

A New Algorithm for Minimizing Chemical Noise in LC-MS: Matched Filtration with Experimental Noise Determination (MEND)

Victor P. Andreev, Tomas Rejtar, Hsuan-shen Chen, Eugene V. Moscovets, Alex R. Ivanov and Barry L. Karger

Barnett Institute, Northeastern University, Boston, MA 02115

Introduction

Noise in LCMS is both random (white) and chemical (colored). In MALDI-MS, chemical noise results from matrix clusters, in ESHMS from mobile phase impurities. Noise can cause either false negative or false positive identifications of sample components by masking or mimicking the signal. In addition, chemical noise can reduce mass accuracy by shifting the centroids of MS peaks. Chemical noise is more difficult to remove than white random noise because it has a pattern in the m/z domain similar to that of the signal. Our new algorithm (MEND-matched filtration with experimental noise determination) is able to minimize both chemical and random noise by exploiting the differences in the patterns of noise and signal in the chromatographic time domain. By performing filtration in the chromatographic time domain, MEND avoids distortion of peak shapes in the m/z domain and by minimization of noise improves accuracy of peak centroids.

Algorithm

A typical LC-MS data set consisted of 3000 spectra with 130,000 m/z data points. First, MEND determined noise characteristics from "vacant" extracted ion chromatograms (EIC) that contained no chromatographic peaks. Noise characteristics were found to be m/z dependent (for both LC-MALDI-MS and LC-ESI-MS), so transfer functions for matched filtration were separately determined for a number of m/z regions (~200). Then, transfer functions were used for matched filtration of EICs for each m/z value. Chemical and random noise were substantially reduced with no distortion in MS peak shapes. Next, peak picking was performed based on the examination of denoised data both in the chromatographic time and m/z domains.

Results

MEND was used for denoising of data sets from LC-MALDI-MS (AB 4700 TOF/TOF) and LC-ESI-MS (AB Maine) of tryptic digests of model protein mixtures and SXC fractions of a yeast lysate. MEND minimized chemical and random noise both for LC-MALDI-MS (Figs. 1-4) and LC-ESI-MS (Fig. 5) and increased SN by a factor of 5-8 relative to original spectra and by a factor of 2-3 relative to matched filtration with white noise assumption. By minimizing chemical noise, MEND improved the mass accuracy of peak centroids. Distances between the first and second isotopes of the isotopic clusters in the denoised data set were shown to be within 1 ppm mass accuracy of the theoretically predicted value of 1.003 (Fig. 4).

MEND was used for analysis of an LC-MALDI-MS data set of the tryptic digest of a cleavable ICAT labeled model mixture of 10 proteins. MEND found a higher number (by a factor of 1.3) of ICAT pairs relative to a program that only minimized white noise. This increase resulted from S/N improvement (detection of low abundant labeled peptides, Fig. 6), from mass accuracy improvement (more precise determination of the distance between the light and heavy labeled peptides within the pair, Fig. 7) and from removal of matrix related peaks masking the members of ICAT pairs (Fig. 8). As a result, 28 additional ICAT labeled peptides were identified (Fig. 9). Good agreement (within 30%) between experimentally determined and expected abundances for each of 10 model proteins was found, even in the case of heavy/light ratio 10:1 and 1:10. (See also posters: MPK 214, WPW 438, WPW 449)

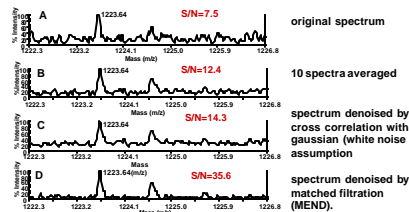


Fig. 1. Spectra from LC-MALDI-MS denoised by various approaches. Sample: mixture of 10 standard peptides. Spectral region of the peak corresponding to Tyr8-Bradykinin.

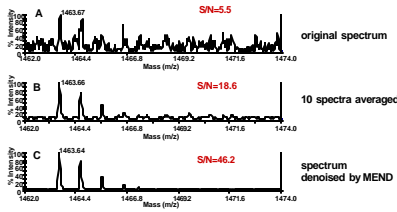


Fig. 2. Chemical noise suppression by MEND in spectra from LC-MALDI-MS of complex mixture. Sample: SXC fraction of the tryptic digest of yeast sample. Only limited spectral region is presented for clarity.

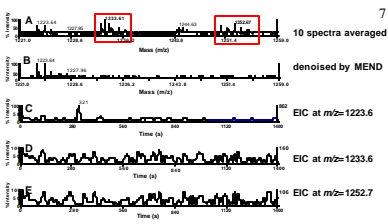


Fig. 3. Removal of matrix related peaks in spectra from LC-MALDI-MS by MEND. Sample: mixture of 10 standard peptides. Spectral region of the peak corresponding to Tyr8-Bradykinin presented. $m/z=1233.6$ corresponds to $[M+2H]^+$ from internal standard

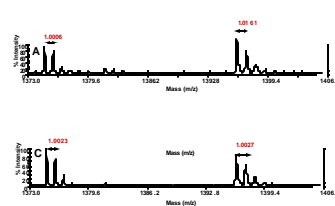


Fig. 4. Influence of denoising on mass accuracy in LC-MALDI-MS. Sample: SXC fraction of tryptic digest of yeast sample. Limited spectral region presented for clarity. A - 10 spectra averaged; B - spectrum denoised by (MEND). Theoretically distance between isotopes in the isotopic cluster should be equal to 1.003.

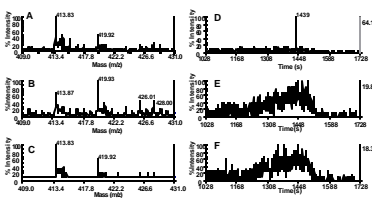


Fig. 5. Representative example of denoising of LC-ESI-MS data by MEND. Sample: tryptic digest of the mixture of 10 proteins. Limited spectral region is presented for clarity. A - original spectrum; B - 10 spectra averaged; C - spectrum denoised by MEND; D - EIC at $m/z=413.8$; E - EIC at $m/z=426.0$; F - EIC at $m/z=428.0$

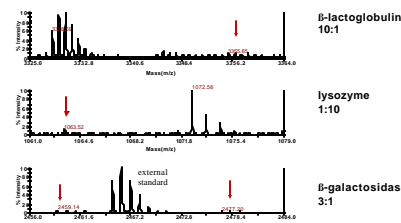


Fig. 6. Examples of additional ICAT pairs picked by MEND. Sample: tryptic digest of the mixture of 10 ICAT labeled proteins

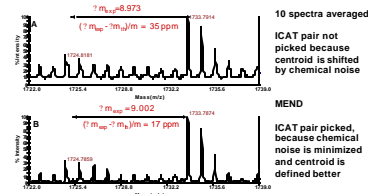


Fig. 7. Picking additional ICAT pairs due to improved mass accuracy. Theoretically $\delta m = 9.03$. Mass tolerance used for selection of pairs 20 ppm.

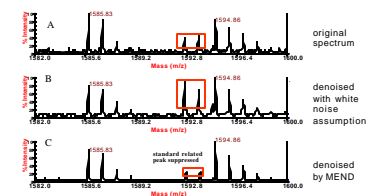


Fig. 8. ICAT pair selected by MEND due to suppression of sodium adduct of the internal standard. Sample: tryptic digest of the mixture of 10 dICAT labeled proteins. Spectral region of the ICAT pair of peptide SVIPSDGPSVACVK from human transferrin. Light - $m/z=1585.83$, heavy - $m/z=1594.96$. Peak at $m/z=1592.8$ is a sodium adduct of internal standard - Glu1-Fibrinopeptide B ($m/z=1570.67$).

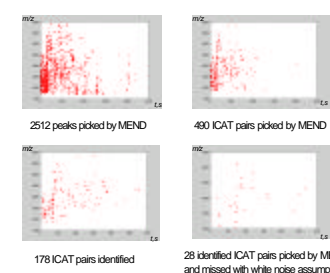


Fig. 9. Distribution of peaks and ICAT pairs picked by MEND for the tryptic digest of a dICAT labeled 10 protein mixture

ADVANTAGES OF MEND

- 5-8 times improvement in S/N relative to original spectra
- 2-3 times improvement in S/N relative to 10 spectra averaging
- Non-distortion of MS peaks due to denoising in the chromatographic time domain, improvement in mass accuracy, better determination of peak ratio in the isotopic cluster
- Minimization of peaks originating from matrix
- Increase in the number of selected ICAT pairs due to S/N and mass accuracy improvement
- Applicability of MEND principles to other hyphenated techniques, e.g. LC-MALDI/TOF, CE-ESI-MS, CE/ESI-MALDI, LC(GE)-NMR, LC(GE)-diode array UV-VIS detection