PROLINE TUTORIAL

PROLINE BASICS : WORKING WITH MS/MS IDENTIFICATIONS

I/ START PROLINE

A. AUTHENTICATION

The connection window appears when Proline starts. Fill the requested fields with the right information depending on your installation (see note below).

Server host: address of Proline server
User: your login
Password: your password

Note	*If you are running Proline Zero, the server host is "localhost" and the user/password are "proline/proline". These information are prefilled.*

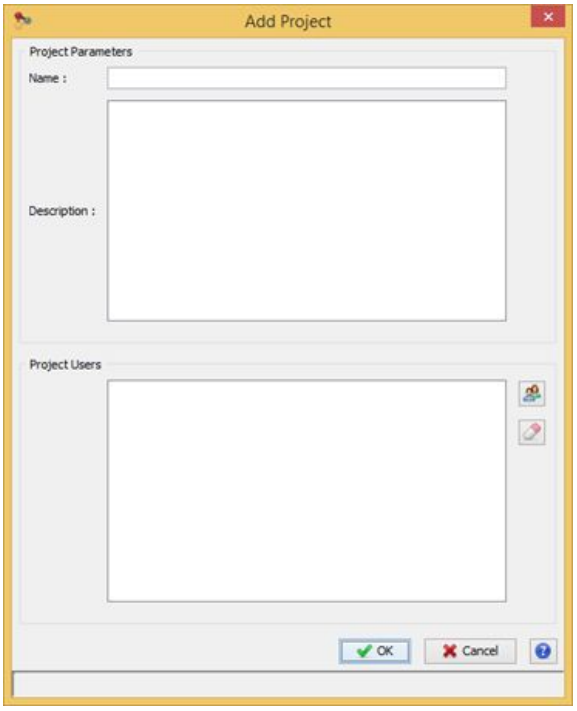
Note	*Raw data and Mascot identification result files used throughout this tutorial are available for download at http://www.profiproteomics.fr/proline/#downloads, under **sample datasets.** Once downloaded, mzDB files and mascot dat files must be, respectively, copied into the mzdb_files and result_files mount points configured in Proline server. If you are running Proline Zero, these folders are*

located in <Proline-Zero-Folder>/data/mzdb and <Proline-Zero-Folder>/data/mascot respectively.

B. Create a new project

To create a new project in Proline, click on the green cross in the Projects pane on the left.

This opens the following window:



In Proline Server mode, a project can be shared with other users. To do so, choose the user in the project edition window. The project will then be visible for these users with restricted functionalities.

Action	Give a name and a description for the project and confirm

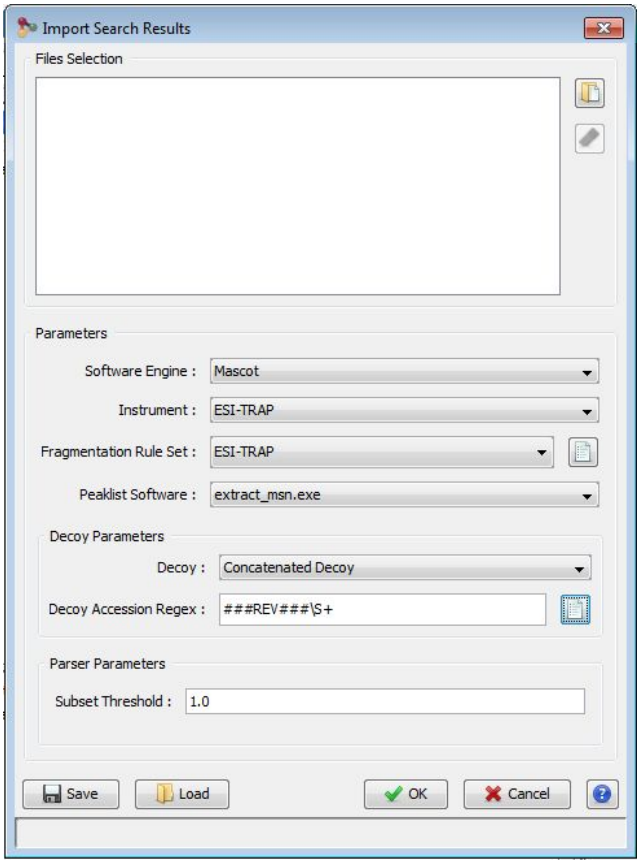
II/ Import files

The files that will be imported in this tutorial correspond to analyses of UPS1 proteins (standard equimolar solution of 48 proteins) spiked (10 fmol) in 2 µg of yeast and injected 3 times on a VELOS ETD mass spectrometer.

A. IMPORT F083064.DAT FILE

To import a file in Proline, right-click on the Identifications node and select Import *Search Result*.

This opens the following window:



File selection: one or more files to import

Parameters

Software engine: Search engine used to generate the file to import. It is automatically selected from the selected file

Instrument: Mass spectrometer which allowed to perform the analysis

Fragmentation Rule Set: The fragmentation rules specified in software. by clicking on you could visualize all rules for a specific rule set. This is necessary to generate spectrum matches.

Peaklist software: Software used to generate the peaklist

Decoy: Target/Decoy strategy used during the search

Decoy Accession Regex: Regular expression enabling target and decoy protein differentiation in a concatenated bank

Use icon to choose a predefined regular expression.

Parser Parameters

Subset Threshold: Possibility to put a filter on the subsets (the default value is 1)

Import parameters can be saved and reused.

Note	*See Proline Help for more details on the various parameters*

Action	Select: • File *F083064.dat* • *ESI-TRAP for instrument and fragmentation rule set* • Peaklist Software: *extract_msn.exe* • Decoy parameters: *Concatenated Decoy* with Decoy Accession Regex: *###REV###\S+* Monitor *Task Log*

B. Import files F083066.dat and F083067.dat

Now please import two other files that are replicates of the former one. They will be processed in Part IV. You can continue the tutorial without waiting for the end of imports.

Action	Import files with the same settings as before.

Note	*More than one file can be imported at the same time.*

C. View the contents of file F083604.dat

In Proline, the content of a file before validation is called *Search Result*. To view the *Search Result*, you can right-click on the icon representing the imported file and then click *Display Search Result* and *PSMs*...

A default representation of the data opens but each view in Proline Studio can be customized. A graphical représentation can then be added to each tabular view of the data.

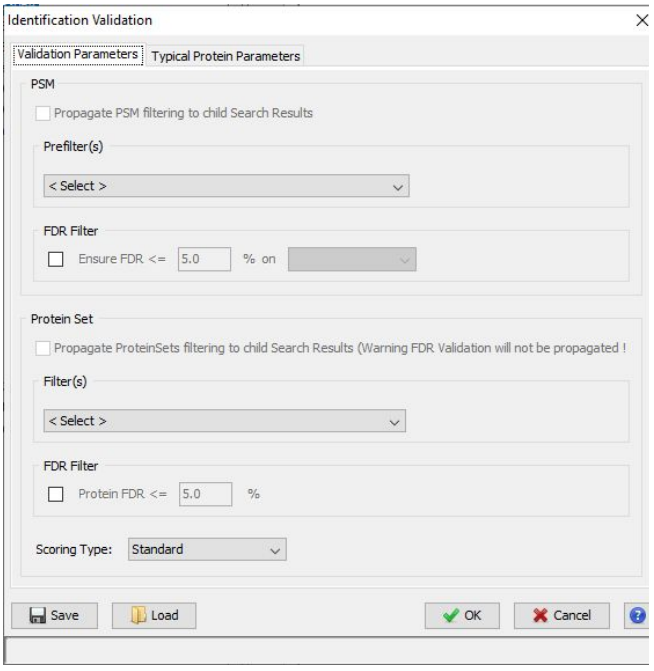
Note	*If the icon representing the Search result doesn't appear, double click on the "All imported" icon. This will open a new panel showing all Search Results imported into the project. An imported file can then be drag and dropped from that panel to the project in the identifications tree located on the left of the window.*

III/ Validating file F083064.dat

A. Validation on rank 1

VALIDATION CRITERIA

To validate a file, right-click on the Search Result underneath the project node and then on *Validate Search Result…*



PSM

- **Prefilter(s)**: Allows filtering according to various criteria (rank, length, score…)
- **FDR Filter**: Allows filtering PSMs to obtain the FDR requested by varying a parameter

Protein Set

- **Filter(s)**: Allows filtering proteins according to various criteria (number of specific peptides…)
- **FDR Filter**: Allows filtering proteins to get the request FDR by varying the score
- **Scoring type**: Method chosen to calculate protein score

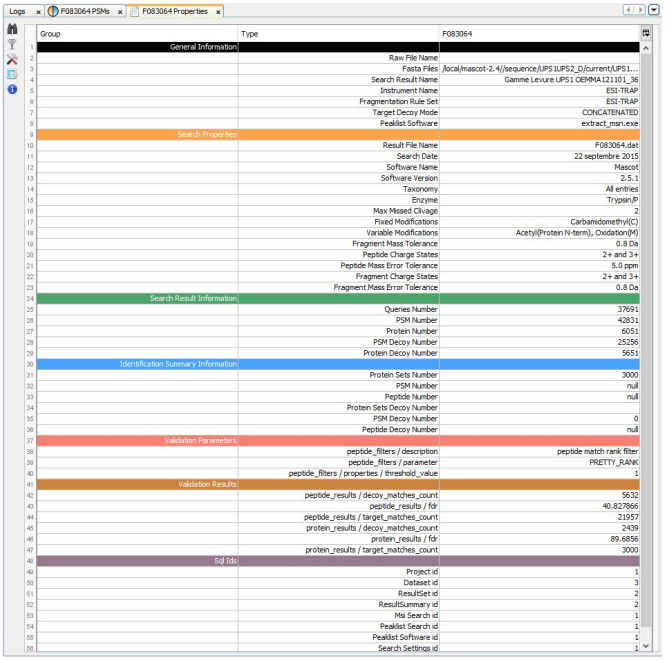
Validation parameters can be saved and reused.

Action	Validate the file using a first set of criteria: ● Only keep pretty rank 1 PSMs ● Scoring type: « *Standard* »

Validation results

In Proline, a validated result is called an *Identification Summary*. An identification summary can be browsed within Studio by right clicking the dataset. Click *Display Identification Summary* and *PSMs* to browse all target PSM. The opened view can be customized, for instance to add a graphical representation of the PSM table. To add this panel, click on the  icon in the lower right corner to add a *Graphics* panel. Choose to display an histogram plot of the PSM scores, it indicates that the average score after validation is 27,68.

Useful information about the Search Result or the Identification Summary can also be found in the Properties view: right click on the dataset and click *Properties*. For instance, number of validated target and decoy PSM, peptide level FDR, number of validated protein sets or protein level FDR can be found in this view.



Search for a valid Protein Set

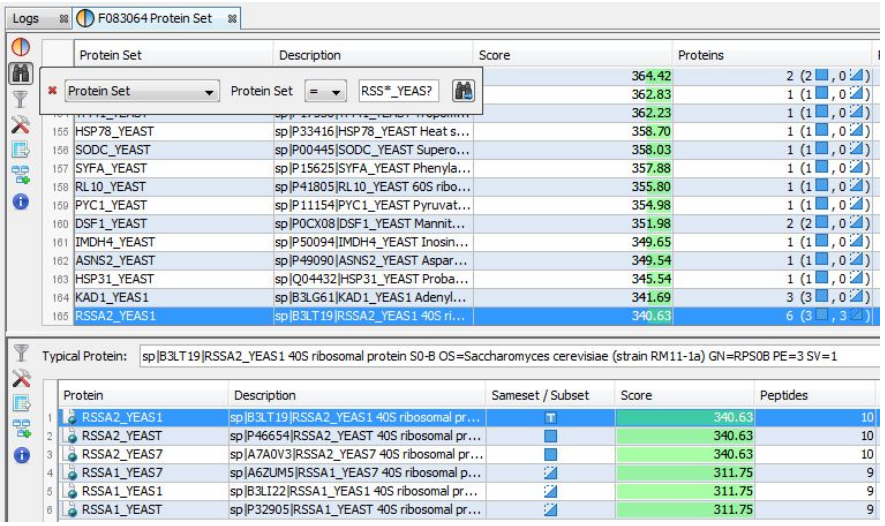
Click *Display Identification Summary* and *Protein Sets* to browse all Protein sets. *Proteins sets* represent the set of proteins corresponding to a validated *peptide set*. A protein called typical protein is chosen to depict a *protein set*.

Each table provides a Search tool (Binoculars icon) and a Filter tool (Funnel icon). In the *protein set* view, these tools will only consider the typical protein of each *protein set*.

Action	Search for RSSA2_YEAST Search for RSS*_YEAS?

Note	*Wildcards "?" and "*" are available in both tools. The interrogation mark matches for any character once, the star character matches for zero or more characters. Searching for "**RSS*_YEAS?**" will return all Protein Sets starting with "RSS", containing "_YEAS" with only one letter after it.*

Even if a typical protein is selected by the algorithm, proteins matching the same set of peptides or a subset of those peptides are shown in the protein set view. RSSA2_YEAS1 is the typical protein of a protein set composed of 6 proteins. Three of them are *sameset* proteins and three are *subset* proteins matching 9 peptides out of 10 peptides in the peptide set. The sameset/subset icons indicates that RSSA2_YEAST is a sameset protein but not the typical one.



Typical protein choice

The typical protein can be chosen using parsing rules. These rules can be set during the validation step, or afterwards (right-click on the *Result File*, then *Change Typical Protein…*).

For example Swissprot proteins can be favoured in a mixed protein bank. In our case, we want YEAST protein to be favoured.

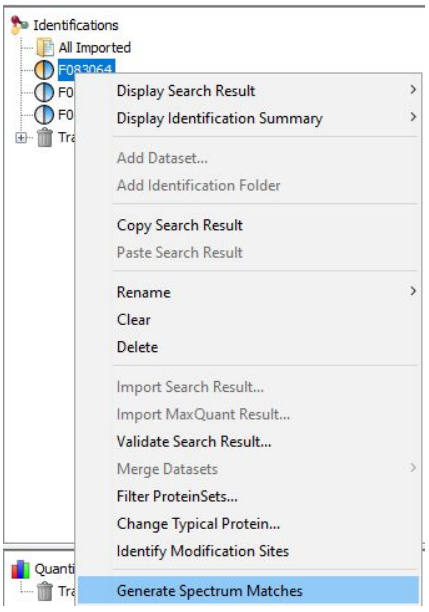
Action	Change typical proteins to favor YEAST labeled ones (Enter “YEAST” in Rule 0). Reload the Protein Set view Search for RSSA2_YEAST again

Annotate spectrum and generate fragmentation table

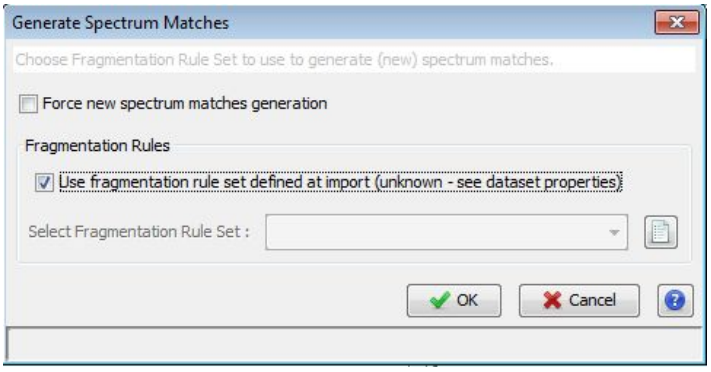
To get an annotated spectrum and the associated theoretical fragmentation table, *Spectrum Matches* have to be generated. This can be done in two ways:

For each spectrum by clicking on the icon on the left

Or on the whole set of valid spectra



In both cases, the following dialog will be opened to define which fragmentation rule sets to use.
As we have already specified “ESI Trap” at import we can select “Use fragmentation rule sets defined at import...”



Note	You can zoom a graphic using the mouse wheel, reset zoom by right-clicking the mouse button and drag toward top left the graphic.

B. Validation at peptidic level

Validation criteria

Properties view indicates the number of validated target PSM (= 21957). Now Revalidate the same dataset (F083064) using following parameters

Action	Revalidate the file using a second set of criteria: • Only keep pretty rank 1 PSMs • Select an FDR of 5% PSM based on the score • Open the Properties view to compare with the previous results

Validation results

The Properties view shows that this new validation criteria improves the identification results. The number of validated target PSMs drops from 21957 to 13510 and the peptide level FDR is now below the 5% requested threshold. However, the protein FDR still remains high (more than 32%).

D. Validation at peptidic and proteic level

Validation criteria

Action	Revalidate the file using a third set of criteria: • Only keep pretty rank 1 PSMs • Select an FDR of 5% for the PSMs based on the score • Select an FDR of 1% for the proteins (standard scoring)

Again, the Properties view gives useful information about this new validated dataset.

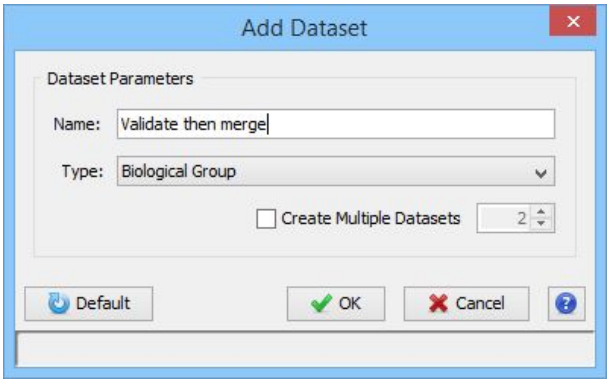
IV/ Merge Result Files

A. Merge validated files

Create a dataset

To create a dataset in Proline, right click on Identification and select *Add* and then *Dataset*

This open the following window:



Action	Name the dataset « Validate then merge »

Trick	*The box « Create Multiple Datasets » allows to create a series of datasets in one click*

Place a copy of each files in the dataset

Each imported file in Proline can be reused with no needs to load it again. *All Imported* allows retrieving the imported files.

Action	Right-click on *All Imported* and then *Display List* (or double click on *All imported*) Select the files to copy (F083064, F083066, F083067), drag and drop them in the created dataset

Validate the three files

Action	Validate the files with the same settings as before (third set of criteria)

Trick	*To validate several files with the same criteria at once, select all the files together, right-click on them and select "Validate Search Result…"*

Merge the three files

To merge files, right click on the dataset and select *Merge Datasets/ Aggregation*.

Note: After merging, the dataset can be filtered but not revalidated.

Note	The two merge methods are explained in the “Proline User Guide”, “Concepts & principles” part.

View the results

The result of a merge can be seen the same way as the files that make it up.

Action	Look for RL26A_YEAST protein among the validated *Protein Sets*.

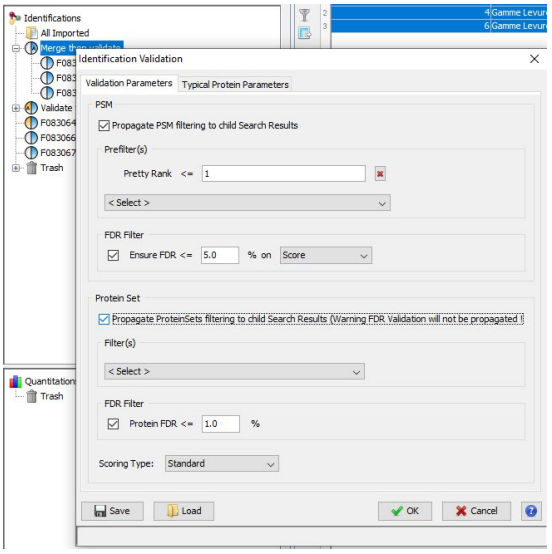
B. Validate merged files

Create a second dataset but merge before validation

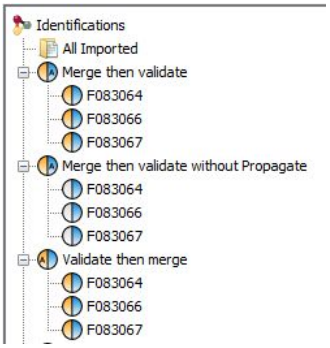
Action	Run the previous steps again but in the following order: • Create a new dataset “Merge then validate” • Put a copy of the three files in this dataset • Merge the dataset • Validate the dataset with the same settings as before

Note	After a merge, the dataset can be revalidated but not the files that make it up. However, the validation done on the parent dataset could be propagated to its childs. The same parameters or calculated parameters will be used.

You can choose *Propagate* on PSMs and/or Protein Sets



Merge then Validate with/without propagate



View the result

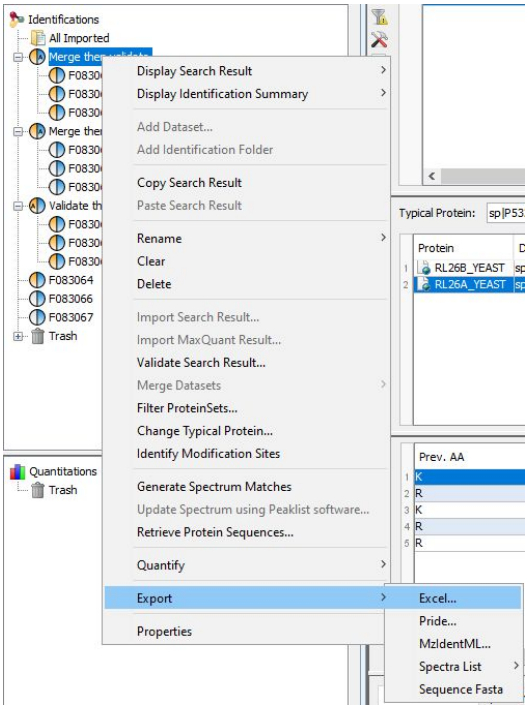
Action	Look for RL26A_YEAST protein among the validated *Protein Sets*.

Note	Look for RL26* *Proteins Sets* and look for samesets-/ subsets

V/ EXPORT RESULTS

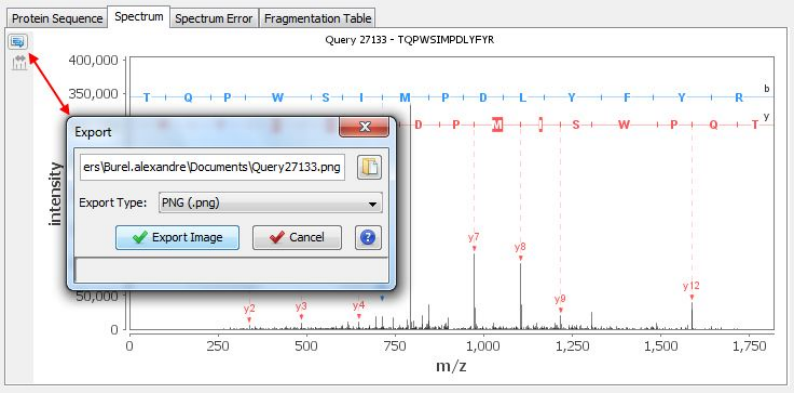
Proline proposes many ways to export data:

A global one is available after validation.



Each table can be exported, either with the button on the top of the table either by copying and pasting table rows in an Excel sheet.

Each graphic element (spectra, histogram) can be exported as a picture. The export types proposed by Proline are PNG (standard one) et SVG (Vectorial picture).



Trick	*The global export generates an Excel file that represents an Identification Summary. It includes generation file parameters, validation criteria and the results at PSM, peptidic, protein and protein set levels.* *Proline Help describes all these options.*

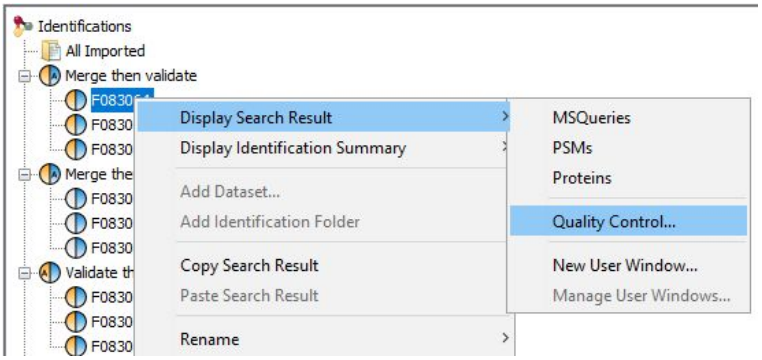
Action	Select the last dataset you have created • Export only Decoy PSM before validation • Export an annotated spectrum

- Export the complete validated results for this dataset

VI/ To go further

A. Quality control

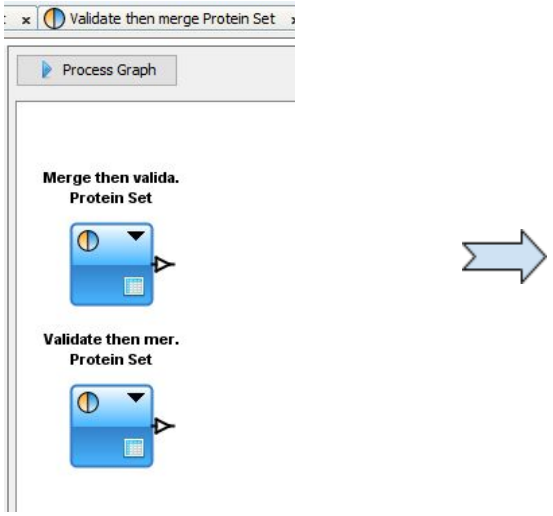
Quality control enables a transversal view on a Search Result. Rather than visualising the results per PSM or Proteins, results are sorted according to ranges of score, M/z, charge state, target or decoy…



B. Compare the results with Data Analyzer

The *Data Analyzer* tool can run a large amount of computation on any kind of data. Every view of any Search Result and Identification Summary proposes a button ("*Add data to Data Analyzer*...") which will open the *Data Analyzer* window. There you can run a series of functions on any column; join or compare two tables and calculate statistics. Some of these statistical functions will be further developed in other tutorial.

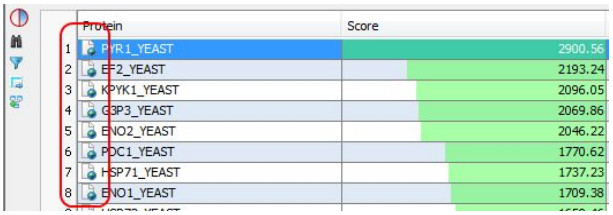
Action	Open "Validate then merge" and "Merge then validate" *Protein Set* view. On each view, click on the "*Add data to Data Analyser*..." button. Use the Join function with default settings then display the results (click on table icon).



C. Other functions of Proline Studio

Uniprot link

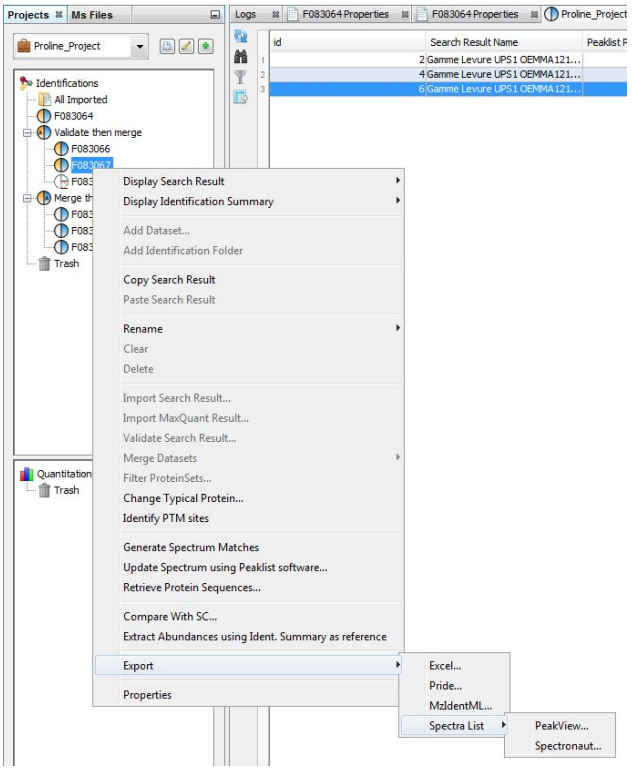
An icon enables direct link to Uniprot website. This icon is available in protein lists.



Export a spectrum library

A list of validated and annotated spectra can be exported to build a spectrum library that can be used in Skyline software for example.

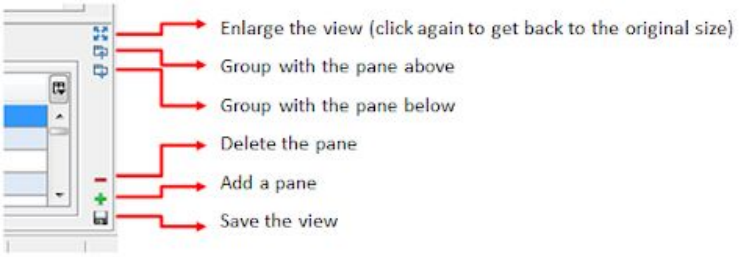
You can choose to export a spectra List compatible with PeakView or Spectronaut.



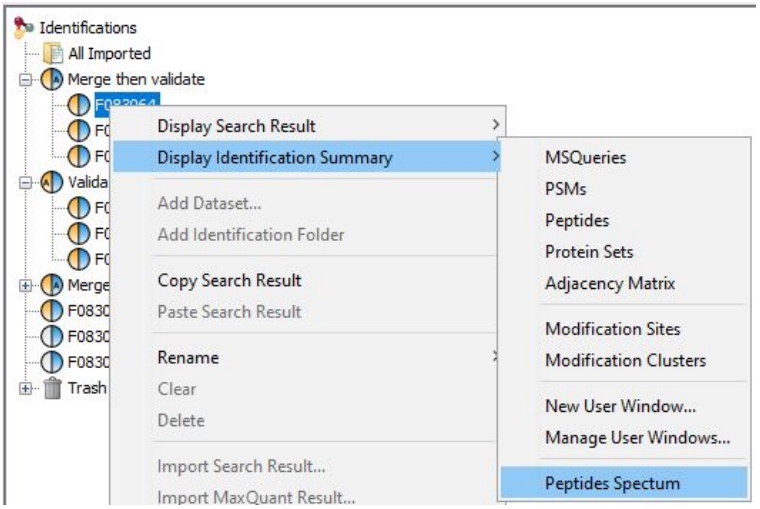
Note	*You should have run the "**Generate Spectrum Matches**" on the selected dataset before exporting to spectrum library. Otherwise an error may occur.*

Custom Views

A view can be designed to suit user needs. Each window is composed of several panes that can be expanded, moved or suppressed.



Saved views can be opened just like any other views : right-click on your Result File and select either *Search Result* or *Identification Summary*. You can also create a new view from scratch and select which pane should compose it.



Action	Create a new view reproducing the *Search Result* PSM view, but using the *Identification Summary* instead. This view should contain: • The list of validated PSMs on top • In the middle pane: the spectrum, spectrum error, fragmentation table and spectrum values • The corresponding proteins Save this view as PSMs Custom.

E. EXPORT TO PRIDE REPOSITORY

Proline export allows you to deposit your **validated** DDA Proteomics Data to the ProteomeXchange Repository as a complete submission data [1,2,3]. See documentation for more details on required information.

(1) Martens, L.; Hermjakob, H.; Jones, P.; Adamski, M.; Taylor, C.; States, D.; Gevaert, K.; Vandekerckhove, J.; Apweiler, R. PRIDE: the proteomics identifications database. Proteomics 2005, 5 (13), 3537–3545.

(2) Ternent, T.; Csordas, A.; Qi, D.; Gómez-Baena, G.; Beynon, R. J.; Jones, A. R.; Hermjakob, H.; Vizcaíno, J. A. How to submit MS proteomics data to ProteomeXchange via the PRIDE database. Proteomics 2014, 14 (20), 2233–2241.

(3) Vizcaíno, J. A.; Csordas, A.; Del-Toro, N.; Dianes, J. A.; Griss, J.; Lavidas, I.; Mayer, G.; Perez-Riverol, Y.; Reisinger, F.; Ternent, T.; et al. 2016 update of the PRIDE database and its related tools. Nucleic Acids Res. 2016, 44 (D1), D447–D456.

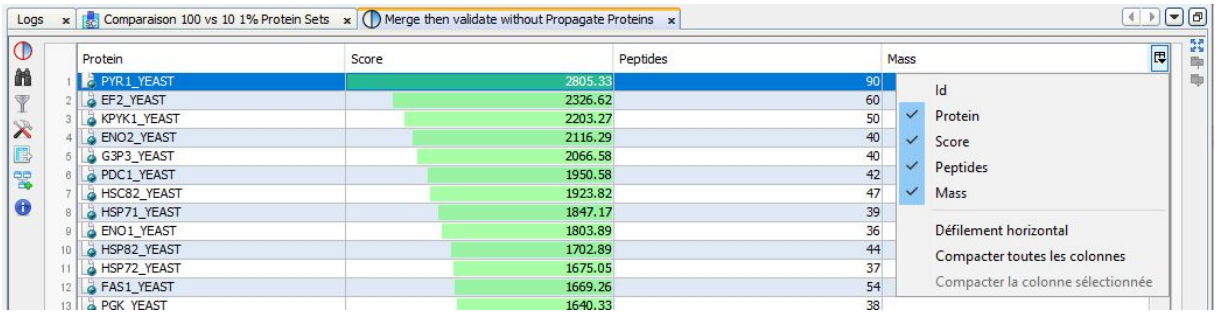
Note	*It is also possible to export datasets into **mzIdentML format**. This is the new required format for submission on ProteomeXchange Repository.*

F. Using Marker Bar

In all table views, you can mark several rows by clicking on its row number. An overview of all marks in the table is available on the right side and you can easily reach them by clicking on it. These marks will remain after filtering/sorting.

G. Set Column Visibility

In all table views, you can customize which column to display using the icon at the top-right corner.



A more complete configuration is available using icon.