ASMS 2006

A NOVEL ADAPTIVE BACKGROUND SUBTRACTION ALGORITHM FOR REDUCTION OF CHEMICAL NOISE IN MASS SPECTRA.

Martin R Green[1], Keith Richardson[1], Martin Palmer[1], Richard Denny[1], Iain Campuzano[1], John Skilling[2].
[1]Waters Corporation, Manchester, England; [2]Maximum Entropy Data Consultants Ltd, Kenmare, Ireland.

OVERVIEW

PURPOSE
Reduction of chemical noise using a novel 'Adaptive' Background Subtraction (ABS) algorithm.

METHODS
Statistical analysis of the local intensity distribution for each nominal mass region. Estimation and removal of intensity distribution due to repetitive background.

RESULTS
Comparison of databank search results.
a) MALDI-MS PMF 500 attomoles Enolase –1
Peptide coverage without ABS……. **26%**
Peptide coverage with ABS……. **42%**

b) LC-ESI-MS "Protein Expression System" *E. Coli* Digest.
Proteins identified >95% confidence without ABS.. **57**
Proteins identified >95% confidence with ABS……**105**

INTRODUCTION

For many mass spectrometric techniques detection limits are restricted by the presence of chemical background. The precise nature of this background is rarely known. In Electrospray, chemical background may arise from clustering of solvent and analyte ions. In MALDI, chemical background may arise from matrix cluster ions. In these techniques the chemical background is complex, partially resolved and, commonly, periodic in nature. Here we describe a novel background subtraction algorithm capable of reducing chemical noise in mass spectra. Improvements in signal to noise are illustrated for complex protein digest samples using both MALDI and ESI modes .

EXPERIMENTAL

All results obtained using a Waters® Q-Tof™ Premier mass spectrometer.
MALDI: Sample—500 attomoles Yeast Enolase tryptic digest.
Matrix—Alpha cyano-4-hydroxy-cinnamic acid.
SwissProt protein databank.
ESI:
Sample mixture— 4ug of E. coli tryptic digest, 500fmol Yeast Enolase, 500fmol Bovine Serum Albumin,500fmol Yeast Alcohol Dehydrogenase,500fmol Rabbit Glycogen Phosphorylase.
Chemistry—Waters Atlantis® 300µm x 15cm dC18.
Conditions—Flow rate 4µL/min Gradient - 40% acetonitrile (0.1% formic acid) over 90mins.
Custom databank containing: *E. coli* proteins, 12,885 random protein sequences, yeast alcohol dehydrogenase , rabbit glycogen phosphorylase, bovine serum albumin and yeast enolase.
Processing: Waters ProteinLynx™ Global SERVER 2.2.5 (PLGS), and Waters Protein Expression Informatics.

ABS SUBTRACTION METHOD

A unique 'background' intensity distribution is calculated for each nominal mass bin in the mass spectrum. The method used to calculate the background distribution for a single nominal mass channel M is as follows.

1. Spectra are divided (conceptually) into m nominal mass channels centered on nominal mass channel M. (**Figure 1**)
2. Each of the m nominal mass channels are further subdivided into n discreet channels.
3. The intensity data within the first of the n subdivisions for each of the m nominal mass channels is collected into a single data set and an intensity value corresponding to the 50% quantile calculated. (**Figure 2**)
4. This procedure is repeated for each of the corresponding n subdivisions for each of the m nominal mass channels defining a new intensity distribution.
5. The calculated intensity distribution is subtracted from the input data centred on the nominal mass channel M. (**Figure 3**).

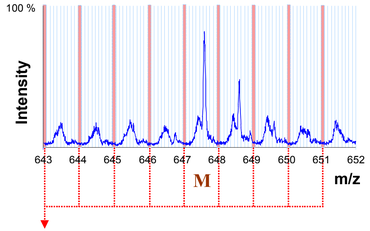

Figure 1. Example of an ESI mass spectrum subdivided into 9 nominal mass channels centered on nominal mass channel M. Intensity data from the first of the n sub-divisions is shown in figure 2.

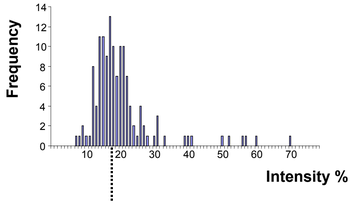

Figure 2. Distribution of intensities produced after collapsing data from first sub-division of each nominal mass channel shown in Figure 1.

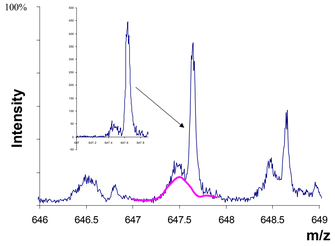

Figure 3. Portion of ESI mass spectrum showing input data, calculated background distribution and data after subtraction (Inset).

RESULTS-MALDI PMF

MALDI PMF (Peptide Mass Fingerprint) search results (SwissProt protein database) from 500 attomoles of Yeast Enolase tryptic digest were compared with and without application of adaptive background subtraction. Both PLGS and MASCOT® (Matrix Science, UK))search engines were used through the 'ProteinLynx' Browser.

Figure 4 A. Shows the MALDI MS spectra from 500 attomoles of Yeast Enolase-1 tryptic digest.
Figure 4 B. Shows the same spectra after application of the adaptive background subtraction algorithm.
Figure 5 A. Shows an expanded region of the spectrum shown in figure 4A prior to subtraction.
Figure 5 B. Shows an expanded region of the spectrum shown in figure 4B after subtraction.

Successful protein identification by MALDI-PMF, at low concentrations is often limited by the presence of background chemical noise. In this example the adaptive background subtraction algorithm efficiently reduces the strongly periodic chemical noise improving identification confidence.

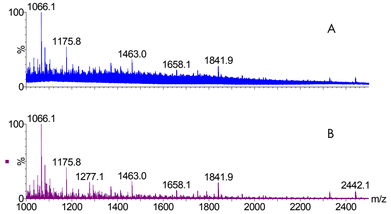

Figure 4. MALDI-MS 500 attomoles Yeast Enolase tryptic digest. A = before ABS. B = After ABS

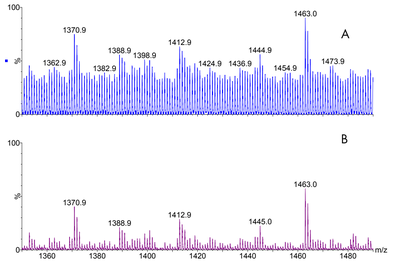

Figure 5. MALDI-MS 500 attomoles Yeast Enolase tryptic digest. A = before ABS. B = After ABS

Figure 6 A. Shows the PLGS databank search results without adaptive background subtraction.
Figure 6 B. Shows the PLGS databank search results after adaptive background subtraction. The signal to noise ratio of the spectra submitted for database searching is clearly improved.

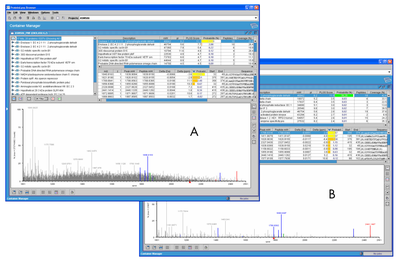

Figure 6. PLGS database search A. = without ABS. B. = With ABS.

Table 1 Shows a summary of the PMF results for Yeast Enolase –1 obtained using the PLGS and MASCOT search engines. The number of peptides identified and the percentage peptide coverage is shown. In both cases application of adaptive background subtraction resulted in significant improvement.

Search Engine	Number of Peptides NO ABS	Number of Peptides WITH ABS
PGLS	10 (26%)	19 (42%)
MASCOT	6 (17%)	10 (30%)

Table 1. Summary of PLGS and MASCOT databank search results for the protein Yeast Enolase -1 with and without ABS.

RESULTS LC-ESI-MS

LC-ESI-MS data was acquired from a sample of *E. Coli* tryptic digest containing four non *E. Coli* standard protein digests spiked at low concentrations, using the Waters Protein Expression System. Precursor and Fragment data were acquired by alternating between low (MS) and elevated (MS^E) collision energy. The data was processed using Waters Protein Expression System Informatics before and after application of ABS. Results were searched using the PLGS search engine against a custom databank containing *E. Coli* proteins, the four spiked proteins and a number of randomly generated protein sequences.

Figure 6. Shows the PLGS search result window for the Protein Expression analysis using ABS.

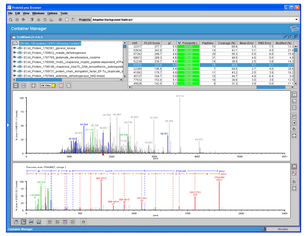

Figure 6. PGLS database search results. Protein expression with ABS.

Figure 7. Shows a Venn diagram illustrating the number of proteins identified (>95% confidence). Approximately twice as many proteins were confidently identified when ABS was used compared to processing without ABS.

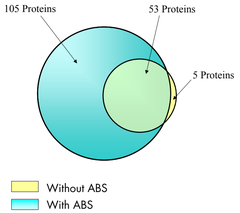

Figure 7. Proteins identified with >95% confidence.

Figure 8. Shows comparison of peptide coverage for each of the 53 proteins common to the data sets shown in Figure 7. Consistently higher peptide coverage is reported after processing with ABS .

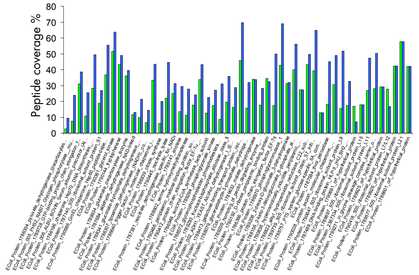

Figure 8. Comparison of peptide coverage assigned proteins.
• = With ABS
• = Without ABS

Table 2 Shows a summary of the PLGS search results (> 95% confidence) for the four spiked protein standards. The number of peptides identified and the percentage peptide coverage is shown. Using ABS all four target proteins were identified. Without ABS only three were identified. In addition the results after processing with ABS show significantly higher peptide coverage for those proteins identified.

Protein	Number of Peptides NO ABS	Number of Peptides WITH ABS
BSA	4 (6%)	14 (24%)
PhosB	5 (8%)	14 (22%)
ADH	3 (9%)	7 (25%)
Enolase	X	5 (15%)

Table 2. Summary of PLGS search results (> 95% confidence) for protein standards spiked into E Coli Tryptic digest.

CONCLUSION

- **The Adaptive Background Subtraction (ABS) algorithm is capable of dramatically reducing background with short range (~ 1Da) periodicity and longer range intensity variation.**
- **The algorithm adapts to the nature of the background throughout the mass spectrum resulting in optimum signal to noise throughout the acquired mass range.**
- **For very complex protein digest samples application of ABS consistently improves peptide coverage and protein identification confidence from protein databank searches.**

720001724EN

TO DOWNLOAD A COPY OF THIS POSTER VISIT WWW.WATERS.COM/POSTERS