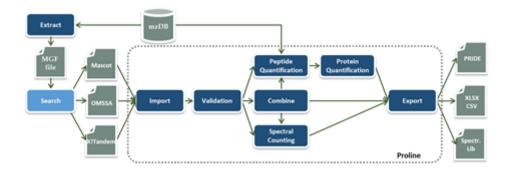


# **Proline User Guide**

## Release 2.2.0

### **Proline Software Suite**

Proline is a production grade software suite, which provides an environment for large-scale MS data management, visualization, analysis and curation with the main objective of promoting the production and sharing of high quality proteomic datasets. Proline can be used (i) to produce reliable identification and quantification results through robust automated processes, (ii) for data curation, (iii) to systematically save and keep track of metadata from processing steps, parameters and generated data, and (iv) to submit highly qualified datasets to public repositories.



A workflow in Proline is implemented as a collection of tasks (see Figure) that can be performed by the user through the graphical user interface. Users can import multiple identification results corresponding, for example, to fractions and replicates of a biological sample and combine them before or after validation. The resulting datasets can then be compared or quantified using spectral counting or DDA label-free quantification, before exporting the results in different file formats.

The software suite is based on two main components: a **server** handling processing tasks and based on **relational database management system** storing the data generated and two different graphical user interfaces, both allowing users to start tasks and visualize the data: **Proline Studio** which is a rich client interface and **Proline Web** the web client interface. An additional component called **ProlineAdmin** is used by system administrators to set up and manage Proline. **Proline Zero** is an all-in-one, "zero installation" solution containing the **server** and **Proline Studio**.

### Document organization

This document is organized in two sections:

- The <u>Concepts & Principles</u> section presents the main concepts and algorithms implemented in Proline as well as the different parameters of these algorithms.
- The <u>How to</u> section gives more details on how to proceed with the Proline Studio graphical interface.



# **Proline Concepts & Principles**

- Dataset types:
  - o <u>Result File</u>
  - o <u>Search Results</u>
  - o Decoy Searches
  - o Identification Summary
  - o <u>Modification Dataset</u>
- Data Processing:
  - o <u>Protein Inference</u>
  - o Protein and Proteins Sets scoring
  - o Validation Algorithm
  - o FDR Estimation
  - o Protein Sets Filtering
  - o <u>Combining datasets</u>
  - o Identify Modification Sites or Clusters
  - o <u>Compare with Spectral Count</u>
  - o Label Free LC-MS quantitation workflow
  - o <u>Post-processing of LC-MS quantitative results</u>
  - o Aggregation of Label-Free quantitative results
- Data Import/Export:
  - o Identification Summary Export
- Advanced Features:
  - o Allow multiple imports in parallel

### **Identification results**

Proline considers different types of identification data: Result Files, Search Results and Identification Summaries which will be defined in the following sections.

### **Result File**

A **result file** produced by a search engine can be imported into Proline in their native format. OMSSA (.omx files), Mascot (.dat files) and X!Tandem (.xml files) search engines are currently supported. In addition, the mzIdentML format is supported to allow the output from any other search engine compatible with this standard to be imported (e.g. MS-GF+). A first version for MaxQuant support has been implemented. It is possible to import only search results or to import search results as well as quantitation (beta version) values from MaxQuant files.

Search engines may provide different types of searches for MS and MS/MS data. It is important to highlight that Proline only supports MS/MS ions searches at this point.

### Search Result

A **search result** is the raw interpretation of a given set of MS/MS spectra given by a search engine. It contains one or many peptides matching the submitted MS/MS spectra (**PSM**, i.e. Peptide Spectrum Match), and the protein sequences these peptides belong to. The **search result** also contains additional information such as search parameters, used protein sequence databank, etc.



A **search result** is created when a r<u>esult file</u> is imported in Proline. During this step, no filtering or thresholding is applied: along with the search parameters, all submitted spectra, peptide spectrum matches (PSMs) and protein hits suggested by the search engine are retained in the Proline database to allow subsequent validation of putative identifications. In the case of a target-decoy search, two **search results** are created: one for the target PSMs, one for decoy PSMs.

#### Content of a search result

Importing a result file creates a new search result in the database which contains the following information:

- Search Settings: software name and version, parameters values
- Peaklist and Spectrum information: file name, MS level, precursor m/z, ...
- Search result data:
  - o Protein sequences
  - o Peptide sequences
  - o Spectra
  - o Two kinds of Matches:
    - Peptide Spectrum Matches (PSM), i.e. the matching between a peptide and a spectrum, with some related data such as the score, fragment matches...
    - Protein Matches, i.e. the proteins in the databank corresponding to the PSMs identified by the search engine

#### Search engine specificities

- Mascot
  - $\circ$   $\;$  The PSM score corresponds to the Mascot ion score.
- OMSSA
  - The PSM score corresponds to the negative common logarithm of the E-value: Score = -log10(E-value). Note that Proline only supports OMSSA Result Files generated with the 2.1.9 release.
- X!Tandem
  - The X!Tandem standard hyperscore is used as a PSM score. Note that Proline supports X!Tandem Result Files generated with the Sledgehammer release (or later).

### **Decoy Searches**

Proline handles decoy searches performed from two different strategies:

- Concatenated searches:
  - A protein databank is created by concatenating target protein sequence to decoy protein sequence.
     Decoy could be created using reverse or random strategy. A unique search is done using that databank.
  - o When importing a Search Result from a decoy concatenated databank, decoy data are extracted from the Result File and stored in Proline databases as a decoy Search Result independent of the target Search result. Nevertheless both searches are linked to each other.
- Separated searches:
  - Two searches are done using the same peaklist, one on a target protein databank and one on a decoy protein databank. These searches are then combined to retrieve useful information such as FDR. Mascot allows the user to check a decoy option and automatically creates a decoy databank.



o The two performed searches are stored in Proline databases and are linked together.

### **Identification Summary**

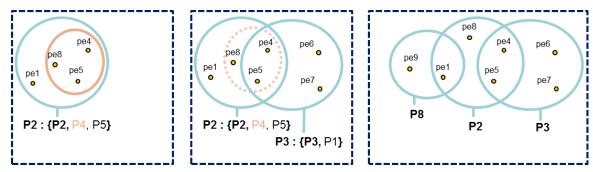
An **Identification Summary** is a set of identified proteins inferred from a subset of the PSM contained in the search <u>result</u> that have been declared valid. The subset of PSM taken into account are the PSM that have been validated by a filtering process (example: PSM fulfilling some specified criteria such as score greater than a threshold value).

#### Content of an identification summary

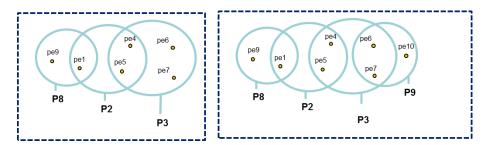
- Peptide Set
  - A set of peptides identifying one or more proteins.
- Protein Set
  - Typical Protein or representative protein. Protein "chosen" to represent the ProteinSet.
  - Sameset : all proteins identified by the same set of peptides than the typical one.
  - Subset: Proteins identified by a subset of peptides. All these peptides should also identify typical protein.

#### **Protein Inference**

All peptides identifying a protein are grouped in a Peptides Set. A same Peptides Set can identify many proteins, represented by one Proteins Set. In this case, one protein of this Protein Set is chosen to represent the set, it is the Typical Protein. If only a subset of peptides identify a (or some) protein(s), a new Peptide Set is created. This Peptide Set is a subset of the first one, and identified Proteins are Subset Proteins.



- In the first example, P2 and P5 are identified by the same peptide set {pe1, pe4, pe5, pe8}. P2 was chosen as typical protein. One subset composed of {pe4, pe5, pe8} identifies subset protein P4.
- In the second example, another protein set represented by P3 shares some peptides with the protein set represented by P2. Both protein sets have specific peptides.
- Sharing could involve many protein sets as shown in example 3.



All peptides sets and associated protein sets are represented, even if there are no specific peptides. In both cases above, no choice is done on which protein set / peptide set to keep. These protein sets could be filtered after inference (see <u>Protein sets filtering</u>).



### **Modification Dataset**

From an Identification Summary, users can ask to Identify Modification Sites and Clusters, thus creating a Modification Dataset containing the result. This dataset contains

- Modification Sites: a modification type at a given location on a given protein. They are only extracted for the
  proteins that are representative of a validated protein set. Only modifications of interest are used in this
  process.
- Modification Clusters: co-localized modification sites are grouped into clusters. Identify Modification Sites and Clusters describe this grouping process.

### Proteins and Protein sets scoring

There are multiple algorithms that could be used to calculate the Proteins and Protein Sets scores. Proteins scores are computed during the importation phase while Protein Sets scores are computed during the validation phase.

#### Protein

Each individual protein match is scored according to all peptide matches associated with this protein, independently of any validation of these peptide matches. The sum of the peptide matches scores is used as protein score (called standard scoring for Mascot result files).

#### **Protein Set**

Each individual protein set is scored according to the validated peptide matches belonging to this protein set (see <u>inference</u>).

#### **Scoring schemes**

#### **Mascot Standard Scoring**

The score associated with each identified protein (or protein set) is the sum **of the score of all peptide matches** identifying this protein (or protein set). In case of duplicate peptide matches (peptide matched by multiple queries) only the match with the best score is considered.

#### **Mascot MudPIT Scoring**

This scoring scheme is also based on the sum of all non-duplicate peptide matches score. However the score for each peptide match is not its absolute value, but the amount that it is above the threshold: the score offset. Therefore, peptide matches with a score below the threshold do not contribute to the protein score. Finally, the average of the thresholds used is added to the score. For each peptide match, the "threshold" is the homology threshold if it exists, otherwise it is the identity threshold. The algorithm below illustrates the MudPIT score computation procedure:

```
Protein score = 0
For each peptide match {
    If there is a homology threshold and ions score > homology threshold {
        Protein score += peptide score - homology threshold
    } else if ions score >= identity threshold {
        Protein score += peptide score - identity threshold
    }
}
Protein score += 1 * average of all the subtracted thresholds
```

- if there are no significant peptide matches, the protein score will be 0.
- homology and identity threshold values depend on a given p-value. By default Mascot and Proline compute these thresholds with a p-value of 5%.



- In the case of separated target-decoy searches we obtain two values for each threshold: one for the target search and another one for the decoy search. In order to obtain a single value we apply the following procedure:
  - o the homology threshold is the decoy value if it exists else the target value
  - o the identity threshold is the mean of target and decoy values.

The benefit of the MudPIT score over the standard score is that it removes many of the junk protein sets, which have a high standard score but no high scoring peptide matches. Indeed, protein sets with a large number of weak peptide matches do not have a good MudPIT score.

#### **Mascot Modified MudPIT Scoring**

This scoring scheme, introduced by Proline, is a modified version of the Mascot MudPIT one. The difference with the latter is that it does not take into account the average of the substracted thresholds:

```
Protein score = 0
For each peptide match {
    If there is a homology threshold and ions score > homology threshold {
        Protein score += peptide score - homology threshold
    }
    Protein score += peptide score - identity threshold
    }
}
```

This score has the same benefits than the MudPIT one. The main difference is that the minimum value of this modified version will be always close to zero while the genuine MudPIT score defines a minimum value which is not constant between the datasets and the proteins (i.e. the average of all the subtracted thresholds).

#### **Fisher**

This scoring scheme rely on the Fisher's test to define protein scores and p-value from the scores of the best subset of peptides. The scoring is extensively described in (<u>https://doi.org/10.1021/acs.analchem.0c00328</u>).

#### Search result validation

Once a result file has been imported and a search result created, the validation is performed in four main steps:

- 1. <u>Peptide Matches filtering and Validation</u>
- 2. Protein Inference (peptides and proteins grouping)
- 3. Protein and Proteins Sets scoring
- 4. Protein Sets Filtering and Validation

Finally, the <u>identification summary</u> issued from these steps is stored in the identification database. Different validation of a Search Result can be performed and a new Identification Summary of this Search Result is created for each validation.

When validating a merged Search Result, it is possible to propagate the same validation parameters to all childs Search Results. In this case <u>Peptide Matches filtering and validation</u> will be applied on childs as well as Protein Sets filtering. <u>Note:</u> actually, Protein Sets validation is not propagated to childs Search Results.

#### Peptide matches filtering

Peptide Matches identified in search results can be filtered using one or multiple predefined filters (described hereafter). Only validated peptide matches will be considered for further steps.



#### **Score filter**

All PSMs with a score lower than a given threshold are discarded. For some search engines, Proline computes the PSM score value itself by applying a mathematical transformation to another PSM property. For instance, the score values for PSMs from X!Tandem search results correspond to the log10 transformation of the PSMs expectation values.

#### **Pretty Rank filter**

This filter is applied after having temporarily joined target and decoy PSMs corresponding to the same query. For each query, target/decoy PSMs are then sorted by score. As in Mascot, a pretty rank is computed for each PSM depending on their ranking: PSM with almost equal score (difference < 0.1) are assigned the same rank. All PSMs with a pretty rank greater than the cut-off specified are discarded.

#### **Minimum sequence Length filter (Length)**

PSMs corresponding to peptide sequences shorter than the cut-off stipulated will be discarded when this parameter is applied.

#### Mascot e-Value filter (e-Value)

This filter is used to select PSMs based on the Mascot expectation value (e-value) which reflects the difference between the PSM's score and the Mascot identity threshold (p=0.05). PSMs with an e-value greater than the threshold specified are discarded.

#### Mascot adjusted e-Value filter (Adjusted e-Value)

Proline can compute an adjusted e-value. It first selects the lowest threshold between the identity and homology e-values (p=0.05). Then, it computes the e-value using this selected threshold. PSMs for which the adjusted e-value is greater than the specified cut-off are discarded.

#### Mascot p-Value based on identity filter (Identity p-Value)

Given a specific p-value, the Mascot identity threshold is calculated for each query and all peptide matches associated with the query for which the score is lower than the identity threshold calculated are discarded.

#### Mascot p-Value based on homology filter (Homology p-Value)

Given a specific p-value, the Mascot homology threshold is inferred for each query and all peptide matches associated with the query which have a score lower than the calculated homology threshold are discarded.

#### Single PSM per MS Query filter

This filter validates only one PSM per Query. To select a PSM, following rules are applied:

For each query:

- Select PSM with higher score.
- If several PSM with same score:
  - o Choose PSM which identify the protein which have the max number of valid PSM
  - o If still equality choose the PSM with the lower delta moz
  - o If still equality, select the first PSM using alphabetic order
  - o If still equality, select the first one. Warning : may be different in different versions of Prolline...

#### Single PSM per Rank filter

This filter selects only **one PSM per pretty rank**, which is already the case when a given pretty rank is associated with a single PSM. When multiple PSMs have the same pretty rank, the same selection than "Single PSM per Query" is used



- Select PSM that identify the protein which have the max number of valid PSM.
- If several PSM are equals
  - o Choose the PSM with the lower delta moz.
  - o If still equality, select the first PSM using alphabetic order.
  - o If still equality, select the first one. Warning : may be different in different versions of Proline...

Thus, if this filter is combined with the "Pretty rank" filter, the result obtained should be identical to the result of the "Single PSM per MS query" filter.

#### Single Sequence per Rank filter

This filter selects the best **PSM per pretty rank** and all PSMs that have the **same sequence**, without considering Post translational modification. The best PSM is selected using the same rules than Single PSM per Rank filter described above.

#### **Isotope Offset filter**

This filter validates PSMs which have been validated using an isotope offset less or equal than specified one. This corresponds to identification search engine parameters such as Mascot isotope error and could be used to invalidate PSM identified using one or more 13C.

#### Peptide matches validation

In addition to these filters a PSM level validation step may be specified using:

- A target-decoy approach (Elias and Gygi, 2007), which can be performed by adjusting a user-specified validation criterion until it reaches a user-specified false discovery rate (FDR). The search engine score can be used as a generic validation criterion for any of the search engines supported by Proline. For results obtained with the Mascot search engine, other criteria can be used to control the FDR: Mascot e-Value, Mascot adjusted e-Value, Mascot p-Value based on identity threshold, or Mascot p-Value based on homology threshold.
- A method based on the Benjamini-Hochberg procedure to control the FDR proposed in (<u>https://doi.org/10.1021/acs.analchem.0c00328</u>)

#### FDR estimation for Target Decoy validation

There are several ways to calculate FDR depending on the database search type. In Proline the FDR is calculated at PSM and protein levels using the following rules:

• if the Search has been done on a concatenated Target/Decoy bank or if rank filter has been used during validation :

$$FDR = 2 \frac{nbr \, DecoyPSM}{nbr \, TargetPSM + nbr \, DecoyPSM}$$

<u>Note</u>: when computing PSM FDR, peptide sequences matching a Target Protein and a Decoy Protein are taken into account in both cases.



• if the Search has been done on a separated Target/Decoy bank :

$$FDR = \frac{nbr DecoyPSM}{nbr TargetPSM}$$

#### Protein sets filtering

Any Identification Summary, generated by a validation process or by merging datasets could be filtered.

Filtering consists in invalidating Protein Sets which doesn't follow specified criteria. Invalidated Protein Sets are not taken into account for further algorithms or display.

Available filtering criteria are defined below.

#### **Specific peptides filter**

This filter invalidates protein sets that don't have at least x peptides identifying only that protein set. The specificity is considered at the DataSet level.

This filtering goes through all Protein Sets from worse score to best score. For each, if the protein set is invalidated, associated peptides properties are updated before going to the next protein set. Peptide property is the number of identified protein sets.

#### **Peptides count filter**

This filter invalidates protein sets that don't have at least x peptides identifying that protein set, independently of the number of protein sets identified by the same peptide.

This filtering goes through all Protein Sets. For each, if the protein set is invalidated, associated peptides properties are updated before going to the next protein set. Peptide property is the number of identified protein sets.

#### Peptide sequence count filter

This filter invalidates protein sets that don't have at least x different peptide sequences (independently of PTMs) identifying that protein set.

This filtering goes through all Protein Sets from worse score to best score. For each, if the protein set is invalidated, associated peptides properties are updated before going to the next protein set. Peptide property is the number of identified protein sets.

#### Protein set score filter

This filter invalidates protein sets which score is below a given value.

#### Protein sets validation

Once prefilters (see above) have been applied, a validation algorithm can be run to control the FDR. In the same way as for PSM, FDR may be controlled using :

- the Benjamini-Hochberg procedure, see (<u>https://doi.org/10.1021/acs.analchem.0c00328</u>)
- a target decoy strategy.

#### Target-Decoy Strategy:

See how FDR is calculated.

At the moment, it is only possible to control the FDR by changing the Protein Set Score threshold. Three different protein set scoring functions are available.



Given an expected FDR, the system tries to estimate the best score threshold to reach this FDR. Two validation rules (R1 and R2) corresponding to two different groups of protein sets (see below the detailed procedure) are optimized by the algorithm. Each rule defines the optimum score threshold allowing to obtain the closest FDR to the expected one for the corresponding group of protein sets.

Here is the procedure used for FDR optimization:

- protein sets are segregated in two groups, the ones identified by a single validated peptide (G1) and the ones identified by multiple validated peptides (G2), with potentially multiple identified PSMs per peptide.
- for each of the validation rules, the FDR computation is performed by merging target and decoy protein sets
  and by sorting them by descending score. The score threshold is then modulated by using successively the
  score of each protein set of this sorted list. For each new threshold, a new FDR is computed by counting the
  number of target/decoy protein sets having a score above or equivalent to this value. The procedure stops
  when there are no more protein sets in the list or when a maximum FDR of 50% is reached. It has to be
  noted that the two validation rules are optimized separately:
  - o G2 FDR is first optimized leading to the R2 score threshold. The validation status of G2 protein sets is then fixed.
  - o final FDR (G1+G2) is then optimized leading to the R1 score threshold. Only the G1 protein sets are here used for the score threshold modulation procedure. However the FDR is computed by taking into account the G2 validated target/decoy protein sets.

The separation of proteins sets in two groups allows to increase the power of discrimination between target and decoy hits. Indeed, the score threshold of the G1 group is often much higher than the G2 one. If we were using a single average threshold, this would reduce the number of G2 validated proteins, leading to a decrease in sensitivity for a same value of FDR.

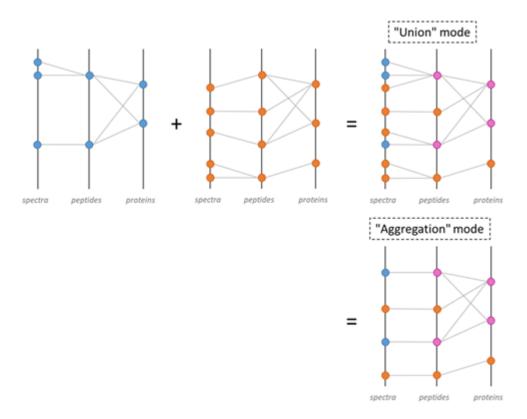
### **Combining datasets**

Identification results can be combined to construct a parent dataset, and create a non-redundant list of identified peptides and proteins. This combination can be performed either before validation (on search results) or after validation (on identification summaries). Since this operation could be recursively performed, it leads to hierarchical structuring of search results and/or identification summaries. On the one hand, combination before validation (taking into account all PSMs identified by the search engine) may, for example, be relevant when analyzing results obtained after peptide fractionation: in that case, several peptides belonging to the same protein may be spread across different result sets; these sets should be merged before protein validation. On the other hand, merging identification summaries is appropriate when seeking to group the validated results from series of individual samples to be compared or when combining data from different search engines.

When datasets are combined, their PSMs are collected to generate a non-redundant set of peptides before recomputing protein inference. Additionally, the mappings between peptides and FASTA entries observed across the different datasets are also collected and merged into a single final mapping list. This list reflects thus the whole set of peptide and protein matches that were observed in the individual datasets.

Users can combine search results or identification summaries. The main difference is the set of spectra and peptides (and thus PSMs) considered. When combining search results, all spectra, peptides and PSMs in the dataset are considered, whereas when combining identification summaries, only validated PSMs are taken into account. In addition, Proline can be used to control how PSMs are collected in the parent dataset: in union mode, PSMs originating from combined datasets are added, while in aggregation mode, all PSMs identifying the same peptide are aggregated into a single representative PSM.





Combining datasets in Proline. Datasets are represented as a tripartite graph composed of spectra, peptides and proteins; edges between spectrum and peptides represent PSM. When the blue and orange datasets are combined, PSMs from both datasets are collected together, generating a non-redundant set of peptides. The combination can be performed in 2 modes: union or aggregation mode.

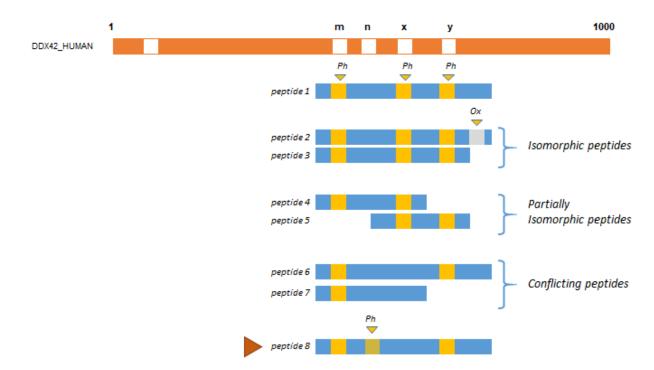
### **Identifying Modification sites and clusters**

A list of Post translational modification (PTM) sites identified among the peptides of an identification summary can be extracted by Proline. A modification site is characterized by a modification type, at a given location on a given protein. The list of modification sites extracted by the software is restricted to the proteins that are representative of a validated protein set and to the modifications of interest specified by the user. This means that a peptide identified with two different modifications of equal interest to the user will appear twice in the list, one for each modification location.

In a second phase, co-localized modification sites are grouped into clusters as soon as evidence of their co-existence exists. The required evidence is a peptide sequence identified in the dataset with all the clusterized modification sites.

The figure below represents different peptides (blue rectangles), co-localized on a protein sequence (in red). In this example, modifications of interest (Phosphorylation (Ph)) are shown in orange. Peptide 1 proves that Phosphorylation at positions m, x and y occurs simultaneously. Peptides 2 and 3 are considered as isomorphic since the oxidation of peptide 2 is ignored (only Phosphorylations have been declared of interest by the user in this example). Peptides 4 and 5 are partially isomorphic: they confirm Phosphorylation respectively at position (m, x) and (x, y). Conversely peptides 6 and 7 are not in accordance, suggesting that there are two other proteoforms, one with Phosphorylations at position (m,y) but no modification at x and one with a Phosphorylation at position m but not at x. These two peptides could be grouped into another cluster.

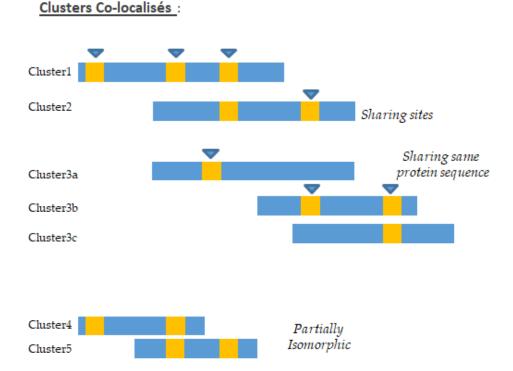




The user can choose between two different modification clustering methods:

- Partially Isomorphic Matching (aka Exact Position matching): Two sequence matches are clusterized if they are Isomorphic or Partially Isomorphic. Partially Isomorphic matches could belong to multiple clusters.
- Only Isomorphic Matching: two sequence matches are clusterized if they are Isomorphic.

In Proline, clusters co-localisation concept is used. Co-localized clusters are clusters that have overlapping on protein sequence. These clusters may share one or more sites, but they could also cover the same protein sequence without any common site. See co-localized clusters example below.



12



### **Spectral Counting**

### Principles

Proline can be used to compare protein sets based on spectral counts through a previously presented algorithm (Hesse et al., 2016). This algorithm notably computes a weighted spectral count metric (called adjusted spectral count in the original publication). Basically, the algorithm takes both unique and shared peptides into account, and for each shared peptide, the proportion of MS/MS spectra that should be attributed to the different protein sets is determined. This proportion (also called weight) is based on the spectral counting of proteotypic (or specific) peptides identifying the different protein sets sharing the peptide to be attributed.

Spectral counting is calculated from a hierarchy of identification summaries. The parent identification summary (at the top of the hierarchy) is where the list of protein sets to compare and the list of specific peptides are created. The list of specific peptides is then used to compute the protein sets respective weights but users can choose any "child" dataset where the weights must be calculated.

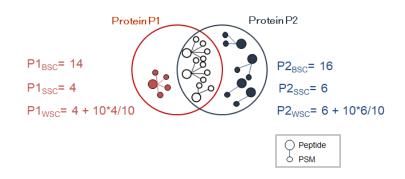
First, Proline compute the peptide spectral count at each level of the dataset hierarchy using the following rules:

- If the dataset is a "leaf" identification summary of the hierarchy (not issued from a merge, no child dataset), the peptide spectral count is the number of MS/MS spectra matching a peptide (equal to the number of peptide spectrum matches).
- If the dataset is a merged identification summary, the peptide spectral count is the sum of it's child peptide spectral count.
- If the dataset is a merged of search results, the peptide spectral count is the sum of validated children peptide spectral count. Validated children are PSMs that meet validation criteria applied to parent Identification Summaries.

Once, peptide spectral count is calculated for each peptide, the protein spectral count is computed using the following rules:

- Protein basic spectral count (BSC) is the sum of the peptide spectral count of all peptides matching a protein.
- Protein specific spectral count (SSC) is the sum of the peptide spectral count of specific peptides. A specific peptide, is a peptide which does not identify any other protein sets in the context of the identification summaries.
- Protein weighted spectral count (WSC) is the Protein specific spectral count (SSC) + weighted peptide spectral count of shared peptides.

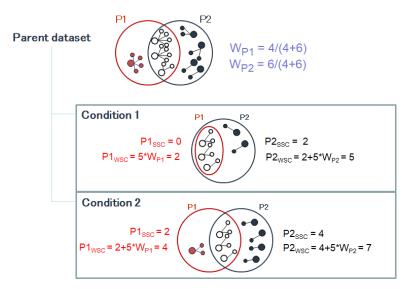
#### Simple example





### Protein sets weights

The protein set respective weights computation is based on the proteotypic peptides. The level in the dataset hierarchy where these weights are calculated can be chosen by the user, it could be the top level of the dataset hierarchy or at a lower level. In the following example, the weights ( $W_{P1}$  and  $W_{P2}$ ) are calculated at the "parent dataset" level. At this level, P1 and P2 are respectively identified by the red and dark blue peptides/psm, each protein set weight is calculated using their specific spectral counting. These weights are thus used to calculate the weighted spectral count of P1 and P2 in the two "child" dataset "Condition 1" and "Condition 2".



### FAQ

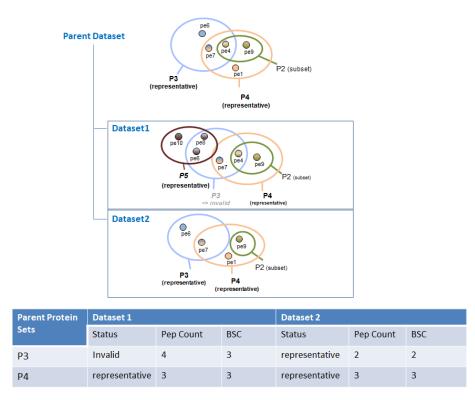
#### Why is the BSC less than Peptide Count ?

When running SC even on a simple hierarchy (1 parent, 2 childs) in some cases we obtain a BSC value smaller than the peptide count of the protein set. This occurs only for *invalid protein sets*. *Invalid protein sets* are the one that are present at the parent level but are filtered at child level (if a *specific peptide* filter have been applied for example).

Indeed, the peptide count value is read in the child protein sets. On the other hand, the BSC is calculated by getting the spectral count information at child level for each peptide identified at parent level. If a protein set is invalid, its peptides are not taken into account during the merging so some of them could be missing at parent level if they were not identified in the other child.

This case is illustrated in the following figure:





### Label-free LC-MS quantitation workflow

Proline detects chromatographic peaks from raw data **converted to the mzDB format** (Bouyssié et al., 2015). The <u>converter</u>, named <u>raw2mzdb</u> is based on ProteoWizard, ensuring compatibility with a wide range of instrument vendors.

After a first **signal extraction** step, the algorithm associates the chromatographic peaks detected with validated PSMs, first by retrieving the corresponding MS/MS spectra acquired during the peptide elution, and then by **matching the precursor** m/z value of these spectra to the chromatographic peak m/z value. After the **deisotoping** step, the abundance of each ion is estimated from the apex of the chromatographic peak, which corresponds to the theoretically most abundant isotopologue (inferred from the peptide's atomic composition). The software then **aligns** the retention time of these annotated ions for all the LC-MS runs to be compared, and uses this information to **cross-assign** MS/MS data to ions (i.e. chromatographic peaks) that were detected but not identified in other runs . The resulting ion abundances are finally stored in the Proline database, making them available for rapid data visualization and further post-processing.

Finally, peptide ion measurements can be **summarized as protein abundances** using different computational methods. The user can opt to perform additional operations such as excluding peptides or ions based on their characteristics (missed cleavages, variable modifications, sequence specificity, etc.) or normalizing peptide and protein abundances between runs. These post-processing steps can be executed on-demand using different parameters or methods; there is no need to repeat the whole quantification process when changes are made.

### Signal extraction

During an LC-MS experiment, the m/z and intensity values for each peptide ion detected are recorded in MS1 scans acquired during the elution of this peptide from the chromatographic column. Most existing peak picking algorithms analyze these MS scans individually or sequentially. The Proline algorithm performs the signal detection in a



different way. It first takes advantage of the mzDB format to detect chromatographic peaks in spectrum slices (5 m/z wide by default) across the whole chromatographic time. Then, in a given slice, m/z peaks are sorted in decreasing order of intensity. Thus, starting from the most intense m/z peak (apex), the algorithm searches for a peak with the same m/z value in the previous and subsequent MS1 scans, while applying a user-defined m/z tolerance. This lookup procedure stops when the ion signal is absent from more than a predefined number of consecutive scans. The [RT, m/z, intensity] peak list obtained, which is comparable to an extracted ion chromatogram (XIC), is then smoothed using a Savitzky-Golay filter (Savitzky and Golay, 1964). The resulting smoothed chromatogram is then split into the time dimension to form chromatographic peaks, by applying a peak picking procedure that will search for significant minima and maxima of signal intensity. When the signals of two ions overlap in the time dimension, a minimum is generally surrounded by two maxima. If the corresponding valley is deep enough, i.e. at least 66% of the lower surrounding maximum, this minimum will be considered significant (and thus will trigger the generation of two peaks). Once the smoothed chromatogram has been fully analyzed, the algorithm removes the corresponding detected peaks from the current spectrum slice, and performs another lookup using the next available apex. The result of this whole procedure is a list of chromatographic peaks defined by an m/z value, an apex elution time and an elution time range.

#### Parameters

These parameters are used by signal extraction algorithms.

- Extraction m/z tolerance: this corresponds to the error tolerance used when searching for a m/z peak in the previous and subsequent MS1 scans (see above).
- Use previous/last peakel detection: If a quantitation has already been done on a run, use the previously detected peakels file for this run. **Warning**: In this case, the extraction m/z tolerance parameter is ignored since the signal extraction is already done !

### PSM assignment and deisotoping

In a single run, validated PSMs are MS/MS spectra assigned to a peptide sequence, and each spectrum is characterized by a precursor mass, a charge state and a retention time (RT). The algorithm assigns PSMs to detected chromatographic peaks by matching the spectrum precursor m/z ratio to the chromatographic peak m/z and verifies that the spectrum retention time falls within the peak time range. The PSM charge state assigned is then used to search for chromatographic peaks corresponding to the ion's isotopologues, considering the precursor mass-to-charge ratio of the spectrum as the monoisotope. The peptide ion intensity is summarized by retaining only the apex of the peak corresponding to the theoretically most abundant isotopologue (inferred from the peptide's atomic composition). All peptide ion signals (a.k.a. LC-MS features) extracted from an mzDB file and assigned to a PSM are used to construct an LC-MS map. For the sake of simplicity no distinction is made between LC-MS runs and LC-MS maps in this manuscript.

#### Parameters

• PSM/Peakel matching m/z tolerance is the m/z error allowed to match a peakel to a PSM

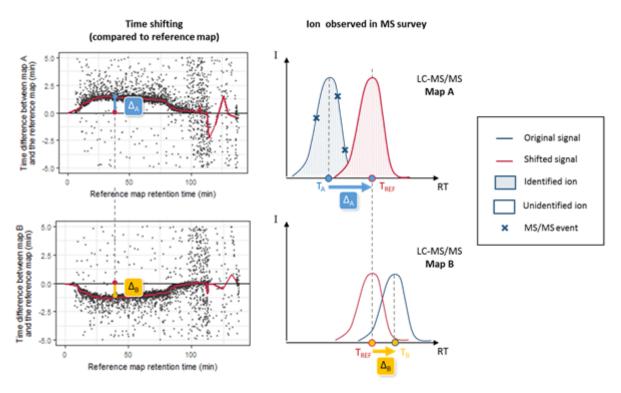
Clustering is applied to group peakels that are matching the same identified ion.



- Cluster time computation: you have the choice between 2 computation methods: most intense or median. For most intense method, the cluster time corresponds to the time of the most intense feature composing the cluster. For the median method, cluster time is the median of the feature times forming the cluster.
- Cluster intensity computation: you have the choice between 2 computation methods: most intense or sum.
   For most intense method, the cluster intensity corresponds to the intensity of the most intense feature of features forming the cluster. For the sum method, cluster intensity is the sum of the intensities of features composing the cluster.

#### Retention time alignment

As soon as the PSM are matched to peakels, the software then aligns the retention time of the annotated ions for all the LC-MS runs to be compared, and uses this information to cross-assign MS/MS data to ions (i.e. chromatographic peaks) that were detected but not identified in other runs. Because chromatographic separation is not completely reproducible, LC-MS runs must be aligned. The retention time alignment procedure is a critical step in MS1 label-free quantification. Proline's alignment algorithm selects a **reference run** and generates a set of functions that will be used to predict the RT (retention time) for missing features from another run. These functions are obtained by performing pairwise alignments between the different runs to be compared (Bylund et al., 2002; Jaitly et al., 2006; Sadygov et al., 2006). The first step consists in computing a scatter plot (see Figure below) of the observed time difference between the two runs as a function of the reference run's time-scale. This mapping can be based on the **peptide identity** (same sequence and same post-translational modifications) of identified features, or by mapping the detected features of the pair of runs, taking user-defined time and mass error ranges into account (the default feature mapping time and m/z tolerance values are set to 600 seconds and 5 ppm, respectively).



**RT prediction using computed alignments.** The two scatter plots on the left correspond to computed run alignments between the reference run and two other runs (A and B). The red curves on these plots correspond to the median RT prediction for each alignment, obtained by applying a moving median calculation. The graphs on the right illustrate the case of a peptide ion that is present in runs A and B, but has only been fragmented by MS/MS in run A. Knowing the retention time in run A ( $T_A$ ), we can predict  $T_B$  by two consecutive time conversions.  $T_A$  is first converted to the reference run scale ( $T_{REF} = T_A + D_A$ ) using the first run alignment, then to  $T_B = T_{REF} + D_B$  using the alignment for the second run.



RT prediction functions are then obtained by **smoothing** these scatter plots using a moving median calculation or a local regression. To decrease the number of alignment combinations (i.e., pairs of maps), the reference run is determined by an iterative method. The algorithms begin by selecting a random run as a reference and compute all alignments against this map. The algorithm then determines a new reference run by selecting the run with the smallest sum of RT differences in the resulting run alignments. The iteration stops after a user-specified maximum number of iterations or when the reference run remains unchanged between two iterations. The software can also be configured to compute all possible RT alignment combinations (all possible pairs of maps, "exhaustive" option), but this can be very computationally expensive when there is a high number of maps to be compared. The alignments computed can then be used to predict the retention times for peptide ions in a specific sample where they were not identified.

#### Parameters

#### Alignment

- Map Alignment: alignment can be disabled. In this case the cross assignment step could be executed but without any retention time prediction
- Method : is the name of the method used to determine the reference run
  - Exhaustive: the exhaustive algorithm computes the distance between maps for each possible couple of maps and selects the map with the lowest sum of distances to be the reference map.
  - Iterative: the algorithm iteratively selects the best reference run as described above.
    - Maximum number of iterations: this option is available only for the iterative method. This is
      a stop condition of the iterative algorithm, when the algorithm has reached its maximum
      number of iterations, it stops.
- Ignore alignment error : allow or not to continue quantitation even if there are some errors in alignment.

#### Feature mapping

- Feature mapping is needed to build the alignment scatter plot of the observed time difference between two runs. The method used could be:
  - Feature coordinates : the mapping is based on features RT and m/z coordinates taking into account a tolerance for each of these values.
    - m/z tolerance : m/z window used to match features between two compared maps.
  - Peptide Identity : the mapping is based on peptides identification (same sequence and same post-translational modifications)
- Time tolerance in seconds: time window used to match features/peptides between two compared maps.

#### **Alignment smoothing**

When features of two runs are matched, a trend can be extracted from the scatter plot by using a smoothing method.

- Smoothing method: you have the choice between three smoothing methods: LOESS, time window or landmark range:
  - LOESS : no specific parameter
  - Landmark range



- Number of landmarks/window size: time of aligned map is corrected using median computed on windows containing a specified number of landmarks. The run is divided into windows of size the specified number of landmarks. You have to provide the number of landmarks by window. The smoothing method is applied considering the number of landmarks present in the window, and computes the median point for this window.
- Sliding window overlap: overlap is used to compute the step to move the smoothing window forward to calculate a smoothing point for this new smoothing window. Overlap gives the percentage of overlapping between two consecutive windows. For example, if window size is 200 (seconds or landmarks depending on which smoothing method is selected) and overlap is 20%, the step forward = 200\*((100-20)/100) = 160 seconds or landmarks, i.e. the smoothing window is moved forward by a step of 160, so two successive windows overlaps each other by a step of 40 seconds or landmarks corresponding to 20% of 200
- Time window : time of aligned map is corrected using median in a time window.
  - Window size/time interval: You have to provide the time interval. This time interval corresponds to the window size in which time median will be computed.
  - Minimum number of landmarks: This allows you to specify the minimum number of landmarks a window must contain to compute a median on it, it is not significant to compute a median on less landmarks.
  - Sliding window overlap: overlap is used to compute the step to move the smoothing window forward to calculate a smoothing point for this new smoothing window. Overlap gives the percentage of overlapping between two consecutive windows. For example, if window size is 200 (seconds or landmarks depending on which smoothing method is selected) and overlap is 20%, the step forward = 200\*((100-20)/100) = 160 seconds or landmarks, i.e. the smoothing window is moved forward by a step of 160, so two successive windows overlap each other by a step of 40 seconds or landmarks corresponding to 20% of 200.

### **Cross Assignment**

Proline uses a hybrid approach to retrieve intensity values for ions that were not identified. As indicated above, identified and quantified features are obtained by detecting chromatographic peaks in raw files without a-priori, using an identification-based deisotoping method. During the PSM assignment step, the identification data provides the monoisotopic mass and the charge state for the ion, guiding the deisotoping procedure to group together the detected chromatographic peaks. These grouped peaks are then removed from the list of peaks to be assigned, thereby reducing the data density when annotating subsequent chromatographic peaks during the cross-assignment step. Ions that were not identified in a run can then be sought out in this restricted list of detected peaks using their m/z and RT coordinates. The m/z value is the theoretical m/z value obtained following identification of this ion in another run. The RT value is predicted from the apex RT of the feature detected in the run providing the highest identification score. This RT prediction computation may involve the use of one or two alignment functions. Using these two coordinates, associated with user-defined m/z and RT tolerances, the algorithm seeks a corresponding signal among the chromatographic peaks that have not already been assigned to an identified ion. To avoid the propagation of erroneous cross-assignments between runs, an additional but optional control (named "use only confident features") is applied to ensure that this peak is the monoisotope of a peptide ion with a charge state identical to the master feature one. This is done by fitting its observed isotope pattern to a theoretical one.



#### Parameters

- allow cross assignment between all runs or only between runs of the same groups: the user can completely disable the cross assignment or can control runs between which cross assignment is applied:
  - Between all runs
  - Within groups only: cross assignment is applied only between runs belonging to the same group.

#### **Feature mapping**

- m/z tolerance (ppm or dalton): when mapping features from two different maps of the map set, delta m/z between features must be lower than the m/z tolerance to be considered as the same feature seen on two different maps.
- Time/RT tolerance (seconds): when mapping features from two different maps of the map set, delta time between features must be lower than the time tolerance to be considered as the same feature seen on two different maps.

#### Filtering/Correction

- Use only confident features: is applied to ensure that the cross assigned peak is the monoisotope of a peptide ion with a charge state identical to the master feature one. This is done by fitting its observed isotope pattern to a theoretical one.
- Master feature intensity filter (optional): a filter can be applied to the map features to keep the best features (above threshold) to build the master map.
  - Two methods are available to filter features: the filter can be applied directly on intensity values (*Intensity method*) or it can be a proportion of the map median intensity (*Relative intensity method*).
  - intensity threshold/value: this provides the threshold for the filtering method. Only features above or below (depending on the *operator*) this threshold are considered for the master map building process.

If you choose Relative intensity for master feature filter type, the only possibility you have is percent, so features which intensities are beyond the relative intensity threshold in percentage of the median intensity are removed. If you choose Intensity for master feature filter type, you also have only one possibility at the moment of the intensity method: basic. Features which intensities are beyond the intensity threshold are removed and not considered for the master map building process.

### Normalizing LC-MS maps

The comparison of LC-MS maps is confronted to another problem which is the variability of the MS signals measured by the instrument. This variability can be technical or biological. Technical variations between MS signals in two analyses can depend on the injected quantity of material, the reproducibility of the instrument configuration and also the software used for the signal processing. The observed systematic biases on the intensity measurements between two successive and similar analysis are mainly due to errors in the total amount of injected material in each case, or the LC-MS system instabilities that can cause variable performances during a series of analysis and thus a different response in MS signal for peptides having the same abundance. Data may not be used if the difference is too important. It is always recommended to do a quality control of the acquisition before considering any computational analysis. However, there are always biases in any analytic measurement but they can usually be fixed by normalizing the signals. Numerous normalization methods have been developed, each of them using a different mathematical approach (Christin, Bischoff et al. 2011). Methods are usually split in two categories, linear and



non-linear calculation methods, and it has been demonstrated that linear methods can fix most of the biases (Callister, Barry et al. 2006). Three different linear methods have been implemented in Proline by calculating normalization factors as the ratio of the sum of the intensities, as the ratio of the median of the intensities, or as the ratio of the median of the intensities.

#### Sum of the intensities

How to calculate this factor:

- 1. For each map, sum the intensities of the features
- 2. The reference map is the median map
- 3. The normalization factor of a map = sum of the intensities of the reference map / sum of the intensities of the map

#### Median of the intensities

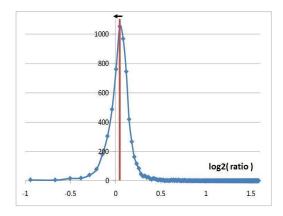
How to calculate this factor:

- 1. For each map, calculate the median of the intensities in the map
- 2. The reference map is the median map
- 3. The normalization factor of a map = median of the intensities of the reference map / median of the intensities of the map

#### Median of ratios

This last strategy has been published in 2006 (Dieterle, Ross et al. 2006) and gives the best results. It consists in calculating the intensity ratios between two maps to be compared then set the normalization factor as the inverse value of the median of these ratios (cf. figure below). The procedure is the following:

- 1. For each map in a "map set", sum the intensities of the features
- 2. The reference map is the median map
- 3. For each feature of the master map, ratio = intensity of the feature in the reference map / intensity of the feature for this map
- 4. Normalization factor = median of these ratios



Distribution of the ratios transformed in log2 and calculated with the intensities of features observed in two LC-MS maps. The red line representing the median is slightly off-centered. The normalization factor is equal to the inverse of this median value. The normalization process will refocus the ratio distribution on 0 which is represented by the black arrow



Proline makes this normalization process for each match with the reference map and has a normalization factor for each map, independently of the choice of the algorithm. The normalization factor for the reference map is equal to 1.

### **Post-processing of LC-MS quantitative results**

This procedure is used to compute peptide and protein abundances. Several filters can also be set to increase the quality of quantitative results.

Here is the description of the parameters that can be modified by the user.

#### **Peptide filters**

- **Peptides selection method:** define which peptides are to be considered for quantitation.
  - All peptides: all peptides are taken into account at this step. Further filters may invalidate some peptides.
  - **specific peptides**: peptides shared between different protein sets are discarded for protein set abundance calculation and statistical analysis.
  - **Razor and specific:** In addition to specific peptides, quantitation of shared peptides are considered for the "best" ProteinsSet. The best is actually based on the score.
- **Discard missed cleaved peptides**: if checked, peptides containing missed cleavages are discarded from the statistical analysis. It has to be noted that perfect tryptic peptides whose sequence is included in an observed missed cleaved peptide are also discarded if this option is enabled.
- **Discard modified peptides**: if checked, peptides containing specific modifications are discarded for protein set abundance calculation and the statistical analysis.
  - only those modified peptides or also non-modified peptides whose sequence is the same as these peptides may be discarded depending on the chosen filtering method.

#### Summarize peptides ions into peptide abundance

To calculate peptide abundance, associated peptide ion abundances can be summed or the best ion is used. To choose the best peptide ion following rules are defined :

- Select the peptide ions identified and quantified in the maximum of quant channels(runs)
- If more than one, select peptide ions with the maximum of PSMs count
- If more than one, choose the peptide ion with the higher abundance.

#### Peptide and protein common parameters

• Normalization: the normalization factors are computed as the median of the ratios distributions between each run and a run of reference. A similar procedure is used for the <u>normalization of LC-MS features.</u>

#### Summarize peptides abundances into protein abundance

Peptide abundances can be summarized into protein abundances using several mathematical methods:

- **sum**: for each quantitative channel (raw file) the sum of observed peptides abundances is computed
- mean: for each quantitative channel (raw file) the mean of observed peptides abundances is computed

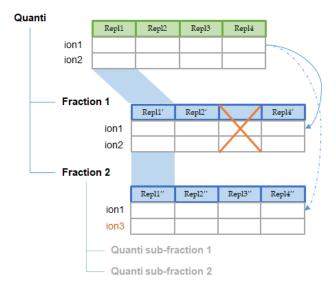


- mean of TOP3 peptides: same procedure but applied on the 3 most abundant peptides. Peptides are sorted by descending median abundances (computed across all compared samples for peptide). Then the 3 first peptides are kept.
- median: for each quantitative channel, the median of observed peptides abundances is computed
- **median profile**: a matrix of peptide abundance ratios is first computed (rows correspond to peptides and columns to quantitative channels). The median of these ratios is then computed for each column. The relative values are then converted back into absolute values using a scaling factor. This factor is computed as the maximum value from the **means of TOP3 peptides** abundances.
- **normalized median profile**: matrix of peptide abundance ratios is first computed (rows correspond to peptides and columns to quantitative channels). This matrix is then normalized and then summarized using the median method described above. The obtained median abundances are then adjusted by using a scaling factor. This factor is computed as the maximum value from the **means of TOP3 peptides** abundances.
- Median Ratio Fitting: MRF computes a matrix of abundance ratios calculated between any two runs from ion abundances for each protein. For each pair-wise ratio, the median of the ion ratios is then calculated and used to represent the protein ratio between these two runs. A least-squares regression is then performed to approximate the relative abundance of the protein in each run in the dataset. This abundance is finally rescaled to the sum of the ion abundances across runs.

### **Aggregation of Label-Free quantitative results**

Two or more quantitations can be combined such that an ion quantified in multiple aggregated quantitations is represented only once in the aggregation result. The abundance of this ion is a combination of its abundance measured in the different aggregated quantitations. This could be useful to combine for example quantitation of fractions into a single quantitation result.

The experimental design of the aggregation is based on the experimental design of the aggregated quantitation: the number of group/condition and the number of replicates per condition remains the same. However, the user can modify the correspondence between the groups and replicates if needed. In the following example the abundance of ion1 in the aggregated quantitation (in green) is based on the quantitation of this same ion in "Fraction 1" and "Fraction 2". Since ion2 is quantified only in "Fraction 1", its abundance values in the aggregated quantitation are the same as the abundances measured in "Fraction 1".





In this simple example, the correspondence between experimental designs is such that the abundance of ion1 in the replicate Repl1 is based on the measured abundance of ion1 in Repl1' in "Fraction 1" and Repl1'' in "Fraction 2". This could be modified by the user to take into account differences in the replicates order in aggregated quantitations or to account for the absence of a replicate (see for example replicate 3 in "Fraction 1").

In the current version, the abundance at the aggregation level is the sum of the abundances in aggregated quantitations.

### **Identification Summary Export**

When exporting a whole Identification Summary in an excel file, the following sheets may be generated:

- Search settings and info : Contains information on project and search settings parameters
- Import and filters : Summary of used parameters during import, filtering and validation process
- *Protein sets* : List of all Protein Sets, valid or invalidated (configurable in custom option) during <u>Protein Sets</u> <u>Filtering</u>. Some columns description :
  - o *#sequences* (*#specific sequences* ) : number of different peptide sequences identifying the Protein Set (specific : which does not identify any other valid Protein Set )
  - o *#peptides* (*#peptides*) : number of different peptide (sequence + PTM) identifying the Protein Set (specific : which does not identify any other valid Protein Set )
  - o *#spectral\_count* (*specific\_spectral\_count*) : number of different peptide spectrum matches identifying the Protein Set (specific : which does not identify any other valid Protein Set )
- *Protein matches in protein set* : list of Protein Matches in each Protein Set. A same Protein Match could thus appear a few times if it belongs to different Protein Sets. (same column as protein set)
- *Best PSM from protein sets* : List of best peptide spectrum matches (a single PSM per peptide is listed) for each Protein Set. Some columns description :
  - o *#psm\_prot\_sets* : number of Valid Protein Sets identified by this PSM.
  - o *#psm\_prot\_matches* : number of Protein Match, which belong to at least 1 valid Protein Set, identified by this PSM.
  - o #psm\_db\_prot\_matches : number of Protein Match, validated or not, identified by this PSM. This is equivalent to the number of proteins in fasta files containing the PSM.
- All PSMs from protein sets : List of all peptide spectrum matches for each Protein Set. (same column as best PSM from protein sets)
- PTM Cluster (filled only if "identify Modifications sites" has been run): All clusters (protein-set of localised ptms) identified in this identification summary.
- *Dataset statistics* : Some statistical values for the exported Identification Summary : number of Protein Set, modified peptides ...

### **Advanced features**

### Allowing multiple imports

By default, Proline does not allow importing multiple results files at the same time. This was due to the obsolete Peptide database which has been removed in Proline version 2.0. Nevertheless, if multiple results files could now be imported at the same time there is still a restriction at project level. Indeed, result files should be inserted sequentially in the same project. In addition, depending on Proline Server configuration, there could not be as many import threads as wanted or as the number of projects.



This is why a configuration allows the user (administrator) to specify groups of projects which will belong to the same "import thread". For instance, if Proline Server enables 3 threads for import services, you can specify 2 groups of projects:

- Import Thread 1 : project 1, project 3, project 5
- Import Thread 2 : project 6, project 7
- Import Thread 3 : all other project (and new projects)

This means that for projects 1, 3 and 5 imports will be sequential but in parallel of import to project 6 or 7...

If you specify more groups than allowed thread, some project groups will be grouped into a single one.

To define a project group, you should add the following string to *project.serialized\_properties* in uds\_db database. Currently this could be done only with PGAdmin (or any database server administration tool) :

{"import\_group":"1"}

In the following example, the user has defined 3 specific groups, and the default group is always defined for all other projects.

9	uds_d	b [	📒 Lo	ocalhost 📑 project 🛛						
■	🖽 Propriétés 🖳 Données 📙 ER Diagram 🦉 uds_db 🌻									
⊞	🚍 project 💈 Entrez une expression SQL pour filtrer les résultats (utilisez Ctrl+Espace)									
Grille		12 <mark>3</mark> id	۲:	ABC name	1 ABC description	🥑 creation_timestamp - 👣	<b>ABC</b> serialized_properties $\[mathcal{T}\]$	123 owner_id 🏾 🕄 🕻		
ق	1		26	20220211_XPLnoFAIMS_import parallèle	20220211_XPLnoFAIMS_import	2022-02-11 16:32:17	{"import_group":1}	2 🗹		
▦	2		27	20220211_XPLFAIMS_import parallèle	20220211_XPLFAIMS_import part	2022-02-11 16:33:14	{"import_group":2}	2 🗹		
e	3		28	20220211_Batch4_import parallèle	20220211_Batch4_import paralle	2022-02-11 16:33:36	{"import_group":3}	2 🗹		

# Raw2mzdb

### Raw file conversion to mzDB

The conversion is done using <u>raw2mzDB</u>.

#### Installation

- 1. get the zip archive on Proline download section
- 2. install of MSFileReader from Thermo (<u>here</u>, will install all necessary c++ redistribuables)
- 3. ensure your regional settings parameters are '.' for the decimal symbol and ',' for the list separator

#### Use case procedure

Open a command line window in the directory containing raw2mzdb.exe

Enter:

#### raw2mzdb.exe -i <rawfilename> -o <outputfilename>



By default, the raw file will be converted in the "fitted" mode for the MS1 (MS2 is often in centroid mode and can not be converted in fitted mode). If the MS2 (or superior) are acquired in high resolution (i.e in profile mode), you could specify that you want to convert several MSs in the required mode: **raw2mzdb.exe** -**i** <**rawfilename**> -**o** <**outputfilename**> -**f** 1-2 will try to convert MS1 to MS2 in fitted mode.

There are two other available conversion modes:

- 1. "profile", the command line is then: raw2mzdb.exe -i <rawfilename> -o <outputfilename> -p 1 (means you want profile mode for MS1, others MS will be stored as they were stored in the raw file)
- 2. "*centroid*" : **raw2mzdb.exe -i <rawfilename> -o <outputfilename> -c 1** (means you want centroid mode for MS1, others MS will be stored as they were stored in the raw file)



# **Proline Studio**

### How to

Note: Read the <u>Concepts & Principles documentation</u> to understand main concepts and algorithms used in Proline.

List of Abbreviations

#### **Creation/Deletion**

- Open a session and access to my projects
- <u>Create a new project</u>
- <u>Create a Dataset</u>
- Import a Search Result
- Delete Data and clean project
- <u>Connection Management</u>

#### Display

- Display MSQueries, Peptides/PSM or Proteins of a Search Result
- <u>Display MSQueries</u>, PSM, Peptides, Protein Sets, PTM Protein Sites or Adjacency Matrices of an Identification Summary
- Display Modification Sites or Clusters of a Modification Dataset
- Display Search Result & Identification Summary Properties
- Display Spectral Counts
- Display XIC
- <u>Create and Save a User Window</u>
- Frame Toolbars Functionalities
- <u>Filter tables</u>
- <u>Search tables</u>
- Graphics : Scatter Plot / Histogram
- <u>Statistical Reports (MSDiag)</u>
- <u>MS Files (MzScope)</u>
- <u>General Settings</u>

#### Save, import and export

- Import Search Result file (Mascot/OMSSA/X!Tandem/MzIdent/MaxQuant)
- Export data
- Generate Spectrum Matches

#### Algorithm and other operation

- Validate a Search Result
- Filter Protein Sets
- <u>Change Typical Protein of a Protein Set</u>
- <u>Combine datasets</u>



- Data Analyzer
- <u>Calculator</u>
- Update Spectrum using Peaklist software
- Identify Modification Sites/Cluster
- Annotate and Edit Modification Clusters
- <u>Administration</u>

#### Quantitation

- <u>Spectral Count</u>
- <u>XIC Quantitation</u>
- <u>XIC Aggregation</u>
- <u>Compute Post Processing on abundances</u>



### **List of Abbreviations**

Calc. Mass: Calculated Mass Delta MoZ: Delta Mass to Charge Ratio Exp. MoZ: Experimental Mass to Charge Ratio Ion Parent Int.: Ion Parent Intensity Missed Cl.: Missed Cleavage Modification D. Mass: Modification Delta Mass Modification Loc.: Modification Location Next AA: Next Amino-Acid Prev. AA: Previous Amino-Acid Protein Loc.: Protein Location of the Modification Protein S. Matches: Protein Set Matches PSM: Peptide Spectrum Match **PTM:** Post Translational Modification PTM D. Mass: PTM Delta Mass **RT**: Retention Time SC: Spectral Counting

### **Server Connection**

When you start Proline Studio for the first time, the Server Connection Dialog is automatically displayed.

Server Connection
Server Parameter Server Host Host
User Parameters User : Username
Password : ••••••

You must fill the following fields:

- Server Host: this information must be asked to your IT Administrator. It corresponds to the Proline server name
- User: your username (an account must have been previously created by the IT Administrator).
- Password: password corresponding to your account (username).



If the field "Remember Password" is checked, the password is saved for future use. Server connection dialog continues to open with Proline Studio, the user though does not need to fill in his password, unless the last one is changed after his last login.

### **Create a New Project**

Projects %	-
< Select a Project >	-
	7

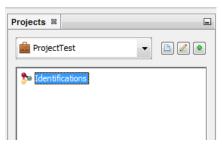


To create a Project, click on "+" button at the right of the Project Combobox. The Add Project Dialog opens. Fill the following fields:

- Name: name of your project
- Description: description of your project

You can specify other people to share this new project with them. Then click on OK Button

Creation of a Project can take a few seconds. During its creation, the Project is displayed grayed with a small hourglass over it.



### **Create a Dataset**

In the Identification tree, you can create a Dataset to group your data

To create a Dataset:

- right click on Identifications or on a Dataset to display the popup.
- click on the menu "Add Dataset..."



Projects × Ms	Files	
Proline_Projec	t	
🏂 Identification	s	
	Display Search Result	>
📩 💼 Trash	Display Identification Summary	>
	Add Dataset	
	Add Identification Folder	
	Copy Search Result	
	Paste Search Result	
	Rename	>
	Clear	
	Delete	
	Import Search Result	
	Import MaxQuant Result	
	Validate Search Result	

On the dialog opened:

- fill the name of the Dataset
- choose the type of the Dataset
- optional: click on "Create Multiple Datasets" and select the number of datasets you want to create

Add Datase	t		23
Dataset	Parameters		
Name:	Replicate		
Type:	Biological Sample		<b>-</b>
		Create Multiple Datasets	3 🜩
Defa	ult	🖌 ОК 💽 🗸 Са	ancel

Let's see the result of the creation of 3 datasets named "Replicate":

Projects 🕷	•
ProjectTest	
Identifications          Identifications         Imported         Replicate 1         Replicate 2         Replicate 3         Trash	

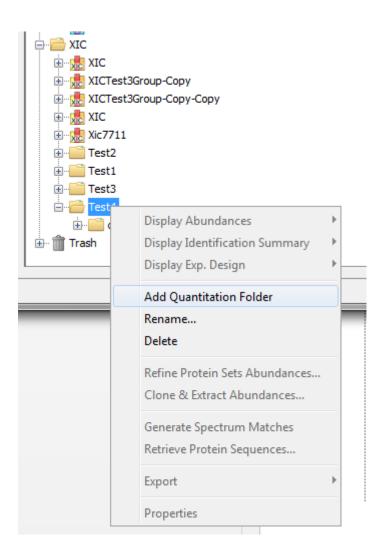


### **Create a Folder**

In both Identification and Quantitation tree, you can create Folders to organize your data

To create a Folder :

- right click on Identifications, Quantitations or on a Folder to display the popup.
- click on the menu "Add Identification Folder..." or "Add Quantitation Folder..."



### **Import a Search Result**

See Concept & Principle section

### Import Mascot/X!Tandem/OMSSA/MzIdentML

There are two possibilities to import Search Results:

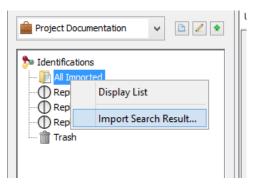
- import multiple Search Results in "All Imported" and put them later in different datasets.
- import directly a Search Result in a dataset.



#### Import in "All Imported"

To import in "All Imported":

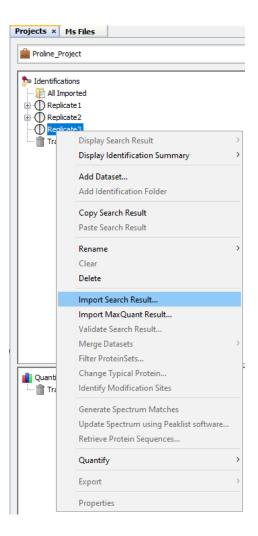
- right click on "All Imported" to show the popup
- click on the menu "Import Search Result..."



#### Import directly in a Dataset

It is possible to import Search Results directly in a Dataset. Even in this case, Search Results are available in "All Imported".

To import a Search Result in a Dataset, right click on a dataset and then click on "Import Search Result..." menu. Same dialog and parameters as in "Import in "All Imported"" above will be displayed.





#### **Import Search Result Parameters**

In the Import Search Results Dialog:

- select the file(s) you want to import thanks to the file button (the Parser will be automatically selected according to the type of file selected)

- select the different parameters (see description below)

- click on OK button

**Note 1**: You can only browse the files accessible from the server according to the configuration done by your IT Administrator. Ask him if your files are not reachable. (Look for *Setting up Mount-points* paragraph in <u>Installation & Setup</u> page).

Note 2: Proline is able to import OMSSA files compressed with BZip2.

#### Parameters description:

- Software Engine: the software which generated your Result File (this parameter will be automatically set when files are selected or you can select it)
- Instrument: mass-spectrometer (with specific configuration) used for sample analysis
- Fragmentation Rule Set: The fragmentation rules specified in software. by clicking on in you could visualize all rules for a specific rule set. This is necessary to generate spectrum matches.
- Peaklist Software: the software used for the peaklist creation (mandatory for LCMS XIC quantitation)
- Decoy Strategy: The type of decoy search which was performed.
  - o "No Decoy": if the search was performed against a target database only.
  - o "Concatenated Decoy": if target and decoy sequences were merged into a single database.
  - o "Software Engine Decoy": if the decoy sequences were generated on-the-fly by your search engine.
  - o Decoy Accession Regex: for concatenated searches only. Select the rule to apply for the discrimination of target and decoy protein matches.
- Parser Parameters: according to your Software Engine, some extra-parameters are displayed:
  - o Mascot:
    - Subset Threshold: the percentage of score between a given protein match and the master protein match (superset). Protein matches with a relative score lower than
       Master\_protein\_score \* (1-subset threshold)
      - won't be imported.
  - Omssa:
    - Usermods file path: an XML file containing the definitions for each user defined PTM used in the OMSSA search.
    - PTM Composition file path: a text file containing the chemical composition for each user defined PTM. This is required for PTMs not already imported in another Search Result. The format is the following:

PTM name=<PTM composition> (one per line).

Example: Acetyl peptide N-term=H(-6) C(-7) O(-1)

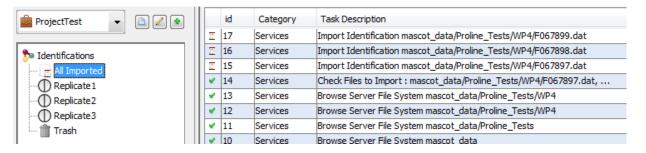
• X!Tandem:



■ Protein parsing rule: Specify the rule (regular expression) to be used to retrieve protein accession from protein info. As an example, to use the Uniprot *Entry Name* as protein accession, use the following rule: ...\|[^|]\*\|([^]\*)

nport Search Results	Classe and M. Sterney, N.	x			
Files Selection					
	nascot_data/Proline_Tests/SmallRuns/F071232.dat				
mascot_data/Proline_Tests/SmallRuns/F071233.dat mascot_data/Proline_Tests/SmallRuns/F071234.dat					
mascot_data/Proline_rests	/smailkuns/r0/1234.uat	2			
Parameters					
Software Engine :	Mascot	<b>-</b>			
Instrument :	LTQ-ORBITRAP XL (A1=FTMS F=CID A2=TRAP)	•			
Fragmentation Rule Set :	ESI-TRAP (A1=TRAP F=CID A2=TRAP)				
Peaklist Software :	Mascot Distiller	-			
Decoy Parameters					
Decoy :	Concatenated Decoy	•			
Decoy Accession Regex :	###REV###\\$+				
Parser Parameters					
Subset Threshold : 1.0	)				
Save Door	d V OK X Cancel				

Importing a Search Result can take some time. While the import is not finished, the "All Imported" or "selected dataset" is shown grayed with an hourglass and you can follow the imports in the Tasks Log Window (Menu Window > Tasks Log to show it).



To show all the Search Results imported, double click on "All Imported", or right click to popup the contextual menu and select "Display List"

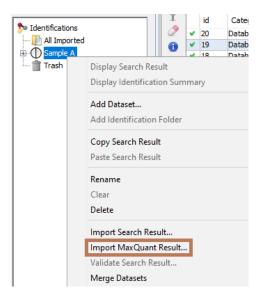


If needed, from the All Imported window, you can drag and drop one or multiple Search Result to an existing dataset.

💼 ProjectTest 🛛 👻 🛅 🌌			id Search Result Name P	Peaklist Path	MSISearch F	MS
		1	2 CR_WP2112 (Calib 25-10-12 / Col reprosil 172) D:	:\\Data\\Clair	r F067897.dat	/mnt/
so Identifications		2	4 CR_WP2112 (Calib 25-10-12 / Col reprosil 172) D:	:\\Data\\Clair	F067898.dat	/mn
All Imported		3	10 K12 DH5 QEx T12 QEx2_002086.raw (DH5_10) D:	:\\MSData\\D	F075556.dat	/mn
- Replicate 1		4	15 CR_WP2112 (Calib 25-10-12 / Col reprosil 172) D:	:\\Data\\Clair	F067897.dat	/mn
- Replicate2		5	17 CR_WP2112 (Calib 25-10-12 / Col reprosil 172) D:	:\\Data\\Clair	F067898.dat	/mn
- Replicate3		6	19 CR_WP2112 (Calib 25-10-12 / Col reprosil 172) D:	:\\Data\\Clair	F067899.dat	/mn
	& Dro	D	)			

### Import MaxQuant result

To import a MaxQuant Search Result, right click on a dataset and then select "Import MaxQuant"



**Note 1:** MaxQuant import will generate a dataset hierarchy with the result from the different acquisition.

The following dialog will be displayed



		Load Identification Da
🞾 Import MaxQuant Results	$\times$	Load Data for Datase
Files Selection		Load Data for Datase
		Load Quantitation Da
		Load Identification Da
		Load Projects for Use
	11	Check User dupierris
		Load Projects for Use
		Check User hesse
		Get Server File Syste
		Load Data for Datase
		Load Data for Datase
	🗫 Ouvrir	×
	Rechercher dans	: 📙 🗸 🤌 📂 🛄 -
	Documents r	50-5 MQ_YO
Parameters	Bureau	
Instrument : < Select >		
Accession regexp :	Documents	
Import quantitation values	Ce PC	
🖬 Save 🕕 Load 🖌 🖌 🖌		
	Réseau	Nom du dossier : D:\DEV\Proline_Data
		Type de fichier : Tous les fichiers $\checkmark$

- select the directory containing the files generated by MaxQuant. This folder should look like:

<root\_folder>\mqpar.xml <root\_folder>\combined\txt\summary.txt <root\_folder>\combined\txt\proteinGroups.txt <root\_folder>\combined\txt\parameters.txt <root\_folder>\combined\txt\msmsScans.txt <root\_folder>\combined\txt\msms.txt

- select the Instrument: mass-spectrometer used for sample analysis different parameters
- specify, if needed, the regular expression to extract protein accessions from MaxQuant protein ids.
- you can choose to import also quantitative data
- click on OK button

# **Delete Data**

You can delete Search Results, Identification Summaries and Datasets in the data tree. You can also delete XIC or Spectral Counts in the quantitation tree.

**Delete** the Datasets (identification or quantitation...) **from the tree view** (Search Result always accessible from "All Imported" view...).

There are two ways to delete data: use the contextual popup or drag and drop data to the Trash.

### Delete Data from the contextual popup

Select the data you want to delete, right-click to open the contextual menu and click on delete menu.

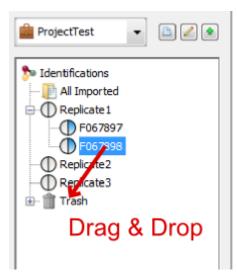


Projects 28 MzDB Fil	es
Project Document	ation 🔹 🗈 🖉 😢
<ul> <li>Identifications</li> <li>All Imported</li> <li>All Imported</li> <li>All State</li> <li>All Imported</li> <li>All I</li></ul>	Dieplay Source Pocult
Replicate2	Display Search Result  Display Identification Summary
⊞ m Trash	Add Dataset Rename
	Clear
	Delete
	Import Search Result
	Validate Search Result
	Merge Datasets
	Filter ProteinSets
	Change Typical Protein
	Generate Spectrum Matches
	Compare With SC
	Export 🕨
	Properties

The selected data is put in the Trash. So it is possible to restore it while the Trash has not been emptied.

### Delete Data by Drag and Drop

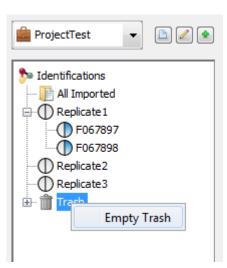
Select the data you want to delete and drag it to the Trash. It is possible to restore data while the Trash has not been emptied



### **Empty the Trash**

To empty the Trash, you have to Right click on it and select the "Empty Trash" menu.





A confirmation dialog is displayed and if accepted Dataset will be removed from the Trash.

Search Results are not completely removed, you can retrieve them from the "All Imported" window.

### **Delete a Project**

It is not possible to delete a Project by yourself. If you need to do it, ask your IT Administrator.

# **Connection Management**

Once user is connected (see Server Connection), it is possible to:

• Reconnect with a different login

Server Connect	tion
Server Paran	neter
Server Host :	GRE046180
User Parame	ters
User :	newUserName
Password :	•••••
	Remember Password
🕑 Default	V OK X Cancel

Change password

Change Password	×
Old password :	••••
New password :	•••••
Confirm new password :	•••••
	V OK Cancel



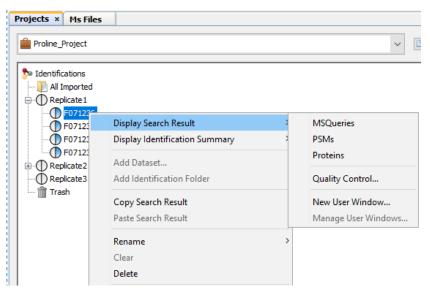
# Display MS Queries, Peptides/PSM or Proteins of a Search Result

All information, validated or not, can be accessible from this menu. Indeed, Search Result contains all data imported from a result file without any validation consideration.

### **Functionality Access**

To display data of a Search Result:

- right click on a Search Result
- click on the menu "Display Search Result >" and on the sub-menu "MSQueries" or "PSM" or "Proteins"



### **MSQueries Window**

If you click on MSQueries sub-menu, you obtain this window:

	Initial Id	Charge	m/z	#PSMs	First Scan	Last Scan	First Time	Last Time	Spectrum Title	
6841	20 813	3	763.3928	2	25212	25212	105.8300	105.8300	22101: Scan 25212 (rt	t=105.83) [D:/Data/MS
6842	20 812	3	763.3925	2	25001	25001	104.8920	104.8920	21908: Scan 25001 (rt	t=104.892) [D:/Data/M
6843	20 811	3	763.3915	2	24916	24916	104.5090	104.5090	21838: Scan 24916 (rt	t=104.509) [D:/Data/M
6844	20 8 10	2	1144.5831	1	24929	24929	104.5670	104.5670	21849: Scan 24929 (rt	t=104.567) [D:/Data/M
6845	20 809	3	763.3789	0	22071	22071	91.7158	91.7158	19135: Scan 22071 (rt	t=91.7158) [D:/Data/M
6846	20 808	3	763.3782	0	13836	13836	55.5833			t=55.5833) [D:/Data/M
6847	20 807	3	763.3781	0	22155	22155	92.0904	92.0904	19208: Scan 22155 (rt	t=92.0904) [D:/Data/M
6848	20 806	3	763.3739	0	18214	18214	74.5564			t=74.5564) [D:/Data/M
6849	20 805	3	763.3493	0	26964	26964	113.6810			t=113.681) [D:/Data/M
6850	20 804	3		-		17478			14808: Scan 17478 (rt	t=71.3421) [D:/Data/M
3851	20 803	3	763.0883	2	28919	28919	122.3630	122.3630	25749: Scan 28919 (rt	t=122.363) [D:/Data/M
3852	20 802	2	1144.1276	1	28910	28910	122.3260	122.3260	25742: Scan 28910 (rt	t=122.326) [D:/Data/M
3853	20 80 1	2	1144.0883	1	19772	19772	81.4564	81.4564	16985: Scan 19772 (rt	t=81.4564) [D:/Data/M
3854	20 800	3	763.0597	1	28202	28202	119.2060	119.2060	25069: Scan 28202 (rt	t=119.206) [D:/Data/M
3855	20 799	2	1144.0857	1	28210	28210	119.2400			t=119.24) [D:/Data/MS
3856	20 798	3	763.0583	0	23012	23012	95.9305	95.9305	20060: Scan 23012 (rt	t=95.9305) [D:/Data/M
3857	20 797	3	763.0579	1	21384	21384	88.6501	88.6501	18482: Scan 21384 (rt	t=88.6501) [D:/Data/M
3858	20 796	3	763.0460	1	8310	8310	32.6783	32.6783	6254: Scan 8310 (rt=3	32.6783) [D:/Data/MSD
3859	20 795	4	572.5167	1	3993	3993	16.5747	16.5747	2221: Scan 3993 (rt=1	16.5747) [D:/Data/MSD
3860	20 794	3	763.0040	1	13152	13152	52.6821	52.6821	10784: Scan 13152 (rt	t=52.6821) [D:/Data/M
3861	20 793	2	1144.0024	1	13235	13235	53.0366			t=53.0366) [D:/Data/M
3862	20 792	3	763.0038	1	13240	13240	53.0584	53.0584	10869: Scan 13240 (rt	t=53.0584) [D:/Data/M
3863	20 791	2	1144.0020	1	13150	13150	52.6731	52.6731	10782: Scan 13150 (rt	t=52.6731) [D:/Data/M
3864	20 790	3	762.7336	1	11987	11987	47.7521	47.7521	9685: Scan 11987 (rt=	=47.7521) [D:/Data/MS
3865	20 789	2	1143.5953	1	12780	12780	51.1014	51.1014	10440: Scan 12780 (rt	t=51.1014) [D:/Data/M
3866	20 788	4	572.2874	0	8157	8157	32.0894	32.0894	6112: Scan 8157 (rt=3	32.0894) [D:/Data/MSD
_										
Pep	otide Length	PTMs	Score Ca	c. Mass	Exp. MoZ Ppm	Charge	Missed Cl. Rank	RT Ion Pare [	Decoy MsQu	Spectrum Title



**Upper View**: list of MSQueries. Some columns may not be (correctly) filled if the Peaklist software were not correctly specified during import. It is possible to change this information using '<u>Update Spectrum ..</u>'

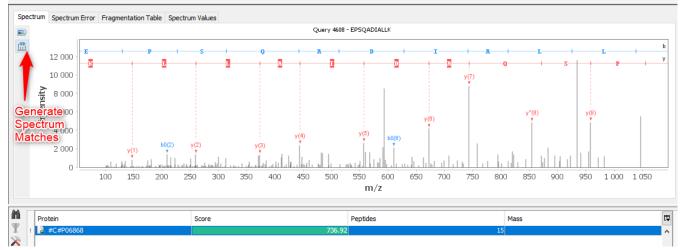
Bottom Window: list of all Peptides linked to the current selected MSQuery.

Note: Abbreviations used are listed here

### PSMs (Peptides) Window

If you click on PSMs sub-menu, you obtain this window:

	Peptide	Length	PTMs	Score	Calc. Mass	Exp. MoZ	Ppm	Charge	Missed Cl.	Rank	RT	Ion Par	MsQuery	Spectrum Title	E,
7756	VTFLAWQYR	9		2.35	1182.6185	592.3171	0.82	2	0	3	28.0916		4595	5178: Scan 714	44 /
7757	RTGSLALFYR	10		2.74	1182.6509	592.3338	1.81	2		1	42.9726		4596	8636: Scan 108	35!
7758	IIPTVDRTLR	10		0.26	1182.7085	592.3578	-6.13	2	1	6	48.3877		4597	9823: Scan 121	13
7759	LIKIQEGNLR	10		0.26	1182.7084	592.3578	-6.10	2	1	6	48.3877		4597	9823: Scan 121	13
7760	LLDSITVPVAR	11		62.26	1182.6973	592.3578	3.35	2	0	) 1	48.3877		4597	9823: Scan 121	13
7761	QRGSITPRIR	10		16.95	1182.6945	592.3578	5.63	2	2	2 2	48.3877		4597	9823: Scan 121	13
7762	EVCFACVDGK	10	Carbamidometh	32.08	1183.5002	592.7587	2.17	2	0	) 1	22.3758		4599	3740: Scan 560	)1
7763	EVCFACVDGK	10	Carbamidometh	27.59	1183.5002	592.7593	3.32	2		1	22.0342		4600	3657: Scan 550	)9
7764	AIEIYTDMGR	10	Oxidation (M8)	30.56	1183.5543	592.7853	1.47	2	0	) 1	20.4658		4604	3236: Scan 507	70
7765	AYSMLTITER	10		46.98	1183.5907	592.8040	2.26	2	0	) 1	39.9774		4606	7942: Scan 101	12
7766	GLTVMFEIMK	10	Oxidation (M9)	0.45	1183.5981	592.8040	-3.99	2		2	39.9774		4606	7942: Scan 101	12
7767	EPSQADIALLK	11		52.48	1183.6448	592.8315	3.11	2	. (	) 1	34.3582		4608	6664: Scan 873	32
7768	HLGTLNFGGIR	11		2.83	1183.6462	592.8315	1.95	2	0		34.3582		4608	6664: Scan 873	32
7769	XAAAPDDLALLK	12		19.63	1183.6448	592.8315	3.11	2		2	34.3582		4608	6664: Scan 873	32
7770	AIEAVAISPWK	11		52.44	1183.6600	592.8387	2.44	2	0	) 1	51.6265		4609	10550: Scan 12	291
7771	MMPTPVILLK	10	Acetyl (Protein	11.63	1183.6709	592.8431	0.59	2		1	104.5440		4610	21845: Scan 24	<del>1</del> 9:
	<														>



Upper View: list of all Peptide Spectrum Matches

**Middle View**: Spectrum, Spectrum Error, Spectrum Values and Fragmentation Table of the selected PSM. If no annotation is displayed, you can generate Spectrum Matches by clicking on the according button

Bottom Window: list of all Proteins identified by the currently selected Peptide.

Note: Abbreviations used are listed here

### **Proteins Window**

If you click on Proteins sub-menu, you obtain this window:



		Proteir	ı		:	Score				Peptides			Mass				E
	1	2 PY	R1_YEAST						3235.37			77	7			245040.7	7
	2	👌 KP1	YK1_YEAST						2937.61			47	7			54544.63	3
	3	🔓 EF:	2_YEAST						2860.99			57	7			93289.18	3
	4	👌 G3	P3_YEAST						2760.79			35	5			35746.6	7
	5	🔓 EN	02_YEAST						2651.52			37	7			46914.1	7
	6	👌 PD	C1_YEAST						2510.19			38	3			61495.3	9
	7	👌 HS	P71_YEAST						2405.36			37	7			69657.2	5
	8	🔓 HS	P72_YEAST						2325.76			40	)			69470.0	1
	9	🔓 EN	01_YEAST						2261.95			33	3			46816.14	4
	10	👌 HS	C82_YEAST						2260.61			43	3			80899.7	5
	11	🍃 PG	K_YEAST						2073.21			34	1			44738.4	3
	12	🔓 HSI	P82_YEAST						2068.16			39	•			81406.40	D
	13	🔓 HS	P75_YEAST						2056.18			35	5			66601.5	9
	14	🔓 HSI	P76_YEAST						1989.48			34	1			66594.5	7
	15	🔓 FA	S2_YEAST						1913.23			48	3			206946.94	4
	16	🔓 FA	S1_YEAST						1903.58			43	3			228691.20	3
	17	👌 G3	P2_YEAST						1890.95			28	3			35846.8	5
1		v. AA	Peptide HILSIK	Next AA	Score 30.48	Start 1912	Stop 1 917	MsQuery 274		Calc. Mass 709.45	Exp. MoZ 355.73	Ppm 0.13	Charge 2	Mis	Ion P	PTM	E.
2			YGLEVR	M	18.41		2 202	647				-0.94	2	0			h
3			QVIAEAK	т	15.38	1 628	1 634	687	-		379.72	1.62	2	0			
4			WSPFNK	D	15.63		1 801	886			389.70	-0.54	2	0			1
5			VSINEPK	1	25.70		205	976			393.72	-2.60	2	0			-U
6	_		LVTLELK	D	5.22		27	1405			408.27	3.31	2	0			
- P			VDVGMGIK	D	8.74		1 741	1434			409.73	-0.04	2	0			
- Z 🛛			SLLATGEK	L	12.02		1 357	1658	1		418.75	-1.22	2	0			
7			LSNVLEAK	v	36.64	287	294	2192	1	872.50	437.26	-0.09	2	0			
- H	R		LOINVLEAK		1 30.04												-111
8 9			VIECNVR	A										0		Carbamidom	
8 9 10	к			A G	17.33	1 259	1 265	2433 2650		888.45	445.23	-0.72 -1.34	2	0		Carbamidom	•
8 9 10 11	K R		VIECNVR		17.33	1 259	1 265	2433	1	. 888.45 . 903.47	445.23	-0.72	2			Carbamidom	•
8 9 10 11	K R K		VIECNVR EGVLDLMK	G	17.33 17.75	1 259 1 941	1 265 1 948	2433 2650 2823	1	888.45 903.47 913.56	445.23 452.74	-0.72 -1.34	2	0		Carbamidom	•
8 9 10 11 12 13	K R K R		VIECNVR EGVLDLMK LLIEAISR	G	17.33 17.75 46.36	1 259 1 941 1 480	1 265 1 948 1 487	2433 2650 2823	1	888.45 903.47 913.56 933.52	445.23 452.74 457.79	-0.72 -1.34 -0.07	2 2 2	0		Carbamidom	•
8 9 10 11 12 13	K R K R		VIECNVR EGVLDLMK LLIEAISR DYPVVITK	G	17.33 17.75 46.36 2.96	1 259 1 941 1 480 1 167	1 265 1 948 1 487 1 174	2433 2650 2823 3152	1	888.45 903.47 913.56 933.52 958.55	445.23 452.74 457.79 467.77	-0.72 -1.34 -0.07 -0.24	2 2 2 2	0 0 0		Carbamidom	
8	K R K R R		VIECNVR EGVLDLMK LLIEAISR DYPVVITK ILAIDVGMK	G N Y Y	17.33 17.75 46.36 2.96 22.82	1 259 1 941 1 480 1 167 229 815	1 265 1 948 1 487 1 174 237	2433 2650 2823 3152 3599	1 1 2 1 1	888.45 903.47 913.56 933.52 958.55 964.49	445.23 452.74 457.79 467.77 480.28	-0.72 -1.34 -0.07 -0.24 -0.02	2 2 2 2 2 2	0 0 0		Carbamidom	
8 9 10 11 12 13 14 15	K R K R R R K		VIECNVR EGVLDLMK LLIEAISR DYPVVITK ILAIDVGMK TFEEAIQK	G N Y Y	17.33 17.75 46.36 2.96 22.82 47.98	1 259 1 941 1 480 1 167 229 815	1 265 1 948 1 487 1 174 237 822	2433 2650 2823 3152 3599 3670		888.45 903.47 913.56 933.52 958.55 964.49 966.47	445.23 452.74 457.79 467.77 480.28 483.25	-0.72 -1.34 -0.07 -0.24 -0.02 -0.45	2 2 2 2 2 2 2 2	0 0 0 0		Carbamidom	-

#### Upper View: list of all Proteins

Bottom View: list of all Peptides identifying the selected Protein.

Note: Abbreviations used are listed here

# Display MS Queries, PSM, Peptides, Protein Sets or Adjacency Matrices of an Identification Summary

From this menu, all displayed information is Identification Summary data, which has been validated according to user specified rules. To view the raw information as defined at import, use the Search Result sub menu.

### **Functionality Access**

To display data of an Identification Summary:

- right click on an Identification Summary
- click on the menu "Display Identification Summary >" and on the sub-menu "MSQueries", "PSM", "Peptides", "Protein Sets", "PTM Protein Sites" or "Adjacency Matrix"



Didentifica	ported	>	
🛛 🖶 🍈 t 📩			MSQueries
Lăt	Display Identification Summary		
	Add Dataset		PSMs
	Add Identification Folder		Peptides
			Protein Sets
	Copy Search Result		Adjacency Matrix
	Paste Search Result		Modification Sites
	Rename	1	Modification Clusters
	Clear		New User Window
	Delete		Manage User Windows
	Import Search Result		Manage Oser Willdows
	Import MaxQuant Result		
	Maliata Carala Danula		

# **MSQueries Window**

If you click on MSQueries sub-menu, you obtain this window:

	Charge	m/z		#Peptide Mat	ches First Sca	an	Last Scan	First Time	L	ast Tin	ne Spe	ectrum Title	2					
	1	3	1004.2088		3	0	0	100	0.7630		249	55: Scan 3	7188 (rt=10	0.763) [D:	/Data/Claire/OB	EMMA 121101	_43b.raw]	
	2	2	710.8333		1	0	0	35	5.8804		766	5: Scan 11	718 (rt=35.	3804) [D:/	Data/Claire/OE	MMA121101_	43b.raw]	
	3	2	433.2238		6	0	0	31	L.9367		663	0: Scan 10	122 (rt=31.	9367) [D:/	Data/Claire/OE	MMA121101_	43b.raw]	
	4	3	393.2524		1	0	0	69	.5741		166	46: Scan 2	5427 (rt=69	.5741) [D:	/Data/Claire/OB	EMMA121101	_43b.raw]	
	5	4	550.5493		0	0	0	58	3.1151		136	11: Scan 2	0740 (rt=58	.1151) [D:	/Data/Claire/OB	EMMA 121101	_43b.raw]	
	6	2	374.7152		0	0	0	48	3.7761		110	77: Scan 1	6949 (rt=48	.7761) [D:	/Data/Claire/OB	EMMA 121101	_43b.raw]	
	7	2	703.8414		3	0	0	37	7.0394		795	6: Sum of :	2 scans in ra	nge 12189	(rt=37.0394) 1	to 12217 (rt=	37.1083) [[	):/Data
	8	2	905.9039		0	0	0	88	3.8100		220	01: Scan 3	3032 (rt=88	.81) [D:/D	ata/Claire/OEM	MA121101_4	3b.raw]	
	9	4	487.7362		2	0	0	72	2.0339		173	06: Scan 2	6438 (rt=72	.0339) [D:	/Data/Claire/OB	EMMA 121101	_43b.raw]	
1	0	2	771.8703		1	0	0	50	.9328		116	26: Scan 1	7831 (rt=50	.9328) [D:	/Data/Claire/OB	EMMA 121101	_43b.raw]	
1	1	2	594.8139		2	0	0	21	L.4588		394	3: Scan 59	60 (rt=21.4	588) [D:/D	ata/Claire/OEM	MA121101_4	3b.raw]	
1	2	3	884.1173		4	0	0	109	9.4620		265	82: Scan 3	9515 (rt=10	9.462) [D:	/Data/Claire/OB	EMMA 121101	_43b.raw]	
1	3	4	848.9335		1	0	0	136	5.2080		302	09: Scan 4	5171 (rt=13	6.208) [D:	/Data/Claire/OB	EMMA 121101	_43b.raw]	
1	4	3	585.6480		2	0	0	63	3.5996		150	92: Scan 2	2981 (rt=63	5996) [D:	/Data/Claire/OB	EMMA 121101	_43b.raw]	
1	5	4	411.2476		0	0	0	62	2.9898		149	28: Scan 2	2727 (rt=62	.9898) [D:	/Data/Claire/OB	EMMA 121101	_43b.raw]	
1	6	2	768.9158		7	0	0	30	0.9361		635	6: Sum of 3	2 scans in ra	nge 9734	(rt=30.9361) to	9756 (rt=30	).9912) [D:/	Data/C
1	7	2	386.7380		10	0	0	10	.7135		109	1: Scan 18	50 (rt=10.7	135) [D:/D	ata/Claire/OEM	MA121101_4	3b.raw]	
1	8	1	568.3451		0	0	0	23	3.7393		450	5: Scan 68	33 (rt=23.7	393) [D:/D	ata/Claire/OEM	MA121101_4	3b.raw]	
1	9	2	633.3152		3	0	0	12	2.7999		161	5: Scan 26	09 (rt=12.7	999) [D:/D	ata/Claire/OEM	MA121101_4	3b.raw]	
2	0	3	756.0681		7	0	0	113	3.8170		273	09: Scan 4	0618 (rt=11	3.817) [D:	/Data/Claire/OB	EMMA 121101	_43b.raw]	
2	1	2	763.8879		6	0	0	37	7.8659						Data/Claire/OE			
2		3	663.3690		0	0	0		2.0440						/Data/Claire/OB			
2	3	3	748 7043		0	0	0	4'	4960		038	0- Scan 14	380 (++-47 )	1860) [D+/	Data/Claire/OEI	MMA 121101	43b rawl	
Pe	eptide		PTM	Score	Calc. Mass	Exp.	MoZ	Ppm	Charg	je M	issed Cl.	Rank	RT	Protei	Protein Sets	Ion Par	Decoy	Valida
	SVGFNEVVAS			0.00	3009.5903		1004.2088		-	3	1		2 100.7630					
	LQIVQFINEPS			50.22	3009.5962		1004.2088			3	0		100.7630		SSZ1_YEAST			×
з   ИК	DNPLDPYIKDI	4YSFIS	SII	0.05	3009.6128		1004.2088	3 -3.0	2	3	2	2	2 100.7630	(	)		×	

Upper View: list of MSQueries.

Bottom Window: list of all Peptides linked to the current selected MSQuery.

Note: Abbreviations used are listed here

This view contains all MSQueries even if it doesn't bring an identification.

### **PSMs Window**

If you click on PSM sub-menu, you obtain this window:



	Peptide	Score	MsQuery	Rank	Calc. Mass	Exp. MoZ	Ppm	Charge	Missed Cl.	Ion P	PTM	Protein Sets	1
1	ILFQK	21.38	113	2	647.40	648.41	-0.42	:	1 C			PUT1_YEAST	ľ
2	LIFAGK	28.56	113	1	647.40	648.41	-0.42	1	1 0	)		RL402_YEAST	. P
3	LLFQK	21.38	113	3	647.40	648.41	-0.42	1	1 0	)		ARGR2_YEAST	٦
4	LVDLIK	25.99	182	1	699.45	350.73	-0.28	2	2 0	1		PHSG_YEAST	
5	LVDLLK	25.99	182	2	699.45	350.73	-0.28	2	2 0	1		BRR2_YEAST	٦
6	LVDLIK	24.51	185	1	699.45	350.73	0.63	2	2 0	1		PHSG_YEAST	
7	LVDLLK	24.51	185	2	699.45	350.73	0.63	2	2 0			BRR2_YEAST	٦
8	LVDLIK	22.23	187	1	699.45	350.73	1.17	2	2 0	1		PHSG_YEAST	
9	LVDLLK	22.23	187	2	699.45	350.73	1.17	2	2 0	1		BRR2_YEAST	П
10	VPFGGVK	31.28	200	1	702.41	352.21	0.24	2	2 0	1		ALDH6_YEAST	
11	AAFIER	28.59	223	1	705.38	353.70	-0.37	2	2 0	)		RIR4_YEAST	П
12	AAFIER	23.53	225	1	705.38	353.70	1.47	2	2 0			RIR4_YEAST	
13	LADFLK	24.19	227	1	705.41	353.71	-0.91		2 0			SYYC_YEAST	
14	AIDLFK	25.18	230	1	705.41	353.71	2.06	2	2 0			SYRC_YEAST	
15	AIDLFK	25.12	232	1	705.41	353.71	2.65	2	2 0			SYRC_YEAST	
16	LADFLK	20.84	232	2	705.41	353.71	2.65	2	2 0	1		SYYC_YEAST	
17	AGAFITK	34. <mark>0</mark> 3	239	1	706.40	354.21	-0.79	2	2 0			PYRD_YEAST	Т
18	AGAFITK	41.53	240	1	706.40	354.21	0.20	2	2 0			PYRD_YEAST	
19	KAMITK	27.74	240	2	706.40	354.21	-4.54	2	2 1		Oxidation (M3)	FMC1_YEAST	
20	FGMDLK	34. <mark>44</mark>	256	1	709.35	355.68	-0.34	2	2 0	1		RIR4_YEAST	
21	FGMDLK	23.01	258	1	709.35	355.68	0.61		2 0	1		RIR4_YEAST	٦
22	FGMDLK	20.96	259	1	709.35	355.68	0.70	2	2 0	1		RIR4_YEAST	
23	DHLLGR	24.18	264	1	709.39	355.70	-0.91		2 0	)		RL16B_YEAST	
24	HILSIK	22.97	268	1	709.45	355.73	-0.69	2	2 0			PYR1_YEAST	
25	HLISIK	22.97	268	2	709.45	355.73	-0.69	2	2 0			OTC_YEAST	
26	HILSIK	23. 1 <mark>6</mark>	269	1	709.45	355.73	-0.58	2	2 0			PYR1_YEAST	
27	HLISIK	23. 1 <mark>6</mark>	269	2	709.45	355.73	-0.58	2	2 0			OTC_YEAST	
28	HLSLLK	21.12	269	3	709.45	355.73	-0.58	2	2 0			SKP2_YEAST	
29	HILSIK	24.81	272	2	709.45	355.73	-0.18	2	2 0			PYR1_YEAST	
30	HLISIK	24.81	272	3	709.45	355.73	-0.18	2	2 0			OTC_YEAST	
31	HLSLLK	28.22	272	1	709.45	355.73	-0.18	2	2 0			SKP2_YEAST	
32	HILSIK	30. <mark>4</mark> 8	274	2	709.45	355.73	0.13	2	2 0			PYR1_YEAST	
33	HLISIK	30. <mark>4</mark> 8	274	3	709.45	355.73	0.13		2 0			OTC_YEAST	
34	HLSLLK	32.31	274	1	709.45	355.73	0.13	2	2 0			SKP2_YEAST	
35	HILSIK	25.34	275	2	709.45	355.73	0.60	2	2 0			PYR1_YEAST	
36	HLISIK	25.34	275	3	709.45	355.73	0.60	2	2 0			OTC_YEAST	
37	HLSLLK	26.93	275	1	709.45	355.73	0.60	2	2 0			SKP2_YEAST	٦.

Note: Abbreviations used are listed here

# Peptides Window

If you click on Peptides sub-menu, you obtain this window:

Peptide	Lengt	h PTMs	Sco	re	Calc. Mass	Exp. MoZ	Ppm	Charge	Missed Cl.	Rank	RT	F	Protein Se	Protein Sets	PSM Count	MsQuer	y Spectrum Title
1 LVQDVANNTNEEAG	DGTTTAT	25		216.83	2559.2412	1280.6292	1.	01	2		1	28.4886		CH60 MOUSE		1	73704 17395: Scan 192
2 SCSGVEFSTSGSSN	TDTGKVSGTLE	27 Carbamidomet		216.10	2722.2239	1362.1228	2.	59	2	l	1	19.9109		2 G3UX26_MOUSE		3	78783 9695: Scan 1212
3 METYCNSGSTDTSS	VINAVTHALT	31 Carbamidomet		205.80	3318.5132	1107.1788	0.	39	3 (		1	103.8870		BDH_MOUSE		4	93848 83798: Scan 766
4 MMGGPYGGGNYGP	GGSGGSGGYG	25 Oxidation (M2)		204.69	2204.8931	1103.4567	2.	63	2	1	1	15.5690		1 ROA2_MOUSE		1	58750 5154: Scan 8165
6 METYCNSGSTDTSS	VINAVTHALT	31 Oxidation (M1)		202.35	3334.5083	1112.5104	0.	35	3	1	1	101.4950		1 BDH_MOUSE		1	94179 81821: Scan 749
6 LLLAGYDDFNCNVW	/DALKADR	21 Carbamidomet		196.45	2468.1794	1235.0977				L	1	93.2148		1 GBB1_MOUSE		6	70303 74683: Scan 689
7 GGGGSFGYSYGGGS	GGGFSAS	32		196.04	2704.1538					1	1	59.3665		1 #C#P35527		2	78238 44642: Scan 435
8 LCYVALDFEQEMAT		23 Carbamidomet		192.61	2549.1665						1	96.8304		ACTG_MOUSE, #		3	73330 77800: Scan 715
8 EQGSSGLGSGSSGG		28		192.56	2401.0742				2		1	22.5157		1 AAK1_MOUSE		1	67623 11996: Scan 142
		25		101.01	0101 0017	1001 5015					*	10.0000				*	56600 0000 0 6006
Protein Set	Description	Score			Proteins	Peptides	5	Obs	ervable Peptide	s	pectral Co	ount	Specific	Spectral Count	Sequence Coun	t	Mass
				2115.78	4 (1	.34)		28					73	1		2	6 31
DAC2_MOUSE	sp  Q60930  VD	AC2_MOU		2115.70													
3UX26_MOUSE	tr  G3UX26 G3L	JX26_MO	T	2113.76 1851.23 Prev. AA	1 (1	eptide	Next /	24	Length	17 16	PTMs		63 Score	:	2 Start	2 Si	14 30 top C
DAC2_MOUSE 3UX26_MOUSE ical Protein: us musc	tr  G3UX26 G3L	JX26_MO	Ţ	1851.23	1 (1	, 0 🖌 )	Next /	24	Length		PTMs		63	:	2	2	4 30
3UX26_MOUSE	tr  G3UX26 G3L tulus OX=10090 GN=	JX26_MO •Vdac2 PE=1 SV=2	Ţ	1851.23	1 (1	, 0 🖌 )		24	Length				63 Score	:	2 Start	2	4 30
3UX26_MOUSE	tr  G3UX26 G3L	JX26_MO	₹ ×	1851.23	1 (1 Pe SC	eptide	rgky	24	Length	16	27 Carba		63 Score	:	2 Start	2 Si	14 30 top C
3UX26_MOUSE cal Protein:	tr [G3UX26]G3U tulus OX = 10090 GN = Description sp [Q60930]VDAC	JX26_MO ■Vdac2 PE=1 SV=2 2_M	¥ *	1851.23 Prev. AA	1 (1 Pe SC SC	eptide	rgky	24	Length	16	27 Carba	midomethyl (	63 Score	216.10	2 Start	2 51 47	4 30 top C 73
Cal Protein: Js musc cal Protein: Js musc Protein VDAC2_MOUSE A0A286YCR8_M	tr  G3UX26 G3U tulus OX = 10090 GN = Description sp  Q60930 VDAC tr  A0A286YCR8 A	-Vdac2 PE=1 SV=2 Same: CP 2_M ^	않	1851.23 Prev. AA 1 K 2 K	1 (1 Pe SC SC	eptide SGVEFSTSGSSNTDT SGVEFSTSGSSNTDT	rgky rg v	24	Length	16	27 Carba 19 Carba	midomethyl (	63 Score	216.10 138.60	2 Start	2 51 47 47 109 109	4 30 top C 73 65 121 122
Cal Protein: Js musc Protein: Js musc VDAC2_MOUSE A0A286YCR8_M D3YZT5_MOUSE	tr  G3UX26 G3U ulus OX = 10090 GN = Description sp  Q60930 VDAC tr  A0A286YCR8 A tr  D3YZT5 D3YZT	-Vdac2 PE=1 SV=2 Same: ₽ 2_M 0A		1851.23 Prev. AA 1 K 2 K 3 K	1 (1 Pe SG SC LTT LTT	o Z) eptide SGVEFSTSGSSNTDT SGVEFSTSGSSNTDT FDTTFSPNTGK	rgk <mark>y</mark> rg v K S	24	Length	16	27 Carba 19 Carba 13 14	midomethyl (	63 Score (C2)	216.10 138.60 108.67 92.23 82.73	Start	2 47 47 109 109 45	4 30 top C 73 65 121 122 65
Cal Protein: Js musc Protein: Js musc VDAC2_MOUSE A0A286YCR8_M D3YZT5_MOUSE	tr  G3UX26 G3U tulus OX = 10090 GN = Description sp  Q60930 VDAC tr  A0A286YCR8 A	-Vdac2 PE=1 SV=2 Same: ₽ 2_M 0A	않	1851.23 Prev. AA 1 K 2 K 3 K 4 K 6 K 6 R	1 (1 Pe SC SC LTT LTT TK PM	aptide SGVEFSTSGSSNTDT SGVEFSTSGSSNTDT POTTESPNTGK FDTTESPNTGKK SCSGVEFSTSGSSNTT CIPPPYADLGK	TG V K S D V A	24	Length	16	27 Carba 19 Carba 13 14 21 Carba 13 Oxidat	midomethyl ( midomethyl ( midomethyl ( tion (M2); Ca	63 Score (2) (2) (2) (2) (2) (2) (2) (2) (2) (2)	216.10 138.60 108.67 92.23 82.73 82.31	Start	2 47 47 109 109 45 12	top C 73 65 121 122 65 24
Cal Protein: Js musc Protein: Js musc VDAC2_MOUSE A0A286YCR8_M D3YZT5_MOUSE	tr  G3UX26 G3U ulus OX = 10090 GN = Description sp  Q60930 VDAC tr  A0A286YCR8 A tr  D3YZT5 D3YZT	-Vdac2 PE=1 SV=2 Same: ₽ 2_M 0A	않	1851.23 Prev. AA 1 K 2 K 3 K 4 K 6 K 6 R 7 K	1 (1 Pe SC LTF LTF WW	eptide SGVEFSTSGSSNTDT SGVEFSTSGSSNTDT POTTFSPNTGK FDTTFSPNTGK SCSGVEFSTSGSSNTI CIPPPYADLGK VTDNTLGTEIAIEDQI	TGY TGV K S TDV A LCL	24	Length	16	27 Carba 19 Carba 13 14 21 Carba 13 Oxida 22 Carba	midomethyl ( midomethyl ( midomethyl ( tion (M2); Ca midomethyl (	63 Score (2) (2) (2) (2) (2) (2) (2) (2) (2) (2)	216.10 138.60 108.67 92.23 82.73 82.73 82.31 80.75	Start	2 51 47 47 109 109 45 12 87	top C 73 65 121 122 65 24 108
Cal Protein: Js musc Protein: Js musc VDAC2_MOUSE A0A286YCR8_M D3YZT5_MOUSE	tr  G3UX26 G3U ulus OX = 10090 GN = Description sp  Q60930 VDAC tr  A0A286YCR8 A tr  D3YZT5 D3YZT	-Vdac2 PE=1 SV=2 Same: ₽ 2_M 0A	않	1851.23           Prev. AA           1           K           2           K           4           K           6           R           7           8	1 (1 Pe SC LTT TK PM WW AE	aptide SGVEFSTSGSSNTDT SGVEFSTSGSSNTDT FDTTFSPNTGK FDTTFSPNTGK SCSGVEFSTSGSSNTT CIPPPYADLGK VTDNTGTELAIEDQI CCVPVCPRPMCIPPP	IGKY IG V K S D V A IC L PY A	24	Length	16	27 Carba 19 Carba 13 14 21 Carba 13 Oxida 22 Carba 23 Acety	midomethyl ( midomethyl ( midomethyl ( tion (M2); Ca	63 Score (2) (2) (2) (2) (2) (2) (2) (2) (2) (2)	216.10 138.60 108.67 92.23 82.73 82.31 80.75 64.14	Start	2 51 47 109 109 45 12 87 2	top C 73 65 121 122 65 24 108 24
Cal Protein: Js musc Protein: Js musc VDAC2_MOUSE A0A286YCR8_M D3YZT5_MOUSE	tr  G3UX26 G3U ulus OX = 10090 GN = Description sp  Q60930 VDAC tr  A0A286YCR8 A tr  D3YZT5 D3YZT	-Vdac2 PE=1 SV=2 Same: ₽ 2_M 0A	않	1851.23           Image: Prev. AA           K	1 (1) SCS SCS LTT LTT TCK PM WM AE GFi	ptide Soverstscssntdt PDTFspNtck PDTFspNtck CIPPPYADLck VTDNTLGTEAIEDQI CrVPV-RPMCIPP GFGLVRLDW	TGKY TGV K S DV A C L PY A T	24	Length	16	27 Carba 19 Carba 13 14 21 Carba 13 Oxidat 22 Carba 23 Acety 12	midomethyl ( midomethyl ( midomethyl ( tion (M2); Ca midomethyl ( ( Protein N-tr	63 Score (C2) (C2) (C2) (C4) (C4) (C18) (C18) (C18)	216.10 138.60 108.67 92.23 82.73 82.31 80.75 64.14 61.21	Start	2 47 47 109 109 45 12 87 2 33	top C 73 65 121 122 65 24 108 24 108 24
3UX26_MOUSE cal Protein: _Js musc Protein VDAC2_MOUSE A0A286YCR8_M 2 D3YZT5_MOUSE	tr  G3UX26 G3U ulus OX = 10090 GN = Description sp  Q60930 VDAC tr  A0A286YCR8 A tr  D3YZT5 D3YZT	-Vdac2 PE=1 SV=2 Same: ₽ 2_M 0A	않	1851.23  Prev. AA  K  K  K  K  K  K  K  K  K  K  K  K	1 (1 Pe SC SC LTT LTT TK PM WM AE GFT WW	, o 24) stokessmooth Soverstisossmooth Soverstisossmooth Forthespantak Soverstisossmooth Compensational Soverstisossmooth Compensational Soverstisossmooth	TGKY TGV K S TDV A TCL PYA T W	24	Length	16	27 Carba 19 Carba 13 14 21 Carba 13 Oxidat 22 Carba 23 Acetyl 12 11 Carba	midomethyl ( midomethyl ( midomethyl ( tion (M2); Ca midomethyl ( (Protein N-b midomethyl (	63 Score (22) (23) (24) (24) (21) (24) (21) (22) (22) (22) (22) (22) (23) (23) (24) (24) (25	216.10 138.60 108.67 92.23 82.73 82.31 80.75 64.14 61.21 57.73	Start	2 47 47 109 45 12 87 2 33 76	44 30 top C 73 65 121 122 65 24 108 24 44 86
3UX26_MOUSE cal Protein: _Js musc Protein VDAC2_MOUSE A0A286YCR8_M 2 D3YZT5_MOUSE	tr  G3UX26 G3U ulus OX = 10090 GN = Description sp  Q60930 VDAC tr  A0A286YCR8 A tr  D3YZT5 D3YZT	-Vdac2 PE=1 SV=2 Same: ₽ 2_M 0A	않	1851.23 Prev. AA 1 2 K 3 K 4 K 6 R 7 K 8 M 8 K 1 1 1 1 1 1 1 1 1 1 1 1 1	1 (1)	ptide Soverstisessmith Soverstisessmith PDTTspNrtek Soverstisessmith PDTTspNrtek Soverstisessmith CPPPPADLek UTDNTLGTEIAIEDQI CCPVPCPRMCIPPP GFQLWQDWK EPGLTFTEK CCVPVCPR	TGKY TGV K S TDV A ICL L DYA T W P	24	Length	16	27 Carba 19 Carba 13 14 21 Carba 13 Oxida 22 Carba 23 Acety 12 11 Carba 10 Acety	midomethyl ( midomethyl ( tion (M2); Ca midomethyl ( (Protein N-b midomethyl ( (Protein N-b	63 Score (2) (2) (2) (2) (2) (2) (2) (2)	216.10 138.60 108.67 92.23 82.73 82.73 80.75 64.14 61.21 57.73 56.98	Start	2 47 47 109 109 45 12 87 2 33 76 2	top C 73 65 121 122 65 24 108 24 108 24 44 86 11
3UX26_MOUSE cal Protein: _Js musc Protein VDAC2_MOUSE A0A286YCR8_M 2 D3YZT5_MOUSE	tr  G3UX26 G3U ulus OX = 10090 GN = Description sp  Q60930 VDAC tr  A0A286YCR8 A tr  D3YZT5 D3YZT	-Vdac2 PE=1 SV=2 Same: ₽ 2_M 0A	않	1851.23 1851.23 1 1 1 2 K 2 K 3 K 4 K 6 K 0 R 7 K 8 M 9 K 10 1 1 1 1 M	1 (1)	, o 24) sptide Soverstisdsshift of Soverstisdsshift Soverstisdsshift OTTFsphildk Sc Soverstisdsshift OTTFsphildk Sc Soverstisdsshift CVPV/DRPMCIPP GrowLDW ErGLIFTEK CCVPV/DRPMCIPPR	TGKY TGV K S TDV A ICL L DYA T W P	24	Length	16	27 Carba 19 Carba 13 14 21 Carba 13 Oxida 22 Carba 23 Acetyl 12 11 Carba 10 Acetyl 23 Acetyl 23 Acetyl	midomethyl ( midomethyl ( midomethyl ( tion (M2); Ca midomethyl ( (Protein N-b midomethyl (	63 Score (2) (2) (2) (2) (2) (2) (2) (2)	216.10 138.60 108.67 92.23 82.73 82.33 80.75 64.14 61.21 57.73 56.98 54.31	Start	2 47 47 109 45 12 87 2 33 76 6 2 2 2	top C 75 65 121 122 65 24 24 104 108 108 108 11 24
3UX26_MOUSE cal Protein: _Js musc Protein VDAC2_MOUSE A0A286YCR8_M 2 D3YZT5_MOUSE	tr  G3UX26 G3U ulus OX = 10090 GN = Description sp  Q60930 VDAC tr  A0A286YCR8 A tr  D3YZT5 D3YZT	-Vdac2 PE=1 SV=2 Same: ₽ 2_M 0A	않	1851.23 1851.23 1 Prev. A4 1 K 2 K 4 K 6 6 R 7 K 6 8 6 K 10 K 11 1 M 12 M 12 M 5 K 6 K 6 K 6 8 K 6 8 8 8 8 8 8 8 8 8 8 8	1 (1   See SCC LTT LTT TCC PM MW AE GFT WW AE GFT	spide Soverstisgsshift Soverstisgsshift Soverstisgsshift Dittrsphrick PDTTsphrick PDTTsphrick PDTTsphrick PDTTsphrick PDTTsphrick CVPV:PRPMCIPP GFQ:UKUDW CVPV:PRPMCIPP GFQ:UKUDW	rg v rg v k s D v A tC L PY A T W P P P Y A	24	Length	16	27 Carba 19 Carba 13 14 21 Carba 13 Oxida 22 Carba 23 Acetyl 12 11 Carba 10 Acetyl 23 Acetyl 8	midomethyl ( midomethyl ( midomethyl ( tian (M2); Ca midomethyl ( (Protein N-b midomethyl ( (Protein N-b (Protein N-b	63 Score (C2) (C4) (C4) (C4) (C4) (C4) (C18) (C18) (C18) (C18) (C18) (C18) (C18) (C18) (C18) (C18) (C19)	216.10 138.60 108.67 92.23 82.33 82.33 80.57 64.14 61.21 57.73 56.98 54.31 53.99	Start	2 47 47 109 109 45 12 87 2 33 76 2 2 33 76 2 2 33	top C 73 73 73 121 122 65 24 24 44 86 11 24 44 44 44 44
Cal Protein: Js musc Protein: Js musc VDAC2_MOUSE A0A286YCR8_M D3YZT5_MOUSE	tr  G3UX26 G3U ulus OX = 10090 GN = Description sp  Q60930 VDAC tr  A0A286YCR8 A tr  D3YZT5 D3YZT	-Vdac2 PE=1 SV=2 Same: ₽ 2_M 0A	않	1851.23 1851.23 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 (1) Pe SG SG SG LTT LTT TG PM WW AB GFP WW AB GFP MG AB GFP MG AB	ptide Soversiscessito Soversiscessito Soversiscessito Diffsputck Diffsputck CVPV-PRPMCIPP Ground Diff CVPV-PRPM_IPP Ground	TG Y TG V K S D V A TC L PY A P PY A L W	24	Length	16	27 Carba 19 Carba 13 14 21 Carba 13 Oxida 22 Carba 23 Acetyl 12 11 Carba 10 Acetyl 23 Acetyl 8 13 Carba	midomethyl ( midomethyl ( tion (M2); Ca midomethyl ( (Protein N-ti (Protein N-ti (Protein N-ti midomethyl (	63 Score (C2) C22 C4) C4) C18) c19 c22	216.10 138.60 108.67 92.23 82.73 82.73 82.73 82.73 82.73 82.73 82.75 64.14 61.21 57.73 56.98 54.31 53.99 46.69	Start	2 477 477 109 45 12 87 2 33 76 2 2 33 76 2 2 33 774	top C C 75 65 121 122 65 121 122 65 121 122 65 121 122 65 121 122 65 124 106 124 146 15 124 146 146 146 146 146 146 146 146 146 14
3UX26_MOUSE	tr  G3UX26 G3U ulus OX = 10090 GN = Description sp  Q60930 VDAC tr  A0A286YCR8 A tr  D3YZT5 D3YZT	-Vdac2 PE=1 SV=2 Same: ₽ 2_M 0A	않	1851.23 1851.23 1 Prev. A4 1 K 2 K 4 K 6 6 R 7 K 6 8 6 K 10 K 11 1 M 12 M 12 M 5 K 6 K 6 K 6 8 K 6 8 8 8 8 8 8 8 8 8 8 8	1 (1	spide Soverstisgsshift Soverstisgsshift Soverstisgsshift Dittrsphrick PDTTsphrick PDTTsphrick PDTTsphrick PDTTsphrick PDTTsphrick CVPV:PRPMCIPP GFQ:UKUDW CVPV:PRPMCIPP GFQ:UKUDW	rg v rg v k s D v A tC L PY A T W P P P Y A	24	Length	16	27 Carba 19 Carba 13 14 21 Carba 13 Oxida 22 Carba 23 Acetyl 12 11 Carba 10 Acetyl 23 Acetyl 8 13 Carba	midomethyl ( midomethyl ( midomethyl ( tian (M2); Ca midomethyl ( (Protein N-b midomethyl ( (Protein N-b (Protein N-b	63 Score (C2) C22 C4) C4) C18) c19 c22	216.10 138.60 108.67 92.23 82.33 82.33 80.57 64.14 61.21 57.73 56.98 54.31 53.99	Start	2 47 47 109 109 45 12 87 2 33 76 2 2 33 76 2 2 33	top C 73 73 73 121 122 65 24 24 44 86 11 24 44 44 44 44

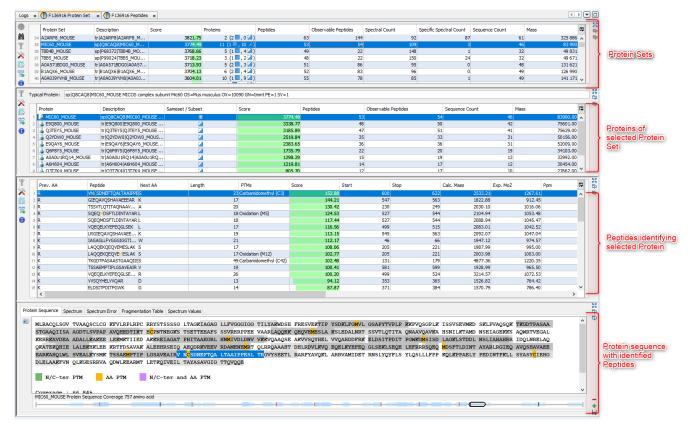
Upper View: list of all Peptides with best PSM information (charge, score ...)



Middle View: list of all Protein Sets identified by the selected peptide. Bottom Left View: list of all Proteins of the selected Protein Set Bottom Right View: list of all Peptides of the selected Protein Note: Abbreviations used are listed <u>here</u>

# Protein Sets Window

If you click on Protein Sets sub-menu, you obtain this window:



View 1 (upper): list of all Protein Sets of the identification Summary

Note: In the column *Proteins*, 8 (2, 6) means that there are 8 proteins in the protein set : 2 in the sameset, 6 in the subset.

View 2: list of all Proteins of the selected Protein Set, sameset or subset.

**View 3**: list of all Peptides of the selected Protein. If a subset is selected only peptides matching that protein will be listed.

**View 4a**: Protein Sequence of the previously selected Protein and Spectrum of the selected Peptide. Other tabs display Spectrum, Spectrum Error and Fragmentation Table.

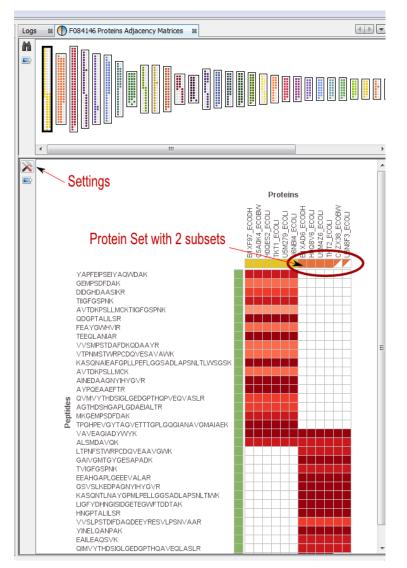
View 4b: Graphic representation of the Protein with matching peptide and associated modifications.

Note: Abbreviations used are listed here



# Adjacency Matrix Window

If you click on Adjacency Matrix sub-menu, you obtain this window:



View 1: All the matrices. Each matrix corresponds to a cluster composed of linked Proteins/Peptides.

Note: use the Search tool to display an Adjacency Matrix for a particular Protein or Peptide

View 2: The currently selected matrix.

In the example, you can see two different protein sets which share only two peptides.

Thanks to the settings you can hide proteins with exactly the same peptides.

# Display Modification Site or Cluster of a Modification Dataset

### **Functionality Access**

To display modification data associated to an Identification Summary:

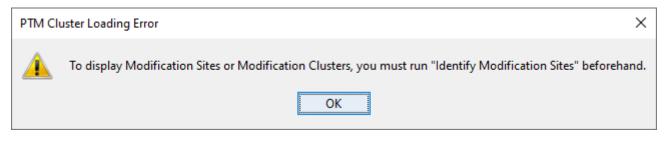
1 - 3	2/ 0 /	NPCA_DUMA
Display Search Result	>	TF3C2_HUM/
Display Identification Sympose		RGS12_HUM
 Display Identification Summary	·	DDED 2 HUM
Display Modifications	2	Sites
Add Dataset		Clusters 0
Add Identification Folder		Annotated Clusters
Copy Search Result		RN126_HUM
		OTUD3_HUM
Paste Search Result		DOCK6_HUM
-		DOCK6_HUM
Rename	>	DOCK6_HUM
Clear Validation		DOCK6_HUM
Clear All		00000 1000



- right click on the Dataset
- click on the Display Modification > "Sites"; "Clusters"; "Annotated Clusters"

If you click on one of *Display Modifications* sub-menu, you can obtain the following warning dialog.

This is due to the fact that you must run beforehand the "Identify Modification Sites" process. To do that, mouse right click on your Identification Summary and select the "Identify Modification Sites" menu.



The Annotated Clusters menu may also cause the display of this warning. Indeed, Annotated Clusters display is only available after Clusters have been modified and saved.

All displays, Sites and Clusters, are structured in the same way. In Sites windows, the upper view will list all individual Sites while in Clusters windows, Sites will be clustered using rules specified by user (see <u>Identify Modification Sites</u>)

# **Clusters/Sites Windows**

1       2004       bits bits Photen       Proceeding of the second stars constrained by the second stars constrained stare second stars constrained by the second st	ogs 🗶	Platcher	Dataset P	TMs Clusters 🙁															
2       2008       Dortsol John Rubertsol Links       Display (19)       30.31       40.000       79.9060       J.102.08.15. San 1         3       175       BSYE66_HCMNA       NELGS: ADEDDODDDDENNITTEN       Phospho (56)       99.97       55.85 %       160.41       5       1       42.9       200.000       79.9663       121.537.85 cm 1         4       1916       BSYE66_HCMNA       NEAGLSPYSUSDAPSSFELR       Oxidation (M1); Phospho (S17)       90.74 %       66.71       4       1       79       95.9612       1424.553.85 cm 2         6       1026       RRARS, HUMAA       APPEADINTPEPTR       Carbamdomethyl (C7); Phospho (S       90.97 %       50.614       4       1       129       100.00       95.9612       153.9878       153.9878       153.9878       153.9878       153.9878       153.9878       153.9878       153.9878       153.9878       153.9878       153.9878       155.9878       150.00       156.9878       153.9878 <th>1</th> <th>Id</th> <th>Status</th> <th>Protein</th> <th>Peptide</th> <th>PTMs</th> <th></th> <th>P</th> <th>TMs Confid.(MD</th> <th>Score</th> <th></th> <th>Peptide count</th> <th>Site count</th> <th>t Sites Loc.</th> <th>Sites C</th> <th>PTM D.Mass</th> <th>Spect</th> <th>rum title</th> <th>R</th>	1	Id	Status	Protein	Peptide	PTMs		P	TMs Confid.(MD	Score		Peptide count	Site count	t Sites Loc.	Sites C	PTM D.Mass	Spect	rum title	R
3       175       B97E61_HCMVA       MEAGL:PYSVSSDAPSSFELVR       Oxidation (M1); Phospho (5:0)       99.97 %       126.19       4       1       59       99.97       95.9612 16323; Scan 1         4       191       B87E61_HCMVA       MEAGL:PYSVSSDAPSSFELVR       Oxidation (M1); Phospho (5:17)       00.74 %       66.71       4       1       70       91.80       95.9612 16323; Scan 1         5       222       B87E61_HCMVA       MEAGL:PYSVSDAPSFELVR       Carbamidomethy (C1; Phospho (5:       99.95 %       100.00 4       4       1       165       93.598 121 2532; Scan 2         7       1405       B87EA6_HCMVA       RRD2PGC/PEPFERR       Carbamidomethy (C1; Phospho (S       99.95 %       100.00 4       4       1       169       95.9878 1312; Scan 2         1484       B87EA8_HCMVA       RQD2PGC/PEPFSWER       Carbamidomethy (C0; Phospho (S       93.04 %       65.35 4       1       55.66.67       163.6978 5811: Scan 2         1485       B87EA8_HCMVA       RQD2PGC/PEPFSWER       Carbamidomethy (C0; Phospho (S       93.04 %       65.35 4       1       55.67 %       100.00,       201.9432 18952; Scan 2       201.9432 18952; Scan 2 </td <td>2</td> <td>2.00</td> <td>1/ 🗂</td> <td>B8YE68_HCMVA</td> <td>LEGSSADEDDDDDDDDEKNIE I PIK</td> <td>Phospho (54</td> <td>)</td> <td></td> <td>93.91 %</td> <td>6</td> <td>128.01</td> <td>5</td> <td>1</td> <td>1 426</td> <td>100.00</td> <td>/9.</td> <td>9663 13780</td> <td>: Scan 1</td> <td></td>	2	2.00	1/ 🗂	B8YE68_HCMVA	LEGSSADEDDDDDDDDEKNIE I PIK	Phospho (54	)		93.91 %	6	128.01	5	1	1 426	100.00	/9.	9663 13780	: Scan 1	
3       15       0       B87E61_PUMA       PEAGLPT/SSDAPSELUR       Ovidation (M1) Phospho (S17)       99.97 %       126.19       4       1       99.97 /9       95.6612 [162/43: Scan 1         4       131       0       B87E61_PUMA       AFEAGLPT/SSDAPSELUR       Ovidation (M1) Phospho (S17)       90.74 %       66.71       4       1       70       91.80       95.6612 [162/43: Scan 1         6       1026       B87E61_PUMA       AFEAGLPT/SSDAPSELUR       Ovidation (M1) Phospho (S17)       90.74 %       66.71       4       1       625       100.00       136.9878 [132: Scan 23         7       1465       B87EA8_HOMA       RQD2FGPD/EPSGVER       Phospho (S17)       90.94 %       65.33       4       1       56       66.87       136.9878 [131: Scan 32         8       1484       5       B87EA8_HOMA       QALAPGPLATTRK       Carbamidomethyl (C17) Phospho (       75.30 %       116.67       4       2       5/6       100.00;       201.9432 [1856: Scan 2         10       1546       PP71_HCMVA       SQASSPECEPSEAAAISSEAAASSER A Acetyl (Protein N+term); Phospho (       75.30 %       116.67       4       2       5/6       100.00;       201.9432 [1857: Scan 2         11       1549       PP71_HCMVA	J.	2 00	18 🖒	B8YE68_HCMVA	RLFGSSADEDDDDDDDEKNIFTPIK	Phospho (S6	)		95.85 %	6	160.41	5	1	1 427	100.00	79.	9663 12158	: Scan 1	
5       222       B8YE61_HCM/A       AFMEANGNHPEQLCRS/PPPPPR       Carbamidomethyl (C14); Phospho (       100.00 %       74.87       4       1       625       100.00       136.9878 [350: Scan 76         7       1405       B8YE61_HCM/A       RRDS-FGUPPEPFSWER       Phospho (S1; Oxidation (M8)       100.00 %       50.44       4       1       136       99.95       135.9878 [131: Scan 32         8       1484       6       B8YEA8_HCM/A       QOALPGP-LATTRK       Carbamidomethyl (C20); Phospho (S       93.04 %       65.35       4       1       555       66.87       135.9878 [321: Scan 32         1485       6       B8YEA8_HCM/A       QOALPGP-LATTRK       Carbamidomethyl (C20); Phospho (S       93.04 %       65.35       4       1       555       66.87       135.9878 [321: Scan 30         1156       6       P71_HCM/A       SQASS/PGECPSSEAALSEAEAASGSFRR       Acetyl (Protein N+term); Phospho (       75.30 %       116.67       4       2       55       100.00;       201.91432 [3855: Scan 3         12       1552       6       P71_HCM/A       SQASS/PGECPSSEAALSEAEASGSFRR       Acetyl (Protein N+term); Phospho (       75.30 %       116.67       4       2       67       100.00;       201.91432 [3857: Scan 3      <		17	′5 🖒	B8YE61_HCMVA	MEAGLSPYSVSSDAPSSFELVR	Oxidation (M	1); Phospho (S6)		99.97 %	6	126.19	4	1	1 59	99.97	95.	9612 15373	: Scan 1	
0       1026       RRAS2_HUMAN       KPC262EPP.PEPTR       Carbamidomethyl (C2); Phospho (S       99.95 %       100.04       4       1       186       99.95       136.9978 [135: Scan 32         7       1484       6       B8FEA8_HOMA       QDALPGP:LATTPK       Carbamidomethyl (C8); Phospho (S       93.04 %       655.35       4       1       554       93.04       136.9978 [135: Scan 32         1       1484       6       B8FEA8_HOMA       QDALPGP:LATTPK       Carbamidomethyl (C10); Phospho (S       93.04 %       655.35       4       1       554       93.04       136.9978 [135: Scan 32         1       1484       6       B8FEA8_HOMA       HQDALPGP:LATTPK       Carbamidomethyl (C10); Phospho (S       93.04 %       655.35       4       1       555       66.67       136.9978 [135: Scan 2         1       1496       6       P71_HOMA       SQASSPREEPSEAALSEAEAASGSFRR       Acetyl (Protein N+term); Phospho (       43.95 %       120.23       4       2       67; 100.00;       201.91432 [1857: Scan 2         12       1554       6       P71_HOMA       SQASSPREEPSEAGSFRR       Acetyl (Protein N+term); Phospho (       43.95 %       92.20       4       2       67; 100.00;       201.9432 [1857: Scan 2 <t< td=""><td>2</td><td>19</td><td>1</td><td>B8YE61_HCMVA</td><td>MEAGLSPYSVSSDAPSSFELVR</td><td>Oxidation (M</td><td>1); Phospho (S17)</td><td></td><td>90.74 %</td><td>6</td><td>66.71</td><td>4</td><td>1</td><td>1 70</td><td>91.80</td><td>95.</td><td>9612 14624</td><td>: Scan 1</td><td></td></t<>	2	19	1	B8YE61_HCMVA	MEAGLSPYSVSSDAPSSFELVR	Oxidation (M	1); Phospho (S17)		90.74 %	6	66.71	4	1	1 70	91.80	95.	9612 14624	: Scan 1	
0       1026       RRAS2_HUMAN       KPC262EPP.PEPTR       Carbamidomethyl (C2); Phospho (S       99.95 %       100.04       4       1       186       99.95       136.9978 [135: Scan 32         7       1484       6       B8FEA8_HOMA       QDALPGP:LATTPK       Carbamidomethyl (C8); Phospho (S       93.04 %       655.35       4       1       554       93.04       136.9978 [135: Scan 32         1       1484       6       B8FEA8_HOMA       QDALPGP:LATTPK       Carbamidomethyl (C10); Phospho (S       93.04 %       655.35       4       1       554       93.04       136.9978 [135: Scan 32         1       1484       6       B8FEA8_HOMA       HQDALPGP:LATTPK       Carbamidomethyl (C10); Phospho (S       93.04 %       655.35       4       1       555       66.67       136.9978 [135: Scan 2         1       1496       6       P71_HOMA       SQASSPREEPSEAALSEAEAASGSFRR       Acetyl (Protein N+term); Phospho (       43.95 %       120.23       4       2       67; 100.00;       201.91432 [1857: Scan 2         12       1554       6       P71_HOMA       SQASSPREEPSEAGSFRR       Acetyl (Protein N+term); Phospho (       43.95 %       92.20       4       2       67; 100.00;       201.9432 [1857: Scan 2 <t< td=""><td>2</td><td>5 22</td><td>12 🖒</td><td>B8YE61_HCMVA</td><td>AFMEANGNHPEQICRSPPPPLPPR</td><td>Carbamidom</td><td>ethyl (C14); Phospł</td><td>no (</td><td>100.00 %</td><td>6</td><td>74.87</td><td>4</td><td>1</td><td>1 625</td><td>100.00</td><td>136.</td><td>9878 5850:</td><td>Scan 76</td><td></td></t<>	2	5 22	12 🖒	B8YE61_HCMVA	AFMEANGNHPEQICRSPPPPLPPR	Carbamidom	ethyl (C14); Phospł	no (	100.00 %	6	74.87	4	1	1 625	100.00	136.	9878 5850:	Scan 76	
1       1		1 02	16 🖒	RRAS2_HUMAN	KFQEQECPPSPEPTR	Carbamidom	ethyl (C7); Phospho	o (S	99.95 %	6	100.04	4	1	1 186	99.95	136.	9878 1132:	Scan 29	
1       1       45       0       887EAB_HOWA       HCQALPCP-LLASTPK       Carbamidomethyl (210); Phospho (       66.87 %       43.73       4       1       555       66.67       136.9878 (1245); Carbamidomethyl (210); Phospho (       75.30 %       116.67       4       2       556       100.00;       201.9432 (1865); Scan 2         1       1       1549       0       P71_HCMVA       SQASSSPGECPSSEAALSEAEAASGSFGR       Acetyl (Protein N term); Phospho (       75.30 %       110.67       4       2       567       100.00;       201.9432 (1865); Scan 2         1       1       1559       0       P71_HCMVA       SQASSSPGECPSSEAALSEAEAASGSFGR       Acetyl (Protein N term); Phospho (       36.48 %       112.73       4       2       57.100,00;       201.9432 (1894); Scan 2         1       1559       0       P71_HCMVA       SQASSSPGECPSEAALSEAEAASGSFGR       Acetyl (Protein N term); Phospho (       36.48 %       112.73       4       2       71314 (100,00;       201.9432 (1894); Scan 2         1       1559       0       P871_HCMVA       SQASSSPGECPSEAALSEAESGSFGR       Acetyl (Protein N term); Phospho (S12)       144.95 (160,0);       119.99327 (160,7); Scan 1         2       100       0       0.00       0.00       0.00 <td></td> <td>1 40</td> <td>15 🖒</td> <td>B8YEA6_HCMVA</td> <td>RRDSPGGMDEPPSGWER</td> <td>Phospho (S4</td> <td>); Oxidation (M8)</td> <td></td> <td>100.00 %</td> <td>6</td> <td>50.44</td> <td>4</td> <td>1</td> <td>1 139</td> <td>100.00</td> <td>95.</td> <td>9612 1513:</td> <td>Scan 32</td> <td></td>		1 40	15 🖒	B8YEA6_HCMVA	RRDSPGGMDEPPSGWER	Phospho (S4	); Oxidation (M8)		100.00 %	6	50.44	4	1	1 139	100.00	95.	9612 1513:	Scan 32	
1       1		1 48	14 🖒	B8YEA8_HCMVA	QDALPGPCIASTPK	Carbamidom	ethyl (C8); Phospho	o (S	93.04 9	6	65.35	4	1	1 554	93.04	136.	9878 5811:	Scan 76	
11       1 5 99 (a)       PP71_HCM/A       SQASS:PGECPSEFAAAISEAEAASCSFR, Acetyl (Protein N-term); Phospho ( 43.95 %       120,23       4       2       6;7       100.00;       201.943211892: Scan 2         13       1 552 (b)       PP71_HCM/A       SQASS:PGECPSEFAAAISEAEAASCSFR, Acetyl (Protein N-term); Phospho ( 36.46 %       112.73       4       2       7;13       100.00;       201.943211892: Scan 2         14       2 057 (b)       B87E68_HCM/A       LISPMTTTSTSQKPULGK       Phospho (S12)       144.95 %       92.20       4       2       157.46       59.962121630? Sccan 1         16       2 088 (b)       B87E68_HCM/A       LISPMTTTSTSQKPULGK       Phospho (S12)       93.95 %       98.02       4       1       764       50.00       95.962121630? Sccan 1         16       2 088 (b)       B87E68_HCM/A       LISPMTTTSTSQKPULGK       Phospho (S10)       93.95 %       78.24       4       1       99.9653216652can 3         16       2 130 (b)       B87E68_HCM/A       RSTGTAAVGSPW/STIGMK       Phospho (S10)       99.96 %       78.24       4       1       991       99.96 %       79.96633980: Scan 280         Improvide 10 10 10 10 10 10 10 10 10 10 10 10 10	÷.	1 48	15 🖒	B8YEA8_HCMVA	HRQDALPGPCIASTPKK	Carbamidom	ethyl (C10); Phosph	no (	66.87 %	6	43.73	4	1	1 555	66.87	136.	9878 1234:	Scan 30	
12       1522       PP71_HCMVA       QASS:PGEGP::SEAAAISEAEAASGSFGR       Acetyl (Protein N-term); Phospho (       36.48 %       112.73       4       2       7;13       100.00;       201.9432       18572: Scan 2         14       1554       IP71_HCMVA       SQASS:SPGEGP::SEAAAISEAEASGSFGR       Phospho (S12)       44.95 %       92.20       4       2       7;13       100.00;       193.9327 (dots) Scan 1         14       2087       B8YE68_HCMVA       LSSP/HTTSTSQKPVLGK       Phospho (S2)       0.00       95.612 (252: Scan 31)         15       2088       B8YE68_HCMVA       LSSP/HTTSTSQKPVLGK       Phospho (S2)       81.32 %       84.74       4       1       76.5       81.32       79.9663 (296: Scan 41)       V         16       2 130       B8YE68_HCMVA       LSSP/HTTSTSQKPVLGK       Phospho (S10)       99.96 %       78.24       4       1       991       99.56       79.9663 (390: Scan 280)       V         Phospho:643       Multiple Sites:165       V<	/ 1	1 54	ю 🖒	PP71_HCMVA	SQASSSPGEGPSSEAAAISEAEAASGSFGR	Acetyl (Prote	in N-term); Phosph	no (	75.30 %	6	116.67	4	2	2 5;6	100.00;	201.	9432 18865	: Scan 2	
1       152       P71_HCMA       SQASSPECEP:SEAAAISSERA Acctyl (Protein N       20.19422 (352): Scan 2         1       1534       IP71_HCMA       SQASSSPECEP:SEAAISSERA       10.19422 (152): Scan 2         1       1534       IP71_HCMA       SQASSSPECEP:SEAAISSERA       10.19422 (152): Scan 2         14       2087       B87E68_HCMA       L:SPMTTTSTSQKPLGK       Phospho (S1)       44.95 (56)       92.20       4       2       1314 (100 0)       19.9927 (1607): Scan 1         14       2087       B87E68_HCMA       L:SPMTTTSTSQKPLGK       Phospho (S2)       0.00 %       98.02       4       1       76.5       81.32       79.9663 (2162): Scan 1         16       2130       B87E68_HCMVA       L:SPMTTTSTSQKPLGK       Phospho (S1)       84.74       4       1       79.9663 (216: Scan 41         16       2130       B87E68_HCMVA       L:SPMTTTSGKPLGK       Phospho (S1)       99.96       78.24       4       1       991       99.96       79.9663 (390: Scan 280       N         Phospho:643       Multiple Sites: 165       Ste Peptide       Ste Pospon (S)       Phospho(S)	1 1	1 54	19 🖒	PP71 HCMVA	SQASSSPGEGPSSEAAAISEAEAASGSFGR	Acetyl (Prote	in N-term); Phosph	no (	43.95 %	6	120.23	4	2	2 6;7	100.00;	201.	9432 18942	: Scan 2	
14       2 087 (S)       B87E68_HCMVA       LS:PMITTSTSQKPULGK       Phospho (S2): Oxidation (MS)       50.00 %       98.02       4       1       764       90.00       95.9612 (262: Scan 31)         16       2 088 (L)       B87E68_HCMVA       LS:PMITTSTSQKPULGK       Phospho (S2):       81.32 %       84.74       4       1       765       81.32 79.9663 (2166: Scan 41)         16       2 130 (L)       B87E68_HCMVA       LS:PMITTSTSQKPULGK       Phospho (S1):       99.96 %       78.24       4       1       991       99.96 %       79.9663 (2166: Scan 41)       79.9663 (216: Scan 41)       79.9663 (21		1 55	2 🖒	PP71_HCMVA	SQASSSPGEGPSSEAAAISEAEAASGSFGR	Acetyl (Prote	in N-term); Phosph	10 (	36.48 %	6	112.73	4	2	2 7;13	100.00;	201.	9432 18572	: Scan 2	
1       2087 (_6]       B87E68_HCMVA       LSSPMTTTSTSQKPULGK       Phospho (S2); Oxidation (MS)       50.00 %       98.02       4       1       764       50.00       95.0612 (262: Scan 31)         1       2088 (_6]       B87E68_HCMVA       LSSPMTTTSTSQKPULGK       Phospho (S2); Oxidation (MS)       61.32 %       84.74       4       1       765       81.32       79.9663 (2196: Scan 41)         1       2088 (_6]       B87E68_HCMVA       LSSPMTTTSTSQKPULGK       Phospho (S1)       99.96 %       78.24       4       1       991       99.96       79.9663 (2196: Scan 41)       99.96 %       79.966	1	1 55	4 👍	PP71 HCMVA	SOASSSPGEGPSSEAAAISEAEAASGSFGR	Phospho (S1	2); Phospho (S13)		44.95 %	6	92.20	4		2 13;14	100.00;	159.	9327 16807	: Scan 1	
In         2 130         B87668_HCMVA         RSTGTAAVGSPVKSTTGMK         Phospho (\$10)         99.96 %         78.24         4         1 991         99.96 %         79.9663 (980: Scan 280           Phospho:643         Multiple Sites:165           Site Peptides         Site PSMs           Image: Site Site Site Site Site Site Site Site	1	1 2 08	7 🖒	B8YE68 HCMVA	LSSPMTTTSTSQKPVLGK	Phospho (S2	); Oxidation (M5)		50.00 %	6	98.02	4	1	1 764	50.00	95.	9612 1262:	Scan 31	
Int       2 130       B8YE68_HCMVA       RSTGTAAVG:PVKSTTGMK       Phospho (\$10)       99.96 %       78.24       4       1 991       99.36 (\$3980: Scan 280         Phospho:643       Multiple Sites:165         Site Peptides       Site PSMs         B8YE68_HCMVA       RSTGTAAVG:PVKSTTGMK       Phospho:(610)       99.96 %       78.24       4       1 991       99.36 (\$300: Scan 280       N         Phospho:643       Multiple Sites:165       Site PsMs       File       Site Poptide       Site Poptide       Site Poptide       Site Poptide       No       No </td <td>1</td> <td>2 08</td> <td>8 🖒</td> <td>B8YE68 HCMVA</td> <td>LSSPMTTTSTSOKPVLGK</td> <td>Phospho (S3</td> <td>)</td> <td></td> <td>81.32 9</td> <td>6</td> <td>84.74</td> <td>4</td> <td></td> <td>1 765</td> <td>81.32</td> <td>79.</td> <td>9663 2196:</td> <td>Scan 41</td> <td></td>	1	2 08	8 🖒	B8YE68 HCMVA	LSSPMTTTSTSOKPVLGK	Phospho (S3	)		81.32 9	6	84.74	4		1 765	81.32	79.	9663 2196:	Scan 41	
Phospho:643         Multiple Sites:165           Site Peptide         Site PSMs           Image: Site Site Site Site Site Site Site Site	1	2 13	0 13	BSYE68 HCMVA		Phospho (S1	, n)		99 96 9	6	78.24	4		1 991	99 96	79	9663 980 - 5	can 280	۰.
Image: Second	Ph	ospho:643	muruple 5	ites: 165															
Estensistic GEOF STERARISERERARISET ON         Lintrix         2         % SQASSS         86.01         1030         3 Oxidat         175.9         20.00         16954         40.00         20.00           Estensistic GEOF STERARISET ON         MSQASSS         86.01         1030         3 Oxidat         175.9         20.00         20.00           MSQASSS         PATA         102.5         3 Phosp         159.9         159.9         39.99         20.03           Estensistic GEOF STERARISET ON TABLE         100.20         20.00         159.9         100.00         62.9         31.21	Site Pe																		
Image: Second	Site Pe	ptides Si	te PSMs	13 14 P P		-								yl(Protein N	. Phospho			Prob	Ţ
E 4 COASSSP 100 78 995 4 3 Aretyl 201 9 31 21 17064 100 00 62 42 31 21	Site Pe	ptides Si	te PSMs	13 14 P P	ISEAEAASGSFGRLI	TNVE T	1 SQASSSP	92.20 98	31.3 3 P	hosp	159.9	44.95 1680	7:	yl(Protein N	. Phospho	58.71	44.95	Prob	₹
	Site Pe	ptides Si	te PSMs	13 14 P P	ISEAEAASGSFGRLI	TNVE 🕅	1 SQASSSP( 2 2 MSQASSS 8	92.20 98 36.01 10	31.3 3 P 30 3 C	hosp xidat	159.9 175.9	44.95 1680 20.00 1695	7: 4:	yl(Protein N	. Phospho	58.71 40.00	44.95 20.00	Prob I	₽ ^
	Site Pe	ptides si MS(	te PSMs	13 14 P P	ISEAEAASGSFGRLI	TNVE X	1 SQASSSP 2 2 MSQASSS 8 3 MSQASSS 7	92.20 98 86.01 10 74.78 10	31.3 3 P 330 3 C 325 3 P	hosp xidat hosp	159.9 175.9 159.9	44.95 1680 20.00 1695 19.99 1672	7 <mark>:</mark> 4: 7:			58.71 40.00 39.99	44.95 20.00 20.03	Prob	₹₽ ^
	Site Pe	ptides si MS(	te PSMs	13 14 P P	ISEAEAASGSFGRLI	TNVE R	1 SQASSSP 2 2 MSQASSS 8 3 MSQASSS 7	92.20 98 86.01 10 74.78 10	31.3 3 P 330 3 C 325 3 P	hosp xidat hosp	159.9 175.9 159.9	44.95 1680 20.00 1695 19.99 1672	7 <mark>:</mark> 4: 7:			58.71 40.00 39.99	44.95 20.00 20.03	Prob	₽
0	Site Pe	ptides Si MS(	te PSMs	13 14 P P	ISEAEAASGSFGRLI	TNVE X	1 SQASSSP 2 2 MSQASSS 8 3 MSQASSS 7	92.20 98 86.01 10 74.78 10	31.3 3 P 330 3 C 325 3 P	hosp xidat hosp	159.9 175.9 159.9	44.95 1680 20.00 1695 19.99 1672	7 <mark>:</mark> 4: 7:			58.71 40.00 39.99	44.95 20.00 20.03	Prob	₽
θ	Site Pe	ptides Si MS(	te PSMs	13 14 P P	ISEAEAASGSFGRLI	TNVE R	1 SQASSSP 2 2 MSQASSS 8 3 MSQASSS 7	92.20 98 36.01 10 74.78 10	31.3 3 P 330 3 C 325 3 P	hosp xidat hosp	159.9 175.9 159.9	44.95 1680 20.00 1695 19.99 1672	7 <mark>:</mark> 4: 7:			58.71 40.00 39.99	44.95 20.00 20.03	Prob	₽¥ ∧
θ	Site Pe	ptides Si MS(	te PSMs	13 14 P P	ISEAEAASGSFGRLI	TNVE R	1 SQASSSP 2 2 MSQASSS 8 3 MSQASSS 7	92.20 98 36.01 10 74.78 10	31.3 3 P 330 3 C 325 3 P	hosp xidat hosp	159.9 175.9 159.9	44.95 1680 20.00 1695 19.99 1672	7 <mark>:</mark> 4: 7:			58.71 40.00 39.99	44.95 20.00 20.03	Prob	₽ <b>₽</b>
θ	Site Pe	ptides Si MS(	te PSMs	13 14 P P	ISEAEAASGSFGRLI	TNVE R	1 SQASSSP 2 2 MSQASSS 8 3 MSQASSS 7	92.20 98 36.01 10 74.78 10	31.3 3 P 330 3 C 325 3 P	hosp xidat hosp	159.9 175.9 159.9	44.95 1680 20.00 1695 19.99 1672	7 <mark>:</mark> 4: 7:			58.71 40.00 39.99	44.95 20.00 20.03	Prob	₽ ~
	Site Pe	ptides Si MS(	te PSMs	13 14 P P	ISEAEAASGSFGRLI	TNVE TNVE TNVE	1 SQASSSP 2 2 MSQASSS 8 3 MSQASSS 7	92.20 98 36.01 10 74.78 10	31.3 3 P 330 3 C 325 3 P	hosp xidat hosp	159.9 175.9 159.9	44.95 1680 20.00 1695 19.99 1672	7 <mark>:</mark> 4: 7:			58.71 40.00 39.99	44.95 20.00 20.03	Prob	₽ ^
C     P721 HCMVA Protein Sequence Coverage 554 amino acid (calculated <= protein length)	Site Pe	ptides Si MSC	te PSMs	GEGPSSEAAA		TNVE TNVE TNVE	1 SQASSSP 2 2 MSQASSS 8 3 MSQASSS 7	92.20 98 36.01 10 74.78 10	31.3 3 P 330 3 C 325 3 P	hosp xidat hosp	159.9 175.9 159.9	44.95 1680 20.00 1695 19.99 1672	7 <mark>:</mark> 4: 7:			58.71 40.00 39.99	44.95 20.00 20.03	Prob	₽

**Upper View**: This view lists all Modification Sites or Clusters. In this example, a Cluster view is shown. For each of them, many information is given : the number of Peptides belonging to the Cluster, how many sites have been clustered, The Confidence for the site combination, ...

**Bottom Left view**: Graphic illustration of the cluster/site peptides mapped to the protein sequence. All modification locations are represented above the protein sequence. An overview graphic is also given below.

Bottom Right view: Cluster/Site peptides list. The PSMs list is accessible in a hidden tab.



# Display Additional Information on Search Result/Identification Summary

# **Functionality Access**

To display properties of a Search Result or Identification Summary:

- right click on a Search Result/Identification Summary
- click on the menu "Properties"

Note: it is possible to select multiple Search Results/Identification Summaries to compare the values.

# **Properties Window**

Property window opened:

**General Information:** Various information on the analysis (instrument name, peaklist software...)

**Search Properties:** Information extracted from the Result File (date, software version, search settings...)

Search Result Information: Amount of Queries, PSM and Proteins in the Search Result.

Identification Summary Information: Information obtained after validation process

Validation xxx: Information on validation process : parameters used to validate and result

Sql Ids: Database ids related to this item

<u>Note:</u> Identification Summary Number may differ from Validation Results. Indeed, on one hand, peptide matches count in Validation Results takes into account all PSMs that have been validated. On the other hand, the PSM Number in "Identification Summary Information" section considers only PSMs that identify a valid Protein Sets.

jects × Ms Fil	PS	_
,		
Proline_Project	~	
<ul> <li>Identifications</li> <li>Imported</li> <li>TEST</li> <li>Repliacat1</li> <li>Repliacat2</li> </ul>		
	Display Search Result	>
Repliac	Display Identification Summary	,
🗄 📺 Trash	Display lucitation summary	
	Add Dataset	
	Add Identification Folder	
	Copy Search Result	
	Paste Search Result	
	Rename	>
	Clear	
	Delete	
	Import Search Result	
	Import MaxQuant Result	
	Validate Search Result	
	Merge Datasets	>
	Filter ProteinSets	
	Change Typical Protein	
	Identify Modification Sites	
	Generate Spectrum Matches	
	Update Spectrum using Peaklist software	
	Retrieve Protein Sequences	
	Quantify	>
	Export	>
	Properties	-



Gro	up	Туре	F083067
	General Information		
2		Raw File Name	
3		Fasta Files	/local/mascot-2.4//sequence/UPS1UPS2_D/current/UPS1UPS2_D
1		Search Result Name	Gamme Levure UPS1 OEMMA121101
5		Instrument Name	
3		Fragmentation Rule Set	
3		Target Decoy Mode Peaklist Software	CONCATENA
	Search Properties	Peakilst Software	extract_msn.
2	ocurentioperacs	Result File Name	F083067.
		Search Date	
2		Software Name	
3		Software Version	2.
1		Taxonomy	All ent
5		Enzyme	Tryps
3		Max Missed Clivage	
		Fixed Modifications	
3		Variable Modifications Fragment Mass Tolerance	Acetyl(Protein N-term), Oxidation 0.8
		Pragment Mass Tolerance Peptide Charge States	
		Peptide Mass Error Tolerance	
2		Fragment Charge States	
3		Fragment Mass Error Tolerance	
1	Search Result Information		
5		Queries Number	37
3		PSM Number	42
		Protein Number	6
-		PSM Decoy Number	25
		Protein Decoy Number	5
	Identification Summary Information	Protein Sets Number	
2		Protein Sets Number PSM Number	9
-		Peptide Number	2
1		Protein Sets Decoy Number	
5		PSM Decoy Number	
3		Peptide Decoy Number	
1		is_coverage_updated	1
3		result_summary.ptm_dataset	defi
	Validation Parameters		
-		peptide_filters#1 / description	
-		peptide_filters#1 / parameter peptide_filters#1 / properties / threshold_value	PRETTY_R/
-		peptide_inters#17 properties / unearind_value peptide_filters#2 / description	
1		peptide_filters#2 / parameter	SC(
5		peptide_filters#2 / properties / threshold_value	
3		protein_filters / description	protein set filter on peptide co
		protein_filters / parameter	PEP_CO
		protein_filters / properties / threshold_value	
	Validation Results		
-		peptide_results / decoy_matches_count	
		peptide_results / fdr peptide_results / target_matches_count	1.1953
-		peptide_results / target_matches_count protein_results / decoy_matches_count	
		protein_results / decoy_matches_count protein_results / fdr	
-		protein_results / target_matches_count	
	Sql Ids		
		Project id	
3		Dataset id	
)		ResultSet id	
		ResultSummary id	
		Msi Search id	
2		Peaklist Search id	
3		Peaklist Software id	

Property window opened with multiple Identification summaries selected:

#### The color of the type column indicates if the values are the same (white) or different (yellow)

Group	Туре	F083066	F083064	>
	PSMI	Decoy Number	342	346
	Peptide I	Decoy Number	192	194
	is_cove	rage_updated	true	true
Validatio	n Parameters			
Como Valuas	peptide	_expected_fdr	5.0	5.0
Same Values	peptide_filters#	1 / description	eptide match rank filter	peptide match rank filter
	peptide_filters#	1 / parameter	RANK	RANK
	peptide_filters#1 / propertie	es / threshol	1	1
	peptide_filters#	2 / description p	eptide match score filter	peptide match score filter
Different Value	peptide_filters#	±2 / parameter	SCORE	SCORE
	pepude_inters#27properue	es / threshol	15.69	15.66
Valic	ation Results			
	peptide_results / decoy_n	natches_count	342	346
	peptid	e_results / fdr	4.980341	4.9942265
	peptide_results / target_n	natches_count	13392	13510
	protein_results / decoy_n	natches_count	186	187
	protei	n_results / fdr	33.0373	32.324978
	protein_results / target_n	natches_count	940	970



# **Display a Spectral Count**

You can display a generated Spectral Count by using the right mouse popup.

Quantitation	S		
SC- ds2	Display Abundances	Þ	Peptides Ions
🕀 🕁 🕀	Display Identification Summary	1	Peptides
SC - ds 🏦 Trash	Display Exp.Design		Proteins Sets
III III	Rename Delete		New User Window Manage User Windows
	Refine Protein Sets Abundances Clone & Extract Abundances		
	Export Excel		
	Properties		

To have more details about the results, see spectral count result

	Protein Set	Overview	Descriptio	n #Peptide	#Quant. Peptide	Peptides Coun YOC_SP-	1 Basic SC YOC_SP-	Specific SC YOC_SP-	Weighted S YOC_SP-	Peptides Count YOC_SP-	Basic SC YOC_SP-	Specific SC YOC_SP-	Weighted SC YOC_SP- Y
1	B8YE68_HCMVA		tr B8YE68	B8 23	3 233	161	1044	1044	1044.00	164	1020	1020	1020.00 Ty A
2	B8YEA8 HCMVA		tr B8YEA8	B8 10-	4 104	73	398	398	398.00	73	338	338	338.00 Ty
3	B8YE61 HCMVA		tr B8YE61	B8 134	4 134	73	204	204	204.00	73	195	195	195.00 Ty
4	PP71_HCMVA		sp P06726	P 9	3 98	63	242	11	217.94	52	226	1	202.56 Ty
5	B8YEA7_HCMVA		tr B8YEA7	B8 9	5 95	52	231	0	24.06	51	225	0	23.44 Su
6	B8YEB1_HCMVA		tr B8YEB1	B8 60	0 60	37	73	73	73.00	32	61	61	61.00 Ty
7	B8YE83_HCMVA		tr B8YE83	B8 5	3 58	31	71	71	71.00	29	67	67	67.00 Ty
8	B8YEA5_HCMVA		tr B8YEA5	B8 5	1 51	23	39	39	39.00	29	51	51	51.00 Ty
9	B8YE91_HCMVA		tr B8YE91	B8 3	4 34	20	43	43	43.00	21	36	36	36.00 Ty
10	B8YE82_HCMVA		tr B8YE82	B8 2	7 27	15	36	36	36.00	13	35	35	35.00 Ty
11	A0A024R 1N1_H	U	tr  A0A024	R1 26	5 26	15	21	20	20.99	14	23	21	22.75 Ty
12	FLNC_HUMAN		sp Q14315	5 F 3	1 31	7	7	7	7.00	9	9	9	9.00 Ty
13	🍐 #C#P60712		SWISS-PRO	OT 2	3 23	14	32	9	25.49	15	43	11	33.21 Ty
14	ACTG HUMAN		sp P63261	A 2	3 23	13	28	5	10.02	14	34	2	8.76 Ty
- h						Ŷ		DC_SP- YOC_S	SP-	9-1		•	<b>⊢</b> 52
1	I VAELER		50.36	2 424 2556	B8VE91	1		2		- "	X III		<b>⊢</b> 52
	LVAFLER TIRSEAEDSYHFSSAK		50.36 60.58	2 424.2556 3 609.9552		1 1	OC_SP- YC	DC_SP- YOC_S	sp	•			-52
2					88YE91	1 1	1	i T i	<u>^</u>				-52
2 3	TIRSEAEDSYHFSSAK		60.58	3 609.9552	88YE91 88YE91		 1 0			8			-50
2 3 4	TIRSEAEDSYHFSSAK DTSLQAPPSYEESV		60.58 100.35	3 609.9552 2 1050.4804 2 720.3463	88YE91 88YE91		1 0 7	0 7	^ <u>&amp;</u>	8			
2 3 4 5	TIRSEAEDSYHFSSAK DTSLQAPPSYEESV SSNVFDLEEIMR		60.58 100.35 77.28	3 609.9552 2 1050.4804 2 720.3463	88YE91 88YE91 88YE91 88YE91 88YE91		- 1 0 7 1	0 7 1		8 - ·····			-50
2 3 4 5 6 7	TIRSEAEDSYHFSSAK DTSLQAPPSYEESV SSNVFDLEEIMR MTATFLSK GLDDLMSGLGAAGK NIICTSMKPINEDLDEC	Oxidation (	60.58 100.35 77.28 23.56 100.74 32.05	3 609.9552 2 1050.4804 2 720.3463 2 449.7386 2 660.8271 3 900.1106	88YE91 88YE91 88YE91 88YE91 88YE91 88YE91 88YE91			7 1 0 7 1 0 2 0		8 - ····· 6 - ·····			-50
2 3 4 5 6 7 8	TIRSEAEDSYHFSSAK DTSLQAPPSYEESV SSNVFDLEEIMR MTATFLSK GLDDLMSGLGAAGK NIICTSMKPINEDLDEC VLELYSQK	Oxidation ( Carbamido	60.58 100.35 77.28 23,56 100.74 32.05 50.19	3 609.9552 2 1050.4804 2 720.3463 2 449.7386 2 660.8271 3 900.1106 2 490.2770	88YE91 88YE91 88YE91 88YE91 88YE91 88YE91 88YE91 98YE91		1 0 7 1 0 2 0 1	1 0 7 1 0 2 0 1		8 - ····· 6 - ·····			-50
2 1 3 4 5 6 7 8 9 1	TIRSEAEDSYHFSSAK DTSLQAPPSYEESV SSNVFDLEEIMR MTATFLSK GLDDLMSGLGAAGK NIICTSWKPINEDLDEC VLELYSQK DTSLQAPPSYEESVYN	Oxidation ( Carbamido Phospho (S	60.58 100.35 77.28 23.56 100.74 32.05 50.19 50.22	3         609.9552           2         1050.4804           2         720.3463           2         449.7386           2         660.8271           3         900.1106           2         490.2770           2         1090.4628	88YE91 88YE91 88YE91 88YE91 88YE91 88YE91 98YE91 88YE91 88YE91 88YE91		0 7 1 0 2	1 0 7 1 0 2 0 1 0 1 0		8 - ····· 6 - ·····			-50
2 1 3 0 4 5 6 0 7 1 8 0 9 0	TIRSEAEDSYHFSSAK DTSLQAPPSYEESV SSNVFDLEEIMR MTATFLSK GLDDLMSGLGAAGK NIICTS <sup>MK</sup> PINEDLDEC VLELYSQK DTSLQAPPSYEESVYN FAQCYSSYSR	Oxidation ( Carbamido	60.58 100.35 77.28 23.56 100.74 32.05 50.19 50.22 72.08	3 609.9552 2 1050.4804 2 720.3463 2 449.7336 2 660.8271 3 900.1100 2 490.2770 2 1090.4628 2 634.7721	B8YE91		1 0 7 1 0 2 0 0 1 0 1 0	0 7 1 0 2 0 1 0 1 0 1		8 - ····· 6 - ·····			-50 -48 -46
2 1 3 4 5 6 7 1 8 0 9 0 1 4	TIRSEAEDSYHFSSAK DTSLQAPPSYEESV SSNIFDLEEIMR MTATFLSK GLDDLMSGLGAAGK NIICTSMRPINEDLDEC VLELYSQK DTSLQAPPSYEESVYN FAQCYSSYSR AQQNGTDSLDGQ	Oxidation ( Carbamido Phospho (S Carbamido	60.58 100.35 77.28 23.56 100.74 32.05 50.19 50.22 72.08 143.66	3 609.9552 2 1050.4804 2 720.3463 2 449.7386 2 660.8271 3 900.1100 2 490.2770 2 1090.4628 2 634.7721 3 962.4702	B8YE91		1 0 7 1 0 2 0 1 1 0 1 1 9	0 7 1 0 2 0 1 0 1 9		Abundance			-50 -48 -46
2 1 3 4 5 6 7 1 8 7 9 0 1 4 2 1	TIRSEAEDSYHFSSAK DTSLQAPPSYEESV SSNVFDLEEIMR MTATFLSK GLODUMSGLGAAGK NIICTS <sup>W</sup> KPINEDLDEC VLELYSQK DTSLQAPPSYEESVYN FAQCYSSYSR AQQNGTDSLDGQ DTSLQAPPSYEESVY	Oxidation ( Carbamido Phospho (S Carbamido Phospho (S	60.58 100.35 77.28 23.56 100.74 32.05 50.19 50.22 72.08 143.66 74.94	3 609.9552 2 1050.480- 2 720.3463 2 449.7386 2 660.8271 3 900.1106 2 490.2770 2 1090.4625 6 634.7721 3 962.4700 2 1090.4615	B8YE91		1 0 7 1 0 2 0 0 1 0 0 1 9 9 0	0 7 1 0 2 0 1 0 1 9 9 0		8 - ····· 6 - ·····			-50 -48 -46 -44
2 1 3 2 5 1 6 0 7 1 8 0 9 2 1 1 2 2 3 1	TIRSEAEDSYHFSSAK DTSLQAPPSYEESV SSNVFDLEEIMR WITATFLSK GLDDLMSGLGAAGK NILITSK/KPINEDLDEC VLELYSQK DTSLQAPPSYEESVYN FAQCYSSYSR AQQNGTDSLDQC DTSLQAPPSYEESVY KQEVIM/SDSALDCVR	Oxidation ( Carbamido Phospho (S Carbamido Phospho (S Oxidation (	60.58 100.35 77.28 23.56 100.74 32.05 50.19 50.22 72.08 143.66 74.94 77.28	3 609.9552 2 1050.480 <sup>2</sup> 2 720.3463 2 449.7386 2 660.8271 3 900.110 2 490.2770 2 1090.4622 6 634.7721 3 962.4700 2 1090.4615 3 813.3811	B8YE91		1 0 7 1 0 2 0 1 1 0 1 9 9 0 1	0 0 7 1 0 2 0 1 0 1 9 0 1 1					-50 -48 -46 -44
2 1 3 4 5 6 7 1 8 9 0 1 2 1 3 4	TIRSEAEDSYHFSSAK DTSLQAPPSVEESU SSINFDLEEIMR MTATFLSK GLDULYSGLGAAGK NIICTSV-KPINEDLDEC VLELYSQK DTSLQAPPSVEESVYD FAQCYSSVSR AQQNGTDSLDGQ DTSLQAPPSVEESVYD TSLQAPPSVEESVT VCSVAQGTDLIR	Oxidation ( Carbamido Phospho (S Carbamido Phospho (S	60.58 100.35 77.28 23.56 100.74 32.05 50.19 50.22 72.08 143.66 74.94 77.28 75.88	3         609.9552           2         1050.4804           2         249.7384           2         660.8271           3         900.1106           2         490.2772           1090.4622         634.7721           3         962.4702           1090.4628         1090.4628           2         634.7721           3         952.4702           2         634.33811           2         683.8245	B8YE91		1 0 7 1 0 2 0 0 1 1 0 1 9 9 0 0 1 1	0 7 1 2 0 1 2 0 1 9 9 0 1 1 1		Abundance			-50 -48 -46 -44 -44 -42
2 1 3 4 5 1 6 7 1 8 0 7 1 8 0 7 1 8 0 7 1 8 0 9 0 7 1 8 0 8 0 9 0 7 1 8 0 8 0 8 0 8 0 9 0 8 0 9	TIRSEAEDSYHFSSAK DTSLQAPPSYEESU SSNVFDLEEIMR MTATFLSK GLDDLMSGLGAAGK NII.TS*/KPINEDLDEC VLELYSQK DTSLQAPPSYEESVYN FAQCYSSYSR AQQNGTDSLDGQ DTSLQAPPSYEESVY KQEVIM/SDSALDCVR VCSMAQGTDLIR GLDDLMSGLGAAGK	Oxidation ( Carbamido Phospho (S Carbamido Phospho (S Oxidation (	60.58 100.35 77.28 23.56 100.74 32.05 50.22 72.08 143.66 74.94 77.28 75.88 89.70	3         609.9552           2         1050.4800           2         149.7386           2         649.7386           2         649.7386           2         640.8271           3         900.1106           2         1090.4622           3         962.4700           2         1090.4615           3         813.8311           2         652.8307	B8YE91		0 7 1 0 2 0 1 1 0 1 9 0 1 1 9 0 0	0 7 1 0 2 0 1 1 0 1 9 9 0 1 1 1 0 0		8			-50 -48 -46 -44 -44 -42
2 1 3 4 5 6 7 1 8 9 0 1 4 1 2 1 3 1 4 1 5 0 6 1	TIRSEAEDSYHFSSAK DTSLQAPPSYEESV SSNVFDLEEIMR MTATFLSK GLDDLMSGLGAAGK (GLDDLMSGLGAAGK UTSLQAPPSYEESVYN FAQCYSSYSR AQQNGTDSLDGQ DTSLQAPPSYEESVY KQEVM/SDSALDCVR V.SMAQGTDLIR GLDDLMSGLGAAGK KGPGPPSSDASTA	Oxidation ( Carbamido Phospho (S Carbamido Phospho (S Oxidation (	60.58 100.35 77.28 23.56 100.74 32.05 50.19 50.22 72.08 143.66 74.94 77.28 75.88 89.70 87.25	3         609.9552           2         1050.4804           2         449.7386           2         660.8271           3         900.1106           4         490.2770           2         1090.4628           2         634.7721           3         962.4702           1         962.4702           3         813.3811           2         653.82430           3         1068.1981	B8YE91		0 7 1 0 2 0 0 1 1 0 1 9 9 0 1 1 1 0 0 0 0 0 0	0 7 1 0 2 0 1 1 0 1 9 0 1 1 1 0 0 1 1 0 0 0 1 1 0 0 0 0					-50 -48 -46 -44 -42 -42 -40
2 1 3 0 5 1 6 0 7 1 8 0 9 0 1 2 3 1 3 1 5 0 6 1 5 0 6 1 7 0	TIRSEAEDSYHFSSAK DTSLQAPPSVEESU SSINFDLEEIMR MTATFLSK GLDDL/SGLGAAGK NILCTS/KRINEDLDEC VLELYSQK PAQCYSSVSR AQQNGTDSLDGQ DTSLQAPPSVEESVY DTSLQAPPSVEESVY DTSLQAPPSVEESVY UCSMAQGTDLIR GLDUMSGLGAAGK KGPGPPSDASTA GSTWLYR	Oxidation ( Carbamido Phospho (S Carbamido Phospho (S Oxidation ( Carbamido	60.58 100.35 77.28 23.56 100.74 32.05 50.19 50.22 72.08 143.66 74.94 77.28 75.88 89.70 87.25 28.60	3         609.9552           2         150.460           2         120.3465           2         449.7386           3         900.100           2         490.2770           2         1909.462           2         493.4772           3         962.4702           2         193.381           3         863.8245           2         652.8300           3         1068.1981           2         441.7267	B8YE91		<pre></pre>	0 0 7 1 0 2 2 0 1 1 0 1 1 9 9 0 1 1 1 0 0 0 0 0 0 0 0		8			-50 -48 -46 -44 -44 -42 -40 -38
2 1 3 4 5 6 7 1 8 9 6 9 6 1 4 1 4 5 6 1 4 5 6 1 5 7 7 8 8 5 7 7 8 8 8	TIRSEAEDSYHFSSAK DTSLQAPPSYEESU SSNVFDLEEIMR MTATFLSK GLDDLMSGLGAAGK NIICTS <sup>WK</sup> PINEDLEOU VLELYSQK DTSLQAPPSYEESVYN FAQCYSSYSR AQQNGTDSLDGQ DTSLQAPPSYEESVYN KQEVN4SDSALDCVR GLDDLMSGLGAAGK KGPGPPSSDASTA GSTWLYR SVYSQHVTSSEAVSH	Oxidation ( Carbamido Phospho (S Carbamido Phospho (S Oxidation ( Carbamido	60.58 100.35 77.28 23.56 100.74 32.05 50.19 50.22 72.08 143.66 74.94 77.28 89.70 87.25 28.60	3         609.9552           2         1050.4800           2         1494.7386           2         640.8271           3         900.1106           4         490.2777           2         1090.4622           3         62.4702           3         62.4702           2         1090.4615           3         813.3811           2         652.8307           3         1068.1983           3         618.6105	B8YE91           B8YE91		0 0 7 1 0 2 0 0 1 1 9 0 1 1 1 0 0 1 1 0 0 1 1 0 0 0 0	0 0 7 1 0 2 0 1 1 0 1 9 9 0 1 1 1 0 0 1 1 0 0 0 0 0		B = B =			-50 -48 -46 -44 -44 -42 -40 -40 -38 -35
2 1 3 4 5 5 1 6 6 6 9 6 10 7 11 4 12 6 13 4 14 5 16 6 17 6 10 7 10 7 11 4 13 8 14 5 16 7 10 7 1	TIRSEAEDSYHFSSAK DTSLQAPPSYEESV SSNVFDLEEIMR MTATFLSK GLDDLWSGLGAAGK NILTS*KPINEDGU VELYSQK DTSLQAPPSYEESVYN FAQCYSSYSR AQQNGTDSLDGQ DTSLQAPPSYEESVY KQEVM*SDSALDCVR VCSMAQGTDLIR GLDDLMSGLGAAGK KGPGPPSSDASTA GSTWL/R SVYSQHVTS:EAVSH TMQLIPDDYSNTH	Oxidation ( Carbamido Phospho (S Carbamido Phospho (S Oxidation ( Carbamido	60.58 100.35 77.28 23.56 100.74 32.05 50.19 50.22 72.08 143.66 74.94 77.28 75.88 89.70 87.25 28.60 46.29 43.17	3         609.9552           2         1050.4800           2         499.7386           2         660.8271           3         900.1106           4         490.2770           2         1090.4628           2         634.7721           3         813.3811           2         652.8301           3         1068.1981           2         441.7265           3         816.6103           3         626.9603	B8YE91		0 0 7 1 0 2 0 1 1 0 1 0 1 1 0 1 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 7 1 0 2 0 1 1 0 1 9 9 0 1 1 1 0 0 1 1 0 0 0 0 0		B = B =	Hoc hoc	Koc Ko	-50 -48 -46 -44 -42 -40 -38 -35
2 1 3 0 6 7 1 8 9 0 1 2 0 1 2 0 1 2 0 1 4 9 1 4 9 1 4 9 1 7 0 1 7 0	TIRSEAEDSYHFSSAK DTSLQAPPSYEESU SSNVFDLEEIMR MTATFLSK GLDDLMSGLGAAGK NIICTS <sup>WK</sup> PINEDLEOU VLELYSQK DTSLQAPPSYEESVYN FAQCYSSYSR AQQNGTDSLDGQ DTSLQAPPSYEESVYN KQEVN4SDSALDCVR GLDDLMSGLGAAGK KGPGPPSSDASTA GSTWLYR SVYSQHVTSSEAVSH	Oxidation ( Carbamido Phospho (S Carbamido Phospho (S Oxidation ( Carbamido	60.58 100.35 77.28 23.56 100.74 32.05 50.19 50.22 72.08 143.66 74.94 77.28 89.70 87.25 28.60	3         609.9552           2         150.460           2         120.3465           2         449.7386           3         900.100           2         490.2770           2         1909.462           2         490.2770           2         1909.462           3         962.4702           3         962.4702           2         1909.4612           3         813.3811           2         683.8245           2         552.8300           3         1068.1981           2         441.7267           3         626.9003           2         570.8361	B8YE91		0 0 7 1 0 2 0 0 1 1 9 0 1 1 1 0 0 1 1 0 0 1 1 0 0 0 0	0 0 7 1 0 2 0 1 1 0 1 9 9 0 1 1 1 0 0 1 1 0 0 0 0 0		B = B =	HOC SD. Fr. Cerev	Voc. 50-17-Co 80-17-Co-80-17-Co 80-17-Co-80-17-Co 80-18-19-19-10-10-10-10-10-10-10-10-10-10-10-10-10-	-50 -48 -46 -44 -42 -40 -38 -35

The overview is based by default on the weighted spectral count values. (Note: if you sort on the overview column, the sort is based on max (value-mean (values))/mean (values). So, you will obtain the most homogenous and confident rows first)

For each compared dataset, are displayed:

- status ( typical, sameset, / )
- peptide numbers



- the basic spectral count
- the specific spectral count
- the weighted spectral count
- the selection level

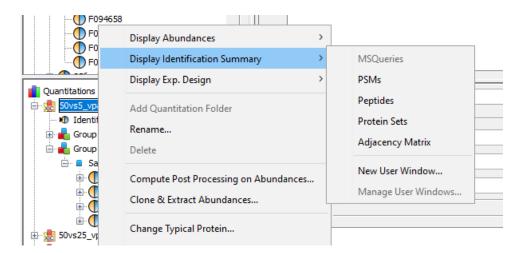
User can change the information displayed by the overview using the table settings icon (  $\gtrsim$  ).

# **Display a XIC**

To display a XIC, right click on the selected XIC node in the Quantitation tree, and select "Display Abundances", and then the level you want to display:

Quantitations	t	
🕕 🕕 Gra	Display Abundances	Peptides lons
E dro	Display Identification Summary	Peptides
Phosph	Display Exp. Design	Proteins Sets
Quant .	Add Quantitation Folder	Modification Sites
🛃 re 🛃 test	Rename	Modification Clusters
- 🛃 Quant	Delete	New User Window
⊷ 📩 Quant … 🏦 Trash	Compute Post Processing on Abundances	Manage User Windows
	Clone & Extract Abundancer	

Note: You can also display the identification summary used as reference for the quantitation from the popup menu in the quantitation tree:



#### **Display Protein Sets**

#### **Protein Sets**

By clicking on "Display Abundances" / "Protein Sets", you can see all quantified protein sets. For each quantified protein set, you can see below all peptides linked to the selected protein set and peptides lons linked to the selected peptide. For each peptide lon, you can see the different features and the graph of the peakels in each quantitation channel.



B         B         C         B         D <thd< th=""> <thd< th=""> <thd< th=""> <thd< th=""></thd<></thd<></thd<></thd<>	P	rotein Set	Overview	Description	ı	#Peptide	e #Quan Peptide		#Quant. PSM 83064	s Ab	undance 3064	= #Qua F08306	ant. PSMs i6	<ul> <li>Abundance</li> <li>F083066</li> </ul>	<ul> <li>#Quant. PSMs</li> <li>F083067</li> </ul>	Abundance F083067	#Quant. PSMs F083068	Đ
star       22       61       172 525 136       57       174 800 400       49       174 372 976         star       22       61       172 525 136       57       174 800 400       49       174 372 976         star       19975 (2244 54       3       220 46       3       220 46       3       220 46       3       220 78 0.12         star       100 7027 Augritres (-H	6	G3P2_YEAST		sp P00358	G3P2_YEA.		31	3		9	12 056 782	2	1	11 491 365	17	11 211 114		1
a Perfection       solp 41000 Hose 76, YE       30       1       3       2 624 444       3       2 802 476       2       2 78 012         a METE_LEXAT       solp 4000 Hose 76, YE       31       24       44       3       2 802 476       2       2 78 012         a METE_LEXAT       solp 4000 Hose 76, YE       28       20       0       3 99724       40       58 647 992       34       37 20 158       1       3 451 923         a FEA_MEAL       solp 41500 Hose 76, YE       31       25       25       59 944 50 26       60 255 992       28       60 316 072         a FEA_MEAL       solp 91320 HOSE 70, YE       23       18       39       76 49 532       27 1 45 272       33       69 91 1000         a FEA_MEAL       solp 91320 HOSE 70, YE       24       19       29       65 596 7060       31       81 49 104         a FEA_MEAL       solp 91320 HOSE 70, YE       24       19       29       65 596 7060       31       81 49 4104         a FEA_MEAL       solp 1300 HOSE 70, YE       24       19       20       85 596 496       24       62 16 5000       31       81 49 4104         a FEA_MEAL       solp 1300 HOSE 70, YE       20       24 596 70	7 [	ALBU_HUMAN_UPS		P02768ups	ALBU_HU		36	24		12	7 190 832	2	10	7 328 332	10	7 633 882		T
Status         Poptide Seque         PTM         Score         Charge         M/2         2 370 456 172         337 465         309 459 1208           Status         Poptide Seque         PTM         Score         Charge         m/2         2 370 456 172         337 465         309 1208         1         345 192           Status         Poptide Seque         PTM         Score         Charge         m/2         Overview         #Abundance	8	HSP75_YEAST		sp P11484	HSP75_YE.		31	22		61 1	72 525 136	i	5	7 174 860 400	49	174 372 976		
1       3       4       3       3       4	9 [	HSP76_YEAST		sp P40150	HSP76_YE.			1		-	2 624 544	ł						
1       0       23       73       94       45105       60       94       653       95       <		METE_YEAST		sp P05694	METE_YE			24		45	60 299 724	ł	4	58 647 992	34	57 221 908		
1       1       25       25       99       9	1.1									-								
1       0																		
6       0		-																
<ul> <li>KGEFLIVEAST             </li> <li>KGEFLIVEAST             </li> <li>KGEFLIVEAST             </li> </ul> Status        Peptide Seque       PTMs       Score       Charge m/z       Overview       Abundance       #Abundance																		
Status       Peptide Seque       PTMs       Score       Charge       m/z       Overview       Abundance       * Abundance																		
Status       Peptide Seque       PTMe       Score       Charge       m/z       Overview       - Abundance       - Abundance       - Abundance       - Abundance         Status       Peptide Seque       PTMe       Score       Charge       m/z       Overview       - 5401 614       5496 122       495 r.       5.006 f.       -				sp P16861	K6PF1_YE.		25	23		34	36 112 208	3	34	4 34 291 688	30	35 702 136		
WHURPEPTTIG.       70.52       2       1054.5780       1022190       1119.563       9.         C       MUNPLEMK       82.43       2       588.3205       2005.4571       12022190       1119.563       9.         C       MUNPLEMK       82.43       2       588.3205       2005.4571       12022190       1119.563       9.         C       MUNPLEMK       82.43       2       588.3205       2005.4571       12022192       215       2005.4571       1202.158       2005.4571       1202.158       2.566.60.66.60.60.60.60.60.60.60.60.60.60.6	<																,	>
C       VFNLPLFPTTTIG       76.92       2       1054.5780       5401614       5496122       497       7       7       770.47       2       788.926       1002 140       1192550       200       7	st	atus Peptide Seque.	PTMs	Score	Charge	m/z	Overview		ndance Al	bundance 3066	e 🛛 Abun F08306		1		X Axis: Quant. Q	Channel 🗸 Y A:	xis: Abundance	
Control       2019       208.9161       1022 190       1 119 563       9         Control       AVOVTALEWK       82.48       2 588.305       2 000 437       123 550       200       7       6.665         Control       State       97.10       2 050 569       2 784 552       2 58       9       6.665       9       1000 52       1000 52       <	n		_	76.92	2	1054 5780		5	401.614 5	496 122	4 90 A		8,0	E6 7				
AVD/TALEMK       82.43       2       588.3205       2005457       1923550       205         PULCEVSTUPLOK       53.22       2       482.7714       2600569       2.784632       2.55         PULCEVSTUPLOK       63.08       2       751.9542       370.486       3094900       2.7         GUVAALHVDFW       61.04       2       697.4023       791.030       668 523       7.7         VICEPUSTUPLAGK       73.91       2       798.931       1300.722       1.374616       1.8         ADROCIDEPTS       41.51       3       1019.5620       706.692       6.667       6.67         ADROCIDEPTS       41.51       3       1019.5620       706.672       6.67       6.67         ADROCIDEPTS       41.51       3       1019.5620       706.672       6.67       6.67         ADROCIDEPTS       41.52       3       570.6732       789.240       385.172       3.       2.       587.240       385.172       3.       2.       587.240       385.172       3.       2.       587.240       385.172       3.       2.       587.240       385.172       3.       2.       587.6732       789.366       798.3061       798.306       798.306												1	7,0	E6		/	-6,8	BE7
Implement       37,21       2       482,7714       2620 569       2.784 632       2.5         Implement       63,88       2       751,9542       3.370 486       3.994 900       2.77         Implement       697,4023       1.99 5101       3.0066       2.77       3.006       4.665       7.5         Implement       5,0067       1.00       2.697,4023       1.00       2.666       1.3         Implement       5,0067       3.0066       1.3       1.00       2.163 46       2.4014       2.2         Implement       5,0067       3.0066       1.00       2.6067       6.007       1.000       2.6067       6.007       1.000       2.6067       6.007       1.000       2.6067       6.007       1.000       2.6067       6.007       1.000       2.0068       1.000       2.0068       1.000       2.000       1.000       2.000       1.000       2.000       1.000       2.000       1.000       2.000       1.000       2.000       1.000       2.000       1.000       2.000       1.000       2.000       1.000       2.000       1.000       2.000       1.000       2.000       1.000       2.000       1.000       1.000       2.000       0.000       0	1 The												6.0	6				
C       PVILGPVSYLFLCK       63.58       2       751.9542       3 370.486       3 094 900       2 76         GLPVAALHOPVR       61.04       2       697.4023       791.030       668 623       77         V       70.93.91       799.899       1 300 722       1 374 616       1.8         V       70.405202478       41.51       3 1015.5620       77       3 351 2       508.3057       668 623       77         ADXOSCICERS       41.51       3 1015.5620       77       1300 722       1 374 616       1.8       77       3.066       6.067       6.0693067       6.083069       6.083069 <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td>_</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td>DE/</td>								_										DE/
C       YDLSPIDILFMARR       73,91       2       79,894       179 523       180 183       17         P       TOMAXLAR.       95,81       3       624 1059       130 722       130 720       130 720       130 720       130 720       130 720       130 720       130 720       130 720       130 720       130 720       130 720       130 720       130 720       100 720       100 720			к										_ <b>2</b> 5,0	E6 -			-6.4	4E7
C       YDLSPIDILFMARR       73.91       2       79.894       1.79       1.79       22.12       1.83       1.7         Provide Line       95.11       3       024.1059       1.70       1.90       72.2       1.374.66       1.3         Provide Line       41.51       3       1019.5620       704.059       616.967       6         Diverviting       28.51       3       2.508.3057       2.163.346       2.44 014       2.2       3.         Provide Line       32.16       4       570.6732       1.372.40       3.8       2.066       -													5 <b>5</b> 4,0	E6 -				
Image: Constraint of the second se	2	YDLSPIDTLFAM	R	73.91	2			1	799 523 1	801 183	175		1				-6,2	2E7
LDEVWITK       S8.13       2       508.057       2163.346       2.244.014       2.2         VIRPPINCCLSRX       28.85       3       570.6732       S87.240       385.172       3.         E CHTRPLICE       32.16       4       578.0911       212.220       215.346       2       2       2         I.OE6       Posso64       Posso65       Posso65       Posso66       Posso65       Posso66	9	TQAMQLALALR.		59.51	3	824.1059		1	300 722 1	374 6 16	13		<b>Z</b> 3,0					
DEWVITK       S8.13       2       503.057       216.3346       2.24014       2.21         PROPUNCULSION       20.85       3       570.6732       S87.240       385.172       3       1.066       1.066       P033064       F083066       F083069	2			41.51	3	1019.5620			704 059	616 967	61		2,0	E6 - · · · · Pere			-6,0	0E7
Image: Second status       20.803       3       570.6732       1       387.240       305 172       x         Image: Second status       32.16       4       578.0911       1       212.220       215 340       x       y         Image: Second status       32.16       4       578.0911       1       212.220       215 340       x       y         Image: Second status       Total sta	2	LDEVVVITK		58.13				2	163 346 2	244 0 14	2 21		1.0	=6 -				
Image: Status       PTMs       Score       Charge       m/z       RT       Proc       Protei       Abundance       Abundance       Abundance       Abundance       Image: Abu	2			28.85	3	570.6732			587 240	385 172	3.					<u>a</u>	5,8	SE/
Comparison       Comparison <td>9</td> <td>ELGIHTRPVLLGF</td> <td></td> <td>32.16</td> <td>4</td> <td>578.0911</td> <td></td> <td></td> <td>212 220</td> <td>215 340</td> <td>2. v</td> <td></td> <td></td> <td>E083064 E</td> <td>-083066 E083067</td> <td>E083068 E08306</td> <td>9 E083070</td> <td></td>	9	ELGIHTRPVLLGF		32.16	4	578.0911			212 220	215 340	2. v			E083064 E	-083066 E083067	E083068 E08306	9 E083070	
b. Peptides Ions XIC Features          Peptide Status       PTMs       Score       Charge       m/z       RT       Pro       Protei       Abundance       Abundance       Abundance       Posso64       F083066       Posso64       Posso66       Posso66       Posso64       Posso66       Posso64       Posso66       Posso64       Posso66       Posso64       Posso66       Posso66       Posso64       Posso64       Posso66       Posso64       Poss	<										>			1 000001				
I       Peptide       Status       PTMs       Score       Charge       m/z       RT       Prot       Protei       • Abundance       • Abundance       • Abundance       • Baundance													alian I	in and Dist. M	_		-14.	_
Peptide       Status       PTMs       Score       Charge       m/z       RT       Pro       Protei	ti. Pe	eptides Ions XIC Feat	ures										apnic: L	inear Plot 🗸 X	Axis: RI	Axis: Inter	isity 🗸	
2.5E6 -	P	eptide Status	PTMs S	core Charge	m/z	RT Pr	o Protei.	- At	oundance A 3064 F08	Abundano 33066	-		3,5E6 -					
2.0E6 - 1.5E6 - 1.0E6 -	1	VDELFK 🍫		37.21	2 482.771	0 75.97	1 METE_Y	2	620 569		^		3,0E6 -		MAA			
1.066 -											(	10	2,5E6 -		MALIA	MA		
1.026 -											(	sity 📨	2,0E6 - ···			NOV N		
1.066 -												Inten	1,5E6 -	0	r 1 h			
5.0ES -													1,0E6 -	M	1 4()	(W	h	
													5,0E5 -	M	, Ke	<u>(</u> )	N.	

The overview is based by default on the abundance values.

Note: if you sort on the overview column, the sort is based on max (value-mean (values))/mean (values). So, you obtain the most homogenous and confident rows first.

For each quantitation channel, are displayed:

- the raw abundance
- the peptide match count
- the abundance
- the selection level

By clicking on the using the "table setting" icon  $\aleph$ , you can choose the information you want to display or change the overview.

#### **Peptides**

The middle part of the window lists all peptides of the selected Protein set with the same kind of quantitative data. The status column indicates whether the peptide was used or not for protein set abundances. On the right part, a graph allows you to see the variations of the abundance (or raw abundance) of a peptide in the different quantitation channels.



#### **Features and Peakels**

You can see the different features in the different quantitation channels and the graph of the peakels:

XIC F	Pep	tides Ions XIC Featu	res								
Featu	res	(4)									
T		Мар	Quant. Channel	m/z	Charge	Elution Time (min)	Apex Intensity	Intensity	Duration (sec)	Predicted El. Time (min)	<b>F</b>
di.	1	OEMMA121101_45b	F067909	727.9713	3	18.56	315 826	545 185	37	18.50	<b>)</b>
	2	OEMMA121101_58b	F067911	727.9704	3	18.11	575 144	1 114 286	47		
	3	OEMMA121101_47b	F067901	727.9710	3	18.49	287 700	615 248	40	18.31	i –
W	4	OEMMA121101_43b	F067900	727.9710	3	18.95	372 579	695 966	43	18.71	l

By clicking on 💷 you can display either:

- the peaks of isotope 0 in all quantitation channels
- all isotopes for the selected quantitation channel:

(6)						<u> </u>				
Мар	Quant. Channel	m/z	Charge	RT	rş.	₽	6,0E6 -			
OEMMA121101_36	b 🛛 F083064	1054.5767		2	^	Ⅲ	5.0E6 -	$\wedge$		
OEMMA121101_38		1054.5762		2		5500	5,020			
DEMMA121101_40		1054.5759		2			4,0E6 -	N		
DEMMA121101_61		1054.5762		2		<b></b>	sity		N I	
OEMMA121101_63		1054.5783		2				/ /		
OEMMA121101_65	b = F083070	1054.5776		2			3,0E6 -			
<					~		2,0E6 -	Time: 96,23	Time: 97,3	97,2
<b>`</b>							20,2	20,4 20,0	1.	27,2

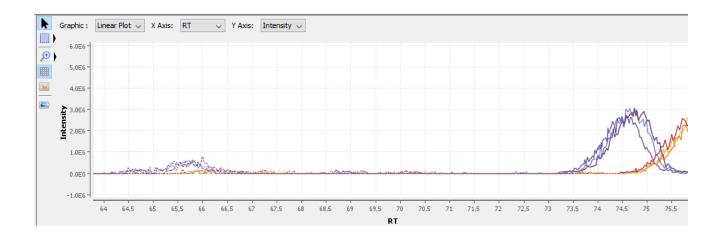
By clicking on 💷 you can see the chromatograms of the features and their first time scan and last time scan in mzScope. For more details see the mzScope section.

It is also possible to extract the corresponding chromatogram for one or all of the features.

	s (6)				
	Мар	Quant. Channel		m/z	
1	OEMMA121101_36b				482.7
2	OEMMA121101_38b				482.7
3	OEMMA121101_40b				482.7
	OEMMA121101_61b				482.7
	OEMMA121101_63b				482.7
6	OEMMA121101_65b	= F083070	Сору	cell	
L			Selec	t All	
L			Displ	ay	>
Ŀ	<	_	View	Selected Dat	a
	×		View	All Data	
			Extra	ct Selected X	IC
			_	ct All XIC	

The resulting chromatograms will be displayed in the same windows as peakel.





#### **Display Peptides**

By clicking on "Display Abundances" / "Peptides", you can see:

- identified and quantified Peptides
- non identified but quantified peptides
- identified but not quantified peptides (linked to a quantified protein)

	Sta	Peptide Sequence	PTMs	Score	Charge	m/z	RT	Prot	Protei	Overvie	W PSMs		Abundance F083064	Quant. PSMs	<ul> <li>Abundan</li> <li>F083066</li> </ul>	PSMs count	Abund F083063	
24	ß	NLSVEDAAR		49.59	2	487.7487	23.34	1	CATA			0	643 183		641 97	4 0		62
25	ຝ	DLFNAIATGK		42.76	2	525.2849	69.99	1	CATA			0	88 747	(	205 92	0 0		1
26	ຝ	DPASDQMQHWK		68.12	2	671.7960	27.77	1	CATA			0	0	(	) (	0 0		
27	₫	NFTEVHPDYGSHIQ/	AL	35.75	3	728.6957	74.97	1	CATA			0	0	(	97 62	3 0		
28	6	DPILFPSFIHSQK		23. <mark>95</mark>	3	510.2776	84.59	1	CATA			0	372 169	(	261 54	0 0		3
29	-	GPLLVQDVVFTDEM/		53.71	3	730.3622	91.96	1	CATA			0	0	(	) (	0 0		
30	6	NPVNYFAEVEQIAF	PSN Oxi	41.53	3	1064.8340		1	CATA			0	206 530	(	185 36	9 0		1
31	6	DPASDQMQHWK	Oxi			679.7931			CATA			0	98 339	(	122 31	7 0		
32	6	FYTEDGNWDLVGN		53.12	3	840.0742			CATA			0	129 663		133 16			
33	6	GAGAFGYFEVTHDI	к	71.78	2	856.9185			CATA			0	92 758		) (	0 0		
34	<u>د</u>	LNVITVGPR		55.22	2	484.7981	48.58	1	CATA			1	729 576	(	865 70	4 0		9
Pe	eptide S	equence Stat	us PTMs	Score	Charg	ge m/z	RT	Pr	rot Pro	otein Sets	Pep. match count		Abundance Pe 83064 mat			■ Pep. match co F083067		Abund F083067
FY	TEDGN\	VDLVGNNT 👩		32	51	2 1259.6	5108	97.96	1 CA	FA_HUM		0	125 180	0	185 452		0	
FY	TEDGN	VDLVGNNT 🏑		53.	12	3 840.0	0740	97.94	1 CA1	FA_HUM		0	129 663	0	133 165		0	

The lower view lists all peptide ions (specific charge) of selected peptide. The status column indicates if the ion is valid or not and if it was used for peptide quantitation.

#### **Display Peptides Ions**

By clicking on "Display Abundances" / "Peptides Ions", you can see:

- all identified and quantified Peptides Ions
- non identified but quantified peptides lons



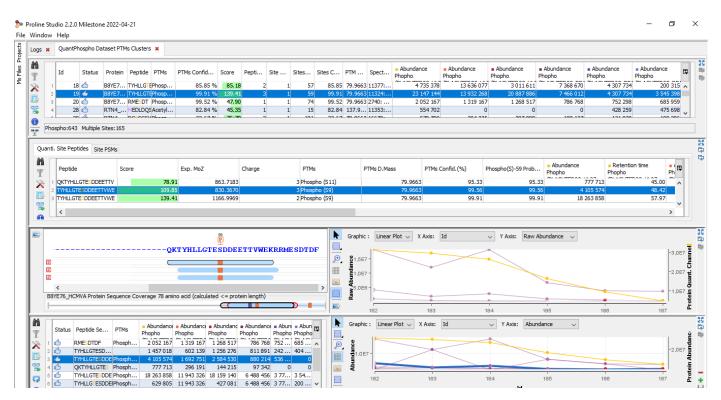
#### Logs 🗴 📩 quant\_100vs10 Protein Sets 🗴 📩 quant\_100vs10 Peptides 🗴 📩 quant\_100vs10 Peptides Ions 🗴

	Peptide Sequence	PTMs	Score	Charge	m/z	RT	Protein S	Protein Sets	Pep. match	Abundance F083064	Pep. match count	Abundance F083066	∎ Pe F083
1	VTSNVVLVSGEGER		32.56	2	723.3838	39.10	1	SKP1_YEAST	0	1 065 472	0	1 028 804	
2	ETYGEMADCCAK	Carbami	70.19	2	717.7703	23.53	1	ALBU_HUMAN	0	0	0	0	
3	TFESEAAHGTVTR		100.21	2	703.3387	20.95		IDHC_YEAST,	4	1 405 549	6	1 185 391	
4	KLNEIDNKR		43.48	2	565.3196	10.08	1	YP260_YEAST	0	0	0	0	
5	YFLDALPVALLGMNAD		57.91	3	1063.5338	111.69	1	RIR2_YEAST	0	1 674 545	1	1 510 852	
6	GDTPDQGHLQTR		<b>44.99</b>	2	662.8149	11.26	1	CLH_YEAST	1	104 071	1	129 788	
7	VEFLEDTSR		45.63	2	548.2687	42.03	1	RS28A_YEAST	2	5 523 036	4	5 571 094	
8	NANPWGGYSQVQSK		65.91	2	768.3644	41.60	1	COX5A_YEAST	0	387 750	2	494 283	
9	NDLTASQLSDKINDVR		29.19	2	894.9583	53.25	1	BGL2_YEAST	0	293 968	0	239 616	
10	VGQAVETVGQAGRPK		26 <mark>.44</mark>	3	499.6104	18.59	1	RPN1_YEAST	1	1 446 326	0	1 192 612	
11	VGQAVETVGQAGRPK		81.99	2	748.9123	18.56	1	RPN1_YEAST	0	278 058	2	341 131	
12	ADIHLVELLYYVEELDS		41.68	3	1036.2253	115.45	1	GSTA1_HUMA	0	64 956	0	86 812	
13	NATFPGVQMK		3 <mark>8.70</mark>	2	546.7785	45.86	1	PDC1_YEAST	2	33 662 596	2	33 767 420	
14	WAGNANELNAAYAAD		106.89	2	999.4583			PDC1_YEAST	10	36 896 336	7	33 997 268	
15	WAGNANELNAAYAAD		<b>4</b> 3.12	3	666.6412	63.14	1	PDC1_YEAST	1	3 922 132	2	4 059 540	
16	LQLLKPFKPWDGK		24.11	4	393.2344	82.01	1	ACON_YEAST	0	0	0	0	
17	QLQLIQVEIK		3 <mark>8.78</mark>	2	606.3718	79.83	1	HS104_YEAST	0	899 909	0	1 005 325	
18	RKPVTEAR		27.94	2	478.7854	7.06	1	TRFL_HUMAN	0	18 459	0	18 459	
19	LYQTEPSGIYSSWSAQ		91.45	2	1122.5495	79.47	1	PSA7_YEAST	1	605 001	1	598 297	
20	TNANLPEGEKK		46.44	2	600.8143	10.06	1	THRC_YEAST	1	417 456	2	396 757	
21	TNANLPEGEKK		43.02	3	400.8786	10.06	1	THRC_YEAST	2	777 439	0	640 429	
22	LPNSNVNIEFATR		79.11	2	737.8857	60.35	1	VDAC1_YEAST	4	2 909 117	5	2 801 511	

#### **Display Modification Sites/Clusters**

By clicking on "Display Abundances" / "Modification XXX" (Site, Clusters or Annotated Clusters), you can see the same Clusters/Sites information as in "Identification Summary/Display / Modification XX", See <u>Display Modification</u> description

In addition, quantitative information is computed and displayed.



**Upper View**: This view lists all Modification Sites or Clusters. For each of them, in addition to cluster information, quantitative data is given. Abundance is calculated using peptide's abundance sum.

**2nd view**: Cluster/Site peptides list. The PSMs list is accessible in a hidden tab. Peptide quantitative information are displayed the same way it is done in other quantitative views.



**3rd view** : Graphic illustration of the cluster/site peptides mapped to the protein sequence is shown on the left. On the right part, a graph allows you to see the variations of the abundance (or raw abundance) of these peptides in the different quantitation channels.

Bottom Right view: This view is the same as "Peptide" view in "Display Xic Protein Sets" above.

#### **Display Experimental Design and Parameters**

By clicking on "Exp. Design > Parameters", you can see the experimental design and the parameters of the selected XIC.

Task	s Log 🛛 🕺 📩 XIC Exp.	Design 🛛
	Exp.Design Exp. Para	ameters
	Fo6790     Fo6791     Fo6791     Fo6791     Fo6792     Fo6792	01 w File : OEMMA121101_47b.mzdb 00 w File : OEMMA121101_43b.mzdb 02 w File : OEMMA121101_56b.mzdb 20 9 w File : OEMMA121101_45b.mzdb 11 w File : OEMMA121101_58b.mzdb
Tasks	Log 🛛 😹 XIC Exp. Des	sign %
1	Exp.Design Exp. Paramet	ters
	XIC Parameters	
	Extraction moz tolerance	(ppm): 5.0
	Extract XIC from:	All detectable features
	✓ Deisotoping Identifica	ation Based
		Normalization Master Map
	Feature Clustering	<b>9</b>
	moz tolerance (ppm):	5.0
	time tolerance (s):	15.0
	time computation:	Most Intense
	intensity computation:	Most Intense

If you have launched the "compute post processing ..." on the XIC, you can also display the corresponding parameters.



Logs	🗴 🛃 quant_100vs10 Exp. Design 🔹 🧏 quant_100vs10 Protein Sets 🔹 📩 quant_100vs10 Peptides		
	Exp. Design Exp. Parameters Compute Post Processing		器庫
	Pep. selection Pep. configuration Prot. configuration		-
	Pentides Selection Method: Razor and Specific Specify peptides to consider for quantitation Discard Miss Cleaved Peptides	~	
	Discard Modified Peptides		
	Z Acetyl (Protein N-term)		
	Carbamidomethyl (C)		
	Oxidation (M)		
	Modified Peptide Filtering Method: Discard all forms	$\sim$	
	Discard Pep Ions Sharing Peakels		

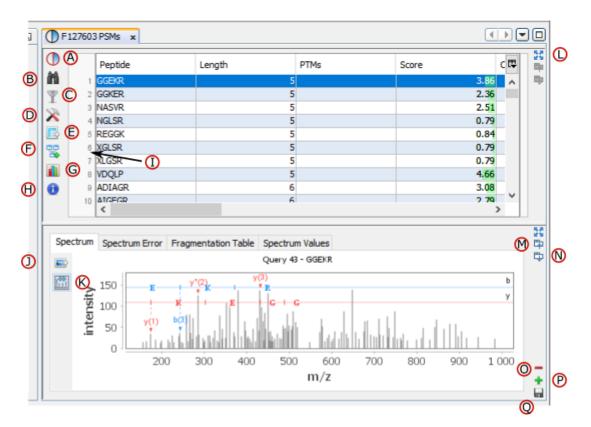
#### **Display Map Alignment**

By clicking on "Exp. Design > Map Alignment", you can see the map of the variation of the alignment of the maps compared to the map alignment of the selected XIC. You can also calculate the predicted time in a map from an elution time in another map.





# **Frame Toolbars Functionalities**



A: Display Decoy Data.

- B: Search in the Table. (Using \* and ? wild cards)
- C: Filter data displayed in the Table
- D: Display settings dialog (you can modify displayed columns and perform double sorting)
- E: Export data displayed in the Table
- F: Send to Data Analyzer to compare data from different views
- G: Create a Graphic : histogram or scatter plot . Only on PSMs table
- H: Display number of entities in the table (number of PSMs / Peptides / Proteins...)
- I: Right click on the marker bar to display Line Numbers or add Annotations/Bookmarks
- J: Export view as an image
- K: Generate Spectrum Matches (specific to spectrum grahic)
- L: Expands the frame to its maximum (other frames are hidden). Click again to undo.
- **M**: Gather the frame with the previous one as a tab.
- N: Split the last tab as a frame underneath
- O: Remove the last Tab or Frame
- P: Open a dialog to let the user add a View (as a Frame, a Tab or a splitted Frame)
- Q: Save the window as a user window, to display the same window with different data later



# **Create a custom User Window**

You can lay out your own user window with the desired views.

You can do it from an already displayed window, or by using the right click mouse popup on a dataset like in the following example (Use menu "Search Result>New User Window..." or "Identification Summary>New User Window...")

ojects %			-		Tasks	Log	
ProjectTest(	63)				₹ ⊘		ic 14
b Identification						•	13
All Impo		_				-	1
⊕ 💼 Trasl						-	10
	Identification Summary		PSM				9
	Properties		Peptides				8 7
	Add		Protein Sets				6
	Merge		New User Wind	ow	/		5
	Validate		Manage User W				4
	Change Typical Protein			Т		╞	3 2
	Generate Spectrum matches						1
	Compare With SC					×	0
	Export						
	Compute Statistical Reports (MSDiag)						
	Change Description						
	Rename				-		
	Delete				Task:	L	
					[ Tim	esta	amp
					Ask	: Tim	e:

In the example, the user has clicked on "Identification Summary>New User Window..." and selects the Peptides View as the first view of his window.

5-	User Defined Window	×
Window Title		
Title : Peptides		
Views		
name	description	T\$
PTM Protein Sites	PTM Protein Sites of an Identification Summary	^
MSQueries	All MSQueries of an Identification Summary	
Protein Set	All Protein Sets of an Identification Summary	
Peptides	All Peptides of an Identification Summary	
PSM	All PSM of an Identification Summary or corresponding to a Peptide Ins	
		•
	V OK X Car	ncel

You can add other views by using the '+' button.



	Peptide	Score	Calc. Mass	Exp. MoZ	Ppm	Charge	Missed Cl.	Protein Set Count	RT	PTM	C,
1	GVLGYTEDAV	172.61	3568.72	1190.58	3.25	3	0	1	110.56		
2	VAVDDPSVLA	164.99	2309.20	1155.61	1.92	2	0	1	105.21		-
3	AANLGGVAVS	162.46	1971.98	987.00	0.02	2	0	1	76.36		
4	QLSLWGADN	162.36	2091.92	1046.97	0.36	2	0	1	74.59		
5	RYGASAGNV	159.58	3412.71	1138.58	2.36	3	1	2	111.50		
6	IIYVDDGVLSF	156.29	2265.18	1133.60	-0.71	2	0	1	130.23		
7	FAIPAINVTSS	153.57	2159.16	1080.59	2.34	2	0	1	94.27		
8	VILFIGDGSLQ	152.44	2432.33	1217.18	0.72	2	0	1	122.24		
9	TPGLSNATQV	149.59	2121.07	1061.54	0.87	2	0	1	76.14		
10	FIAEGSNMGS	149.52	2297.10	1149.56	2.21	2	0	1	87.82		
11	FQYIAISQSD	149.20	1947.85	974.93	-0.59	2	0	1	54.46	Carbamidom	
12	VINDAFGIEE	149.11	. 2574.30	1288.16	2.05	2	0	1	96.50		
13	SEGITDIEES	148.57	2110.95	1056.48	0.30	2	0	1	73.00	Acetyl (Prote	
14	AVGDNDPIDV	147.54	2516.26	1259.14	1.68	2	0	1	96.63		
15	SIVPSGASTG	146.03	1839.91	920.96	0.10	2	0	2	64.56		
16	EAQADAAAE	145.20	2370.08	1186.05	0.10	2	1	1	70.72		
17	SINPDEAVAY	144.42	2576.26	1289.14	4.15	2	0	2	93.19		
18	AAEEADADA	142.60	2340.97	1171.49	0.67	2	0	1	82.40		
19	HVDELLAECA	142.13	1611.76	806.88	-0.69	2	0	1	48.23	Carbamidom	
20	GYLADDIDAD	141.97	2652.22	1327.12	2.55	2	0	1	90.14		
21	LNLPTNSSISV	141.06	2072.07	1037.05	0.91	2	0	1	84.74		
22	VGQFDQVLNL	140.29	3580.70	1194.57	1.57	3	0	1	98.42	Carbamidom	1
23	AVGSLTFDEN	139.62	2213.08	1107.55	1.47	2	0	1	86.63		
24	SGETEDTFIA	139.06	1820.92	911.46	-0.14	2	0	2	130.29		
25	QAFDDAIAEL	139.04	2143.98	1073.00	2.03	2	0	2	92.62		Ļ

In this example, the user has added a Spectrum View and he saves his window by clicking on the "Disk" Button.

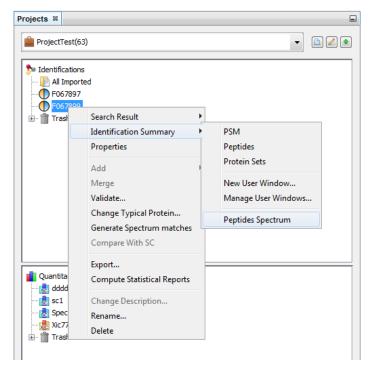
	Peptide	Score	Calc. Mass	Exp. MoZ	Ppm	Charge	Missed Cl.	Protein Set Count	RT	PTM	<b>₽</b>
	1 GVLGYTEDAV	. 172.61	3568.72	1190.58	3.25	3	0	1	. 110.56		
	2 VAVDDPSVLA	164.99	2309.20	1155.61	1.92	2	0	1	. 105.21		
	3 AANLGGVAV	162.46				2		1	. 76.36		
	4 QLSLWGADN					2		1	. 74.59		
	5 RYGASAGNV	159.58				3		2	111.50		
	6 IIYVDDGVLSF					2		1	. 130.23		
	7 FAIPAINVTS	153.57				2	0	1	. 94.27		
	8 VILFIGDGSL	152.44				2		1	. 122.24		
	9 TPGLSNATQV	. 149.59				2		1	. 76.14		
	10 FIAEGSNMGS					2		1	87.82		
	11 FQYIAISQSD 12 VINDAEGIEE	149.20				2	0	1	. 54.46 96.50	Carbamidom	-
	250 000 -			Query 2836	8 - GVLGYTEDAV	/SSDFLGDSHSSI	FDASAGIQLSPK				
	200 000 -										
÷	150 000										
6	<u><u><u></u><u></u><u></u><u></u><u></u><u></u><u></u><u></u><u></u><u></u><u></u><u></u><u></u><u></u><u></u><u></u><u></u><u></u></u></u>										
intensity	-			ha tri		th					

The user selects 'Peptides Spectrum' as his user window name



	Peptide	Score	Calc. Mass	Exp. MoZ	Ppm	Charge	Missed Cl.	Protein Set Count	RT	PTM	r,
1	GVLGYTEDAV	. 172.61	3568.72	1190.58	3.25	3	0	1	l 110.56		-
2	VAVDDPSVLA	. 164.99	2309.20	1155.61	1.92	2	0	1	105.21		
3	AANLGGVAV	162.46	1971.98	987.00	0.02	2	0	1	76.36		
	QLSLWGADN	162.36	2091.92			2	0	1	. 74.59		
	RYGASAGNV	159.58		1138.58		3		2	111.50		
-	IIYVDDGVLSF					2			130.23		
	FAIPAINVTS	153.57	2159.16	1080.59		2			94.27		
elect	t Window Name	e			0.72	2			122.24		
					0.87	2			76.14		
Win	dow Name : Pe	ptides Spectrum			2.21	2			87.82		
					-0.59	2	0		96.50	Carbamidom	-
_			✓ ОК		GVLGYTEDAV	/SSDFLGDSHSSI	FDASAGIQLSPK				
	200 000 -				L						
intensity	150 000 -										
inte	100 000 -										
	50 000 -										

Now, the user can use his new 'Peptides Spectrum' on a different Identification Summary.



# **Filter Tables**

You can filter data displayed in the different tables thanks to the filter button at the top right corner of a table.



Logs	Logs × F024297 Proteins ×								
$\bigcirc$		Protein							
H.	1	ap  P 19097  FAS2_YEAST							
$\nabla$	2	splP07149 FAS1_YEAST							
20	3	sp P10592 HSP72_YEAST							
<b> </b> X	4	sp P16521 EF3A_YEAST							
	5	sp P00549 KPYK1_YEAST							
22	6	sp P10591 HSP71_YEAST							
•	7	sp P32324 EF2_YEAST							
	8	sp P07259 PYR1_YEAST							
	9	sp P11484 HSP75_YEAST							
	10	sp P40150 HSP76_YEAST							
	11	SOLDOOSEO DOK VEAST							

When you have clicked on the filter button, a dialog is opened. In this dialog you can select the columns of the table you want to filter thanks to the "+" button.

In the following example, we have added two filters:

- **one on the Protein Name column** (available wildcards are \* to replace multiple characters and ? to replace one character)

- one on the Score Column (Score must be at least 100 and there is no maximum specified).

Filters
Filter(s)
Protein = V GLPK*
AND 100.0 <= Score <=
Peptides 🗸
V OK Cancel

The result is all the proteins starting with GLPK (correspond to GLPK\*) and with a score greater or equal than 100.

Note: for String filters, you can use the following wildcards: \* matches zero or more characters, ? matches one character.

Protein	Score	
		2503.8
GLPK_ECOLI		2503.8
GLPK_ECOBW		2503.8

# **Search Tables**

In some tables, a Search Functionality is available thanks to the search button at the top right corner.



Logs	×	F024297 Protein Set ×	
		Protein Set	Description
M ┥	€₄	SPIPOU549 KPYK1_YEAST	Pyruvate kinase 1 OS=Saccharom
	5	sp P16521 EF3A_YEAST	Elongation factor 3A OS=Sacchar
2	6	sp P10591 HSP71_YEAST	Heat shock protein SSA1 OS=Sac
<u>×</u>	7	sp P32324 EF2_YEAST	Elongation factor 2 OS=Saccharo
	8	sp P07259 PYR1_YEAST	Protein URA2 OS=Saccharomyces
먫	9	sp P11484 HSP75_YEAST	Heat shock protein SSB1 OS=Sacc
•	10	sp P40150 HSP76_YEAST	Heat shock protein SSB2 OS=Sacc
	11	sp P00560 PGK_YEAST	Phosphoglycerate kinase OS=Sac
	12	sp P38972 PUR4_YEAST	Phosphoribosylformylglycinamidin
	4.0	ID00005 IENOD VEACT	Factors 2.00. Carebonness and

When you have clicked on the search button, a floating panel is opened. In this panel you can select the column searched and fill in the searched expression, or the value range.

For searched expressions, two wild cards are available:

- '\*' : can replace all characters
- '?' : can replace one character

In the following example, the user searches for a protein set whose name contains "PGK".

Logs	×	F024297 Protein Set	×	
		Protein Set		Description
	×F	Protein Set 🗸 🗸	Protein S	et = ↓ *PGK*
	0			near mock protein 35AT 05-5ac
	7	sp P32324 EF2_YEAST		Elongation factor 2 OS=Saccharo
	8	sp P07259 PYR1_YEAST		Protein URA2 OS=Saccharomyces
2	9	sp P11484 HSP75_YEAST		Heat shock protein SSB1 OS=Sacc
	10	sp P40150 HSP76_YEAST		Heat shock protein SSB2 OS=Sacc
	11	sp  P00560  PGK_YEAST		Phosphoglycerate kinase OS=Sac
	12	sp P38972 PUR4_YEAST		Phosphoribosylformylglycinamidin
	13	sp P00925 ENO2_YEAST		Enolase 2 OS=Saccharomyces cer

You can do an incremental search by clicking again on the search button of the floating panel, or by pressing the Enter key.

# **Graphics**

### Create a Graphic

There are two ways to obtain a graphic from data:

- 1. In the windows with PSM of a Search Result or of an Identification Summary, you can ask for the display of a histogram in a new window to check the quality of your identification.
- 2. In any window, you can click on the '+' button to add a graphic (Scatter Plot or Histogram) as a view in the same window



	Peptide	Score	MsQuery	Rank	Calc. Mass	Exp. MoZ	Ppm	Charge	Missed Cl.	Ion	PTM	Protein Sets	E
1	ILFQK	21.38	113	2	647.40	648.41	-0.42	1	0				-
2	LIFAGK	28.56	113		. 647.40	648.41	-0.42					RS27A_YEAST	
3	LLFQK	21.38	113		647.40	648.41	-0.42						
4	LVDLIK	25.99	182		. 699.45	350.73	-0.28					PHSG_YEAST	
5	LVDLLK	25.99	182		699.45	350.73	-0.28						
6	LVDLIK	24.51	185	1	699.45	350.73	0.63	2	0			PHSG_YEAST	
7	LVDLLK	24.51	185	1	699.45	350.73	0.63	2	0				
8	LVDLIK	22.2 <mark>3</mark>	187	1	. 699.45	350.73	1.17	2	0			PHSG_YEAST	
و	LVDLLK	22.23	187	1	699.45	350.73	1.17	2	0				
10	VPFGGVK	31.28	200	1	. 702.41	352.21	0.24	2	0			ALDH6_YEAST	
11	AAFIER	28.59	223		. 705.38	353.70	-0.37					RIR4_YEAST	
12	AAFIER	23.5 <mark>3</mark>	225	1	. 705.38	353.70	1.47	2	0			RIR4_YEAST	
13	LADFLK	24.19	227	1	. 705.41	353.71	-0.91	2	0			SYYC_YEAST	
14	AIDLFK	25.18	230	1	705.41	353.71	2.06	2	0			SYRC_YEAST	
15	AIDLFK	25.12	232	1	705.41	353.71	2.65	2	0			SYRC_YEAST	
16	AGAFITK	34.03	239	1	. 706.40	354.21	-0.79	2	0			PYRD_YEAST	
17	AGAFITK	41. <mark>5</mark> 3	240	1	706.40	354.21	0.20	2	0			PYRD_YEAST	
18	KAMITK	27.74	240	2	706.40	354.21	-4.54	2	1		Oxidation (M3)		
19	FGMDLK	34.44	256	1	. 709.35	355.68	-0.34		0			RIR4_YEAST	
20	FGMDLK	23.01	258	1	. 709.35	355.68	0.61	2	0			RIR4_YEAST	
21	DHLLGR	24.18	264	1	. 709.39	355.70	-0.91	2	0			RL16B_YEAST	
22	HILSIK	22.97	268	1	. 709.45	355.73	-0.69	2	0			PYR1_YEAST	
23	HLISIK	22.97	268	1	. 709.45	355.73	-0.69	2	0				
24	HILSIK	23.1 <mark>6</mark>	269	1	. 709.45	355.73	-0.58	2	0			PYR1_YEAST	
25	HLISIK	23.16	269	1	709.45	355.73	-0.58	2	0				
26	HILSIK	24.81	272	2	709.45	355.73	-0.18	2	0			PYR1 YEAST	

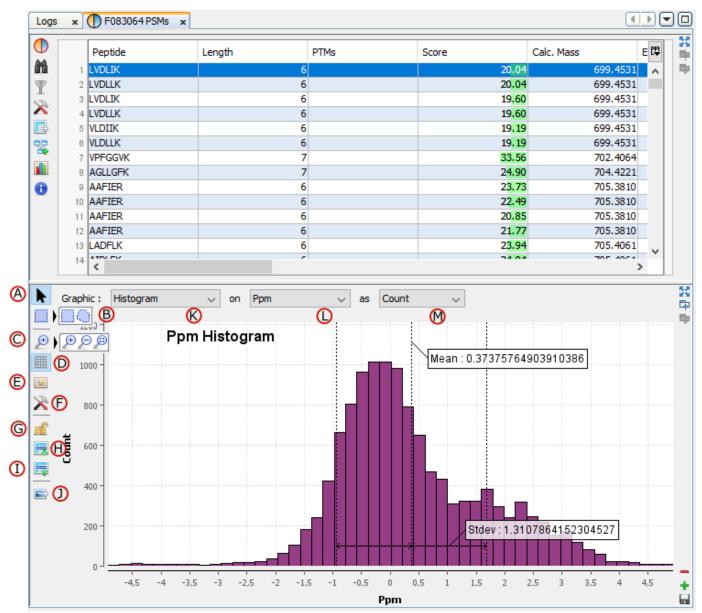
If you have clicked on the '+' button, the Add View Dialog is opened and you must select the Graphic View

og	🛛 🗶 F067												•	
	Peptide	Score	MsQuery	Rank	Calc. Mass	Exp. MoZ	Ppm	Charge	Missed Cl.	Ion	PTM	Protein Sets	E.	
1	ILFQK	21.38	113	2	647.40	648.41	-0.42	1	. 0				-	Ę
2	LIFAGK	28.56	113		647.40	648.41	-0.42					RS27A_YEAST		
3	LLFQK	21.38	113		647.40	648.41	-0.42							
(		75.00	100	1	500 AE	250 72	95.0		-					
1	Add a View													
10	Views													
Ш.	name						escription					Ę		
н.	Graphic						raphics : Histo	aram / Centte	v Dlat					
	Proteins						raphics : Histo roteins for a P	- · · ·	r Piot					
	FIOLEINS													
1	Spectrum							-						
1 1 1 1	Spectrum						pectrum of a P	-						
1 1 1 1 1 1 1 1								-						
1 1 1 1 1 1 1 1 1	Spectrum View Positio	on : 🔊 Bek	ow 🔘 Tab	bed 💿 Sp	itted			-				Ţ		
1 1 1		on: 🔿 Bele	ow 🔘 Tab	bed 💿 <u>Sp</u>	itted			-			✓ OK	▼ Cancel		
1 1 1		on: 🔿 Bek	ow 🔘 Tab	bed 💿 Sp	itted			-		[	✓ OK			
1 1 1 2 2		on : 🔘 Bek	ow 🔿 Tab	bed 🍥 Spl	itted			-		[	✓ ОК			
1 1 1 2 2		on : ⊚ Bek	ow ○ Tab 268	bed 💿 Sp	itted 709.45	S		eptide		[	✓ ОК			
1 1 1 2 2 23	View Positio		268	bed () 50		355.73	Dectrum of a P	reptide 2			✓ ОК	Cancel		
1 1 1 2 2 23 24	View Positio	22.97	268 269	bed () 50	709.45	355.73 355.73	Dectrum of a P	eptide 2 2	0		✓ ОК			
1 1 2 2 23 24 25	View Positio	22.97 23.16	268 269	1 1 1	709.45 709.45 709.45	355.73 355.73 355.73	-0.69 -0.58 -0.58	eptide 2 2 2 2	0		✓ ОК	Cancel		



# **Graphic options**

#### Graphic Toolbar



A: When this button is selected, you have the "Pointer Mode" activated.

In this mode :

- If you move with the left mouse button pressed on the middle of the graphic, you can scroll along the X and Y Axis.

- If you move with the right mouse button pressed from the top/left corner to the bottom/right corner, a zooming rectangle is displayed. When you release the mouse button, a zoom in according to the zooming rectangle is performed.

- If you move with the right mouse button pressed from the bottom/right corner to the left/top corner, a view all is done.

B: When this button is selected, you have the "Selection Mode" activated.

By clicking on the black right arrow, you can switch between the square selection mode and the lasso selection mode.



In this mode:

- Use the left mouse button to do a square or lasso selection.
- The right mouse button works as with the "Pointer Mode"

**C**: Zoom out / Zoom in / View all. Click on the black right arrow to select the zooming mode.

D: Display/Remove Grid toggle button

**E**: Display/Hide View All Map. The goal of this map is to display the whole graphic in a small zone even when you have zoomed

F: Open a settings dialog for the graphic. You can modify for example colors or bins of an histogram.

**G**: Lock/Unlock incoming data. If it is unlocked, the graphic is updated when the user applies a new filter to the previous view (for instance Peptide Score >= 50) If it is locked, changing filtering on the previous view does not modify the graphic.

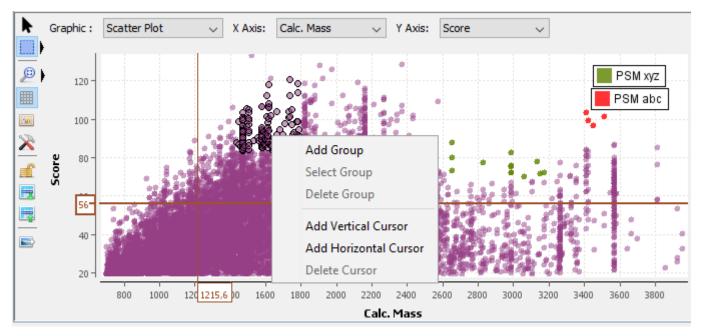
H: Select Data in the graphic according to data selected in the table in the previous view.

I: Select data in the table of the previous view according to data selected in the graphic.

J: Export graphic to image

K: Select the graphic type: Scatter Plot / Histogram

L/M: Select data used for X / Y axis.



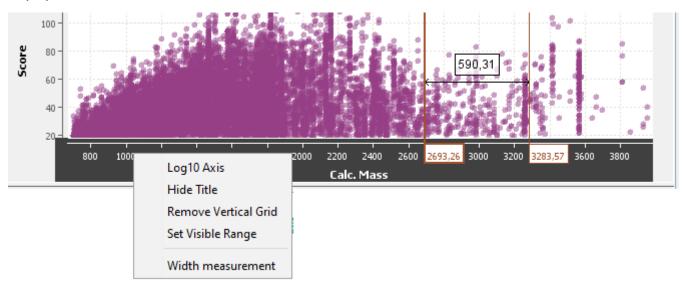
#### Popup Menu on Graphic Area

By right mouse click on the graphic area, you get a popup with several menus:

- Add Group: after selecting some data, you can create a group defined by a name and a color (there are two groups in the example : PSM xyz and PSM abc)
- Select Group: Right click on a data group to select them.
- Delete Group: Right click on a data group and delete it
- Add Vertical Cursor: Add a vertical cursor, its X position is displayed on the X Axis.
- Add Horizontal Cursor: Add an horizontal cursor, its Yposition is displayed on the Y Axis.



- Delete Cursor: Right click on a cursor and delete it.



#### Popup Menu on Axis

By right mouse click on an axis, you get a popup with several menus:

- Log 10 / Linear Axis : toggle the display of the axis between log and linear.
- Hide / Display Title : hide the axis title if you want more space for the graphic
- Remove / Add Vertical Grid
- Set Visible Range : to set the minimum and maximum values of the axis
- Width measurement : display 2 cursors and the measurement of the space between the cursors.

#### Zooming

There are several ways to perform zoom actions.

#### Zoom in:

- Click on the Zoom in button in the toolbar.
- Press the right mouse button and drag to the right bottom direction. A red box is displayed. Release the mouse button when you have selected the area to zoom in.
- Put the mouse over an axis and roll upward the mouse wheel. The zoom in is centered around the middle of the axis. If you press on Ctrl Key at the same time, the zoom in is performed around the mouse position.

#### Zoom out:

- Click on the Zoom out button in the toolbar.
- Put the mouse over an axis and roll downward the mouse wheel. The zoom out is centered around the middle of the axis. If you press on Ctrl Key at the same time, the zoom out is performed around the mouse position.



#### View All:

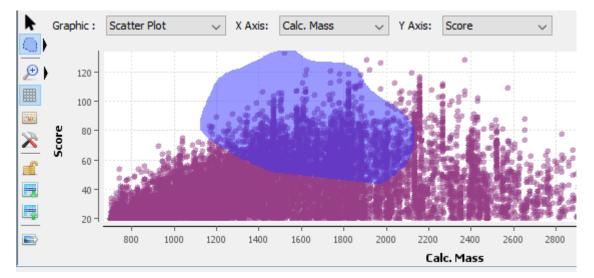
- Click on the View All button in the toolbar.
- Press the right mouse button and drag to the left top direction. When you release the mouse button, the zooming is reset to view all.

#### Selecting

To be able to select data, you must be in "Selection Mode"

**Select**: You can select with a rectangle area or with a lasso according to the selected button. Press the left mouse button and drag the mouse to surround the data you want to select. When you release the button, the selection is done. Or left click on the data you want to select. It is possible to use the Ctrl key to add to the previous selection.

**Unselect**: Left click on an empty area to clear the selection.



#### View All Map



Click on View All Map button to display the map. This map always displays the whole graphic and the zoomed area. You can directly zoom on the view all map. You can resize it and move it.



# **Quality Control**

# Search Result QC

You can run a Quality Control on any leaf Search Result, that is to say an imported Result File not a merged search result. It consists in a transversal view of the imported data: rather than visualising the results per PSM or Proteins, results are sorted according to the score, charge state...

#### Choose the menu option:

rojects × MzDB	Files	- Task	s Log X Tutoriel Formation De
in Tutoriel Format	ion Déc 2016 🗸 🖉 💌		Exp. MoZ per charge and score
<ul> <li>Identifications</li> <li>Al Importe</li> <li>Vald and n</li> <li>Posso</li> <li>Posso</li> <li>Posso</li> <li>Posso</li> <li>Posso</li> </ul>	d erge 64 57	- 1	Resultset Target PSM Decoy PSM
- C FOR	Display Search Result		MSQueries
0 108	Display Identification Summary	1	PSM
8- 🎬 Trash	Add Dataset		Proteins
	Rename		Quality Control
	Clear Delete		New User Window Manage User Windows
		-	

#### Settings

Configure some settings before launching the process

- Score windows: you can split your data in different groups based on the score. The default groups are : less than 20, between 20 and 40, between 40 and 60, over 60
- Max rank: data can be filtered to get a view focused on the best ranks. Default is to consider only the first rank.

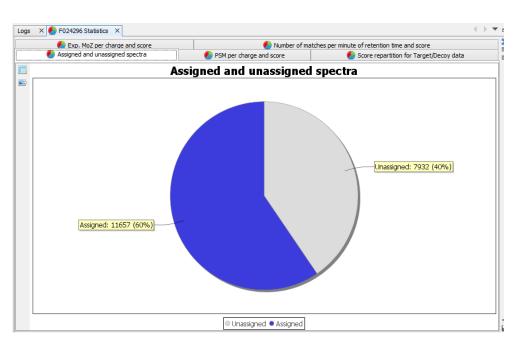
*	Define settings for Statistical Reports	×
Statistical	Reports Settings	
General	Settings	¥
Statistic	al Settings	
Score	windows (ex: 20-40-60) : 20-40-60	
	Max rank : 1	
Save	Cancel	9
Save	V K K Cance	U

#### QC results

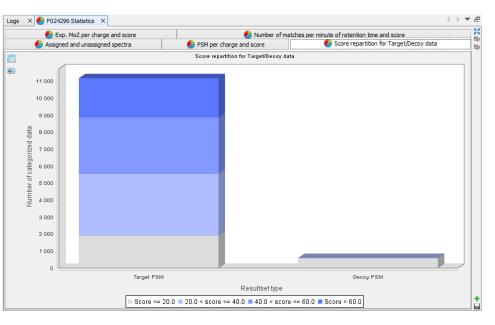
The report will appear in a matter of seconds (depending on the amount of data to be processed). You will get the following tabs:



#### Assigned and unassigned spectra: Pie chart presenting the ratio of assigned spectra

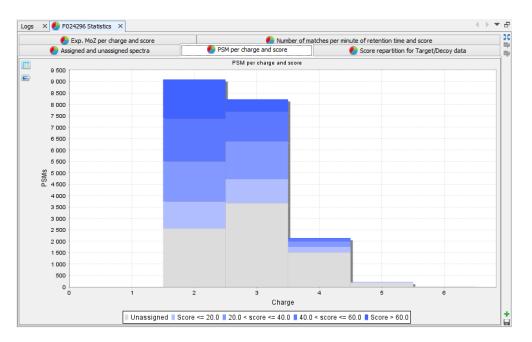


#### Score repartition for Target/Decoy data: Histogram presenting the amount of PSM per group of score, separating target and decoy data

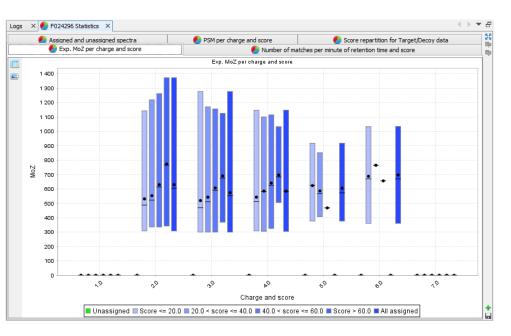




PSM per charge and score: Histogram presenting the amount of PSM per group of score and charge state

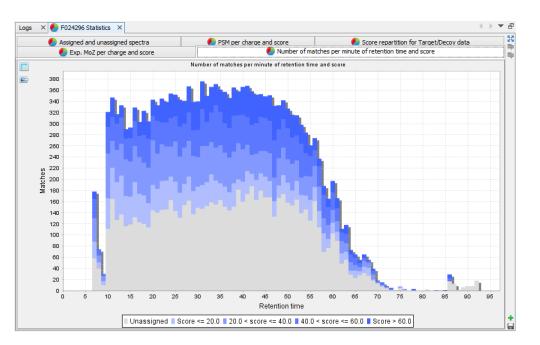


ExperimentalM/zperchargeandscore:BoxplotpresentingM/zinformationforeachcategoryofscoreandchargestatestate





Number of matches per minute of RT and score: histogram presenting the amount of PSM and per score retention time. This view is only calculated when retention time is available.



Each graph is also available in a table view

	🕖 Exp. MoZ per (	charge and score	Number of matches per minute of retention time and score					
(	🐌 Assigned and una	ssigned spectra	PSM per charge and score	Score repartition for the second s	or Target/Decoy data			
d	Resultset	Score <= 20.0	20.0 < score <= 40.0	40.0 < score <= 60.0	Score > 60.0	₽		
Target PSM 1918			3 652	3 306	2 304	~		
_	Decoy PSM	58	9 14	0	0			

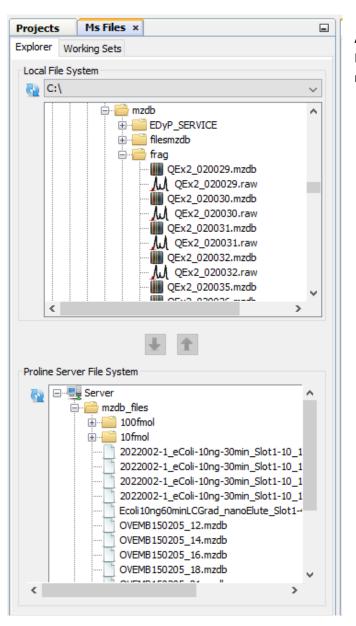


# **Ms Files Tab**

In order to facilitate different actions on Ms Files, Proline Studio contains an homonym tab providing the end user with a view over his local and server remote file system, called Local File System and Proline Server File System respectively.

Furthermore, on local file system a series of actions can take place, through an appropriate popup menu, on the encountered .mzdb and .raw files, including among others the:

- Conversion of a .raw file to an .mzdb file
- The upload of an .mzdb file
- View .mzdb files
- Detect its peakels

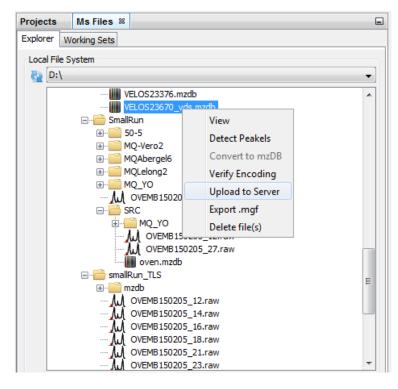


Apart from the popup menu supported functionality, since Proline Studio 1.5, uploads can be triggered via drag and drop mechanism.



#### mzDB File Upload

As mentioned earlier, after selecting a number of files, the user can either drag and drop them inside the remote site, or use the popup menu as shown in the following screenshot. It is important to precise that both approaches are not compatible with a selected group consisting of different file types.



As we can see, clicking on upload opens a dedicated dialog packing a series of uploading options:

C: \Users\AK249877\Documents\Andreas\QEx2_007990b.mzdb	Files Selection		
Delete mzdb file after a successful upload  Create Parent Directory in Destination  Server's mounting point : mzdb_PROD	C: \Users\AK249877\Docu	ments\Andreas\QEx2_007990b.mzdb	
Create Parent Directory in Destination Server's mounting point : mzdb_PROD	Upload Options		
Server's mounting point : mzdb_PROD		Delete mzdb file after a successful upload	
		Create Parent Directory in Destination	
	Server's mounting point :	mzdb_PROD	•
VK K Cancel		V OK K Cancel	

• The deletion of the file(s) after the successful upload

• The creation of the file(s) parent directory in destination: only the direct parent folder will be created. Otherwise, the file will be uploaded at the mount point root.

• The mounting point at the server: depending on server mounting points configuration.



Furthermore, the dialog permits us to add or remove .mzdb files to upload. The uploading tasks status could be viewed in the Logs tab.

#### **Raw File Conversion**

In the same way, when the user desires to convert and upload a raw file, he or she can use the respective dialog through the popup menu.

Files Selection	
O:\301-Projets_BGE\301.6	5-EDYP-Echange\Andrea\raw files\QEKAC141027_25.raw
Conversion & Upload Optic	ons
Converter (.exe) :	D:\mzdb_x64_0.9.8d\raw2mzDB.exe
Output Path :	
	Delete raw file after a successful conversion
	Upload .mzdb file successful conversion
Server's mounting point :	mzdb_PROD 👻
	▼ Delete .mzdb file after a successful upload
	Create Parent Directory in Destination
	V OK K Cancel

For the upload step, the same options as described above are shown.

For the conversion, the path to the converter exe file should be specified. This value will be saved upon different executions. A default path may be specified in the <u>general settings</u> dialog. The same way mzdb file may be deleted after a successful upload, raw files could be deleted after a conversion.

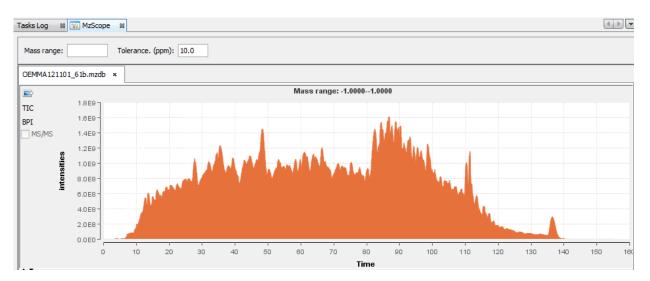
### TIC or BPI chromatogram

When the user chooses to "View" an mzDB file, the MzScope window is opened.

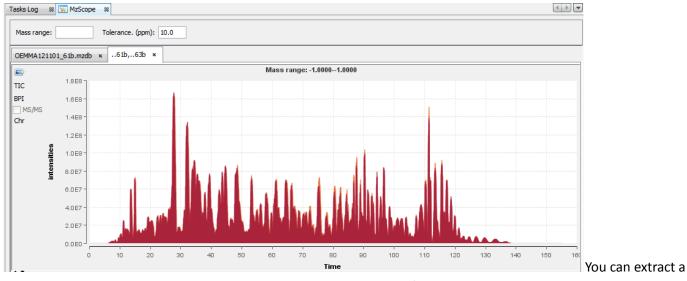
By default, the TIC chromatogram is displayed. You can click on "BPI" to see the best peak intensity graph.

By clicking in the graph, you can see below the scan at the selected time.



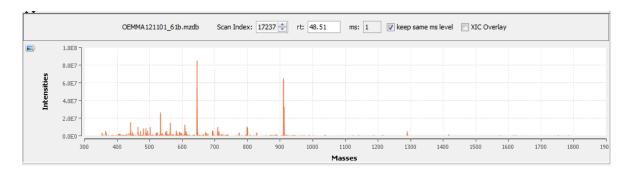


You can choose to display 2 or more chromatograms on the same graph, by selecting 2 files and clicking on "View"



chromatogram at a given mass by entering the specified value in the panel above.

Scan

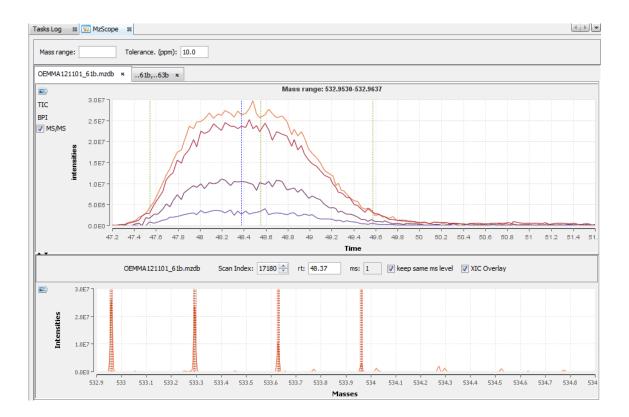


You can navigate through the scans

- by increasing or decreasing the scan Ids
- by entering a retention time
- by clicking the keys arrows on the keyboard (Ctrl+Arrows to keep the same ms level)



By double clicking on the scan, the corresponding chromatogram is displayed above (The Alt key or the check box "XIC overlay" allows you to overlay the chromatograms in the same graph).





### Peakels

By selecting a file, you can click on "Detect Peakels" in the popup menu.

Projects	MzDB File	es %		
MzDB Files L	ocation			
D: <b>\</b> ocal				
Raw files				
OEMMA1211	101_45b.m	zdb		
OEMMA1211	OEMMA121101_47b.mzdb			
OEMMA1211	101_50b.m;	zdb		
OEMMA1211	101_56b.m	zdb		
OEMMA1211	101_58b.m	zdb		
OEMMA121	101_6 <u>1b.m</u> ;	zdb		
OEMMA1211	101_6	View data		
OEMMA1211	101_6			
		Detect Peakels		

A dialog allows you to choose the parameters of the peakels detection: the tolerance and eventually a range of m/z, or a m/z value:

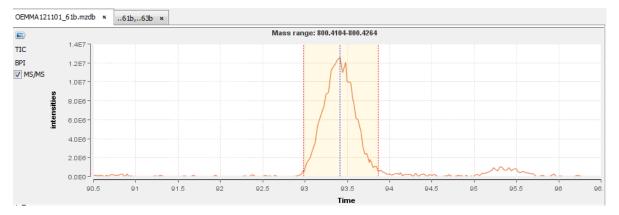
🗫 Detect Peakels Parameters 🛛 🗙 🗙					
Processing Parameters					
Intensity percentile:	0.9				
min peaks count:	5				
max consecutive gaps:	3				
min/max distance (count):	3				
min/max ratio (0-1):	0.75				
remove baseline	smooth peake	ls			
Precursor mz					
m/z tolerance (ppm): 10.0					
No m/z bounds					
◯ Enable m/z bounds					
minimum m/z: 0.0					
maximum m/z: 0.0					
O Detect at m/z: 0.0					
Ok	ancel				



#### The results are displayed in a table:

1MA 1	21101_61b.mzdb × OEMMA1	21101_63b.mzdb ×					
	m/z	Elution	Duration	Apex Int.	Area	MS Count	đ
1	800.4038	53.88	1.41	30,457,534	1,368,053,504	40	
2	800.9050	53.93	1.46	25,768,910	1,132,491,520	41	i
3	800.4013	84.90	0.85	18, 198, 898	429,421,056	28	3
4	800.4184	93.41	0.88	12,576,231	330,627,712	37	7
5	800.0841	93.44	0.81	10,422,337	276,852,832	33	3
6	800.4133	51.48	0.99	9,245,809	294,092,576	32	2
7	800.5230	35.59	1.63	8,889,166	428,618,624	56	از
8	800.7528	93.44	0.72	8,827,408	223,339,616	29	)
9	800.9026	84.94	0.69	8,689,033	191,712,608	24	
10	800.8957	55.23	1.21	4,470,434	156,532,816	32	2
11	800.9154	51.50	1.10	4,029,301	130,792,512	36	5
12	800.3785	75.01	1.35	2,855,941	121,684,904	41	1

You can double-click (or through the popup menu) on a row to display the peakel in the corresponding raw file:



# Export Data / Image

There are many ways to do an export:

- Export a Table using the export button (supported formats: {xlsx, xls, csv})
- Export data using Copy/Paste from the selected rows of a Table to an application like Excel.
- Export all data corresponding to an Identification Summary, XIC or Spectral Count
- Export an image of a view
- Export Identification Summary data into MzIdentML format (for ProteomeXchange).
- Export Identification Summary spectra list.



## 1. Export a Table

Projects 🕺 MzDB Files	Task	s Log 🛛 🕅 F07560	5 PSM 🕺		
💼 Project 1 🔹 🗈 🖉 💌		Peptide	PTM	Score	Calc. Mass
	· · · · · · · · · · · · · · · · · · ·	1 FTEGAFK		29.00	798.3912
n Identifications		2 FTEGAFK		10.26	798.3912
	a l	3 LVADLIR		26.56	798.4960
		4 LVIAVER		8.48	798.4958
🖶 🕞 ds1		5 RLGVQLD		23.15	799.4549
F075605		6 IEGILEK		8.55	800.4644
dsTest		7 LLEGELK		12.21	800.4644
dsLocal		8 WANIVAK		0.17	800.4542
		9 TVDVTIR		28.54	802.4543
F054967		10 TVDVTIR		28.01	802.4543
		11 GEPLSFR		20.10	804.4130
ds_test_xic		12 GEPLSFR		23.08	804.4130
F083342		13 TVDALMR		8.41	804.4161
F075556		14 LGIDFLK		39.14	804.4745
		15 AVLLSFR		30.85	804.4855
test new version		16 EQFNLR		30.07	805.4083
test old version		17 EQGVAFR		5.07	805.4080
		18 EQVMRK	Oxidation (M4)	5.07	805.4114
		19 MMADVPK	Oxidation (M1)	0.38	806.3664

To export a table, click on the Export Button at the left top of a table.

An Export Dialog is opened, you can select the file path for the export and the format of the export (supported formats: {xlsx, xls, csv}).

In case that the selected format is either .xls or .xlsx, the user has now the ability to maintain in his exported excel document any rich text format elements (color, font weight etc.) apparent on the original table in Proline Studio. Choice is done using the checkbox shown on the following screenshot.

Export X	Export ×
C:\vero\DATA\Autre\Report.xlsx	C:\vero\DATA\Autre\Report.xlsx
Export Type: Excel (.xlsx)	Export Type: Excel (.xlsx) ~
Export Decorated	28 %
✓ Export X Cancel	Export X Cancel

To perform the export, click on the Export Button. The task can take a few seconds if the table has a lot of rows and so a progress bar is displayed.

## 2. Copy/Paste a Table

To copy/Paste a Table:

- Select rows you want to copy
- Press Ctrl and C keys at the same time. The column titles are also copied

- Open your spreadsheet editor and press Ctrl and V keys at the same time to paste the copied rows. If paste is done in a text editor, the column separator used is the tabulations.



### 3. Export an Identification Summary, a XIC or a Spectral Count

💼 Test Vero			~	1
🏂 Identification				
F123	Display Search Result Display Identification Summary	> >		
() F130 () Trasl	Add Dataset Add Identification Folder			
	Copy Search Result Paste Search Result			
	Rename Clear	>		
	Delete Import Search Result			
	Import MaxQuant Result Validate Search Result			
uantita	Merge Datasets Filter ProteinSets	>		
🗄 📆 Tota 🖶 📆 Phos	Change Typical Protein Identify Modification Sites			
	Generate Spectrum Matches Update Spectrum using Peaklist software Retrieve Protein Sequences			
🗈 📩 Quar	Quantify	>		
	Export	>	Excel	
	Properties		Mzldent Spectra I Sequence	.ist →

To Export all data of a dataset (Identification Summary, XIC or Spectral Count), right-click on the dataset to open the contextual menu and select the "Export" menu and then "Excel..." sub-menu.

You can also export multiple dataset simultaneously, if they have the same type (Identification Summary or XIC or Spectral Count).

An Export Dialog is opened, you can select the file path and the type of the export : Excel (.xlsx) or Tabulation separated values (.tsv).

You can export with the default parameters or perform a custom export. To enable custom export, click on the tick box located on the right of the dialog:

🗫 Export			×
Export to file:	C:\vero\DATA\Autre		
Export Type:	Excel (.xlsx)		$\sim$
Custom Op	tions	🖌 Export	🗙 Cancel 👔

Custom export allows a number of parameters in addition to the file format to be chosen.

- Sheets to be created. You can **enable/disable** sheets to be exported, **rename** them (by double clicking on the title) or **change their order** by drag and drop. Available sheets are dependent on the identification summary type, but you can export which parameters were used (search settings or import and filters), the list of protein sets, all proteins, PSMs..., modification Clusters or statistics data.
- For each sheet, you can define the fields to be exported: change the title used in the exported file if needed, change the order of the fields using drag and drop of the corresponding row at the desired location.
- Symbol to use for decimal output (comma ',' or point '.')
- Timestamp format to use



- Protein sets to export ('all' or 'validated only')
- Export profile ('best' or 'all)

Description of the exported file is available here.

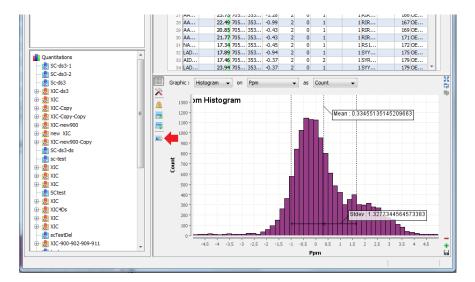
<u>Note</u>: The parameters could be saved and loaded in further export using the Save / Load buttons of the export dialog.

🏷 Export				×
Export to file: C:\vero\DATA\Autre	•			
Export Type: Excel (.xlsx)				$\sim$
<ul> <li>Custom Options</li> </ul>				
Date format: yyyy:MM:dd	HH:mm:ss 🗸 Protein set	validated only		~
Number separator: .	<ul> <li>Export pro</li> </ul>	file: Best		~
Search settings and infos	☑ Import and filters	Protein sets	Protein matches in protein sets	Bes
Orientation: rows	~		Select/Unselect all fields 🛛 🗸	
Internal field Displayed field r	na Export			
information_p project_name			$\bigtriangledown$	
information_r result_set_name				<u>^</u>
information_s search_title				
information_s search_date				
information_r raw_file_name				
information_p peaklist_file_pat	h			
information_r result_file_name				
information_r result_file_direct				
information_jojob_number				
information_u user_name				
information_u user_email				
information_g gueries_count				
information_s searched_seque	nc			
information_s software_name				
information_s software_version	n			
information_in instrument_conf	ig			
information_d database_names	S			
information_d database_releas	es			
information_d database_seque	nc			
information_t taxonomy				
information_e enzymes				
information_m max_missed_clea	ava			
information_fi fixed_ptms				
information_v variable_ptms				
information p peptide charge	st			¥
📊 Save 🕕 Load			🖌 Export	Cancel

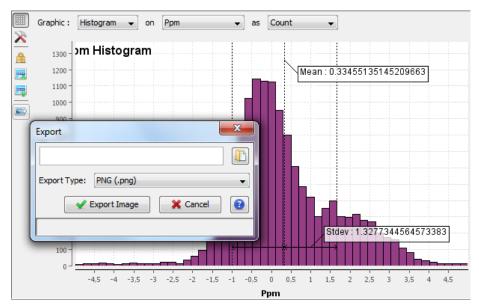
### 4. Export an Image

Any graphics in proline can be exported. Click on the Export Image Button 🔤 at the left top of the image.





An Export Dialog is opened where you can select the file path and the export type. Available formats are PNG or SVG formats. SVG format produces a vector image that can be edited and resized afterwards.



### 5. Export Identification Summary for ProteomeXchange

Actually it is possible to export Identification Summary into MzIdentML, Pride isn't supported any more.

<u>Note</u>: Before exporting data all spectrum matches should have been generated. To do so, right click on the dataset and select "Generate Spectrum Matches".

Test Vero			× 🖹 🖊 单
Identifications	ed		
	Display Search Result	>	
🗄 🍈 Co	Display Identification Summary	>	
F 1238	Add Dataset		
• F1304	Add Identification Folder		
Trash	Copy Search Result		
	Paste Search Result		
	Rename	>	
	Clear		
	Delete		
	Import Search Result		
	Import MaxQuant Result		
	Validate Search Result		
	Merge Datasets	>	
Quantitatic	Filter ProteinSets		
- 📩 TotalQ	Change Typical Protein		
Phospi	Identify Modification Sites		
📲 Quant	Generate Spectrum Matches		
- 🛃 test	Update Spectrum using Peaklist software		
👷 Quant	Retrieve Protein Sequences		
Trash	Quantify	>	
	Export	>	Excel
	Properties		MzldentML

Right click on the dataset you want to export and select the "Export" menu and then "MzIdentML..." sub-menu

A dialog is opened where user information may be specified (name, organization ...)



🗫 Export to Ma	zldentML format X
•	dentML parameters
	trum Matches should have been generated before exporting to MzidentML format. also recommended to jave run 'Retrieve proteins sequences'.
First Name*:	John
Last Name*:	DOE
Email:	john.doe@mail.com
URL:	
Organization Name	*: org
URL:	
Save	🕕 Load 🎐 Next 🔀 Cancel 😢

The file name and path should be specified in the next step. A progress bar is shown until the file is generated. The generated file contains identification and validation data issues from the dataset. All meta information including instrument configuration as well as search engine parameters are also extracted from dataset associated data.

### 6. Export Identification Summary spectra list

To export valid PSM Spectra from an Identification Summary or from a XIC Dataset. The exported tsv file is compatible with Peakview.

Note: all Spectrum Matches must be generated first.

# **Generate Spectrum Matches**

When importing a Search Result in Proline, users can view PSM with their associated Spectrum but by default no annotation is defined. Users need to generate (and save) this information explicitly.

- For a single PSM, select the icon near the Spectrum (see <u>Display Peptide and PSM</u>)
  - For a whole Search Result, Identification Summary or Quantitation Result :
    - o right click on a Dataset

•

o select "Generate Spectrum Matches"

In both cases, the following dialog will be opened. User can

- Choose to force spectrum matches to be calculated even if it is defined. This could be useful if previous spectrum matches generation was done with the wrong rules set.
- Use the fragmentation rules set defined at import. In some cases, this information is not accessible and the user should then specify it.
- Specify the fragmentation rules set to use without taking into account the rules specified at import.



Generate Spectrum Matches		×
	ion Rule Set that will be used to generate spectrum er set, new generated spectrum matches will overwr	
Force new spectrum matches	generation	
Fragmentation Rules		
Use fragmentation rule set	defined at import (unknown - see dataset propertie	s)
Select Fragmentation Rule Set	:	~ 📋
	ESI-4SECTOR	^
	ESI-FTICR	
	ESI-QUAD	
	ESI-QUAD-TOF	
	ESI-TRAP	
	ETD-TRAP	
	FTMS-ECD	
	MALDI-ISD	¥

Once executed, the dataset views need to be loaded again to effectively view the spectrum matches.

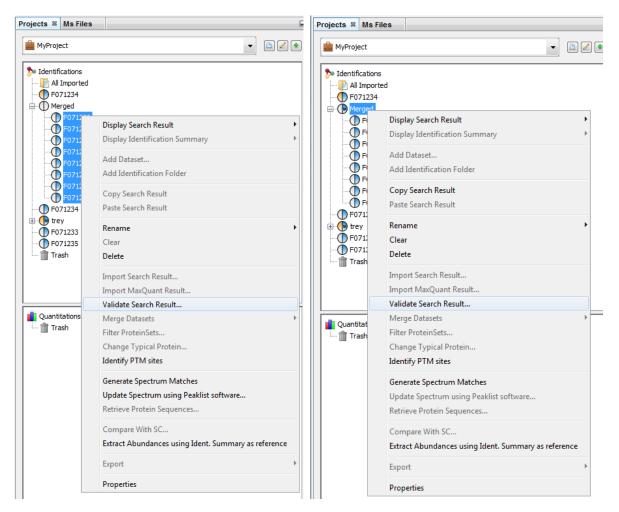
# How to validate a Search Result

See description of Validation Algorithm.

It is possible to validate identification Search Result or merged ones. In the latest case, the filters and validation threshold can be propagated to child Search Results.

### **Starting Validation**





To validate a Search Result:

- Select one or multiple Search Results to validate
- Right Click to display the popup
- Click on "Validate..." menu



### Validation Dialog

entification Validation	Х
/alidation Parameters Typical Protein Parameters	
PSM	
Propagate PSM filtering to child Search Results	
Prefilter(s)	
Length >= 6	
AND Score >= 25.0	
AND Single PSM per MS Query	
< Select >	
< select >	
FDR PSM Filter	
✓ ensure BH	
Optimisation based on Score 🗸	
Peptide	
Filter(s)	
< Select > v	
FDR Peptide Filter	
BH $\checkmark$ Peptide FDR <= 1.0 %	
Protein Set	
Propagate ProteinSets filtering to child Search Results (Warning FDR Validation will not be propagated	!
Filter(s)	
Specific Peptides >= 1	
< Select > V	
FDR Protein Filter	
BH V Protein FDR <= 5.0 %	
Scoring Type: Fisher 🗸	
📊 Save 🕕 Load 🖌 🖌 Cancel	0
	_

In the Validation Dialog, fill the different Parameters (see Validation description):

- you can add multiple PSM Prefilter Parameters (Rank, Length, Score, e-Value, Identity p-Value, Homology p-Value) by selecting them in the combobox.

- you can ensure a FDR on PSMs using BH or Taget/Decoy strategy. If using the latest, FDR will be reached according to the variable selected (Score, e-Value, Identity p-Value, Homology p-Value,...)

- you can add a Protein Set Prefilter on Specific Peptides count, peptides or peptides sequence count or on Protein Sets score.



- you can also ensure a FDR on protein Sets or peptides using BH or Target/Decoy strategy. The strategy must be consistent between all levels where an FDR control is done (PSM, Peptide and ProteinSet)

Note: FDR can be used only for Search Results with Decoy Data.

If you run validation on a merged Search Result, you can choose to propagate it to child Search Result. Specified prefilters will be used as defined. For the FDR Filter, it is the threshold found by the validation algorithm which will be used for childs, as a prefilter.

In the second tab, you can define rules for choosing the Typical Protein of a Protein Set by using a match string with wildcards (\* or ?) on Protein Accession or Protein Description. (see <u>Change Typical Protein of Protein Sets</u>).

Identification Valid	dation ×
Validation Parameters Typical Protein Parameters	
Set Typical Protein Match	
	✓ Using rules (in priority order):
Rule 0	
Typical Protein Match : *YEAST	on Protein Accession 🗸
advanced RegEx	
Rule 1	
Typical Protein Match : *	on Protein Accession 🗸
advanced RegEx	
Rule 2	
Typical Protein Match : *	on Protein Accession 🗸
advanced RegEx	
( * = any string ; ? = any character )	
Save Load	V OK X Cancel

<u>Note</u>: All validation parameters can be saved and loaded using appropriated buttons.



### Validation Processing

Projects 🕷 두	Tas	ks Log	86	
💼 ProjectTest 🕒 💽		id	Category	Task Description
	22	13	Services	Validation of Search Result F067899
So Identifications	Ξ	12	Services	Validation of Search Result F067898
All Imported	-	11	Services	Validation of Search Result F067897
□ □ □ Replicate 1	2	10	Database A	Load Search Result and Identification Su
F067897		9	Database A	Load Data for Dataset Replicate3
□ □ □ Replicate2	2	8	Database A	Load Data for Dataset Replicate2
F067898		7	Database A	Load Data for Dataset Replicate1
□ □ □ Replicate3	2	6	Database A	Load Data for Project ProjectTest
E067899		5	Database A	Load Data for Project ProjectTest
1 Trash	2	4	Database A	Load Projects for User menetrey
		3	Database A	Connection to UDS Database

Validating a Search Result can take some time. While it is not finished, the Search Results are shown greyed with an hourglass over them. The tasks are displayed as running in the "Tasks Log Dialog".

### Validation Done

Projects 🕷 💻
💼 ProjectTest 🔹 🗈 🖉 😒
Identifications All Imported Peplicate 1 F067897 Replicate 2 F067898 Replicate 3 F067899 Trash

When the validation is finished, the icon becomes orange and blue. Orange part corresponds to the Identification Summary. Blue is for the Search Result part.

# **How to filter Protein Sets**

See description of Protein Sets Filtering.

The protein sets windows are not updated after filtering Protein Set. You should close and reopen the window



## Starting filtering

Identification         Image: All Import         Image: All Import	ted	
Trash	Search Result	•
ind an	Identification Summary	•
	Properties	
	Add	•
	Merge	
	Validate	
	Filter ProteinSets	
	Change Typical Protein	
	Generate Spectrum matches	
	Compare With SC	
_	compare manoe	

To filter Protein sets of Identification Summaries:

- Select one or multiple Identification Summaries to filter
- Right Click to display the popup
- Click on "Filter ProteinSets..." menu

### Filtering Dialog

ProteinSet Filtering			×
Filter(s)			
Specific Peptides >=	1	×	
< Select >		~	
< Select >			]
Peptides count			Cancel 🕢
Peptide sequence count			
Protein Set Score			

you can add multiple filters (Specific Peptides, Peptide count, Peptide sequence count, Protein Set Score) by selecting them in the combobox.

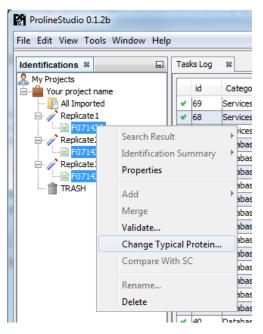
Once the filtering is done, you will have to open a new protein sets window in order to see modification.

# **Change Typical Protein of Protein Sets**

• The protein sets windows are not updated after changing Typical Protein. You should close and reopen the window ()



## Open the Dialog



To change the Typical Protein of the Protein Sets of an Identification Summary:

- Select one or multiple Identification Summaries
- Right Click to display the popup
- Click on "Change Typical Protein..." menu

### **Dialog Parameters**

Change Ty	pical Protein				×
0			noosing typical p vill be tested the		in, in priority order: le 1
	rotein Match : anced RegEx	sp*		on	Protein Description $\!$
	rotein Match : anced RegEx	*		on	Protein Accession 🗸
	rotein Match : anced RegEx	*		on	Protein Accession 🗸
(*= any s	tring ; ? = any (	character )	V Ok	<	🗶 Cancel 😵

You can set the choice for the Typical Protein of Protein Sets by using a match string with wildcards (\* or ?) on Protein Accession or Protein Description.

For Advanced users, a fully regular expression could be specified. In this case, check the corresponding option.



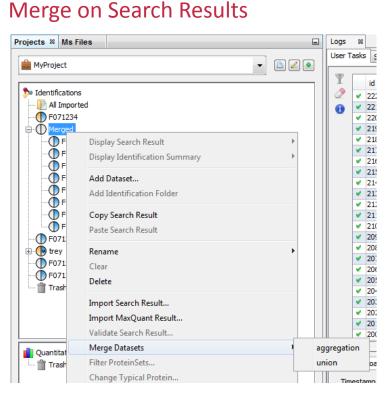
Three rules could be specified. They are applied in priority order, i.e. if no protein of a protein set satisfies the first rule, the second one is tested and so on.

### Processing

The modification of Typical Proteins can take some time. During the processing, Identification Summaries are displayed grayed with an hourglass and the tasks are displayed in the Tasks Log Window

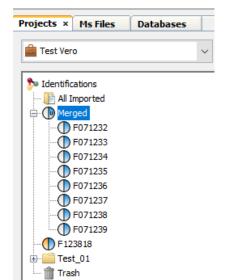
# **Combine datasets**

Merge can be done on Search Results or on Identification Summaries. You have also to specify which merge mode is to be used (aggregation or union). See description for <u>combining Search Results or Identification Summaries</u>.



To merge a dataset with multiple Search Results:

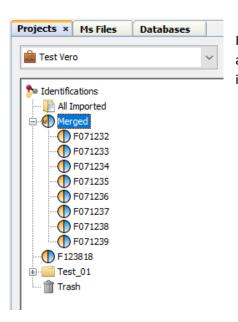
- Select the parent dataset
- Right Click to display the popup
- Click on "Merge" menu



When the merge is finished, the dataset is displayed with an U or A in the blue part of the icon, indicating that the merge has been done using Union or Aggregation at a Search Result level.



## Merge on Identification Summaries



If you merge a dataset containing Identification Summaries. The merge is done on an Identification Summary level. Therefore the dataset is displayed with an U or A in the orange part of the icon.

# **Data Analyzer**

The purpose of the Data Analyzer is to easily do calculations/comparisons on data.

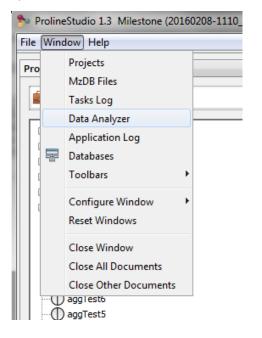
To open the data analyzer, you have two possibilities:

- you can use the dedicated button that you can find in the toolbar of all views. If you use this button, the corresponding data is directly sent to the data analyzer.



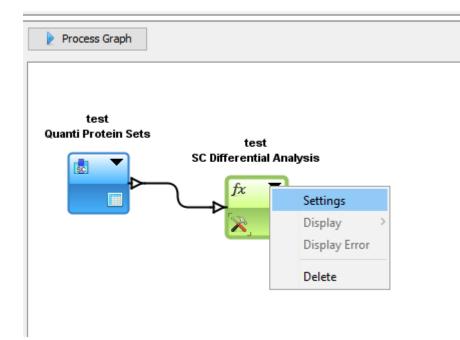
Task	s Log	🛿 🕕 F067	897 PSM 🛛 🕷	() F067899 F	PSM ≌	
$\bigcirc$		Peptide	Score	MsQuery	Rank	Calc. Mas
m	1	ILFQK	21.38	113	2	647
T	2	LIFAGK	28.56	113	1	647
	3	LLFQK	21.3 <mark>8</mark>	113	2	647
2	4	LVDLIK	25.99	182	1	699
_	5	LVDLLK	25.99	182	1	699
	6	LVDLIK	24.51	185	1	699
	7	LVDLLK	24.51	185	1	699
	8	LVDLIK	22.23	187	1	699
	9	LVDLLK	22.23	187	1	699
	10	VPFGGVK	31.28	200	1	702
	11	AAFIER	28. <mark>5</mark> 9	223	1	705
	12	AAFIER	23.5 <mark>3</mark>	225	1	705
	13	LADFLK	24. 1 <mark>9</mark>	227	1	705
	1.4	ATDLEK	25.10	220		705

- you can use the menu "Window > Data Analyzer"



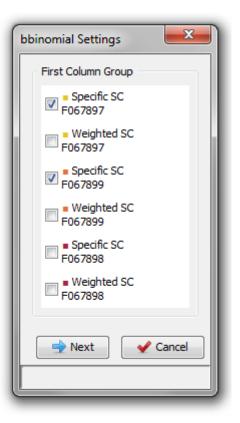
In the Data Analyzer view, you can access all data views, to some functions and graphics. In the following example, we create a graph by adding by Drag & Drop the Spectral Count Data and the corresponding differential analysis function (beta-binomial BBinomial). Then we link them together.





You have to specify the parameters of the Function: right click on the function and select the "settings" menu

In the settings menu, select the two groups of columns on which you want to perform the BBinomial function. When the parameters are set, the calculation is started immediately and an hourglass icon is shown.



When the calculation is finished: the hourglass icon becomes a green tick, and the user can right click and select the "Display" menu to see the result (or click on the "table" icon).



PSN		F067899 PSM 8	🛃 newSC Protein	Sets 🛛 📽 🚏 Data M	ixer 🛛 🖾 newSC	bbinomial 🛚 🖇	•	
T		Protein Set	Overview	Basic SC F067897	Basic SC F067899	Basic SC F067898	bbinomial	E
	1	BYR1_YEAST		126	124	119	0,722	_
	2	G EF2_YEAST		156	147	141	0,509	
	З	KPYK1_YEAST		253	279	276	0,527	
	4	G3P3_YEAST		501	573	521	0,771	
	5	PDC1_YEAST		175	198	168	0,757	
	6	BNO2_YEAST		446	465	483	0,438	
	7	BP71_YEAST		128	130	141	0,605	
	8	BP72_YEAST		124	114	132	0,75	
	9	BC82_YEAST		124	124	118	0,755	
	10	G FAS1_YEAST		61	78	83	0,219	
				-		10	0.440	

### Available Functions

### STATISTICS FUNCTIONS

#### **FDR Computation Function**

This function is used by ProStar Macro to compute the FDR.

More information: <u>http://bioconductor.org/packages/release/bioc/vignettes/Prostar/inst/doc/Prostar\_UserManual.pdf</u>

### PValue Adjustment Function / calibration Plot

Calibration Plot for Proteomics is described here: <u>https://cran.r-project.org/web/packages/cp4p/index.html</u>

#### SC Differential Analysis Function

beta binomial function, useful for Spectral Count Quantitations

#### **Xic Differential Analysis Function**

This function is used by ProStar Macro. Two tests are available: Welch t-test and Limma t-test.

More information:

http://bioconductor.org/packages/release/bioc/vignettes/Prostar/inst/doc/Prostar\_UserManual.pdf

#### **Missing values filter Function**

This function is used by ProStar Macro to remove rows with too many missing quantitative values.

The available missing values algorithm are:

- Whole Groups: The lines (across all groups) in the quantitative dataset which contain less non-missing value than a user-defined threshold are deleted.
- For every group: The lines for which each condition contains less non-missing value than a user-defined threshold are deleted.
- At least one group: The lines for which at least one condition contains less non-missing value than a user-defined threshold are deleted.

#### **Missing values imputation Function**

This function is used by ProStar Macro to impute missing values.



#### More information: <u>http://bioconductor.org/packages/release/bioc/vignettes/Prostar/inst/doc/Prostar\_UserManual.pdf</u>

#### **Normalization Function**

This function is used by ProStar Macro to normalize quantitative values.

More information on algorithms:

http://bioconductor.org/packages/release/bioc/vignettes/Prostar/inst/doc/Prostar\_UserManual.pdf

#### TABLE FUNCTIONS

#### Join Function

Join data from two tables according to the selected key.

#### **Difference Function**

Perform a difference between two joined table data according to a selected key. When a key value is not found in one of the data source tables, the line is displayed as empty. For numerical values a difference is done and for string values, the '<>' symbol is displayed when values are different.

#### **Columns Filter Function**

Columns filter, let the user remove unnecessary columns in a matrix. A combobox, with prefix and suffix of the columns allows to select multiple similar columns to filter them rapidly.

#### **Rows Filter Function**

Rows filter function lets the user filter some rows of a matrix according to settings on columns.

#### Log Function

Create a column by calculating the Log (2 or 10) of an existing column.

#### CSV/TSV Import

This module lets you import data from a CSV or TSV file. Then you can do calculations and display these data directly in Proline Studio.



CSV/TSV Import Settings						
		CSV/TSV Import	-			
CSV/TSV File :	D:\Max	Quant\MaxQuant.csv				
Separator						
🔘 Tab	Co	omma				
Semicolor	n 🔘 Sp	ace				
© Other						
Preview						
log2(Id)		Protein Set	log2(Raw abundanc 🛱			
	14.1910	B2RQQ1_MOUSE	27.6478 🔺			
		B2RXX9_MOUSE	26.9022			
		ECHA_MOUSE	28.5793			
		NDUS1_MOUSE	27.4212			
	14, 1902	AT2A2_MOUSE	27.7462			
•			• • •			
			✓ OK X Cancel			

The separator is automatically selected according to the csv file. But you can modify it.

The preview zone displays the first lines of the file as it will be loaded.



### **Expression Builder**

Expression Builder Settings		×
	F067897 Expression Builder	
Expression Builder fx = mean(Calc. Mass) fx Functions b abs b mean b log2 b log10 b std	Variables Id Peptide PTM Score Calc. Mass Calc. Mass Exp. MoZ Ppm Charge Missed Cl. Rank RT Protein Set Count Protein Sets	7 8 9 / 4 5 6 * 1 2 3 - 0 . ± + ( ) , and or not = != < > >= <=

The expression builder lets you create an expression with built-in functions or comparators and variables (columns from the linked matrix). In the example, we calculate the mean of a column in the matrix.



## Prostar Macro

Refresh Data		Process Graph	Do	calculation	S					
ata Analyzer Data Windows	<u>^</u>					Рто	itar			
Marce Marce Marce Marce Marce Marce Marce Marce Marce Arcolumet Arcolumet Arcolumet Arcolumet Arcone Ar	Bureact     Febre     S'     n: Dubler     s      s	Lade - 00:186.1, Quard Protein Sets		kdd3. 30:198.1 Rown Filter ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓			tude: 20 (19.4) Bornalization R C C C C C C C C C C C C C C C C C C C	toted - XIC 186.1. Bissing Values Impuration frequencies of the second s	turds-XIC (40.4. ICC Differential Analysis	tudo XIC 100.1. Calibridion Pot ALL Calibridion Pot ALL Calibridion Pot ALL Calibridion Calibridion Calibridion Calibridion Calibridion Calibridion Calibridion Calibridion Calibridion Calibridion Calibridion Calibridion Calibridion Calibridio Calibridion Calibridio Calibridion Calibridio Calibridion Calibridio Calibridion Calibridio Calibridion Calibridio Calibridion Calibridio Calibridion Calibridio Calibridion Calibridio Calibridion Calibridio C
ProStar					· · · · · · · · · · · · · · · · · · ·	Plor FOR		Value  0.00344154470	74609143%	<b>(</b> )
	1 0.2 xm(bg2/Abur, Lum	0.4 0.6 a PValue dog10	0.8 10 Lioma Plake) Lioma k	15 og Ratio Differential		1 FDR	X Axis: Limma log Ratio	0.00344154470	74509143% ( Axis: [dog10(Jamma PW	
R ProStar 0	orm(log2(Abun Limm -6.4338 -5.7713 -4.2448 -5.7684	a PValue -log 10 0.0020 0.0015 0.0905 4.49E-4	Limma PValue) Limma lo 2,6894 2,8319 1,0428 3,3482	og Ratio -0. 1664 -0. 1712 -0.0628 -0.2855		Graphic: Scatter Flot +	X Avis: Lumma log Ratio Selection	0.00344164470	(Axis: Hog10(Jamma PW	
R ProStar 00 00 00 00 00 00 00 00 00 0	em(log2(Abun Limm -6.4338 -5.7713 -4.2448 -5.7684 -5.7684 -6.4642 -4.3606 -9.8677	a PValue -log 10 0.0020 0.0015 0.0905	Linma PValue) Linma lo 2.6894 2.8319 1.0428 3.3482 3.0388 2.3231 2.9779	og Ratio Differential -0. 1664 -0. 1712 -0.0628	Proteins 0	Graphic: Scatter Flat •		0.00344164470 • ]	(Axis: Hog10(Jamma PW	
R ProStar 000 000 000 000 000 000 000 0	em(log2(Abun Limm -6.4338 -5.7713 -4.2448 -5.7684 -6.4642 -4.3606	a PValue -log ID 0.0020 0.0015 0.0906 4.995-4 0.0048 0.0048 0.0011	Linma PValue) Linma k 2.6894 2.8319 1.0428 3.3482 3.0388 2.3231	og Ratio 0. 1664 0. 1712 0. 0628 0. 2355 0. 2759 0. 1449 0. 2526	Proteins 0	Graphic: Scatter Flat		0.00344164470	(Axis: Hog10(Jamma PW	

1 or 2: Add XIC Data to Data Analyzer from the Protein Set View or by importing data from a csv file.

3: Add Prostar Macro by a drag and drop and link XIC Data to the Macro. And do the calculation by clicking on the button Process Graph.

During the process, the Data Analyzer will ask you settings for each function.

4: Filter unnecessary columns from your data if. Settings can be validated with no parameters if you don't need it.

5: Filter is needed only if you want to remove contaminants. Settings can be validated with no parameters if you don't need it.

6: Log is needed to log abundances (Data from Proline). For Data coming from MaxQuant, data is already logged.

7 to 13: follow the settings asked (you can find some help in Prostar documentation, or information in corresponding functions.)

During the process, results will be automatically displayed:

14: FDR Result

15: Calibration Plots

16: Result Table with differential Proteins Table and the corresponding scatter plot. You can select differential proteins in the table, to import them in the scatter plot and create a colored group with them.

If you want to look at other results, right click on a function and select "Display in New Window"

Prostar User Manual:

http://bioconductor.org/packages/release/bioc/vignettes/Prostar/inst/doc/Prostar\_UserManual.pdf

Prostar Tutorial :

http://bioconductor.org/packages/release/bioc/vignettes/Prostar/inst/doc/Prostar\_Tutorial.pdf



# **Calculator**

Calculator lets you write python scripts to manipulate freely viewed data.

1) To open the calculator, click on the calculator icon (not available on all views for the moment)

On the left part of the calculator, you can access all viewed data, double click to add a table or a column to the script.

- 2) Write your python script on the text area
- 3) Execute it by clicking on the green Arrow.

4) When the script has been executed, the results of the calculations (variables, new columns) are available in the "Results" tab. Double click on a new column to add it to the table. Or like in the example, directly add the column to the table programmatically.

Protein Set	Overview	#Peptide	#Quant. Peptide	Status F067897	Peptide Number F067897	Specific SC F067897	log(specificSC)
PYR1_YEAST		93	80	Typical	67	125.00	-1
EF2_YEAST		60		Typical	51	156.00	
KPYK1_YE		50		Typical	46	248.00	
G3P3_YEAST		42	37	Typical	34	315.00	5
💮 🛄 3: r	Results PSM Peptides Quanti Protein Sets ewSC Quanti Peptid ewSC Quanti Peptid obinomial	s # ge t = 1 # ge # mu spec # nu nb = # lou for i # v if el # se spec # ac t.ad	t the Table 3 whic Table.get(3)	+ corresponds umn 10 of the be able to mo nutable() he column ) the column NaN values for = float(NaN') = math.log(v) = which will be i nnName('log(s) umn to the tabl CCol)	r errors) used to the user pecificSC)")		3

#### **Examples**

#### Script to calculate a log column

#### Algorithm to calculate the logarithm of a column ####

# get the Table 3 which corresponds to table newSC Quanti Protein Set t = Table.get(3)

# get the constant column 10 of the table t ( Specific SC column)
# mutable() is called to be able to modify data



specificSCCol = t[10].mutable()

```
# number of rows of the column
nb = len(specificSCCol)
# loop on the data of the column
for i in range (0,nb):
    # calculate the log (NaN values for errors)
    v = specificSCCol[i]
    if v <= 0:
        specificSCCol[i] = float('NaN')
else:
        specificSCCol[i] = math.log(v)
```

# set the column name which will be used to the user specificSCCol.setColumnName("log(specificSC)")

# add the created column to the table t
t.addColumn(specificSCCol)

#### Script to perform a difference and a mean between two columns

#### Algorithm to perform a difference and a mean between two columns ####

```
t = Table.get(9)
colAbundance1 = t[3]
colAbundance2 = t[5]
# difference between two columns
colDiff = colAbundance1-colAbundance2
# set the name of the column
colDiff.setColumnName("diff")
# mean between two columns
colMean = (colAbundance1+colAbundance2)/2
# set the name of the column
colMean.setColumnName("mean")
# add columns to the table
t addColumn(colDiff)
```

t.addColumn(colDiff) t.addColumn(colMean)

#### Script to perform a perform a pvalue and a ttd on a XIC quantitation table

```
#### Algorithm to perform a pvalue and a ttd on abundances column of a XIC quantitation
####
t = Table.get(1)
pvalueCol = Stats.pvalue( (t[2], t[3]), (t[4],t[5]) )
ttdCol = Stats.ttd( (t[2], t[3]), (t[4],t[5]) )
pvalueCol.setColumnName("pvalue")
ttdCol.setColumnName("ttd")
t.addColumn(pvalueCol)
t.addColumn(ttdCol)
```

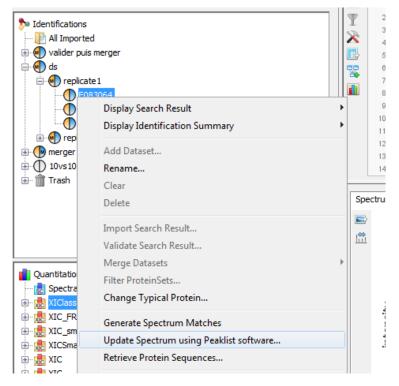


# **Update Spectrum using Peaklist software**

When <u>importing a search result</u>, the software used for the peaklist creation has to be specified. This parameter is mandatory for the <u>XIC quantitation</u> as it is used to find scan number or RT in the spectrum title. Indeed, this information is then used to extract abundances in the *raw files*.

If an invalid software has been specified when importing, it is possible to change the peaklist software afterwards. This option is only valid for Identification DataSets.

Right click on the identification DataSet, and select "Update Spectrum using Peaklist software"



The following dialog will be displayed allowing the user to select the peaklist software to use.

🐤 Update Spectrum Parar	ieters	x
Peaklist Softwares		
Choose Peaklist Software to	use to generate (new) spectrum para	ameters.
		-
		â.
extract_msn.exe		
Mascot Distiller		ΞĤ
mascot.dll		
MaxQuant		
Proline 1.0		
Protein Pilot 4.5		
Proteome Discoverer		-

# **Spectral Count**

See description of Compare Identification Summaries with Spectral Count.



## Generate a Spectral Count

To obtain a spectral count, right click on a Dataset with merged Identification Summaries and select the "Quantify  $\rightarrow$  Spectral Counting" menu in the popup. This Dataset is used as the reference Dataset and Protein Set list as well as specifics peptides are defined there.

				Y		id	Category	Criticalit
🏂 Identification	ns			0	-	18	Services JMS	High
All Impor	rted			0	1	17	Services JMS	High
				<b>•</b>	~	16	Services JMS	High
i ⊨	Display Search Result		>			15	Database A	Medium
-	Display Identification Summary		>			14	Database A	Medium
						13	Database A	Low
	Add Dataset					12	Database A	Low
	Add Identification Folder				2	11	Database A	
					1	10	Database A	
	Copy Search Result				2	9	Database A	
	Paste Search Result				-	8	Database A	
					V	7 6	Database A Database A	
	Rename		>		÷	5	Services JMS	Low
	Clear				-	4	Services JMS	Low
Tes						3	Services JMS	Medium
Tra	Delete					2	Database A	High
III 114	Import Search Result					1	Services JMS	High
						0	Services JMS	High
	Import MaxQuant Result							
	Validate Search Result							
	Merge Datasets		>					
	Filter ProteinSets							
	Change Typical Protein							
	Identify Modification Sites							
	Generate Spectrum Matches							
	Update Spectrum using Peaklist software.			<u> </u>	_			
Quanti	Retrieve Protein Sequences			Task:				
🗄 🖳 📩 Tot	Quantify		>		S.	ettel C	ounting	
🗄 📩 Pho	Quantity		-					
🖶 间 Qu	Export		;		Lat	oel Free.		
			-		Res	idue La	beling	
es tes	Properties				lso	baric La	beling	
🖶 🏓 Quant		П	пL	LIN		<b>L</b> .	_	

In the Spectral Count window, fill the name and description of your Spectral Count and press Next.

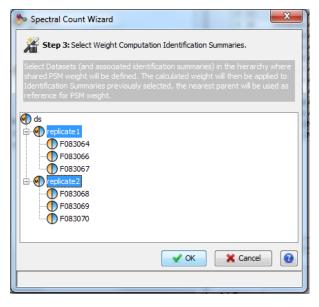
🎭 Spectral Co	unt Wizard	x
X Step 1:	Define spectral count name and description.	
Spectral Cour	nt	
Name:	Spectral Count agg2	
Description:	Your Description	
	🚽 Next 🛛 🗶 Cancel	0



Then select the Identification Summaries on which you want to perform the Spectral Count and press Next.

spectral Count Wizard
Step 2: Select Identification Summaries.
Select the Identification Summaries for which Spectral Count will be calculated. The common list of protein sets and peptide specificity will be calculated at parent level, from which spectral count has been run.
ⓓ ds □-ⓓ replicate 1
F083070
Next X Cancel

Finally select the DataSet where shared peptides spectral count weight should be calculated and press OK.



A Spectral Count is created and added to the Quantitations Panel.



Projects 🕷 MzDB Files	
Tutoriel	-
Identifications All Imported Identifications All Imported Identifications Id	
Quantitations Spectral Count agg2	

### **Display a Spectral Count**

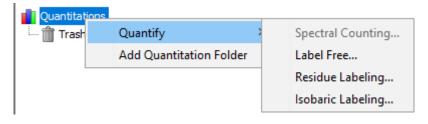
You can then display a Spectral Count, see Display a Spectral Count

# **XIC Quantitation**

For description on LC-MS Quantitation you can first read the principles in this page: Quantitation: principles

### Create a New XIC Design

You can create a new XIC design by clicking on the "Quantify  $\rightarrow$  Label Free" menu from Quantitation or from a dataset in the Identification tree. In this case, the selected dataset will be used as reference for the quantitation (from where the list of peptides and protein is extracted).



#### Settle Design

When you start the creation of a new XIC Design, the following dialog is opened:



ا (میلا)	Drag & Drop	
Create Sample Group L101_43b.raw Delete L101_47b.raw OEMMA121101_56b.raw OEMMA121101_50b.raw OEMMA121101_50b.raw	Identifications         F083342         Image: Construction of the second secon	

In the left Area, you can create your design :

- by adding Group and Samples (thanks to mouse right click popup)
- by renaming every items ( thanks to F2 or mouse right click popup)
- by rearranging order of groups and samples (thanks to drag & drop)

From the right Area, you can drag and drop desired identifications to the left Area. If you drop in the XIC or in a Group node, Group and Sample nodes will be automatically added. But you can drop directly in a Sample node.

Note : it is recommended that the XIC node is renamed.

When your design is ready, click on Next Button.

#### Link to Raw Files

The following panel is displayed:



files associatior	1					
Group Group Fr1 Group Fr1	Sample Sample Fr1 Sample Fr1	Sample Analysis F127609 F127603	mzDB File 20070918_CL_Orbi2 20070918 CL_Orbi2		QEx2_007990D.m2db	^
Group Fr1 Group 2	Sample Fr 1 Sample Fr 1	F12760922 F127614	20070918_CL_Orbi2 20070919_CL_Orbi2	. 20	Sabrina SaliRuns	
Group 2 Group 2	Sample Fr1 Sample Fr1	F127617 F127620	20070919_CL_Orbi2 20070919_CL_Orbi2		VELOS23138.mzdb     VELOS23140.mzdb	
<ul> <li>utomatic MS file</li> </ul>	association			>	VELOS23142.mzdb           VELOS23144.mzdb           VELOS23144.mzdb           VELOS23374.mzdb	
files dropped.	0 files were associated.		Drop Zone our .mzdb files & folder here	rs	VELOS23376.mzdb VELOS23378.mzdb VELOS23380.mzdb VELOS23864.mzdb VELOS23666.mzdb VELOS23666.mzdb VELOS23666.mzdb VELOS23668.mzdb VELOS236670.mzdb VELOS23670.vds.mzdb	

In order to be able to perform any XiC design, all participating Sample Analysis must be associated with a corresponding raw file. The association is done at the second step of the XIC design by :

- drag & drop the "mzDB File" into the raw corresponding to the correct sample analysis. It must be noted
  that although it is possible to overwrite an existing association (done in a previous quantitation using the
  same sample analysis dataset), there is no verification done to ensure compatibility between the mzDb file
  and the sample analysis dataset.
- using the Drop Zone. This feature can be quite helpful in cases where multiple associations are missing or when the plethora of uploaded .mzdb files intimidates the user from manually searching the files one by one. The feature itself is extremely easy to use as it just requires dragging a set of files or folders containing .mzdb files from the user part. As soon as a drop takes place, all missing connections will be automatically created as long as a matching .mzdb file has been dropped in the Drop zone. It must also be noted that since version 1.5, users have now at their disposal indices about the association source. Furthermore, in order to protect from a possible corruption of an existing association in the database, the latter ones cannot be overwritten.

When all participating Sample Analysis are associated with a raw file, click on Next Button.

#### **XIC Parameters**

Now you must select the parameters. See <u>Label-free LC-MS quantitation configuration</u> to have more details about the different parameters.

The XIC parameters are not all displayed. You can display a complete set of parameters by clicking on the "Advances Parameters".



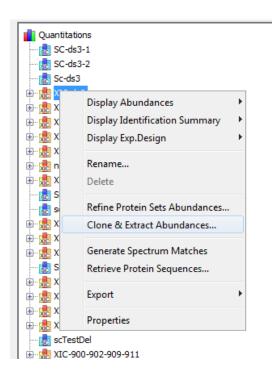
Se XIC Quantitation Wizard
Step 3: Specify quantitation parameters.
Advanced Parameters
XIC Parameters
Moz tolerance (ppm): 5.0
Map Alignment
method: Iterative 🗸
max iteration: 3
Smoothing
method:
alignment time interval (s): 200
minimum number of landmarks : 50
Feature Mapping
time tolerance (s): 600.0
Master Map
Feature Mapping
time tolerance (s): 120.0
Save Load 🖉 Back V CK K Cancel 😨

Note: all the parameters are already set with default values.

## Clone a XIC Design

You can copy a XIC design by selecting the "Clone & Extract Abundances" option from an existing quantitation result. The new XIC is generated using an existing Experimental Design and parameters setting.





## **Display XIC Results**

Newly generated XIC designs are immediately added to the Quantitation Tree. Through the latter one, and via a popup menu, the end user has the capacity either to view a design's properties as seen to the following screenshot, or to apply a series of actions on it, including among others:

- Delete a XIC Design, see how to Delete Data
- Rename a XIC Design, by clicking on "Rename..." in the popup menu.
- Export the XIC results, see how to Export Data



	Туре	Group General Information	Gr
361	XIC id	General Information	
		Identification Summary	
14	Identification Summary id		
	Description		
20 mai 201	Date		
		Quantitation Processing Config	
EXHAUSTIV	aln_method_name		
5.	aln_params / ft_mapping_params / moz_tol		
PP	aln_params / ft_mapping_params / moz_tol		
600.	aln_params / ft_mapping_params / time_tol		
2000	aln_params / mass_interval		
	aln_params / max_iterations		
TIME_WINDO	aln_params / smoothing_method_name		L-
5	aln_params / smoothing_params / min_windo		L
2	aln_params / smoothing_params / window_o		L
20	aln_params / smoothing_params / window_size		L
MOST_INTENS	clustering_params / intensity_computation		$\vdash$
10.	clustering_params / moz_tol		<u> </u>
PP	dustering_params / moz_tol_unit		<u> </u>
MOST_INTENS	clustering_params / time_computation		L
15.	clustering_params / time_tol		<u> </u>
fals	detect_features		L
tru	detect_peakels		L
5.	extraction_params / moz_tol		L
PP	extraction_params / moz_tol_unit		
INTENSIT	ft_filter / name		
G	ft_filter / operator		
0.	ft_filter / value		
10.	ft_mapping_params / moz_tol		
PP	ft_mapping_params / moz_tol_unit		
120.	ft_mapping_params / time_tol		
MEDIAN_INTENSIT	normalization_method		
tru	start_from_validated_peptides		
fals	use_last_peakel_detection		
		Quantitation Method	
ree based on the extraction of feature.			
label_fre	Туре		
feature_intensit	Abundance Unit		
		Master Quantitation Channel 1	
73	Master Quantitation id		L
XI	Name		L
	Serialized Properties		
		Quantitation Channel 1	
288	Quantitation Channel id		L
OEMMA121101_43b.ra	Result File Name		_
D:\Data\Claire\OEMMA121101_43b.ra	Raw File Path		
OEMMA121101_43b.mzd	Mzdb Raw File Name		
11	Identification Summary Id		
		Biological Sample 1	
202	Biological Sample id		
Group ds3-1Sample ds3-	Biological Sample Name		
		Quantitation Channel 2	
288	Quantitation Channel id		
OEMMA121101_45b.ra	Result File Name		
/_D01\Shares\301-Projets_BGE\301.1	Raw File Path		
OEMMA121101_45b.mzd	Mzdb Raw File Name		
12	Identification Summary Id		
		Biological Sample 2	
202	Biological Sample id		
Group ds3-2Sample ds3-	Biological Sample Name		



# **Xic Aggregation**

It is possible to aggregate data from different XIC designs.

First you must select two or more XIC Designs in the Quantitation tree and do a mouse right click to obtain a popup and then select the menu "Aggregate Quantitations".

Quantitations	
⊡… <mark>…</mark> Quant-2_ ⊡…	Aggregate Quantitations
	Compute Post Processing on Abundances
	Export Excel
	Retrieve Protein Sequences
	Properties

#### The following dialog is opened:

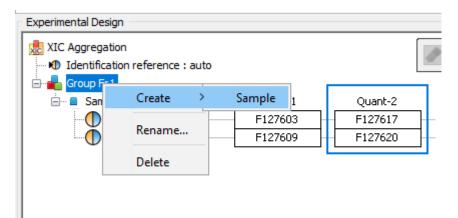
Se Aggregate Quantitation Wizard	×
<ul> <li>Left Panel: Define the aggregation experimental design</li> <li>The following experimental design was inferred from the quantitation that will be aggregated. Group, samples and channels (replicate entities can be modified         <ul> <li>Change entities order by drag and drop</li> <li>Rename entities by contextual menu (right click)</li> <li>Create or delete entities from the contextual menu</li> </ul> </li> <li>Center Panel: Define quantitation channels mapping</li> <li>Each quantitation channel of the aggregation will correspond to sample analyses of aggregated quantitations. The following modification be made:                 <ul> <li>Change association by dragging and dropping sample analysis from the right panel to a cell or from another cell</li> <li>Remove association by using contextual menu or toolbar</li> <li>Move analyses up or down by using contextual menu or toolbar</li> </ul></li></ul>	
Experimental Design XIC Aggregation Group Fr1 Channel 2 Experimental Design Quant-1 Quant-1 Quant-2 F127603 F127617 F127620 Quant-1 Quant-2 F127617 F127620 Quant-1 Quant-2 F127617 F127620 Channel 2 Channel 2 C	
V OK X Cancel	0



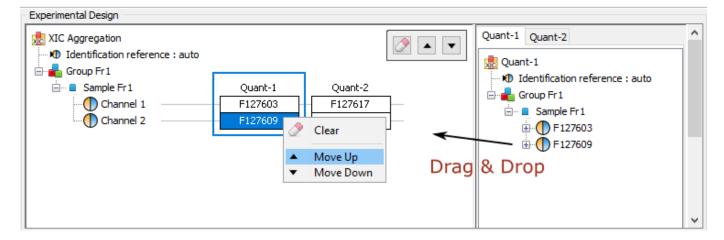
This dialog is divided in three areas:

- **left area**: aggregated experimental design. This design is inferred from the quantitations that will be aggregated.
- **center area** : quantitation channels mapping. Each quantitation channel of the aggregation will correspond to sample analysis of aggregated quantitations.
- right area : XIC Sources

You can completely modify the inferred aggregated experimental design. You can use drag & drop to modify the order of the items and the right mouse popup to create, delete or rename items.

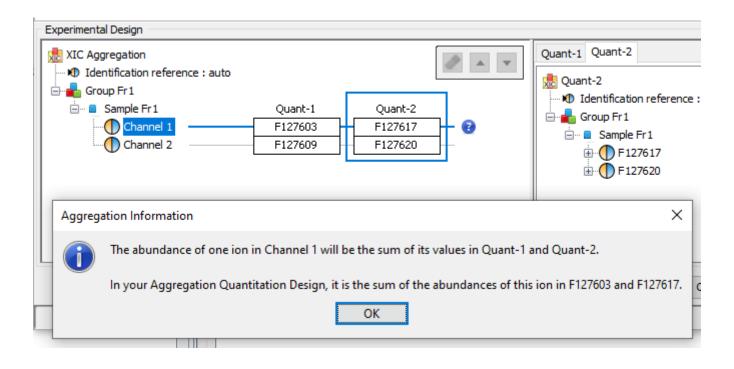


You can modify the quantitation channels mapping by drag and drop from the right area to the center area. You can also remove or reorder the mapping thanks to the right mouse popup or thanks to the floating panel.



When you select a Channel, you can click on the help button to check what the aggregation will exactly do.





# **Compute Post Processing on abundances**

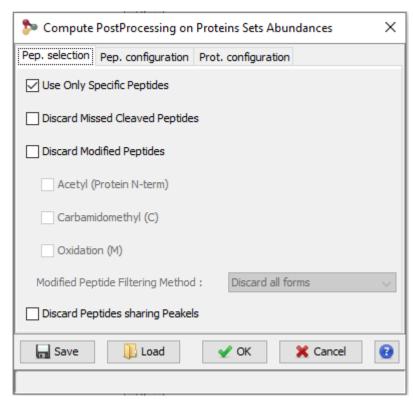
### Advanced Protein Sets abundances

Right click on the selected XIC node in the Quantitation tree, and select "Compute Post Processing on Abundances..."



Quanti	tation	s	
≝… <u>xic</u> sr 1		Display Abundances	>
⊕ <u>ಹ</u> Qu		Display Identification Summary	>
👥 sr 1 👥 sr 1	Display Exp. Design	>	
🗈 <u>に</u> sr 1 🔐 💼 Tra		Add Quantitation Folder	
		Rename	
		Delete	
		Compute Post Processing on Abundances	
		Clone & Extract Abundances	
		Change Typical Protein	
		Identify Modification Sites	
		Generate Spectrum Matches	
		Retrieve Protein Sequences	
		Export	>
		Properties	

## Configuration



In the dialog, you can:

- specify peptides to consider for quantitation



- configure parameters used for peptides quantitation
- configure parameters used for proteins quantitation
For more details, see <u>Post-processing of LC-MS quantitative results</u>

## Advanced XIC results

You can see the results by displaying the XIC (Display a XIC) or export them Export Data



# **General Settings**

Since version 1.4 Proline Studio includes a general settings dialog which can be accessed from the top menu bar clicking on "General Settings" as shown in the screenshot below.

File Window Help										
Change User		Logs	88					4		-
Change password		User T	asks	Se	rver Tasks					
General Settings										5.7
Admin		T		id	Category	Criticality	Task Description	Progress	Ę	
Upload .dat File(s)	s_BGE\301.6-EDYP-Echange  ojets BGE\301.6-EDYP-Echan	2	٠		Database Access	Low	Load Data for Datase			-
Exit	ojets BGE\301.6-EDYP-Echan	0	۷.	-	Services JMS	Low	Browse Server File Sy			
	rojets BGE\301.6-EDYP-Echan		٠.	-	Database Access	Low	Load Data for Datase		_	
	rojets_BGE\301.6-EDYP-Echan			-	Database Access	Medium	Load Quantitation Da		Ξ	
	rojets_BGE\301.6-EDYP-Echan		*	-	Database Access	Medium	Load Identification Da		_	
	rojets BGE\301.6-EDYP-Echan			-	Database Access	Medium	Load Projects for Use		_	
	01-Projets BGE\301.6-EDYP-Ec		× ×		Services JMS	Low	Browse Server File Sy		_	
	01-Projets BGE\301.6-EDYP-Ec		_	-	Services JMS	Medium	Get Server File Syste		-	
	01-Projets BGE\301.6-EDYP-Ec *		-	2	Database Access	High	Connection to UDS D		-	
<ul> <li>III</li> </ul>	+		•			III		- P		
	05207_3-2.mzdb		esta Time	÷.,					] []	) <b>2</b> <b>4</b>
1QEx2_0 1QEx2_0	05210_4-1.mzdb 05211_4-2.mzdb 05212_4-3.mzdb		rt Tir				Start Delay:			
		End	d Tim	e:			Duration:			
CAVEN 10	1635.mzdb 1637.mzdb 1639.mzdb	Err	or Me	essag	je					_
· · · · · · · · · · · · · · · · · · ·	4									+

The latter one consists of a constantly but slowly growing number of user preferences regarding various aspects of the utilization of Proline Studio. Based on their context, for the time being, preferences are organized into the following four tabs:

- JMS Settings
- Conversion/Upload Settings
- Table Parameters
- General Application Settings

### **JMS Settings**

JMS Settings tab contains parameters that concern the exchange of messages between your local machine and the JMS Server. It should be made clear that compared to other preferences, preferences that are included in this tab should be treated with caution. Mistreating a communication preference can lead either to communication/connection problems or to users' confusion to whether they are connected to the correct server version.



Proline Studio Settings	Parameters Service Request Queue Name :	ProlineServiceRequestQueue	
			V OK X Cancel

#### Service Request Queue Name

Parameter can be seen as a name which represents a server address. The parameter's existence is justified by the fact that multiple server versions might run on the same server machine imposing the need to be able to canalize Studio's messages appropriately.

### **Table Parameters**

Table parameters' tab encapsulates a short list of preferences regarding all tables generated throughout Proline Studio. More specifically those preferences control the arrangement of the participating columns as well as their respective width.

Proline Studio Settings	Parameters Columns Arrangement : Column Width :		]
		▼ OK Cancel	•

#### **Columns Arrangement**

This field dictates the spatial arrangement of table columns. Three arrangements are possible:

- Fixed Column Size
- Automatic Columns Size
- Smart Columns Size

When "Automatic Columns Size" is used, all columns are width-wise readjusted in a way that they all fit to their container. Given that it is a "fit-to-screen" approach, it lacks scrollbars and does not guarantee the readability of the presented date, especially when the number of columns is high.



ects 🕷 MzDB Files		Logs	88 (	🔵 ds 1 Prote	ins %		4	Ŀ
Project 1 👻 🗈 🖉	• •			Protein	Score	Peptides	Mass	r,
		m	1	FCGB				-
Identifications		T	2	🔓 CO3				Ξ
- All Imported		$\otimes$	3	GFAH	3193.83	65		
		-	4	💊 A8K5T				
			5	MUC5	2480.67			
dsTest		22	6	69PBJ	2480.67			
dsLocal	=		7	🎍 #C#P		58		
Here ds3			8 9	ALBU		58 67		
			10	A/19J				
			11	EFG E				
			12	EFG E				
			13	U5M6	2369.66	43		
			14	A H0Q7	2369.66	43	77581.30	
test new version			15	6ND	2369.66	43	77581.30	
test old version			16	EFG_E	2369.66	43	77581.30	
dsMat			17	🖕 H6VR	2282.58	38	66052.76	
- Пртм	-		18	👌 K2C1	2282.58	38	66038.73	Ŧ
	•	<b>#1</b>						
								₽
Quantitations	ÂIJ	T	1 R	L		2 0 3		
SC-ds3-1		$\mathbf{X}$						
🛃 SC-ds3-2			3 <b>R</b>					Ξ
🛃 Sc-ds3	=				••• ••• ••• ••			
🗠 📩 XIC-ds3		-	5 R		••• ••• ••• •			
- 😹 XIC			6 K					
- 🛃 XIC-Copy			7 R 8 R		•••• ••• ••• •			
XIC-Copy-Copy			8 R 9 K					
XIC-new900			10 R					
new XIC			11 R			. 3 1 1		
XIC-new900-Copy			12 R					
SC-ds3-ds			13 K					
			14 K					
			15 R	S		. 2 0 1		
XIC			16 K	L		3 0 2		
XIC			17 R	L		2 0 1		
- 📩 XIC	-		18 R	C		2 0 1		Ŧ

On the other hand a simpler approach that guarantees readability is to select "Fixed Column Size". In this case all table columns have a fixed width, explicitly dictated by the user using the parameter "Column Width".



jects  MzDB Files		Logs	🕷 🕕 ds1 Proteins 🛛 🕷		4	⊧ l
🖹 Project 1 🔹 🔳 🖉	🔹 🥑	$\square$	Protein	Score	Pep	<b>₽</b>
		i i i i i i i i i i i i i i i i i i i	1 🛃 FCGBP_HUMAN	550	5.39	-
🏂 Identifications		T	2 CO3_HUMAN	337.		Ξ
All Imported		$\mathbf{x}$	3 GFAH_HUMAN	3193		
			4 🛃 A8K5T0_HUMAN			
in transformed to the second			5 MUC5B_HUMAN	2480		
i dsTest		2	6 6 E9PBJ0_HUMAN	2480		
dsLocal	=		7 <b>#C#P02768-1</b>	2449		
⊕			8 ALBU_HUMAN	2449		
Gamme Levure UPS1			<ul> <li>8 A7Y939_HUMAN</li> <li>10 A8K9P0_HUMAN</li> </ul>			
⊞ ① ds_test_xic			10 A8K9P0_HUMAN	2369		
			12 BEFG ECOLI	236		
			13 U5M6M8_ECOLI	2369		
			14 HOQ782_ECOLI	2369		
test new version			15 6 U6NDZ2_ECOLI	2369	9.66	
test old version			16 SEFG_ECODH	2369	9.66	
⊕ () dsMat			17 A HOVRES HUMAN	2282		Ŧ
	-		۰ III		•	
< m	•	<b>M</b>	[		_	
Quantitations		T	Prev. AA	Peptide	Next /	P
			1 R	LPVSLSEGR	L	-
SC-ds3-1		$\mathbf{X}$	2 R	TPDGSLLVR	Q	Ξ
SC-ds3-2			3 R	GNPAVSYVR	V	
	=		4 R 5 R	GNPAVSYVR GNPAVSYVR	V	
🖅 🚾 XIC-ds3		-	6 K	FYPAGDVLR	v	
🗄 🔜 📩 XIC			7 R	AQDESPCYG	-	
🗄 👷 XIC-Copy			8 R	SRLPVSLSEGR	L	
🗄 👷 XIC-Copy-Copy			9 K	AIGYATAADCGR	Т	
			10 R	VVAEVQICHGK	T	
En 👷 XIC-copy-copy			11 R	RVSYVGLVTVR	A	
				NMVLQTTKGLR	L	
			12 R		-	
xIC-new900 			12 R 13 K	LTYNHGGITGSR	G	
<ul> <li>Transformer Stress</li> <li>Transformer Stre</li></ul>				LTYNHGGITGSR LDSLVAQQLQSK	G N	
<ul> <li>XIC-new900</li> <li>mew XIC</li> <li>mew XIC-new900-Copy</li> <li>XIC-new900-Copy</li> <li>SC-ds3-ds</li> <li>sc-test</li> </ul>			13 K			
XIC-new900 			13 K 14 K	LDSLVAQQLQSK	Ν	
<ul> <li>Transformed Stress</li> <li>Transformed Stress</li> <li>Transformed Stress</li> <li>Transformed Stress</li> <li>Sc-test</li> </ul>	Ŧ		13 K 14 K 15 R	LDSLVAQQLQSK GEVGFVLVDNQR	N S	Ŧ

The less clear option is "Smart Column Size" which serves as a trade off between the aforementioned ones. It tackles with the cases that either we have too many columns to visualize using the "Automatic" approach or too few and the selected default and globally applied width imposes unneeded scrollbars ending in hiding some columns at the same time. In this context, "Smart Column Size" can be seen as a simple rule based on the ratio between the mean column width needed in case of "Automatic Column Size" for a specific table and the globally selected width. For the sake of simplicity we have set a threshold of 0.7 or 70% which on its turn determines which one of the two first modes will be used given a table. If the ratio is smaller than 0.7 then the table in question will be presented in "Fixed Column Size" mode. On contraire if ratio is equal or greater than 0.7, then we consider that using "Automatic Column Size" mode is more appropriate as it balances between a possible slightly smaller than desired width and the possibility of hiding a column using scrollbars.



#### **Column Width**

The second preference is more or less self explanatory and corresponds to the globally desired columns' width when fixed rune is applied, either directly or as a result of the smart mode.

## **General Application Settings**

Settings Categories JMS Settings Conversion/Upload Settings Table Parameters General Application Settings	Parameters Default Search Result Name Source :	Hide Getting Started Dialog On Startup Msi Search Filename Export Decorated Use dataset type to create Xic Design by DnD	
			V OK X Cancel

In this tab we can find a diverse set of preferences regarding various tasks encountered in Proline Studio. For the time being those preferences are:

- → Hide gettings Started Dialog
- → Default Search Name Source
- → Export Decorated
- → Use dataset type to create a XiC design by DnD

#### **Default Search Name Source**

Unlike the last three, Hide Getting Started Dialog is pretty much self explanatory. The second preference on the other hand, "Default Search Name Source" affects the way identification datasets are named on importation. For this preference we have three possible options:

- → Search Name (E.g Gamme Levure UPS1)
- → Peaklist (E.g OEMMA121101\_36.raw)
- → MSI Search Filename (E.g F054967)
- → Mascot Rule (...)

#### **Export Decorated**

This parameter affects the .xls and .xlsx files that are produced in the process of export (client side). It could be easily described as a preservation of any existing Rich Text Feature in a table. (Colors, Font Weight etc.)



### Use dataset type to create a XiC design by DnD

## **MsFiles**

🗫 General Settings			×
Settings JMS MsFiles Tables General Plots	Parameters Converter (.exe) : Working Set Entry Label :	È:\Local\Programs\HomeTools\raw2 Absolute Path	mzDB_0.9.10_
		🖌 ок	Cancel

#### Converter (.exe)

Corresponds to the default raw2mzDB converter. While it is left at the discretion of the user, which version to choose, it must be noted that different versions do tend to work better with specific type raw files. Said that, it is also important to understand that in order for a conversion to be successful within Proline Studio, all system requirements set by the specific raw2mzDB version must be met.

# **Identify Modification Sites or Clusters**

For description on sites and clusters you can first read the principles in this page: <u>Identifying Modification sites and</u> <u>clusters</u>

To identify Modification sites and create Modification clusters, right click on the Identification Summary and select *Identify Modification sites*.



	Display Search Result	>
±()	Display Identification Summary	>
	Add Dataset	
📄 Tes 🕕 Sma	Add Identification Folder	
l Tra	Copy Search Result	
	Paste Search Result	
	Rename	>
	Clear	
	Delete	
	Import Search Result	
	Import MaxQuant Result	
	Validate Search Result	
	Merge Datasets	>
	Filter ProteinSets	
	Change Typical Protein	

A dialog is displayed where you can choose Modification of interest and configure the clustering method (see <u>Identifying Modification sites and clusters</u> for more details).

Select the list of modifications of interest (other modifications will be ignored during clustering) and set	
the method's parameters that will be used to clusterize modification s	ite
ng method	
In localizations inferred from PSMs: this clustering method is site localization inferred from the validated PSMs, regardless reported localization probabilities.	
Group fully deaved with missed deaved peptides in same cluster	
Separate fully cleaved and missed cleaved peptides in different clusters	
Finterest	
Acetylation	
iodoacetamide derivative	
Dxidation or Hydroxylation	
Phosphorylation	
V OK Cancel	
	In localizations inferred from PSMs: this dustering method e site localization inferred from the validated PSMs, regardless reported localization probabilities. Group fully cleaved with missed cleaved peptides in same cluster Separate fully cleaved and missed cleaved peptides in different clusters finterest Acetylation Codoacetamide derivative Dividation or Hydroxylation



# **Annotate and Edit Modification Clusters**

After running Identify Modification for a specific Identification Summary or Quantitation dataset (see above and Identifying Modification sites and clusters for more details), users can annotate and edit these clusters.

In the associated display, few operations are allowed.

- Change the status of the cluster by clicking on the discon. The following dialog appears, allowing user to
  - Validate/Invalidate the cluster
  - Add a confidence level : any number which signification is user specific
  - Add a free description

🐎 Modi	ify Modification Cl	uster Status X							
0	* the Status it-s * the Status co	Change the status of selected modification Cluster. Three status properties can be modified * the <b>Status</b> it-self : validated or invalidated * the <b>Status confidence</b> : a number indicating the confidence you have in the current state * a <b>free description</b> on current cluster status.							
Modificatio	n Cluster Status:	● Validated 🖒							
		🔿 Invalidated 🛛 🖓							
Status Con	fidence level:								
		Exact Position Matching							
Status Con	fidence description:								
		V OK X Cancel							

Edit the selected cluster by clicking on the icon. In this case, the status dialog is displayed in addition with the list of peptides of the cluster. Same status annotation as described above can be done. Users can also remove some peptides from the cluster. The removed peptides will be removed from the whole "PTM Dataset". They will not be added to an existing or new cluster. Note: Peptide used as reference could not be deleted, a single peptide could be deleted at a time.



≫ E	dit Cluster 43 [PP6R3_H	IUMAN / IQQFDDGGSI	DEEDIWEEK]				
6	Edit Cluster						
	* the Status it-se * the Status cor * a free descrip Change peptides	elf : validated or invalida <b>ifidence</b> : a number indi <b>tion</b> on current duster s list.	cating the confidence you tatus.	, have in the current stat	US.	er can not be removed.	
Clust	ter Status						
lodi	fication Cluster Status:	🖲 Validated 🛛 💰	,				
		🔿 Invalidated 🛛 🖓	1		tde used to represent the duster can not be removed.		
tati	us Confidence level:						
Statu	us Confidence description	Exact Position	on Matching				<b>^</b>
lust	ter Peptides						
ĸ	Id	Peptide	Score	Exp. MoZ	Charge	PTMs	PTMs D.Mass [
		IQQFDDGGSDEEDIWEE	41.69			4 TMTpro (Any N-term)	688.3806
	25291	IQQFDDGGSDEEDIWEE	107.86	943.4362		3 TMTpro (Any N-term)	688.3806
	<						>
L						🖌 ОК	X Cancel

- Merge co-localized clusters by clicking on icon. Users must select clusters that should be merged. These clusters must be co-localized, see <u>Identify Modification Sites or Clusters</u>.
- Save edition by clicking in 🖬 icon. All modifications done on clusters will not be saved in datastore until user explicitly ask for it. When saving annotation, an Annotated Modification Dataset is created and will be accessible through the *Annotated Clusters* menu. This permits users to still access the initial Modification Dataset. Warning : there is only one Annotated Modification Dataset per Modification Dataset. If new annotation is done and saved from initial Modification Dataset, previous Annotated Modification Dataset will be erased

# **Administration**

Some administration views and operations are accessible through the Proline Studio *File > Admin* menu. Edit functions are only permitted to advanced users, who at the same time hold the status of Admin.

### **User Accounts**

The "User Accounts" tab, list all registered users with the group they belong to. Admin users can add new users or modify existing one. Modifications consist in changing user groups or passwords.



User	Group	E
lupierris	User	
dyp	User	
erre	User	
2p	User	
atem	User	
esse	User	
raut	User	
nartin	User	1
nenetrey	User	
ieter	User	1
amus	User	
ardif	User	
estjpm1	User	
	🛨 Add User Account 📄 📷 Modify User Accourt	

Login :	trauchesse	2C		
Group :	O User	) Admin		
Pasword :	*****			
	(	🖌 ОК	🗶 Cancel	0

User account modification dialog

## Peaklist Softwares

	-	Fragmentation Rule Sets		
		Version		₽
		5.3		~
		4.0		
		4.1		
		1.0		
		2.0		v
Test			🗗 Test	
	+	Add Peaklist Software	😤 View Peaklist Software	•
	Test		5.3 4.0 4.1 1.0 2.0 2.1	5.3 4.0 4.1 1.0 2.0 7 Test



Peaklist Softwares tab permits users to view the list of all available Peaklist Software. A detailed view could be displayed when selecting the "View Peaklist Software" button.

eaklist Software				
	Name :	Data Analysis		
	Version :	4.1		
Spectrum Title Parsing	g Rules			
Raw File Identifier :			< Select Predefined >	
First Cycle :			< Select Predefined >	
Last Cycle			< Select Predefined >	
First Scan :	Cmpd. +MS\d. +, \d+\.\d+ min #(\d+)		< Select Predefined >	
.ast Scan :	Cmpd. +MS\d. +, \d+\.\d+	min #(\d+)	< Select Predefined >	,
First Time :	Cmpd. +MS\d. +, (\d+\. \d+	-) min	< Select Predefined >	
Last Time :	Cmpd. +MS\d. +, (\d+\. \d+	-) min	< Select Predefined >	
Spectrum Title Parsing	a Test			
Spectrum Title:				🗗 Test
			🖌 ОК	💥 Cancel

In both windows, It is possible to test a spectrum title to verify which data are extracted for a given Peaklist Software.

Jser Accounts	Peaklist Softwares	Projects and Databases	Fragmentation Rule Sets				833.993
Peaklist Softwa							491.287
Calabe bortina					Spe	ectrum T	itle Par 🗙
Peaklist Softw	are		Version		E		
extract msn.e	xe					Raw File	e Identifier :
 Mascot Distiller	r					First Cy	cle ·
mascot.dll							
MaxQuant						Last Cy	cle :
Proline			1.0	Test result		First Sc	an :3408
Protein Pilot				restresult			
Proteome Disc	overer					Last Sc	an :3408
ProteoWizard			2.0			First Tir	me :10.1026
ProteoWizard			2.1				
ProteoWizard			3.0		1	Last Tir	ne :10.1026
Spectrum Mill					-		
Spectrum Tit	e Parsing Test						X Close
Spectrum Tit	le: 348: Scan 3408	3 (rt=10.1026) [D:/Data/N	ISData/All/HF1_010328.ra	w] 🖓 1	Test		478.313
	1						499.786
		) 🕂 A	dd Peaklist Software	Modify Peaklist Sof	twar	e	Calc. Mass
Enter	title to test						Calc. Mass
					_		
				Close	e	•	

Admin users also have the possibility to add new Peaklist Software definition (name, version and extraction rules using regular expression) or to modify existing definitions by changing the name or version. Rules are not modifiable since some already parsed data may not be consistent with the new definition. New rules definition may be done by creation of a new peaklist software.



Peaklist Software	
	Name :
	Version :
Spectrum Title Parsing	Rules
Raw File Identifier :	< Select Predefined >
First Cycle :	< Select Predefined >
Last Cycle	< Select Predefined >
First Scan :	< Select Predefined >
Last Scan :	< Select Predefined >
First Time :	< Select Predefined >
Last Time :	< Select Predefined >
Spectrum Title Parsing	,Test
Spectrum Title:	Test
	V OK X Cancel

### **Projects and Databases**

This tab lists all projects defined on the Proline Suite.

ojects		Peaklist Softwares		gmentation Rule Se									
3	Id	Project	Description	Size (MB)	Owner	Raw Files Cou	nt	Databases		Dataset Date		Properties	1
		2 Vero Dev(2)	Base de test "prod" p.		dupierris		5	lcms_db_project	2, m	11 févr. 2021			-
¥ 📊		246 VDS_29012019	d	NaN	dupierris			lcms_db_project_				{"is_active":false}	
		172 VDS_201706	Test New Prj (RC 1.5	) 19782	dupierris		10	lcms_db_project_	172	16 déc. 2021			
		29 TachesDeVert	Analyses ponctuelles	NaN	tardif		0	lcms_db_project_	29,	16 nov. 2015		{"is_active":false,	٦,
		153 REYNOIRD	Analyses pour Nicolas	NaN	tardif		20	lcms_db_project_	153	21 févr. 2017		{"is_active":false,	<b>"</b> a
<													>
\$	C	Creation Date	Identifier	Raw File Name	Raw File I	Directory	Project I	[ds	Project	s Count	Prop	erties	
	6	févr. 2015	OVEMB150205_21.raw	OVEMB150205_21.	raw <mzdb_pa< td=""><td>ath&gt;/VDS_TEMP</td><td>172</td><td></td><td></td><td></td><td>1</td><td></td><td>T</td></mzdb_pa<>	ath>/VDS_TEMP	172				1		T
	6	févr. 2015	OVEMB150205_23.raw	OVEMB150205_23	raw <mzdb_pa< td=""><td>ath&gt;/VDS_TEMP</td><td>172</td><td></td><td></td><td></td><td>1</td><td></td><td></td></mzdb_pa<>	ath>/VDS_TEMP	172				1		
	6	févr. 2015	OVEMB150205_25	OVEMB150205_25.	.mzdb <mzdb_pa< td=""><td>ath&gt;/VDS_TEMP</td><td>172</td><td></td><td></td><td></td><td>1</td><td></td><td></td></mzdb_pa<>	ath>/VDS_TEMP	172				1		
	6	févr. 2015	OVEMB150205_27	OVEMB150205_27.	.mzdb <mzdb_pa< td=""><td>ath&gt;/VDS_TEMP</td><td>172</td><td></td><td></td><td></td><td>1</td><td></td><td></td></mzdb_pa<>	ath>/VDS_TEMP	172				1		
	3	oct. 2015	QEx2_006341	QEx2_006341.mzd	b <mzdb_pa< td=""><td>ath&gt;</td><td>66,172</td><td></td><td></td><td></td><td>2 {"mz</td><td>db_file_path":"mz</td><td>•</td></mzdb_pa<>	ath>	66,172				2 {"mz	db_file_path":"mz	•
	3	oct. 2015	QEx2_006343	QEx2_006343.mzd	b <mzdb_pa< td=""><td>ath&gt;</td><td>66,172</td><td></td><td></td><td></td><td>2 {"mz</td><td>db_file_path":"mz</td><td></td></mzdb_pa<>	ath>	66,172				2 {"mz	db_file_path":"mz	
	- 2	oct. 2015	OEx2 006345	OEx2 006345.mzd	b <mzdb pa<="" td=""><td>ath&gt;</td><td>66,172</td><td></td><td></td><td></td><td>2 {"mz</td><td>db file path":"mz</td><td>Л</td></mzdb>	ath>	66,172				2 {"mz	db file path":"mz	Л

**Upper view:** List of projects with associated generic information (owner, description) and more administration information such as the name of the databases containing data, size of these databases, last dataset creation... The satus column, first one, indicates if the project is active (green), inactive (blue) or archived (grey)

**Lower view:** for selected project, list of the mzdb files used for quantitation. This is useful to do some cleanup on disk when archiving projects. The status indicated if this mzdb file is used in active projects only (green), in active and archived projects (yellow) or in archived projects only (grey).

### **Fragmentation Rule Sets**

This tab show all fragmentation rule sets defined in Proline Suite and the rules they use. The display is similar to the one defined by Mascot <sup>®</sup>



ed x g x	ESI-4SECT	x	ESI-QUAD (	x	x		x	LTQ-ORBIT	
g x									
		x	x	x	x		x		
g									
x	x								
x	x								
x									
x									
x	x	x	x	x	x				
x	x	x	x	x	x				
x	x	x	x	x	x				
						x	x		
x	x	x	x	x	x	x	x		
x		x	x	x	x				
x		x	x	x	x				
	x								
x	x								
x	x								
u									
u									
						x	x		
						x	x		
	x x x x x x x x x x x x x x x x x x x	X     X       X     X	XXX-X-XX	XXImage: select of the s	X     X <td>X     X<td>NameNa</td><td>NameNa</td><td>NNN</td></td>	X     X <td>NameNa</td> <td>NameNa</td> <td>NNN</td>	NameNa	NameNa	NNN

This information is used for <u>generating spectrum matches</u>. Custom fragmentation rule set can be added if needed by "admin" user.