PROLINE TUTORIAL

PROLINE BASICS : WORKING WITH MS/MS IDENTIFICATIONS

I/ START PROLINE

A. AUTHENTICATION

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

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

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

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

B. CREATE A NEW PROJECT

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

Projects ×	
< Select a Project >	VEZO

This opens the following window:



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

Action

Give a name and a description for the project and confirm

II/ IMPORT FILES

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

A. IMPORT F083064.DAT FILE

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

This opens the following window:

Files Selection		File selection: one or more files to import
		Parameters
		Software engine: Search engine used to
		generate the file to import. It is
		automatically selected from the selected
		file
		Instrument: Mass spectrometer which
		allowed to perform the analysis
Parameters		Fragmentation Rule Set: The
Software Engine :	Mascot 🔹	fragmentation rules specified in software.
Instrument :	ESI-TRAP 👻	by clicking on 📃 you could visualize all
Fragmentation Rule Set :	ESI-TRAP 🔹	rules for a specific rule set. This is
Peaklist Software :	extract_msn.exe 🔹	necessary to generate spectrum matches.
Decoy Parameters		Peaklist software: Software used to
Decoy :	Concatenated Decoy 👻	generate the peaklist
Decoy Accession Regex :	###REV###\S+	Decoy: Target/Decoy strategy used during
		the search
Parser Parameters		Decoy Accession Regex: Regular
Subset Threshold : 1.0		expression enabling target and decoy
		protein differentiation in a concatenated
Save Load	→ OK X Cancel 3	bank
		Use icon to choose a predefined

Parser Parameters

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

Import parameters can be saved and reused.

Note

See Proline Help for more details on the various parameters

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

B. IMPORT FILES F083066.DAT AND F083067.DAT

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

Action Import files with the same settings as before.		
Note	More than one file can be imported at the same time.	

C. VIEW THE CONTENTS OF FILE $F083604.\mbox{dat}$

In Proline, the content of a file before validation is called Search Result. To view the Search Result,

you can right-click on the icon representing the imported file and then click *Display Search Result* and *PSMs*...

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

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

III/ Validating file F083064.dat

A. VALIDATION ON RANK 1

VALIDATION CRITERIA

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

dentification Validation X	< PSM		
Validation Parameters Typical Protein Parameters PSM Propagate PSM filtering to child Search Results Prefilter(s) FDR Filter Ensure FDR <= 5.0 % on	 Prefilter(s): Allows filtering according to various criteria (rank, length, score) FDR Filter: Allows filtering PSMs to obtain the FDR requested by varying a parameter 		
Protein Set	Protein Set		
Propagate ProteinSets filtering to child Search Results (Warning FDR Validation will not be propagated ! Filter(s)	• Filter(s): Allows filtering proteins		
< Select > ~	according to various criteria (number of specific peptides)		
PDR Filter Protein FDR <=	• FDR Filter : Allows filtering proteins to get		
Scoring Type: Standard 🗸	the request FDR by varying the score		
☐ Save 〕 Load	• Scoring type: Method chosen to calculate protein score		

Validation parameters can be saved and reused.

Action	 Only keep pretty rank 1 PSMs Scoring type: « <i>Standard</i> »

VALIDATION RESULTS

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

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



SEARCH FOR A VALID PROTEIN SET

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

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

Action Search for RSSA2_YEAST Search for RSS*_YEAS?		
Note	Wildcards "?" and "*" are available in both tools. The interrogation mark matches for any character once, the star character matches for zero or more characters. Searching for " RSS*_YEAS? " will return all Protein Sets starting with "RSS", containing "YEAS" with only one letter after it.	

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

	Protein Set	Description	Score	Proteins	
8	E.			364.42	2 (2 🗖 , 0 🚄)
×	Protein Set	✓ Protein Set = ✓ RSS*_YEAS?		362.83	1 (1 🗖 , 0 🚄)
		opp 2700011112_10101 110pointi		362.23	1 (1 🗖 , 0 🚄)
15	6 HSP78_YEAST	sp P33416 HSP78_YEAST Heat s		358.70	1 (1 🗖 , 0 🚄)
15	6 SODC_YEAST	sp P00445 SODC_YEAST Supero		358.03	1 (1 🗖 , 0 🚄)
15	7 SYFA_YEAST	sp P15625 SYFA_YEAST Phenyla		357.88	1 (1 🗖 , 0 🚄)
15	8 RL10_YEAST	sp P41805 RL 10_YEAST 60S ribo		355.80	1 (1 🗖 , 0 🚄)
15	9 PYC1_YEAST	sp P11154 PYC1_YEAST Pyruvat		354.98	1 (1 🗖 , 0 🚄)
16	0 DSF1_YEAST	sp POCX08 DSF1_YEAST Mannit		351.98	2 (2 , 0)
16	IMDH4_YEAST	sp P50094 IMDH4_YEAST Inosin		349.65	1 (1 🗖 , 0 🚄)
16	2 ASNS2_YEAST	sp P49090 ASNS2_YEAST Aspar		349.54	1 (1 , 0)
16	B HSP31_YEAST	sp Q04432 HSP31_YEAST Proba		345.54	1 (1 , 0)
16	4 KAD1_YEAS1	sp B3LG61 KAD1_YEAS1 Adenyl		341.69	3 (3 , 0)
16	5 RSSA2_YEAS1	sp B3LT19 RSSA2_YEAS1 40S ri		340.63	6 (3 , 3)
Тур	ical Protein: sp B3LT19	PIRSSA2_YEAS1 40S ribosomal protein S0-B OS=Sa Description	ccharomyces cerevisiae Sameset / Subset	(strain RM11-1a) GN=R Score	PS0B PE=3 SV=1 Peptides
1	a RSSA2 YEAS1	sp IB3LT19IRSSA2_YEAS1_40S ribosomal pr	TT I	340.6	53 10
2	RSSA2 YEAST	splP46654IRSSA2_YEAST_40S ribosomal pr		340.6	53 10
3	RSSA2 YEAS7	sp A7A0V3 RSSA2 YEAS7 40S ribosomal pr		340.6	53 10
	BSSA1 YEAS7	sp[A6ZUM5]RSSA1_YEAS7 40S ribosomal p.,	. 🛛	311.7	75 9
4	C NOUNT ILNUT		the second se		
4	RSSA1_YEAS1	sp B3LI22 RSSA1 YEAS1 40S ribosomal pr	97	311.7	75 9

TYPICAL PROTEIN CHOICE

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

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

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

ANNOTATE SPECTRUM AND GENERATE FRAGMENTATION TABLE

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

For each spectrum by clicking on the icon on the left

Or on the whole set of valid spectra



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

Generate Spectrum Matches	
Choose Fragmentation Rule Set to use to ge	enerate (new) spectrum matches.
Force new spectrum matches generation	1
Fragmentation Rules	
Use fragmentation rule set defined at	import (unknown - see dataset properties)
Select Fragmentation Rule Set :	•
	VK X Cancel

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

B. VALIDATION AT PEPTIDIC LEVEL

VALIDATION CRITERIA

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

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

VALIDATION RESULTS

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

D. VALIDATION AT PEPTIDIC AND PROTEIC LEVEL

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

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



A. MERGE VALIDATED FILES

CREATE A DATASET

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

This open the following window:

	1	Add Dataset		
Dataset	Parameters			
Name:	Validate then merge			
Type:	Biological Group			~
		Create Multiple Da	atasets	2 🌲
🕑 Defa	ult	✓ ОК	💥 Cano	cel 🚺

Action	Action Name the dataset « Validate then merge »	
Trick	The box « Create Multiple Datasets » allows to create a series of datasets in one click	

PLACE A COPY OF EACH FILES IN THE DATASET

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

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

VALIDATE THE THRE	EE FILES
Action	Validate the files with the same settings as before (third set of criteria)
Trick	To validate several files with the same criteria at once, select all the files together, right-click on them and select "Validate Search Result…"

Merge the three files

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

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

NoteThe two merge methods are explained in the "Proline User Guide", "Concepts &
principles" part.

VIEW THE RESULTS

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

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

B. VALIDATE MERGED FILES

 $\mathbf{C}_{\mathsf{REATE}}$ a second dataset but merge before validation

Action	 Run the previous steps again but in the following order: Create a new dataset "Merge then validate" Put a copy of the three files in this dataset Merge the dataset
	 Validate the dataset with the same settings as before
	After a merge, the dataset can be revalidated but not the files that make it up.
Note	However, the validation done on the parent dataset could be propagated to its childs. The same parameters or calculated parameters will be used.



Action	Look for RL26A_YEAST protein among the validated <i>Protein Sets</i> .
Note	Look for RL26* <i>Proteins Sets</i> and look for samesets-/ subsets

V/ EXPORT RESULTS

Proline proposes many ways to export data:

A global one is available after validation.



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

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



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

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

• Export the complete validated results for this dataset

VI/ To go further

A. QUALITY CONTROL

Quality control enables а tidentifications All Imported transversal view on a Search Merge then validate Result. Rather than visualising F0830 **F0830 Display Search Result MSOueries** the results per PSM or Proteins, **F0830 Display Identification Summary PSMs** 🖨 🚺 Merge the results are sorted according to Proteins Add Dataset... F0830 ranges of score, M/z, charge Add Identification Folder Quality Control.. **F0830 F0830** state, target or decoy... Copy Search Result New User Window... Validate th Manage User Windows... **F0830** Paste Search Result **F0830** Rename **F0830**

B. COMPARE THE RESULTS WITH DATA ANALYZER

The Data Analyzer tool can run a large amount of computation on any kind of data. Every view of any

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





C. OTHER FUNCTIONS OF PROLINE STUDIO

UNIPROT LINK

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

1	Protein	Score
1	PYR1_YEAST	2900.1
2	E=2_YEAST	2193.3
3	A KPYK1_YEAST	2096.0
4	G3P3_YEAST	2069.8
5	BIO2_YEAST	2046.3
6	PDC1_YEAST	1770.6
7	BP71_YEAST	1737.3
8	ENO1_YEAST	1709.3
	LIGHT WELCH	

EXPORT A SPECTRUM LIBRARY

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

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



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

CUSTOM VIEWS

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



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

F d	Display Search Result	>	
T FC	Display Identification Summary	>	MSQueries
Valida FC FC FC	Add Dataset Add Identification Folder		PSMs Peptides Protein Sets
Merge F0830	Copy Search Result Paste Search Result		Adjacency Matrix Modification Sites
30	Rename	;	Modification Clusters
sh	Clear Delete		New User Window Manage User Windows
	Import Search Result Import MaxQuant Result		Peptides Spectum



E. EXPORT TO PRIDE REPOSITORY

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

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

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

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

Note

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

F. USING MARKER BAR

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

G. SET COLUMN VISIBILITY

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

icon at the top-right

PYR1_YEAST 2805.33 90 Id Protein Id	Protein	Score		Peptides	N	lass	
2 EF2_YEAST 2326.62 60 10 3 EVPK1_YEAST 2203.27 50 ✓ Protein 4 ENO2_YEAST 2116.29 40 ✓ Score 6 G32_YEAST 2066.58 40 ✓ Peptides 6 G92_YEAST 1950.58 42 ✓ Peptides 6 G92_YEAST 1923.82 47 ✓ Mass 6 HSP1_YEAST 1803.89 36 Défilement horizontal 0 ENSP3_YEAST 1702.89 44 Compacter toutes les colon	PYR1_YEAST		2805.33		90	L4	
B & KPYK1_VEAST 2203.27 50 ✓ Protein 1 & ENO2_YEAST 2116.29 40 ✓ Score 50 ✓ Pertein 1 & ENO2_YEAST 2066.58 40 ✓ Score Peptides 0 & PDC1_YEAST 1950.58 42 ✓ Peptides 1 & HSC3_YEAST 1923.82 47 ✓ Mass 2 & HSP71_YEAST 1847.17 39 Officient horizontal 2 & ENSP3_YEAST 1702.89 44 Compacter toutes les colon	EF2_YEAST		2326.62		60	14	
Image: Brook yr EAST 2116.29 40 ✓ Score Image: Brook yr EAST 2066.58 40 ✓ Peptides Image: Brook yr EAST 2066.58 40 ✓ Peptides Image: Brook yr EAST 1950.58 42 ✓ Peptides Image: Brook yr EAST 1923.82 47 ✓ Mass Image: Brook yr EAST 1847.17 39 Defilement horizontal Image: Brook yr EAST 1702.89 44 Compacter toutes les colon	KPYK1_YEAST		2203.27		50	Protein	
G G3P3_YEAST 2066.58 40 Peptides G G3P3_YEAST 1950.58 42 Peptides G HSC8_YEAST 1923.82 47 Mass G HSC8_YEAST 1923.82 47 Mass G HSC9_YEAST 1803.89 36 Défilement horizontal G HSC9_YEAST 1702.89 44 Compacter toutes les colon	ENO2_YEAST		2116.29		40	✓ Score	
PDC1_YEAST 1950.58 42 Peptides 10 PDC1_YEAST 1923.82 47 Mass 10 FISC2_YEAST 1923.82 47 Mass 10 FISC2_YEAST 1847.17 39 10 FISC2_YEAST 1803.89 36 Défilement horizontal 10 FISP82_YEAST 1702.89 44 Compacter toutes les colon	G3P3_YEAST		2066.58		40		
1923.82 47 ✓ Mass 0 A HSP71_YEAST 1924.17 39 0 A HSP71_YEAST 1847.17 39 0 A HSP71_YEAST 1803.89 36 0 A HSP82_YEAST 1702.89 44 Compacter toutes les colon 50 Compacter toutes les colon	PDC1_YEAST		1950.58		42	Peptides	
B HSP71_YEAST 1847.17 39 B ENO1_YEAST 1803.89 36 Défilement horizontal B HSP82_YEAST 1702.89 44 Compacter toutes les colon	HSC82_YEAST		1923.82		47	✓ Mass	
ENO1_YEAST 1803.89 36 Défilement horizontal B HSP82_YEAST 1702.89 44 Compacter toutes les colon Compacter toutes les colon	HSP71_YEAST		1847.17		39		
Compacter toutes les colon Compacter toutes les colon	ENO1_YEAST		1803.89		36	Défilement horizontal	
A LODZD VEACT	HSP82_YEAST		1702.89		44	Compacter toutes les colonne	
1 10/5.05 3/	HSP72_YEAST		1675.05		37	compacter toutes les colorina	
	PGK YEAST		1640.33		38		

A more complete configuration is available using \gg icon.

oles Sorting Paramet	ers Overview Parameters				
umns Arrangement :	Smart Column Size				
Column Width :	120				
C <mark>olumn</mark> s Visibility					
Hidden Columns				Visible Columns	
Id #Quant. Peptide Sel. level F0712: Raw abundance #Quant. Peptide Sel. level F0712: Raw abundance #Quant. Peptide Sel. level F0712: Raw abundance #Quant. Peptide Sel. level F0712:	s F071236 66 F071236 s F071237 77 F071237 is F071238 88 F071238 is F071239 99 F070000	~	<pre> Search for Text> Select from Prefix/Suffix> </pre>	Protein Set Overview Description #Peptide #Quant. Peptide #Quant. PSMs F071236 Abundance F071237 Abundance F071237 #Quant. PSMs F071237 #Quant. PSMs F071238 Abundance F071238 #Quant. PSMs F071239	~