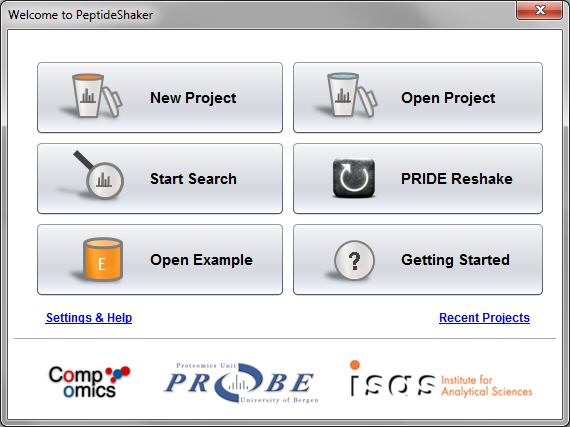
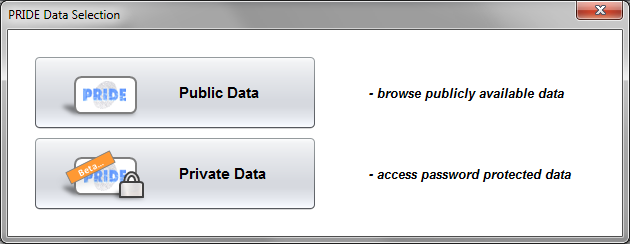
Reprocessing Public Experiments

In the previous chapter, we saw that numerous proteomics experiments are freely available in online public repositories. It can be useful 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 PRIDE Reshake function of PeptideShaker. PRIDE Reshake is found in the PeptideShaker Welcome dialog:

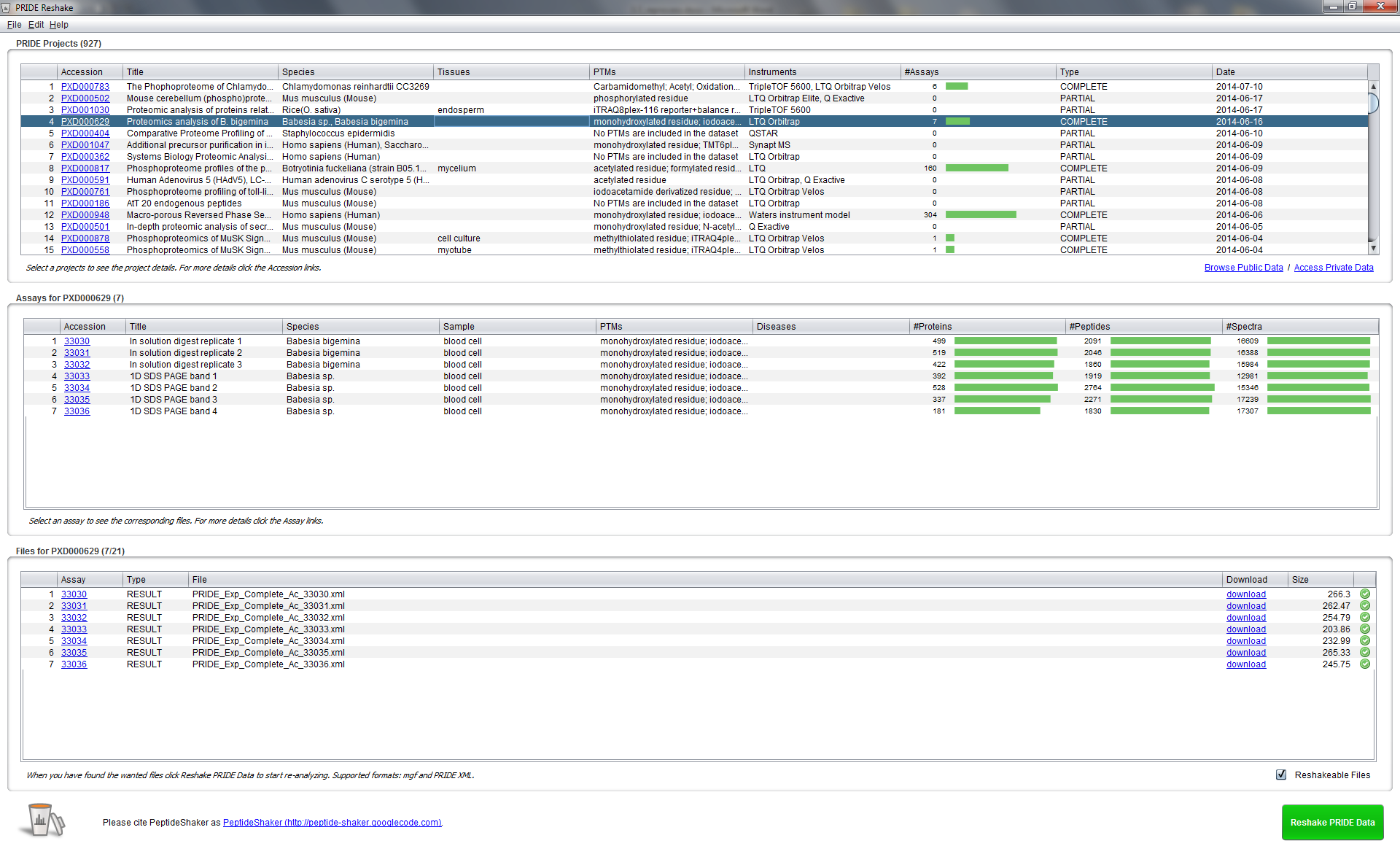


After clicking on ‘PRIDE Reshake’, you will see the following dialog:

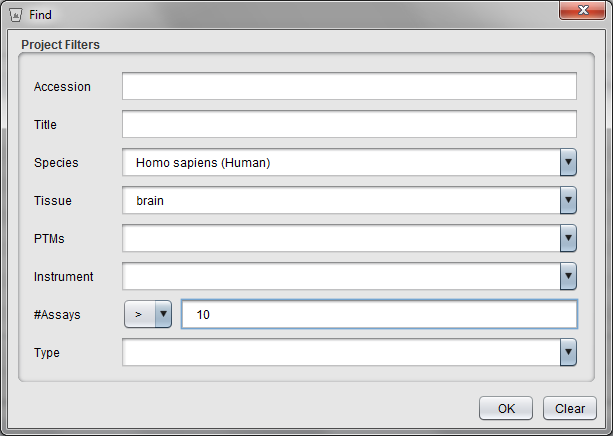


Here you can choose between reprocessing public data or private password protected data, for example as part of a reviewing process. For now click the ‘Public Data‘ option.

This will display the list of publicly available projects in PRIDE[1](#_ENREF_1). By clicking a project you will see the accociated assays and data files, here using the project PXD000629 as an example:



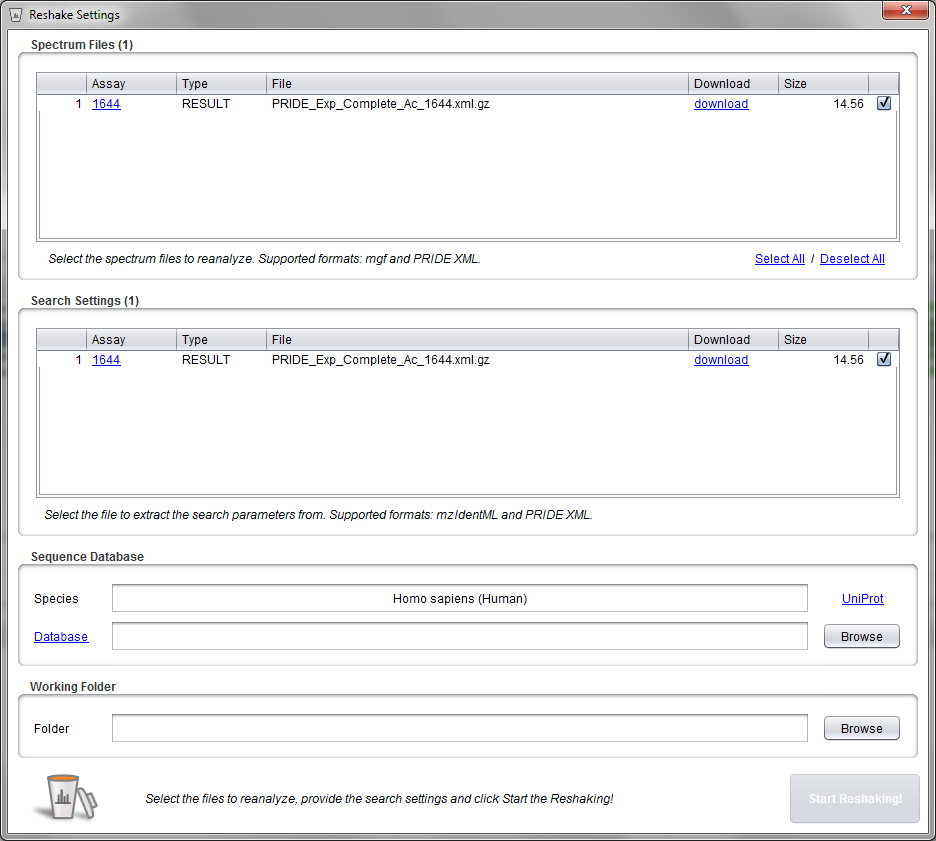
You can search for particular projects of interest using the Find feature located in the Edit menu (or by clicking Ctrl+F). For example:



We are now going to inspect the first project ever loaded in PRIDE, accession 1. The dataset has the accession PRD000001. Click the project accession link for more details about the project. *When was this dataset published? What differences do you see with the example of the tutorial? [3.3a]*

Note that the project contains five assays, named Assay 1-3 and Assay 1643-1644. The reason for this is that for this first submission, only the identified spectra were uploaded. In fact, if you browse other projects, 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.

We are now going to reprocess Assay 1644. This is the same dataset as Assay 1, 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. Then select Assay 1644 in the Assays table, and click the green ‘Reshake PRIDE Data’ button in the lower right corner. A new dialog will then appear, allowing you to customize the settings for the reprocessing:



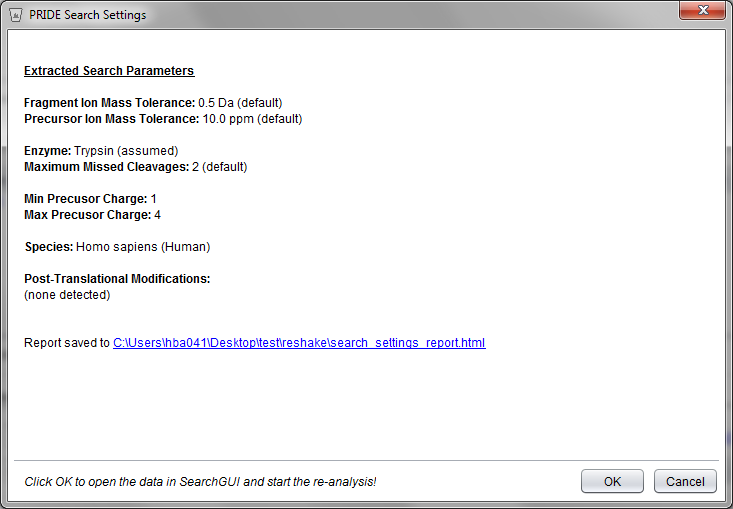
At the top you now the possible source(s) for the spectrum data: PRIDE XML or mgf files. We will here use PRIDE XML file. For the search settings the options are PRIDE XML or mzIdentML files. Again we here only have one option, the PRIDE XML file.

Next, we select the database to search against. The species annotated by the submitters of the dataset is displayed. For single species datasets you can click the ‘Download from UniProt‘ link to easily get the correct database from UniProt as explained in Chapter 1.1. For now you can simply use the same target/decoy database as for the identification tutorial: uniprot-human-reviewed-october-2014\_concatenated\_target\_decoy.fasta.

Finally, you have to select the working folder. This is where the downloaded project files are stored and where the search results will appear (unless you change this later in SearchGUI). Select the folder you created earlier, or create a new empty folder.

Then we are ready to start the reprocessing. As indicated by the ‘Start Reshaking! ‘ button turning green! Click the button to start.

As usual you will be informed of the progress. When the selected files have been downloaded and converted to the correct format, a report with the extracted search parameters will be shown (the report is also saved to the working folder):

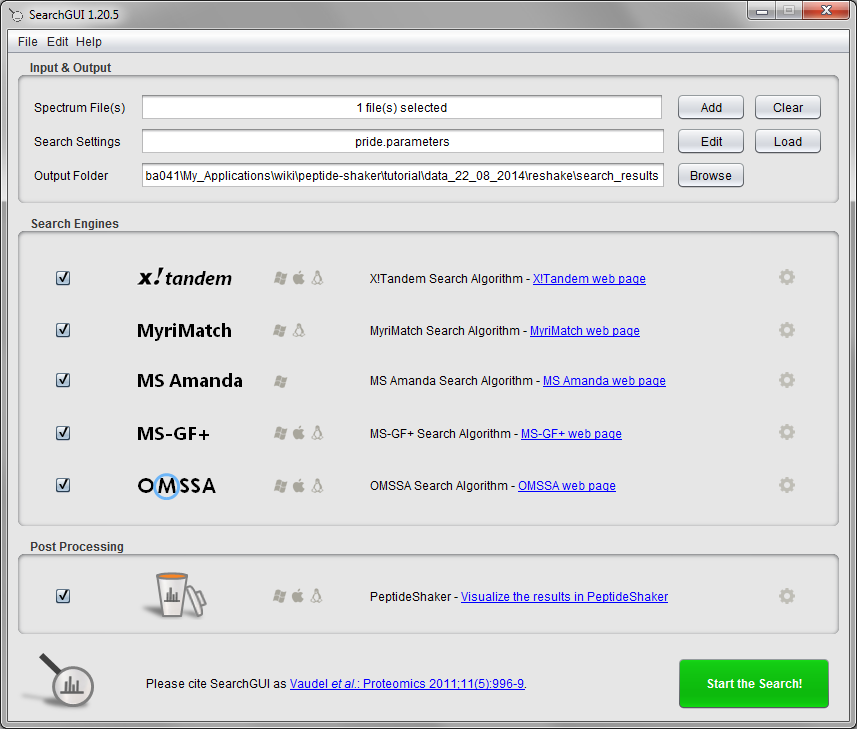


Note that in many cases (most notably for older submissions), the complete search settings were not provided by the user. In such cases, default values are suggested.

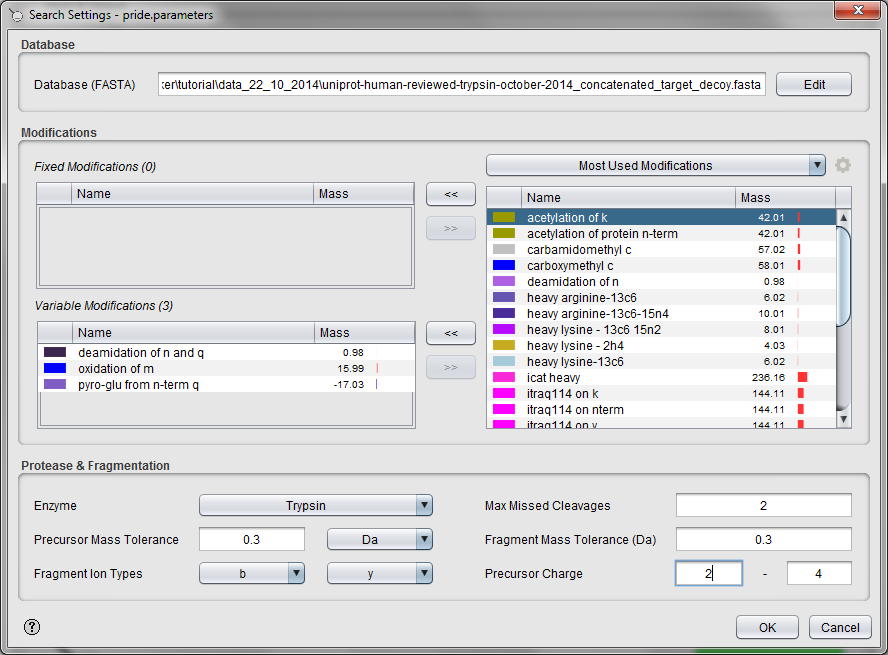
**Tip:**  
*If the download of the files should fail, e.g., due to network issues, you can also download the files directly from the PRIDE website. If you cannot access internet, the files are provided in the resources folder of the tutorial.*

Click ‘OK‘ in the PRIDE Search Setting Dialog to close PeptideShaker and open SearchGUI to start the reprocessing.

SearchGUI will now open with all the settings and files already inserted:

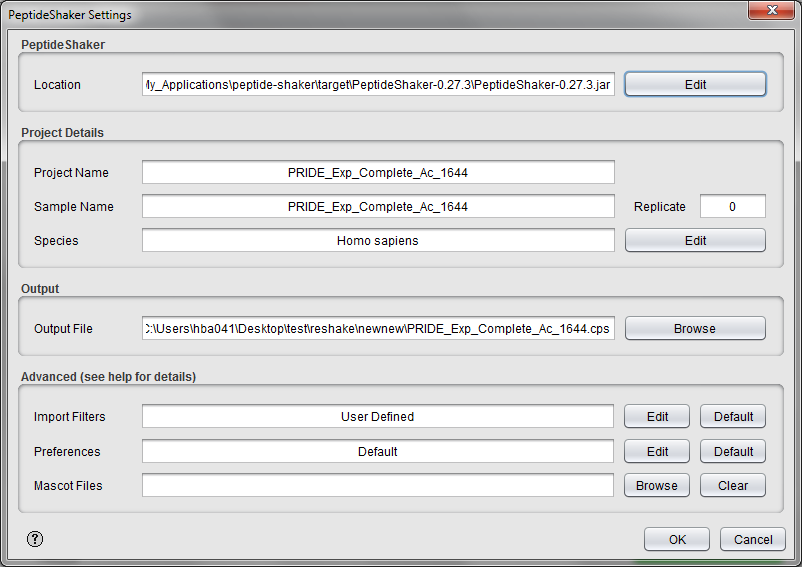


Given that the default settings do not correspond to the ones used in the publication, we are going to change them accordingly. Click the ‘Edit‘ button next to the Settings File field and select ‘oxidation of m’, ‘pyro-glu from n-term q’ and ‘deamidation of n and q’ as variable modifications, change both precursor and fragment ion tolerances to 0.3 Da and the minimum precursor charge to 2. You should now have the following settings:



Click ‘OK‘ to close the Search Settings dialog and choose ‘Yes‘ to the questions about saving the new settings.

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

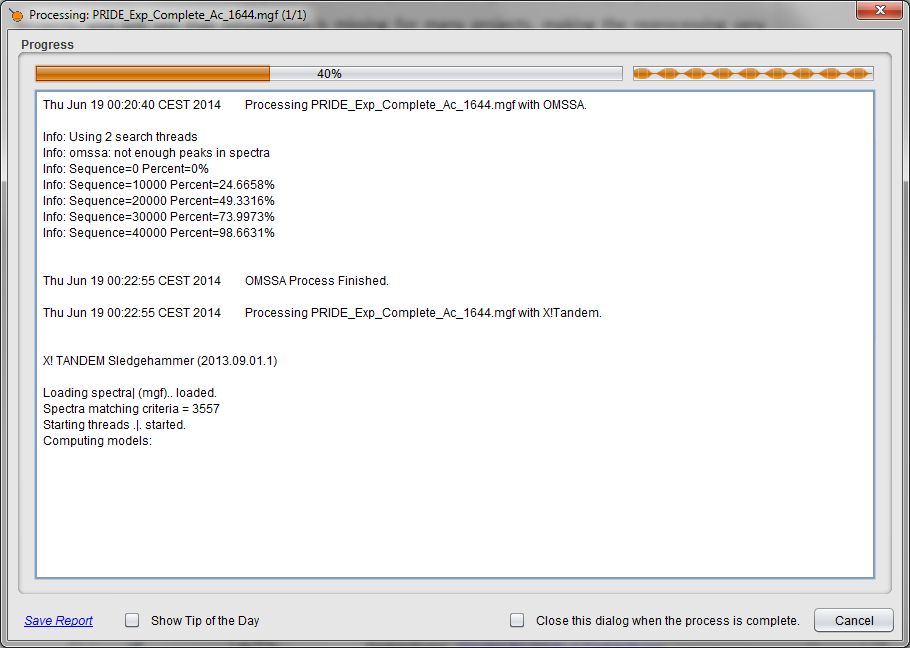


Note that default settings have already been inserted and your PeptideShaker project will be saved automatically to the chosen location.

Click ‘OK‘ to close the PeptideShaker Settings dialog and go back to the SearchGUI main dialog.

To save time, disable MyriMatch, MS Amanda, MS-GF+ and Comet, and then click ‘Start the Search!‘.

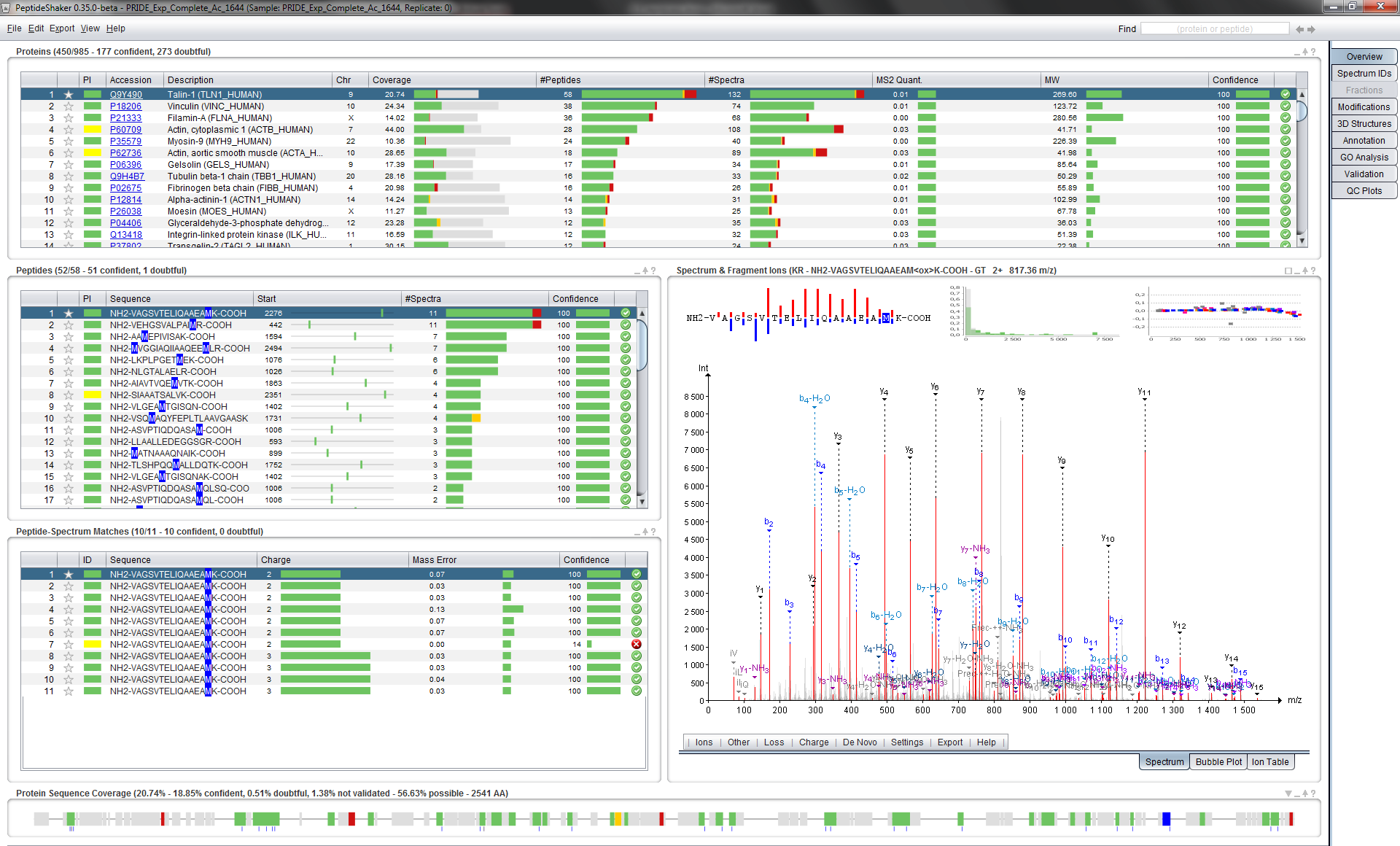
As usual, you will be updated on the progress during the processing:



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).