Peptide Identification with High Mass Accuracy using Lock Mass Calibration

Bioinformatics Solutions Inc.

University of Waterloo

M. Ziaur Rahman[1], Mingjie Xie[1], Denis Yuen[1], Bin Ma[2]
Bioinformatics Solutions Inc, Waterloo, ON [1]
David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, ON [2]

Introduction

High mass accuracy is an important goal in liquid chromatography-mass spectrometry experiments. Some Waters systems allow periodically injecting a calibration compound with a known m/z peak into the mass spectrometer. The variation between the expected and measured m/z peak values can be used to calibrate the m/z values in the sample data. A mass calibration routine has been implemented that accepts raw Waters data as input, calibrate the detected masses and save the data in mzXML format for subsequent analyses. The calibration routine can significantly reduce the mass error. The effect of the lock mass calibration on database search and de novo sequencing is also examined.

Methods

The calibration routine accepts the Waters raw file and the calibration mass as input. It then separates the calibration scans from data scans. Next, for each data scan, the calibration factors are calculated. The calibration factor of a data scan is the weighted average of the ratio of the expected lock mass and the detected lock mass of the neighbouring calibration scans. The calibration factor is then applied to all m/z values of the data scan. The modified data is saved in mzXML format.

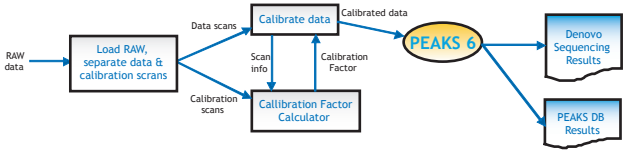

Figure 1. Lock mass calibration and peptide identification

PEAKS 6 software is used for peptide identification with both database search and de novo sequencing. The results on the pre-calibrated and calibrated data are compared to examine the effect of the lock mass calibration.

Results

A high resolution MS dataset with 921 MS/MS spectra were tested. The data was obtained from mammalian mitochondrial cell line on nanoAcquity HPLC coupled with a Waters Q-ToF-Premier mass spectrometer. Acquisition was done with lock spray using Glu-Fib 2+ for the MS data and Leu-Enk 1+ for the MS/MS data as lock mass correction.

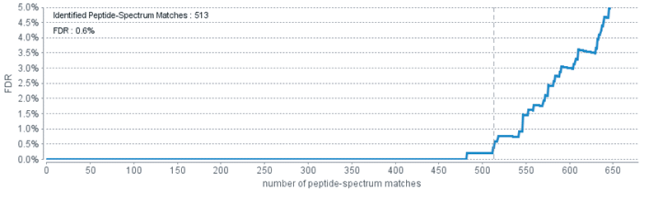

Figure 2. The False Discovery Rate (FDR) curve for pre-calibrated data

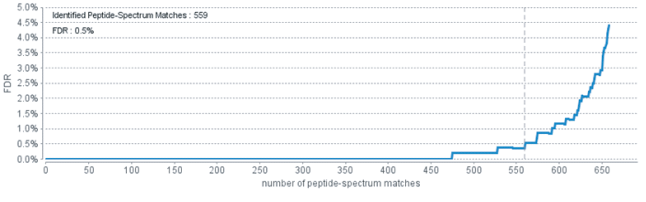

Figure 3. The False Discovery Rate (FDR) curve for calibrated data

The data are searched against the SwissProt sequence database. For the calibrated data, 559 peptide-spectrum matches (PSMs) were identified at 0.5% false discovery rate (FDR). 75% of the identified PSMs have a precursor mass error below 5 ppm. Whereas for the pre-calibrated data, 513 PSMs were identified with 0.5% FDR, where 9% and 77% of the identified PSMs have a precursor mass error below 5 ppm and 15 ppm respectively. Figure 2 and Figure 3 depict the FDR curves for pre-calibrated and calibrated data respectively, where FDR is plotted against the number of PSMs. Figure 4 and Figure 5 show the precursor mass error distribution of the PSMs for pre-calibrated and calibrated data respectively, where the precursor mass error (in ppm) is plotted against the number of PSMs and m/z. The result shows that the calibration reduced the mass error by more than half, and increased the number of confidently identified PSMs by 9%.

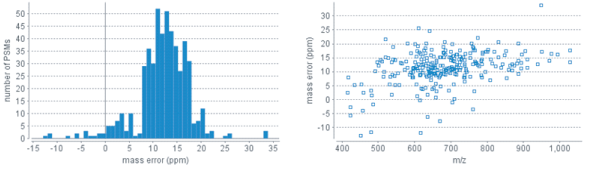

Figure 4. Precursor Mass Error (Observed m/z - Theoretical m/z) Distribution of the PSMs in Pre-Calibrated Data

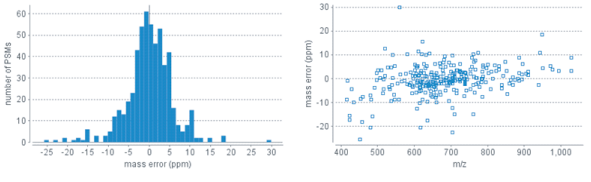

Figure 5. Precursor Mass Error (Observed m/z - Theoretical m/z) Distribution of the PSMs in Calibrated Data

The performance of de novo sequencing is also slightly improved by the recalibration. For calibrated data, 164 of 921 spectra have de novo confidence scores (ALC) greater than 70%; whereas the number is 158 for the pre-calibrated data.

Conclusions

A lock mass calibration procedure is implemented, improving peptide identification with high mass accuracy using database search and de novo sequencing as demonstratd by PEAKS 6.

Acknowledgement

The authors would like to thank Dr. Randy Whittal and his group at University of Alberta for making available the data and valuable comments.