

## **Clustering of Liquid Chromatography Tandem** Mass-Spectrometry Data for Peptide Analysis

Beer, I.<sup>2</sup>, Barnea, E.<sup>1</sup>, Tamar Ziv, T.<sup>1</sup> and Admon, A.<sup>1</sup> <sup>1</sup>The Smoler Protein Center, Department of Biology, Technion <sup>2</sup>IBM Research Laboratory, Haifa, Israel

Liquid chromatography (LC) and tandem mass spectrometry (MS/MS) are commonly combined for analysis and comparison of complex peptide mixtures such as obtained during proteome analysis. The resulting datasets include very large amounts of data combining the full mass spectrum of the peptides and the ms/ ms data of selected peptides. A typical mass spectrometer produces hundreds of MS and MS/MS spectra in one run. Even in small-scale proteomics projects, dozens of LC-MS/MS analyses with tens of thousands mass spectra of peptides can be generated, which is beyond the analysis capacity of a human being. The existing peptide identification computer programs only provide a partial solution. We show here how the clustering of similar spectra from multiple LC-MS/MS runs helps manage these data and discover interesting properties of the peptides, the peptide mixtures, and the cells from which the peptides originated. Clustering-based operations contribute to peptide identification by improving spectra quality and providing decision-supporting information. Clustering also facilitates the comparison of peptide mixtures, alleviating the need to identify individual peptides beforehand. In addition, it can be used to correlate the retention time scales of multiple LC runs and to predict peptide retention times from peptide sequences. We implemented the clustering-based methods in a software tool, Pep-Miner. Using the tool, we catalogued the repertoires of MHC Class-I peptides displayed by various human cancer cell types and discovered several cancer-specific peptide candidates for immunotherapy. The methods, however, are not limited to these applications and have the potential to be used for general proteomics.