A procedure to decompose high resolution mass spectra

In the present work, we will propose a strategy to decompose high resolution SELDI/MALDI-TOF mass spectra obtained from a mixture of peptides/proteins by grouping the different mass/charge ratios corresponding to the same protein.

The algorithm analyses a set of N mass spectra coming from samples of a tissue or body fluid of N different subjects (SELDI-TOF studies for differential analysis of protein expression) or from N different scans of the same sample (MALDI-TOF studies for protein identification by PMF).

After a quite standard procedure of preprocessing of every spectrum (baseline subtraction, smoothing filtering and normalization), the proposed methodology considers the median spectrum and looks for all the isotopic peaks contained in the median spectrum by computing the position of all the local maxima.

Then, the procedure is able to associate every peak to an isotopic distribution (with a specific ionizing charge) by some regression models, finding in this way the masses of the peptides/proteins measured.

The mass/charge ratios associated to every isotopic distribution are only the ones whose intensities have a good correlation with the intensities of the most-likely peak on all the spectra. Therefore, our binning algorithm associates to every isotopic distribution, for every spectrum, the intensity given by the sum of the intensities corresponding to the most correlated mass-charge ratios.

Finally the isotopic distributions produced by the same peptide/protein (different ionizations, PTMs, fragments, etc...) are grouping by the coefficient of correlation calculated on the data after binning.