Evaluation of a Strategy for Parallel Quantitative Protein Phosphorylation and Expression Analysis

Glibert P. [1] *, Dhaenens M. [1] *, Meert P. [1], Colaert N. [2,3], Van Nieuwerburgh F. [1], Martens L. [2,3], Deforce D. [1]

[1]Laboratory for Pharmaceutical Biotechnology, Ghent University, Harelbekestraat 72, B-9000 Ghent, Belgium
[2]Department of Medical Protein Research, VIB, Ghent, Belgium
[3]Department of Biochemistry, Ghent University, Ghent, Belgium
*Authors contributed equally

Background

Reversible protein phosphorylation is one of the major molecular mechanisms behind cell signaling and regulates complex protein functions. Analytical methods in phosphoproteomics however, face many obstacles since phosphorylation has a **transient, dynamic character** and a **low stoichiometry**. Hence, enrichment strategies before mass spectrometry analysis are often mandatory but induce the loss of quantitative information *(Figure 1)*.

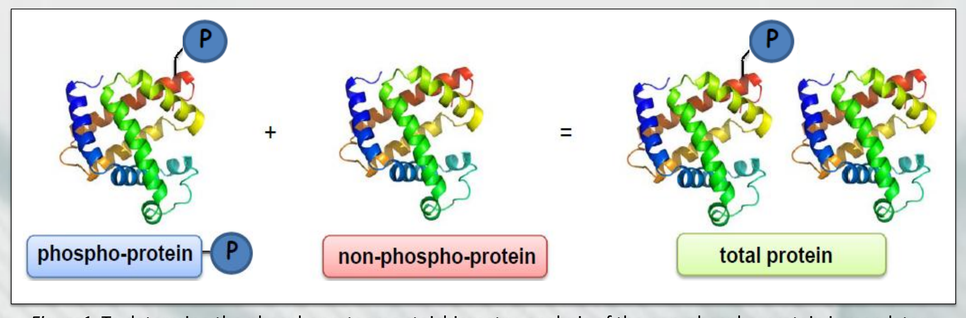


Figure1: To determine the phosphoproteome stoichiometry, analysis of the non-phospho-protein is mandatory.

Here we present a strategy which **simultaneously identifies phosphopeptides and measures the biologically important phosphorylation stoichiometry**[1,2]. This iTRAQ-based approach allows to compare both the phosphorylation and the **expressional differences between** different **samples** in a single experiment. However, extensive evaluation of the data is still essential to see beyond the **smoke and mirrors**.

Principle

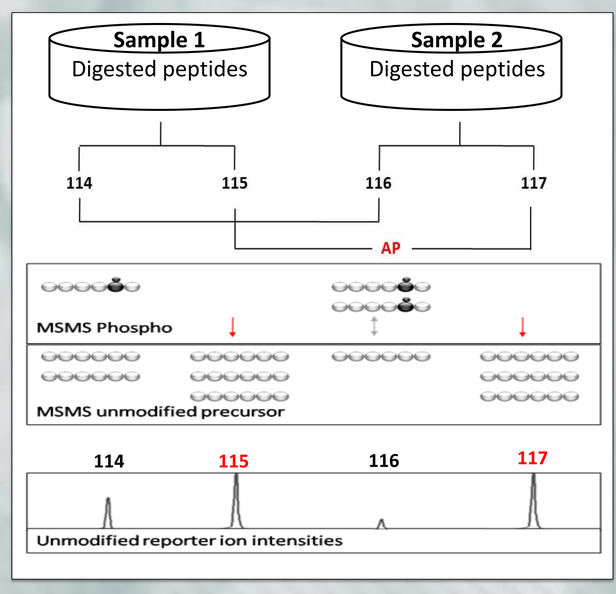


Figure 2: Principle. A peptide mixture of each biological state is split in two identical parts and each part is labeled separately. One half of each sample is dephosphorylated by alkaline phosphatase (AP), one half serves as a control. The iTRAQ reporter ions in the MSMS spectrum increase proportionally to the amount of phosphorylated peptide. The reporter ratios 114/115 & 116/117 thus define stoichiometry at the peptide level, i.e. to calculate the phosphorylation level of a peptide in sample 1 the formula is $[1-(114_{area}/115_{area})]$. Ratios 114/116 & 115/117 define expression at the protein level between different samples.

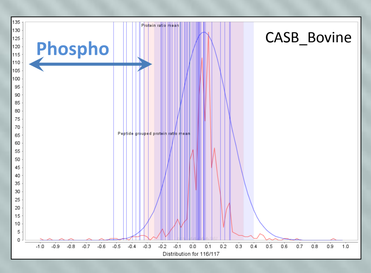
This phospho-mapping strategy focuses on the **unphosphorylated peptide precursor mass**. The approach is based on the mass alteration of a phosphopeptide that is induced by ***in vitro* dephosphorylation**, by the use of a phosphatase. This peptide thus ends up in the **same MSMS spectrum as its unmodified equivalent** *(Figure 2)*.

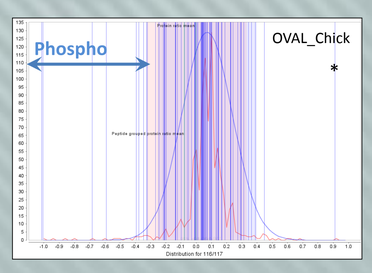
In each sample, the unphosphorylated peptide precursor mass serves as the abundant template for peptide identification, while the skewed ratio of the **iTRAQ reporter labels** indicates the **phosphosite stoichiometry**. The iTRAQ multiplex provides the opportunity to compare both the phosphorylation and the **expressional differences** between multiple biological states.

Rover defines outliers

To reveal the iTRAQ ratios of the peptides that deviate from the normal distribution, the data is imported in the **CompOmics software tool Rover**[3]. The Rover wizard displays the distribution of the peptide ratio's from selected proteins and uses robust statistics to account for the **outliers of the dataset**.

The principle of this analytical LC-MS/MS technique was initially tested on a protein mixture containing non phosphoproteins and the known phosphoproteins **β-casein and ovalbumin** *(Figure 3)*. Peptides with a Z-score <1.96 are identified as phophopetides.





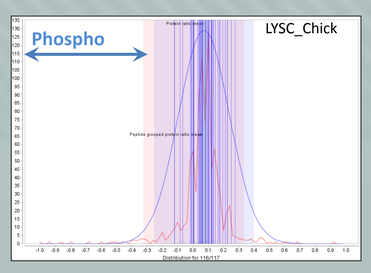


Figure 3: With the Rover software the peptide ratio's that deviate from the normal distribution are characterized. *Left panel*: On a non-phosphoprotein (e.g. LYSC_Chick) all peptides have a 1:1 ratio. *Upper panels:* Peptides of the proteins CasB_Bovine and Oval_Chick with a Z-score < 1.96 are identified as phosphopeptides. In Oval_Chick one still intact phosphorylated peptide mass with a Z-score > 1.96 is identified *(asterisk)*.

Increasing coverage

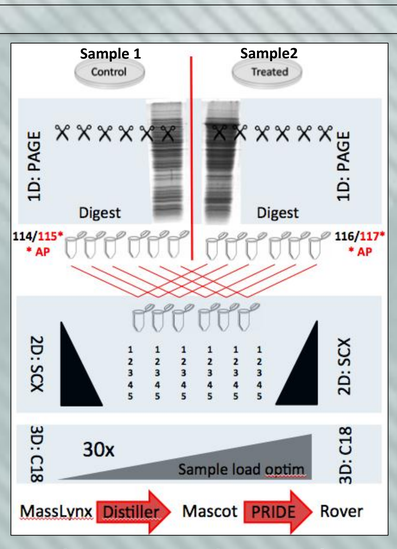
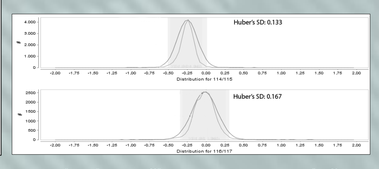


Figure 4. Left panel: Overview of the complete protocol including three fractionation steps. 1D SDS-PAGE to include hydrophobic proteins is followed by SCX (2D) and C18 (3D) LC fractionation of the labeld and partly dephosphorylated peptide mixture.
Lower panel: The 12.250 annotated spectra displayed a very small standard deviation as given by Rover.

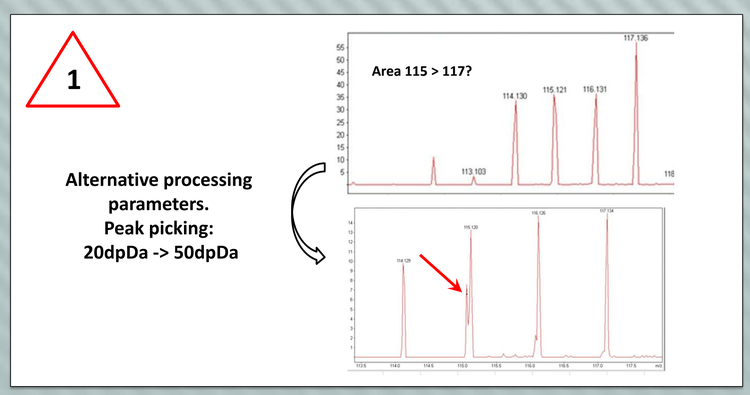


In this approach peptides are only briefly separated for labeling and dephosphorylation, which allows for **unlimited prefractionation** steps *(Figure 4)*. Including a 1D SDS PAGE prefractionation step allows for the discovery of phosphorylation events on **hydrophobic proteins** which are lost in most MS-based approaches.

To test the impact of these steps on the distribution of the ratio's we used the model system of forskolin stimulated and non-stimulated **HepG2 cells**. For the 114/115 ratios, 209 peptides with a Z-score < -1,96 were found. For 116/117 this resulted in 230 peptides, of which 17 peptides were equally phosphorylated in the non-stimulated HepG2 cell line.

iTRAQ: Pitfalls emerge

Drawing quantitative conclusions of an iTRAQ experiment on the **peptide level** however, puts high demands on the **data quality** as manual in depth analysis of the processed data revealed. Aberrant, low quality spectra might **bias the result** when focusing on the outliers of a large dataset. **Alternative data processing** and different search algorithms are currently being validated to minimize the observed issues.



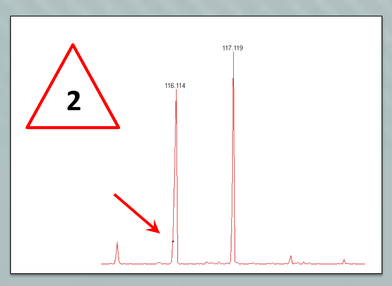


Figure 5: Even after deleting low quality spectra with low reporter peak areas from the iTRAQ dataset, difficulties remain in reliably quantitating on the peptide level.
Upper panel: Adjusting the parameters of the Mascot Distiller processing revealed a contaminating peak which severely influences the area of the 115 reporter ion. *Left panel:* Shoulder and tail annotations encompassing the iTRAQ reporters hamper accurate quantitation.

Conclusion

- **Focusing on the unphosphorylated peptide precursor mass** overcomes most of the challenges in phosphoproteome analysis.
- The iTRAQ-based approach is capable of measuring **phosphopeptide stoichiometry** and the **expressional differences** between multiple biological states in a single experiment.
- **Rover** is highly suitable to define the deviant peptide ratios of a large dataset.
- Drawing **quantitative conclusions of an iTRAQ experiment on the peptide level** puts high demands on the **data quality and data processing** algorithms.

References
1.Pflieger D, Jünger MA, Müller M, Rinner O, Lee H, Gehrig PM, Gstaiger M and Aebersold R. Quantitative Proteomic Analysis of Protein Complexes. Molecular and cellular proteomics 2008, 7(2): 326-346
2.Wu R, Haas W, Dephoure N, Huttlin EL, Zhai B, ME and Gygi SP. A large-scale method to measure absolute protein phosphorylation stoichiometries. Nature Methods 2011, 8(8): 677–683
3.Colaert N, Helsens K, Impens, F, Vandekerckhove J, Gevaert K. Rover: a tool to visualize and validate quantitative proteomics data from different sources. Proteomics 2010, 10(6):1226-9