MasSPIKE (Mass SPectrum Interpretation and Kernel Extraction) for Biological Samples

Parminder Kaur, Konstantin Aizikov, Bogdan Budnik and Peter B. O'Connor*
Mass Spectrometry Resource and Cardiovascular Proteomics Center, Department of Biochemistry,
Boston University School of Medicine, Boston

A suite of data reduction algorithms, called MasSPIKE, has been developed. It includes methods for modelling noise, isotopic cluster identification, charge state determination, picking isotopic peaks, and alignment of experimental and theoretical isotopic distributions for isotopically resolved mass spectra. The final output is an accurate and complete list of the monoisotopic mass information in the spectrum, which can be used to search the protein databases with high mass accuracy. If the sequence of the biomolecule under consideration is known, the observed masses are matched against the possible fragment ions and common losses and adducts.

Previously developed algorithms[1] for isotopic cluster identification were improved in order to identify overlapping clusters of very low and high charge states. For a given molecular mass, elemental composition was determined using the averagine[2] model for proteins. Rockwood's Mercury6[3] algorithm was incorporated into the program to generate theoretical isotopic distributions for a given elemental composition. The previously reported[1] charge state determination routine (Matched Filter approach) was modified using improved normalization of the experimental and theoretical isotopic distributions before determining the charge state. Special focus has been given to resolving overlapping isotopic distributions with different charge states. All these methods have been integrated into MasSPIKE.

MasSPIKE was successful in resolving 9 overlapping isotopic distributions, with multiple distributions sharing isotopic peaks, in the spectrum of a biologically derived sample (Ubch10 protein from an immunoprecipitated whole cell lysate). The complete analysis of a top down spectrum of Bovine Carbonic Anhydrase elucidated 165 isotopic distributions which were assigned to the closest theoretical masses of the fragments. A typical case is shown in Fig 1. The Matched Filter approach used for charge state determination was found to work correctly in 91% cases when a total of 775 isotopic distributions of myoglobin with charge states ranging from 8-22 and S/N ranging from 2-100 were investigated. An isotope alignment module has been developed to align the experimental and theoretical distributions and was tested using 3150 Monte Carlo simulations, with each simulation generated from 100 ions. These tests revealed that the Maximum Likelihood alignment method correctly aligns 85% of the time, as compared to the least squares error method which gave 76% correct results. MasSPIKE was applied to the intact protein and top-down ESI spectra obtained from a number of samples obtained from biological sources. Alpha chain of Hemoglobin from human blood samples was used to distinguish between a mutated alpha variant and a normal person. Beta chain of Hemoglobin revealed differences between a beta sickle and normal beta chain. Monoisotopic masses for large intact proteins (20-23 kDa) were determined from a single isotopic cluster with an accuracy of less than 5 ppm.

References

[1] P Kaur; K Aizikov; P B O'Connor, Proceedings of the 52nd Conference of the American Society of Mass Spectrometry 2004.

[2] M W Senko; S C Beu; F W McLafferty, "Automated Assignment of Charge States from Resolved Isotopic Peaks for Multiply Charged Ions", J. Am. Soc. Mass Spectrom.; 1995; 6, 52-56

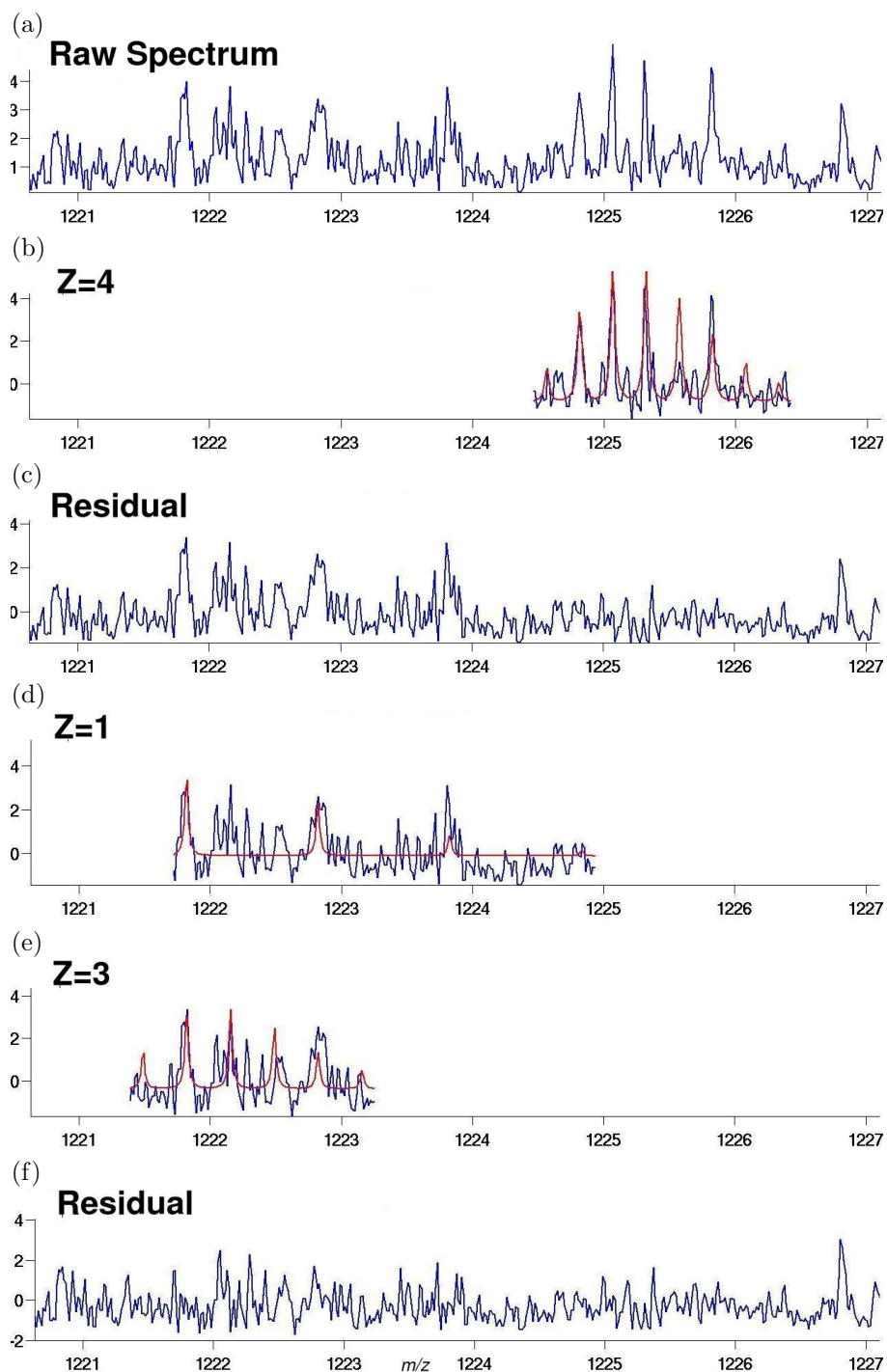[3] A L Rockwood, "Ultrahigh-Speed Calculation of Isotope Distributions", Anal Chem; 1996; 68; 2027-2030

Figure 1: (a) EID of a BCA fragment (b) Z=4 being assigned, r=0.731 (c) Residual after subtraction (d) Z=1 being assigned, r=0.54 (e) Z=3 being picked sharing 2 peaks with Z=1, r=0.495 (f) Residual after subtraction of assigned charge states