

Poster presentation

A procedure to decompose high resolution mass spectra

Nicola Barbarini*, Paolo Magni and Riccardo Bellazzi

Address: Department of Computer Science and Systems, University of Pavia, Italy

Email: Nicola Barbarini* - nicola.barbarini@unipv.it

* Corresponding author

from Third International Society for Computational Biology (ISCB) Student Council Symposium at the Fifteenth Annual International Conference on Intelligent Systems for Molecular Biology (ISMB)
Vienna, Austria. 21 July 2007

Published: 20 November 2007

BMC Bioinformatics 2007, 8(Suppl 8):P6 doi:10.1186/1471-2105-8-S8-P6

This abstract is available from: <http://www.biomedcentral.com/1471-2105/8/S8/P6>

© 2007 Barbarini et al; licensee BioMed Central Ltd.

Background

The proposed procedure can be applied to two different studies. The first one is discovering proteomic patterns by SELDI-TOF mass spectra. The application of the procedure has the aim of finding how many proteins are represented in the spectrum of the sample, among which to search the biomarker.

The second study is the protocol composed by 2D-GE separation, MS MALDI-TOF and PMF (Peptide Mass Fingerprinting) algorithm, used to identify proteins in a sample. The aim of the application of the proposed procedure to PMF data is the automatic choice of the masses to use as input for PMF algorithm.

Methods

The algorithm analyses a set of N mass spectra already pre-processed by a quite standard procedure (baseline subtraction, smoothing filtering and normalization). Then it looks for all the isotopic peaks contained in the median spectrum by computing the position of all the local maxima. Next it extracts all the isotopic distributions which are in the median spectrum, analyzing every isotopic peak from the one with highest intensity using a model based on chemical knowledge and statistical properties (the coefficient of correlation).

At last the isotopic distributions with the coefficient of correlation among spectra above a threshold are grouped together.

Results

First, the procedure was applied to a dataset of SELDI-TOF mass spectra from the human serum of 216 different subjects. With this algorithm we found 7216 isotopic peaks and then 118 isotopic distribution which we assembled in 10 groups with a coefficient of correlation threshold of 0,72. Every group seems to be associated with a single protein.

Then, we analyzed with the same procedure a MALDI-TOF spectrum of PDIA1_MOUSE enzymatically digested composed of 162 scans. We found 4700 isotopic peaks and then we assembled the first 60 isotopic distributions in 4 groups with a coefficient of correlation threshold of 0,75. The group not produced by the matrix contains the masses of the peptides.

Conclusion

The correlation coefficient among spectra of different subjects/scans is a mean never utilized by the current algorithms for the extraction of isotopic distributions. This solution reduces the speed of the algorithm but it seems to manage very well the situations of overlapping situations and the hard cases.

The most innovative part is the phase of isotopic distribution grouping by correlation coefficient. The application of this procedure shows some good results both for grouping isotopic distributions of the same proteins and for finding the masses of the enzymatically digested peptides in PMF experiment.