PosterREPRINT

Waters

AN AUTOMATED METHOD TO SELF-CALIBRATE AND REJECT NOISE FROM MALDI PEPTIDE MASS FINGERPRINT SPECTRA

Jeffery M Brown, Neil Swainston, Dominic O. Gostick, Keith Richardson, Richard Denny, Steven Leicester and Phillip Young

Waters Corporation, Floats Road, Wythenshawe, Manchester, M23 9LZ United Kingdom

Presented at ASMS 2002, Orlando, Florida, 3rd-6th June, 2002

Overview

Purpose

Investigation of a new algorithm for self calibration and determination of noise levels of MALDI-TOF peptide mass fingerprint spectra.

Methods

The algorithm which uses the known mass sufficiency function of peptides is applied to data from several standard protein digests.

Improvements in mass measurement accuracy and noise level determination are measured using a new data base search engine.

Results

Mass measurement accuracy for data acquired without internal calibration improved from 28.9 ppm RMS to 13.1 ppm RMS.

Optimal noise level determination provides more confident data base identifications for low signal:noise data and minor components of mixtures.

Introduction

MALDI peptide mass fingerprinting (PMF) is an established method for identification of proteins extracted from 2D electrophoresis gels. Successful identification by matching of observed peptides against those theoretically derived from protein data banks depends on the mass measurement accuracy and signal to noise of the peak lists considered. Usually an intensity threshold is applied to the spectral data to remove noise peaks from the real peptide peaks. If the threshold is set too high, small components of peptide mixtures are often ignored and remain unidentified. If the threshold is too low a large number of noise peaks may be submitted to the data base search which may "confuse" the search algorithm resulting in random incorrect protein matches.

We have developed a new algorithm based on the mass sufficiency function of peptides that rejects noise from PMF spectra. Furthermore, the mass sufficiency function may be used to "self calibrate" the peptide masses of MALDI PMF spectra without the use of an internal reference.

Using this new algorithm peptide peak lists are optimally filtered for noise and the differences between the theoretical and matched peptide masses are corrected to better than 20 ppm RMS without using internal references.

Methods and Instrumentation

All data were acquired automatically using a Micromass M@LDI-R time of flight mass spectrometer in reflectron mode operating at approximately 12,000 (FWHM) resolution between m/z 1000 and m/z 3000. Data were initially near point lock mass corrected using a single monoisotopic tryptic peptide peak (MH+ 1618.844, VLGIDGGEGKEELFR) acquired from an alcohol dehydrogenase digest loaded near to each sample well.

A variety of protein standards were separated by 1D SDS-PAGE (BioRad) and the resulting bands were digested by porcine trypsin (Promega). 1 micro-liter of each solution (approximately 100fmole) were mixed on the sample plate with 1 micro-liter of matrix alpha-cyano-4-hydroxycinnamic acid (concentration of 10 g/L in 50:50 water acetonitrile).

MICROMASS MS TECHNOLOGIES

Two mixtures were also loaded - 100 fmole:10 fmole mix of BSA and Beta-galactosidase and a 100 fmole:100 fmole mix of Phosphorylase B and Beta Galactosidase.10 of 10 laser shot spectra were acquired and summed from each sample well ("strong data") and the weakest spectra of the 10 spectra were also recorded ("weak data").
Each spectrum was automatically post processed and data base searched using ProteinLynx Global Server software (Micromass, PLGS 2.0). This software smoothes, background subtracts, centroids, de-isotopes, lock mass corrects, creates peak lists and searches against protein digest databanks.

The self calibration algorithm determined the optimal noise threshold and calibration factor. These parameters were applied to the data prior to data base searching.

Description of Self Calibration Algorithm

Peptide masses have been previously shown to have well defined fractional mass values and occupy relatively narrow distribution zones on the mass scale[1].

The fractional mass sufficiency (*sufficiency*) may be calculated from equation (1) where *mass* is the nominal value of the peptide mass and *suff_fact* is the mass sufficiency factor (previously reported as 0.00048).

$$sufficiency = suff\_fact \,.\, mass \qquad \text{(eq1)}$$

The mass distribution zone or window (mass_window) of 95% of all amino acid combinations may be calculated from equation (2).

$$mass\_window = 0.19 + mass/10000 \qquad \text{(eq2)}$$

Figure 1 depicts the zones of all typical and non-typical peptides for the nominal peptide masses of 1000 Da and 1001 Da.

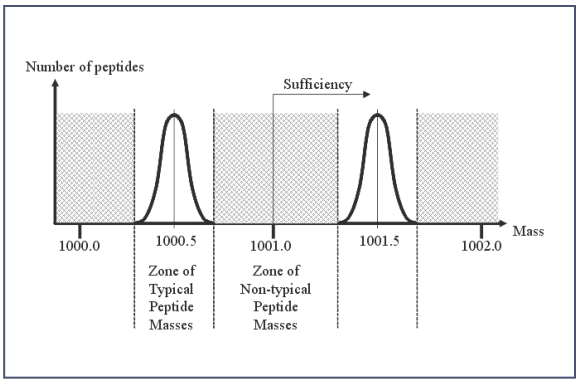

Figure 1. Mass Sufficiency and Mass Scale Zones for Typical and Non-typical Peptides for the Nominal Peptide Masses 1000 Da and 1001 Da

The function *deviation(gain)* shown in equation (3) represents the overall mass deviation between the expected and observed mass sufficiency and is minimised by varying the spectrum gain factor (gain) over a +/-200 ppm window, massn is the mass value of the n'th peak of the uncorrected mass spectrum, (Int is a rounding function to the nearest integer).

$$deviation(gain) = \Sigma\{Int\,[mass_n.(1\text{-}sufficiency).gain] - mass_n.(1\text{-}sufficiency).gain\}^2 \qquad \text{(eq3)}$$

Noise Rejection Threshold

It is assumed that peaks occurring in the zone of non-typical peptide masses are noise and that the noise intensity level extends into the zone of typical peptides.

The ratio of the sum of the intensities of non-typical peptides to the sum of the intensities of typical peptides may be used as a measure of "data quality" or signal:noise. The % intensity threshold (or cut off threshold) may be optimised for each individual spectrum by fixing the required "data quality" at a nominal value of 10%.

PosterREPRINT

Waters

Results

Figures 2a to 8b show the "strong" and "weak" spectra for the 7 tryptic digest samples. For each spectrum both the near point correction and self calibration algorithm mass measurement accuracies are reported in **Table 1**.

The RMS mass accuracy of each spectrum was determined by submitting each peak list against SwissProt (40.6) database and calculating the RMS errors for the differences between the calibrated data and the significant matching peptides for the correct protein match. Searches were performed with a 100 ppm allowable mass error.
Fixed carbomidomethyl cysteines and variable oxidised methionine modifications were considered.

Table 1 also shows the noise threshold for each spectrum as determined by the algorithm.

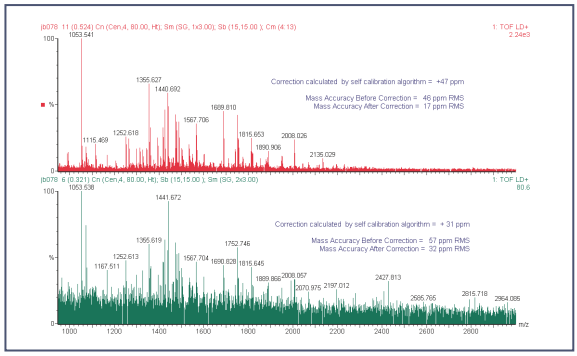

Figure 2. Glycogen Phosphorylase a) 100 shot spectra summed, b) 10 shot spectra

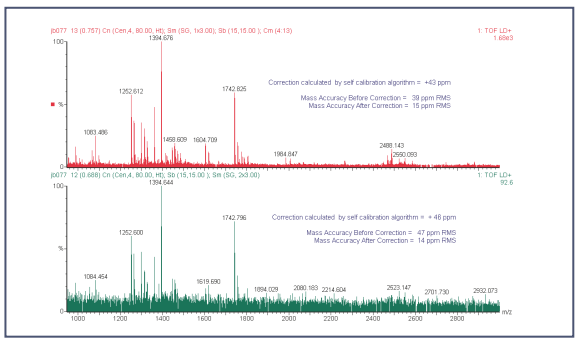

Figure 3. Beta Galactosidase a) 100 shot spectra summed, b) 10 shot spectra

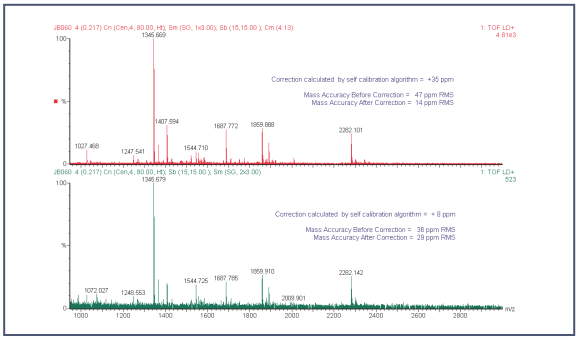

Figure 4. Ovalbumin a) 100 shot spectra summed, b) 10 shot spectra

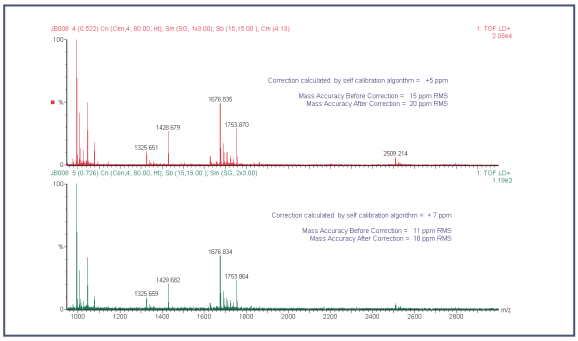

Figure 5. Chicken Lysozyme a) 100 shot spectra summed, b) 10 shot spectra

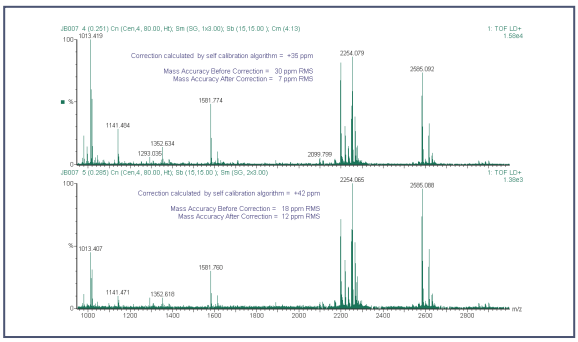

Figure 6. Carbonic Anhydrase a) 100 shot spectra summed, b) 10 shot spectra

Waters

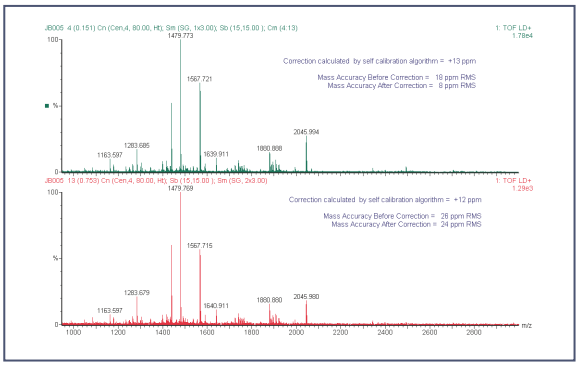

Figure 7. Serum Albumin and Beta Galactosidase (10:1 mix) a) 100 shot spectra summed, b) 10 shot spectra

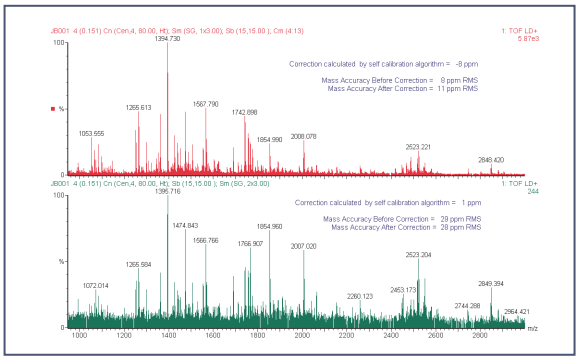

Figure 8. Phosphorylase B and Beta Galactosidase (1:1 mix) a) 100 shot spectra summed, b) 10 shot spectra

		Mass Accuracy (RMS ppm matching peptides)		
		Near Point Lock Mass Calibration	Self Calibration	Predicted Noise Threshold %
Stronger (100 shot) data	Spectrum			
Glycogen Phosphorylase	2 (a)	46	17	3
Beta Galactosidase	3 (a)	39	15	3
Ovalbumin	4 (a)	47	14	1
Chicken Lysozyme	5 (a)	15	20	0.5
Carbonic Anhydrase	6 (a)	30	7	0.5
Serum Albumin+Beta Galactosidase	7 (a)	18	8	0.25
Glycogen Phosphorylase+Beta Galactosidase	8 (a)	8	11	1
Mean		29.0	13.1	
Weaker (10 shot) data				
Glycogen Phosphorylase	2 (b)	57	32	29
Beta Galactosidase	3 (b)	47	14	14
Ovalbumin	4 (b)	36	29	6
Chicken Lysozyme	5 (b)	11	16	1.5
Carbonic Anhydrase	6 (b)	18	12	2
Serum Albumin+Beta Galactosidase	7 (b)	26	24	1
Glycogen Phosphorylase+Beta Galactosidase	8 (b)	29	28	10
Mean		32.0	22.1	

Table 1. Mass measurement accuracy, near point calibration and self calibration algorithm and predicted noise thresholds

		Thresholding Method		
	Figure	Peaks used above 5% threshold	40 most intense peaks used	Peaks used above threshold determined by algorithm
Glycogen Phosphorylase as top hit from weak spectrum	2 (b)	NO	YES	**YES**
(Serum Albumin) as top hit Major component of mixture	7 (a)	YES	NO	**YES**
(Beta Galactosidase) Identified as 2nd hit Minor component of mixture	7 (a)	NO	NO	**YES**

Table 2. Success of data base searching using various noise threshold methods

Figures 9 to 13 show PGS 2.0 database search results for 2 sets of data. The weak spectrum **Figure 2 (b)** Glycogen Phosphorylase was identified correctly when the noise threshold applied was determined by the algorithm (at 29%) or if the 40 most intense peaks were searched. (searching NRDB, +/-50ppm tolerance). The sample was not identified when using a 5% noise threshold.
For the stronger spectrum **Figure 7(a)** (mix of 10:1) BSA and Beta-Galatosidase, both components were identified correctly (1st and 2nd hits) when the applied noise threshold was determined by the algorithm at 0.25%, using peaks above a 5% noise threshold or 40 most intense peaks failed to give the correct results. (searching SwissProt, +/-50ppm tolerance).

Figure 9. Weak data from Figure 2b - 5% noise threshold - Glycogen Phosphorylase - 2nd hit (1st hit incorrect random match)

Figure 10. Weak data from figure 2b - 29% noise threshold (determined by algorithm) matching peptides from Glycogen phosphorylase as correct top hit

Figure 11. Binary Mixture 5% noise threshold - matching peptides from Serum Albumin top hit, 2nd hit incorrect

Figure 12. Binary Mixture 0.25% noise threshold (determined by algorithm) - matching peptides Serum Albumin top hit, (2nd hit Beta Galactosidase -see Figure 13)

Figure 13. Binary Mixture 0.25% threshold (as determined by algorithm) - Serum Albumin top hit, matching peptides from correct 2nd hit - Beta Galactosidase shown

Conclusion

Improvements in mass measurement accuracy using an axial MALDI-TOF mass spectrometer have been observed by the application of a new self calibration algorithm to peptide mass fingerprint spectra.

Mass measurement accuracy improved from 29.0ppm RMS (external calibration) to 13.1ppm RMS for higher signal:noise data. For lower signal:noise data the mass accuracy improved from 32.0ppm RMS to 22.1ppm RMS. This improvement enables data base searching using a mass error window of +/- 50ppm rather than +/- 100ppm thereby improving specificity and reducing the probability of erroneously matching random noise peaks.

The algorithm also predicts the optimum noise threshold for each spectrum. Peaks above this threshold were used to identify proteins from poor signal:noise data. Low level components of a binary mixture were also identified with greater confidence than if a fixed threshold or 40 most intense peaks had been submitted to the database.

References

[1] Mann, Possible peptide masses. Proceedings of the 43rd ASMS, Atlanta, 1995

PosterREPRINT

Author to whom all correspondence should be addressed:
Jeff Brown
Waters Corporation
(Micromass UK Limited)
Floats Road, Wythenshawe
Manchester, M23 9LZ
Tel: + 44 (0) 161 946 2400
Fax: + 44 (0) 161 946 2480
e-mail: jeff.brown@micromass.co.uk

WATERS CORPORATION
34 Maple St.
Milford, MA 01757 U.S.A.
T: 508 478 2000
F: 508 872 1990
www.waters.com

Made in the United Kingdom

Waters
RIGHT ON TIME.

MICROMASS®
MS TECHNOLOGIES

Certificate No: 951387