Efficient Phosphopeptide Identification from Tandem Mass Spectra

TIENGO A, BARBARINI N, MAGNI P
Dipartimento di Informatica e Sistemistica, Università degli Studi di Pavia, Pavia

Motivation
Phosphorylation is a chemical reaction taking place into the cell, in which a phosphate group is added by a protein kinase to the amino acids serine, threonine and tyrosine of a protein increasing their molecular mass. Identification of phosphorylated peptides, called also phosphopeptides, is an interesting issue in proteomic, due to their implication in both normal and pathological conditions [1]. Tandem Mass Spectrometry (MS/MS) coupled with liquid chromatography (LC) is the most widely used approach to study phosphorylated proteins. MS/MS is used to measure, with high precision, the mass charge ratio (m/z) of charged molecule. The output of MS/MS is the representative mass spectrum of each peptide analyzed, i.e. a list of all m/z values of the ions generated from the fragmentation of the peptide with their intensity. Although there are many software tools in the literature able to perform MS/MS identification, phosphopeptide MS/MS spectra still require manual interpretation. This step represents the main bottleneck of the analysis in terms of time and reliability of the obtained results, because of the large amount of data generated with mass spectrometer and of the complexity of the signals acquired.

Methods
A Perl software tool called PhosphoHunter was specifically designed to analyze phosphopeptide MS/MS spectra. The main steps of PhosphoHunter are:
1. generation of a database of true and random phosphorylated peptides from a suitable set of proteins of interest;
2. processing of the acquired spectra to select those probably referred to phosphopeptides;
3. identification of the selected spectra;
4. attribution of a p-value to each identification, testing the null hypothesis of coming from chance.

In the first step, a database of peptides is created starting from a list of FASTA protein sequences coming from a private or public database (e. g. SwissProt), reproducing in silico the digestion of the amino-acid sequences to obtain the theoretical peptides of each protein [2]. To test the statistical significance of the results, a database of random sequences is created reversing the protein sequences. The sequences are in silico digested and the two databases are merged. Because in MS/MS experiments phosphorylated peptides produce a characteristic spectrum often dominated by an intense peak cor-
responding to a neutral loss of phosphoric acid (see Figure 1), PhosphoHunter includes a preliminary step (step 2.) to filter out those spectra that do not show this characteristic. In the identification (step 3.), each experimental spectrum is compared with the theoretical MS/MS spectra stored in the database previously created, by using a sigmoid score function. All the peptides in the database are, then, ranked on the basis of their scores and the experimental MS/MS spectrum is assigned to the peptide with the highest score. Finally, the statistical significance of each score in the ranked list is tested (step 4.), by comparing it with the distribution of the scores obtained on the random sequences.

Results
A dataset of 1500 MS/MS spectra has been used to test PhosphoHunter and its performance was compared with a very popular software tool: Mascot [3-6]. Results show that the performance of PhosphoHunter is globally better than that of Mascot. The main advantages of using PhosphoHunter are the minor number of false positive phosphopeptides (e.g. the MS/MS spectra wrongly identified that do not correspond actually to phosphorylated peptides) and the greater efficiency of the analysis, since less MS/MS spectra are processed in the identification step thanks to the filtering of the step 2. Finally, because PhosphoHunter is an open source tool implemented in Perl, it can be easily extend and customize by the user including also additional features.

References

Contact email
alessandra.tiengo@unipv.it
Figure 1 - Phosphopeptide MS/MS spectrum