

# **X!TandemPipeline** 3.4.0 (Elastine Durcie)

validating, filtering and grouping MSMS identifications

Olivier Langella  
langella@moulon.inra.fr  
PAPPSO - <http://pappso.inra.fr/>



28 June 2016

## **Abstract**

**X!TandemPipeline** is a standalone, easy to install, light and powerfull software to validate and group peptide/protein identifications from MS/MS mass spectra. **X!TandemPipeline** can read Mascot dat files as well as **X!Tandem** XML result files. It features a rich graphical user interface (GUI) that helps users to browse their MS/MS identification results. It helps also to launch **X!Tandem** analysis with the full support of all **X!Tandem** parameters.

**X!TandemPipeline** performs database searching and matching on many MS/MS runs in one shot. It focuses on the ease of use to select databases and identification parameters.

**X!TandemPipeline** also performs filtering of data according to statistical values at peptide and protein levels. Moreover, redundancy of protein databases are fully filtered as follows :

- proteins identified without specific peptides compared to others are eliminated;
- proteins identified with the same pool of peptides are assembled;
- proteins are grouped by function (identified with at least one common peptide), and the specific peptides for each sub-group of proteins are indicated.

**X!TandemPipeline** allows to view and edit the filtered results, compute the false discovery rate, ... The results can be exported into TSV (Tab Separated Values) files or directly to a spreadsheet software format using ODS (Open Document Spreadsheet).

# Contents

<b>1</b>	<b>Installation</b>	<b>5</b>
1.1	License . . . . .	5
1.2	Requirements . . . . .	5
1.3	Third party softwares for Windows and Mac . . . . .	5
1.4	Third party softwares for Linux . . . . .	5
1.5	Start <b>X!TandemPipeline</b> . . . . .	6
1.6	Configuration . . . . .	6
<b>2</b>	<b>X!Tandemanalysis</b>	<b>7</b>
2.1	Parameters . . . . .	7
2.2	Running analysis . . . . .	7
2.3	Peak-lists . . . . .	8
2.4	Databases . . . . .	8
<b>3</b>	<b>Processing the results</b>	<b>9</b>
3.1	Three modes of analysis . . . . .	9
3.2	Filter parameters . . . . .	9

<b>4</b>	<b>View and edit the results</b>	<b>11</b>
4.1	Main window . . . . .	11
4.2	Proteins List . . . . .	11
4.3	Protein Abundance Index (PAI) computation . . . . .	12
4.4	Exponentially Modified Protein Abundance Index (emPAI) computation . . . . .	13
4.5	Protein Details . . . . .	13
4.6	Peptides List . . . . .	13
4.7	Peptides Details . . . . .	14
<b>5</b>	<b>Save and Load X!TandemPipeline project</b>	<b>15</b>
<b>6</b>	<b>Exporting the results</b>	<b>16</b>
6.1	Export settings . . . . .	16
6.2	Spreadsheet report . . . . .	16
6.3	<b>X!TandemPipeline</b> informations table . . . . .	18
6.4	Proteins table (Fig 11) . . . . .	18
6.5	Peptides table (Fig 12) . . . . .	19
6.6	Spectra table (Fig 13) . . . . .	19
6.7	Peptides position table (Fig 14) . . . . .	20

6.8	MS samples table . . . . .	21
6.9	Compar spectra table (Fig 15) . . . . .	22
6.10	Compar specific table (Fig 16) . . . . .	23
6.11	Compar unique table . . . . .	23
6.12	Compar specific unique table . . . . .	24
6.13	Compar PAI table (Fig 17) . . . . .	24
6.14	PhosphoIsland table (Fig 18) . . . . .	25
6.15	PhosphoPeptides table (Fig 19) . . . . .	25
6.16	Phosphopeptide mode compar Spectra table (Fig 20) . . . . .	27
6.17	Phosphopeptide mode Spectra table (Fig 21) . . . . .	27
6.18	Phospho stats table . . . . .	28
6.19	SequenceLI column . . . . .	28
6.20	Files *fdr.txt . . . . .	29
<b>7</b>	<b>Changelog</b>	<b>30</b>
7.1	"Elastine" branch . . . . .	30
7.2	"Myosine" branch . . . . .	30
7.3	"Tubuline" branch . . . . .	30
7.4	"Kératine" branch . . . . .	31

# 1 Installation

## 1.1 License

Copyright (C) 2010 Olivier Langella and Benoit Valot

**X!TandemPipeline** program is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the [GNU General Public License](#) for more details.

## 1.2 Requirements

**X!TandemPipeline** works on all platforms (Linux, Windows and Mac). Java 1.7 must be installed (it can be at : <http://java.com/fr/download/index.jsp>).

## 1.3 Third party softwares for Windows and Mac

Download and install the **X!Tandem** executable from the **X!Tandem** site (<http://www.thegpm.org/tandem/>).

## 1.4 Third party softwares for Linux

### Debian or Ubuntu

- Follow instructions on how to install the PAPPISO Debian repository [http://pappso.inra.fr/bioinfo/install\\_debian\\_jessie.php](http://pappso.inra.fr/bioinfo/install_debian_jessie.php).
- Install the *tandem-mass* package.
- You can also install the *xtandempipeline* package to run **X!TandemPipeline** instead of using the jlnp link.

### Other distributions

- Please visit the **X!Tandem** site, and follow instructions about getting and compile the source code.



## 1.5 Start **X!TandemPipeline**

To run **X!TandemPipeline**, simply :

- Open **X!TandemPipeline** by using this [link](#)
- Wait for the program to execute
- The main window will appear (Fig 5)

## 1.6 Configuration

At the first start, the application open the configuration path window:

- Open the menu *Option* → *Configuration Path* (Fig 1).
- Define the path to the **X!Tandem** executable
- Choose the folder where to store the **X!Tandem** parameters (or used default one).
- Choose the folder where the MS/MS data, the protein databases and the **X!Tandem** results are stored

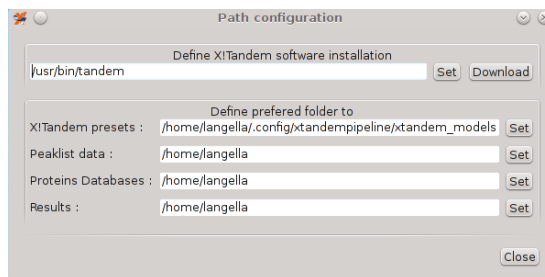


Figure 1: Configuration window

## 2 X!Tandemanalysis

**X!TandemPipeline** allows you to analyze peak-lists files by searching a list of protein databases using the **X!Tandem** software. Three successive graphical boxes help you select first the mzXML files or other peak-lists, then the protein databases and finally the folder where the results will be stored. The databases must be protein ones, **X!Tandem** does not work on DNA databases.

### 2.1 Parameters

To perform database searching, you must create or edit a model XML file (stored in the xtandem models folder). Open the menu *Option* → **X!Tandem preset** (Fig 2).

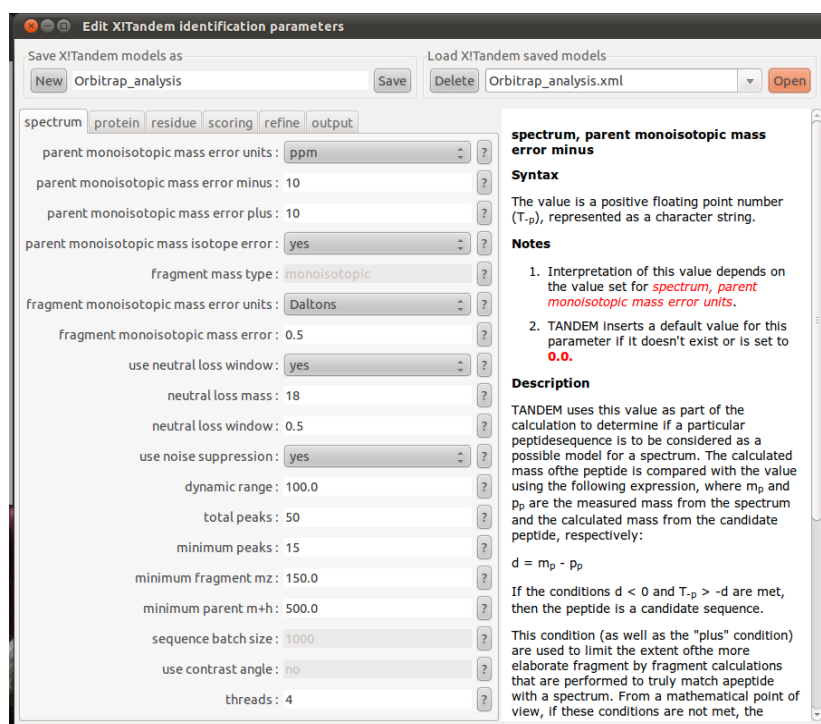


Figure 2: X!Tandem preset window

To use complete performance of your computer, specify the number of CPU in the model : spectrum → threads.

### 2.2 Running analysis

To perform analysis, start the menu *File* → **X!Tandem** → *Analysis*. Select on the window (Fig 3) :



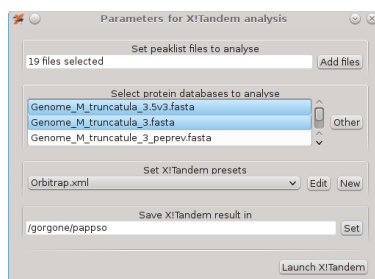
1. Peak-list files to be analyzed (See 2.3)
2. Database files to be searched (See 2.4)
3. Searching parameters model (See 2.1)
4. Folder where to store the result files

## 2.3 Peak-lists

**X!Tandem** works with open peak-list files like mzXML, mgf, mzData, mzML or pkl files.

## 2.4 Databases

**X!Tandem** software uses only protein databases in fasta format. It doesn't work with EST <sup>1</sup> sequences. You can transform your database using our application *Database Manager*, available at <http://pappso.inra.fr/bioinfo/pdm>.



<sup>1</sup>Expressed Sequenced tag

Figure 3: **X!Tandem** parameter window



## 3 Processing the results

**Warning:** To process results, **X!TandemPipeline** needs to have **X!Tandem** result files (.xml) or Mascot result files (.dat). The names of the files are used as **sample names**.

### 3.1 Three modes of analysis

You can filter the MS/MS identification results and export them in three different modes : (menu *File* → *Load Result*)

#### Individual mode

Each MS/MS result file is processed individually.

You cannot perform comparison by using this process because each sample as its own protein/peptide list.

#### Combined mode

The MS/MS result files are combined in one result file, and this file is filtered / exported.

This mode is useful to compare different results : there is only one protein/peptide list for all samples.

#### Phosphopeptide mode

Same as the combined mode analysis except that only phosphopeptides are conserved and the result is oriented in order to validate phosphosites.

In all modes, you have to define the filter parameters.

### 3.2 Filter parameters

The filter window (Fig 4) defines the automated filtering process parameters :

#### Add files

At this stage, you can add other MS/MS result files to the analysis. If two files have the same name, they are combined in one result file. Interesting if one wants to combine **X!Tandem** and/or Mascot results of the same LC-MS/MS run using different modification parameters or protein databases.

#### Peptide E-value

Defines the E-value above which a peptide is considered as valid.

#### Peptide Prophet probability

Defines the Peptide Prophet minimum probability threshold. This is only relevant if you are loading pepXML files. In other cases, this parameter will be simply ignored.

#### Peptide iProphet probability

Defines the minimum Peptide iProphet probability threshold. This is only relevant if you are loading pepXML files, and if the iProphet probability was computed. In other cases, this parameter will be simply ignored.



**Peptide number**

Defines the number of valid unique<sup>2</sup> peptides necessary to validate a protein.

**Protein E-value**

Defines the E-value above which a protein is considered as valid.

- The protein E-value is the product of its valid unique peptide E-values and it is different from the protein E-values determined by **X!Tandem**.
- The values are expressed in  $\log(E\text{-value})$ .

**Apply protein filter to all samples together**

Defines how protein filter is performed when MS/MS results are combined :

**Disabled** To validate a protein, the 2 parameters (peptide number and protein E-value) must be valid in at least one result. Interesting if one wants to compare SDS-PAGE-LC-MS/MS results, where peptides from a protein are in the same LC-MS/MS run.

**Enabled** To validate a protein, the 2 parameters (peptide number and protein E-value) must be valid in the sum of all results. Interesting if one wants to compare 2DLC-MS/MS results, where peptides from a protein are split in different LC-MS/MS runs.

**Contaminants**

When you perform an analysis using different fasta databases, you can remove the result from one database by selecting this database. Interesting because it allows you to always include the same contaminant proteins during the database search, and because it removes the contaminant proteins from the results.

<sup>2</sup>Unique peptides are defined as peptides with different sequences. This excludes peptides with different modifications.

## 4 View and edit the results

After loading the results, you can select the result to view in the main window (see 4.1). After this selection, you can navigate in this result in four different windows listed in the menu *Windows* :

### 4.1 Main window

- First frame "Identification Results" : choose the result to edit, displays the current number of samples, groups and subgroups.
- False Discovery Rate : estimates an FDR using a reverse/decoy database
- Mass precision : computes the standard deviation between theoretical and observed mass of peptides
- Filter identification results : choose criterium to validate identifications as described in 3.2

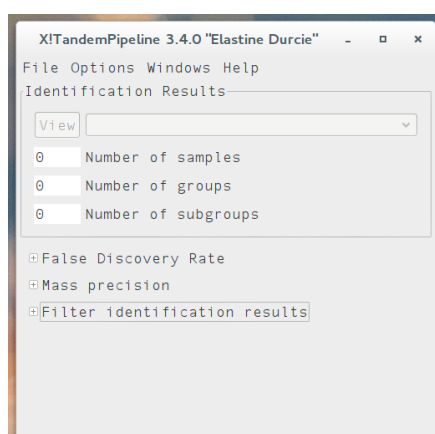


Figure 5: Main window

### 4.2 Proteins List

View the list of protein identified on the result. For more details on column see Fig 4.2.

- Filter the protein by description;
- Click on a protein to view the corresponding peptides list (see 4.6) and protein details (see 4.5);
- The checkbox on each protein line allow to validate or unvalidate corresponding peptides;
- **Apply modification** to validate the edition.

Protein list

Combine results

Search

Cancel

Apply modifications

	group	subgroup	Protein Description	Spectra	Specific Uniques	Uniques	PAI
	1	1.01	GRMZM2G083841_P01 P04711 Phosphoenolpyruvate 241	52	56	1.6666666	
	1	1.02	GRMZM2G069542_P01 NP_001105503 phosphoenolpyruvate 32	9	16	0.37209302	
	1	1.02	GRMZM2G074122_P01 NP_001130365 hypothetical protein 32	9	16	0.37209302	
	1	1.02	GRMZM2G074122_P02 NP_001130365 hypothetical protein 32	9	16	0.37209302	
	1	1.03	GRMZM2G473001_P01 P51059 Phosphoenolpyruvate 25	10	14	0.28846154	
	2	2.01	GRMZM2G385622_P01 P00827 ATP synthase subunit 145	7	28	1.64	
	2	2.01	GRMZM2G0808402_P01 P00827 ATP synthase subunit 145	7	28	1.64	
	2	2.02	GRMZM2G448142_P02 P06670 NAD(P)H-quinone oxidoreductase 110	1	22	1.7368422	
	2	2.02	GRMZM2G448142_P03 P06670 NAD(P)H-quinone oxidoreductase 110	1	22	1.7368422	
	2	2.03	GRMZM2G113408_P01 P19023 ATP synthase subunit 45	2	18	0.8888889	
	2	2.04	GRMZM2G041275_P01 seq=translation; coencoded 44	2	18	0.85714287	
	2	2.05	GRMZM2G021331_P01 NP_001151807 ATP synthase subunit 40	1	16	0.8148148	
	2	2.06	GRMZM2G550611_P01 seq=translation; coencoded 12	1	2	0.75	
	2	2.07	GRMZM2G109613_P01 seq=translation; coencoded 7	2	4	2.0	
	3	3.01	GRMZM2G089136_P01 NP_001147628 phosphoglycerate kinase 75	11	23	1.1724138	
	3	3.01	GRMZM2G089136_P02 NP_001147628 phosphoglycerate kinase 75	11	23	1.1724138	
	3	3.02	GRMZM2G083016_P01 NP_001151968 metacaspase 62	7	20	1.1538461	

Protein groups : 1031 - Protein subgroups : 1306 - Proteins : 2314

Figure 6: Proteins List

Protein list columns :

**group**

Protein group number : proteins are grouped together if they share at least one peptide.

**subgroup**

Protein subgroup number : proteins inside a group sharing exactly the same set of peptides (indistinguishable).

**Protein Description**

The protein description as given by the identification engine.

**Spectra**

Total number of scans (spectra) leading to the identification of this protein.

**Specific Uniques**

Number of unique amino acid sequence specific to this protein (or subgroup).

**Uniques**

Total number of unique amino acid sequence leading to the identification of this protein.

**PAI**

Protein Abundance Index as described in 4.3.

## 4.3 Protein Abundance Index (PAI) computation

The PAI is an estimation of the abundance of the protein. PAI was defined as the number of sequenced peptides (fragmentation spectra assigned with significant score and as the top match to an individual identified protein) divided by the number of observable peptides per protein<sup>3</sup>.

In **X!TandemPipeline**, Readily observable tryptic peptides were taken to be those in the mass range 800 to 2500 Daltons. Fragmentation spectra matching the same peptide sequence but with different charge states and containing missed cleavage sites are counted separately. For this reason, the index can be > 1.

<sup>3</sup>Rappaport, J., et al., 2005. Large-scale proteomic analysis of the human spliceosome. Genome. Res. 12, 1231-1245

## 4.4 Exponentially Modified Protein Abundance Index (emPAI) computation

the emPAI<sup>4</sup> is a PAI transformation such as :  $emPAI = 10^{PAI} - 1$ .

## 4.5 Protein Details

View the protein sequence and coverage on a identified protein. To view this window, you must open it in the menu *Windows* → *Protein details*.

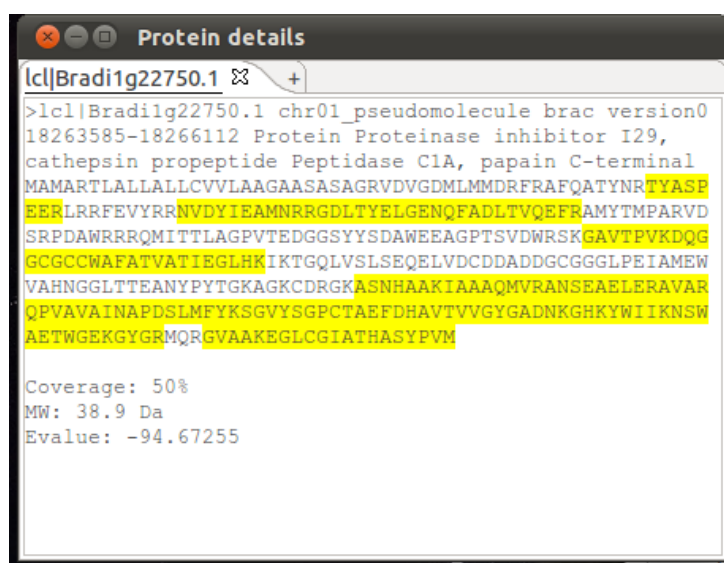


Figure 7: Protein details

## 4.6 Peptides List

View the peptides identifying a protein..

- Filter the peptide by different options;
- Click on a peptide to view the corresponding MS/MS spectra (see 4.7);
- Uncheck peptide to unvalidate it.

Peptide list columns :

<sup>4</sup>Ishihama, Y., et al., 2005. Exponentially Modified Protein Abundance Index (emPAI) for Estimation of Absolute Protein Amount in Proteomics by the Number of Sequenced Peptides per Protein. Mol Cell Proteomics 4, 1265–1272

Peptide list

GRMZM2G083841\_P01 8

GRMZM2G083841\_P01 P04711 Phosphoenolpyruvate carboxylase 1 (PEPCase 1)(PEPC 1)(EC 4.1.1.31) seq=translation; coord=9:61296279..61301686:1; parent\_transcript=GRMZM2G083841\_T01; parent\_gene=GRMZM2G083841

Sample	Scan	Rt	Charge	Sequence	Modifs	Used	Sub-groups	E-value	MH+	Theo	Delta MH+
20120906_bal	1976	11.11	2	AQEEMAQVAK	M5;+15;-	1.01		9.7E-4	1120.5306396484	-0.0012000000569969416	
20120906_bal	2769	14.0	2	HHSIDAQLR	-	1.01		5.9E-4	1076.5600585937	-0.001500000013038516	
20120906_bal	3051	14.8	2	RAQPTPQDEMR	-	1.01		3.7E-6	1328.6379394531	-6.000000284984708E-4	
20120906_bal	3076	14.86	3	RAQPTPQDEMR	-	1.01		1.8E-5	1328.6379394531	1.9999999494757503E-4	
20120906_bal	3555	16.16	2	SATPETEYGR		2/3	1.01 1.02	8.7E-5	1110.5065917968	-3.9999998989515007E-4	
20120906_bal	3586	16.25	2	AQEEMAQVAK	-	1.01		2.4E-4	1104.53576660151	-6.99999975040555E-4	
20120906_bal	3805	16.83	2	SATPETEYGR		2/3	1.01 1.02	2.3E-5	1110.5065917968	-3.9999998989515007E-4	
20120906_bal	4040	17.48	2	AQPTPQDEMR	-	1.01		1.6E-9	1172.5368652343	0.0	
20120906_bal	4056	17.51	2	SATPETEYGR		2/3	1.01 1.02	5.6E-4	1110.5065917968	-3.9999998989515007E-4	
20120906_bal	4205	17.93	2	RPGGGITTLR	-	1.01		8.1E-4	1027.6010742187	-6.99999975040555E-4	
20120906_bal	4220	17.96	2	GKQVMVGYSDSG	-	1.01		3.8E-10	1483.7214355468	-0.0010000000474974513	
20120906_bal	4223	17.98	3	GKQVMVGYSDSG	-	1.01		4.3E-7	1483.7214355468	-0.0024000000113993883	
20120906_bal	4292	18.16	2	QQVMVGYSDSGK	M4;+15;-	1.01		1.2E-4	1314.5998535156	-0.003700000001117587	
20120906_bal	4536	18.78	2	AQPTPQDEMR	-	1.01		2.5E-7	1172.5368652343	-0.0	
20120906_bal	4549	18.81	2	QIPPNEPYR	-	1.01		3.7E-5	1113.5692138671	5.000000237487257E-4	
20120906_bal	4817	19.48	2	QIPPNEPYR	-	1.01		0.0021	1113.5692138671	-0.0020999999925121665	
20120906_bal	5189	20.38	2	QIPPNEPYR	-	1.01		0.0024	1113.5692138671	-0.0020999999925121665	

Sample Search Cancel 241 scans

Figure 8: Peptides List

**sample**

The identification engine result filename from which the peptide identification was found.

**Scan**

Scan number (spectrum reference) that leads to this peptide identification

**Rt**

Retention time in minutes

**Charge**

Charge of the identified peptide

**Sequence**

Amino acid sequence of the identified peptide

**Modifs**

Positions and mass delta of amino acid modifications

**Used**

Number of subgroups in which this peptide was identified. Empty string if this peptide was not shared by other subgroups (specific). If the peptide is shared between subgroups, this indicates the number of subgroups concerned on the total number of subgroups in this group.

**Sub-groups**

Subgroups in which this peptide was identified.

**E-value**

The E-value for this peptide spectrum match (PSM).

**MH+ Theo**

The theoretical peptide ion mass with  $z=1$ .

**Delta MH+**

The difference between the observed and theoretical peptide mass given by the identification engine (observed minus theoretical).

## 4.7 Peptides Details

View the MS/MS spectra of an identified peptide.



- Click on spectra to zoom.
- Save MS/MS annotated spectra on png or svg.

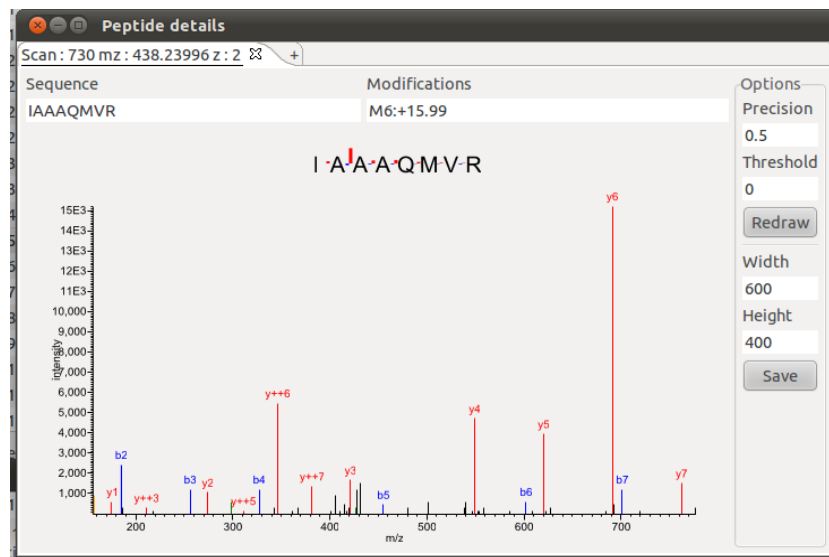


Figure 9: Peptides Details

## 5 Save and Load X!TandemPipeline project

You can save all the current results using menu *File* → *Save Project*, or load a previous one using menu *File* → *Load Project*. The extension of created files is *\*.xpip*.

## 6 Exporting the results

You can export the result in different formats in menu *File* → *Export*.

### 6.1 Export settings

The export window (Fig 10) shows the different types of available exports :

#### Spreadsheet report

Produces a tabulated report either in an ODS file <sup>5</sup> or using a classical TSV file<sup>6</sup>. this output is detailed in the Spreadsheet report section 6.2.

#### Fasta

Creates a fasta file containing the validated proteins.

#### FDR

Creates two tabulated files containing the number of valid peptides or valid proteins for the different E-values in each database (see 6.20). Allows you to determine the E-value above which FDR value is acceptable.

#### PepNovo

Creates a XML file containing the peptide results to be removed for an automated *De Novo* interpretation in sequence using our [DeNovo pipeline](#).

#### MassChroQ

Creates a MassChroQ compatible XML file, so you can perform quantitative analysis using our **MassChroQ** software.

#### PROTICdb

Creates a PROTICdb compatible XML file, so you can store results in [PROTICdb](#) proteomic database.

#### MassChroqPRM

XML format dedicated to a PRM PAPPSO software that is not yet available for public use.

### 6.2 Spreadsheet report

**Individual mode** In this mode, the spreadsheet report will produce six tables. Proteins are listed for each samples.

**X!TandemPipeline informations** see 6.3.

**proteins** see 6.4.

**peptides** see 6.5.

**spectra** only available in expert mode. see 6.6.

**peptides\_pos** only available in expert mode. see 6.7.

**MS samples informations** see 6.8.

---

<sup>5</sup>ODS is an open format, similar to xls, readable with MS Office or LibreOffice

<sup>6</sup>Tab Separated Value





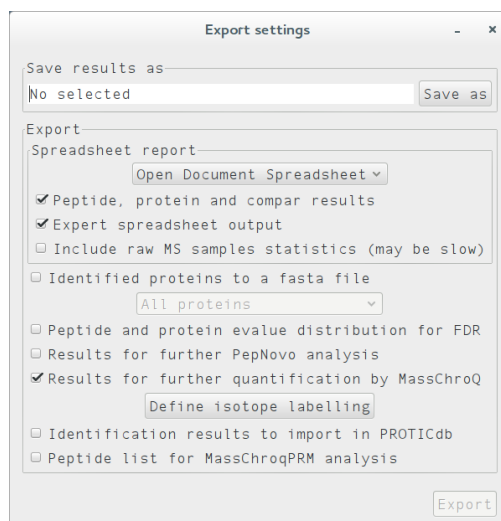


Figure 10: Export window

**Combined mode** In this mode, the spreadsheet report will produce eleven tables. There is a unique protein list, the same for all samples.

**X!TandemPipeline informations** see 6.3.

**proteins** see 6.4.

**peptides** see 6.5.

**compar\_Spectra** only available in expert mode. see 6.9.

**compar\_Specific** see 6.10.

**compar\_Unique** see 6.11.

**compar\_Specific\_Unique** see 6.12.

**compar\_PAI** see 6.11.

**spectra** only available in expert mode. see 6.13.

**peptides\_pos** only available in expert mode. see 6.7.

**MS samples informations** see 6.8.

**Phosphopeptide mode** In this mode, only peptides containing one or more phosphorylation site are taken into account. These peptides are called phosphopeptides. Phosphopeptides sharing the same peptide sequence (possible overlaps) are represented by subgroups. A phosphoIsland is a set of phosphopeptides sharing the same peptide sequence, specific to a protein. Thus PhosphoIslands in the same subgroups share the same set of phosphopeptides. Subgroups are grouped together if they share at least one protein. In this mode, the ODS report contains seven tabs :

**X!TandemPipeline informations** see 6.3.



**phosphoislands** see 6.14.

**phosphopeptides** see 6.15.

**compar.Spectra** only available in expert mode. see 6.9.

**spectra** only available in expert mode. see 6.13.

**phospho stats** see 6.18.

**MS samples informations** see 6.8.

### 6.3 X!TandemPipeline informations table

This table contains general informations about the **X!TandemPipeline** process : version, filters, mode...

### 6.4 Proteins table (Fig 11)

**Group ID** The Group to which the protein belongs. All the proteins in a group have at least one peptide in common.

**Sub-Group ID** The Sub-Group to which the protein belongs. All the proteins in a sub-group are identified with the same valid peptides.

**Protein ID** A single reference to the protein in this grouping experiment (unique within a sample in individual mode).

**Description** Protein description as it appears in the header of the fasta file.

**log(E-Value)** Protein E-value expressed in log.

- Statistical value representing the number of times this protein would be identified randomly.
- Calculated as the product of unique peptide E-values in the sample.

**Coverage** % of protein coverage.

**MW** Molecular weight of the protein expressed in KDa.

**size** The number of amino acid of the protein sequence.

**Spectra** The total number of spectra assigned by the identification engine to this protein.

**Specific** The total number of spectra that are specific to this subgroup of proteins. It is only available if there are more than one subgroup within a group.

**Unique** The number of unique peptide sequence assigned to this protein.

**Specific uniques** The number of unique peptide sequence specific to this subgroup of proteins. It is only available if there are more than one subgroup within a group.

**Uniques peptide-mod-charge** The number of unique peptide assigned to this protein, counting separately sequences obtained with different modifications and charges.

**Theoretical number of tryptic peptides** The theoretical number of tryptic peptides in this protein<sup>7</sup>.

**PAI** Protein Abundance Index as described in 4.3.

**emPAI** the emPAI as described in 4.4.

**spectral count AI** Abundance index computed as the number of spectra assigned to the protein divided by the number of theoretical tryptic peptides.

**Sub-group proteins** The number of proteins in this subgroup.

**number of MS samples** The number of MS samples in which this protein was identified. Only usefull in the combined mode.

<sup>7</sup>Theoretical peptides correspond to the peptides resulting from the theoretical digestion of the protein sequence by trypsin and that are visible in mass spectrometry ( $800 < MH < 2500$ )

Figure 11: Protein results

## 6.5 Peptides table (Fig 12)

A peptide stands for a unique peptide sequence and theoretical  $MH^+$  (the charge state is not taken into account). This means that one peptide may be identified by several spectra in several samples (if running in combined mode). A unique peptide sequence may be listed several times as the  $MH^+$  differs (modified peptides). **WARNING** : isotope tags are not taken into account. This means that peptides tagged with isotopic modifications are merged : the expert mode is needed to have full details on all spectrum match. **WARNING** The sequenceLI column stands for a peptide sequence in which all Leucines are converted into Isoleucines. This is required because the L and I amino acids has exactly the same mass (see 6.19).

**Group ID** The Group ID to which the peptide belongs. A peptide only belongs to one Group.

**Peptide ID** A unique reference to this peptide (sequence +  $MH^+$ ).

**SequenceLI** The peptide sequence in which all L are converted to I (warning see 6.19)

**Modifs** Textual informations on modifications (if any)

**Charge** The charge state of the peptide identification featuring the best E-value among the same peptides (sequence +  $MH^+$ ).

**$MH^+$  theo** The theoretical  $MH^+$  for this peptide.

**Number of subgroups** The number of subgroups in which this peptide is present.

**Subgroup ids** Subgroup ID list in which this peptide is present.

**Number of spectra** The number of spectra (peptide spectrum match) identifying this peptide.

## 6.6 Spectra table (Fig 13)

The spectra tab contains details on each spectrum match. It is only available the expert ODS output was chosen.



MS Sample	Group ID	Peptide ID	Sequence	Modifs	Charge	MH+ Obs	MH+ Theo	Number of Subgroups	Number of spectra
2008_07_17_PAPPSO_325_1_03_1348									
1	al	pepta1	AAPDQDT		2	1612.84802	1	al	1
3	al	pepta2	ADIEWQIE		2	2096.04712	1	al	2
4	al	pepta3	ADIEWQIE	M5+15.99	2	2112.03711	1	al	2
5	al	pepta4	AEESIVY		2	1381.94905	1	al	1
6	al	pepta5	ETTEDVP		2	1099.97095	1	al	2
7	al	pepta6	GGGFSSG		2	1450.83501	1	al	1
8	al	pepta7	GGGFSSG		2	2090.81909	1	al	1
9	al	pepta8	GGGFSSG		3	2792.11304	1	al	1
10	al	pepta9	GGGFSSG		2	1099.97095	1	al	1
11	al	pepta10	GGGFSSG		2	1707.77295	1	al	1
12	al	pepta11	IDNNEQIS		2	1423.70703	1	al	1
13	al	pepta12	IKYENQAP		2	1234.67899	1	al	1
14	al	pepta13	IRENEQIT		2	1434.77039	1	al	1
15	al	pepta14	ISAEQIQ		2	1427.78601	1	al	1
16	al	pepta15	ITDQANIQ		2	1784.95105	1	al	1
17	al	pepta16	IKYENQAP		2	2267.26294	1	al	2
18	al	pepta17	INVAEIEQ		2	1797.01195	1	al	1
19	al	pepta18	INVTGQV		3	2872.39307	1	al	1
20	al	pepta19	INVTGQV	M26+15.99	3	2889.38306	1	al	1
21	al	pepta20	INVTGQV	M11+15.99	3	2894.37305	1	al	1
22	al	pepta21	OSIASIAE	M11-17.026	2	1373.65503	1	al	1
23	al	pepta22	OSIASIAE		2	1390.68103	1	al	1
24	al	pepta23	OSVEADIN		2	1201.81804	1	al	1
25	al	pepta24	OSVEADIN		2	1357.71875	1	al	1
26	al	pepta25	SGGYGGG		2	1432.66101	1	al	1
27	al	pepta26	SIIEGEOS		2	1262.59802	1	al	1
28	al	pepta27	SKETTED		2	2212.0879	1	al	1
29	al	pepta28	SOYEQAE		2	1565.64001	1	al	1
30	al	pepta29	SOYEQAE		2	1493.73499	1	al	2
31	al	pepta30	SSGGFSSG		2	1132.50195	1	al	1
32	al								

Figure 12: Peptide results

**Group ID** The Group ID to which the spectrum belongs.

**Peptide ID** The Peptide ID to which the spectrum belongs.

**Sample** The spectrum's sample name (the MS/MS run file name).

**Scan** The spectrum's scan number within this MS/MS run.

**Rt (minutes)** The retention time in minutes.

**Sequence (top)** The sequence of the best peptide (best E-value) assignment for this spectrum.

**Modifs (top)** The modifications (if any) of the best peptide (best E-value) assignment for this spectrum.

**Number of subgroups** The number of subgroups in which this spectrum is present.

**Sub-groups Ids** The subgroup id list in which this spectrum is present.

**Best E-value** The E-value of the best peptide assignment for this spectrum.

**Best hyperscore** The hyperscore (only available for **X!Tandem** analysis) of the best peptide assignment for this spectrum.

**m/z Obs** Observed m/z for the precursor of this MS/MS spectra.

**Charge** The charge state of the best peptide assignment for this spectrum.

**MH+ Obs** Observed  $MH^+$  (Monoisotopic observed mass for the peptide + one proton) of the best peptide assignment for this spectrum.

**MH+ theo** Theoretical  $MH^+$  of the best peptide assignment for this spectrum.

**DeltaMH+** Error in the precursor mass between observed and theoretical data (Da)

**Delta-ppm** Error in the precursor mass between observed and theoretical data (ppm)

## 6.7 Peptides position table (Fig 14)

List of all peptides (unique sequence +  $MH^+$ ) per identified protein, containing the position of this peptide in a protein. It is only available the expert ODS output was chosen.

**Peptide ID** The Peptide ID to which the spectrum belongs.

The screenshot shows a spreadsheet titled 'pappso\_indiv\_expert.ods - LibreOffice Calc'. The data is organized into columns: A (MS Sample), B (Peptide ID), C (Sample), D (Scan), E (RT), F (Sequence (Modifs (top))), G (Number of Subgroups), H (Best Eval), I (Best Hypers), J (Charge), K (MH+ Obs), L (MH+ theo), M (DeltaMH+), N (Data ppm). The data is organized into rows for various peptides and their modifications.

Figure 13: Spectra results

**Protein ID** The protein ID in this grouping experiment (unique within a sample in individual mode).

**accession** The protein accession (database reference).

**description** The protein description.

**Sequence** The peptide sequence.

**Modifs** Textual informations on modifications (if any).

**Start** The start position of this peptide on this protein.

**Stop** The stop position of this peptide on this protein.

**MH+ theo** Theoretical  $MH^+$  of this peptide.

The screenshot shows a spreadsheet titled 'pappso\_indiv\_expert.ods - LibreOffice Calc'. The data is organized into columns: A (MS Sample), B (Peptide ID), C (Protein ID), D (accession), E (description), F (Sequence), G (Modifs), H (Start), I (Stop), J (MH+ theo). The data is organized into rows for various peptides and their modifications.

Figure 14: peptides\_pos results

## 6.8 MS samples table

This table contains the list of MS samples considered in this experiment. The number of groups and subgroups are the same accross samples in the combined and phosphopeptide mode.



**MS sample name** MS sample name.

**Number of groups** The number of groups found by **X!TandemPipeline** concerning this sample.

**Number of subgroups** The number of groups found by **X!TandemPipeline** concerning this sample.

More informations about samples are available if the option “include raw MS samples statistics” was checked (Fig 10). If **X!Tandem** was used as the identification engine, this will display :

**X!tandem xml result file name** **X!Tandem** result file name.

**X!Tandem version** **X!Tandem** version.

**X!tandem xml model file name** **X!Tandem** parameter file.

**total spectra used** Number of used spectra as found in **X!Tandem** result file.

**total spectra assigned** Number of assigned spectra as found in **X!Tandem** result file.

**assigned/used percent** Percent of assigned spectra vs used.

**total unique assigned** See **X!Tandem** documentation.

**database file names** Space separated list of sequence databases used for this identification in **X!Tandem**.

**MS/MS data source file name** MS raw data file name (if available).

**MS/MS data source file path** MS raw data file complete file path (if available).

**MS level 1** Number of spectrums acquired in MS level 1.

**MS level 2** Number of spectrums acquired in MS level 2.

**TIC mean in MS 1** Mean of total ion count in MS level 1.

**TIC mean in MS 2** Mean of total ion count in MS level 2.

**rt min** Minimum retention time.

**rt max** Maximum retention time.

## 6.9 Compar spectra table (Fig 15)

This tab displays the number of spectra observed by sample, for each top protein of a subgroup. This is convenient to have an idea of the presence/absence of a subgroup of proteins in a sample, and an indication about abundance in each samples.

**Group ID** The Group ID to which the spectrum belongs.

**Sub-Group ID** The Sub-Group to which the protein belongs. All the proteins in a sub-group are identified with the same valid peptides.

**Top Protein Description** Description of the first protein listed in this subgroup

**Number of proteins** Total number of proteins listed in this subgroup

**sample names columns** one column for each sample containing the number of assigned spectra per subgroup

Group ID	Sub-group	Top Protein Description	Number of proteins	2008_07_17_PAPPSO_325_1_03_1348	2008_07_17_PAPPSO_325_1_04_1610
1	a1.a1	P13945(K1C10_HUMAN Keratin, type I cytoskeletal	1	28	28
2	a2.a1	Q86Y29(QORN_HUMAN Ig-capsin - Homo sapien	1	7	0
3	a3.a1	Q5D8Q2(QD862_HUMAN Fasciculin (Faginr	1	3	2
4	a4.a1	Q02413(DSG1_HUMAN Desmoglein-1 precursor	1	4	0
5	a5.a1	Q3SY94(QSY94_HUMAN Keratin-71 - Homo sa	1	2	2
6	a6.a1	Q1RTT2(QRTT2_HUMAN Keratin-78 (Kerat	2	2	2
7	a7.a1	P06702(S10A9_HUMAN Protein S100-A9 (S100	1	1	3
8	a8.a1	P15934(DESP_HUMAN Desmoplakin (DP) (250	1	3	0
9	a9.a1	P14932(PLAK_HUMAN Junction plakoglobin (Des	2	2	0
10	a10.a1	P47920(LEGG_HUMAN Galectin-7 (Gae-7) (MWL	1	2	0
11	b10.a1	P29508(SPB3_HUMAN Sepsin B3 (Squamous c	2	0	2
12	b11.a1	P60709(ACTB_HUMAN Actin, cytoplasmic 1 (Be	10	1	0
13	b12.a1	P05109(S10A9_HUMAN Protein S100-A9 (S100	1	0	1
14	b13.a1	P10598(THO_HUMAN Thoredoxin (Tho) (ATL-d	2	0	1
15	b14.a1	P04409(G3P_HUMAN Glyceraldehyde-3-phosph	2	0	1
16	b15.a1	P12273(PPI_HUMAN Prolactin-inducible protein	1	1	0

Figure 15: compar.Spectra results

## 6.10 Compar specific table (Fig 16)

This tab displays the number of specific spectra observed by sample, for each top protein of a subgroup. A spectra is considered as specific if it is only present in one subgroup and no other subgroups of the same group. If there is only one subgroup in a group, all the spectra of this protein are considered specific.

**Group ID** The Group ID to which the spectrum belongs.

**Sub-Group ID** The Sub-Group to which the protein belongs. All the proteins in a sub-group are identified with the same valid peptides.

**Top Protein Description** Description of the first protein listed in this subgroup

**Number of proteins** Total number of proteins listed in this subgroup

**sample names columns** one column for each sample containing the number of assigned specific spectra per subgroup

## 6.11 Compar unique table

This tab displays the number of unique sequences observed by sample, for each top protein of a subgroup.

**Group ID** The Group ID to which the spectrum belongs.

**Sub-Group ID** The Sub-Group to which the protein belongs. All the proteins in a sub-group are identified with the same valid peptides.

**Top Protein Description** Description of the first protein listed in this subgroup

**Number of proteins** Total number of proteins listed in this subgroup

**sample names columns** one column for each sample containing the number of unique peptide sequence per subgroup

Group ID	Sub-group	Top Protein	Top Protein Description	Number of proteins	2008_07_17_PAPPSO_325_1_03_1348	2008_07_17_PAPPSO_325_1_04_1610
1	a1	a1	P13945K1C10_HUMAN Keratin	1	28	28
2	a2	a2	P08923HORN_HUMAN Keratin	1	7	0
3	a3	a3	Q5D8Q2QSD8Q2_HUMAN Flag-tag	1	3	0
4	a4	a4	Q02413DSG1_HUMAN Desmoglein	1	4	0
5	a5	a5	Q3SY94Q3SY94_HUMAN Keratin	1	1	1
6	a6	a6	Q1RTT2Q1RTT2_HUMAN Keratin	2	1	0
7	a7	a7	P06702S10A8_HUMAN Protein S	1	1	3
8	a8	a8	P15924DESP_HUMAN Desmoglein	1	3	0
9	a9	a9	P14923PLAK_HUMAN Junction	2	2	0
10	a10	a10	P47938LEST_HUMAN Lectin	1	2	0
11	b1	b1	P29508SPB3_HUMAN Serpin B3	2	0	2
12	b2	b2	P60709ACTB_HUMAN Actin	10	1	0
13	b3	b3	P05109S10A8_HUMAN Protein S	1	0	1
14	b4	b4	P10598THO_HUMAN Thrombospondin	2	0	1
15	b5	b5	P04409G3P_HUMAN Glyceraldehyde	2	0	1
16	b6	b6	P12273PFP_HUMAN Prolectin	1	1	0

Figure 16: compar\_Specific results

## 6.12 Compar specific unique table

This tab displays the number of unique sequences specific within a subgroup, observed by sample, for each top protein of a subgroup.

**Group ID** The Group ID to which the spectrum belongs.

**Sub-Group ID** The Sub-Group to which the protein belongs. All the proteins in a sub-group are identified with the same valid peptides.

**Top Protein Description** Description of the first protein listed in this subgroup

**Number of proteins** Total number of proteins listed in this subgroup

**sample names columns** one column for each sample containing the number of unique peptide sequence and specific per subgroup

## 6.13 Compar PAI table (Fig 17)

This tab displays the protein abundance index (PAI4.3) and the exponentially modified PAI (emPAI4.4) by sample, for each top protein of a subgroup. This is convenient to have an idea of the presence/absence of a subgroup of proteins in a sample, and an indication about relative abundance in each samples.

**Group ID** Group ID of this protein

**Sub-Group ID** Sub-Group of this protein

**Top Protein Description** Description of the first protein listed in this subgroup

**Number of proteins** Total number of proteins listed in this subgroup

**PAI** PAI4.3 of the top protein of this subgroup in a particular sample

**emPAI** emPAI4.4 of the top protein of this subgroup in a particular sample



Group ID	Sub-group ID	Top Protein ID	Top Protein Description	Number of proteins	PAI	smPAI
a1	a1.a1	P13659(C10_HUMAN_Kar)	1	1.2727272727	17.7381742286	1.2727272727
a2	a2.a1	Q86VZ3(HORN_HUMAN_Har)	1	0.1111111111	0.281548655	
a3	a3.a1	Q5D862(Q5D862_HUMAN_Ep)	1	0.0483870988	0.1178581778	0.0322580645
a4	a4.a1	Q02419(DSG1_HUMAN_Destr)	1	0.1176470588	0.3111338374	0.07710506
a5	a5.a1	Q3S194(Q3S194_HUMAN_Ka)	1	0.0384615385	0.0230036611	0.0789230769
a6	a6.a1	Q7RTT2(Q7RTT2_HUMAN_Ka)	2	0.0416666667	0.1006941713	0.0416666667
a7	a7.a1	P06702(S10A8_HUMAN_Prote)	1	0.1428571429	0.3894954944	0.4285714286
a8	a8.a1	P15924(DESFP_HUMAN_Destr)	1	0.0201342282	0.047452236	1.8826958
a9	a9.a1	P14829(PLA2_HUMAN_Junc)	2	0.0571428571	0.1485248239	
a10	a10.a1	P47929(LEG_HUMAN_Galec)	1	0.25	0.77827941	
b1	b1.a1	P28508(SPB3_HUMAN_Sepi)	4	0.0666666667	0.0909090909	0.23284674
b2	b2.a1	P00709(AC7B_HUMAN_Acib)	10	0.1659144012		
b3	b3.a1	P05108(S10A8_HUMAN_Prote)	1		0.1428571429	0.38949549
b4	b4.a1	P10599(THO_HUMAN_Thore)	2		0.1666666667	0.46779927
b5	b5.a1	P04406(G3P_HUMAN_Glyce)	2		0.0909090909	0.23284674
b6	b6.a1	P12279(PF_HUMAN_Prosac)	1	0.1111111111	0.281548655	

Figure 17: compar\_PAI results

## 6.14 PhosphoIsland table (Fig 18)

This tab displays the phosphoisland list.

**Group ID** Group ID of this phosphoisland : a group contains all proteins sharing one or more phosphoisland

**Sub-Group ID** Sub-Group of this phosphoisland : a subgroup contains all proteins sharing the same phosphopeptides

**PhosphoIsland ID** PhosphoIsland identifier : a set of phosphopeptides covering the same peptide sequence (possible overlaps) in a protein

**Description** The description of the phosphoisland's protein

**MW** The molecular weight of the phosphoisland's protein

**Phosphosites positions** the phosphorylation sites positions of the phosphoisland's phosphopeptides in this protein

**Spectra** Number of spectra identifying the phosphopeptides of this phosphoisland

**Uniques** Number of unique phosphopeptide sequences

**number of proteins sharing these phosphosites** Number of proteins in this subgroup (sharing the same phosphopeptides)

## 6.15 PhosphoPeptides table (Fig 19)

This tab displays details about each identified phosphopeptides. A phosphopeptide is a merge of identified peptides including phosphorylations and sharing the same unique sequenceLI +  $MH^+$  mass modification. **WARNING** The sequenceLI stands for a peptide sequence in which all Leucines are converted into Isoleucines. This is required because the L and I amino acids has exactly the same mass (see 6.19).

**Group ID** Group ID of this phosphopeptide

**Sub-Group ID** Sub-Group of this phosphopeptide (it regroups overlapping peptide sequences)



Group ID	Sub-group ID	Phosphoisland ID	Description	MW	Phosphosites positions	Spectra	Uniques	number of proteins sharing these phosphosites
2	a1	a1 a1	AT1G01100 11.1000004	102		435	3	6
3	a1	a1 a1	AT1G01100 11.1000004	102		435	3	6
4	a1	a1 a1	AT1G01100 9.8999996	96		435	3	6
5	a1	a1 a1	AT1G01100 11.1000004	102		435	3	6
6	a1	a1 a1	AT3G47700 11.1999998	103		435	3	6
7	a1	a1 a1	AT3G47700 11.1999998	103		435	3	6
8	a2	a2 a1	AT4G00810 11.1999998	103		191	3	2
9	a2	a2 a1	AT4G00810 11.1999998	103		191	3	2
10	a3	a3 a1	AT4G25890 11.8000002	106		151	3	3
11	a3	a3 a1	AT5G57290 11.8000002	107		151	3	3
12	a3	a3 a1	AT5G57290 11.8999998	106		151	3	3
13	a4	a4 a1	AT2G27710 11.3999999	105		102	3	9
14	a4	a4 a1	AT2G27710 11.3999999	105		102	3	9
15	a4	a4 a1	AT2G27710 11.3999999	105		102	3	9
16	a4	a4 a1	AT2G27710 10.9999998	88		102	3	9
17	a4	a4 a1	AT2G27720 11.3999999	105		102	3	9
18	a4	a4 a1	AT2G27720 12.9999999	120		102	3	9
19	a4	a4 a1	AT2G27720 12.6000004	117		102	3	9
20	a4	a4 a1	AT3G44590 10.8999999	101		102	3	9
21	a4	a4 a1	AT3G44590 10.8999999	101		102	3	9
22	a5	a5 a1	AT4G02510 160.600006	630 632		20	1	1
23	a5	a5 a2	AT4G02510 160.600006	589		16	2	1
24	a5	a5 a3	AT4G02510 160.600006	71 73		5	2	1
25	a5	a5 a4	AT4G02510 160.600006	695		5	1	1
26	a6	a6 a1	AT4G05150 52.8000015	292 305 310 311		21	1	1
27	a6	a6 a2	AT4G05150 52.8000015	266 270		7	1	1
28	a6	a6 a3	AT4G05150 52.8000015	277 281 282 284		5	1	1
29	a7	a7 a1	AT3G05420 72.8000015	515 520		20	1	2
30	a7	a7 a1	AT3G05420 73.0999985	516 521		20	1	2
31	a7	a7 a2	AT3G05420 72.8000015	501 503		11	1	2
32	a7	a7 a2	AT3G05420 73.0999985	502 504		11	1	2

Figure 18: phosphoislands results

**Phosphopeptide ID** Phosphopeptide identifier**SequenceLI** Phosphopeptide sequence (warning see 6.19)**Modifs** Modification mass**Number of phosphorylations (phosphosites)** number of phosphorylation sites in this peptide sequence (observed by the identification engine)**Best Position in sequence** The phosphosites position that gives the best peptide E-value**All positions** List of all possible phosphosite positions (observed by the identification engine)**Charges** List of all possible peptides charges (observed by the identification engine)**MH+ theo** Theoretical  $MH^+$  of this phosphopeptide.**Number of proteins** Number of protein identified with this phosphopeptide.**Number of spectra** Number of spectra identifying this phosphopeptide.

Group ID	Sub-group ID	Phosphopeptide ID	Sequence	Modifs	Number of phosphosites	Best Position in sequence	All positions	Charges	MH+ theo	Number of proteins	Number of spectra
2	a1	pepa1a1	DEPPEESDDGGFGF	79.9663315	17	2	7	2	1892.71143	6	6
3	a1	pepa1a2	KDEPAEESDDGGFGF	79.9663315	18	8	2	2020.80627	6	182	
4	a1	pepa1a3	KVDEPAEESDDGGFGF	79.9663315	19	9	2 3	2149.90137	6	247	
5	a2	pepa2a1	DEPPEESDDGGFGF	79.9663315	17	7	2	1892.71143	2	6	
6	a2	pepa2a2	EPHDEPPEESDDGGFGF	79.9663315	111	11	2 3	2406.86957	2	3	
7	a2	pepa2a3	KDEPAEESDDGGFGF	79.9663315	18	8	2	2020.80627	2	182	
8	a3	pepa3a1	EESEEEEGGFGFGF	79.9663315	13	3	2	1915.67969	3	1	
9	a3	pepa3a2	KEESEEEEGGFGFGF	79.9663315	14	4	2	2043.71466	3	70	
10	a3	pepa3a3	KKESEEEEGGFGFGF	79.9663315	15	5	2 3	2171.86963	3	80	
11	a4	pepa4a1	EEKEESDDMGFSIF	79.9663315	16	6	2	1986.72021	9	30	
12	a4	pepa4a2	EEKEESDDMGFSIF	95.9612274	16	6	2	2002.71497	9	5	
13	a4	pepa4a3	EEESDDMGFSIF	79.9663315	13	3	2	1609.54094	9	1	
14	a4	pepa4a4	EEESDDMGFSIF	95.9612274	13	3	2	1616.53503	9	7	
15	a4	pepa4a5	KEEKEESDDMGFSIF	79.9663315	17	7	2 3	2114.81519	9	55	
16	a4	pepa4a6	KEEKEESDDMGFSIF	95.9612274	17	7	2	2130.81096	9	4	
17	a5	pepa5a1	ASSGEAHSDEENK	79.9663315	18	9	3	2284.81406	1	3	
18	a5	pepa5a2	EIDSSSEA/SGNSK	79.9663315	124	22 24	3	2721.12622	1	3	
19	a5	pepa5a3	KVGEESAEEDENK	79.9663315	17	7	3	2724.26147	1	8	
20	a5	pepa5a4	VDSSEETETEMF	159.832693	24 6	4 6	2 3	2520.92529	1	18	
21	a5	pepa5a5	VDSSEETETEMF	175.827551	24 6	4 6	2	2536.92017	1	2	
22	a5	pepa5a6	VGADDSSEKFN	79.9663315	17	7 9	3	2253.02856	1	2	
23	a5	pepa5a7	VVEGDSAEEDENK	79.9663315	16	6	3	2596.1665	1	8	
24	a5	pepa5a1	ASSGEAHSDEENK	79.9663315	18	9	3	2284.81406	1	3	
25	a5	pepa5a2	EIDSSSEA/SGNSK	79.9663315	124	22 24	3	2721.12622	1	3	
26	a5	pepa5a3	KVGEESAEEDENK	79.9663315	17	7	3	2724.26147	1	8	
27	a5	pepa5a4	VDSSEETETEMF	159.832693	24 6	4 6	2 3	2520.92529	1	18	
28	a5	pepa5a5	VDSSEETETEMF	175.827551	24 6	4 6	2	2536.92017	1	2	
29	a5	pepa5a6	VGADDSSEKFN	79.9663315	17	7 9	3	2253.02856	1	2	
30	a5	pepa5a7	VVEGDSAEEDENK	79.9663315	16	6	3	2596.1665	1	8	
31	a5	pepa5a1	ASSGEAHSDEENK	79.9663315	18	9	3	2284.81406	1	3	
32	a5	pepa5a2	EIDSSSEA/SGNSK	79.9663315	124	22 24	3	2721.12622	1	3	

Figure 19: phosphopeptides results

## 6.16 Phosphopeptide mode compar Spectra table (Fig 20)

This tab displays the number of spectra identifying phosphoislands by sample. This is convenient to have an idea of the phosphorylation rate of phosphoislands between samples.

**Group ID** Group ID of this phosphoisland

**Sub-Group ID** Sub-Group of this phosphoisland

**Top PhosphoIsland ID** PhosphoIsland identifier : the first in this subgroup

**Top PhosphoIsland protein** PhosphoIsland's protein description

**Positions** PhosphoIsland's phosphorylation site positions in this protein

**Number of proteins** number of proteins in this subgroup

**sample names columns** one column for each sample containing the number of assigned spectra per phosphoisland

A	B	C	D	E	F	G	H	I	J	K	L	M
Group ID	Sub-group ID	Top PhosphoIsland ID	Top PhosphoIsland protein	Positions	Number of p	20110309_L1	20110309_L2	20110309_L3	20110309_L4	pappso_nostress_nolab	pappso_nostress_nolab	pappso_nostress_nolab
1	a1	a1	AT1G01100.1   Symbols	1605 acids*102	6	114	103	108	110			
2	a2	a2	AT4G00810.1   Symbols	1605 acids*103	2	47	43	52	49			
3	a3	a3	AT4G02680.1   Symbols	1605 acids*106	3	33	30	36	32			
4	a4	a4	AT2G27710.1   Symbols	1605 acids*105	9	29	25	24	24			
5	a5	a5	AT4G02510.1   Symbols	TOC159, T*589	1	4	4	4	4			
6	a6	a6	AT4G02510.1   Symbols	TOC159, T*589	1	0	1	0	2			
7	a7	a7	AT4G02510.1   Symbols	TOC159, T*71 73	1	2	1	2	0			
8	a8	a8	AT4G02510.1   Symbols	TOC159, T*630 632	1	6	4	4	6			
9	a9	a9	AT4G02510.1   Symbols	TOC159, T*630 632	1	1	3	2	1			
10	a10	a10	AT4G02510.1   Symbols	TOC159, T*630 632	1	7	5	4	5			
11	a11	a11	AT4G02510.1   Symbols	TOC159, T*630 632	1	2	2	1	1			
12	a12	a12	AT4G02510.1   Symbols	TOC159, T*630 632	2	5	4	6	5			
13	a13	a13	AT4G02510.1   Symbols	TOC159, T*630 632	2	3	3	2	2			
14	a14	a14	AT4G02510.1   Symbols	TOC159, T*630 632	1	8	7	8	6			
15	a15	a15	AT4G02510.1   Symbols	TOC159, T*630 632	3	6	6	5	9			
16	a16	a16	AT4G02510.1   Symbols	TOC159, T*630 632	2	4	6	6	8			
17	a17	a17	AT4G02510.1   Symbols	TOC159, T*630 632	1	4	2	4	2			
18	a18	a18	AT4G02510.1   Symbols	TOC159, T*630 632	1	0	3	0	0			
19	a19	a19	AT4G02510.1   Symbols	TOC159, T*630 632	1	1	1	1	2			
20	a20	a20	AT4G02510.1   Symbols	TOC159, T*630 632	1	1	2	0	0			
21	a21	a21	AT4G02510.1   Symbols	TOC159, T*630 632	1	1	2	0	0			
22	a22	a22	AT4G02510.1   Symbols	TOC159, T*630 632	1	1	2	1	2			
23	a23	a23	AT4G02510.1   Symbols	TOC159, T*630 632	1	1	2	1	2			
24	a24	a24	AT4G02510.1   Symbols	TOC159, T*630 632	1	3	4	4	4			
25	a25	a25	AT4G02510.1   Symbols	TOC159, T*630 632	6	8	4	5	3			
26	a26	a26	AT4G02510.1   Symbols	TOC159, T*630 632	1	6	4	3	5			
27	a27	a27	AT4G02510.1   Symbols	TOC159, T*630 632	2	4	5	4	4			
28	a28	a28	AT4G02510.1   Symbols	TOC159, T*630 632	1	2	2	2	2			
29	a29	a29	AT4G02510.1   Symbols	TOC159, T*630 632	1	3	2	0	2			
30	a30	a30	AT4G02510.1   Symbols	TOC159, T*630 632	1	0	1	1	0			
31	a31	a31	AT4G02510.1   Symbols	TOC159, T*630 632	2	2	2	2	0			
32	a32	a32	AT4G02510.1   Symbols	TOC159, T*630 632	2	3	2	3	2			

Figure 20: compar.Spectra results (mode phospho)

## 6.17 Phosphopeptide mode Spectra table (Fig 21)

This tab displays details about each spectra (scan number in a particular sample) identifying a phosphorylated peptide. Only available if the expert output was chosen.

**Group ID** Group ID of this spectra

**Sub-Group ID** Sub-Group of this spectra

**Phosphopeptide ID** The phosphopeptide identified by the spectra

**Sample** The sample name of this spectra

**Scan** The scan number of this spectra

**Rt** The retention time (in minutes) of this spectra

**Sequence (top)** The sequence of the best peptide assignment for this spectra

**Modifs (top)** The modifications of the best peptide assignment for this spectra

**Best position in peptide** The phosphorylation site positions of the best peptide assignment for this spectra

**All observed positions in phosphopeptide** All observed positions of phosphorylation sites in the same phosphopeptide

**Number of phosphoislands** number of phosphoislands (proteins) identified by this spectra

**Phosphoislands Ids** list of phosphoislands ids identified by this spectra

**Best e-value** the best peptide E-value assignment for this spectra

**Charge** the charge of the best peptide assignment for this spectra

**MH+ Obs** Observed  $MH^+$  (Monoisotopic observed mass for the peptide + one proton) of the best peptide assignment for this spectrum.

**MH+ theo** Theoretical  $MH^+$  of the best peptide assignment for this spectrum.

**DeltaMH+** Error in the precursor mass between observed and theoretical data (Da)

**Delta-ppm** Error in the precursor mass between observed and theoretical data (ppm)

Figure 21: Spectra results (mode phospho)

## 6.18 Phospho stats table

This table displays miscellaneous statistics on phosphosites.

Total number of sample/scan in the whole experiment containing phosphorylation sites.

Total number of phosphorylations per residues on Threonine, Serine, and Tyrosine.

## 6.19 SequenceLI column

The sequenceLI stands for a peptide sequence in which all Leucines are converted into Isoleucines. This is required because the L and I amino acids has exactly the same mass. Identification engines can thus link the

same MS/MS spectra to 2 proteins having different peptide sequences differing only by L or I. As a consequence the grouping algorithm has to consider these sequences as unique and we have chosen to only show in this tab a sequenceLI in which there is an ambiguity between L or I. To get the *real* peptide sequence, please use in the expert mode the *peptide\_pos* tab (see 6.7). It shows all matching peptides associated to their proteins (this could produce a large output file). The spectra output (see 6.6) only shows the best matching peptide sequence (best evalue).

## 6.20 Files \*fdr.txt

This result file indicates the number of peptides with an E-value less than the E-value indicated in the first column (Fig 22). You just have to divide the number of peptides in the reverse or decoy database by the number of peptides in the normal database to obtain the false discovery rate at each E-value level.

This method could be performed if :

- normal and reverse databases must be saved in different fasta files;
- **X!Tandem** analysis have been performed with reverse option.  
In this case, the column corresponding to the normal and reverse search are indicated as *xtandem normal* and *xtandem reverse*, respectively.

FDR on peptide identification		
Evalue	Normal.fasta	Reverse.fasta
-14.5	0	0
-14	0	0
-13.5	1	0
-13	1	0
-12.5	1	0
-12	3	0
-11.5	3	0
-11	4	0
-10.5	4	0
-10	6	0
-9.5	6	0
-9	7	1
-8.5	7	1

Figure 22: FDR results

## 7 Changelog

### 7.1 "Elastine" branch

#### 3.4.0 Elastine Durcie

### 7.2 "Myosine" branch

#### 3.3.5 Myosine Dopée

**3.3.4 Myosine Bodybuildée** Experimental support for the pepXML format used by the TPP (Trans Proteomic Pipeline) and the Comet MS identification engine. More details available in the "**X!TandemPipeline**information" of the ODS report for each MSrun : number of assigned spectrums, number of MS1, MSn. Lot of internal structure design (FDR outputs using ODS writer, FASTA writer). BUG FIX (introduced in 3.3.1) in the tabulated output for combine and individual mode (ODS or TSV report) : "peptides" and "peptide pos" sheets now only displays valid peptides. BUG FIX in the MassChroQML output (introduced in 3.3.1) : only writes valid peptides to the list of identified peptides.

**3.3.3 Myosine Anabolisée** New XML parser giving better informations to the users (line number, column number, filename containing errors). Better support for high performance clusters through HTCondor. BUG FIX in the "peptide pos" tab of the spreadsheet output in expert mode : only the first protein in each subgroup was displayed.

**3.3.2 Myosine Chargée** Better performances. New ODS output featuring complete summary of MS identification runs. Updated documentation.

**3.3.1 Myosine Famélique** Better performances. Problems concerning the PAI computation by samples are fixed. Updated documentation.

**3.3.0 Myosine Rachitique** Grouping of sub-group has been changed for better performances and to fix overgrouping on large datasets (thanks to M. Blein)  
If you have a very large dataset, we recommend to reload xtandem results to fix errors.

### 7.3 "Tubuline" branch

**3.2.2 Tubuline Étayée** Corrected report of input parameter on X!Tandem output result (thanks to T. Greko).

**3.2.1 Tubuline Étayée** Add new X!Tandem parameters for multiple search of modifications in one analyse and calculation can now be performed on  $z \geq 3$ .

**3.2.0 Tubuline Squelettique** Identification from Mascot dat file can now be imported and filtered. All work as X!Tandem result excepts that protein sequence can not be retrieved : PAI and coverage are absent. Correction of FDR calculation from Reverse/Decoy search.

## 7.4 "Kératine" branch

**3.1.5 Kératine Moustachue** Add support of phosphorylation neutral loss and enhanced ETD detection on MS2 spectra.

Correction of MassChroQ export.

**3.1.4 Kératine Poilue** Add support for viewing ETD spectra after automatic detection.

**3.1.3 Kératine Chevelue** Corrected bug of xtandem preset. Refine analysis was never start instead refine param is set to yes.

Adds a new annotated spectrum renderer and bug fix on ODS export.

**3.1.2** Add export results on Open Document Spreadsheet (.ods) file.

Correction of bugs (Grouping, PepNovo export, ...).

**3.1.1** FDR computation are now compatible with reverse option of X!Tandem.

**3.1.0** Algorithm of grouping have been completely rewritten :

- Older project must be refiltered to be properly grouped.
- Phosphopeptide filtering have been enhanced to correspond to :
  - SubGroup represents the number of phosphosites
  - Group represents the number of phosphoproteins
- Configuration file have been modified and must be parameter again