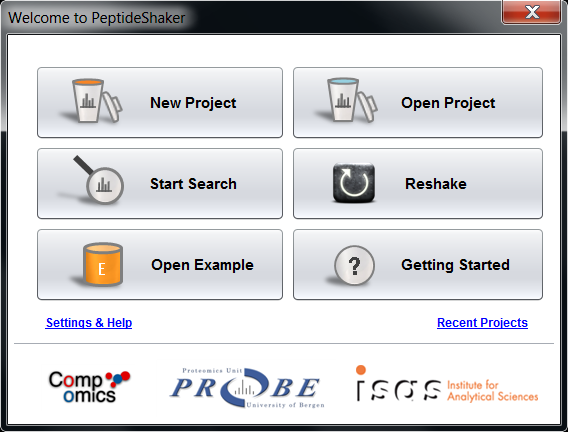
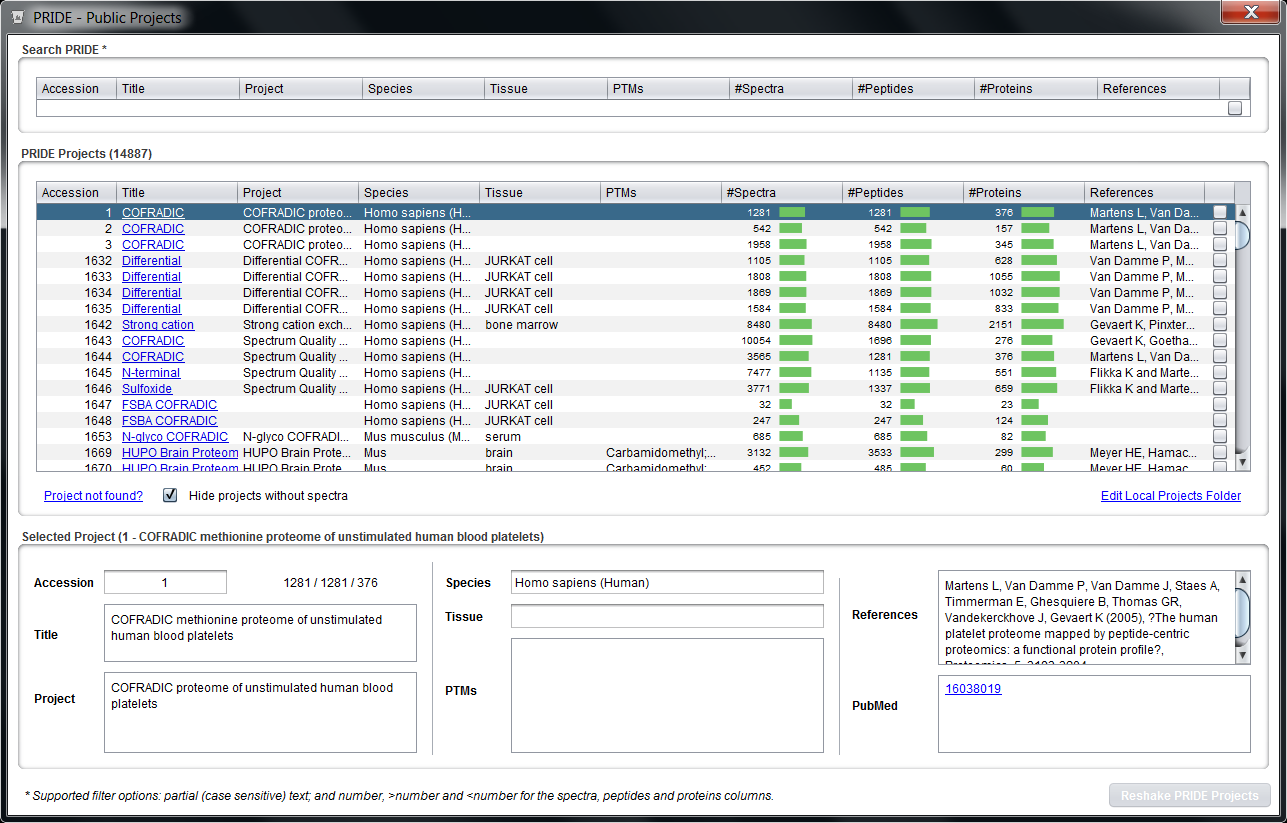
Reprocess Public Experiments

In the previous chapter, we have seen that many experiments are freely available online in public repositories. It can be very interesting to research a project of interest, maybe with different modifications? This is possible via the Reshake function of PeptideShaker.



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



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 database and can select a project of interest for reprocessing.

We are now going to reshake 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?*

*This dataset was part of a publication from 2005 as displayed in the References panel.*[*2*](#_ENREF_2) *Reshaking it will hence give us an impression of what changed in Proteomics since then.*

*One of the major differences comes from the instrumentation: 3,565 MS2 spectra were generated and searched with a tolerance of 0.3 Da. In comparison, the example dataset of the tutorial counts 11,332 MS2 spectra (measured over a longer gradient however) searched with a tolerance of 10ppm/0.01 Da. Since 2005, the resolution of the instrument was hence multiplied by more than 10 without decreasing the scan time.*

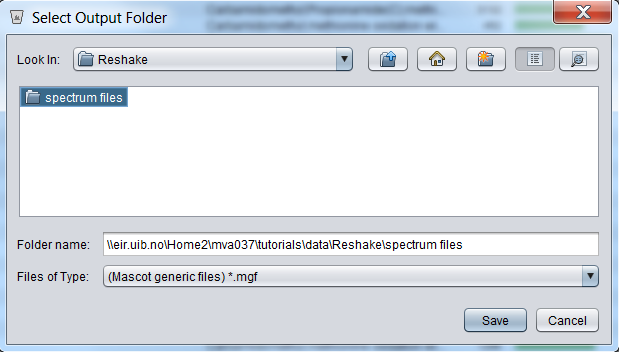
*Secondly, the sequence database used was the International Protein Index (IPI) which does not exist anymore and is now included into UniProt.*[*3*](#_ENREF_3) *You will also notice that the original data interpretation pipeline is complex and requires good computational skills. Especially, there was no user friendly interface allowing the intuitive browsing of protein, peptides and spectra. Finally, note that there is no estimation of the error rate.*

*Finally, you will observe that this project has the same number of spectra than peptides. In fact, only the identified spectra were uploaded then. It is now required to provide all the raw data for publication – this will be further discussed in the tutorial.*[*4*](#_ENREF_4)

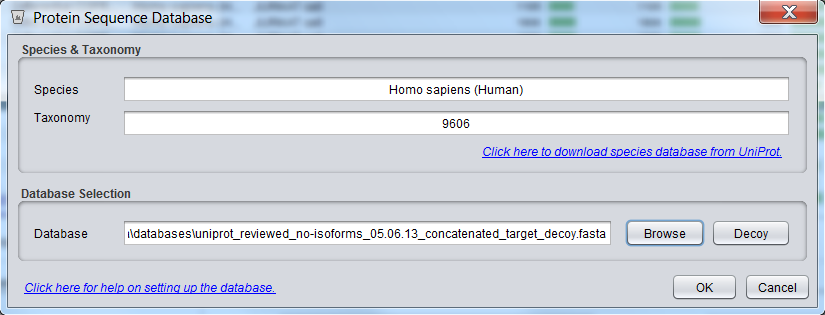
*In the end, the only thing that did not change is the search engine. It is actually quite an issue of the field, we are looking at high resolution data with tools designed on low resolution. Hence, new algorithms specifically designed for high resolution mass spectrometers are being developed and will be included into the present tutorial as soon as technically possible.*

Some of you will have noticed 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 can be 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 1,644, you will see here the same dataset with all spectra uploaded. As you can see, 1,281 spectra were identified out of 3,565 (36%), we are going to see if there is any improvement when reprocessing the data. Create a folder which will be used to store this new project.

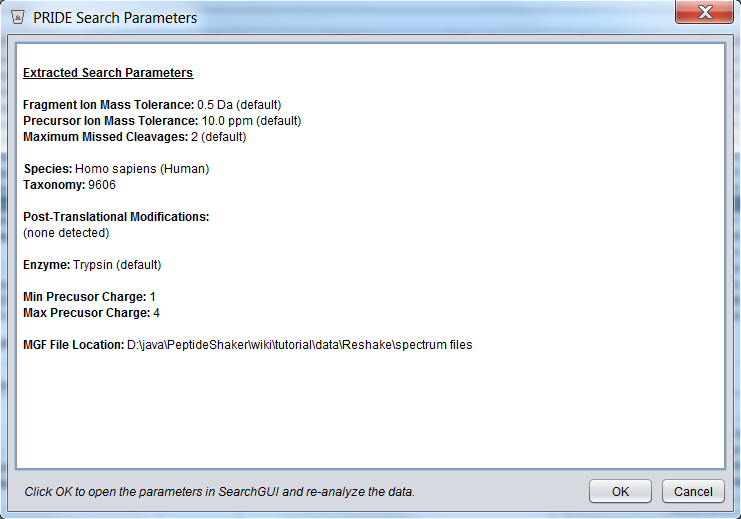
Select accession 1,644 at the end of the line and click ‘Reshake PRIDE Projects’, PeptideShaker will ask you to provide a folder where the mgf files will be stored, select the folder created previously:



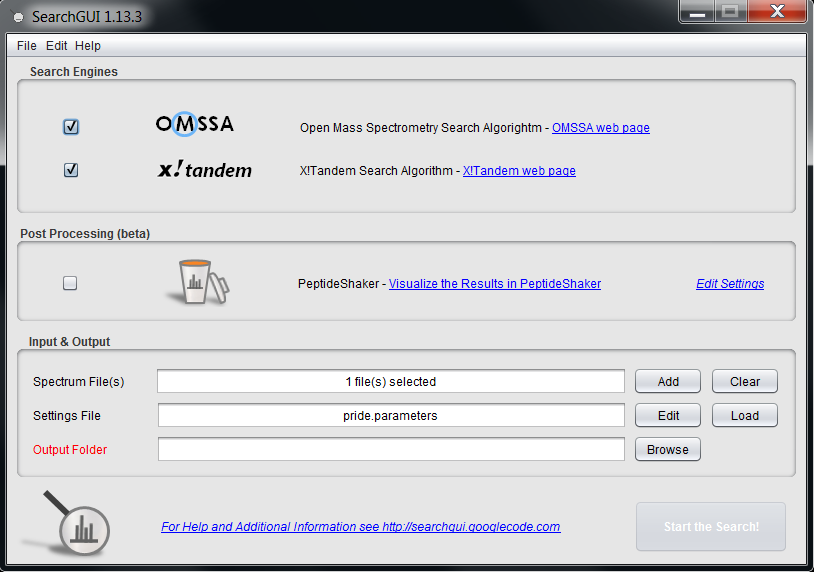
PeptideShaker will now download the pride file corresponding to this project and extract the spectra as mgf file. If this process fails, note that you can download the files yourself from the PRIDE website. If you cannot access internet, the files are provided in the resources folder of the tutorial. Once the spectra are extracted, PeptideShaker will ask you to provide a protein database, use the same target/decoy database as for the identification tutorial: uniprot\_reviewed\_no-isoforms\_05.06.13\_concatenated\_target\_decoy.fasta.



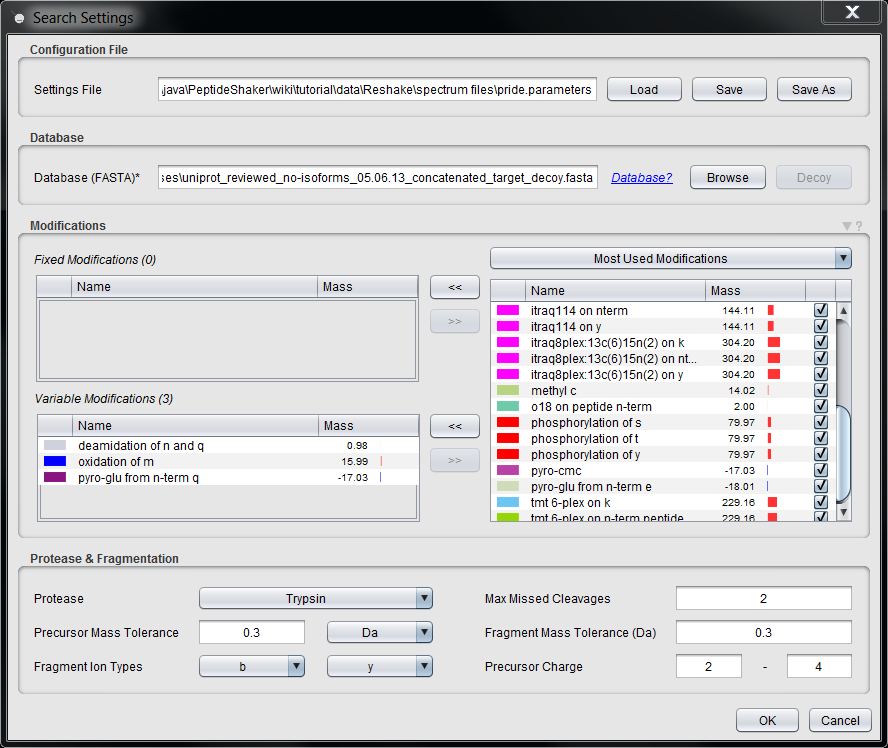
Note that PeptideShaker recognized that it was a human sample. It will also look for the search settings used to generate the file. In most cases however, the search settings were not provided by the user. In such cases, default values are suggested:



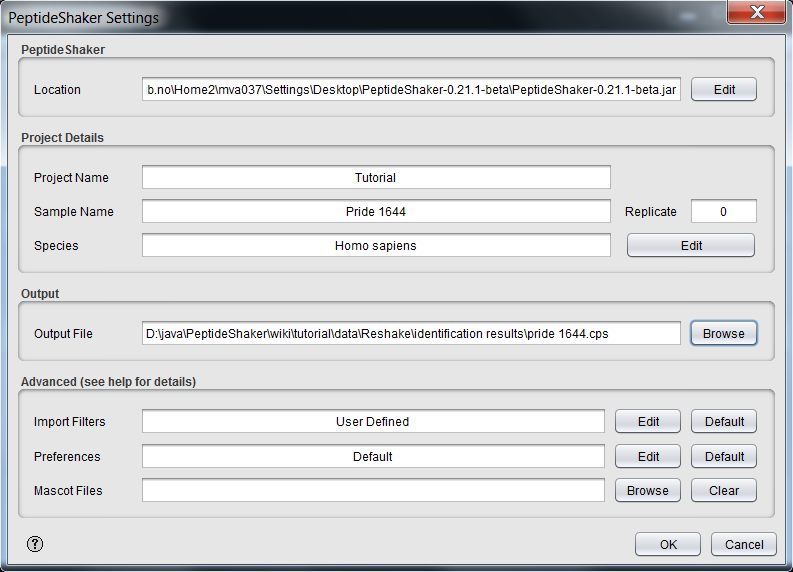
After clicking ‘OK’, SearchGUI is automatically started. Note that the mgf file is selected and the search parameters pre-set:



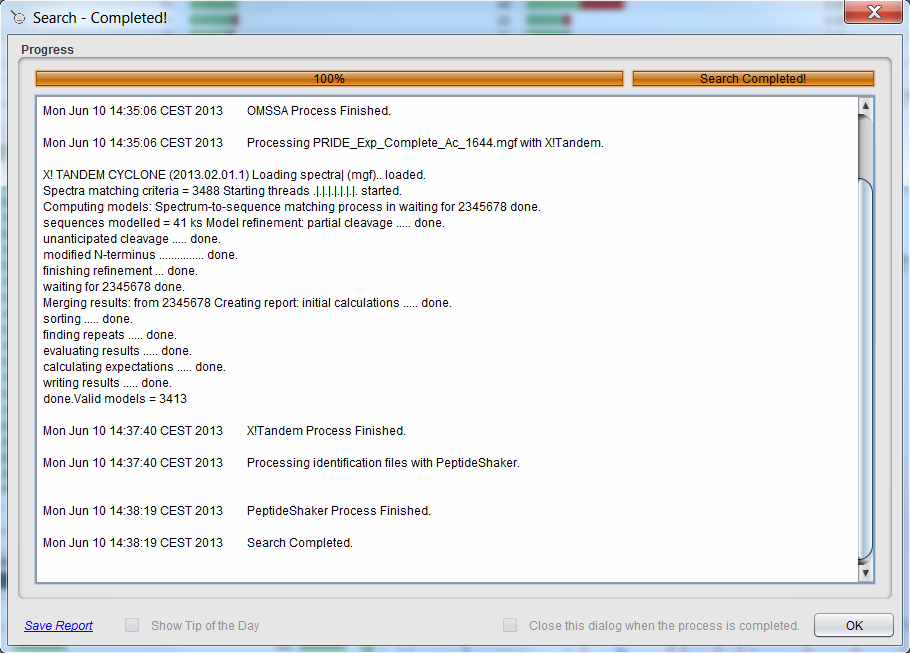
Since the default settings do not correspond to the ones used in the publication, we are going to change them accordingly: select ‘oxidation of m’, ‘pyro-glu from n-term q’ and ‘deamidation of n and q’ as variable modifications and change both precursor and fragment ions tolerances to 0.3 Da:



We are also going to start PeptideShaker directly once the search is finished. Note that this feature is still under testing and that in case of failure you can follow the manual project creation of the identification tutorial. Select ‘PeptideShaker’ in post-processing and add your project details:



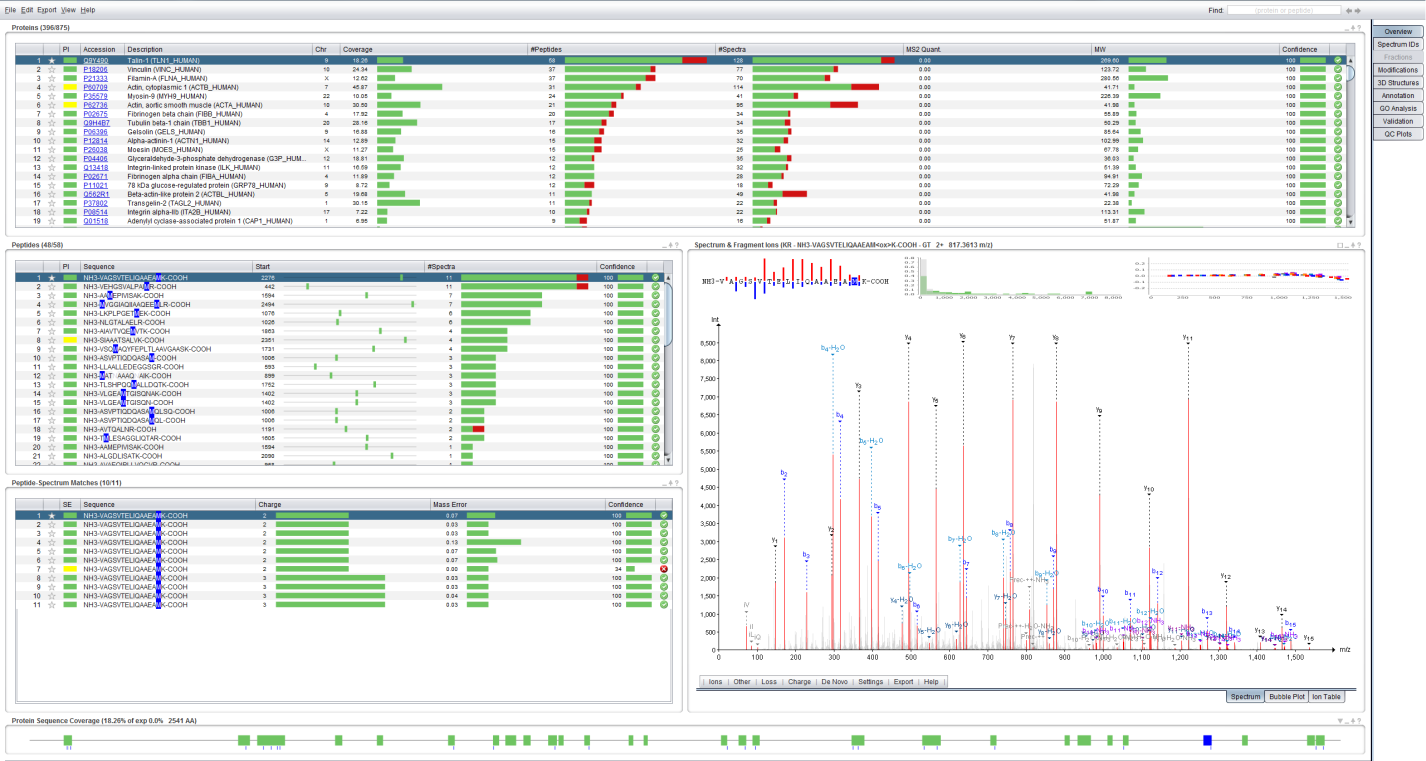
Note that your PeptideShaker project will be saved automatically in the chosen file. Once this is finished, start your search:



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

Note that the complete reprocessing of this dataset took only three minutes on a regular laptop.

After loading in PeptideShaker you should see the following results:



*After this simple reprocessing, what is the new identification rate?*

*If you select the “Spectrum IDs” tab, you will see that 1,768 spectra out of 3,565 (49.6%) identified at 1% FDR.*

Since different validation methods were employed in the original manuscript and during reprocessing, we can obviously not compare the two identification rates. The real interest of the reshake feature is actually that you can investigate this dataset as if it were yours. 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).