

been funded by Eortek and Saiotek grants from the Industry Department of the Basque Government. Proteomics Core Facility-SGIker is supported by UPV/EHU, MICINN, GV/EJ, ESF and ProteoRed agencies.

- [1] Peeper DS, Keblusek P, Helin K, Toebes M, Van der Eb AJ and Zantema A. Phosphorylation of a specific cdk site in E2F-1 affects its electrophoretic mobility and promotes pRB-binding in vitro. *Oncogene* 1995;10:39-48.
- [2] Lin WC, Lin FT and Nevins JR. Selective induction of E2F1 in response to DNA damage, mediated by ATM-dependent phosphorylation. *Genes Dev* 2001;15:1833-44.
- [3] Larsen MR, Thingholm TE, Jensen ON, Roepstorff P and Jørgensen TJ. Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol Cell Proteomics* 2005;4:873-86.
- [4] Blom N, Gammeltoft S and Brunak S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol* 1999;294:1351-62.
- [5] Gnad F, Ren S, Cox J, Olsen JV, Macek B, Oroshi M and Mann M. PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites *Genome Biology* 2007;8:R250.

Establecimiento de un flujo de trabajo efectivo en la caracterización cualitativa y cuantitativa del fosfoproteoma

David Ovelleiro, Montserrat Carrascal, Joaquín Abian

LP-CSIC/UAB, IIBB-CSIC, Edificio M, Campus UAB, Bellaterra, Barcelona, Spain

La determinación de puntos de fosforilación en proteínas es una de las áreas de la proteómica con mayor interés en la actualidad. Los últimos avances en la instrumentación y en las técnicas de purificación están permitiendo la detección de formas fosforiladas de péptidos a muy bajas concentraciones y con gran fiabilidad. La caracterización del denominado fosfoproteoma hace posible la obtención de instantáneas del estado de fosforilación de un determinado organismo y su cuantificación en diferentes estados. No obstante, debido a las utilizaciones de métodos de enriquecimiento y a los barridos específicos utilizados en el análisis por espectrometría de masas, el análisis de los datos generados aumenta sensiblemente en complejidad respecto a otros análisis proteómicos clásicos. Es por tanto necesario el establecimiento de un flujo de trabajo específico para fosfoproteómica con objeto de responder a las diversas necesidades que las técnicas involucradas requieren.

1. Peculiaridades de los datos obtenidos en la identificación de fosfopéptidos

En el análisis de fosfopéptidos mediante técnicas de *shotgun proteomics* deben tenerse en cuenta las siguientes consideraciones:

- La evaluación de la confianza en la identificación de fosfoproteínas en base al número de péptidos identificados no siempre es posible dado que en las etapas de enriquecimiento gran parte de los péptidos no fosforilados se eliminan. La información analítica reportada debe por tanto tener en cuenta todas las proteínas a las que apuntan estos péptidos únicos, tanto a efectos cualitativos como cuantitativos.
- El número de modificaciones a tener en cuenta durante la identificación de los espectros de fragmentación (al menos 3 modificaciones variables para S/T/Y en MS², y dos más para S/T en MS³) y las características espectrales de los fosfopéptidos (pérdida preponderante del grupo fosfato) facilitan la aparición de falsos positivos. La utilización de bases de datos señuelo nos permite obtener una estimación de la proporción de falsos positivos en el conjunto de fosfopéptidos validados (*FDR* o *False Discovery Rate*) [4, 5].
- Un elemento clave en la asignación de puntos de fosforilación es la correcta identificación del aminoácido modificado cuando exista más de una alternativa posible. En estos casos, los buscadores en base de datos como SEQUEST,

Phenyx u OMSSA, suelen tener dificultades para distinguir entre las distintas posiciones de fosforilación y pueden dar lugar a asignaciones incorrectas. Por este motivo son precisos análisis adicionales específicos con objeto de validar estas asignaciones o corregirlas. Entre los métodos descritos para estos análisis [6,7], uno de los más populares es la puntuación Ascore.

2. Formatos para el almacenamiento de información

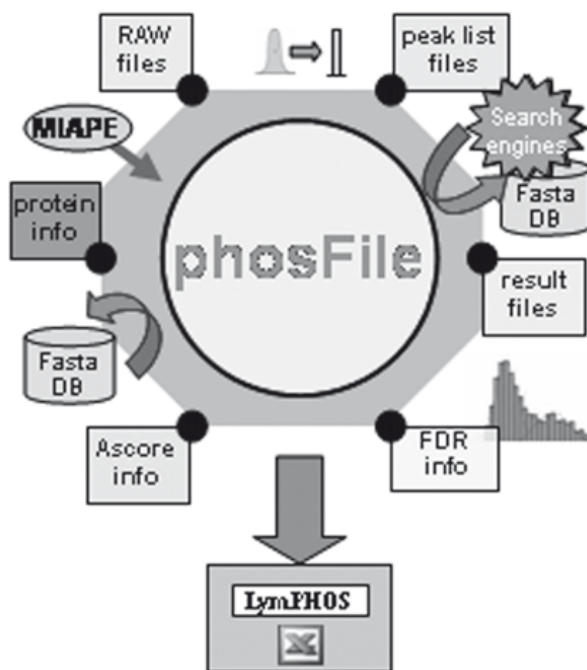


Figura 1. Información generada en el flujo de trabajo y contenida en el archivo de almacenamiento *phosFile*.

El formato estándar más utilizado para almacenar la información analítica en proteómica es el XML. Uno de los estándares más recientes de archivo XML para el almacenamiento de datos analíticos es el *mzIdentML* [8]. Sin embargo, la complejidad y diversidad de la información recopilada en un experimento proteómico hacen que estos estándares sean hoy por hoy insuficientes, por lo que es necesario evaluar otras alternativas de almacenamiento. Por otra parte, es también importante establecer un procedimiento para la incorporación de la información experimental utilizando diferentes MIAPE asociadas al experimento, teniendo especial importancia aquí el modelo MIASPPE [9] (“*Minimum Information About Sample Preparation for a Phosphoproteomics Experiment*”). En la Figura 1 se presenta el esquema general del flujo de trabajo utilizado en la identifica-

ción y cuantificación de fosfopéptidos, así como el almacenamiento de la información obtenida en cada paso en un único formato de almacenamiento llamado “*phosfile*”. Dicho formato de almacenamiento permite una fácil conversión a formatos usables por el investigador (ie. Excel) o la exportación de la información a una base de datos como Lymphos [10].

Referencias

- [1] Nesvizhskii A. I. and Aebersold, R.. Interpretation of Shotgun Proteomic Data. The Protein Inference Problem. *Mol. Cell Proteomics* 2005; 4:1419-1440.
- [2] Wilkins, M. R., Appel, R. D., Van Eyk, J. E., et al., Guidelines for the next 10 years of proteomics. *Proteomics* 2006; 6: 1, 4-8.
- [3] Reiter L, Claassen M, Schimpf, S.P., et al. Protein identification false discovery rates for very large proteomics datasets generated by tandem mass spectrometry. *Mol. Cell Proteomics* 2009.
- [4] Elias, J.E. and Gygi, S.P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* 2007; 4: 207-214.
- [5] Xiuxia Du. et al. Linear Discriminant Analysis-Based Estimation of the False Discovery Rate for Phosphopeptide Identifications. *J. Proteome Res.* 2008 ; 7 : 2195-2203.
- [6] Beausoleil, S. A., Villen, J., Gerber, S. A., et al. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat. Biotechnol.* 2006; 24: 1285-1292.
- [7] Ruttenberg, B.E., Pisitkun, T., Knepper, M.A. and Hoffert, J.D. PhosphoScore: An Open-Source Phosphorylation Site Assignment Tool for MSⁿ Data. *J. Proteome Res.* 2008; 7: 3054-3059.
- [8] Orchard S, Jones P, Taylor C, et al. Proteomic data exchange and storage: the need for common standards and public repositories. *Methods Mol Biol.* 2007; 367: 261-70.
- [9] Carrascal, M., Ovelleiro, D., Casas, V. et al. Phosphorylation Analysis of Primary Human T Lymphocytes Using Sequential IMAC and Titanium Oxide Enrichment. *J. Proteome Res.* 2008; 7: 5167–5176.
- [10] Ovelleiro D, Carrascal M, Casas V and Abian J. LymPHOS: design of a phosphosite database of primary human T cells. *Proteomics* 2009; 9: 3741-51.