MS-Angel:

an automated management of LC-MS/MS acquisition files, and a direct link to Proline

(release 1.5)

Introduction to MS-Angel

The MS Angel software allows you to easily manage your acquisition files. It offers several levels of file processing, including:

- file conversions, from RAW and WIFF files;

- peaklists identifications using one or several search engines (Mascot fully supported, OMSSA is being integrated);

- import of identification results within Proline.

The design of workflows and tasks provides a high level of automation and control.

Install, set up and run the application

1. Download the application

MS-Angel is available at http://proline.profiproteomics.fr/download/ and unzip the archive file.

2. Configuration file

The configuration file is located in the /conf folder. All fields can be modified through the graphical interface, using the *Setup* menu (see dedicated documentation).

However, we recommend to provide the URL to the Proline server before starting the graphical interface. Indeed, the interface needs access to the server and may not start if looking for some information it can't get from it.

The fields dedicated to default paths are not mandatory. If you wish to ignore them, just let the empty value (double quote), but don't delete the fields.

```
proline-config {
   url="http://localhost:9000"
}

mascot-config {
   url="http://www.matrixscience.com/cgi"
}

timers {
   auto-refresh=30000
   http {
        read-timeout=10000
        connection-timeout=10000
  }
}

input-files-dir=""
created-mgf-files-dir = ""
created-mzdb-files-dir = ""
par-files-dir=""
```

3. Start the graphical interface

The interface can be starting by double clicking the JAR file or using one of the staring scripts.

- start.bat (Windows) or start.sh (Linux) will start the interface with a console where the logs will be displayed. This is convenient for admin user.
- start_without_console.bat will start the interface and immediately hide the console. This is meant for regular users. The logs will still be available in the log file of the day in the /logs folder.

4. Finish the setup via the graphical interface

Before using the application and running some workflows with it, there is still some configuration to do. Go to the *Setup* menu, where you will be able to set the last parameters of your configuration (see details below). These fields are not in the configuration file, they are saved server-side. That's why you need to set up the Proline URL first.

File browsing preferences are optional.

File conversion tools path are mandatory to use them.

Mascot server URL is mandatory to run identifications on Mascot.

The Mascot result files mount point is mandatory to import result files into Proline.

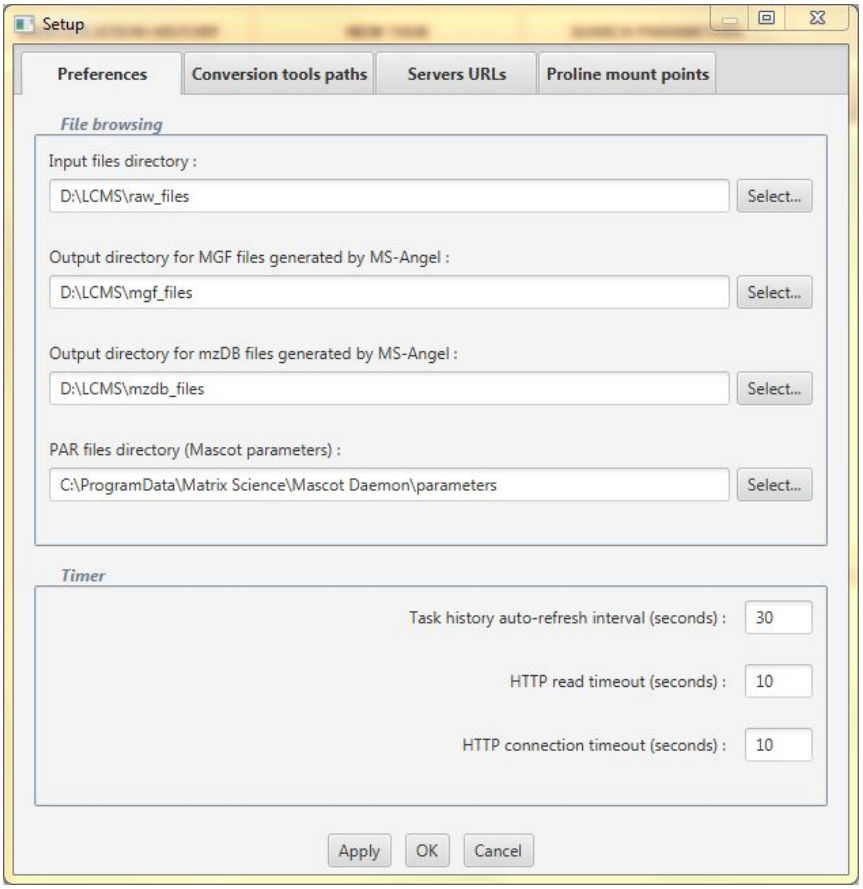
Setup menu

On the menu bar, click Setup to open the dedicated window. It is separated in for tabs. There are 3 buttons at the bottom **(WARNING: you may have to scroll down to see them)**:

- *Apply*: save the changes on this tab, continue editing the setups
- *OK*: save the changes for all tabs and close the window
- *Cancel*: close the window, ignoring the changes

NB: apart from the *Preferences*, the setup is meant for the admin only. Regular users should not modify the setup.

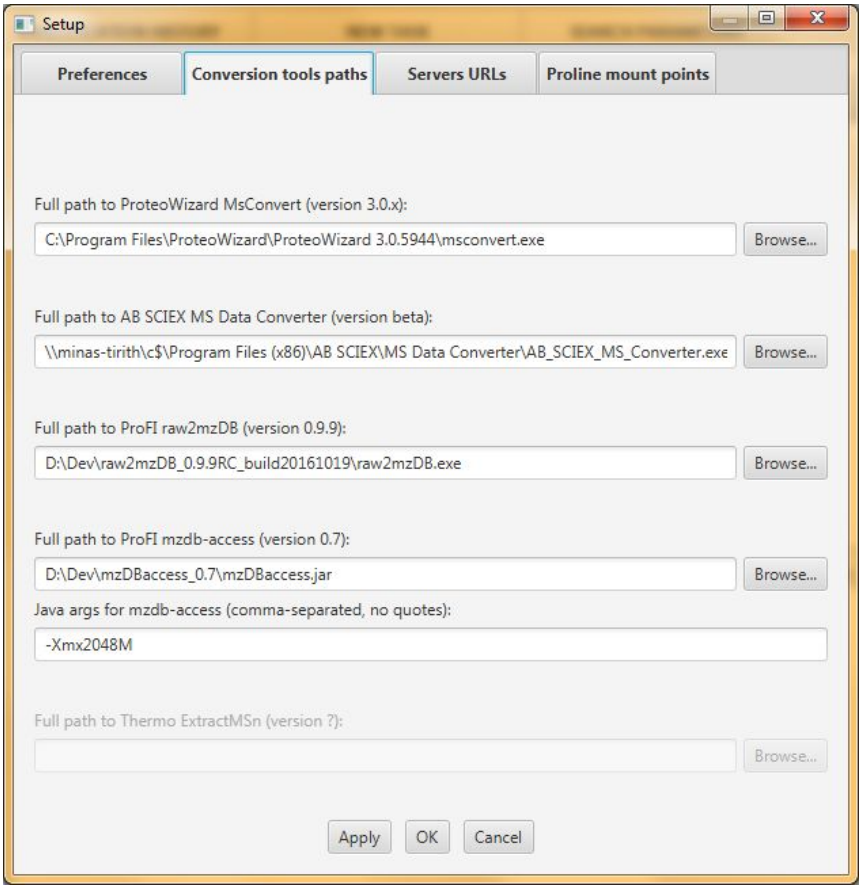
Preferences



The first part is meant to define default path to specific kind of files, in order to save you many redundant clicks in the file browser.

The second part lets you define your preferences in terms of timers:

- Task history auto-refresh: the period at which the interface will automatically update the tasks progress
- HTTP read and connection timeouts



Conversion tools paths

MS-Angel lets you use many tools for file conversion. Yet, in order to have a lighter software, they are only supported by MS-Angel, not embedded in it. This means you must first download a tool on your computer, then set up the full path to its executable (.exe, .jar).

Warning: please mind the indications on the tools versions that are supported. Parameters and usage may defer from one version to another of a software.

Note that you can also attribute some Java parameters to the use of the ProFI mzdb-access converter.

Servers URLs

- Proline server URL:

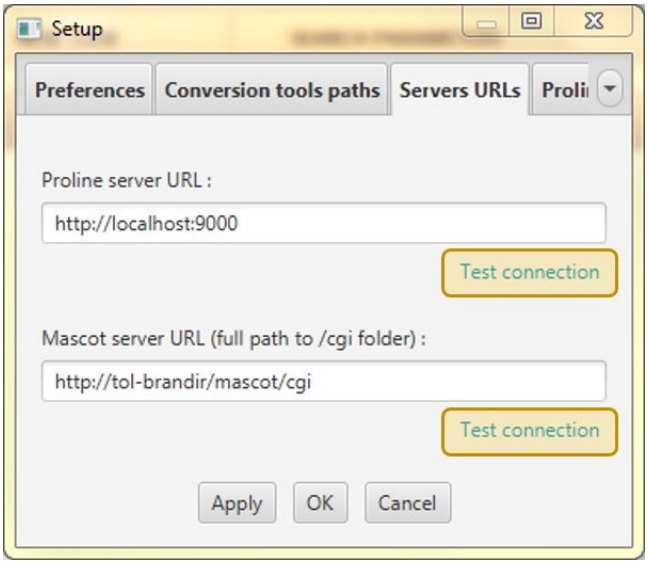
the URL to the Proline server, must be of the type *http://machine_name:port_number*. If the server is installed on the same machine, on the default port, then it will be: *localhost:9000*.

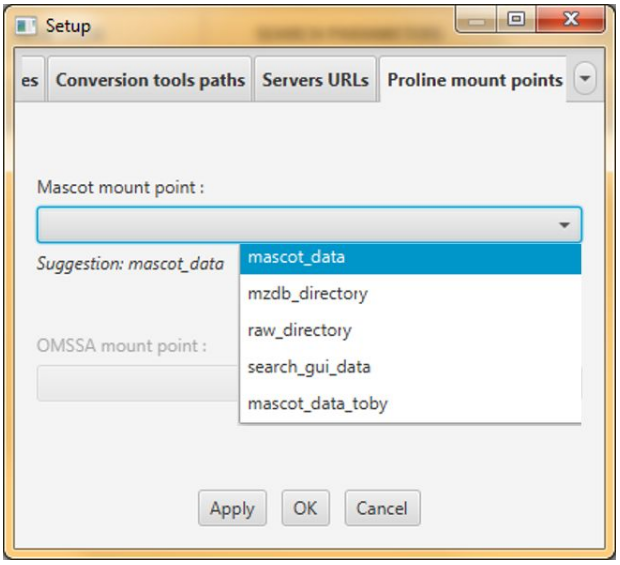
NB: This part is crucial, because MS-Angel can't work without a connection to Proline for now. Thus, we recommend you to specify this URL directly in the configuration file before the first start (this is a job for the administrator, see ***Installation***).

- Mascot server URL:

the full path to the */cgi* folder of your Mascot server. It can be a private server or the public one: http://www.matrixscience.com/cgi (default).

You can click the *Test connection* links to make sure your URL is valid and offering a connection.





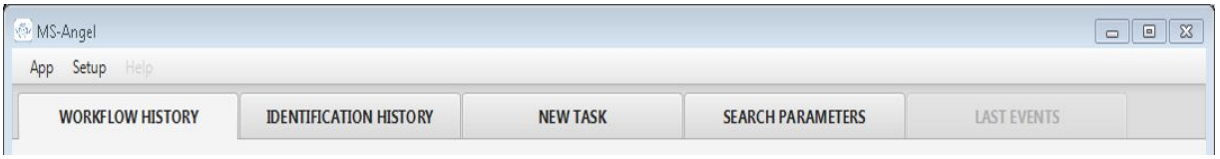
Proline mount points

Proline works with mount point, i.e. alias to specific folders. One of them is the mount point under which the Mascot results will be found. Since many mount points can be designed, one must be selected. Make sure this is the case **before** launching a task requiring it (e.g. including the import of Mascot results into Proline databases).

As you can see, the software try and guess which mount point is the best match.

Mount points are defined in the server configuration file.

The graphical interface in 5 tabs



- *Workflow history:* visualize tasks, giving details about the progression of the workflow on each file of the selected task.
- *Identification history:* visualize tasks including an MS identification step, giving details about the search engine, search parameters, and identification results for each file of the selected task.
- *New task:* design and launch a new task.
- *Search parameters:* create, import, visualize and modify search parameters templates.
- *Log events:* know about the last notable events in tasks execution [not yet implemented].

User case example

We will follow the example of a user willing to launch a task executing :
- the conversion of RAW acquisition files into MGF files (Mascot Generic Format),
- a peaklist identification using Mascot,
- the import of identification results into an existing Proline project.

For this, we will see how to go through each step:
1. Create a search parameters template
2. Create and launch a task (input files, workflow…)
3. Visualize progression and results.

For some of these steps, several options will be suggested and explained.

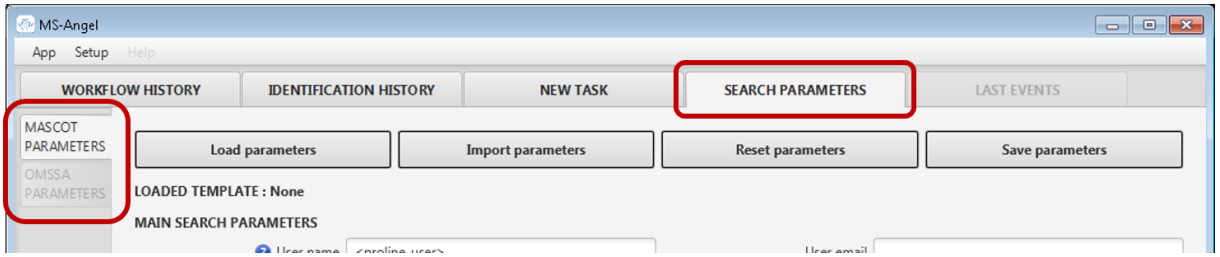
NB: The tasks and templates created in MS-Angel are assigned to an owner, wich is a Proline user. If you don't have a profile on Proline yet, please create it first. If you wish to import your identification results in a Proline project, it also has to be already existing. See how to create a project in Proline.

1. Create a search parameters template

Go to the “Search parameters” tab, then to the search engine-specific subtab.

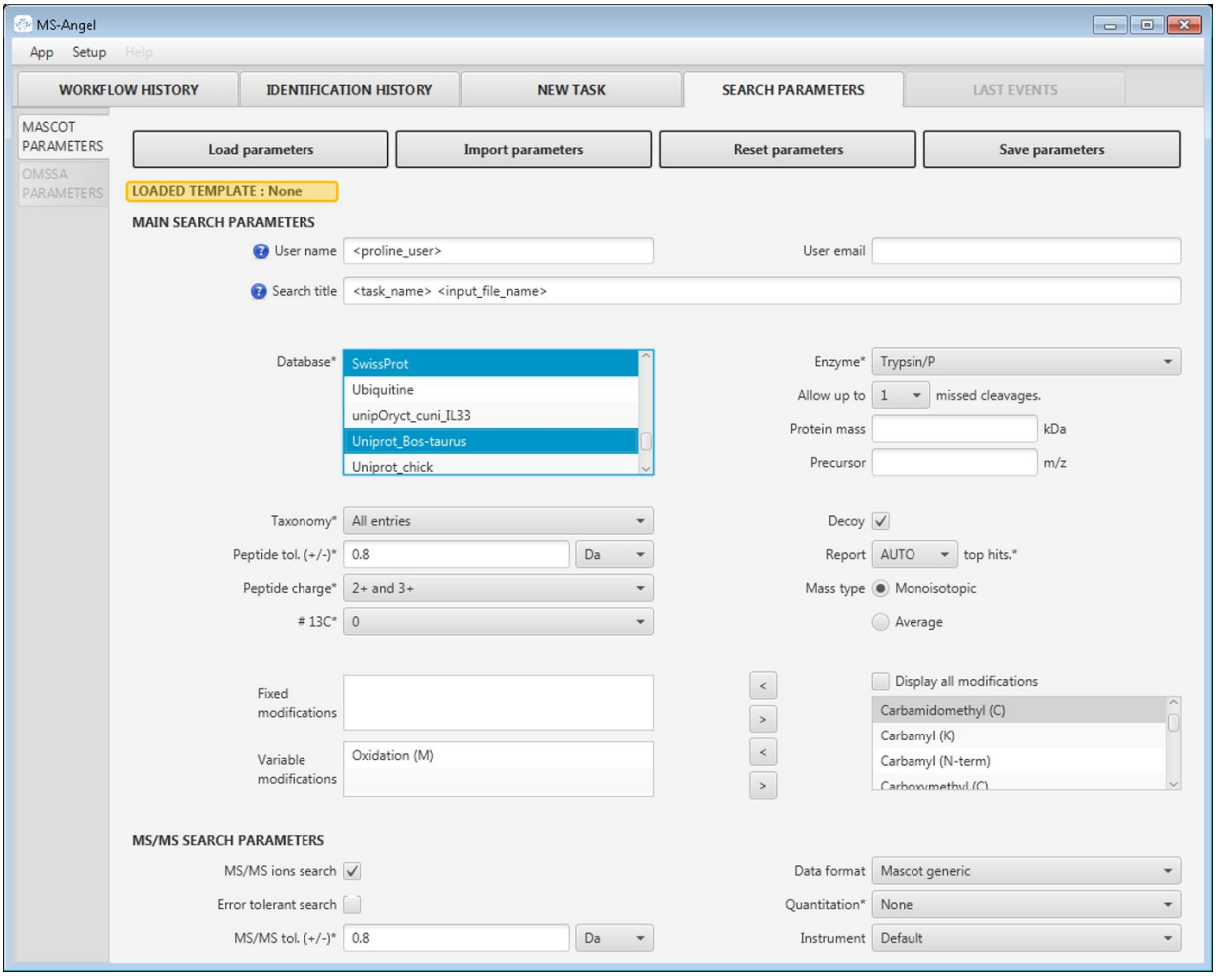
There a several ways to create a template:

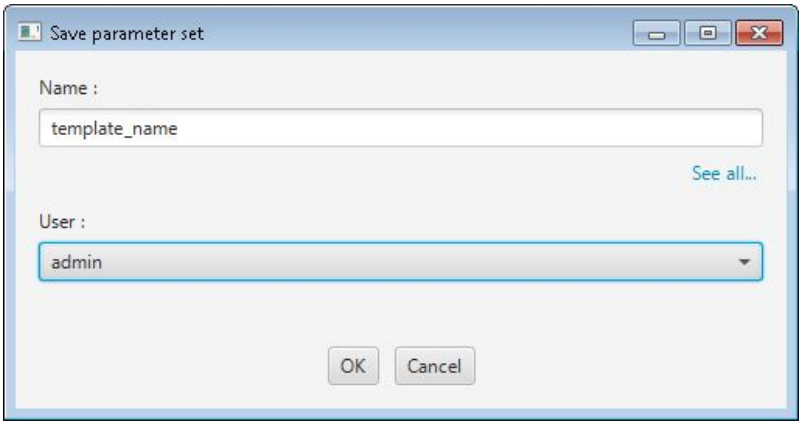
- fill the form by hand
- modify an existing template
- import a Mascot Daemon .PAR file



Option A: Fill the form by hand

Fields whose names are marked by a star are mandatory.
Keywords can be used (within diples) for the fields *'User name'* and *'Search title'*. The list of available keywords and their meaning is displayed by clicking on the help icon.
Several databases can be selected by clicking on the databases names while pressing the shift key (or the Ctrl key to select a range of entries).
To add some PTMs to the *'Fixed modifications'* list, select them in the complete list of modifications (on the right), then click on the top '<' button. To remove modifications from the *'Fixed modifications'* list, select them within the *'Fixed modifications'* list then use the top '>' button. This works the same way for *'Variable modifications'*, using the bottom '<' and '>' buttons.



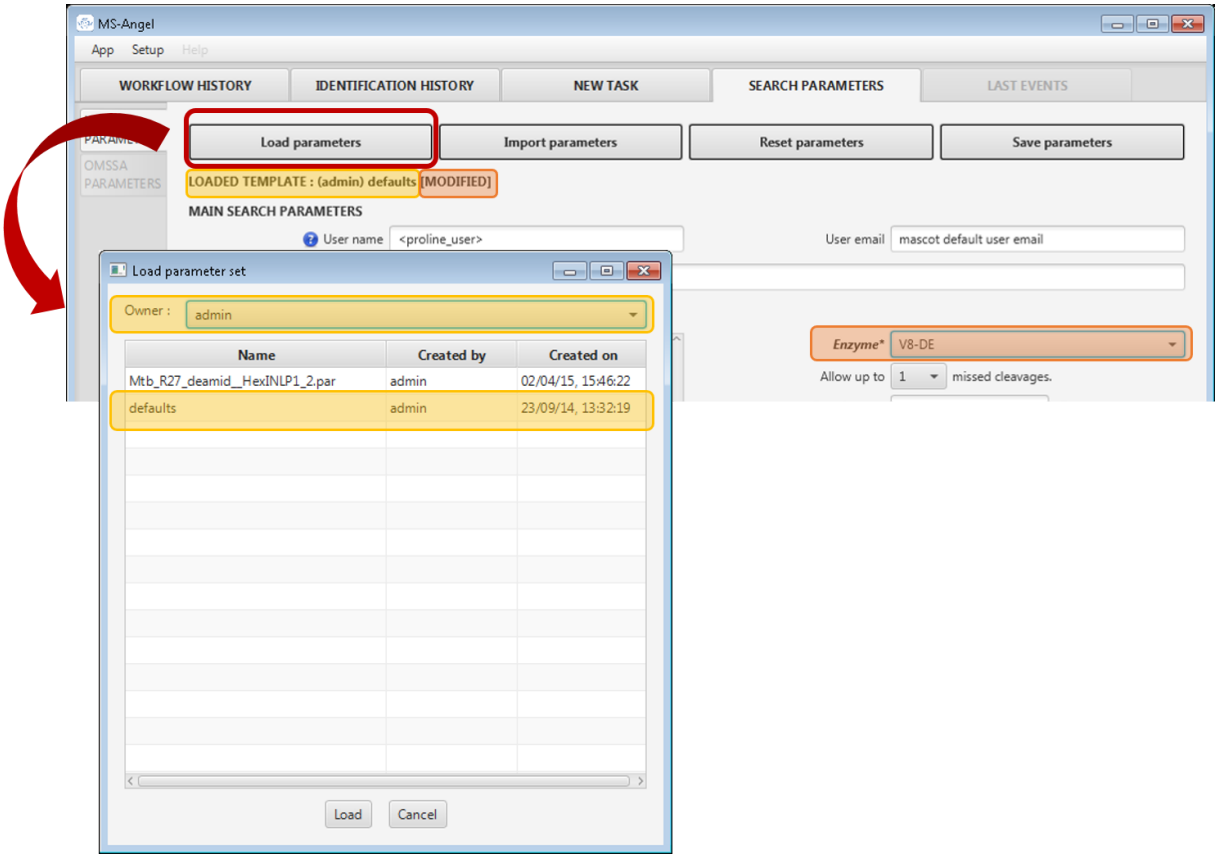


Then save your new template by clicking *'Save parameters'*.

You must provide your template a name, and an owner within Proline users.

Option B: Modify an existing template

Click on *'Load parameters'*. Select the template you want to load. The template list can be filtered by template user, as seen on the screenshot below. You can see all the templates in the database by selecting *Owner: All users*.



Load the template by double-clicking on it, or clicking *'Load'*.

The form fields will then auto-fill with the template values; and the template name and owner will be displayed at the top of the form (yellow zone on screenshot).

You can edit any field. When a value is changed and differs from the actual template, the field name

becomes bold and italic. If any value changes in the template, the text *'[MODIFIED]'* will appear near the template name (orange zones).

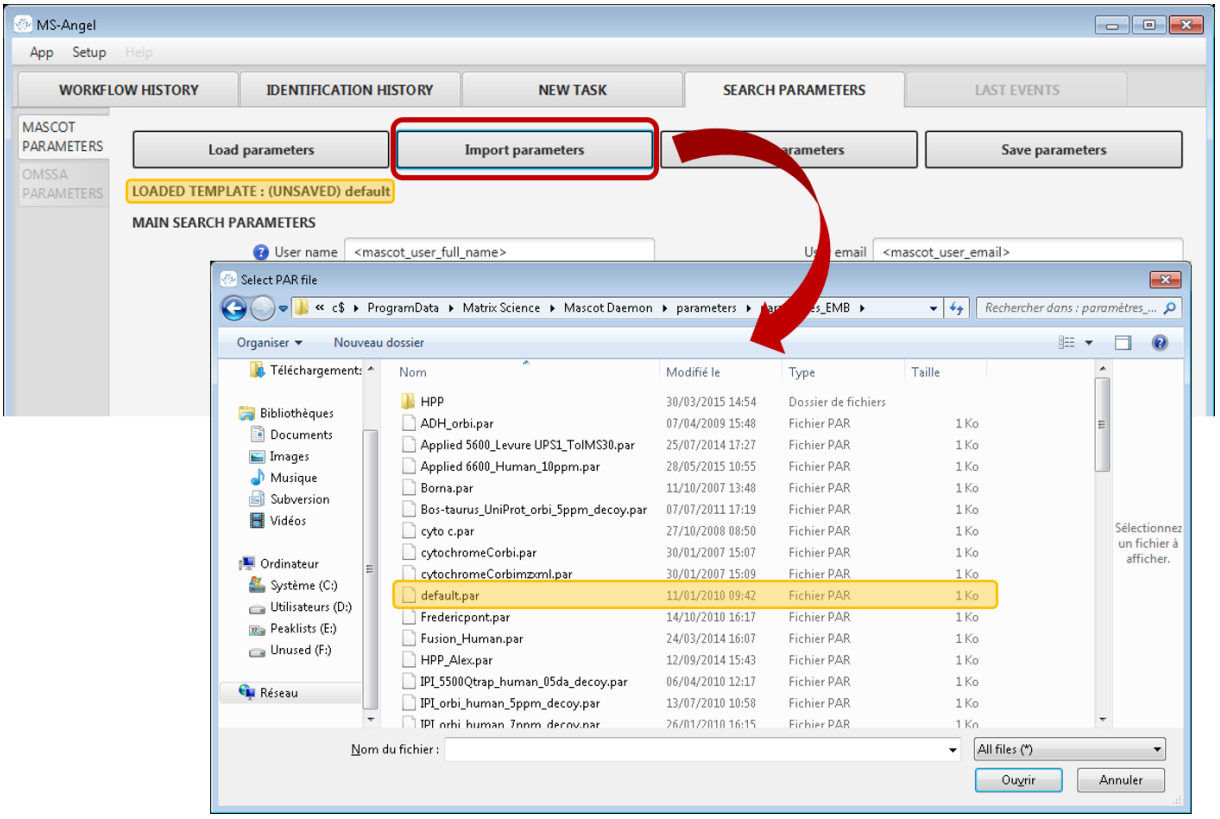
When your changes are done, you can either:
- save it as a new form: click on *'Save parameters'*, change the template name and/or owner.
- override the existing form: click on *'Save parameters'*, and don't change the template name and owner. A confirmation will be asked, since this operation is irreversible.

Option C: Import a Mascot Daemon .PAR file (Mascot parameters only)

Click on *'Import parameters'*. A file browser will open. The default folder for Mascot Daemon .PAR files can be registered in the 'Setup' menu.

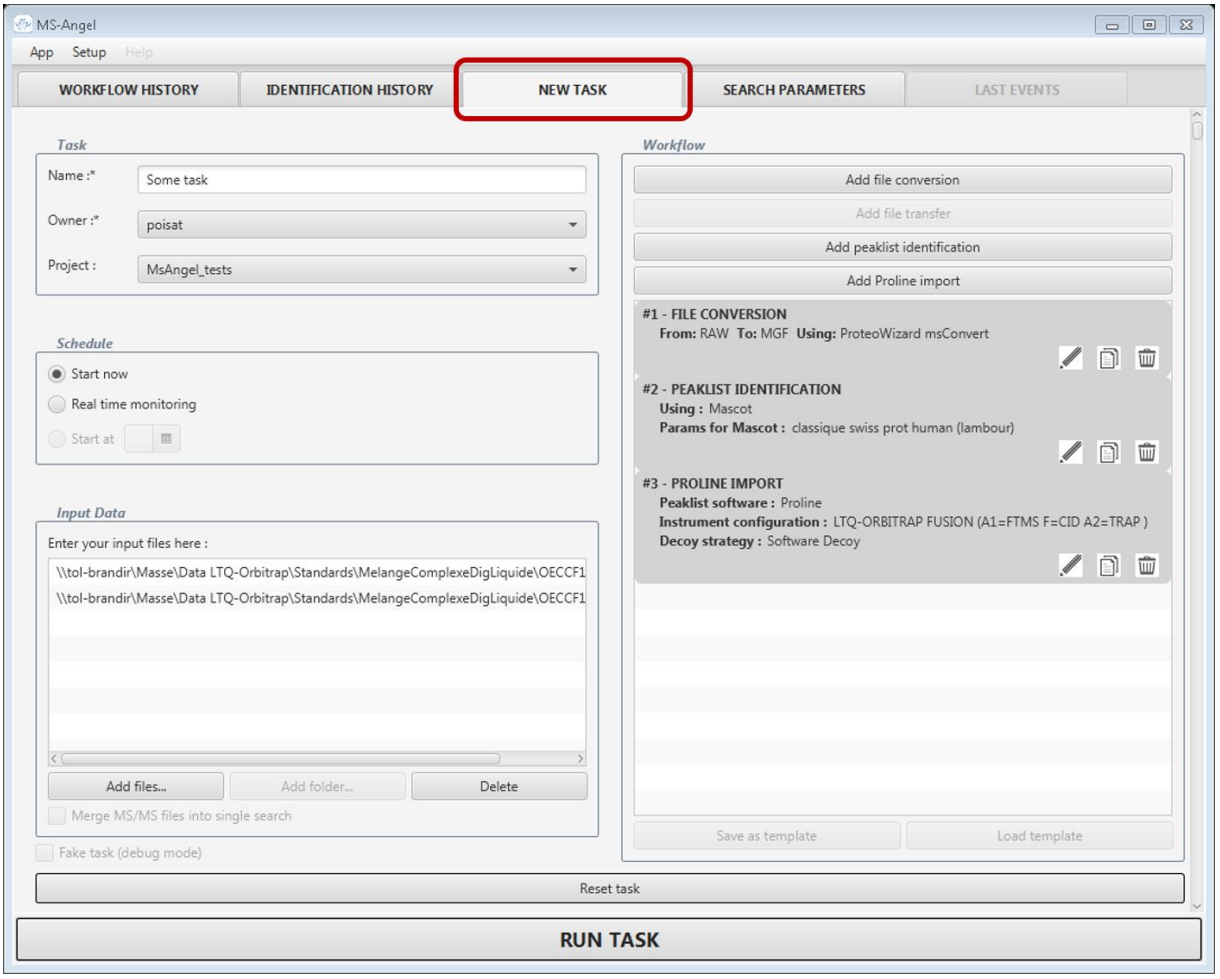
The form will auto-fill when a .PAR file is selected. Note that the default template name will be the .PAR file name, and that no owner is attributed to the template.

As said before, you can edit this form before saving it by clicking *'Save parameters'*.



2. Create and launch a task

Select the '*New task*' tab.



What is a task, what is a search

The tasks and searches are organized as in Mascot Deamon:
- a **search** refers to a single file or, more widely, to the execution of the workflow on this input file.
- a **task** refers to a set of searches: a set of input files that will be processed together, with the same workflow and the same parameters.

Define the global parameters for the task

Consider the '*Task*' part of the '*New task*' tab.

- First, give your task an arbitrary name. This is the name that will represent the task in the '*Workflow history*' and '*Identification history*' tabs.

- Select the task owner (you). This is a Proline user, which means you *must* have an account on your Proline installation already (see how to create a Proline user).

- Select the Proline project on which the task depends. As for the owner, it must be already existing (see how to create a project in Proline).

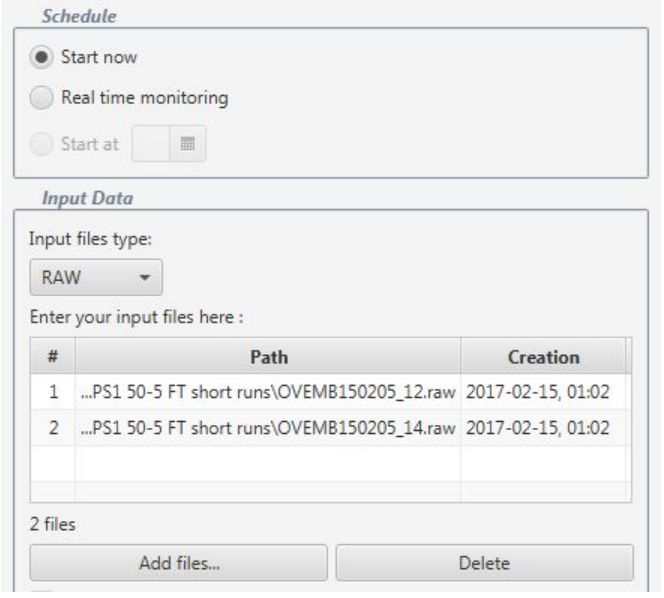
A project is mandatory if you wish to automatically import your search results in Proline. If you don't wish to use this feature, though, you can select 'None'.
Tip: You are strongly recommanded to specify a project when possible, since it will later become a criteria to quikcly filter and find tasks in the history.

Scheduling type, choosing input data

There are two types of task execution in MS-Angel:
- *Start now*: workflow execution in batch on a given set of input files.
- *Real time monitoring*: monitoring of a given data folder; the workflow is executed on each input file appearing in this folder (as soon as it is created).

The '*Input data*' panel depends on the selected schedule type.

Mode 'Start now'



In the '*Schedule*' part, select *Start now*.

In the '*Input data*' part, click on *Add files*. A file browser will open for you to select all your input files. It can be .RAW, .WIFF, or .MGF files. All input files must have the same extension.

The default folder for this file browser can be set in *'Setup' (menu) → Open setup dialog → Preferences (first tab) → File browsing (first section) → Input files directory (first field)*.

As long as you task is not started, you can modify this list by adding more input files, or select and remove some (*Delete* button).

Mode 'Real time monitoring'

In the '*Schedule*' part, select *Real time monitoring*. In the '*Input data*' part, you will be able to set up many parameters:

- *Path to data folder*: the **absolute** path to the folder you want to monitor, e.g. where the input files will be created then handled by MS-Angel. You are advised to use the '*Browse*' button to select your folder.

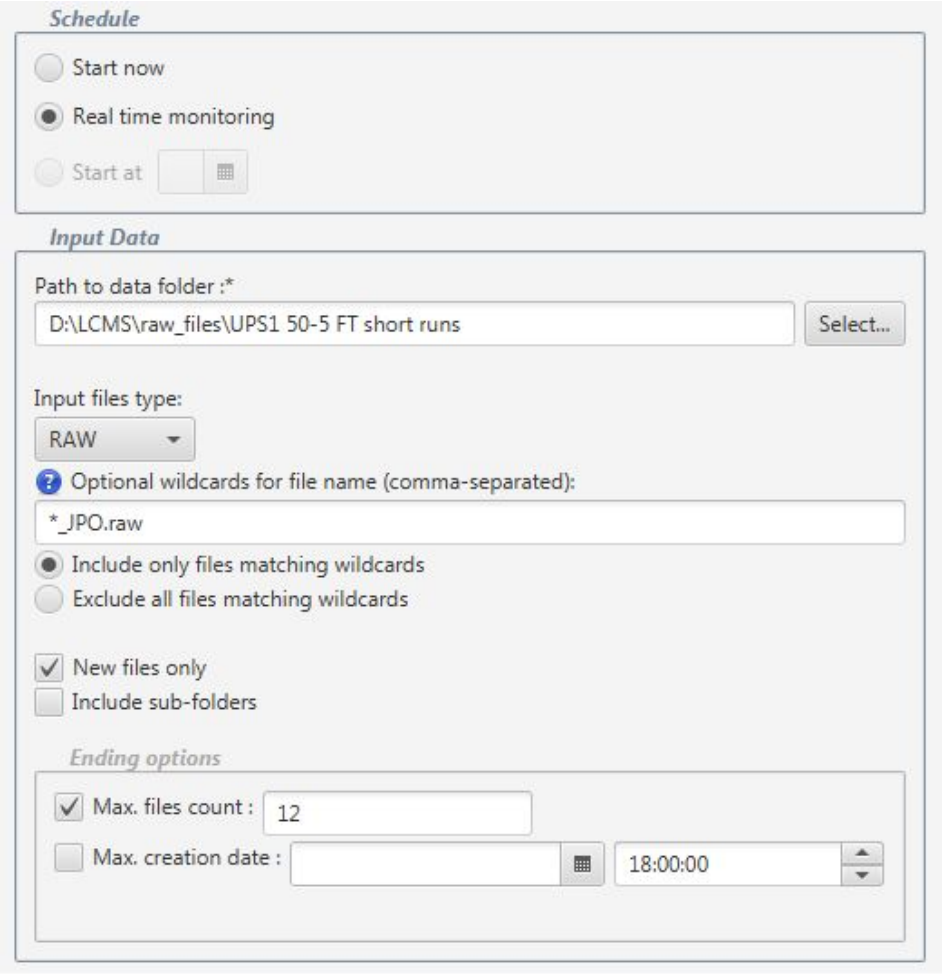
- *Optional wildcard for file(s) name*: you can use this textfield to filter the input files name and/or extension. A star means 'anything'. In the given example, only files whose name is ending with ".JPO.raw" will be taken into account. Only one expression can be described in this field (don't use

comma-separated list of expressions). If you don't want to apply any filter, you can just leave this field blank or with '*' or the initial '*.* '.

- *New files only*: if this option is checked, all files already existing before the task is launched will be ignored.

- *Include sub-folder*: if this option is checked, files created in folders under the chosen data folder will be processed.

- *Ending options*: select the criterium to end up the task. It can be either a finite number of processed files, or a given date, or (if the two are selected) the first to be reached. It is currently impossible to launch a task in Real-time monitoring mode without ending parameters.



Input files (mode 'Start now')

Input files can be added in 3 ways:

- click *Add files…* and browse them
- drag and drop files from the Windows explorer to the dedicated panel
- copy files from the explorer then paste them into the panel using *Ctrl+V*

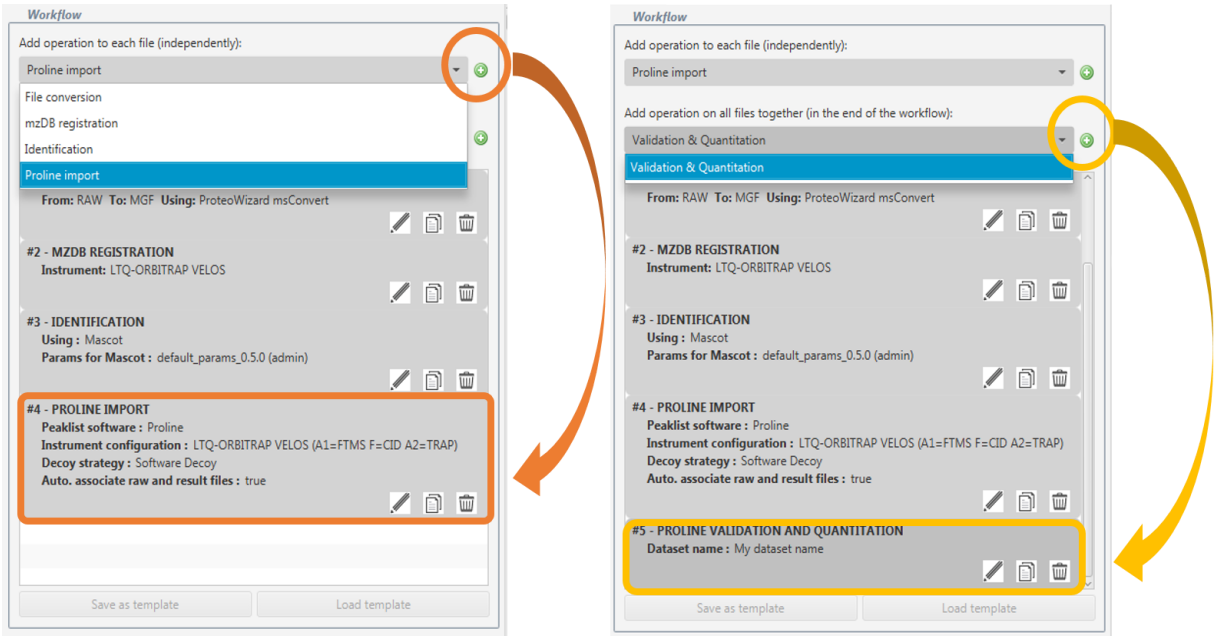
The content of the input files grid can be copied using *Ctrl+C*.

Design the workflow

Let's focus on the '*Workflow*' part. Here you will design the workflow applied on the input files. This is the types of operations that are currently available:

- File conversion
- mzDB registration (register raw and/or mzDB information in Proline databases for quantitation)
- Peaklist identification
- Proline import (import the identification result file in Proline)
- Validation and quantitation

The boxes buttons at the top allow you to create and design new operations.



WARNING: The operations order is crucial: the operations will be chained in this order, so it must be coherent with your file format, and in-between operations. For example, you must convert RAW files into MGF files before submitting them to Mascot, and must run a search on Mascot before import its results in Proline.
When you create an operation, it is placed at the end of the workflow. You can change its position by dragging and dropping it.

Two types of operations

There are now two types of workflow operations.

- The **job operations** occur on each file independently. They are available in the first choice box: file conversion, mzDB registration, peaklist identification, import of results in Proline (see graphics, left part).

- The **task operation** occur on all files altogether, when all job operations finished on all input files. Thus it must be placed at the end of the workflow. It is available in the second choice box: validation and quantitation (see graphics, left right).

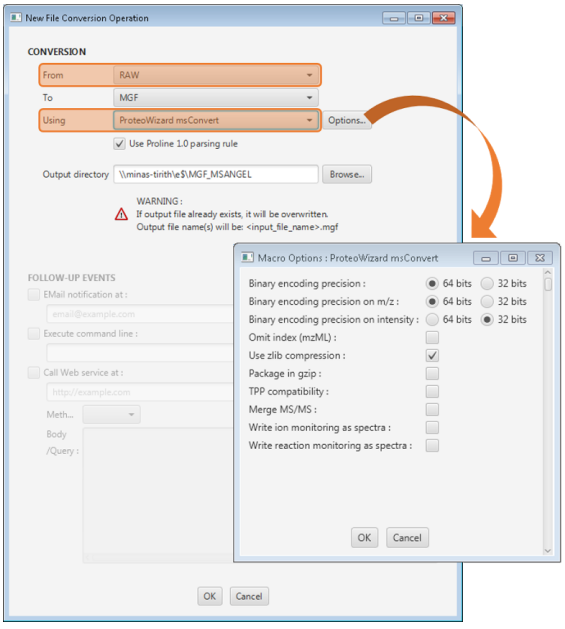
File conversion

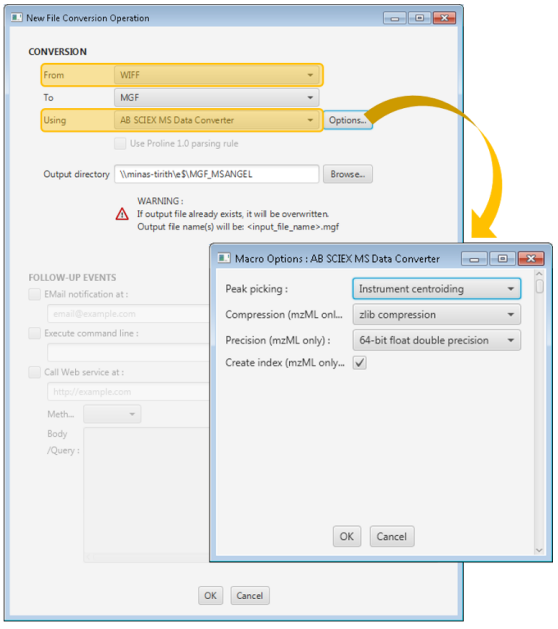
Click on '*Add file conversion*'. Select the input format of your files (depending on the Schedule mode, it may be pre-selected), then the format in which they will be converted. You can then select a conversion tool. If none is displayed, then the conversion you wish to do is not yet handled by MS-Angel. The available tools are:

- ProteoWizard MsConvert (typically for RAW → MGF conversions)
- ABSciex MS Data Converter (typically for WIFF → MGF conversions).

You can see and change the conversion settings by clicking *Options*.

When using MsConvert, you can (and are recommanded to) use the Proline rule for generated spectrum titles by selecting '*Use Proline 1.0 parsing rule*'. Using this rule, your MGF file will contain all the information needed by Proline for further analysis.

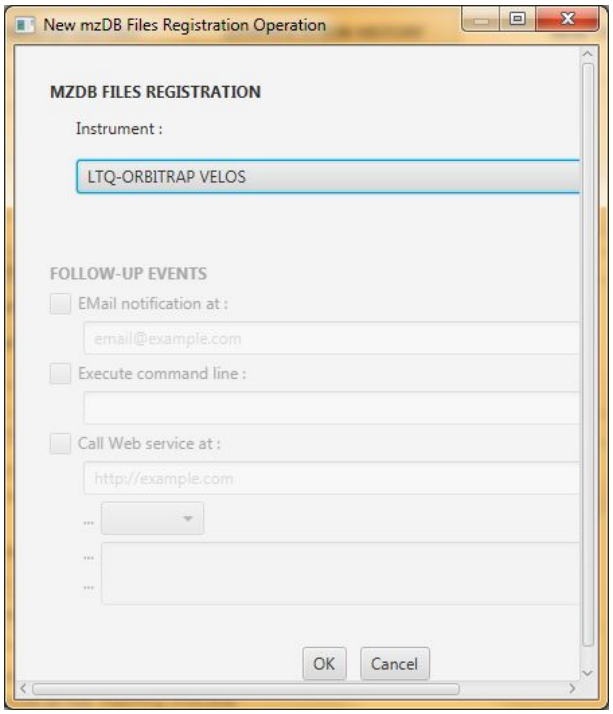




mzDB registration

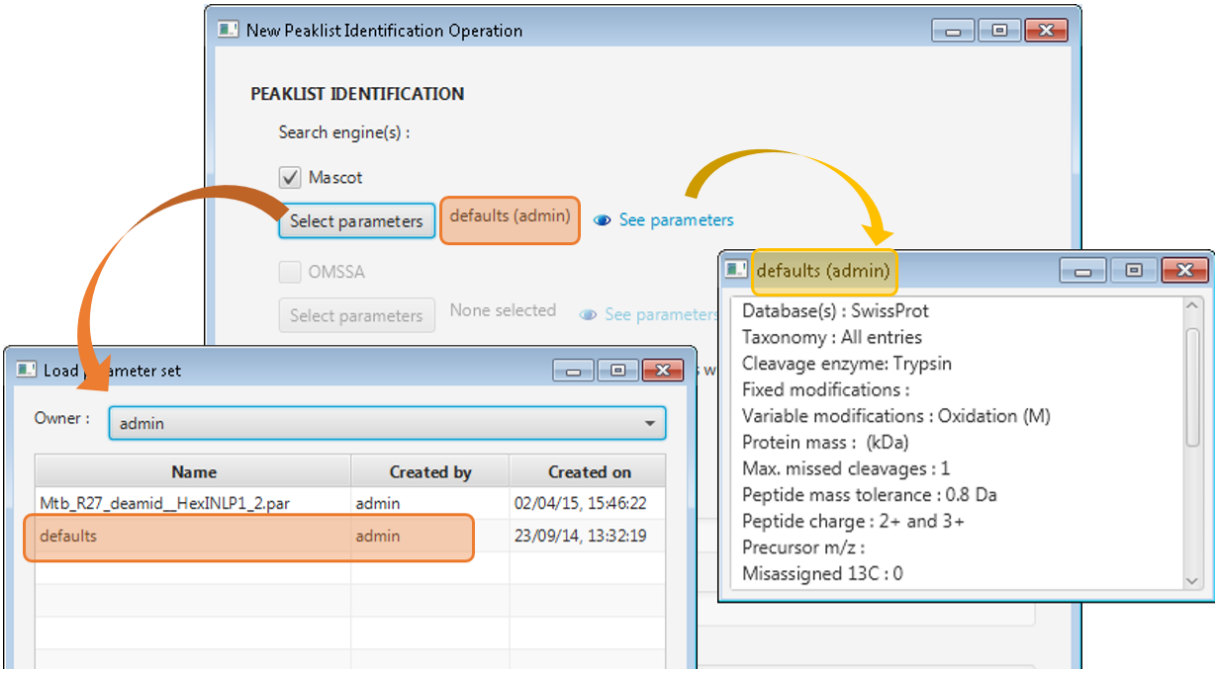
When running a quantitation in Proline, you need your raw data to be registered in the database first. This can be done using this operation. It can be placed anywhere in the workflow before the Validation and Quantitation, as long as the mzDB file has already been produced or provided. If this file has already been registered in Proline, the information will be updated with this file's content. Not that if the raw file generating this mzDB is present in the workflow, the two files will be automatically associated in Proline.

You need only select the type of instrumentation that was needed to generate



the file.

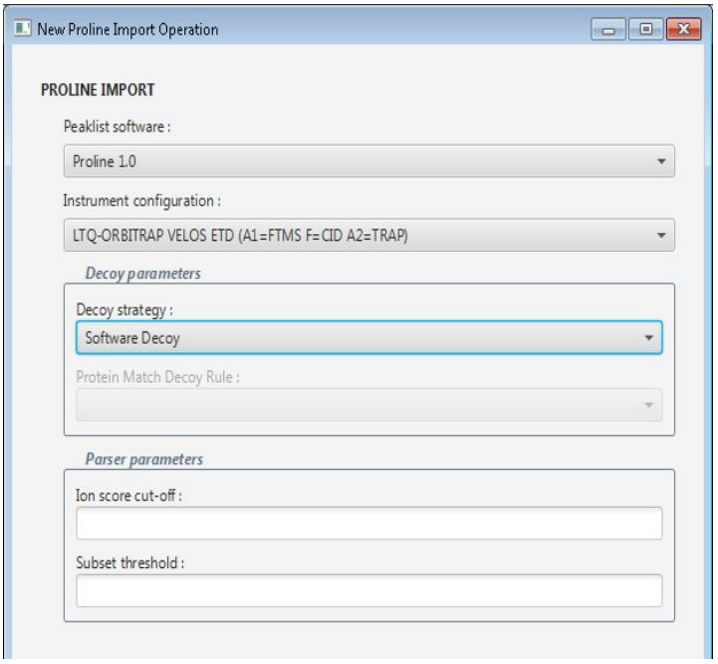
Peaklist identification



Select the search engine(s) you desire for peaklist identification.

For each, selected a parameter set by clicking the corresponding '*Select parameters*' button. When a template is chosen, its name and owner will be displayed, and you will be able to click '*See parameters*' to have a quick look at it.

Proline import

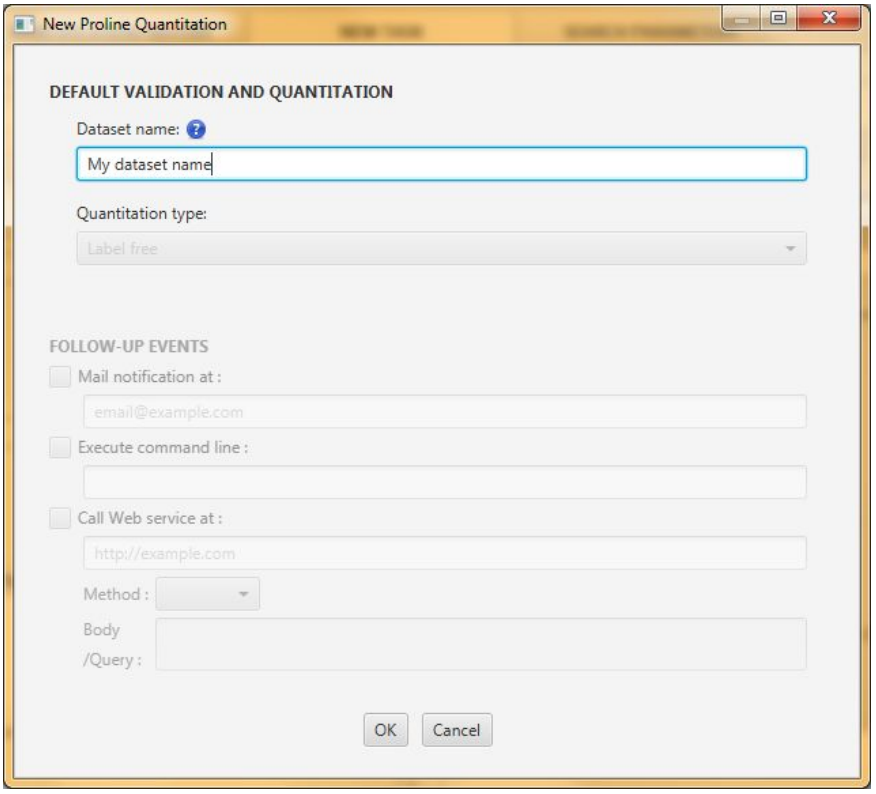


This feature allows you to import each identification result in Proline, as soon as it is created. The parameters are the same than in Proline Studio and in Proline Web.

- *Peaklist software*: the software that was used to create the peaklists (i.e. MGF files). If you used MS-Angel for this conversion, this value will be infered. (NB: If you used the option '*Use Proline 1.0 parsing rule*4, then select '*Proline 1.0*' here.)
- *Instrument configuration*: on which the samples were run
- *Decoy strategy*: if you ran a classic target/decoy search on Mascot, select '*Software Decoy*'.
- *Protein Match decoy rule*: these rules must be defined in Proline before
- *Ion score cutoff* and *Subset threshold* are optional.

Validation and quantitation

After importing the result files in Proline, you can create an identification dataset, validate it then quantify it using this operation.



Enter a dataset name: it will be used for both the identification and the quantitation datasets.

For now, it only supports the default parameters for validation and label-free quantitation.

Going to Proline graphical interfaces (Proline Web, Proline Studio) will let you:

- Visualize your results

- Re-run the validation with custom parameters

- Run the post-processing on the quantitation to access statistical data

- Clone the quantitation to use custom parameters, and optionally a more complex experimental design (this will allow you to compare ratios between groups). If you do clone the quantitation, choose the "*Use previous peakel detection*" option to quantify faster. The signal detection will then be skipped: the algorithm will use the signal found in the first quantitation.

Launch the task

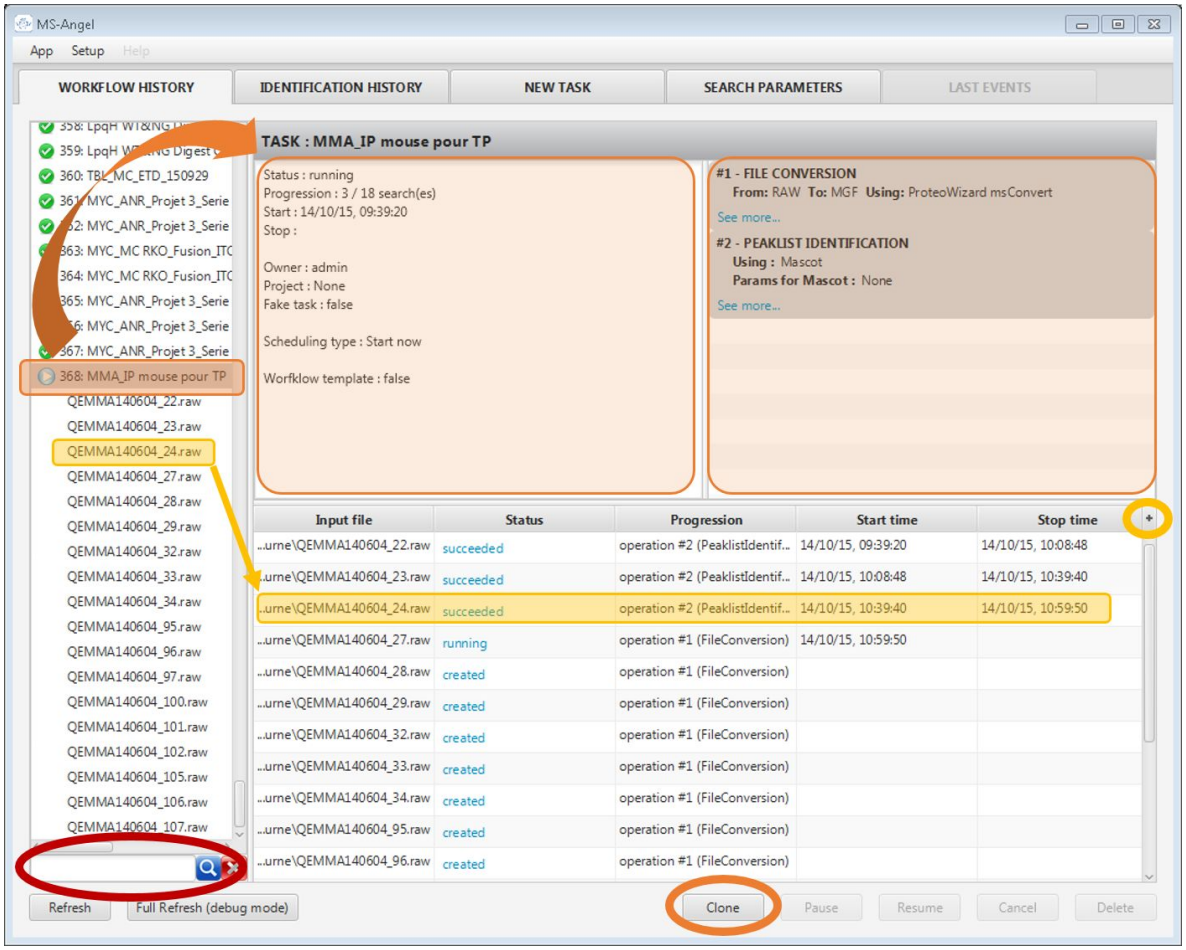
Click on '*RUN TASK*' at the bottom.

A popup window will open to let you know if the task was well launched. Close this window, you will be redirected to the *Workflow History* tab.

3. Visualize progression and results

Two tabs are dedicated to tasks visualization.
In the *Workflow History* tab, all tasks are displayed, and the progression of the workflow can be followed.
In the *Identification History* tab, only tasks with a Peaklist Identification are displayed. The information given in this tab is related to the peaklist identification.

Workflow history



In the tree at the left, all tasks are displayed with an icon pointing out the tasks status (running, succeeded, failed...). By double-clicking the task name (orange), the input files appear in the tree (yellow).
A selected task can be cloned by clicking *Clone*.
The displayed tasks can be filtered on their name using the tool circled in red above.

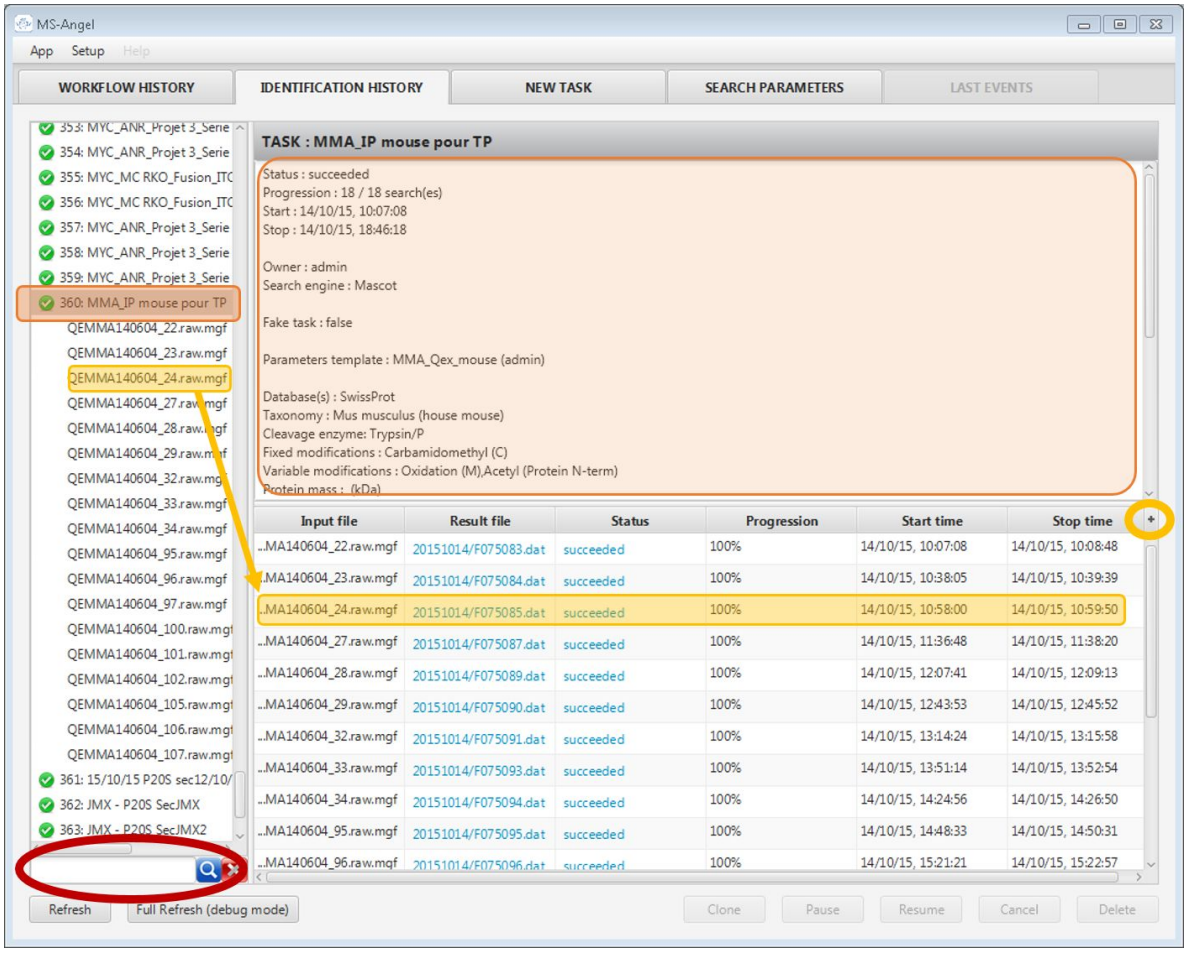
The top-right parts (pale orange) are related to the tasks. It gives detailed information on the parameters (left) and workflow operations (right). The bottom-right table is search-related: one line

per input file. Texts in blue are hyperlinks, offering much details on the search progression. Some columns may be shown or hidden ('+' icon, circled in yellow).

You can also select only some cells to be copied (use the shift key to select a range of cells).

You can switch to *Identification History* while keeping the focus on a given task. To do that, right click on the task name in the tree, then on '*Go to Mascot task*' / '*Go to OMSSA task*'.

Identification history



The tasks in the tree are only those containing a Peaklist Identification in the workflow. So be aware that task number in the two history tabs may be different for a same task. If a search is run on *n* search engines, then there will be *n* associated identification tasks.

The top-right part (pale orange) is related to the identification task: you can se the search parameter template that was used, and the details of these parameters (useful in case the template has been updated since).

In the bottom-right table, search-related, you can see the details of the peaklist identification progression. For Mascot tasks, as in Mascot Daemon, you can click on a result file name (xxx.dat) to be redirected in your default web browser, on the Mascot result page.

As for *Workflow History*, you can copy the table (see below), show or hide columns, and filter the displayed tasks. You can also switch to the *Workflow History* while keeping the focus on a given task; by right-clicking on the task name in the tree, and then on '*Go to Workflow task*'.

Copy result grids

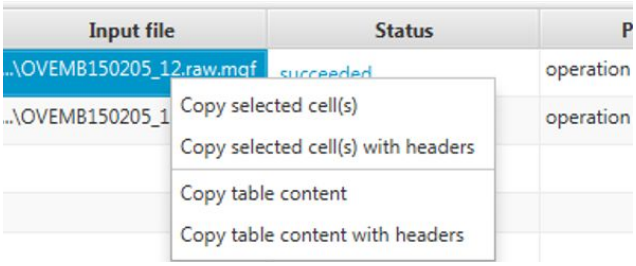
The result grids content can be copied in several ways.

First, select the desired cells using the mouse click, the *shift* key, the arrow keys and/or the *Ctrl+A* combination (selects the whole grid).

Then, you can either:

- Type *Ctrl+C*: the selected cells will be copied, with the columns headers

- Right click on a cell and select one of the copy options. Note that using "*Copy table content*" will also copy hidden columns.



Visualize logs (task events)

Workflow history: Click on a file status to see the details about the progress of the workflow on this file.

In case of error, this is where you should find the explanation of the problem. It will help you or your admin to resolve it.

Msi history: Click on a file status to see the details about the progression of the proteins identification in this file. There are two tabs:

- *Submission trace*: all the logs that were produced by the server and the search engine while processing the file
- *Monitoring trace*: the response the server got each time it asked for the progression of the identification