Combining Bayesian Networks and Markov Models for Positioning Modifications on Phosphopeptides

Oliver Horlacher[1], Frederique Lisacek[1], Markus Muller[1*]
[1]Swiss Institute of Bioinformatics, Geneva, Switzerland
[*]Presenting author

Abstract. Post translational modification (PTM) of proteins, such as phosphorylation, play an important role in many biological processes. Correct assignment of the modification site is essential if mass spectrometry is to be used to analyze these processes reliably. There are two parts to solving spectra of peptides that contain PTMs: identifying the peptide sequence and determining which AA is modified. Generally search engines are good at finding the correct peptide sequence however the positioning of the modifications is less reliable, especially for PTMs that generate complex fragmentation patterns such as phosphorylation. We estimate the positioning error of several popular sequence search tools, and investigate when positioning errors happen.

The assignment of peptide sequence and phosphorylation site is challenged by the lability of the phosphate group during collision-induced disassociation. The phosphate group is readily lost via the neutral loss of either HPO_3 or H_3PO_4 (HPO_3 + H_2O). Both of these neutral losses yield peaks that are indistinguishable from those that an unmodified fragment can generate. Here we outline a new algorithm that takes into account the complex fragmentation pattern of phosphopeptides. The algorithm combines a Bayesian network and Markov model to find the modification site. The Bayesian network is used to classify each position along the peptide sequence as carrying a modification on the b ion, y ion or both ions. The Markov model then uses these classifications to resolve the position of the modification. The algorithm is compared to AScore and PhosphoRS, two commonly used positioning methods.