

FOOTPRINT: A New Algorithm for Charge State Deconvolution of Complex ESI Mass Spectra

Hailing Zhang¹, Kaizhong Zhang¹, Bin Ma¹, Cunjie Zhang², Amanda Doherty-Kirby², Gilles Lajoie²
¹Department of Computer Science, ²Department of Biochemistry
University of Western Ontario, London, ON, Canada N6A 5B7

1. Introduction:

Complex protein mixtures are often difficult to separate due to similar physico-chemical properties of the proteins in mixture. Mass spectrometry has the potential to resolve such mixture and give the mass of each protein component. MALDI-MS generally displays poor resolution at higher mass. By ESI MS each protein produces a series of peaks in the spectrum due to its multiple charge states resulting from the electrospray ionization process. Each peak represents a species of ions. However this feature can yield very complex spectra for mixture of proteins. The signal at each m/z value may be further complicated due to the natural abundance of isotopes and salt adducts. We developed new software called FOOTPRINT to calculate the intact masses of proteins present in the sample by making use of isotopes and adducts of the multiply charged mass spectra.

2. Methods:

The software process a spectrum in three steps:

1) Baseline subtraction

ESI-MS spectra consist of a series of peaks superimposed upon a continuous relatively slowly varying baseline. For protein mass analysis, only those peaks of different charge state are useful. Baseline subtraction method removes the baseline from spectrum to improve the signal-noise ratio.

2) Scoring each mass position

For each mass position m , the software assigns an initial score according to (a) signal intensities at positions of m/z (mass-to-charge ratio), for different charge values z , in the spectrum; and (b) the similarity of the shapes of the isotopic and adduct peaks around different m/z positions. The score for a given mass value represents the possibility and intensity of the analyte with the mass. Our algorithm involves several iteration steps until the result converge. This makes the scoring more accurate.

3) Post-processing

Ions whose masses are n/m times of a real ion may match part or all of the peaks corresponding to the real ion, where n and m are positive integers. So mass values that equal to n/m of the real ion mass can often get high scores too. A post-processing is done to recognize these false positives and remove them from the deconvoluted spectrum.

3. Results:

More than 20 spectra are used to test the deconvolution software. These spectra are obtained from Biological Mass Spectrometry Laboratory at Department of Biochemistry, at the University of Western Ontario. All spectra of them were obtained with a Micromass Q-TOF 2 (Quadrupole-Time-of-Flight) mass spectrometer instrument using a standard ESI source. The masses of sample proteins or protein mixtures range from 100 to 90,000 Dalton. We use the deconvolution software to process all these multiply charged spectra data and get the zero-charged spectra as results. Simultaneous Equations are used to manually determine the molecular weights of the analyte by two neighbor peaks. We use Mann et al.'s 'averaging algorithm' to solve the equations. For all of these test data, the deconvolution software gets all expected molecular weights compared to the manual analysis results.

We also compared our results with the results obtained from deconvolution with MaxEnt1 (Micromass, U.K.) program. For all these spectra, our deconvolution software FOOTPRINT gets equal or better

results. Most of our results have fewer artifacts. Specifically, for one spectrum of protein Eglin (Fig. 1), our software found both the eglin and its dimer (Fig. 2), while MaxEnt1's dimer peak is very weak and difficult to determine. This dimer is known to exist in solution.

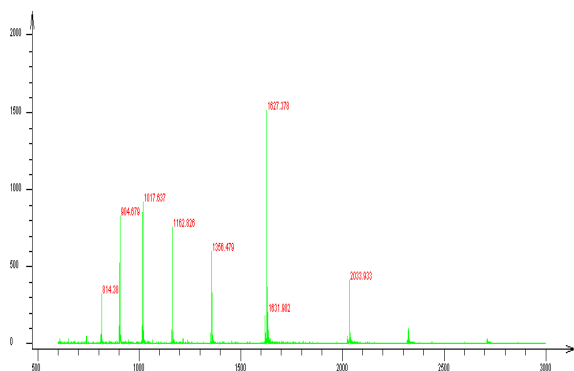


Fig.1: Spectrum of eglin.

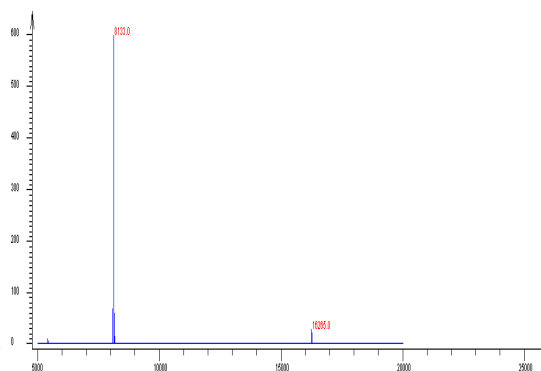


Fig.2: Deconvoluted spectrum of eglin

The most complicated spectrum of our test data is a mixture of four proteins (Fig. 3) with masses ranging from 1000 to 30,000 Da. When we set mass range from 10 to 40,000 Da as parameters, our software can identify three of them (Fig. 4). When we change the parameter of scoring-window size, we get also the other one with the least mass value (Fig. 5). We noticed that MaxEnt1 failed find the one with least mass value because it requires the parameter of minimum molecular mass to lie outside the raw data m/z range.

