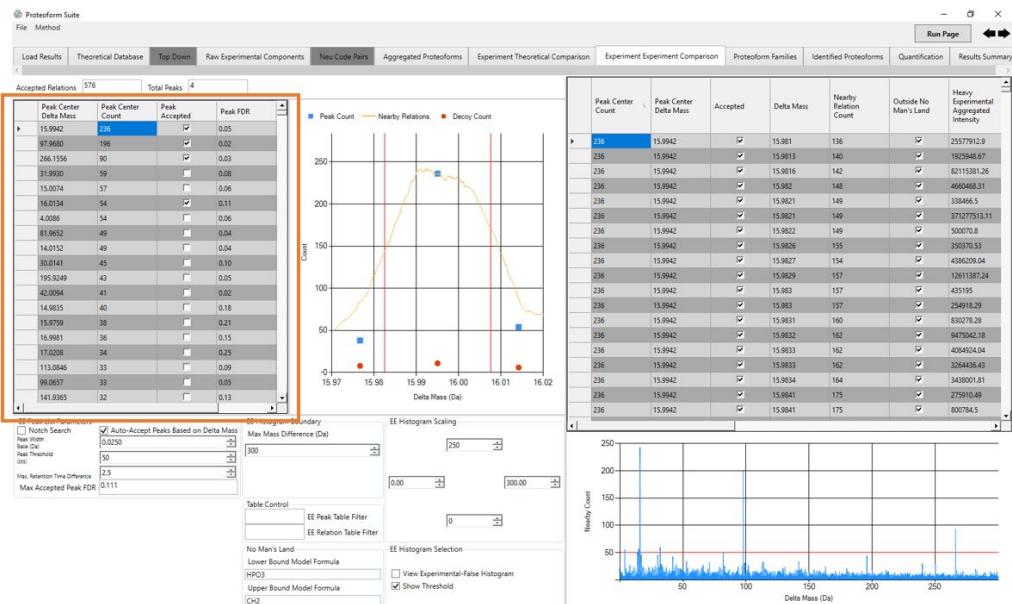
12.3 Set Parameters



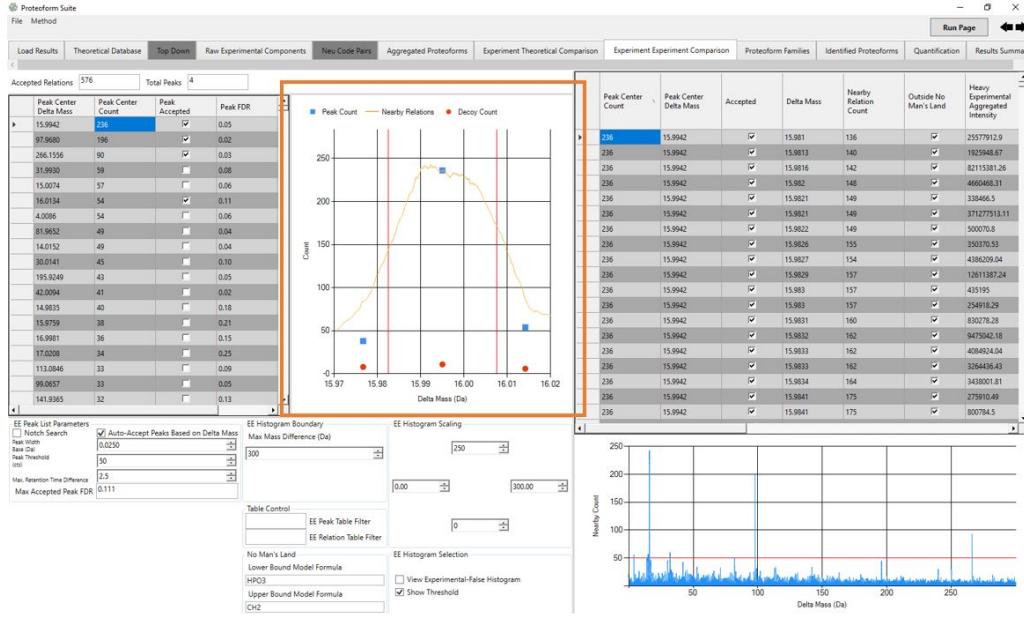
- Notch Search: if checked, a notch search will be performed. For each modification, experiment-experiment pairs will be generated if the delta mass is within the set tolerance from the modification's delta mass
- Auto-Accept Peaks Based on Delta Mass: if checked, peaks with a count above the Peak Threshold will be accepted if they correspond to a common modification
- Peak Width Base (Da): if notch search is unchecked, this is the size of bins used for generating the delta mass histogram from experiment-experiment pair delta masses
- Peak Threshold (cts): the minimum number of experiment-experiment pairs that must belong to a delta mass peak for the peak to be accepted
- Max. Retention Time Difference: maximum allowed retention time difference for two experimental proteoforms to be eligible to be in an experiment-experiment pair
- Notch Tolerance: if a notch search is performed, this tolerance will be used to generate experiment-experiment pairs at each modification delta mass
- EE Histogram Boundary: maximum delta masses to be considered for an experiment-experiment pair to be generated

12.4 Results

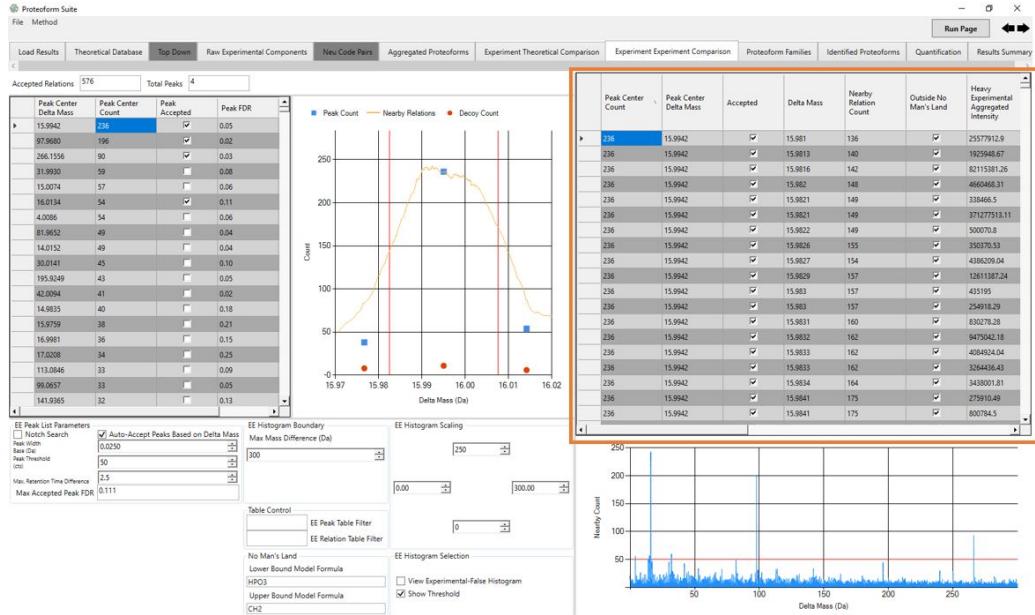
- Accepted Relations: the number of accepted experiment-experiment pairs
- Total Peaks: the number of accepted experiment-experiment delta mass peaks
- Experiment-Experiment Delta Mass Peaks table: the top left table displays the delta mass peaks from the histogram of experiment-experiment pair delta masses. If a notch search is performed, each peak is a different modification delta mass/notch



- Peak Center Delta Mass: delta mass at the center of this delta mass peak
- Peak Center Count: the number of experiment-experiment comparisons delta masses that are part of this peak
- Peak Accepted: checked if peak is accepted (peak center count is above peak threshold in set parameters or manually changed by user). This check box can be checked or unchecked to accept or unaccept a delta mass peak
- Peak FDR: the false discovery rate for this delta mass peak; determined based on the number of experiment-false pairs that fall within this peak delta mass plus/minus half of the peak width base
- Peak Assignment Possibilities: modifications/combinations of modifications that could correspond to the delta mass of this delta mass peak
- Experiment-Experiment Delta Mass Peak Zoom-in Graph: when a peak is selected in the Experiment-Experiment Delta Mass Peak table, this top left graph displays a zoom-in of the peak from the Experiment-Experiment Delta Mass Histogram (bottom right graph). Blue square point is peak count, and yellow circle point is median decoy count for this peak. The red line plots the nearby relations histogram count for each delta mass value

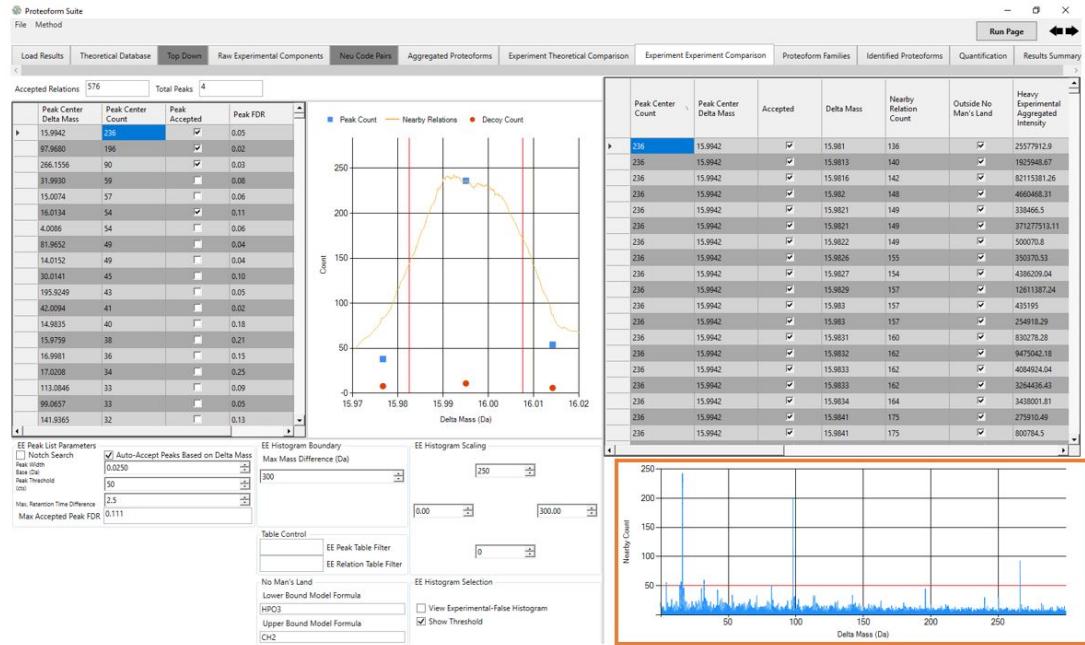


- Experiment-Experiment Pairs: this right table displays all experiment-experiment pairs, consisting of two experimental proteoforms and their mass difference.



- Peak Center Count: if this pair is in a delta mass peak, the number of pairs in the peak
- Peak Center Delta Mass: if this pair is in a delta mass peak, the delta mass at the center of the peak
- Accepted: checked if this pair is in a delta mass peak that is accepted in the Experiment-Experiment Delta Mass Peaks table
- Delta Mass: mass difference between the two experimental proteoforms in this pair
- Nearby Relation Count: number of pairs with a delta mass close to this pair’s delta mass; this value is used to plot the delta mass histogram
- Outside No Man’s Land: checked if this pair is an acceptable delta mass regarding the numbers after the decimal point. Pairs with a delta mass in no man’s land are not joined into delta mass peaks
- Heavy Experimental Aggregated Intensity: sum of intensity of aggregated raw experimental components for the heavier experimental proteoform in this pair
- Aggregated RT Heavy: retention time of heavier experimental proteoform in this pair
- Number Heavy Experimental Observations: number of aggregated raw experimental components (unlabeled) or NeuCode pairs (NeuCode labeled) for the heavier experimental proteoform in this pair. If top-down proteoform, the number of top-down hits
- Heavy Experimental Aggregated Proteoform Mass: mass of the heavier experimental proteoform in this pair
- Heavy Experimental Accession: unique ID given by Proteoform Suite for the heavier experimental proteoform in this pair

- Heavy Abundant Exp. Component for Manual Validation: file information for the most abundant raw experimental component aggregated into the heavier experimental proteoform of this pair
- Light Experimental Aggregated Intensity: sum of intensity of aggregated raw experimental components for the lighter experimental proteoform in this pair
- Aggregated RT Light: retention time of lighter experimental proteoform in this pair
- Number Light Experimental Observations: number of aggregated raw experimental components (unlabeled) or NeuCode pairs (NeuCode labeled) for the lighter experimental proteoform in this pair. If top-down proteoform, the number of top-down hits
- Light Experimental Aggregated Proteoform Mass: mass of the lighter experimental proteoform in this pair
- Light Experimental Accession: unique ID given by Proteoform Suite for the lighter experimental proteoform in this pair
- Light Abundant Exp. Component for Manual Validation: file information for the most abundant raw experimental component aggregated into the lighter experimental proteoform of this pair
- Experiment-Experiment Delta Mass Histogram graph: this bottom right graph shows a histogram of the delta masses for the experiment-experiment pairs



- View Experiment-False Histogram: if checked, delta mass histogram for experiment false pairs will be plotted
- Show Threshold: if checked, a red line on the graph shows the Peak Threshold set

13 Proteoform Families

13.1 Overview

On this page, accepted experiment-theoretical and experiment-experiment pairs are joined to construct proteoform families. Experimental proteoforms are identified by intact-mass analysis; beginning with each theoretical proteoform in each family, connections between proteoforms are traced to identify proteoforms first from experiment-theoretical pairs and then from subsequent experiment-experiment pairs. Decoy proteoform families are constructed from experiment-decoy and experiment-false pairs and are used to calculate a global false discovery rate for intact-mass proteoform identifications. A script for Cytoscape^{9,10} can be exported to visualize proteoform families as a network of nodes (proteoforms masses) and edges (mass differences between proteoforms).

13.2 Run Page

- The Experiment-Theoretical Comparison and Experiment-Experiment Comparison pages must be run before running this page.
 - Set all parameters as desired for current analysis (see below)
 - Click Run Page button (top right)

13.3 Set Parameters

Load Results	Theoretical Database	Top Down	Raw Experimental Components	Neu Code Pairs	Aggregated Proteiforms	Experiment Theoretical Comparison	Experiment Experiment Comparison	Proteiform Families	Identified Proteiforms	Quantification	Results Summary
Family ID	Experimental	Experimental	Theoretical	Top-Down	Gene Count	Theoretical	Theoretical	Gene Names	Relat...		
22	16	E5; E21; E7; E2...	620652; 62225...	1	0	1	P04650_2f0l15...	RL39_YEAST	YIL189W; YOR...	32	
17	13	E0; E180; E24...	1201553; 12113...	4	0	2	P05744_2f0l107...	RL33A_YEAST	YPL143W; YOR...	19	
281	13	E22; E285; E1...	8894; 8890; 95...	0	0	0					
24	11	E16; E20; E50...	109972; 11095...	1	0	1	O14455_2f0l1...	RL36B_YEAST	YPL249C-A	15	
192	12	E17; E371; E1...	15373.16; 1538...	0	0	0					
21	8	E8; E47; E24; E...	144853; 14501...	2	0	2	P0C0W1_2f0l1...	RS22A_YEAST	YIL190C; YLR3...	14	
138	10	E17; E281; E1...	15895.02; 15911...	0	0	0				13	
66	9	E59; E150; E662...	1427278; 1438...	2	0	1	P40046_2f0l120...	YTC1_YEAST	YER072W	12	
494	8	E354; E1199; E3...	10844.66; 10802...	1	0	1	Q96VH3_2f0l97...	MIC10_YEAST	YCL057C-A	11	
20	6	E2; E14; E11; E...	6981.02; 7080.9...	1	0	1	P0C34_2f0l63...	RS20D_YEAST	YOR182C	10	
161	10	E172; E281; E1...	13634.29; 14842...	0	0	0				10	
466	8	E519; F904; F19...	14601.26; 14617...	1	0	1	P13331_2f0l175...	NTE7_YEAST	YPR099W	10	
Experimental	Aggregated	Description	Gene Name	Gene ID	Aggregated RT	Aggregated	Top-Down	Sequence	Manually	Aggregat...	
E5	62065297	605 ribosomal ...	YIL189W	853250	44.97	2404854176.9448	<input type="checkbox"/>	AAQKSFRKIQ...	<input type="checkbox"/>	38	
E21	62225241	605 ribosomal ...	YIL189W	853250	44.62	1379969003.1432	<input type="checkbox"/>	AAQKSFRKIQ...	<input type="checkbox"/>	23	
E71	6238.5197	605 ribosomal ...	YIL189W	853250	44.42	339076738.6866	<input type="checkbox"/>	AAQKSFRKIQ...	<input type="checkbox"/>	17	
E202	6093.4334	605 ribosomal ...	YIL189W	853250	44.41	79822219.4676	<input type="checkbox"/>	AAQKSFRKIQ...	<input type="checkbox"/>	8	
E704	6254.5150	605 ribosomal ...	YIL189W	853250	44.35	1207525.9068	<input type="checkbox"/>	AAQKSFRKIQ...	<input type="checkbox"/>	7	
E24	6304.4784	605 ribosomal ...	YIL189W	853250	44.97	870891022.2363	<input type="checkbox"/>	AAQKSFRKIQ...	<input type="checkbox"/>	11	
E181	6109.4340	605 ribosomal ...	YIL189W	853250	44.25	83492305.3769	<input type="checkbox"/>	AAQKSFRKIQ...	<input type="checkbox"/>	6	
E361	6336.4816	605 ribosomal ...	YIL189W	853250	44.39	56929748.7466	<input type="checkbox"/>	AAQKSFRKIQ...	<input type="checkbox"/>	8	
E106	6320.4791	605 ribosomal ...	YIL189W	853250	44.64	259284576.7770	<input type="checkbox"/>	AAQKSFRKIQ...	<input type="checkbox"/>	15	
E486	6125.4377	605 ribosomal ...	YIL189W	853250	44.10	2395780.4709	<input type="checkbox"/>	AAQKSFRKIQ...	<input type="checkbox"/>	8	
E692	6191.4049	605 ribosomal ...	YIL189W	853250	44.41	14297834.5183	<input type="checkbox"/>	AAQKSFRKIQ...	<input type="checkbox"/>	5	
E162	6402.4449	605 ribosomal ...	YIL189W	853250	44.96	98462029.3908	<input type="checkbox"/>	AAQKSFRKIQ...	<input type="checkbox"/>	10	
E633	6207.4132	605 ribosomal ...	YIL189W	853250	44.24	1251597.6373	<input type="checkbox"/>	AAQKSFRKIQ...	<input type="checkbox"/>	4	
E424	6398.4667	605 ribosomal ...	YIL189W	853250	44.95	3016968.4339	<input type="checkbox"/>	AAQKSFRKIQ...	<input type="checkbox"/>	6	
E830	6148.4460	605 ribosomal ...	YIL189W	853250	44.62	6239405.7759	<input type="checkbox"/>	AAQKSFRKIQ...	<input type="checkbox"/>	3	
E1482	6223.4131	605 ribosomal ...	YIL189W	853250	44.09	2216807.4290	<input type="checkbox"/>	AAQKSFRKIQ...	<input type="checkbox"/>	3	

- Folder for Family Build: folder to build scripts to visualize proteoform families in Cytoscape

- Most Recent Time Stamp: time stamp that will be in the filename of scripts to visualize proteoform families in Cytoscape
- Decimal Rounding for Labels: number of decimal places to round labels in visualized proteoform families in Cytoscape
- Only Assign Common/Known Mods: if checked, intact-mass identifications will only be made for common modifications or for modifications annotated for that theoretical protein in UniProt
- Count Adducts as Identifications: if checked, adducts (oxidation, sulfate adducts, SDS adducts) will be counted as unique identifications
- Build Gene-Centric Families: if checked, all proteoforms connected to theoretical proteoforms from the same gene will be grouped into the same proteoform family
- Build as Quantitative Families: if checked and if quantification was performed, quantitative proteoform families will be built (with pie charts for each experimental proteoform showing abundance ratios)
- Identify from Top-Down Nodes: if checked, top-down proteoforms in addition to theoretical proteoforms will be used as starting points for identification in proteoform families. This requires high-quality top-down identifications to prevent high false discovery rate
- Remove Bad Connections: if checked, connections between proteoforms in identified families that do not lead to identification will be removed, unaccepting the experiment-theoretical or experiment-experiment pairs
- Use Top-Down IDs to Reduce Ambiguity: if checked, top-down IDs will be used to reduce ambiguity in intact-mass identifications; identifications from theoretical proteoforms confirmed by top-down will be prioritized
- Use Annotated PTMs to Reduce Ambiguity: if checked, annotated PTMs in UniProt will be used to reduce ambiguity in intact-mass identifications; identifications with an annotated PTM will be prioritized
- Use ppm tolerance: if checked, this ppm tolerance will be used when making intact-mass identifications in proteoform families
- Highlights for Significant Differences: if checked, a red node border and bold label will be used in quantitative proteoform families to highlight experimental proteoforms with statistically significant quantitative differences

13.4 Results

- Proteoform Families table: the top table displays all constructed proteoform families

The screenshot shows the Proteoform Suite software interface. At the top, there is a menu bar with 'File' and 'Method'. Below the menu is a toolbar with buttons for 'Load Results', 'Theoretical Database', 'Top Down', 'Raw Experimental Components', 'New Code Pairs', 'Aggregated Proteoforms', 'Experiment Theoretical Comparison', 'Experiment Experiment Comparison', 'Proteoform Families', 'Identified Proteoforms', 'Quantification', and 'Results Summary'. A 'Run Page' button is also present.

The main area contains a table titled 'Proteoform Families' with the following columns:

Family ID	Experimental	Experimental	Theoretical	Top-Down	Gene Count	Theoretical	Theoretical	Gene Names	Relation	
22	E5; E21; E2...	E5; E21; E2...	6206;53;6223...	1	0	1	P04630_2full51...	R139_YEAST	32	
17	E0; E189; E214...	E0; E189; E214...	12015;53;1213...	4	0	2	P0744_2full107...	R133A_YEAST...	YPL145W_YOR...	19
281	E222; E225; E1...	E222; E225; E1...	8594;8;86095...	0	0	0				
24	E18; E20; E250...	E18; E20; E250...	10997;28;11095...	1	0	1	O14455_2full10...	R136B_YEAST...	VPL249C_A	15
192	E157; E37; E16...	E157; E37; E16...	15373;16;15389...	0	0	0				
21	E8; E47; E243; E...	E8; E47; E243; E...	14485;82;14501...	2	0	2	P0C1W1_2full...	RS22A_YEAST...	YL190C_YLR3...	14
138	E172; E201; E13...	E172; E201; E13...	15895;02;15911...	0	0	0				
66	E59; E19; E662...	E59; E19; E662...	14277;78;14288...	2	0	1	P40048_2full...	VTC1_YEAST	YR072W	12
494	E354; E119; E3...	E354; E119; E3...	10844;66;10862...	1	0	1	Q96VH5_2full97...	MIC10_YEAST	YCL057C_A	11
20	E2; E14; E13; E1...	E2; E14; E13; E1...	6981;02;7009...	1	0	1	P0CK34_2full63...	RS30B_YEAST...	YOR182C	10
161	E172; E10; E91...	E172; E10; E91...	1334;29;13462...	0	0	0				
446	E133; E946; E19...	E133; E946; E19...	14601;34;14617...	1	0	1	B13231_2full17...	NTE2_YEAST	YER000W	10

On the right side of the interface, there are several filter and search panels:

- Code Selection: Proteoform Families, R139_YEAST
- Code Filter: Mass-Based Spiral
- Code Layout: On Node
- Code Label Positioning: Inferred Theoretical ID
- Code Label Informing: Modification IDs (omits edges with null IDs)
- Code Label Information: Ordered Locus, e.g. YLR13W
- Preferred Gene Label: YLR13W
- Folder for Family Build: Browse
- Most Recent Time Stamp: 2023-09-11T14:27:45Z
- Decimal Rounding for Labels: 2
- Only Assign Common Known Mods: checked
- Count Adds as Identifications: checked
- Build Gene-Centric Families: checked
- Build as Quantitative Families: unchecked
- Identify from Top-Down Nodes: unchecked
- Remove Bad Connections: unchecked
- Use Top-Down IDs to Reduce Ambiguity: unchecked
- Use Annotated PTMs to Reduce Ambiguity: unchecked
- Use ppm tolerance: 10.0
- Highlights for Significant Differences: Red Node Border, Bold Label
- Build All Families in Cytoscape
- Build Selected Families in Cytoscape

At the bottom, there are two status bars:

- Proteoform Families: 599 Identified Families (Correspond to 1 gene), 253 Experimental Proteoforms in Identified Families, 3 Ambiguous Families (Correspond to > 1 gene), 23 Experimental Proteoforms in Ambiguous Families, 295 Unidentified Families (Correspond to no gene), 788 Experimental Proteoforms in Unidentified Families, 1328 Orphaned Experimental Proteoforms (Inact-mass proteoforms not joined with another proteoform)
- Raw Experimental Components in Families: 13053 Raw Experimental Components in Families, 63.6 % of Raw Experimental Components in Families

- Family ID: unique ID assigned by Proteoform Suite for this proteoform family
- Experimental Proteoforms: number of experimental proteoforms in this proteoform family
- Experimental Accessions: semi-colon separated list of accessions for experimental proteoforms in this proteoform family
- Experimental Aggregated Masses: semi-colon separated list of masses for experimental proteoforms in this proteoform family
- Theoretical Proteoforms: number of theoretical proteoforms in this proteoform family
- Top-Down Proteoforms: number of top-down proteoforms in this proteoform family
- Gene Count: number of genes in this proteoform family. Families with more than 1 gene are ambiguous and families with no genes are unidentified
- Theoretical Accessions: semi-colon separated list of accessions for theoretical proteoforms in this proteoform family
- Theoretical Names: semi-colon separated list of protein names for this proteoform family
- Gene Names: semi-colon separated list of gene names in this proteoform family
- Relation Count: number of proteoform relations in this family (experiment-theoretical pairs and experiment-experiment pairs)
- Proteoforms table: the bottom table displays proteoforms for the proteoform family selected in the Proteoform Families table.

The screenshot shows the Proteoform Suite application window. At the top, there's a menu bar with 'Proteoform Suite' and 'File Method'. Below the menu is a toolbar with buttons for 'Load Results', 'Theoretical Database', 'Top Down', 'Raw Experimental Components', 'New Code Pairs', 'Aggregated Proteoforms', 'Experiment-Theoretical Comparison', 'Experiment-Experiment Comparison', 'Proteoform Families', 'Identified Proteoforms', 'Quantification', and 'Results Summary'. A 'Run Page' button is also present.

The main area contains two tables. The first table, titled 'Experimental Proteoforms', has columns: Family ID, Experimental, Experimental, Theoretical, Top-Down, Gene Count, Theoretical, Theoretical, Gene Names, Relation. The second table, titled 'Aggregated Proteoforms', has columns: Experimental, Aggregated, Description, Gene Name, GeneID, Aggregated RT, Aggregated, Top-Down, Sequence, Manually, Aggregate. Both tables have rows corresponding to various family IDs (e.g., 22, 17, 281, 24, 11, 21, 12, 138, 66, 9, 494, 20, 161, 646) and experimental details.

To the right of the tables are several configuration panels:

- Table Selection:** Set to 'Proteoform Families'.
- Table Filter:** Set to 'Mass-Based Spiral'.
- Node Layout:** Set to 'On Node'.
- Node Label Positioning:** Set to 'Inferred Theoretical ID'.
- Edge Label Information:** Set to 'Modification IDs (omits edges with null IDs)'.
- Prefered Gene Label:** Set to 'Ordered Locus, e.g. YLR113W'.

Below these are sections for 'Build for Family Build' (with a 'Browse' button), 'Build All Families in Cytoscape', and 'Build Selected Families in Cytoscape'. There are also checkboxes for 'Only Assign Common Known Mods', 'Count Adds as Identifications', 'Build Gene-Centric Families', 'Build as Quantitative Families', 'Identify from Top-Down Nodes', 'Remove Bad Connections', 'Use Top-Down IDs to Reduce Ambiguity', 'Use Annotated PTMs to Reduce Ambiguity', and 'Use ppm tolerance [10.0]'. A 'Highlights for Significant Differences' section includes 'Red Node Border' and 'Bold Label' checkboxes.

At the bottom, there's a status bar with '99 Identified Families (Correspond to 1 gene)', '53 Experimental Proteoforms in Identified Families', 'Ambiguous Families (Correspond to > 1 gene)', '3 Unidentified Families (Correspond to no gene)', '88 Experimental Proteoforms in Unidentified Families', '328 Orphaned Experimental Proteoforms (Intact-mass proteoforms not joined with another reference)', '3053 Raw Experimental Components in Families', and '3.6 % of Raw Experimental Components in Families'.

- If the Experimental Proteoforms, Experimental Accessions, or Experimental Aggregated Masses columns are selected, the experimental proteoforms for the selected family will be displayed. See the Aggregated Proteoforms table in the **Aggregated Proteoforms** section for column descriptions.
- If the Theoretical Proteoforms, Theoretical Accessions, or Theoretical Names Columns are selected, the theoretical proteoforms for the selected family will be displayed. See the Theoretical Proteoforms table in the **Theoretical Proteoforms** section for column descriptions.
- If the Top-Down Proteoforms column is selected, the top-down proteoforms for the selected family will be displayed. See the Top-Down Proteoforms table in the **Top-Down Proteoforms** section for column descriptions.
- If the Relation Count column is selected, the proteoform relations in the selected family will be displayed. See the Experiment-Theoretical Pairs table in the **Experiment-Theoretical Comparison** section and the Experiment-Experiment Pairs table in the **Experiment-Experiment Comparison** section for column descriptions.
- Table Selection: this drop-down box changes which proteoform families are displayed (target or decoy communities). Can also observe theoretical proteoforms and GO Terms
- Table Filter: filter the Proteoform Families table (top left) by any entered text
- Node Layout: this drop-down box changes the node layout in visualized proteoform families in Cytoscape
- Node Label Positioning: this drop-down box changes the position of node labels in visualized proteoform families in Cytoscape

- Node Label Informing: this drop-down box changes the node label information in visualized proteoform families in Cytoscape
- Edge Label Information: this drop-down box changes the edge label information in visualized proteoform families in Cytoscape
- Preferred Gene Label: this drop-down box changes the gene label information in visualized proteoform families in Cytoscape
- Build All Families in Cytoscape: exports scripts for Cytoscape to visualize all proteoform families
- Build Selected Families in Cytoscape: exports scripts for Cytoscape to visualize proteoform families selected in the Proteoform Families table
- The text box at the bottom right of the page displays information about the results. See the **Results Summary** section for a description of each result

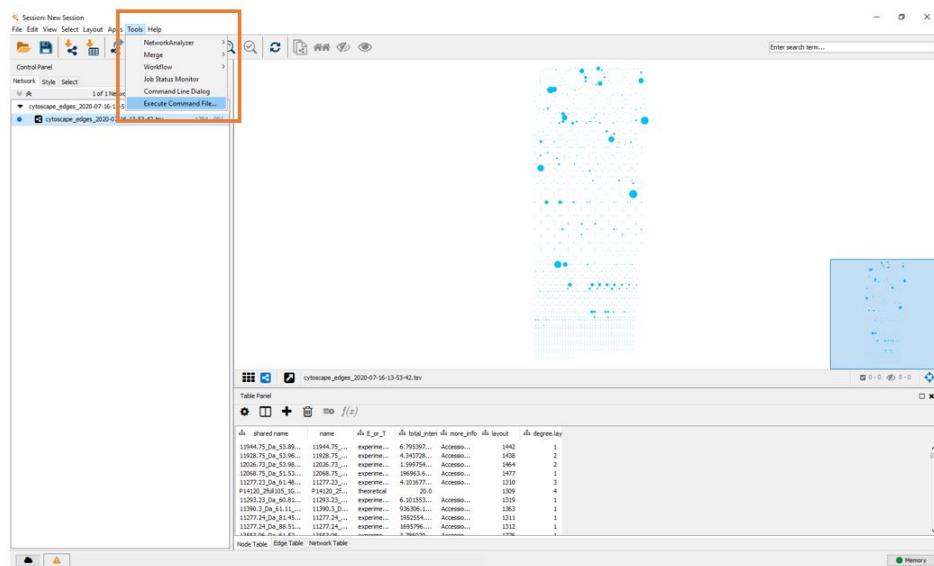
14 Visualizing Proteoform Families in Cytoscape

14.1 Overview

Proteoform families are visualized in the software program Cytoscape.^{9,10} Each node is a proteoform and each edge represents a mass difference between proteoforms, corresponding to a modification or amino acid difference. Proteoform family visualization allows users to visualize all observed proteoforms and modification combinations from each family in a simple graphic. Install Cytoscape version 3.5.0: <https://cytoscape.org/download.html>.

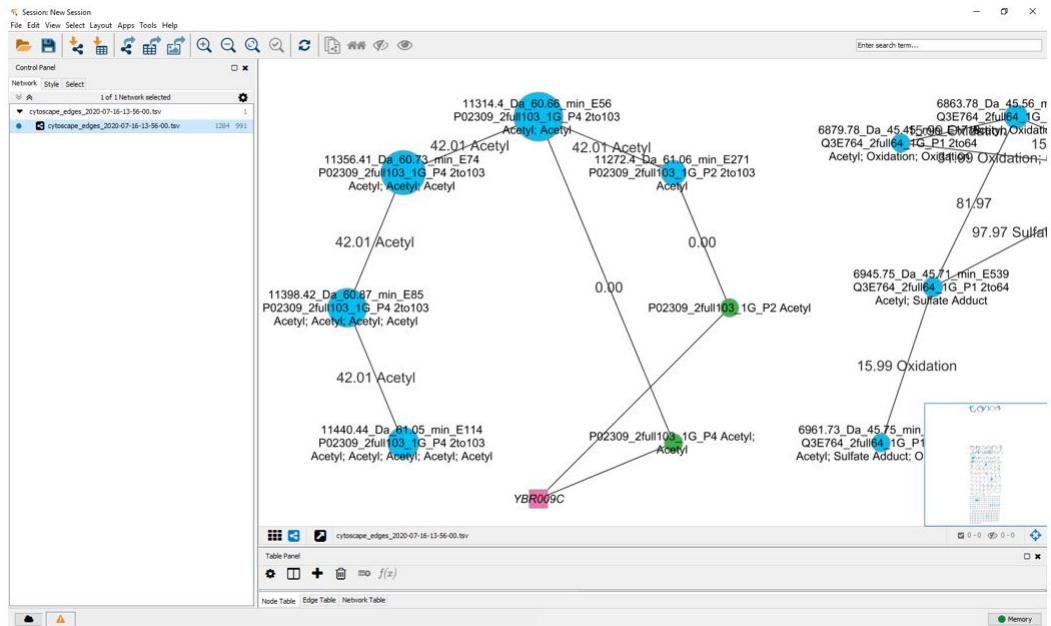
14.2 Visualize Families in Cytoscape

- Export scripts for Cytoscape either on the Proteoform Families page or the Results Summary page
 - Cytoscape_style_timestamp: file Cytoscape uses to correctly style visualized proteoform families
 - Cytoscape_nodes_timestamp: file containing node information for visualized proteoform families
 - Cytoscape_edges_timestamp: file containing edge information for visualized proteoform families
 - Cytoscape_script_timestamp: script for Cytoscape to load in the style, nodes, and edges file and visualize proteoform families
- In Cytoscape, select Tools > Execute Command File....
- Select the cytoscape_script_timestamp file generated by Proteoform Suite
- Proteoform families will appear! Should take about 1 min for larger proteoform families.



14.3 Proteoform Families Key

- Pink square: gene name
- Green nodes: theoretical proteoforms
- Blue nodes: experimental proteoforms; size of node is proportional to summed intensity
- Purple nodes: top-down proteoforms
- Edges: mass differences between proteoforms
- Specialized
 - Quantified Proteoform families: pie chart for each experimental proteoform shows abundance ratio between two conditions (blue and yellow)
 - Bottom-up data: orange nodes are bottom-up peptides. Edges indicate that the peptide could be derived from connected the proteoform node



14.4 Exporting Family Visualizations into Adobe Illustrator

- In Cytoscape, File > Export > Network View as Graphics...
- Select SVG
- This will export everything in view into a format that can be loaded by Adobe Illustrator and polished there for publication

15 Identified Proteoforms

15.1 Overview

On this page, all identified proteoforms are displayed, including intact-mass identification (from proteoform family construction) and top-down identifications (from loaded top-down results). The results automatically refresh each time the page is loaded.

15.2 Results

- Compare With other Top-Down Results: option to select another top-down hit results file and export and Excel file comparing intact-mass identifications with top-down identifications
- Table Filters: filter Identified Intact-Mass Experimental Proteoforms table (top left) and Top-Down Proteoforms table (top right) by any entered text
- Identified Intact-Mass Experimental Proteoforms: the left table displays experimental proteoforms identified by intact-mass analysis (proteoform family construction). See the Aggregated Proteoforms table in the **Aggregated Proteoforms** section for column descriptions.

PFR Accession	Input Filename	Scan	Reported Mass	Theoretical Mass	Retention Time	Uniprot ID	Sequence	Begin	End	PTM Description	Accession	Name	Q-Value	Score	Shared
P04912_91_12—	12-10-16,A17A—	15108	3364.8431	3364.8467	158.29	P04912	NIDELNLNLNL..._91	121		Unmodified	P04912	Histone H2A.2	0.000131	29.231	✓
P04912_98_10—	12-10-16,A17A—	14346	2408.3947	2408.3994	150.36	P04912	LLQNTIAQGG..._98	120		Unmodified	P04912	Histone H2A.2	0.000125	28.236	✓
P04912_84_12—	12-10-16,A17A—	16175	4068.2588	4068.2596	170.20	P04912	HLCLARNDE..._84	120		Unmodified	P04912	Histone H2A.2	2.8E-05	21.132	✓
P04912_91_10—	12-10-16,A17A—	16007	3236.7524	3236.7517	168.32	P04912	NIDELNLNLNL..._91	120		Unmodified	P04912	Histone H2A.2	9.4E-05	33.237	✓
P04912_98_12—	12-10-16,A17A—	13196	2536.4864	2536.4853	138.73	P04912	LLQNTIAQGG..._98	121		Unmodified	P04912	Histone H2A.2	9.5E-05	26.247	✓
P04912_84_97—	12-10-16,A17A—	6035	1677.8803	1677.8798	78.58	P04912	HLCLARNDE..._84	97		Unmodified	P04912	Histone H2A.2	0.000965	20.194	✓

- Top-Down Proteoforms: the right table displays top-down proteoform identifications (aggregated from top-down hit results loaded from top-down results). See the Top-Down Proteoforms table in the **Top-Down Proteoforms** section for column descriptions.

Proteiform Suite

File Method

Load Results Theoretical Database Top Down Raw Experimental Components New Code Pairs Aggregated Proteoforms Experiment Theoretical Comparison Experiment Experiment Comparison Proteoform Families Identified Proteoforms Quantification Results Summary

Compare With other Top-Down Results Table Filters

Identified Experimental Proteoforms Not in Top-Down

Experimental Proteoform ID	Aggregated Mass	Aggregated RT	Aggregated DecoyIntensity	Top-Down Proteoform	Manually Shifted Mass	Aggregated Component Count for Identification	Description
E309	15646.7148	54.19	9257000.3752			1	605 ribosomes
E2107	12068.754	51.53	199691.6341			1	405 ribosomes
E2151	15369.7252	66.24	170661.7159			1	Histone H3_B
E855	7273.9669	53.63	9910407.7250			5	605 ribosomes
E2335	14438.6832	84.50	26785.8016			5	405 ribosomes
E2150	14108.0377	49.39	615206.3994			4	605 ribosomes
E93	8551.5728	54.04	24381154.4552			11	Ubiquitin-405
E1877	15203.7949	67.44	643519.8403			2	Pynovate dec
E977	14402.6634	55.80	9062204.2524			6	605 ribosomes
E1865	14124.6111	57.46	92703.5167			1	605 ribosomes
E238	6531.6582	42.60	1419154.6732			2	605 ribosomes
E2066	15762.793	58.20	269602.0206			1	605 ribosomes
E266	11178.1725	51.79	89776409.1821			16	605 ribosomes
E266	11178.1725	51.79	89776409.1821			16	605 ribosomes
E266	11178.1725	51.79	89776409.1821			16	605 ribosomes

Top-Down Proteoforms

Accession	Modified Mass	Retention Time	Observations	PFR Accession	Original PFR/full-seqenc	Description	Gene Name
P42945_1232_1...	1380.666	62.78	3	P42945_1232_1...	209763	U3 small nucle...	UTP10
P05747_2_36_A...	4001.1950	42.89	30	P05747_2_36_A...	178837	605 ribosomal ...	RPL29
P04912_70e132...	6944.8396	62.64	1	P04912_70_e132...	194429	Histone H2A.2	HTA2
P04456_2_27_A...	2562.5914	41.57	12	P04456_2_27_A...	183476	605 ribosomal ...	RPL25
P03521_257e29...	4652.5521	45.29	1	P03521_257_e29...	209915	605 ribosomal ...	RPL5
P11484_49e61...	1293.8377	58.67	19	P11484_49_e61...	182855	Heat shock pro...	SB81
P05756_261e44...	16230.9453	58.93	3	P05756_2_144...	209463	405 ribosomal ...	RPS13
Q1874_59e67...	8838.5256	117.30	1	Q1874_59_e67...	209915	405 ribosomal ...	RPS21B
P49167_2e767...	8693.0639	53.52	276	P49167_2_76_A...	11017	605 ribosomal ...	RPL38
P49157_2e767...	8690.0648	118.74	94	P49157_2_76_A...	11017	605 ribosomal ...	RPL38
P49157_2e767...	8690.0644	62.15	43	P49157_2_76_A...	11017	605 ribosomal ...	RPL38
P49157_2e767...	8690.0645	71.80	19	P49157_2_76_A...	11017	605 ribosomal ...	RPL38
P49157_2e767...	8690.0574	130.11	1	P49157_2_76_A...	11017	605 ribosomal ...	RPL38

Bottom-Up Peptides from Selected Proteoform

Accession	Input Filename	Scan	Reported Mass	Theoretical Mass	Retention Time	Uniprot ID	Sequence	Begin	End	PTM Description	Accession	Name	Q-Value	Score	Shared
P04912_91_121...	12-10-16_A17A...	15106	3364.8452	3364.8467	158.29	P04912	NODEKLKLLG...	91	121	Unmodified	P04912	Histone H2A.2	0.000131	29.231	✓
P04912_90_120...	12-10-16_A17A...	14346	2408.3947	2408.3904	150.36	P04912	LLGNVTAAGGG...	98	120	Unmodified	P04912	Histone H2A.2	0.000215	28.236	✓
P04912_84_120...	12-10-16_A17A...	16175	4068.2588	4068.2596	170.20	P04912	HQLQARNHDE...	84	120	Unmodified	P04912	Histone H2A.2	2.8E-05	21.132	✓
P04912_84_120...	12-10-16_A17A...	16007	3236.7324	3236.7317	168.32	P04912	NODEKLKLLG...	91	120	Unmodified	P04912	Histone H2A.2	2.8E-05	33.237	✓
P04912_84_121...	12-10-16_A17A...	13196	2536.4854	2536.4853	138.73	P04912	LLGNVTAAGGG...	98	121	Unmodified	P04912	Histone H2A.2	0.5E-05	26.247	✓
P04912_84_97...	12-10-16_A17A...	6035	1677.8803	1677.8798	78.58	P04912	HQLQARNHDE...	84	97	Unmodified	P04912	Histone H2A.2	0.000965	20.194	✓

- Bottom-Up Peptides from Selected Proteoform: the bottom table displays bottom-up peptides from the selected proteoform in either the Identified Intact-Mass Experimental Proteoforms table or the Top-Down Proteoforms table. See the Bottom-Up Peptides table in the **Top-Down Proteoforms** section for column descriptions.

Proteiform Suite

File Method

Load Results Theoretical Database Top Down Raw Experimental Components New Code Pairs Aggregated Proteoforms Experiment Theoretical Comparison Experiment Experiment Comparison Proteoform Families Identified Proteoforms Quantification Results Summary

Compare With other Top-Down Results Table Filters

Identified Experimental Proteoforms Not in Top-Down

Experimental Proteoform ID	Aggregated Mass	Aggregated RT	Aggregated DecoyIntensity	Top-Down Proteoform	Manually Shifted Mass	Aggregated Component Count for Identification	Description
E309	15646.7148	54.19	9257000.3752			1	605 ribosomes
E2107	12068.754	51.53	199691.6341			1	405 ribosomes
E2151	15369.7252	66.24	170661.7159			1	Histone H3_B
E855	7273.9669	53.63	9910407.7250			5	605 ribosomes
E2335	14438.6832	84.50	26785.8016			5	405 ribosomes
E2150	14108.0377	49.39	615206.3994			4	605 ribosomes
E93	8551.5728	54.04	24381154.4552			11	Ubiquitin-405
E1877	15203.7949	67.44	643519.8403			2	Pynovate dec
E977	14402.6634	55.80	9062204.2524			6	605 ribosomes
E1865	14124.6111	57.46	92703.5167			1	605 ribosomes
E238	6531.6582	42.60	1419154.6732			2	605 ribosomes
E2066	15762.793	58.20	269602.0206			1	605 ribosomes
E266	11178.1725	51.79	89776409.1821			16	605 ribosomes
E266	11178.1725	51.79	89776409.1821			16	605 ribosomes
E266	11178.1725	51.79	89776409.1821			16	605 ribosomes

Top-Down Proteoforms

Accession	Modified Mass	Retention Time	Observations	PFR Accession	Original PFR/full-seqenc	Description	Gene Name
P42945_1232_1...	1380.666	62.78	3	P42945_1232_1...	209763	U3 small nucle...	UTP10
P05747_2_36_A...	4001.1950	42.89	30	P05747_2_36_A...	178837	605 ribosomal ...	RPL29
P04912_70e132...	6944.8396	62.64	1	P04912_70_e132...	194429	Histone H2A.2	HTA2
P04456_2_27_A...	2562.5914	41.57	12	P04456_2_27_A...	183476	605 ribosomal ...	RPL25
P03521_257e29...	4652.5521	45.29	1	P03521_257_e29...	209915	605 ribosomal ...	RPL5
P11484_49e61...	1293.8377	58.67	19	P11484_49_e61...	182855	Heat shock pro...	SB81
P05756_261e44...	16230.9453	58.93	3	P05756_2_144...	209463	405 ribosomal ...	RPS13
Q1874_59e67...	8838.5256	117.30	1	Q1874_59_e67...	209915	405 ribosomal ...	RPS21B
P49167_2e767...	8693.0639	53.52	276	P49167_2_76_A...	11017	605 ribosomal ...	RPL38
P49157_2e767...	8690.0648	118.74	94	P49157_2_76_A...	11017	605 ribosomal ...	RPL38
P49157_2e767...	8690.0644	62.15	43	P49157_2_76_A...	11017	605 ribosomal ...	RPL38
P49157_2e767...	8690.0645	71.80	19	P49157_2_76_A...	11017	605 ribosomal ...	RPL38
P49157_2e767...	8690.0574	130.11	1	P49157_2_76_A...	11017	605 ribosomal ...	RPL38

Bottom-Up Peptides from Selected Proteoform

Accession	Input Filename	Scan	Reported Mass	Theoretical Mass	Retention Time	Uniprot ID	Sequence	Begin	End	PTM Description	Accession	Name	Q-Value	Score	Shared
P04912_91_121...	12-10-16_A17A...	15106	3364.8452	3364.8467	158.29	P04912	NODEKLKLLG...	91	121	Unmodified	P04912	Histone H2A.2	0.000131	29.231	✓
P04912_90_120...	12-10-16_A17A...	14346	2408.3947	2408.3904	150.36	P04912	LLGNVTAAGGG...	98	120	Unmodified	P04912	Histone H2A.2	0.000215	28.236	✓
P04912_84_120...	12-10-16_A17A...	16175	4068.2588	4068.2596	170.20	P04912	HQLQARNHDE...	84	120	Unmodified	P04912	Histone H2A.2	2.8E-05	21.132	✓
P04912_84_120...	12-10-16_A17A...	16007	3236.7324	3236.7317	168.32	P04912	NODEKLKLLG...	91	120	Unmodified	P04912	Histone H2A.2	2.8E-05	33.237	✓
P04912_84_121...	12-10-16_A17A...	13196	2536.4854	2536.4853	138.73	P04912	LLGNVTAAGGG...	98	121	Unmodified	P04912	Histone H2A.2	0.5E-05	26.247	✓
P04912_84_97...	12-10-16_A17A...	6035	1677.8803	1677.8798	78.58	P04912	HQLQARNHDE...	84	97	Unmodified	P04912	Histone H2A.2	0.000965	20.194	✓

16 Quantification

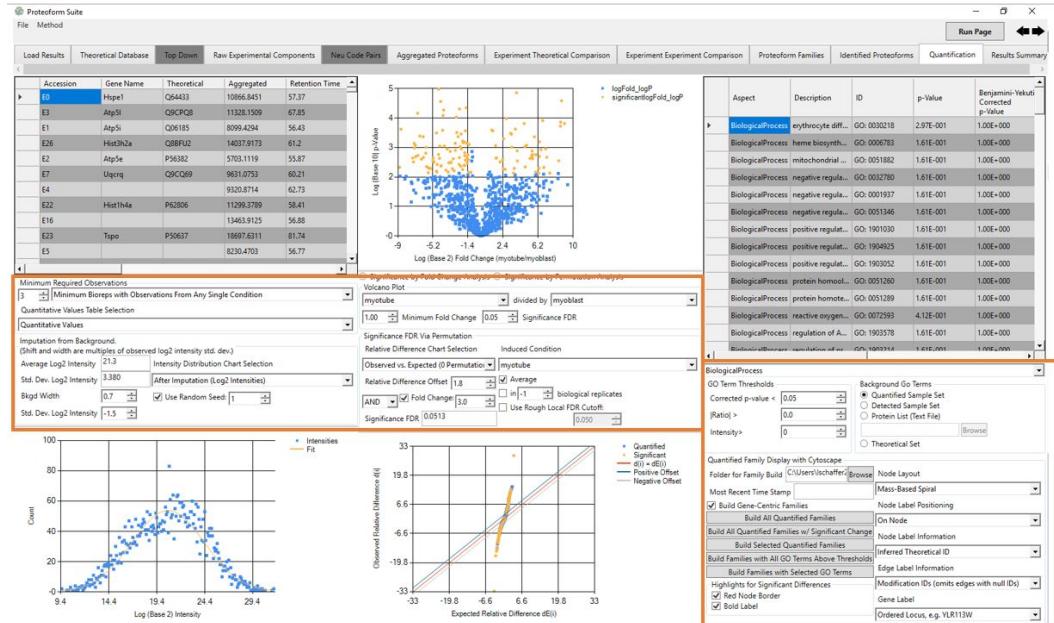
16.1 Overview

On this page, experimental proteoforms are quantified and Proteoform Suite determines experimental proteoforms with statistically significant abundance differences between two conditions. Two separate statistical tests are performed: a permutation analysis based on Tusher et al.¹⁹ and a log2 fold-change t-test with a Benjamini Hochberg multiple testing correction.

16.2 Run Page

- The Proteoform Families page must be run before running this page.
- Set all parameters as desired for current analysis (see below)
- Click Run Page button (top right)

16.3 Set Parameters

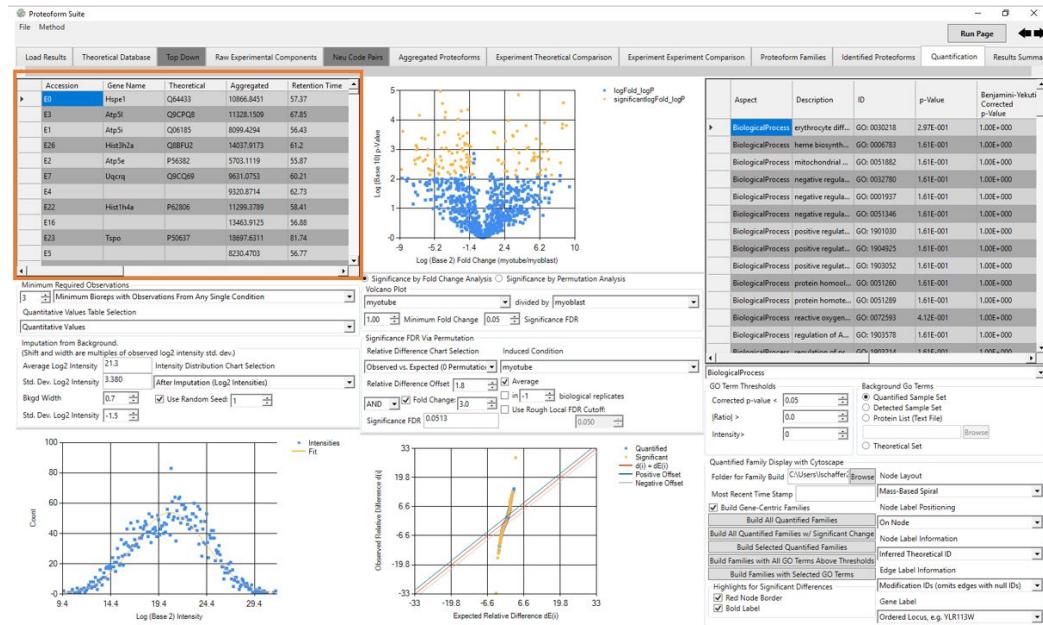


- Minimum Required Observations: set the # (left) and the requirement (right drop down box) to require quantitative observations of an experimental proteoform in more than one file type
- Quantitative Values Table Selection: select what will be displayed in the Quantitative Values table (top right)
- Bkgd Width: background distribution for imputation width (sigma)
- Std. Dev. Log2 Intensity: background distribution for imputation shift, number of sigma from the population mean

- Use Random Seed: a random seed will be used in the random number generator for selecting imputed intensity values, resulting in the same intensity values each time (with the same given parameters)
- Significant by Fold Change Analysis: select this option to have the fold change analysis be what determines which experimental proteoforms have statistically significant abundance changes
- Significance by Perutations: select this option to have the permutation analysis be what determines which experimental proteoforms have statistically significant abundance changes
- Volcano Plot: choose which condition intensity is divided by which condition intensity
- Minimum Fold Change: set the minimum log2 fold change for an experimental proteoform's abundance change to be considered significant
- Significance FDR: set the maximum false discovery rate for an experimental proteoform's abundance change to be considered statistically significant
- Relative Difference Chart Selection: select which permutation analysis chart is displayed in the bottom middle graph
- Induced Condition: select which condition is the induced condition
- Relative Difference Offset: select a minimum relative difference offset from the expected relative difference curve for an abundance change to be considered significant using permutation analysis
- AND/OR: option to use both relative difference offset and/or a minimum fold change value
- Fold Change: select a minimum fold change value for an abundance change to be considered significant
- Average: check this box to use average permutation fold change
- In # biological replicates: check this box and set the # to set a minimum number of biological replicates
- Use Rough Local FDR cutoff: check this box to use a local false discovery rate cutoff in permutation analysis; set the maximum FDR cutoff below
- GO Term Dropbox: select which gene ontology terms to display in the Gene Ontology table (top right)
- GO Term Thresholds: select thresholds for a gene ontology term to be considered significant (maximum p-value, minimum ratio, minimum intensity)
- Background GO Terms: select what should be used for the background gene ontology terms (quantified only, all detected, or a new protein list loaded in with Browse)

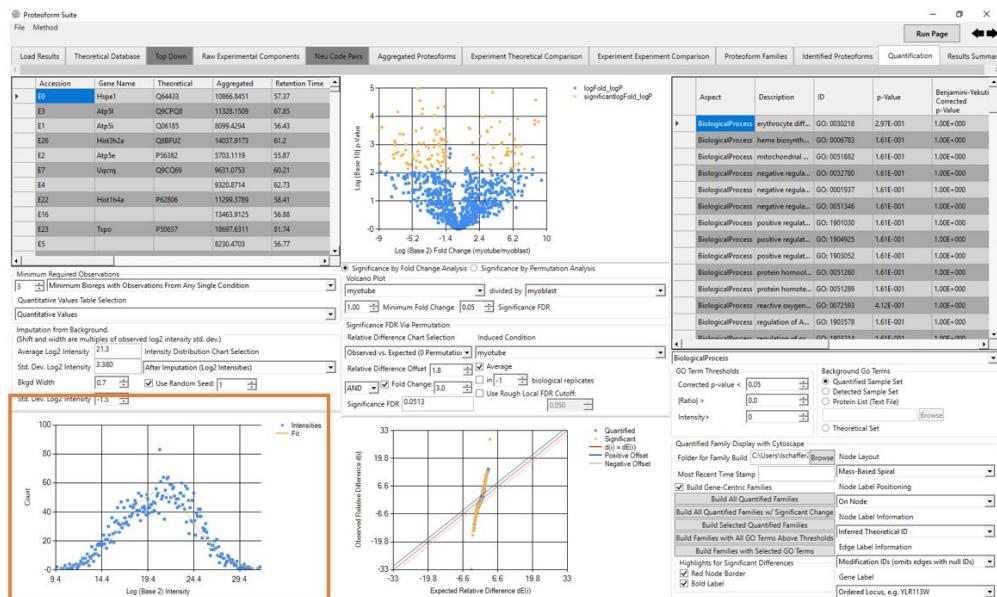
16.4 Results

- Quantitative Values table: this top left table displays quantified experimental proteoforms

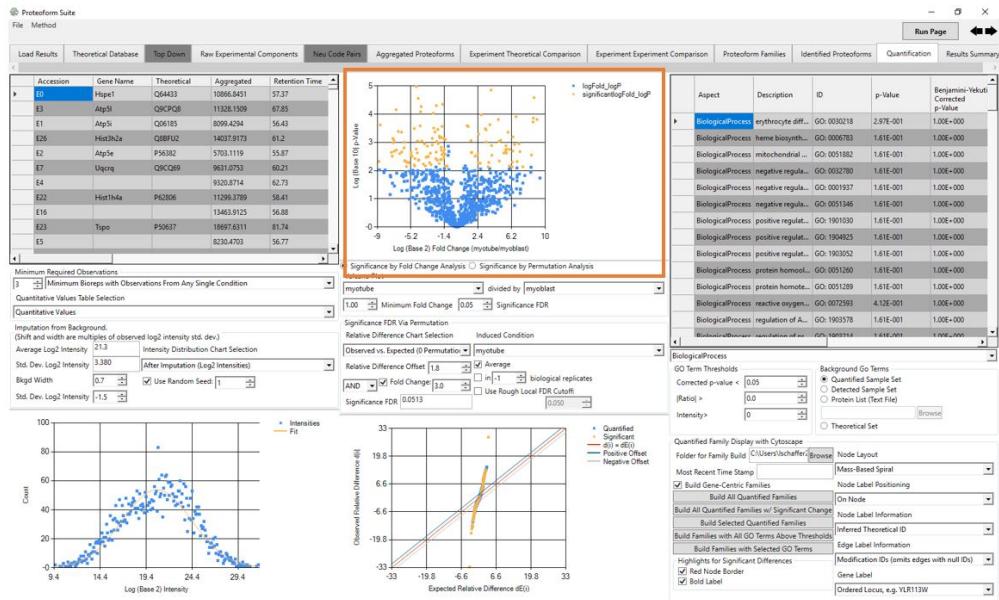


- Accession: unique accession given by Proteoform Suite given to this experimental proteoform
- Gene Name: if identified, gene name for this experimental proteoform
- Theoretical: if identified, theoretical accession from UniProt for this experimental proteoform
- Aggregated Mass: monoisotopic mass of experimental proteoform, weighted average of (light) raw experimental components by intensity
- Aggregated RT: retention time of experimental proteoform, average of raw experimental components (unlabeled) or NeuCode pairs (NeuCode labeled)
- Condition1 Intensity Sum: summed intensity of raw quantitative components for this condition for this experimental proteoform
- Condition2 Intensity Sum: summed intensity of raw quantitative components for this condition for this experimental proteoform
- Intensity Sum: total summed intensity of all raw quantitative components (both conditions) for this experimental proteoform
- Log2 Fold Change: log2 fold change between 2 conditions for this experimental proteoform
- Scatter linear: if significance by permutation analysis, linear intensity
- p-value: if significance by fold change analysis, p-value for this experimental proteoform fold change test statistic

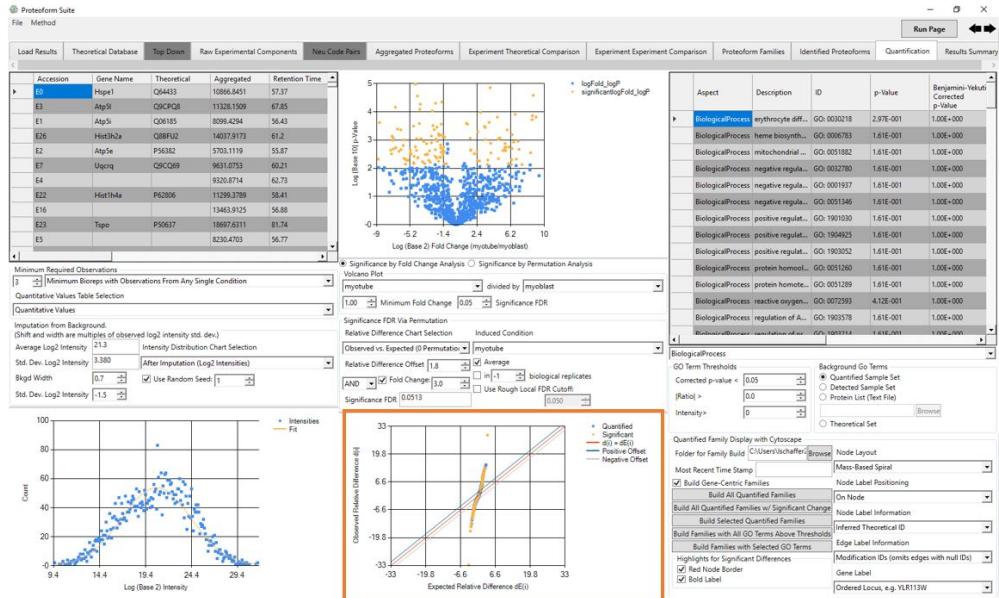
- Benjamini-Hochberg corrected p-value: if significance by fold change analysis, Benjamini-Hochberg corrected p-value for this experimental proteoform fold change test statistic
- Significant: checked if fold change for this experimental proteoform is statistically significant
- Student's t-test Statistic: test statistic for log2 fold change analysis t-test
- Corresponding Avg. Permuted Student's t-Test Statistic: averaged permuted student's test statistic
- Manual Validation: file information for manual validation of most abundant raw quantitative component
- Imputations from Background: this bottom left graph displays info about the log2 intensities of quantified proteoforms, with an option to view before and after imputation.



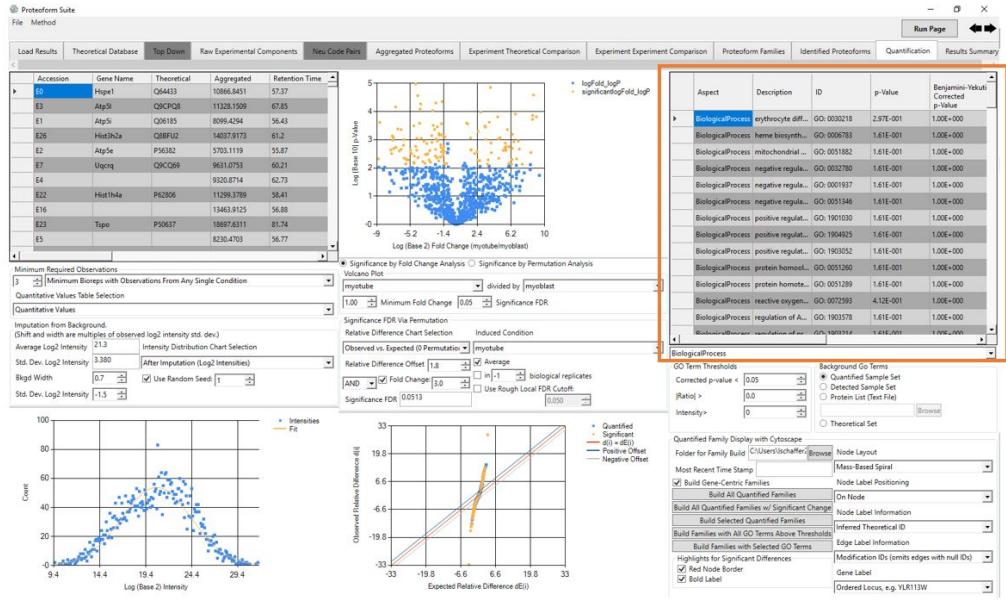
- Volcano plot: this top middle graph shows the a volcano plot from the fold change analysis, p-value vs. log₂ fold change. Proteoforms with a statistically significant fold change are yellow, other proteoforms are blue.



- Permutation Analysis graph: the bottom middle graph shows information from the permutation analysis, including observed relative difference vs. expected relative difference



- Gene Ontology Terms table: the top right table displays the gene ontology terms
 - Aspect: gene ontology term type



- Description: gene ontology term description
- ID: gene ontology term
- P-Value: p-value for this gene ontology term
- Benjamini-Yekutieli Corrected p-Value: Benjamini-Yekutieli corrected p-value for this gene ontology term
- Log Odds Ratio: log odds ratio
- Significant Proteins With This Go-Term: number of unique proteins with this GO term and with an experimental proteoform with a statistically significant abundance change
- Total Significant Proteins: number of unique proteins with an experimental proteoform with a statistically significant abundance change
- Background Proteins With This Go-Term: number of unique background proteins with this GO term
- Total Background Proteins: total number of background proteins
- Folder for Family Build: folder to build scripts to visualize proteoform families in Cytoscape
- Most Recent Time Stamp: time stamp that will be in the filename of scripts to visualize proteoform families in Cytoscape
- Build Gene-Centric Families: if checked, all proteoforms connected to theoretical proteoforms from the same gene will be grouped into the same proteoform family
- Build All Quantified Families: exports scripts for Cytoscape to visualize all quantified proteoform families
- Build All Quantified Families w/ Significant Change: exports scripts for Cytoscape to visualize all quantified proteoform families with at least one experimental proteoform with a statistically significant abundance change

- Build Selected Quantified Families: exports scripts for Cytoscape to visualize proteoform families with an experimental proteoform selected in the Quantification Values table
- Build Families with All GO Terms Above Threshold: exports scripts for Cytoscape to visualize all quantified proteoform families with gene ontology terms above threshold
- Build Families with Selected GO Terms: exports scripts for Cytoscape to visualize proteoform families with a gene ontology term selected in the Gene Ontology Terms table
- Highlights for Significant Differences: if checked, a red node border and bold label will be used in quantitative proteoform families to highlight experimental proteoforms with statistically significant quantitative differences
- Node Layout: this drop-down box changes the node layout in visualized proteoform families in Cytoscape
- Node Label Positioning: this drop-down box changes the position of node labels in visualized proteoform families in Cytoscape
- Node Label Informing: this drop-down box changes the node label information in visualized proteoform families in Cytoscape
- Edge Label Information: this drop-down box changes the edge label information in visualized proteoform families in Cytoscape
- Preferred Gene Label: this drop-down box changes the gene label information in visualized proteoform families in Cytoscape

17 Results Summary

17.1 Overview

This page displays all of the results from the Proteoform Suite analysis. The results automatically refresh each time the page is loaded. Results and tables can be exported.

17.2 Results

- Results Folder: browse for a folder where all results will be exported
- Save All: save all results files in the folder selected
 - ExperimentExperimental_MassDifferences_timestamp.png: image of Experiment-Experiment Delta Mass Histogram
 - ExperimentTheoretical_MassDifferences_timestamp.png: image of Experiment-Theoretical Delta Mass Histogram
 - Presets_timestamp.xml: method file of all set parameters in Proteoform Suite. Can be used in future Proteoform Suite analyses to replicate this analysis
 - Proteoform_bottomup_evidence_timestamp.tsv: if bottom-up data input, this is a list of potential proteoforms inferred with bottom-up evidence
 - Bottomup_results_timestamp.tsv: if bottom-up data input, this is a list of bottom-up peptides
 - Shared_peptide_bottomup_results_timestamp.tsv: if bottom-up data input, this is a list of shared bottom-up peptides
 - Topdown_results_timestamp.tsv: if top-down data input, this is a list of top-down proteoforms
 - AllFamilies_cytoscape_edges_timestamp.tsv: file containing edge information for all proteoform families in Cytoscape visualization
 - AllFamilies_cytoscape_nodes_timestamp.tsv: file containing node information for all proteoform families in Cytoscape visualization
 - AllFamilies_cytoscape_script_timestamp.tsv: script to visualize all proteoform families in Cytoscape
 - AllFamilies_cytoscape_style_timestamp.tsv: file containing style information for all proteoform families in Cytoscape visualization
 - BottomUp_cytoscape_edges_timestamp.tsv: if bottom-up data input, file containing edge information for all proteoform families in Cytoscape visualization
 - BottomUp_cytoscape_nodes_timestamp.tsv: if bottom-up data input, file containing node information for all proteoform families in Cytoscape visualization
 - BottomUp_cytoscape_script_timestamp.tsv: if bottom-up data input, script to visualize peptide-to-proteoform relations in Cytoscape
 - BottomUp_cytoscape_style_timestamp.tsv: if bottom-up data input, file containing style information for all proteoform families in Cytoscape visualization

- Decoy_experimental_results_timestamp.tsv: all decoy intact-mass identifications
- Experimental_intensities_by_file_timestamppe.tsv: intensity for each file for each experimental proteoform
- Experimental_results_timestamp.tsv: experimental proteoform intact-mass identifications
- Summary_timestamppe.txt: summary of all proteoform results, displayed on Results Summary page
- Results Summary
 - Unprocessed Raw Experimental Components: the number of raw experimental components for identification before merging artifacts
 - Raw Experimental Components: the number of raw experimental components for identification after merging artifacts
 - Missed Monoisotopic Raw Experimental Components Merged: the number of raw experimental components for identification artifacts merged due to being missed monoisotopic errors within the set mass tolerance
 - Harmonic Raw Experimental Components Merged: the number of raw experimental components for identification artifacts merged due to being charge state harmonic errors within the set mass tolerance
 - Unprocessed Raw Quantitative Components: the number of raw experimental components for quantification before merging artifacts
 - Raw Quantitative Components: the number of raw experimental components for quantification after merging artifacts
 - Missed Monoisotopic Raw Quantitative Components Merged: the number of raw experimental components for quantification artifacts merged due to being missed monoisotopic errors within the set mass tolerance
 - Harmonic Raw Quantitative Components Merged: the number of raw experimental components for quantification artifacts merged due to being charge state harmonic errors within the set mass tolerance
 - Raw NeuCode Pairs: the total number of NeuCode pairs, each with a heavy and light NeuCode raw experimental component
 - Accepted NeuCode Pairs: the number of accepted NeuCode pairs, which are used in subsequent analysis
 - Top-Down Hit: total number of top-down hits (proteoform spectral matches)
 - Accepted Level 1 and 2 Top-Down Proteoforms: number of aggregated top-down proteoforms. May be greater than the number of unique PFRs if some hits of the same ID fall outside of the retention time tolerance. Only level 1 and 2 top-down proteoform identifications are included
 - Experimental Proteoforms: the total number of experimental proteoforms
 - Accepted Experimental Proteoforms: the number of accepted experimental proteoforms that are used in subsequent analysis

- Accepted Intact-Mass Experimental Proteoforms: the number of accepted experimental proteoforms aggregated from raw experimental components from deconvolution results
- Accepted Level 1 and 2 Top-Down Experimental Proteoforms: the number of accepted experimental proteoforms aggregated from top-down hits from top-down results. Only level 1 and 2 top-down proteoform identifications are included
- Theoretical Proteins: the number of unique proteins
- Expanded Theoretical Proteins: the number of unique protein sequences, including annotated subsequences
- Theoretical Proteoforms: the number of theoretical proteoforms in the database
- Experiment-Theoretical Peaks: the total number of experiment-theoretical delta mass peaks
- Experiment-Theoretical Pairs: the total number of experiment-theoretical pairs, each with a delta mass between the experiment and theoretical proteoforms
- Accepted Experiment-Theoretical Peaks: the number of accepted experiment-theoretical delta mass peaks
- Accepted Experiment-Theoretical Pairs: the number of experiment-theoretical pairs in accepted delta mass peaks, used in subsequent proteoform family construction
- Average Experiment-Decoy Pairs: the average number of experiment-decoy pairs across all decoy databases generated
- Experiment-Experiment Peaks: the total number of experiment-experiment delta mass peaks
- Experiment-Experiment Pairs: the total number of experiment-experiment pairs, each with a delta mass between the two experimental proteoforms
- Accepted Experiment-Experiment Peaks: the number of accepted experiment-experiment delta mass peaks
- Accepted Experiment-Experiment Pairs: the number of experiment-experiment pairs in accepted delta mass peaks, used in subsequent proteoform family construction
- Average Experiment-False Pairs: the average number of experiment-false pairs across all decoy analyses
- Proteoform Families: the number of constructed proteoform families, from accepted experiment-theoretical and experiment-experiment pairs
- Identified Families (Correspond to 1 gene): the number of proteoform families with one unique gene from the theoretical proteoforms in the family
- Experimental Proteoforms in Identified Families: the number of experimental proteoforms in identified proteoform families with 1 gene
- Ambiguous Families (Correspond to > 1 gene): the number of proteoform families with more than one unique gene from the theoretical proteoforms in the family
- Experimental Proteoforms in Ambiguous Families: the number of experimental proteoforms in ambiguous proteoform families with more than 1 gene
- Unidentified Families (Correspond to no gene): the number of proteoform families with no genes (no theoretical proteoforms)

- Experimental Proteoforms in Unidentified Families: the number of experimental proteoforms in unidentified families with no theoretical proteoforms/genes
- Orphaned Experimental Proteoforms (Intact-mass proteoforms not joined with another proteoform): the number of experimental proteoforms that were not part of any accepted experiment-theoretical or experiment-experiment pairs
- Raw Experimental Components in Families: the number of raw experimental components that were aggregated into an experimental proteoform that was a part of a proteoform family (non-orphans)
- % of Raw Experimental Components in Families: the percentage of raw experimental components that were aggregated into an experimental proteoform that was a part of a proteoform family (non-orphans)
- Raw Quantitative Components in Families: the number of raw quantitative components that were aggregated into an experimental proteoform that was a part of a proteoform family (non-orphans)
- % of Raw Quantitative Components in Families: the percentage of raw quantitative components that were aggregated into an experimental proteoform that was a part of a proteoform family (non-orphans)
- Identified Experimental Proteoforms: the number of experimental proteoforms that were identified from proteoform family construction of experiment-theoretical and experiment-experiment pairs
- Average Identified Experimental Proteoforms by Decoys: the average number of experimental proteoforms that were identified from proteoform family construction of experiment-decoy and experiment-false pairs
- Proteoform FDR: the calculated global false discovery rate (average number of decoy identifications / experimental proteoform identifications)
- Identified Experimental Proteoforms (no Ambiguous): the number of experimental proteoforms that were identified from proteoform family construction of experiment-theoretical and experiment-experiment pairs, excluding ambiguous identifications
- Average Identified Experimental Proteoforms by Decoys (no Ambiguous): the average number of experimental proteoforms that were identified from proteoform family construction of experiment-decoy and experiment-false pairs, excluding ambiguous identifications
- Proteoform FDR: the calculated global false discovery rate (average number of decoy identifications / experimental proteoform identifications), excluding ambiguous identifications
- Level 1 and 2 Top-Down Proteoforms Assigned Same Identification by Intact-Mass Analysis: the number of top-down proteoforms that were assigned the same identification through proteoform family construction as the original top-down identification
- Level 1 and 2 Top-Down Proteoforms Assigned Different Identification by Intact-Mass Analysis: the number of top-down proteoforms that were assigned a different identification through proteoform family construction as the original top-down identification

- Level 1 and 2 Top-Down Proteoforms Assigned Ambiguous Identification by Intact-Mass Analysis: the number of top-down proteoforms that were assigned an ambiguous identification through proteoform family construction
- Level 1 and 2 Top-Down Proteoforms Unidentified by Intact-Mass Analysis: the number of top-down proteoforms that were not assigned an identification through proteoform family construction
- Unique Level 1 and 2 Top-Down Protein Identifications (TDPortal): the number of unique protein identifications from the level 1 and 2 top-down identifications
- Total Unique Protein Identifications: the total number of unique protein identifications, top-down + intact-mass
- Unique Intact-Mass Experimental Proteoform Identifications: the total number of unique protein identifications from proteoform family construction
- Ambiguous Intact-Mass Experimental Proteoform Identifications: the number of ambiguous experimental proteoform identification from proteoform family construction
- Unique Level 1 and 2 Top-Down Proteoforms Identifications (TDPortal): the number of unique level 1 and 2 top-down proteoform identifications from the top-down hit results
- Unidentified Intact-Mass Experimental Proteoforms: the number of unidentified intact-mass experimental proteoforms (aggregated from raw experimental components from deconvolution results)
- Total Unique Proteoform Identifications: the total number of unique proteoform identifications from top-down results and intact-mass results
- Quantified Experimental Proteoforms (Threshold for Quantification: 0 = Minimum Bioreps with Observations From Any Single Condition): the number of experimental proteoforms that met the minimum threshold for quantification analysis
- Average Log2 Intensity Quantified Experimental Proteoform Observations: the average log2 intensity of experimental proteoforms that were quantified (calculated from the raw quantitative components)
- Log2 Intensity Standard Deviation for Quantified Experimental Proteoform: the standard deviation fo the log2 intensity of experimental proteoforms that were quantified
- Experimental Proteoforms with Significant Change (Threshold for Significance: Log2FoldChange > 0, & Total Intensity from Quantification > 0, & Q-Value < 0.05): experimental proteoforms with statistically significant fold change from both relative difference and fold change analysis
- Experimental Proteoforms with Significant Change by Relative Difference (Offset of 1 from $d(i) = dE(i)$ line): number of proteoforms with statistically significant relative difference in Tusher analysis with X permutations
- Experimental Proteoforms with Significant Change by Fold Change (Offset of 1 from $d(i) = dE(i)$ line): number of proteoforms with statistically significant fold change in Tusher analysis with X permutations
- FDR for Significance Conclusion (Offset of 1 from $d(i) = dE(i)$ line): the false discovery rate for the relative difference analysis from the Tusher analysis with X permutations

- Proteoform Families with Significant Change: number of proteoform families with at least one experimental proteoform with a statistically significant fold change from the Tusher analysis with X permutations
- Identified Proteins with Significant Change: the number of unique identified proteins with at least one identified experimental proteoform with a statistically significant fold change from the Tusher analysis with X permutations
- GO Terms of Significance (Benjamini-Yekutieli p-value < 0.05; using Experimental Proteoforms that satisfied the criteria: Log2FoldChange > 0, & Total Intensity from Quantification > 0, & Q-Value < 0.05): number of significant gene ontology terms from the Tusher analysis with X permutations
- Experimental Proteoforms with Significant Change (Threshold for Significance: Benjamini-Hochberg Q-Value < 0.05): number of proteoforms with statistically significant fold change from log2 fold change t-test with Benjamini-Hochberg correction
- FDR for Significance Conclusion: FDR value for fold change to be considered statistically significant
- Proteoform Families with Significant Change: number of proteoforms with statistically significant relative difference from log2 fold change t-test with Benjamini-Hochberg correction
- Identified Proteins with Significant Change: the number of unique identified proteins with at least one identified experimental proteoform with a statistically significant fold change from log2 fold change t-test with Benjamini-Hochberg correction
- GO Terms of Significance (Benjamini-Yekutieli p-value < 0.05; using Experimental Proteoforms that satisfied the criteria: Log2FoldChange > 0, & Total Intensity from Quantification > 0, & Q-Value < 0.05): number of significant gene ontology terms from log2 fold change t-test with Benjamini-Hochberg correction
- Identified Proteins with Significant Change: (0 Permutations): list of unique identified proteins with at least one identified experimental proteoform with a statistically significant fold change from Tusher analysis with X permutations
- Identified Proteins with Significant Change: (by log2 fold change analysis): list of unique identified proteins with at least one identified experimental proteoform with a statistically significant fold change from log2 fold change t-test with Benjamini-Hochberg correction
- GO Terms of Significance, Tusher Analysis with 0 permutations (Benjamini-Yekutieli p-value < 0.05): list of statistically significant gene ontology terms from Tusher analysis with X permutations
- GO Terms of Significance, Log2 Fold Change Analysis with 0.05 FDR (Benjamini-Yekutieli p-value < 0.05): list of statistically significant gene ontology terms from log2 fold change t-test with Benjamini-Hochberg correction
- USER ACTIONS: list of user actions, including adding files, changing file labels, accepting/unaccepting delta mass peaks, shifting the mass of experiment-theoretical delta mass peaks
- DECONVOLUTION RESULTS FILES AND PROTEIN DATABASE FILES: list of files loaded on the Load Results page

References

- (1) Shortreed, M. R.; Frey, B. L.; Scalf, M.; Knoener, R. A.; Cesnik, A. J.; Smith, L. M. Elucidating Proteoform Families from Proteoform Intact-Mass and Lysine-Count Measurements. *J Proteome Res* **2016**, *15*, 1213–21.
- (2) Cesnik, A. J.; Shortreed, M. R.; Schaffer, L. V.; Knoener, R. A.; Frey, B. L.; Scalf, M.; Solntsev, S. K.; Dai, Y.; Gasch, A. P.; Smith, L. M. Proteoform Suite: Software for Constructing, Quantifying, and Visualizing Proteoform Families. *J Proteome Res* **2018**, *17*, 568–578.
- (3) Dai, Y.; Shortreed, M. R.; Scalf, M.; Frey, B. L.; Cesnik, A. J.; Solntsev, S.; Schaffer, L. V.; Smith, L. M. Elucidating Escherichia coli Proteoform Families Using Intact-Mass Proteomics and a Global PTM Discovery Database. *J Proteome Res* **2017**, *16*, 4156–4165.
- (4) Schaffer, L. V.; Shortreed, M. R.; Cesnik, A. J.; Frey, B. L.; Solntsev, S. K.; Scalf, M.; Smith, L. M. Expanding Proteoform Identifications in Top-Down Proteomic Analyses by Constructing Proteoform Families. *Anal Chem* **2018**, *90*, 1325–1333.
- (5) Schaffer, L. V.; Renvold, J. W.; Shortreed, M. R.; Cesnik, A. J.; Jochem, A.; Scalf, M.; Frey, B. L.; Pagliarini, D. J.; Smith, L. M. Identification and Quantification of Murine Mitochondrial Proteoforms Using an Integrated Top-Down and Intact-Mass Strategy. *J Proteome Res* **2018**, *17*, 3526–3536.
- (6) Dai, Y.; Buxton, K. E.; Schaffer, L. V.; Miller, R. M.; Millikin, R. J.; Scalf, M.; Frey, B. L.; Shortreed, M. R.; Smith, L. M. Constructing Human Proteoform Families Using Intact-Mass and Top-Down Proteomics with a Multi-Protease Global Post-Translational Modification Discovery Database. *J Proteome Res* **2019**, *18*, 3671–3680.
- (7) Schaffer, L. V.; Tucholski, T.; Shortreed, M. R.; Ge, Y.; Smith, L. M. Intact-Mass Analysis Facilitating the Identification of Large Human Heart Proteoforms. *Anal Chem* **2019**, *91*, 10937–10942.
- (8) Schaffer, L. V.; Millikin, R. J.; Shortreed, M. R.; Scalf, M.; Smith, L. M. Improving Proteoform Identifications in Complex Systems Through Integration of Bottom-Up and Top-Down Data. *Journal of Proteome Research* **2020**, DOI: [10.1021/acs.jproteome.0c00332](https://doi.org/10.1021/acs.jproteome.0c00332).
- (9) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **2003**, *13*, 2498–504.
- (10) Smoot, M. E.; Ono, K.; Ruscheinski, J.; Wang, P. L.; Ideker, T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*, **27**, 431–2.
- (11) Jeong, K.; Kim, J.; Gaikwad, M.; Hidayah, S. N.; Heikaus, L.; Schluter, H.; Kohlbacher, O. FLASHDeconv: Ultrafast, High-Quality Feature Deconvolution for Top-Down Proteomics. *Cell Syst* **2020**, *10*, 213–218 e6.
- (12) LeDuc, R. D.; Taylor, G. K.; Kim, Y. B.; Januszyk, T. E.; Bynum, L. H.; Sola, J. V.; Garavelli, J. S.; Kelleher, N. L. ProSight PTM: an integrated environment for protein identification and characterization by top-down mass spectrometry. *Nucleic Acids Res* **2004**, *32*, W340–5.

- (13) Zamdborg, L.; LeDuc, R. D.; Glowacz, K. J.; Kim, Y. B.; Viswanathan, V.; Spaulding, I. T.; Early, B. P.; Bluhm, E. J.; Babai, S.; Kelleher, N. L. ProSight PTM 2.0: improved protein identification and characterization for top down mass spectrometry. *Nucleic Acids Res* **2007**, *35*, W701–6.
- (14) Toby, T. K.; Fornelli, L.; Kristina, S.; DeHart, C. J.; Fellers, R. T. A comprehensive pipeline for translational top-down proteomics from a single blood draw. *Nature Protocols* **2019**, *14*, 119–152.
- (15) Solntsev, S. K.; Shortreed, M. R.; Frey, B. L.; Smith, L. M. Enhanced Global Post-translational Modification Discovery with MetaMorpheus. *J Proteome Res* **2018**, *17*, 1844–1851.
- (16) Schaffer, L. V. et al. Identification and Quantification of Proteoforms by Mass Spectrometry. *Proteomics* **2019**, e1800361.
- (17) LeDuc, R. D.; Fellers, R. T.; Early, B. P.; Greer, J. B.; Thomas, P. M.; Kelleher, N. L. The C-score: a Bayesian framework to sharply improve proteoform scoring in high-throughput top down proteomics. *J Proteome Res* **2014**, *13*, 3231–40.
- (18) Smith, L. M. et al. A five-level classification system for proteoform identifications. *Nature Methods* **2019**, DOI: 10.1038/s41592-019-0573-x.
- (19) Tusher, V. G.; Tibshirani, R.; Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences of the United States of America* **2001**, *98*, 5116–21.