PTM Analysis

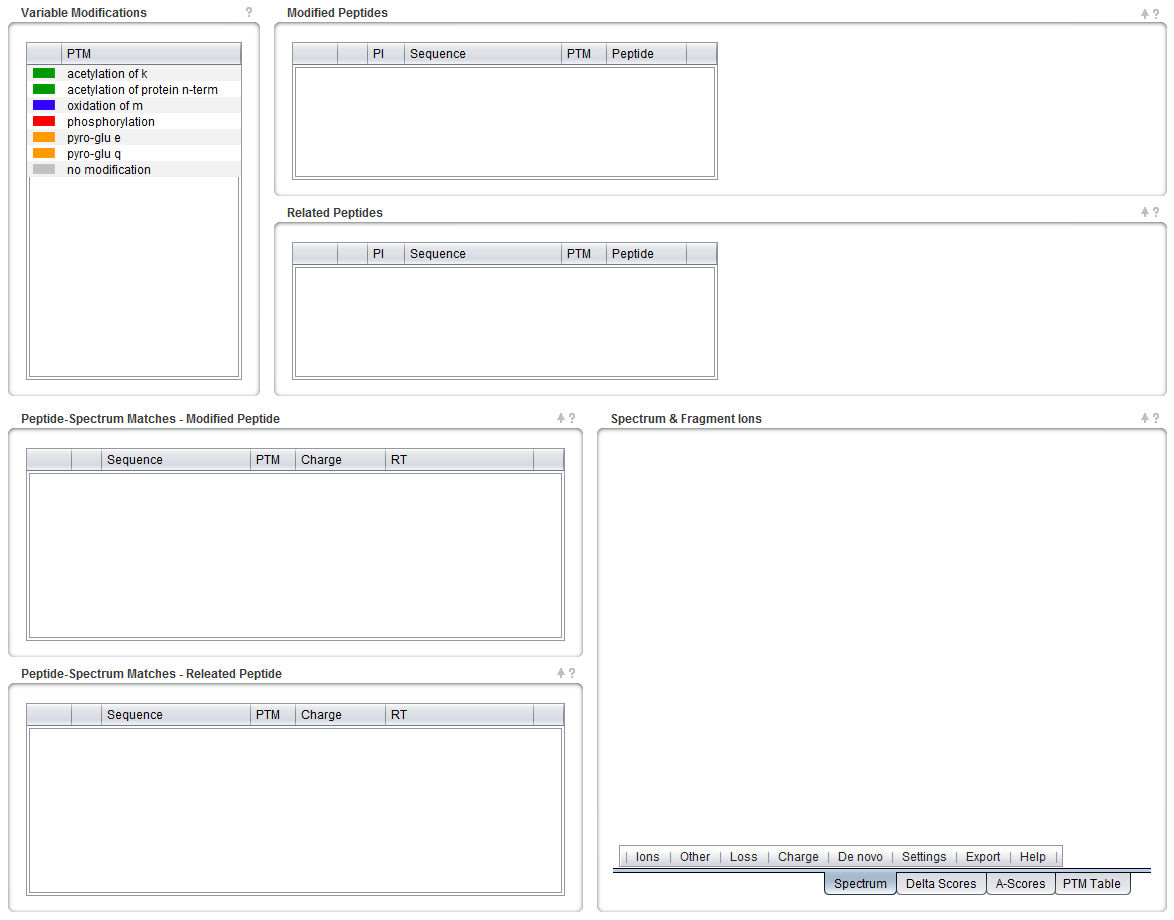
A major challenge in proteomics is the identification of Post Translational Modifications (PTMs). Among the known PTMs, perhaps the most studied is the phosphorylation. PTMs are detected in spectra by the shift in mass they induce. However, the actual position of a phosphorylation in a peptide is not straightforward to confidently assess, simply because these is very little in the spectra that can help us discriminate between potential modification sites.

The 'Modifications' tab in PeptideShaker is dedicated to the study of PTMs. Note that this part of PeptideShaker is under development and will be subject to improvement in the coming month. And as always, any comments are welcome!

In PeptideShaker load the project named ‘phosphorylated\_peptides.cps’ located in the resources folder. This small dataset consists of synthetic peptides among which some are phosphorylated. You will see in the 'Validation' tab that 188 peptides were validated at 1% FDR.

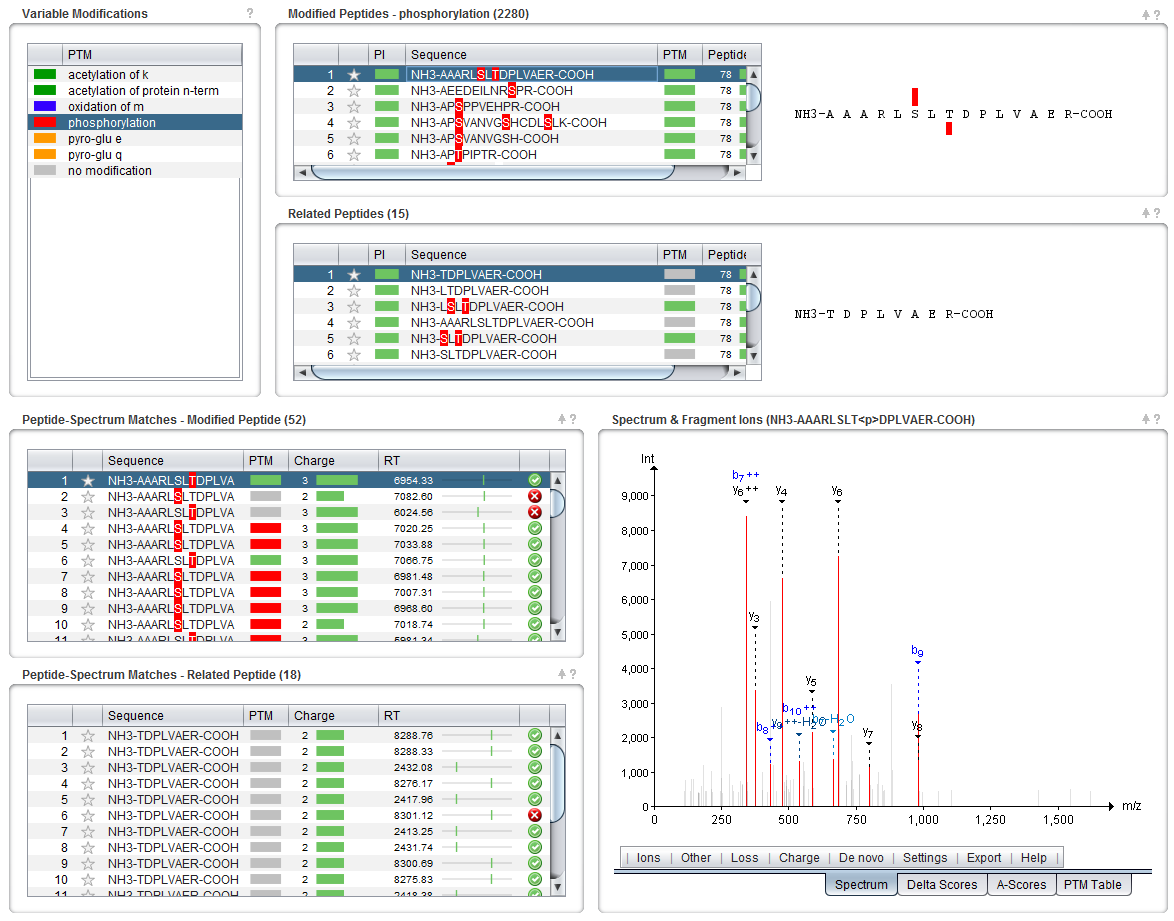
*What do you think of the quality of the statistical estimation? And at the PSM/protein levels?*

Selecting the ‘Modifications’ tab, you should see the following:



On the top left, you can see a list of all the variable modifications found in the results. Note that all PTMs are color coded in PeptideShaker and you can change all colors. *We only looked for oxidation and phosphorylation, why are other modifications present in the results?*

We will focus here on phosphorylations. Note that the method applies to all modifications inducing a mass difference in peptide fragments. Select ‘phosphorylation’ in the top left table and select the first peptide in the top table, the display should now look like this:

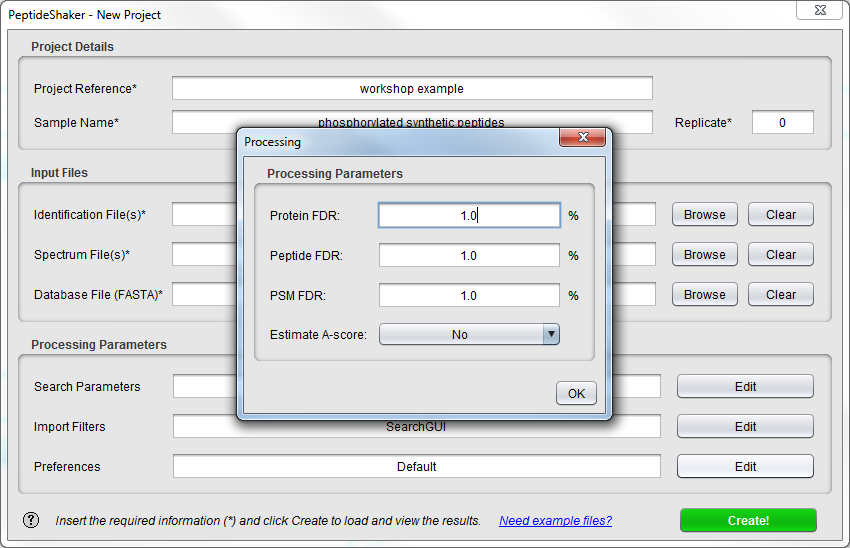


Near the list of variable modifications, you will see two tables containing peptides. The modified peptides table contains all the peptides in the dataset containing the desired modification. Next to the table, the sequence is represented with two bars representing the PTM scores. These scores try to assess the location site.

Two scores are available in PeptideShaker (see More on PTM Scoring at the end of the tutorial):

1. The **delta score** which is simply the difference in confidence between the two best modification locations according to all search engines used.
2. The **A-score**[1](#_ENREF_1), which selects the two best modification locations according to a built-in score and compares them.

Warning: The A-score is turned off by default in PeptideShaker since it can be very time consuming. When creating a project where the A-score is wanted, you can turn it on in the New Project dialog by editing the Processing Preferences at the bottom of the dialog:



In the present example, AAARLSLTDPLVAER, the delta score returns Serine 6 as most likely modification position with a score of 100. The A-score votes for the Threonine 8 with a score of 148.84. *Which possibility would you validate?*

Note that PeptideShaker returned both possibilities as possible phosphorylation sites. These are thus painted in red on the sequence. Under the peptide table, you will see a table grouping so-called related peptides. The related peptides are peptides with a sequence similarity with the peptide selected, e.g., the same sequence but unmodified, or partly overlapping peptides. In general, the presence of related peptides increases the confidence in the peptide identification: it is unlikely that many related peptides were found by chance. Moreover, the modification status of one peptide might help resolve the modification status of another. In the present example, you see the non-phosphorylated version of the peptide. We also see some unexpectedly cleaved peptides. *Looking at the related peptides, which phosphorylation site is the most likely to be in the sample?*

The PSMs of the selected peptides are listed at the bottom left of the screen. As you see, both phosphorylation sites were retrieved by the search engine among 52 spectra. Have a look at the retention times and sort the PSM table according to this metric. *How does the retention time of a spectrum help you here? What about the charge?*

*What are the ions that would prove the phosphorylation to be on the Serine or the Threonine*?

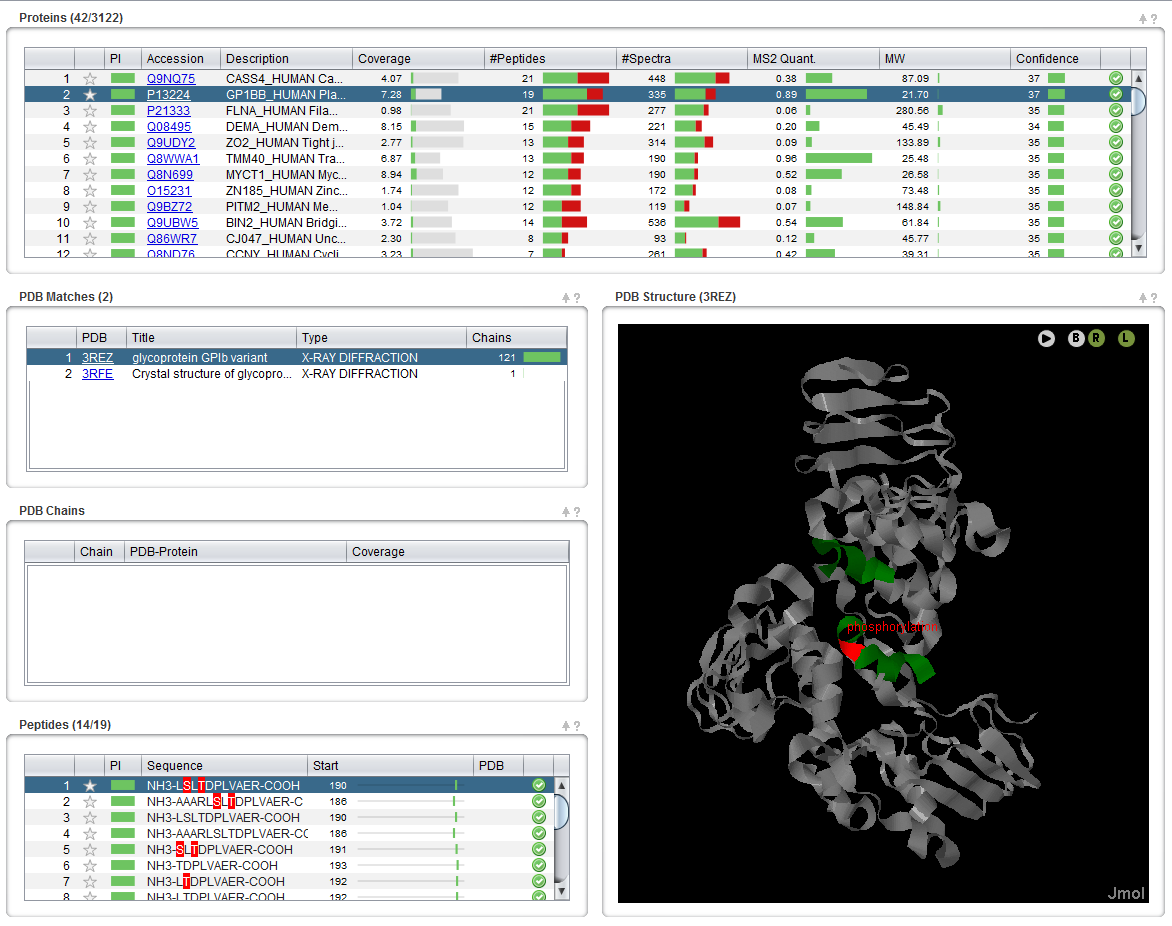
Such ions are called site determining ions. *Do you find them in the selected spectra?*

Go back to the 'Overview' tab. If you look at the sequence coverage display at the bottom, you will see that the protein where this peptide was found has 19 detected peptides, all at the same place in the sequence:



*How many of these peptides are phosphorylated? How many phosphorylation locations do we have in the end?*

Make sure that the example peptide is selected in the 'Overview' tab, and then move to the '3D Structures' tab. When the PDB Data is finished loading, select the second PDB match (Crystal structure of glycoprotein GPIb ectodomain) in the 'PDB Matches' table. You should then see the following:



*Can you see the phosphorylation sites on the structure?*

More on PTM Scoring

Two methods are commonly used in proteomics for PTM site localization: probabilistic scoring and secondary hit comparison. The former is applied by three of the most famous scoring algorithms: MaxQuant’s PTM score[2](#_ENREF_2), A-score[1](#_ENREF_1) and PhosphoRS[3](#_ENREF_3" \o "Taus, 2011 #123) (for a more exhaustive review of algorithms, see Beck et al.[4](#_ENREF_4)). Probabilistic scoring relies on a simple binomial model to establish the probability that a modification is carried by the candidate amino-acid: every theoretical fragment ion (typically an a, b or y-ion) is considered as an independent trial that is successful if that ion was found in the considered peaks of the spectrum. With number of theoretical fragments (*i.e.* number of trials), number of fragments actually matched (*i.e.* number of successes) and a being the probability of matching a peak, the binomial probability is given by

(2)

This represents the probability of randomly matching the given number of fragment ions in the spectrum. A-score and PhosphoRS estimate the cumulative probability to match at least theoretical ions:

(3)

The secondary hit score comparison approaches on the other hand rely on the rationale that when a search engine is unsure about the location of a PTM, it will retrieve the second best location as a secondary PSM with a very similar score to the top hit. For every spectrum of a modified peptide, the PTM location score is then simply based on the comparison of the best candidate score with the score of the best secondary hit presenting the same sequence but another modification site[5](#_ENREF_5),[6](#_ENREF_6). An example of this approach is the Mascot Delta Score[5](#_ENREF_5)*.*

References

(1) Beausoleil, S. A.; Villen, J.; Gerber, S. A.; Rush, J.; Gygi, S. P. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat Biotechnol* **2006**, *24*, 1285.

(2) Olsen, J. V.; Blagoev, B.; Gnad, F.; Macek, B.; Kumar, C.; Mortensen, P.; Mann, M. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **2006**, *127*, 635.

(3) Taus, T.; Kocher, T.; Pichler, P.; Paschke, C.; Schmidt, A.; Henrich, C.; Mechtler, K. Universal and confident phosphorylation site localization using phosphoRS. *J Proteome Res* **2011**, *10*, 5354.

(4) Chalkley, R. J.; Clauser, K. R. Modification site localization scoring: strategies and performance. *Mol Cell Proteomics* **2012**, *11*, 3.

(5) Savitski, M. M.; Lemeer, S.; Boesche, M.; Lang, M.; Mathieson, T.; Bantscheff, M.; Kuster, B. Confident phosphorylation site localization using the Mascot Delta Score. *Mol Cell Proteomics* **2011**, *10*, M110 003830.

(6) Beck, F.; Lewandrowski, U.; Wiltfang, M.; Feldmann, I.; Geiger, J.; Sickmann, A.; Zahedi, R. P. The good, the bad, the ugly: validating the mass spectrometric analysis of modified peptides. *Proteomics* **2011**, *11*, 1099.