Supplementary Information

Pages: 9 Figures: 4 Tables: 0

Supplementary Results

Analysis of the impact of calibration on SEQUEST search results

SEQUEST was modified to allow for the modification of binning parameters (width and offset). During the preprocessing step for each spectrum, SEQUEST bins peaks by *m/z* using consecutive bins of uniform width. The binning offset modifies the phasing of the binning (e.g. a bin offset of 0.1 means that values between 0.1 and 1.1005 *m/z* will be assigned to bin 0). With such a binning scheme, it is possible for mass measurement error to cause a peak to fall into a different bin than its theoretical bin and thus lower the similarity score between a spectrum and its true theoretical spectrum match. This case can be avoided by setting the bin width and offset such that peaks are more likely to fall in the center of a bin. This can be done by using a bin width of ~1.0005 *m/z* such that the edges of bins fall in the "forbidden zones" of the theoretical fragment ion map (**Figure 1, Supplementary Figure 1**) and is the reason that a default bin width of 1.0005 *m/z* was reported in the original SEQUEST paper.

In **Supplementary Figure 4A**, the bin width was fixed at 1.00045475 *m/z* and the offset was varied. For each binning offset, the LTQ-Orbitrap-Velos (Thermo Fisher Scientific, San Jose, CA, USA) data was searched using a target-decoy strategy and the number of unique peptides at a peptide level FDR threshold of

0.01 was determined. Peptide level q-values were calculated using the Xcorr scores reported by SEQUEST.

The binning offset used by SEQUEST can have a large impact on the number of peptides identified. For the control data, the worst binning offset identifies only 1,845 peptides ($q \leq 0.01$) while the best offset identifies 3,007 peptides; a 63% improvement. After calibration with FineTune, the database search is robust across a wider range of binning offsets. The optimal binning offset for the calibrated data is 0.35.

A similar analysis was performed for bin widths 0.50028 (**Supplementary Figure 4B**), 0.33348 (**Supplementary Figure 4C**), 0.250113, and 0.200091 *m/z*. Bin widths 0.500028 *m/z* and 0.33348 *m/z* return the most peptide identifications (3,755, and 3,745 for the calibrated data respectively) at their optimal offset than any other bin width. At this point in time, it is unclear why, at certain binning offsets, the uncalibrated data yields better results than the calibrated data. This could in part be due to the half-height flanking peaks added adjacent to theoretical b and y ion peaks by SEQUEST during the preprocessing step.

It is clear that at certain binning offsets, a SEQUEST database search yields far less peptide identifications (~40%) than at the optimal binning offset (0.35 for bin width ~1.0005 *m/z*). For calibrated data, optimal binning offset/width combinations can be determined *a priori* by placing bin edges in the theoretical forbidden zones[1]. If data has not been calibrated, the optimal binning offset

cannot be determined because systematic mass measurement error can change the optimal offset. Therefore, application of FineTune to a set of MS/MS spectra followed by a SEQUEST database search can avoid unnecessary losses in peptide identifications.

While it is currently not possible to define the binning width and offset in non-modified versions of SEQUEST, the Crux[2] search algorithm produces very similar results to SEQUEST and enables these parameters to be defined.

Supplementary Figure Captions

Supplementary Figure 1. The theoretical ion map. Two theoretical ion maps were generated using monoisotopic, singly charged b and y fragment ion masses from peptide-spectrum matches in the *C. elegans* BiblioSpec library. In both maps, peaks were binned into 0.01 *m/z* wide bins. In **A)** the *m/z* of each bin (x-axis) is plotted against the summed intensity of the peaks falling into each bin (y-axis). In **B)** the mass-to-charge (*m/z*) of each bin (x-axis) is plotted against the number of peaks falling into each bin (y-axis).

Supplementary Figure 2. Improved calibration on the LTQ-Velos Pro. MS/MS spectra were acquired analyzing *C. elegans* lysate on an **A)** LTQ-Velos (Thermo Fisher Scientific, San Jose, CA, USA) and **B)** LTQ-Velos Pro (Thermo Fisher Scientific, San Jose, CA, USA) followed by mass measurement error analysis by database searching. The new LTQ-Velos Pro calibration software (build 1083) corrected most of the non-linear variation in mass measurement error seen in the LTQ-Velos data. This new calibration routine is implemented on all future software releases for the LTQ-Velos, LTQ-Velos Pro and their respective Orbitrap hybrid mass spectrometers (personal communication Jae Schwartz, ThermoFisher Scientific).

Supplementary Figure 3. *De novo* calibration does not impact Mascot search results. The LTQ-Orbitrap-Velos (Thermo Fisher Scientific, San Jose, CA, USA) data was searched with Mascot using a target-decoy strategy to determine q values for each unique peptide from reported expect scores. The number of unique peptides identified by Mascot for control and calibrated data with a FDR threshold of 1% was compared using various fragment ion tolerances in Mascot.

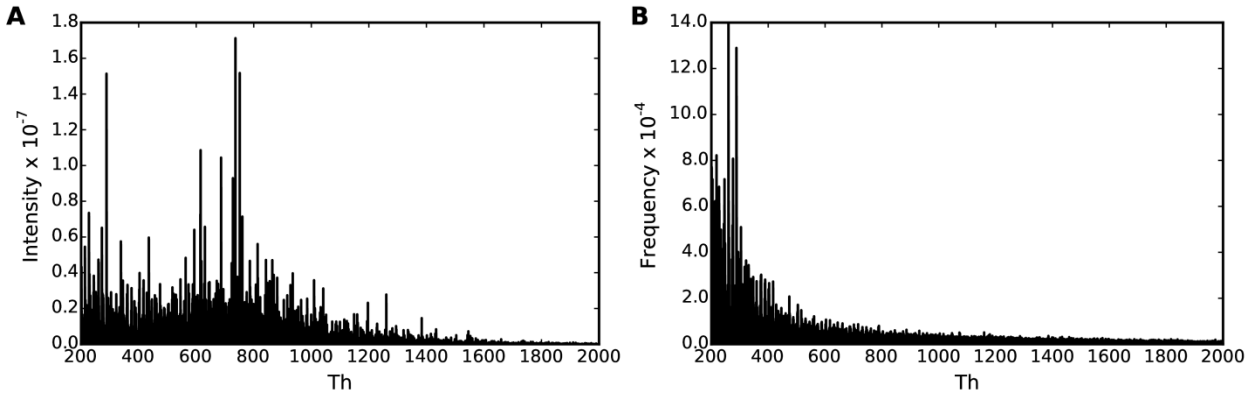
Supplementary Figure 4. Bin offset and bin width impact Sequest search results. The number of unique peptides identified by SEQUEST (q≤0.01) at different binning offsets was determined for bin widths **A)** 1.00045475 *m/z* , **B)** 0.50022738 *m/z* and **C)** 0.33348492 *m/z* using the LTQ-Orbitrap-Velos data before (Control) and after (Calibrated) *de novo* calibration.

References

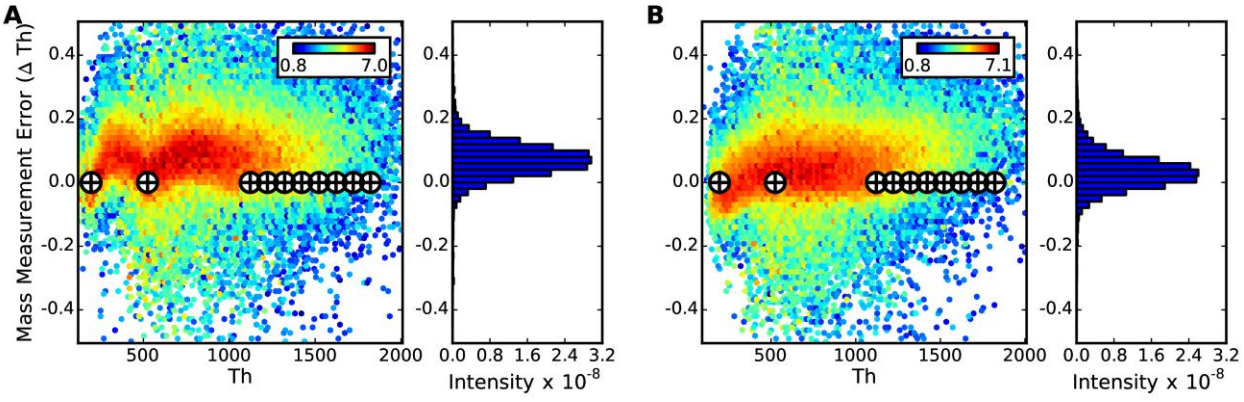
1. Frahm, J. L.; Howard, B. E.; Heber, S.; Muddiman, D. C., Accessible proteomics space and its implications for peak capacity for zero-, one- and two-dimensional separations coupled with FT-ICR and TOF mass spectrometry. *Journal of mass spectrometry : JMS* 2006, *41* (3), 281-8.

2. Park, C. Y.; Klammer, A. A.; Kall, L.; MacCoss, M. J.; Noble, W. S., Rapid and accurate peptide identification from tandem mass spectra. *Journal of proteome research* 2008, *7* (7), 3022-7.

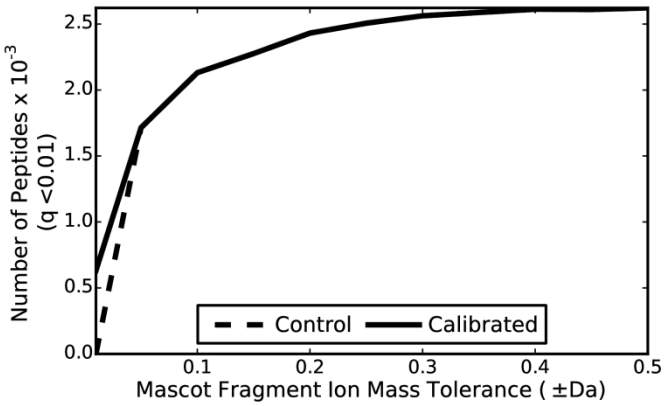
Supplementary Figure 1



Supplementary Figure 2



Supplementary Figure 3



Supplementary Figure 4