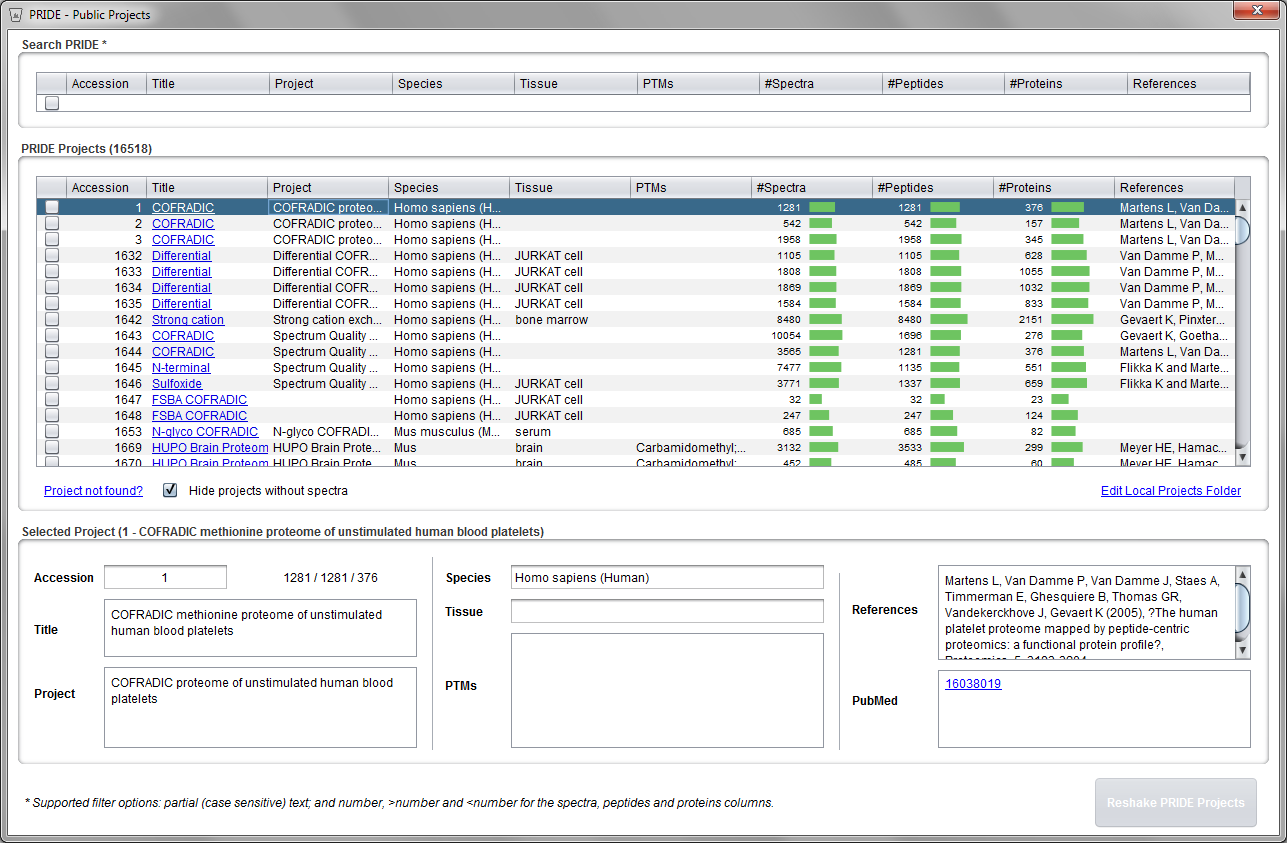
Reprocessing Public Experiments

In the previous chapter, we saw that many proteomics experiments are freely available in online public repositories. It can be very interesting to re-analyze a project of interest, maybe with a different set of modifications or with a different search engine? This is possible via the Reshake function of PeptideShaker. Reshake is found in the PeptideShaker Welcome dialog:



After clicking on ‘Reshake’, you will see the following:



Keen observers will recognize here the list of PRIDE[1](#_ENREF_1) projects from the previous chapter and the projects we inspected on the web interface. In fact, you have here a snapshot of the PRIDE database and can select a project of interest for reprocessing.

**Tip:**  
*You can search for wanted project properties at the top. Simply put in the term or value you are looking for and click Enter. The number fields support simple filters like > and <. Note that the text search is case sensitive, i.e., 'Homo sapiens' is not the same as 'homo sapiens'.*

We are now going to inspect the first project ever loaded in PRIDE, accession 1. *When was this dataset published? What differences do you see with the example of the tutorial? [3.3a]*

Note that the number of PSMs (named peptides in PRIDE) equals the number of spectra. Indeed, for this first upload, only the identified spectra were uploaded. In fact, if you browse the table, you will see that information is missing for many projects, making the reprocessing very difficult. This is one of the reasons why the quality of the dataset annotation is of highest importance when submitting your data – as stressed already in the submission chapter.

Now go to dataset accession 1644. You will see here the same dataset but with all the spectra uploaded: 1281 spectra identified out of 3565 (36%). We are going to see if we can do any better by reprocessing the data. First create a folder which will be used to store this new project.

Select accession 1644 by clicking in the first column and click ‘Reshake PRIDE Projects’ in the lower right corner. PeptideShaker will ask you to provide a folder where the mgf file(s) will be stored, select the previously created folder.

PeptideShaker now downloads the PRIDE file corresponding to this project and extract the spectra into an mgf file. If this process should fail, e.g., due to network issues, note that you can also download the files directly from the PRIDE website. If you cannot access internet, the files are provided in the resources folder of the tutorial.

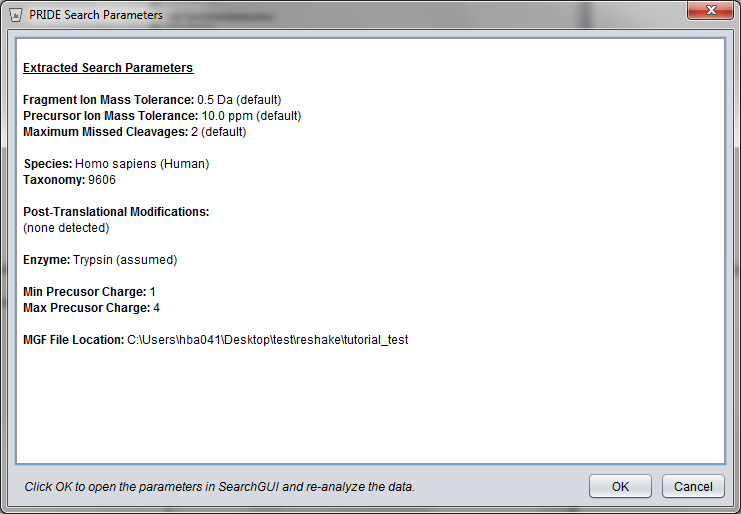
**Tip:**  
*Local or private PRIDE XML files can be added to the table be storing the files in a given folder and then use the 'Edit Local Projects Folder' link found to the bottom right of the table. Your local PRIDE XML files will then appear in the table and are now available for re-analysis.*

Once the spectra have been extracted, PeptideShaker will ask you to provide a protein database, use the same target/decoy database as for the identification tutorial: uniprot\_human\_reviewed\_ no\_isoforms\_october\_2013\_concatenated\_target\_decoy.fasta, and click OK.

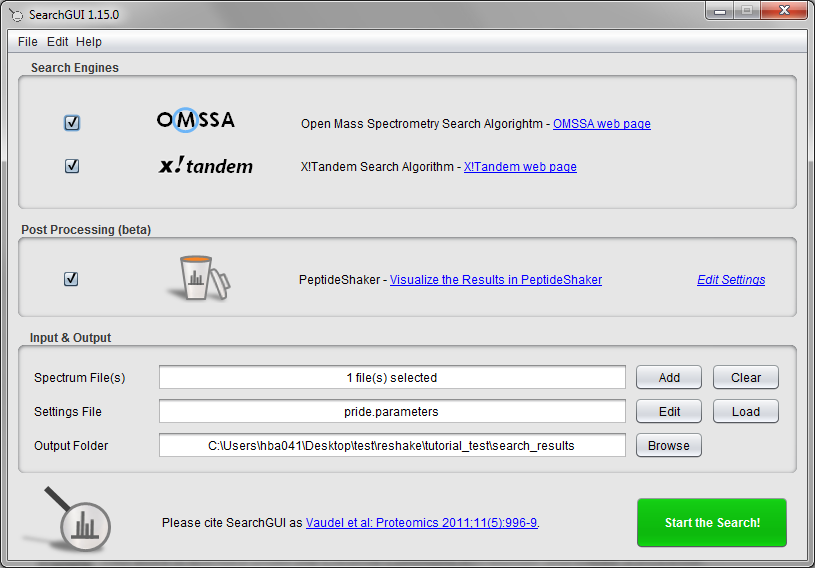
Note that PeptideShaker recognized that it was a human sample:



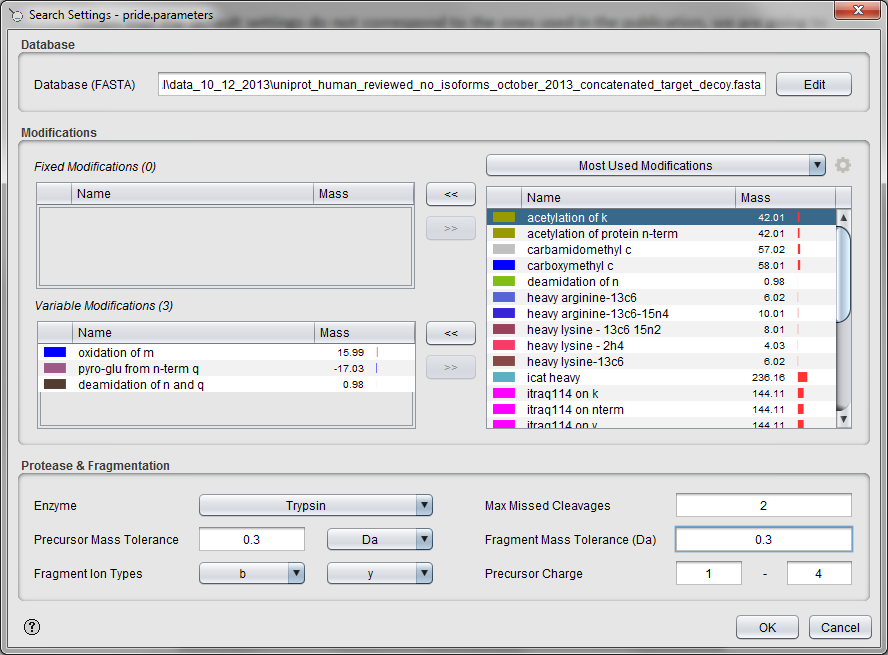
PeptideShaker will also look for the search settings used to generate the file. In many cases however (most notably for older submissions), the complete search settings were not provided by the user. In such cases, default values are suggested



After clicking ‘OK’, SearchGUI is automatically started. Note that the mgf file is already selected and the search parameters are pre-set:

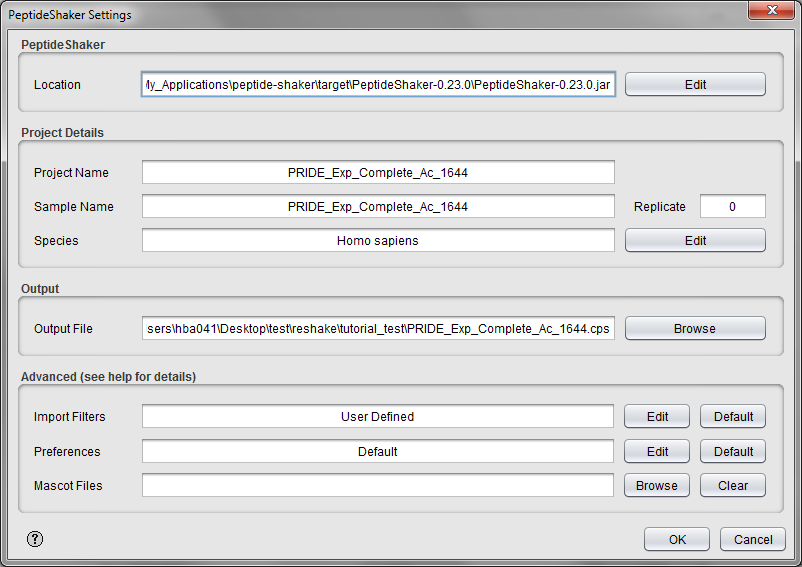


Given that the default settings do not correspond to the ones used in the publication, we are going to change them accordingly. Click the ‘Edit‘ button next to the Settings File field and select ‘oxidation of m’, ‘pyro-glu from n-term q’ and ‘deamidation of n and q’ as variable modifications, and change both precursor and fragment ion tolerances to 0.3 Da. You should now have the following settings:



Click 'OK' to close the Search Settings dialog and chose yes to the question about saving the new settings.

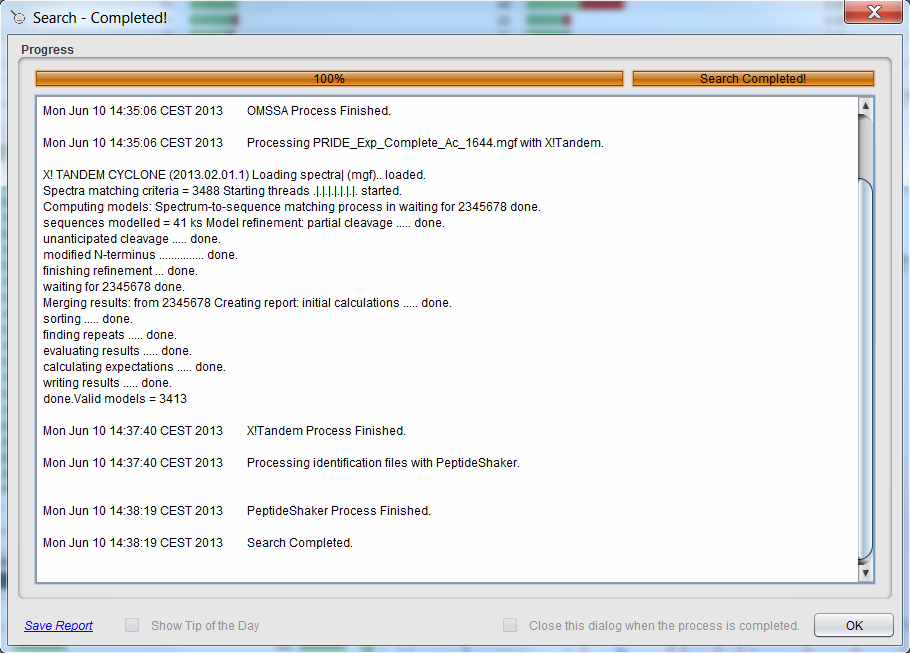
We are also going to start PeptideShaker directly once the search is finished. (Note that this feature is still under testing. But in case of issues you can always load the search results manually in PeptideShaker.) Select ‘Edit Setting’ next to the PeptideShaker post-processing option:



Note that default settings have already been inserted, and your PeptideShaker project will be saved automatically to the chosen location.

Click 'OK' to close the PeptideShaker Settings dialog and go back to the SearchGUI main dialog and click 'Start the Search!'.

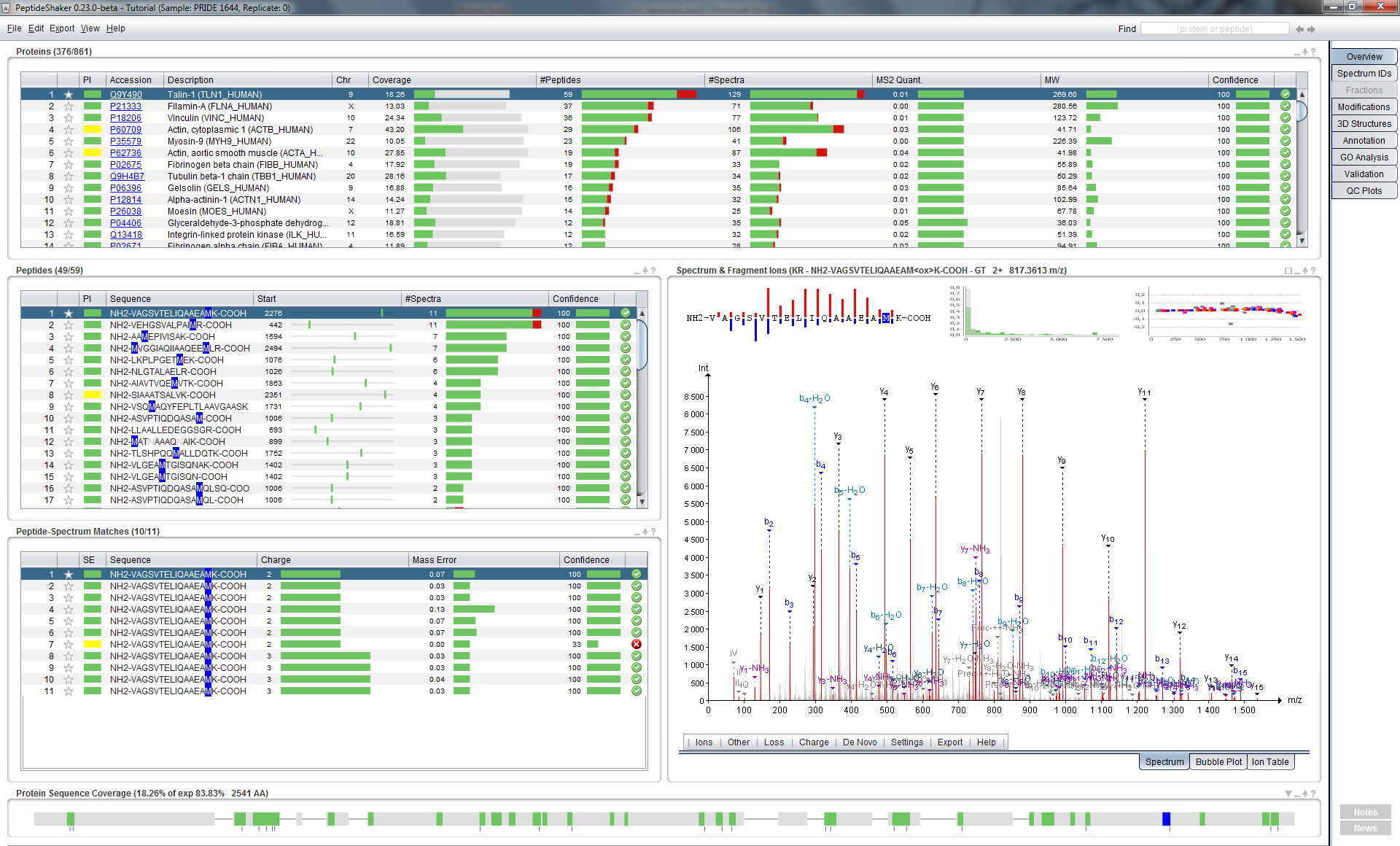
As usual, you will be updated on the progress during the processing:



Note that the complete reprocessing of this dataset can be performed on a regular laptop and does not require any advanced informatics skills.

**Tip:**  
*Run demanding searches and automated post-processing overnight!*

After loading in PeptideShaker you should see the following results:



*After this simple reprocessing, what is the new identification rate? [3.3b]*

Given that different validation methods were employed in the original manuscript and during reprocessing, we can obviously not compare the two identification rates directly.

The real interest of the reshake feature is that you can now investigate this dataset as if it were your own. For instance, you can look for a particular protein or modification.

References

1. Martens, L. et al. PRIDE: the proteomics identifications database. *Proteomics* **5**, 3537-3545 (2005).

2. Martens, L. et al. The human platelet proteome mapped by peptide-centric proteomics: a functional protein profile. *Proteomics* **5**, 3193-3204 (2005).

3. Griss, J. et al. Consequences of the discontinuation of the International Protein Index (IPI) database and its substitution by the UniProtKB "complete proteome" sets. *Proteomics* **11**, 4434-4438 (2011).

4. Martens, L. et al. Do we want our data raw? Including binary mass spectrometry data in public proteomics data repositories. *Proteomics* **5**, 3501-3505 (2005).