

Identification of unknown proteins in protein mixture samples by LC/MS/MS

Jon Hao¹, Yi Liu¹, Gai Yan², and Anton Wellstein²

¹*Poochon Scientific, Frederick, Maryland 21701*

²*Lombardi Comprehensive Cancer Center, Georgetown University, Washington, DC 20057*

Objective

To identify the protein(s) from FPLC fraction samples that demonstrated biological activity in activating AKT/ERK signaling pathway using Liquid Chromatography tandem-Mass Spectrometry (LC-MS/MS).

Background: To study the biological function of an interesting protein. The target protein was overexpressed in human cell line and it was found to be secreted into culture medium. A FPLC method was developed for partial purification of the target from the medium. The different fractions prepared from culture medium by FPLC demonstrated varied activity in stimulating the AKT/ERK signaling pathway (Figure 1). As shown in Figure 1 the fractions 18 and 19 exhibited a stronger activity in term of enhancing the level of phosphorylation of AKT and ERK (Figure 1C). However, the western blot demonstrated that the fraction 25 contains more intact target protein (Figure 1B). To identify the protein(s) or possible posttranslational modifications (PTM) that could be involved in activating the AKT/ERK pathway we analyzed the three fractions using Liquid Chromatography tandem-Mass Spectrometry (LC-MS/MS) technology.

Methods: FPLC fraction samples were digested by trypsin or chymotrypsin and analyzed by nano-LC/MS/MS using a Dionex Ultimate 3000 RSLCnano system coupled with a Thermo Q-Exactive mass spectrometer.

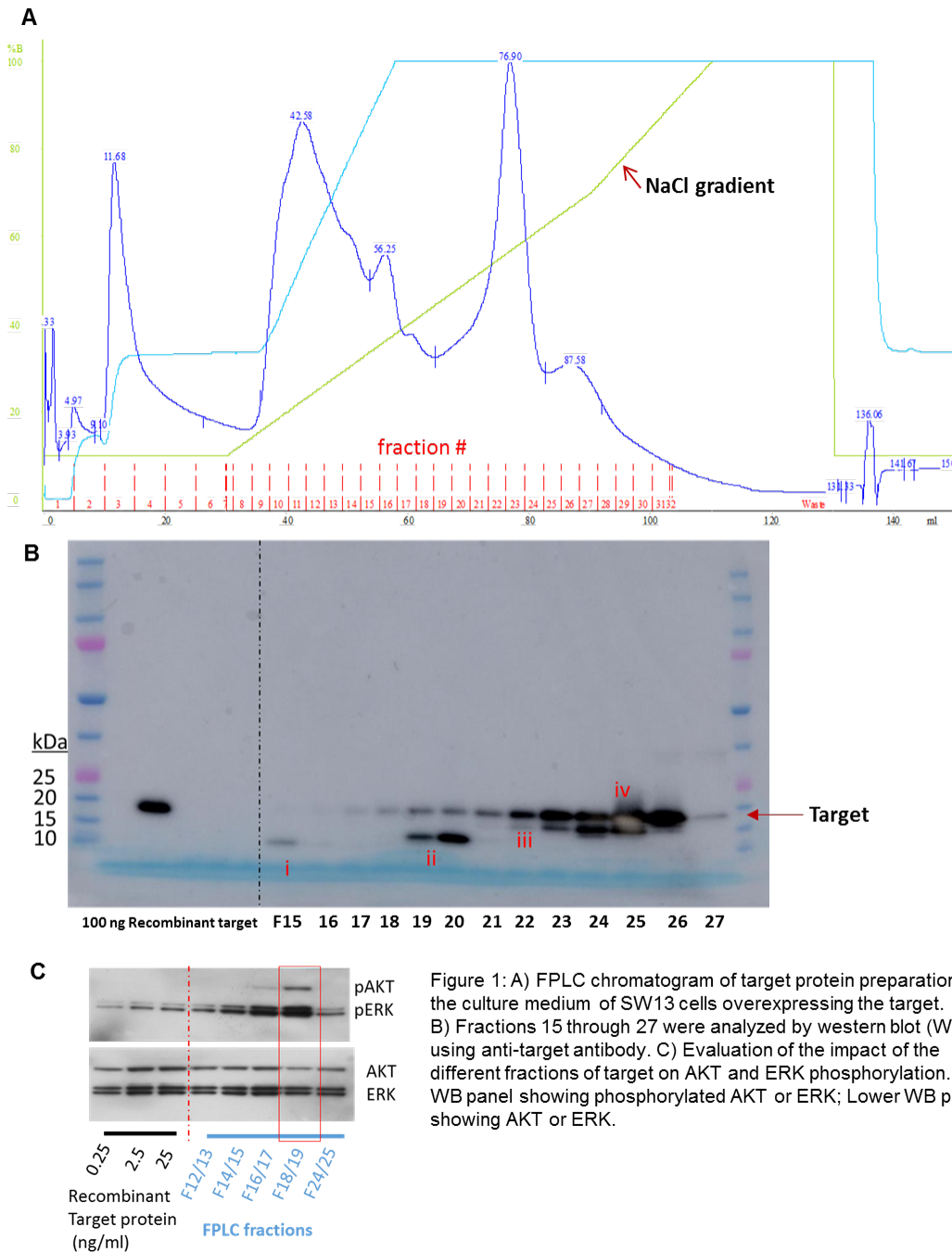
Results: The analyses of FPLC fractions 18, 19 and 25, which were digested by trypsin and chymotrypsin, by LC/MS/MS enabled the identification of the target protein and several growth factors. The analysis also revealed a different posttranslational modification pattern on the target protein from fractions 18 and 19 comparing with the target from fraction 25. In addition, several growth factors were detected in fractions 18 and 19, but they were not or much less detected in fraction 25. The results suggested that the PTM and co-purified growth factors may contribute to the observed activation of AKT/ERK signaling pathway in the functional assay.

Introduction

Biological information flow from the DNA/genome to protein/proteome automatically regulates cell living status in response to internal or external signals. The genome mainly ensures the genetic stability of an organism while the proteins/proteome coded by the genes shapes and guides the operation of living organisms. As the link from the gene to the phenotype, proteins play the key roles in orchestrating all biological processes, ranging from central metabolism to cell structure, maintenance, and replication.

Therefore, it is speculated that all biological processes (on/off) can be regulated and operated by the proteins. The determination of their identities and variations are the most important challenge of modern biology. Recently, Liquid Chromatography Tandem-Mass Spectrometry (LC-MS/MS) has been a predominant technology for quick and precise identification of proteins from biological samples [1–4].

In this report, we evaluated the power of LC-MS/MS in identification of unknown proteins from protein mixture samples.



Materials and Methods

Samples

1. FPLC fraction samples, fraction 18, 19 and 25.

Reagents:

1. Pierce Trypsin Protease and Chymotrypsin Protease (MS grade)
2. Formic Acid Optima LC/MS (A11-50) from Fisher Scientific
3. DTT and Iodoacetamide from sigma
4. LC-MS/MS grade water, Methanol and Acetonitrile from sigma

Equipment

1. Mass spectrometry system: Q-Exactive Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Scientific), Dionex Ultimate 3000 RSLCnano system (Thermo Scientific), Proteome Discoverer 1.4 software (Thermo Scientific). PicoFrit Column, 360 μm OD/75 μm ID, 15 μm tip ID, packed with 5 μm ProteoPep II C18 300A, 20 cm (New Objective).

Preparation of Protein Digestion

Fractionation of culture medium via FPLC was performed by Wellstein Lab at Georgetown University. FPLC fraction samples were separated on a SDS-PAGE gel and the target protein gel-band were digested by trypsin or chymotrypsin according to Poochon's SOP PS6001 (Figure 2). The peptides mixtures were cleaned using C18 Zip-Tip and reconstituted in 25 μl of 0.1% formic acid according to Poochon's SOP PS6002. 15 μl of tryptic peptides was analyzed by LC/MS/MS.

Nanospray LC/MS/MS analysis and database search

The LC/MS/MS analysis of samples were carried out using a Thermo Scientific Q-Exactive hybrid Quadrupole-Orbitrap Mass Spectrometer and a Thermo Dionex UltiMate 3000 RSLCnano System. Peptide mixture sample was loaded onto a peptide trap cartridge at a flow rate of 5 $\mu\text{L}/\text{min}$. The trapped peptides were eluted onto a reversed-phase PicoFrit column (New Objective, Woburn, MA) using a linear gradient of acetonitrile (3-36%) in 0.1% formic acid. The elution duration was 110 min at a flow rate of 0.3 $\mu\text{l}/\text{min}$. Eluted peptides from the PicoFrit column were ionized and sprayed into the mass spectrometer, using a Nanospray Flex Ion Source ES071 (Thermo) under the following settings: spray voltage, 1.6 kV, Capillary temperature, 250°C. For protein identification MS files were analyzed using the Thermo Proteome Discoverer 1.4.1 platform (Thermo Scientific, Bremen, Germany) for peptide identification and protein assembly. Database search against public available protein database obtained from NCBI website is performed based on the SEQUEST and Decoy/Percolator algorithms through the Proteome Discoverer 1.4.1 platform. Carbamidomethylation of cysteines was set as a fixed modification, and Oxidation, Deamidation Q/N-deamidated (+0.98402 Da), S/T/K phosphorylation (+79.966 Da), K acetylation (+42.011 Da), and R/K methylation (+14.016 Da) were set as dynamic modifications. The minimum peptide length was specified to be five amino acids. The precursor mass tolerance was set to 15 ppm, whereas fragment mass tolerance was set to 0.05 Da. The maximum false peptide discovery rate was specified as 0.01 or 0.05. The resulting Proteome Discoverer Report contains all assembled proteins with peptides sequences, possible PTMs and matched spectrum counts. The estimation of relative abundance of protein is based on peptide spectrum match counts (PSM #).

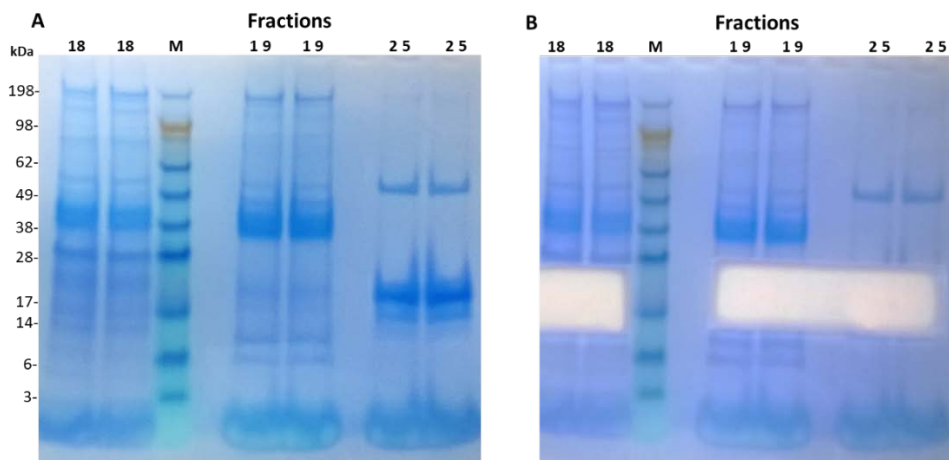


Figure 2: Separation of fraction samples by SDS-PAGE. A) Image of SDS-PAGE gel stained by simply Blue before cutting the target bands for LC/MS/MS analysis. B) Image of the same gel after cutting the target protein gel-bands for LC/MS/MS analysis.

Results

Identification of target protein and possible unknown proteins from FPLC fractions

Three fractions after digestions by chymotrypsin and trypsin respectively were analyzed by LC/MS/MS. The analysis revealed three differences between fractions 18/19 and fraction 25. First, although the target was detected in all three samples but the relative abundance of target in fraction-25 is almost five fold than that in either fraction-18 or 19 (Table 1). This is consistent with the western blot result. Second, the post-translational modification sites on target protein were detected from fractions 18/19 are different from that of fraction 25. One lysine residual was detected with high acetylation (57%) on the target from fraction 25 (data not shown since it has not been published) but this modification was not detected on the target from fractions 18/19. This result implies that this modification may be involved in the regulation of the function of the target protein. Last, several relative abundant growth factors were identified from fractions 18 and 19 but they were not or much less detected in the fraction 25 (Table 1). The growth factors present in fraction 18 and 19 may also contribute to the observed activation of AKT/ERK signaling pathway during the biological function assay.

Table 1: Comparison of the relative abundance of growth factors identified among three fractions (18, 19 and 25)

Accession	Description	# PSM 18	# PSM 19	# PSM 2	# AAs	MW [kDa]	Ratio (18/25)	Ratio(19/25)
xxxxxx	Target	31	16	151	168	18.9	0.21	0.11
116241320	Connective tissue growth factor	211	262	4	349	38.1	52.75	65.50
74717921	Platelet-derived growth factor D	179	74	3	370	42.8	59.67	24.67
113873	Angiogenin	141	35	1	147	16.5	141.00	35.00
74751362	Fibroblast growth factor-binding protein 3	47	119	2	258	27.6	23.50	59.50
74738715	Hepatoma-derived growth factor-related protein 2	45	5	0	671	74.3	infinity	infinity
1708157	Hepatoma-derived growth factor	13	1	0	240	26.8	infinity	infinity
17380528	Vascular endothelial growth factor A	2	12	0	232	27.0	infinity	infinity
146327827	Insulin-like growth factor-binding protein 3	5	8	0	291	31.7	infinity	infinity
124069	Insulin-like growth factor-binding protein 5	9	3	0	272	30.6	infinity	infinity

References

1. Jürgen Cox and Matthias Mann. Quantitative, High-Resolution Proteomics for Data-Driven Systems Biology. *Annu. Rev. Biochem.* 2011. 80:273–99.
2. Paoletti, A.C., et al., Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors. *Proc Natl Acad Sci U S A*, 2006. 103:18928-33.
3. Han, C.L., et al., An informatics-assisted label-free approach for personalized tissue membrane proteomics: case study on colorectal cancer. *Mol Cell Proteomics*, 2011. 10(10):1-15.
4. Gao JL., et al., A conserved intronic U1 snRNP-binding sequence promotes trans-splicing in *Drosophila*. *Genes Dev.* 2015. 29(7):760-71.