

A Perl procedure for the analysis of phosphopeptide tandem mass spectra

Tiengo A¹, Barbarini N¹, Temporini C², Massolini G², Magni P¹

Motivation

Phosphorylation is a protein post-translational modification (PTM) occurring on serine, threonine and tyrosine residues. Phosphorylation has recently become the focus of interest because of its function as “molecular switch” in biological systems, often being representative for the activation/inactivation of a particular disease related pathway. Consequently, the study of the degree of phosphorylation of a protein and the localization of this PTM along the amino-acid chain becomes a key point for proteomic analysis aimed at identifying diagnostics or therapeutic targets. Liquid chromatography coupled to tandem mass spectrometry (MS/MS) is the most widely used analytical approach used. Phosphopeptides which undergoes collision induced dissociation (CID) in MS/MS give representative mass spectra. Although non modified peptides are automatically identified by robust and consolidated algorithms, phosphopeptide MS/MS spectra still require manual interpretation and this step constitutes the main bottleneck of the analysis. A Perl procedure is here proposed to automate the data treatment step.

Methods

To identify phosphopeptides, MS/MS spectra are searched against a suitable peptide database, automatically created from the most recent release of SwissProt. The proteins of the organism(s) of interest are selected and the theoretical peptides are computed from the amino-acid sequences through an *in silico* enzymatic digestion adding also the missed cleavages (MCs) and the PTMs, among which phosphorylation. For each MS/MS spectrum the intensities of all the detected fragment ions are normalized to the most intense peak. When phosphopeptides undergo CID, they are subjected to a neutral loss of phosphoric acid (H₃PO₄), which is often a dominant peak in a MS/MS spectrum. Therefore, to select from all the acquired spectra only those spectra corresponding to phosphopeptides, the Perl procedure analyzes every spectrum to find a peak with intensity over a certain threshold corresponding to the neutral loss of H₃PO₄. When a neutral loss is found in the spectrum, it is searched against the generated database. The theoretical peptides that match the precursor ion (monoisotopic mass) are “fragmented” *in silico* to obtain the correspondent theoretical MS/MS spectrum. The fragmentation

¹ Dipartimento di Informatica e Sistemistica, Università degli Studi di Pavia, Via Ferrata 1, Pavia, I-27100, Italy ² Dipartimento di Chimica Farmaceutica, Università degli Studi di Pavia, Viale Taramelli 12, Pavia, I-27100, Italy

model considers the b and y ions. The theoretical spectrum is then compared with the experimental one through a scoring function. The procedure ranks all the candidate peptides on the basis of this score, identifying the correspondent phosphopeptide and its phosphorylation sites.

Results

The procedure proposed here was designed and validated on a publicly available library containing 2845 MS/MS spectra of human peptides. It was implemented in Perl. First, a database of tryptic peptides was created starting from the release 57.12 of SwissProt. A maximum of 2 MCs for each peptide were allowed and the PTMs considered were phosphorylation of serine and threonine and CAM of cysteine. The mass tolerances used to perform the comparison between acquired and theoretical MS/MS spectra were 7 ppm for the precursor ion masses and 0.8 Da for the fragment ions. The intensity threshold used to select the mass spectra of phosphorylated peptides is a design parameter of the procedure. Tuning it, the performance (sensitivity, specificity, accuracy, etc.) of the Perl procedure changes because the number of peptides that pass the neutral loss filter depends on this threshold. Seven threshold values between 0 to 100 were used to plot the ROC curve of Fig. 1. Note that, the sensitivity switches from 0 for thresholds greater than 100 to 0.586 for a threshold of 100, because of the normalization of the peak intensities to the higher one. The threshold parameter has to be chosen in accordance with the characteristics of the investigation, for example minimizing the false discovery rate or increasing the accuracy.

Contact e-mail

alessandra.tiengo@unipv.it

Image

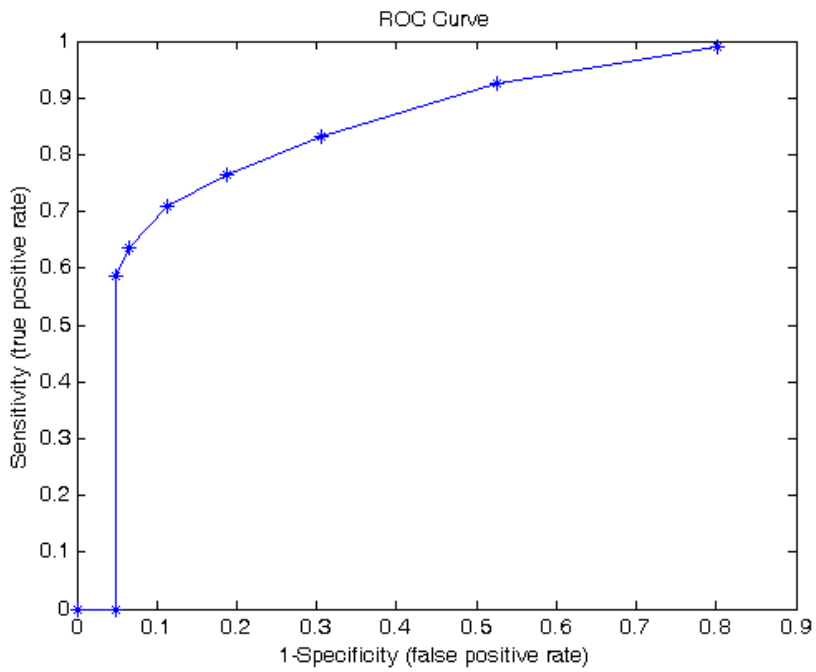


Figure 1– The ROC curve obtained tuning the intensity threshold