Analysis of Glyco Data- Extended

The data from the mapping of glycoproteins and –peptides required some manual interpretation to decide which peptides that were true glycopeptides, and which were unspecifically bound to the hydrazide beads during the protocol. We decided that the criteria for a peptide to be a true glycopeptide was that it had to have one or more deamidated asparagines and contain the N-glyco sequence motif [N][X^P][ST], and in addition a deamidation could not be confidently assigned only to asparagines which were not in the N-glyco sequence motif.

Although deamidation of asparagine can happen spontaneously, a deamidated asparagine in the N-glyco motif strongly indicate the removal of an oligosaccharide chain from the peptide by the action the enzyme Peptide-N-Glycosidase F (PNGase F), leading to a conversion of asparagine to aspartic acid, especially when glycopeptide enrichment (SPEG) is performed.

Peptideshaker (http://peptide-shaker.googlecode.com) provides the position of the deamidated asparagine in the peptide sequence and also assigns a location confidence. The location confidence tells us how confident the software is in assigning the modification to the specific amino acid: *very confident*, *confident*, *doubtful* or *random*. If a peptide has the same amount of deamidations as there are asparagines in the sequence, the location confidence is *very confident*. However, if the number of asparagines in the peptide sequence is higher than the number of deamidations, PTM location scoring, Ascore (1) and D-score (2), is used to assign the modification to the correct asparagine. Ascore is used for peptides with only one deamidation, and D-score is used for peptides with multiple deamidations. An Ascore of 0 result in a *random* annotation, while scores between 0 and the threshold (as set in PeptideShaker) result in *doubtful* annotations, and scores above the threshold result in a *very*

confident annotation. D-scores below 95% give a *doubtful* annotation, and above 95% give a *confident* annotation.

Because of this possible spontaneous deamidation of asparagine, it was important to distinguish such spontaneous deamidated asparagines from the former glycosites. Most of the glycopeptides that we identified had only one asparagine in the sequence (665), and for these the location confidence of the deamidation, and thereby the confidence of the glycosite was *very confident*. Several other glycopeptides, however had more than one asparagine in the peptide sequence (459), and for some of these (202) the deamidation annotation was *doubtful* or *random*, and sometimes assigned to an asparagine which was not in the N-glyco sequence motif. Considering the low probability that peptides with a deamidation and the N-glyco sequence motif are <u>not</u> glycopeptides, and the fact that we performed glycopeptide enrichment, we assume that for these peptides, the deamidation is on the asparagine which is in the N-glyco motif, and we have therefore for these peptides assigned that as the glycosite position.

For only a few glycopeptides (six) the deamidation was confidently or very confidently assigned to an asparagine which was not in the N-glyco motif, although there were other asparagines in the motif elsewhere on the peptide. These particular peptides are likely not true glycopeptides, although they fulfill part of our glycopeptide definition. However, three of them also appear in the dataset with the deamidation placed on the asparagine which is in the motif, and will because of that still be considered as valid glycopeptides. In addition, some glycopeptides (seven) had only one deamidation, but two asparagines in the motif, and the site confidence annotation for the deamidation was *doubtful*. For these we cannot know for sure on which asparagine the glycosite is and have chosen to go with the site assigned by PeptideShaker (*doubtful*, but not *random*) since that is the more likely alternative. All but one of these seven peptides had both sites in question either

confidently identified elsewhere in our dataset, or in SwissProt (with reference), and the remaining one is marked with * in the table of novel glycopeptides.

To investigate if we had identified previously unknown/not verified glycosites, we cross-checked our identified sites against what was contained in Swiss-Prot. Information about all the proteins we had identified in our glycopeptide enrichment experiment was downloaded from UniProt, and the protein accession number was linked with the position of a possible glycosylation in the protein sequence. The same protein accession number – glycosite position linking was done for our dataset and this was then compared. In this way we could see which of our identified sites were listed with a reference in Swiss-Prot. Peptides which were shared between proteins or isoforms, or that occurs at multiple locations in a protein sequence (38 in total) where for obvious reasons not included in this comparison. We considered a glycosite as new if there were no reference to this site being a glycosylation site for this protein.

The glycodata analysis was performed on peptide and protein lists exported from PeptideShaker v0.19.0. Note that new versions of PeptideShaker has since been released, and that peptide and protein identifications may vary slightly between versions.

REFERENCES

1. Beausoleil, S. A., Villen, J., Gerber, S. A., Rush, J., and Gygi, S. P. (2006) A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat Biotechnol* 24, 1285-1292
2. Vaudel, M., Breiter, D., Beck, F., Rahnenfuhrer, J., Martens, L., and Zahedi, R. P. (2013) D-score: a search engine independent MD-score. *Proteomics* 13, 1036-1041