Para la sesión hemos buscado personas con la suficiente experiencia y conocimientos como para ser las personas de referencia en el debate sobre dichos temas:

Alberto Medina (CNB-CSIC), pionero de la proteómica computacional en España, con dilatada experiencia en temas relacionados con la biología de sistemas como la búsqueda de anotaciones y los repositorios de información biológica y proteómica, así como en temas relacionados con estándares.

Alex Campos (PCB), experimentado usuario de herramientas de software libre aplicadas a la proteómica, así como de estándares y herramientas de conversión de formato de datos.

Marco Trevisan (CBMSO-CSIC), uno de los desarrolladores de un software de cuantificación de uso gratuito para instituciones públicas y conocedor de los problemas a resolver con la estadística en la cuantificación de proteínas.

Juan Antonio Vizcaíno (EMBL-EBI) integrante del proyecto PRIDE (*PRoteomics IDEntifications database*) [1], ha participado en estudios de minería de datos de espectrometría de masas y es participante en el desarrollo de estándares de la iniciativa de HUPO-PSI [2].

Cada uno de ellos realizará una muy pequeña introducción de los temas propuestos, y seguidamente se aprovechará la mayor parte del tiempo de la sesión en el intercambio de información entre los asistentes y nuestros "expertos".

Creemos que entre los cuatro forman una mesa muy interesante y esperamos que podamos proporcionar una sesión fructífera para todos.

## Referencias

[1] Vizcaino JA, Cote R, Reisinger F, Foster JM, Mueller M, Rameseder J, et al. A guide to the Proteomics Identifications Database proteomics data repository. Proteomics 2009;9:4276-83.

[2] Orchard S, Hermjakob H y Apweiler R. The proteomics standards initiative. Proteomics 2003;3:1374-6.

# A new versatile file translator for Proteomics standards

*J. Alberto Medina-Aunon[1,2], Salvador Martínez-Bartolomé[1,2], J. Pablo Albar[1,2]*

[1]Proteomics Facility, National Center for Biotechnology (CNB-CSIC), [2]Spanish Proteomics Institute (ProteoRed).

Information derived by proteomics experiments such as sample preparation and separation, mass spectrometry, protein identification, etc..., is represented using different and sometimes incomprehensible file formats, resulting not very suitable for daily life. To deal with this problem, several initiatives have been developed: HUPO-PSI MIAPE and XML-based formats [1], XML-based repository: PRIDE [2], etc... All of them are contributing effectively within a Proteomics scope, but some lacks regarding the translation between some of the previous schemas are still confusing.

In order to contribute to this situation, we have developed a new Sun Java application accessible on the Internet (http://www.proteored.org/). This tool has been designed using common graphics libraries and latest XML parsers offering a friendly and easy interface avoiding complex configuration dialogs.

Currently, the application is able to manage the following file formats -proteomics standards-: MIAPE Gel, MS and MSI documents (MS Excel and XML format), GelML, and PRIDE.

To illustrate this innovative work we realized two tests, both of them based on the translation between GelML and MIAPE Gel schemas. Input files were creating from 1) HUPO PSI Microsoft Excel template and 2) ProteoRed MIAPE web site (www.proteored.org). Both schemas are perfectly

complementary: MIAPE Gel presents a lot of information according to protocols, methods and images and GelML could exchange this information with other bioinformatics tools. In spite of the fact that both present a great combination, some problems appeared during the transformation. For instance, MIAPE presents a lineal schema -well defined pipeline- whereas GelML is a FUGE based one, dividing the information along several linked sections. Furthermore, any assumption can be done about the cardinality of the relationships between both examples. Finally, these issues made us to define a set of configuration files to drive the transformation.

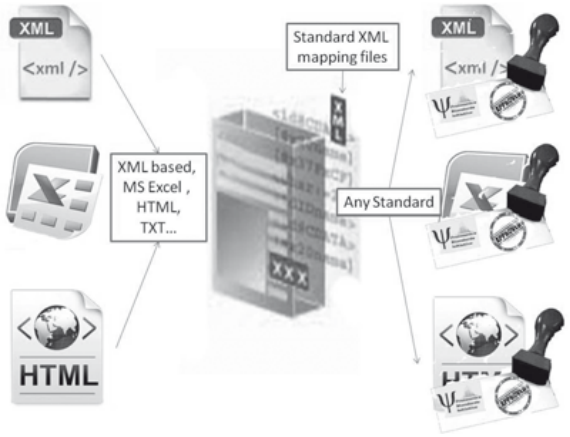


**Figure 1.** *Translation between proteomics standards will be easier.*

The described work results a proper correspondence between both formats, allowing a complete translation between MIAPE Gel and GelML schemas.

## References

[1] Martens L, Orchard S, Apweiler R and Hermjakob H. Human proteome organization proteomics standards initiative: data standardization, a view on developments and policy. Mol Cell Proteomics 2007; 6: 1666-1667.

[2] Jones P, Côté RG, Martens L, Quinn AF, Taylor CF, Derache W, Hermjakob H and Apweiler R. PRIDE: a public repository of protein and peptide identifications for the proteomics community; Nucl Acid Res. 2006; 34 (Database issue): D659-D663.

# Open-source Bioinformatics Solutions for the Analysis of Mass Spectrometry-based Proteomics Data: Pipelines and Quantitation

*Alexandre R. Campos*

Proteomics Platform, Barcelona Science Park, Barcelona, Spain

The proteomics community has been generating openly available software framework for systematic proteomic data analyses and management. Processing and analysis of proteomics data involves a complex, multistage process including raw file pre-processing, peptide assignment to MSMS experimental spectra, protein identification and further validation, and in some cases, MS-based quantitation. Here, I list a number of freely available and open-source computational tools for the analysis of proteomics data. Particularly, I focus on available platforms and pipelines, and software for MS-based quantitation.

## 1. Platforms and Pipelines for LC-MS and LC-MS/MS Data Analysis

With the advent of new generation of mass spectrometers arose the necessity of developing new bioinformatics solutions for mining large data sets. In the past years, we have witnessed a dynamic