A multi-region recalibration approach for obtaining ppm-range mass measurement accuracy for LC-MS analyses of complex proteomics mixtures

Aleksey V. Tolmachev and Richard D. Smith

Pacific Northwest National Laboratory, Richland, WA 99352

Overview

Mass spectrometry combined with on-line separation techniques has become a powerful tool for characterizing complex mixtures, such as protein digests in proteomics studies. Accurate mass measurements of complex mixtures can be compromised due to the variability of mass spectra acquisition conditions over the course of an on-line separation. We have developed an approach that takes these variable mass measurement conditions into account and corrects the mass calibration according to multiple regions of parameters critical for accurate mass measurement of a specific *m/z* peak. We describe the mass accuracy improvement obtained for LC-MS data acquired using both custom and commercial FTICR systems, as well as TOF MS systems and LTQ Orbitrap

Introduction

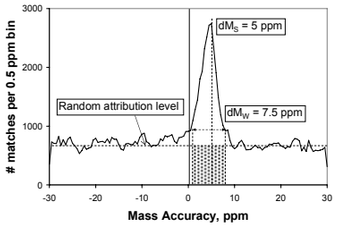
The approach is based on statistical matching of measured masses to a set of putative accurate masses, generally a subset of peptides previously identified by LC-MS/MS methods for developing a database for the organism under investigation [1]. To compensate for calibration variations due to variable ion populations, all mass spectra are grouped according to the total ion current (TIC) measured for each mass spectrum. Similarly, multiple regions of *m/z*, ion abundance, and elution time are used. As a result, multiple calibrations are applied for a single separation-MS dataset; each one is optimized for accurate mass measurements in a narrow range of parameters.

Methods

A table with 10^3 to 10^5 entries of theoretical *m/z* values is likely to have only a fraction of those results match the LC-MS experiment. Creating a histogram of matches between experimentally detected and theoretical molecular masses can be used for statistical evaluation. Random matches contribute to a plateau that represents the probability of false attribution. The peak in the histogram corresponds to true matches. The peak offset (dM_s) and width (dM_w) estimates mass accuracy and precision respectively.

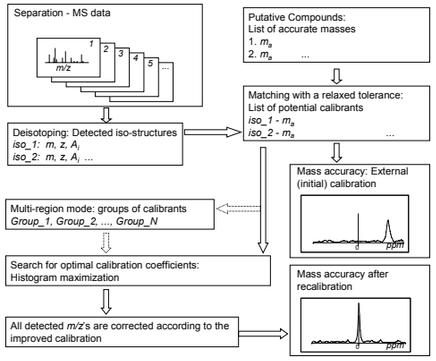
The mass accuracy histogram

- dM_S: systematic mass measurement error (accuracy)
- dM_W: the mass measurement error variation (precision)



Recalibration using the histogram maximization

The aim of recalibration is to move the histogram peak to 0 ppm and to minimize the peak width. This can be achieved by adjusting the calibration coefficients A, B of the FTICR calibration function: $m/z = A / (f + B)$, where f is the cyclotron frequency. The search for the optimal pair of values A, B is performed using small increments for each of the coefficients, covering a sufficiently large scope of all possible calibrations. Each trial calibration is tested using the mass accuracy histogram. The calibration pair that produces the largest number of matches falling into the 0 ppm bin is chosen as the final calibration.



Multi-region recalibration

The next level of mass accuracy improvement can be achieved by means of multi-region calibration. Since the measurement conditions can vary in course of LC-MS measurements, it is likely that the optimal calibration will similarly vary. To address this challenge we have applied a recalibration approach that uses multiple calibrations for a single separation-MS dataset. To compensate the calibration variations due to variable ion population, mass spectra are ranked according to their TIC values and grouped into N_{TIC} groups of spectra having similar TICs. Recalibration applied to each TIC group produces a number (N_{TIC}) of calibrations, each one optimized for a particular TIC range.

The multidimensional recalibration

Along with TIC, other parameters influencing the mass calibration include *m/z*, ion abundance and elution time. The collection of all detected masses is divided into multidimensional structure according to pre-selected ranges of these parameters, and recalibration applied individually to each group. Such a multidimensional recalibration procedure produces a multidimensional array of calibrations, each one optimized for narrow ranges of several parameters.

Relationships used for the recalibration

1. Instrument - specific calibration functions, e.g.:
   $m/z = A/(f + B)$ FTICR
   $m/z = C_{TOF}(t_i - t_0)^2$ TOF MS
2. Linear mass correction:
   $\Delta m/z = \delta_1 \, m/z + C_0$; $\delta_1 << 1$
   or $m/z_t = C_1 \, m/z + C_0$; $C_1 = 1 + \delta_1$
   where *m/z* and m/z_t are initial and corrected *m/z* values
3. Linear correction of the relative mass deviation:
   $\Delta m/z \; / \; m/z = \delta_1 \, m/z + C_0$
   or $x_1 = C_1 \cdot x + C_0$, where $x_1 = (m/z_t)^{-1}$, $x = (m/z)^{-1}$
4. Higher order corrections
   $\Delta m/z = C_0 + \delta_1 \, m/z + C_2(m/z)^2 + ... + C_N(m/z)^N$

Since very small variations are used, mass corrections produce results approaching those with instrument-specific calibration functions. Recalibration based on mass correction is instrument-independent and does not require instrument calibration information and access to raw data, such as measured peak frequencies or ion times of flight.

Results

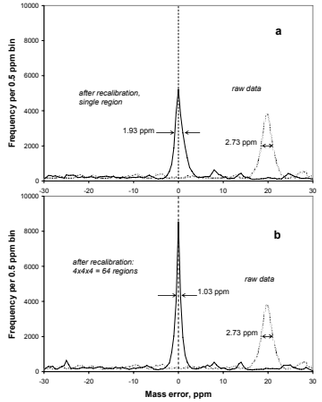
Recalibration applied to 11 tesla FTICR LC-MS data

The solution used for routine quality control (QC) analyses is a mixture of 23 peptides and 12 proteins digested with trypsin [2].

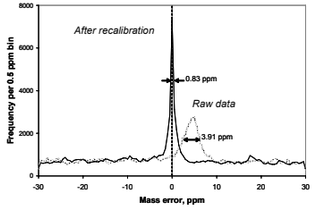
(a) A single calibration is applied, common for all 3382 specs

(b) Multi-region recalibration

4x4x4 = 64 regions for TIC, *m/z* and peak intensity, i.e. 64 different calibrations, applied according to specific conditions



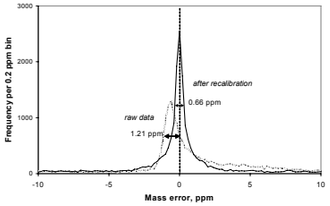
~1 ppm mass accuracy is obtained for complex proteomics LC-MS datasets, including bacterial, human, etc.



AMT dataset for *Neurospora Crassa* [3], 11 tesla LC-FTICR. Calibration regions for TIC, *m/z* and peak intensity: 10x2x10=200

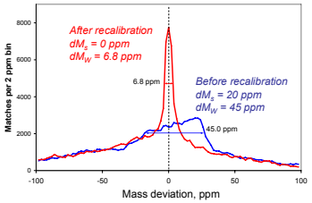
ThermoFinnigan LTQ FT

LC MS quality control data; AGC target set to 500K. Calibration regions for TIC, elution time and peak intensity are 4x16x8=512.



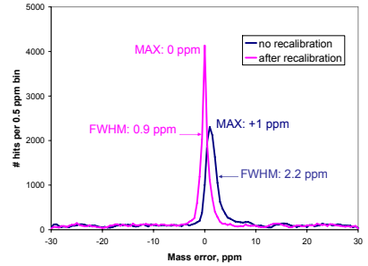
The "raw" mass accuracy histogram (gray) exhibits a wing tailing towards positive mass errors, up to ~10 ppm. The recalibration (black) improves the mass error spread from 1.2 to 0.7 ppm, corrects the histogram maximum position from -0.6 ppm to 0 ppm and the distribution "tail" is compensated to < 3 ppm .

Agilent TOF MS, QC mixture LC-MS data

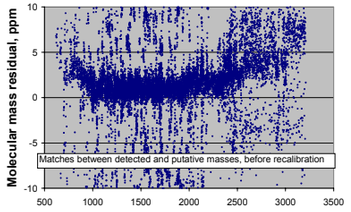


QC data from Agilent TOF MS. Calibration regions for TIC, elution time and peak intensity are 2x32x8=512. Mass accuracy defined as the FWHM of the histogram is 45 ppm (raw, blue curve) and 6.8 ppm (after recalibration, red). The optimal settings use elution time devided into 32 regions, indicating temporal variations of the mass calibration, possibly due to temperature drifts and power supply instabilities [4].

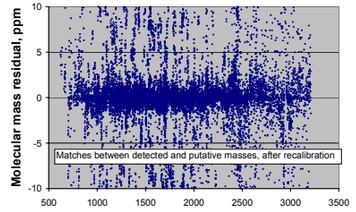
ThermoFinnigan LTQ Orbitrap



Recalibration results for QC mixture LC MS using Thermo LTQ Orbitrap. The separation lasted for 100 minutes and included 3211 MS spectra. The mass error distribution before recalibration (blue) shows +1 ppm systematic mass error and a wing of positive errors tailing up to +10 ppm. The map of hits, mass error vs. spectrum #, shows a clear time-dependent trend.



Recalibration uses 4x16x8 regions for TIC, LC separation time and peak intensity, total 512 3D regions. Precision is improved from 2.2 ppm to 0.9 ppm FWHM, as seen in the upper histogram (magenta curve). Mass accuracy is corrected from +1 to 0 ppm and the positive error wing is fully compensated. The map of matches after recalibration shows that the time-dependent trend is compensated.



Conclusions

- The recalibration procedure corrects the mass calibration according to parameters influencing mass accuracy, such as the spectrum TIC and elution time, individual ion abundance and *m/z* values.
- The procedure involves an automated analysis of the mass accuracy histogram and is robust with respect to large quantities of false matches, thus large datasets of $>10^5$ detected iso-structures and up to $\sim 10^5$ potential mass tags can be used.
- The mass accuracy improvement is obtained for LC-MS data acquired using both custom and commercial FTICR systems, as well as TOF MS systems and LTQ Orbitrap.
- Both the accuracy and precision of mass measurements are improved.
- Sub-ppm mass measurement accuracy is obtained for complex proteomics samples, providing significantly improved identification quality.

Acknowledgements

The authors gratefully acknowledge Drs M. E. Monroe, W. Qian and N. Jaitly for help with PMT databases and informatics support, A. Shukla, T. Clauss and R. J. Moore for LC-FTICR QC data, A. Norbeck and D. G. Camp for providing Neurospora crassa data, J.N. Adkins and N.G. Colton for editing. Portions of this work were supported by the National Center for Research Resources (RR18522), the National Institute of Allergy and Infectious Diseases NIH/DHHS through interagency agreement Y1-AI-4894-01, and the National Institute of General Medical Sciences (NIGMS, R01 GM063883). Experimental portions of this research were performed in the Environmental Molecular Sciences Laboratory, a U.S. Department of Energy (DOE) national scientific user facility located at the Pacific Northwest National Laboratory (PNNL) in Richland, Washington. PNNL is a multiprogram national laboratory operated by Battelle for the DOE under Contract No. DE-AC05-76RLO 1830.

References

1. Smith, R. D., Anderson, G. A., Lipton, M. S., Pas˙a-Tolic´, L., Shen, Y., Conrads, T. P., Veenstra, T. D. & Udseth, H. R. (2002) Proteomics 2, 513–523.
2. Purvine, S.; Picone, A. F.; Kolker, E. OMICS 2004, 8, 79-92.
3. Schmitt S, Prokisch H, Schlunck T, Camp DG 2nd, Ahting U, Waizenegger T, Scharfe C, Meitinger T, Imhof A, Neupert W, Oefner PJ, Rapaport D. Proteomics 2005, 6(1), p.72-80
4. C. Eckers, J.C. Wolff, N.J. Haskins, A.B. Sage, K. Giles, R. Bateman, Anal. Chem. 2000, 72, 3683

Contact Information

Aleksey Tolmachev, Ph. D.
Pacific Northwest National Laboratory
P.O. Box 999, MS K8-98, Richland, WA 99352
Phone: (509) 376-9963 E-mail: tolmachev@pnl.gov