

# Post-Acquisition Mass Calibration and Filtering of High-Resolution Mass Spectrometry Data

Damon May<sup>1</sup>, Matthew Fitzgibbon<sup>1</sup>, and Martin McIntosh<sup>1</sup>  
<sup>1</sup>Fred Hutchinson Cancer Research Center, Seattle WA

## OVERVIEW

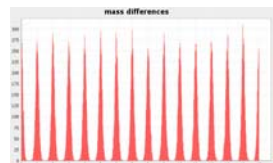
Many modern mass spectrometry analysis workflows depend on reliable LC-MS identification of peptides and accurate ascertainment of ion mass [Zubarev 2006]. Problems with mass spectrometer instrument calibration can reduce mass accuracy greatly.

A new toolset, within the msInspect platform [Bellew 2006], dynamically recalibrates peptide mass spectrometry data post-acquisition, without the need for lock masses. This work is based on an approach described by Wolski et al. [Wolski 2006], building on earlier work by Gay and others [Gay 1999], which relies on theoretical properties of peptide masses rather than mass injections or other preparatory steps.

We demonstrate these tools on mass spectrometry data from ESI-TOF and LTQ-FT instruments, to improve the performance of both LC-MS and LC-MS/MS-based analytical methods.

## METHODS

As others have noted, peptide masses tend to form evenly-spaced clusters across the mass range. These clusters are defined by the differing mass defect values for the elements that make up peptides, and the similar elemental composition of all peptides. We refer to the spacing between clusters as the "mass wavelength". Others have derived this theoretical value ( $w_t$ ) as 1.000455 or 1.000495 [Wolski 2006]. For the population of peptides in our data, we derive a  $w_t$  of 1.000476.

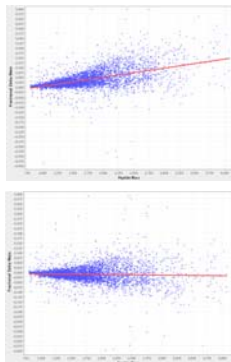


We can determine the deviation from  $w_t$  ( $d_w$ ) for any set of observed peptide features. To do this, we enumerate all pairs of features and use robust linear regression to relate the mass distance between features in a pair, to the deviation from theoretical pair separation: **massError =  $c_0$  +  $d_w$  \* massDistance**. The actual mass wavelength of the dataset, then, is  $w_d = w_t + d_w$ .

We multiply all masses by ( $w_t / w_d$ ) to correct for miscalibration scaling effects on mass values. Then, we calculate the mean difference by which each peptide mass differs from the closest lower theoretical mass cluster. We subtract this value from all masses to correct for any systematic upward or downward shift in masses.

After calibration of LC-MS data, we also remove all peptides whose masses are  $\geq 200$ ppm away from the nearest theoretical mass cluster, as these masses have negligible probability of representing actual peptides [Wolski 2006]. This step often removes a significant number of incorrectly-identified peptides.

## LC-MS/MS APPLICATIONS



The charts to the left demonstrate the effect of mass calibration errors on peptide masses (PeptideProphet score  $\geq 0.9$ ) observed via LC-MS/MS on an LTQ-FT instrument, searched with a mass error of  $\pm 3$ Da. X axis shows peptide mass; Y axis shows the Fractional Delta Mass of the precursor ion (difference between precursor ion mass and calculated neutral peptide mass, modulo 1).

There is a clear, linear relationship between peptide mass and Fractional Delta Mass. In this case, the magnitude of the error is roughly 3.3ppm. After calibration, this error largely disappears.

The Fractional Delta Mass histograms on the right demonstrate the same effect. Before calibration (upper histogram), there is a marked positive skew and a mean value of 0.038Da. After calibration, the values are more normally distributed, with a mean of 0.009Da.



Fractional Delta Mass error affects database search results significantly, especially if, as recent research endorses [Zubarev 2006], accurate masses are utilized in LC-MS/MS peptide identification.

With mass calibration, PeptideProphet sensitivity and error do not change significantly, but the estimated total number of correct peptide assignments increases from 19452.4 to 19796.3 (1.74%).

More strikingly, both the ACCMASS flag and mass calibration increase the number of high-probability ( $\geq 0.9$  peptide IDs). The ACCMASS flag in PeptideProphet, even pre-calibration, increases good IDs by 6.6%. Calibration increases good IDs by 4.7% even when ACCMASS is not used, and by 5.4% when ACCMASS is used. Our recommendation to use both the ACCMASS flag and precursor mass calibration for general use, since this increases the yield of high-probability peptide IDs by a combined 11%.

IDs with PeptideProphet probability $\geq 0.9$	Replicate 1		Replicate 2		Replicate 3	
	Before Calibration	After Calibration	Before Calibration	After Calibration	Before Calibration	After Calibration
Without ACCMASS	5140	5327	4896	5141	4970	5237
With ACCMASS	5409	5708	5249	5531	5343	5629

## LC-MS Applications

Mass calibration is also extremely important in LC-MS-based analytical methods such as Accurate Mass and Time (AMT) that rely on accurate peptide mass. We performed msInspect-based AMT analysis, matching between an AMT database created from LTQ runs and LC-MS features from ESI-TOF data. Before calibration, the data showed a 7.5ppm negative shift in peptide masses.

We determine False Assignment Rate (FAR) as the ratio of AMT matches to a decoy AMT database, to the sum of matches to both the decoy and target databases. Before calibration, FAR was an unusable 0.433. After calibration, this shrank to 0.0989, making AMT analysis feasible.

	True Matches	False Matches	Total Matches	False Assignment Rate
Before Calibration	55	42	97	0.433
After Calibration	1203	132	1335	0.0989

## CONCLUSION

These tools, within the msInspect software platform, can dynamically assess and correct mass calibration errors within mass spectrometry data, post-acquisition. Since these methods rely on intrinsic properties of peptides, they require no mass injections or other actions in sample processing. For this reason, this approach is particularly well-suited to working with data across laboratories. These methods can be used to improve the performance of both LC-MS and LC-MS/MS analysis methods.

## AVAILABILITY

msInspect software and source code are available from our website <http://proteomics.fhcr.org/> under an Apache 2.0 license. The software may be downloaded as an executable Java JAR file, or may be launched via Java Web Start directly from a web browser. The software is in regular use under Windows, GNU/Linux, and Mac OS X.

## ACKNOWLEDGEMENTS

This work was funded by National Cancer Institute contract 23XS144A.

## REFERENCES

- Zubarev R, Mann M. (2006) On the proper use of mass accuracy in proteomics. *Molecular and Cellular Proteomics*. 2006 Dec 12
- Bellew M, Coram M, Fitzgibbon M, Igra M, Randolph T, Wang P, May D, Eng J, Fang R, Lin CW, Chen J, Goodlet D, Whitesaker J, Paulovich A, McIntosh M. (2006) A suite of algorithms for the comprehensive analysis of complex protein mixtures using high-resolution LC-MS. *Bioinformatics* 2006; *Advance Access published June 9, 2006*.
- Wolski WE, Farrow M, Emde AK, Lettrach H, Latowski M, Reinert K. (2006) Analytical model of peptide mass cluster centres with applications. *Proteome Sci* 2006; 4:18
- Gay S, Binz PA, Hochstrasser DF, Appel RD. Modeling peptide mass fingerprinting data using the atomic composition of peptides. *Electrophoresis* 1999; 20(18):3527-3534.