

Tutorial for protein identification and quantification with MaxQuant software platform

Laboratory of Mass Spectrometry, LNBio, CNPEM

Tutorial version 1.0, September 2013

Note: This tutorial was written based on the information available in scientific papers, MaxQuant google groups, local group discussions and it includes our own experiences in the proteomics data analysis performed in our research group.

1 Release information

Tutorial version 1.0, September 2013.

Software versions discussed in the actual tutorial is:

MaxQuant version 1.3.0.5 de 2012.

Software updates and versions mentioned above can be obtained through the following website:

MaxQuant: <http://www.maxquant.org/>

Contents

1	Release information.....	2
	Contents	3
2	Max Quant	4
2.1	Getting started with MaxQuant.....	4
2.2	Installing MaxQuant software.....	4
2.3	Selecting search protein database	5
2.4	Executing the MaxQuant software.....	8
2.5	Uploading experimental RAW data files.....	9
2.6	Defining the experimental design file.....	10
2.7	Selecting protein identification parameters	12
2.7.1	Group-specific parameters.....	12
2.7.2	MS/MS & sequences	14
2.7.3	Identification and quantification.....	15
2.7.4	Miscellaneous	17
2.7.5	Saving parameter settings.....	18
2.7.6	Starting the protein identification process	18
2.7.7	Finishing the protein identification and getting your results.....	20

2 Max Quant

2.1 Getting started with MaxQuant

MaxQuant is a software package developed in the computing language C# (C sharp) for protein identification and quantification from large proteomics datasets obtained through high-resolution mass-spectrometry. It runs in Windows operational systems (Windows 7 or higher) and Windows Vista. It can also run with in Windows server 2008 or 2012. NET framework 4.5 must be installed in your computer before installing the MaxQuant software.

2.2 Installing MaxQuant software

Before starting with the MaxQuant software installation, it is essential to download another program which is intended to read the raw files generated by high-resolution mass spectrometers manufactured by Thermo Scientific.





1. Go to the Thermo Scientific web site (<http://sjsupport.thermofinnigan.com/public/detail.asp?id=703>)
2. Download the file “MSFileReader 2.2.62.zip”.
3. Decompress the file and install it in the same folder where you would like to have MaxQuant installed in your computer (example; “C: User>Programs>MaxQuant”).
4. Go to the web site of MaxQuant (www.maxquant.org)
5. In “Downloads” menu, click in the “**MaxQuant executables**”, which contains the most recent version or you can contact us directly for a copy of the versions of MaxQuant and Perseus software showed in this tutorial.
6. You will be asked to fill a registration form and you will receive by e-mail a code for downloading the MaxQuant software.
7. After downloading is complete, unzip the files in the new created folder named MaxQuant.

TIP 1

Do not open or execute the MaxQuant software now. Follow the section 3.3 if you did not configure the Andromeda search engine.

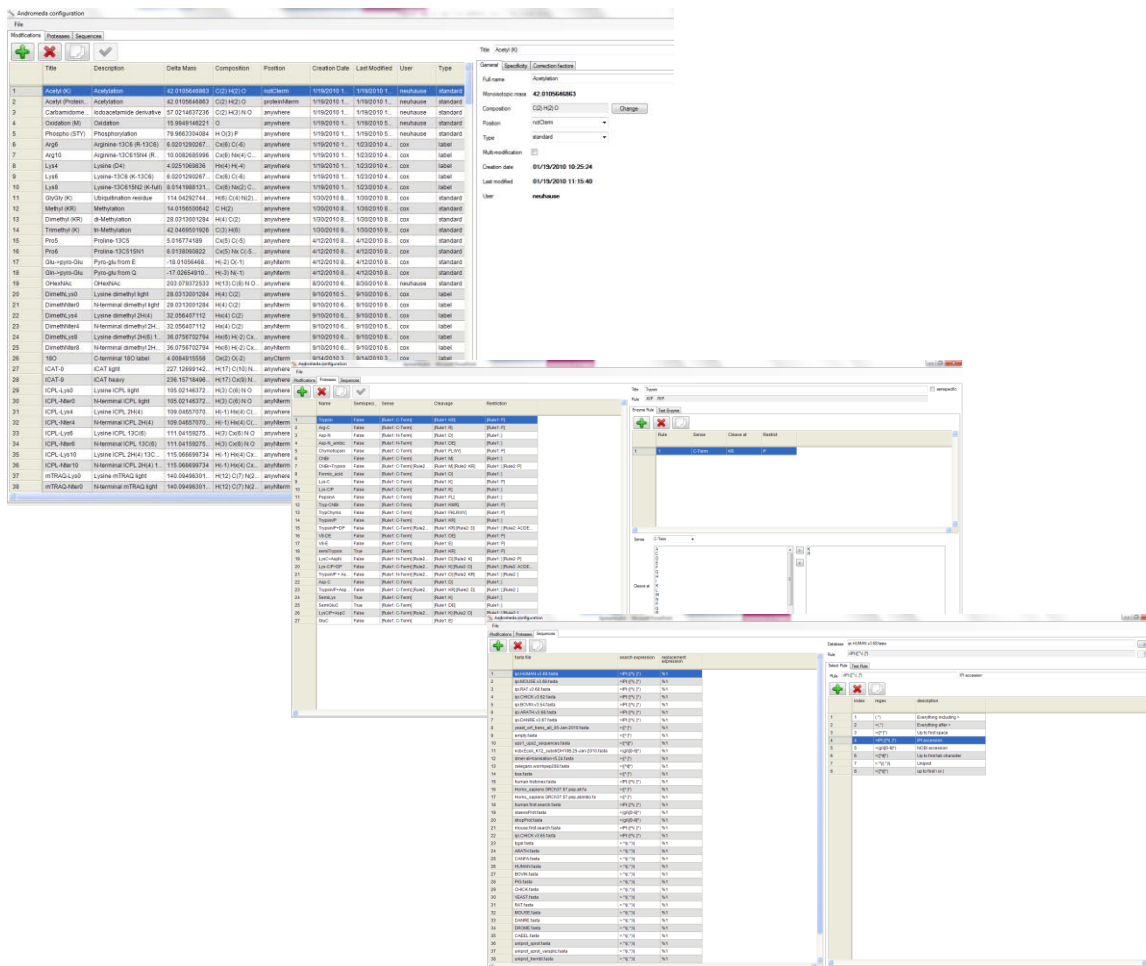
2.3 *Selecting search protein database*

The MaxQuant software package works coupled to the search engine Andromeda, which is provided together with MaxQuant. In the folder where the MaxQuant is installed, you will see other folder named “bin” and the icons of other two executable applications which are part of the MaxQuant package. One of them is the Andromeda, which is the search engine and the Viewer application which performs mass spec chromatogram viewing functions.

 bin	6/17/2013 2:15 PM
 AndromedaConfig.exe	6/17/2013 2:15 PM
 MaxQuant.exe	6/17/2013 2:15 PM
 Viewer.exe	6/17/2013 2:15 PM

Before starting the protein identification using MaxQuant, it is important that you configure the protein database which will be used in the protein search, which is processed by Andromeda engine.




In Andromeda, you will see three distinct options on the upper left part of the window as follow: Modifications, Proteases, Sequences.

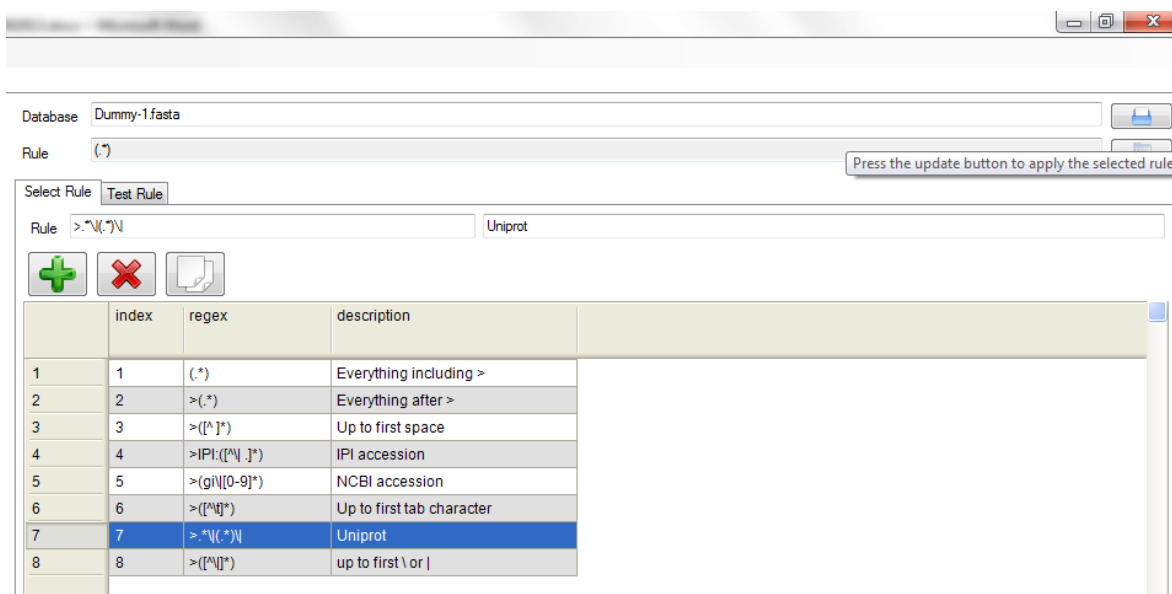


The Modifications option allows to modify or to add new protein modifications you may want to investigate and to integrate as new protein identification parameter. The Proteases option also allows including new enzymes in the default list of possible protein cleavages events which may be considered during the protein identification processing. For the protein identification to be performed you have to be sure that a protein database which contains all annotated protein sequences from the organism you are studying are included in the list of databases registered by default in Andromeda. If the database of the organism

you are studying is not there or if you want to use another protein database, you can include a proteome database into the list of Andromeda protein databases.

Let's do that?

1. Double click on the Andromeda executable file (AndromedaConfig.exe).
2. Select the **Sequences** option.
3. Click on the icon . A phantasy name "**Dummy-1.fasta**" should appear in the right side of the panel in a previous blank cell.
4. Substitute the name "Dummy-1.fasta" by selecting your FASTA protein database file of interest using the symbol . **Choose the FASTA file and click Open.**
5. Click on **Select Rule** menu and choose one of the rules for the protein identifiers which correspond to the FASTA file you have uploaded.
6. Press the button  for updating your selection.
7. Go to **File > Save > All** and click in Yes when a message to update all systems appears.
8. Close Andromeda.

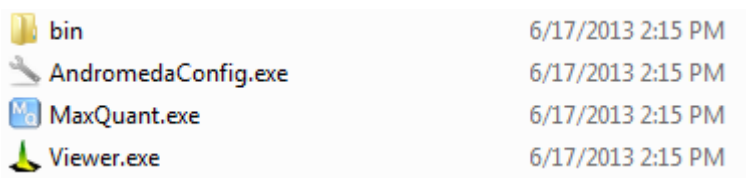


TIP 2

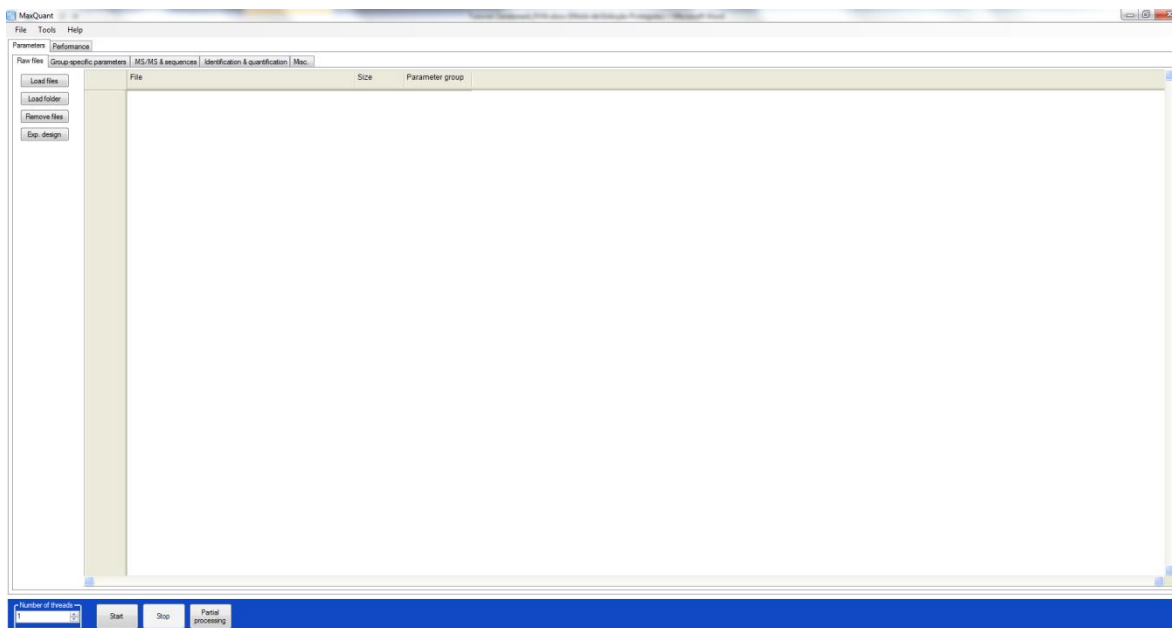
The protein sequence identifiers (IDs) usually observed in the FASTA files correspond to a unique code for each protein sequence you have in the protein database. Different institutions may include different protein IDs to the protein sequences.

2.4 Executing the MaxQuant software

In the folder where you saved the MaxQuant software, **double click in the MaxQuant.exe** executable file.



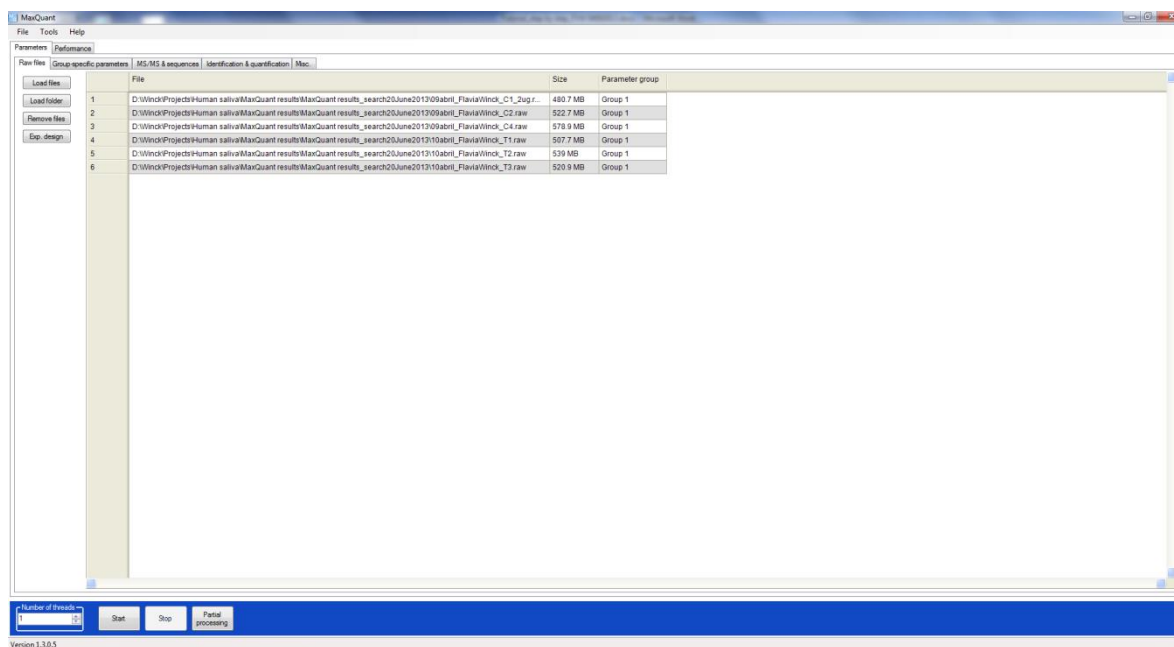
A graphic user interface should open.



2.5 Uploading experimental RAW data files

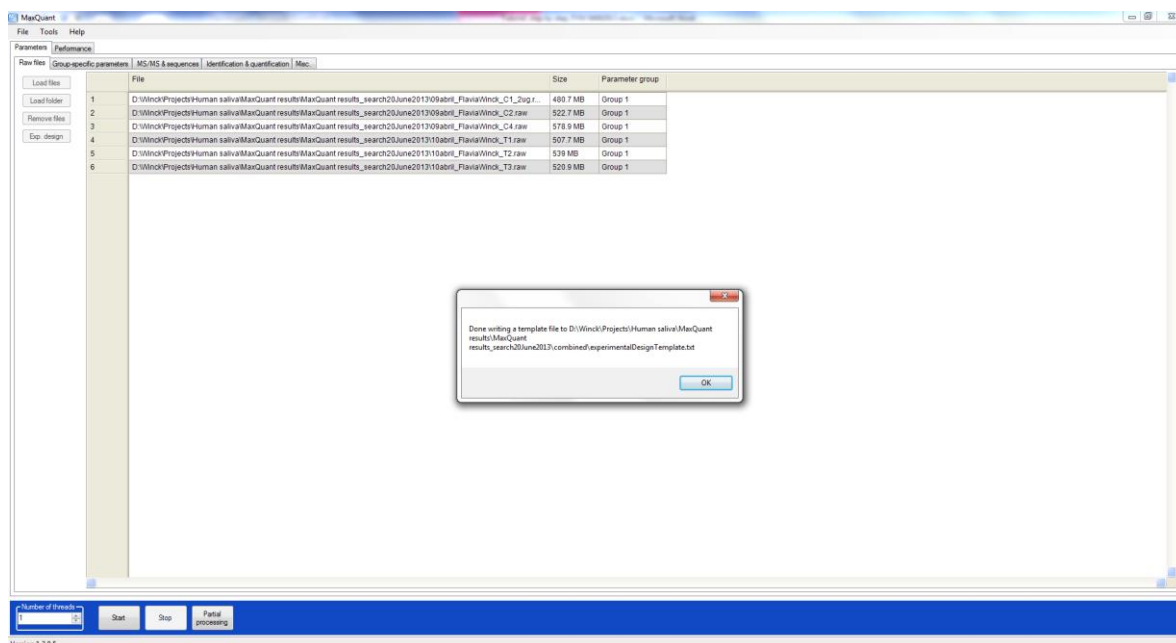
1. In the left side of the GUI, click in **Load Files** for loading individual files or **Load folder** if you want to upload many files located in the same folder location.
2. **Browse the raw files** you want to analyze and press **Open**.

Your files should appear as follow:



2.6 Defining the experimental design file

1. Click on the **Exp. Design** button on the left side menu. A message indicating the file location should be displayed. Memorize the route where the file is located and click **OK**.



- Go to the file location shown in the previous step and open the file named “experimentalDesignTemplate.txt” using Excel. In the Fraction column you must assign if the corresponding files shown in the left column belong to a fraction of a gel fraction. If your data is not obtained through gel-based pre-fractionation you must assign the same number for all files in the column Fraction.

	A	B	C
1	Name	Fraction	Experiment
2	09abril_FlaviaWinck_C1	1	
3	09abril_FlaviaWinck_C2	1	
4	09abril_FlaviaWinck_C4	1	
5	10abril_FlaviaWinck_T1	1	
6	10abril_FlaviaWinck_T2	1	
7	10abril_FlaviaWinck_T3	1	
8			
9			

- In the column named as Experiment if you want to combine all experimental replicates as a single dataset to be analyzed by MaxQuant, you must enter the same identifier for the files which should be concatenated as shown below. However, if

you want each individual file to be treated as a different experiment which you want to compare further, you should then jump step 3 and go for step 4.

	A	B	C
1	Name	Fraction	Experiment
2	09abril_FlaviaWinck_C1	1	Control
3	09abril_FlaviaWinck_C2	1	Control
4	09abril_FlaviaWinck_C4	1	Control
5	10abril_FlaviaWinck_T1	1	Treatment
6	10abril_FlaviaWinck_T2	1	Treatment
7	10abril_FlaviaWinck_T3	1	Treatment
8			
9			

experimentalDesignTempl

- For including each data file as an individual experiment you should assign different identifiers to each of the files, as shown below.

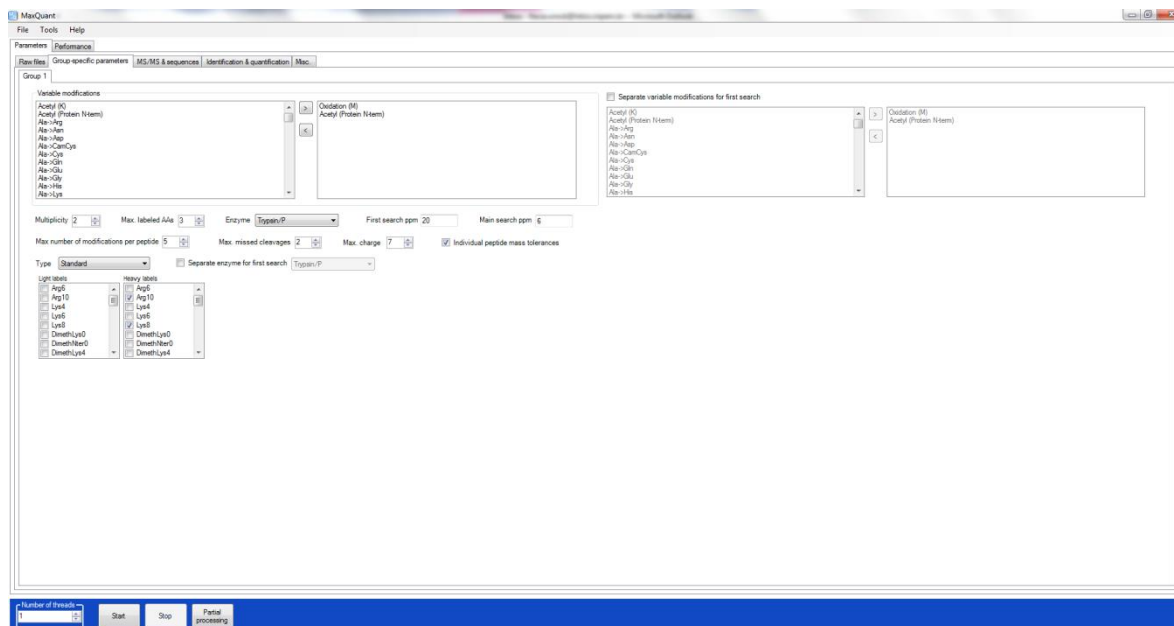
	A	B	C
1	Name	Fraction	Experiment
2	09abril_FlaviaWinck_C1	1	c1
3	09abril_FlaviaWinck_C2	1	c2
4	09abril_FlaviaWinck_C4	1	c3
5	10abril_FlaviaWinck_T1	1	t1
6	10abril_FlaviaWinck_T2	1	t2
7	10abril_FlaviaWinck_T3	1	t3
8			

- Save the document keeping it as .txt file format and do not modified the position of the columns.

2.7 Selecting protein identification parameters

2.7.1 Group-specific parameters

In the MaxQuant interface click in a menu named “Group-specific parameters” which will open a form with several parameters you must select in order to perform the protein identification.



Variable modifications

In this menu, you can include extra variable modifications which should be taken into account for the protein identification by clicking in a modification on the left box “**Variable modifications**” and click in the symbol “ > ” to select the additional modifications. Maximum number of modifications per peptide must be selected (e.g. 5) and the Maximum charge of each peptide must be defined (e.g. 7).

Cleavage and enzyme parameters

You must also indicate the Enzyme used to digest the proteins (e.g. Trypsin/P), the maximal number of missed cleavages (e.g. 2), modifications and charge accepted for each peptide identified. Usually, the default values presented in this menu are compatible with most proteomic applications.

Defining your quantification mode

For **Label-free quantification**, you must select number **1** in the parameter **Multiplicity**.

For other quantification strategy, such as the ones using isotopically labeled peptides, you

must select the multiplicity which corresponds to the number of different isotopes you are using in your analysis. For instance, if you have performed an experiment using SILAC labeling in one sample with a heavy isotope, you must select the number **2** in the parameter **Multiplicity**, and indicate the Light (if any) and Heavy labels present in your experiment and the maximum number of amino acids that could be labeled.

Error tolerance

Maximum error tolerance (ppm) must be defined for the first search (e.g. 20) and for the main search (e.g. 6).

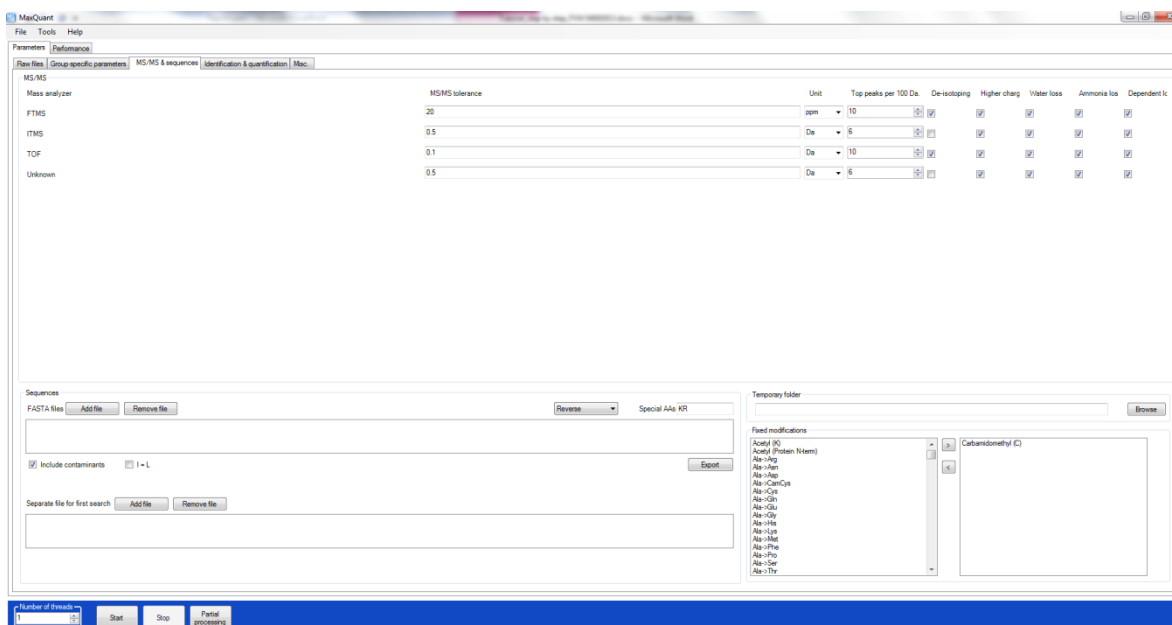
2.7.2 MS/MS & sequences

Now you will fill the parameters of the peptide detection and you will select the database which peptide identification must be performed against.

Click in the “MS/MS & sequences menu”. Select the MS/MS parameters for mass error tolerance either indicating in Da (Daltons) or in ppm. You must also select the number of top peaks you want to identify. Again, the default parameters should be appropriate for ordinary proteomic identification.

In the lower part of the menu, you have to indicate which database file you want to use in your peptide search. This file will be used also to generate the Decoy protein database using the Reverse or Random protein sequences.

1. Click in “Add file” and browse the protein database file which contains all protein sequences from the organism or combination of organisms you want to perform the search. Remember that this file must be in FASTA format.



TIP 3

A Decoy protein database is a new file which is temporarily created from the protein sequence database of the organism you are studying. The original protein sequences will have their amino acids sequence shuffled in a Random way or organized in a Reverse order, so you have the same peptides in a wrong order. This information is used to calculate the False Discovery Rate (FDR) of your resulting protein identifications and it permit to filter or to establish a limiting FDR for reporting peptide or protein identifications.

Indicate in the box “Fixed modifications” which will be the peptide modifications which must be included as permanent modifications of the peptides you will identify. By default Carbamidomethylation is selected. If you want to select other fixed modifications, you must select the modifications from the list in the left box and then click in the symbol “>” to include the additional modifications.

2.7.3 Identification and quantification

We now need to add the parameters which will define which protein identifications will be reported and filtering parameters must be included. In the menu “Identification and quantification” you must define what will be the highest FDR for peptide and protein identification.

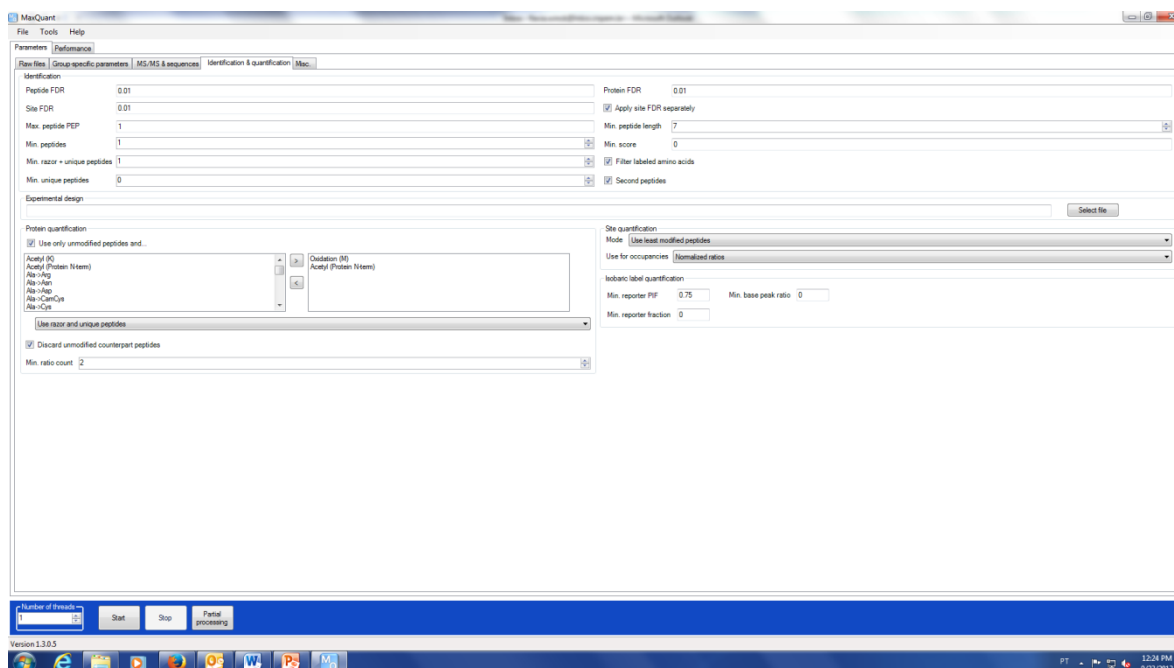
In the identification area, you must choose the **minimal number of peptides** that have to be matched for a certain protein be considered as identified. In this case, you must define the minimal number of unique peptides (peptides that only match one protein sequence) and razor peptides (peptides that might match to more than one protein sequence) that have to be identified during the protein identification. For Label-free quantification, you might also uncheck the default selected option on “**Filtered labeled amino acids**”.

Filtering by **False Discovery Rate** can be applied here by indicating which maximum FDR value you want for you peptide and protein identification. The default values (Peptide FDR = 0.01 and Protein FDR = 0.01), will filter out all identifications which FDR values are higher than 0.01, so only proteins with a FDR less than 1% will be reported.

In the Experimental design area, you have to include the location of the experimental design template file you filled with the data from your experiments. Click on Search File on the right side of the empty box below **Experimental design** and browse the file “experimentalDesignTemplate.txt” you have previously edited.

In the Protein quantification area you have to select some parameters to specify the peptides that will be considered for protein quantification, such as razor and unique or only unique peptides. You can select modified peptides to be quantified, besides the peptides you already selected in the previous menu. For quantitative purposes, you have to define what will be the minimum ratio count for a protein to be reported. For instance, if you define the minimum ratio as 2, only proteins which had a ratio higher than 2 will be reported in the final files. Our default value for Min ratio of protein quantification is 2.

In the Protein identification area, you can select which peptides should be used in the quantitative analysis and you may decide to only use unique peptides in the relative protein quantification analysis. The default value is Use razor and unique peptides.



2.7.4 Miscellaneous

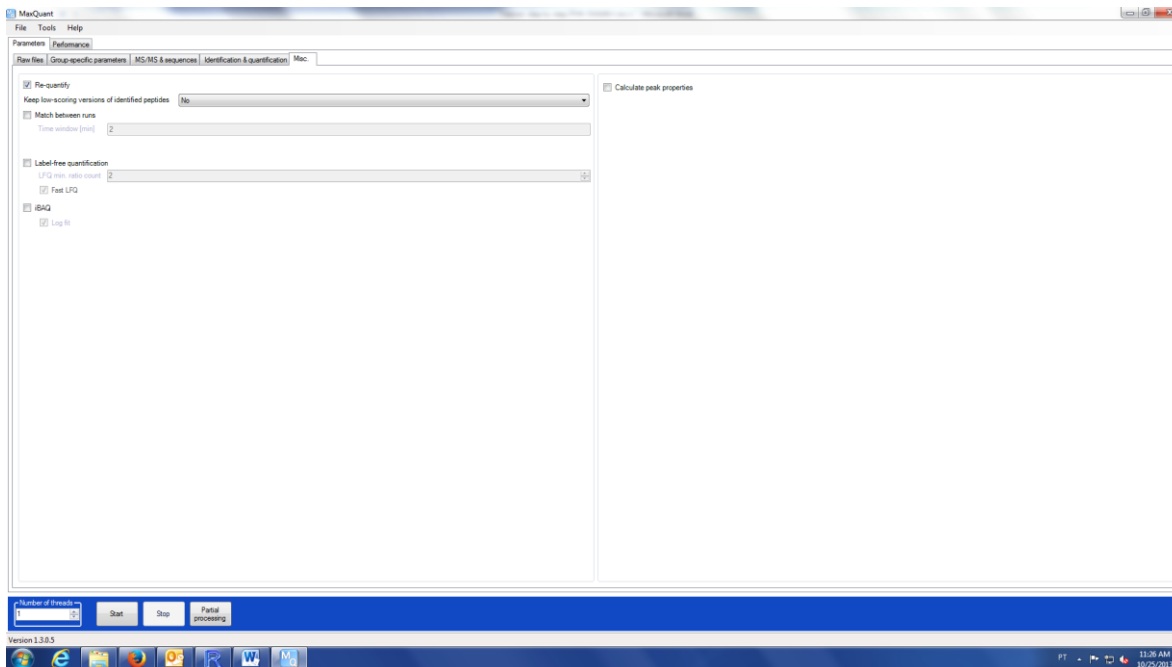
In the Miscellaneous menu you will find four main options you must fill.

Keep the Re-quantify checked so the peaks which had no labeled pairs will be used in a second search based on their location of retention time.

The option Matching between runs is used only when you are performing a Label-free protein quantification and in this case it must be checked. The information of the runs regarding the identifications from one run will be used to screen another run looking for peptide peaks closely located in the same retention time range. This time window can be modified. The default is 2 min and for most purposes it works fine for high reproducible runs.

The option **Label-free quantification** must be checked only if you are performing this quantification strategy. Here you must decide which will be the minimum ratio between the intensity values you want to be reported. The values of protein intensity reported using the LFQ method will appear in the resulting final file "Proteingroups.txt" in a column name

“LFQ Intensity”. The default value for **LFQ min. ratio count** is 2 and LFQ fast must be checked.



2.7.5 Saving parameter settings

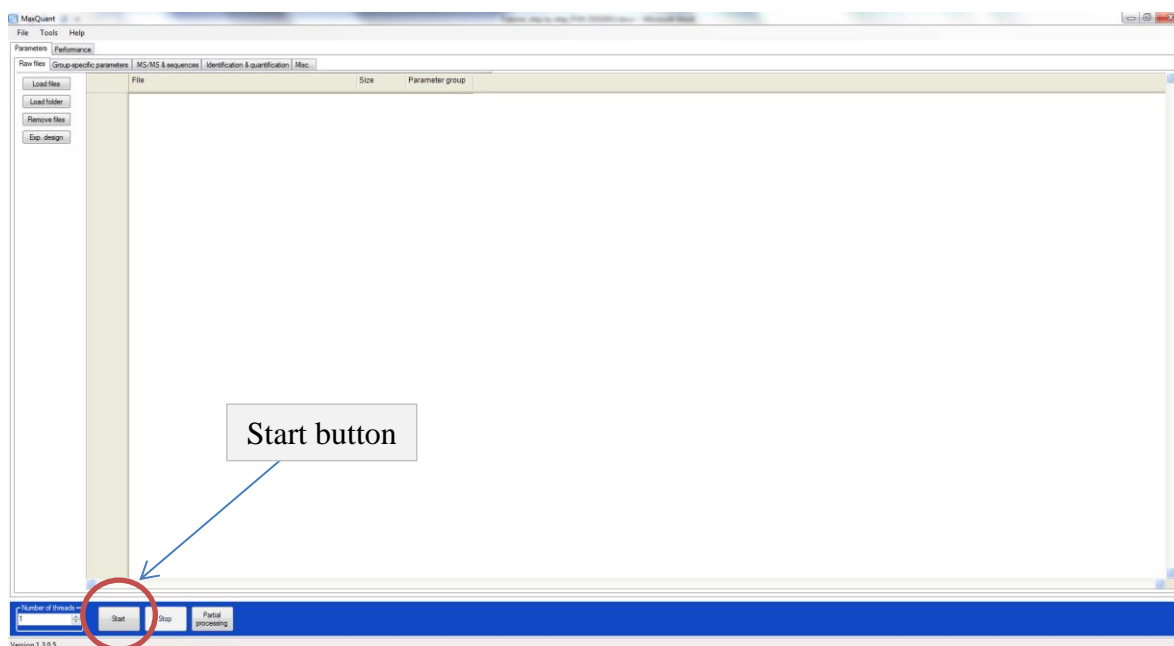
Once you finished deciding which parameters fits best to your research, you may want to save this file as a reference.

1. Click in File.
2. Select Save parameters. The file “mqpar.xml” will appear in the same folder where the .raw files are located. Do not transfer or save this file in another directory.
3. Do not change the name of the file now. Just click SAVE.

2.7.6 Starting the protein identification process

After finishing setting all parameters for the protein identification and quantification, you can now start your identification process.

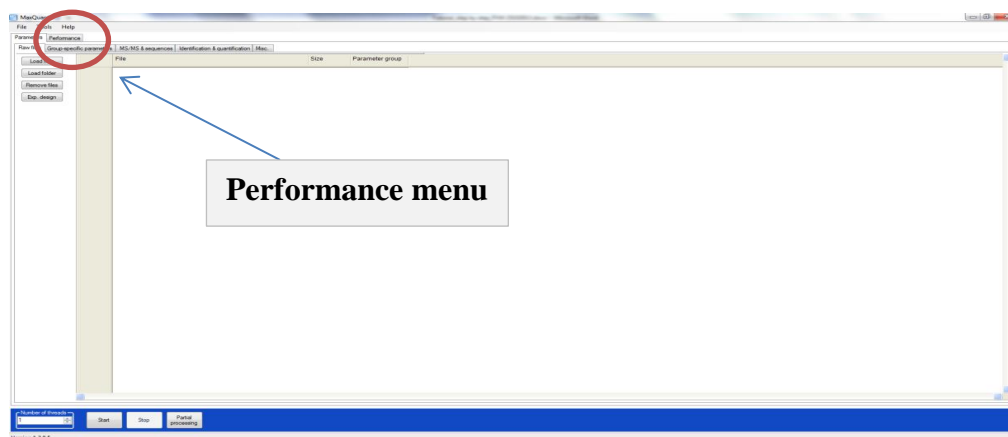
1. Press **Start**, located in the lower area of the interface.



If you have many processors available you can also distribute the job in up to 4 threads.

If you want to do so, modify the number in the left side of the Start button.

In order to verify if the processing steps are running fine, you can follow the steps of your data analysis in the **Performance menu**, located in the right side of the Parameters menu.



2.7.7 Finishing the protein identification and getting your results

After your protein identification and quantification is done, you will see a message written “Done” in the graphical interface of MaxQuant.

Once this message appears your results must locate your files.

1. Go to the directory where your .raw files are located.
2. Open a directory named “combined”.
3. Open a directory named “txt”.

The results of the MaxQuant search will be in .txt files, which contain all information from your search job.

The main file here which you will use to further protein identification, quantification and annotation will be “proteinGroups.txt”.

In our next chapter “**Protein data analysis with Perseus**” you will learn how to process your data and interpreting the results.