

Improved Protein Identification and Biomarker Discovery with Statistical Analysis of Liquid Chromatography/Tandem Mass Spectrometry Profiles

○Takao Kawakami (Clinical Proteome Center, Tokyo Medical University, Tokyo, Japan)

Comprehensive protein identification and quantitation in a complex biological sample are major subjects for most proteomic researches. While peptide-based methodologies using reversed phase liquid chromatography (RPLC) directly coupled with electrospray ionization (ESI)-tandem mass spectrometry (MS/MS) has become popular to perform them, further development of strategies to arrange the resulting huge raw data and remove false positive protein identifications is of a great necessity for biologists in order to easily access and interpret these data.

To validate peptide assignments to product ion spectra resulted from usual searches against primary structure databases, information of peptide elution times on the RPLC was introduced (Kawakami, T. *et al.*, *Proteomics* 2005, 5, 856-864). The validation is based only on correlation between the measured LC elution time of each precursor ion and the deduced elution time from the amino acid sequence assigned to the product ion spectrum obtained from the MS/MS analysis of the precursor. Sufficient numbers of probable assignments gave a highly correlative curve. Any peptide assignments within a certain tolerance from the correlation curve were accepted for the following arrangement step to list identified proteins.

Computational correction of nonlinear warps of the liquid chromatogram in each analytical run enabled to directly group and compare plural RPLC-ESI-MS/MS profiles. In the processing, more than two ion signal positions from internal standard peptides on the profile were used as fixation points to facilitate accurate matching of identical peptide signals (*internal standard-guided Optimal Profile ALignment, i-OPAL*, in a patent application). The algorithm is applicable for statistical analyses including biomarker discovery studies, in which peptide signals with significant differences on their intensity between patient and healthy groups were focused on to identify their sources with clues of the corresponding product ion spectra and the elution correlations. Supported with the quantitative data handling, these identified proteins will be further validated for establishment of multiple disease markers with specificity and solidity. The present approaches make primary proteomic data more reliable and accurate, improving throughput of studies as a whole.