An Instrument-Independent Demultiplexing Method for Computationally Improving the Specificity of Data-Independent Acquisition

Dario Amodei[1], Jarrett Egertson[2], Brendan McLean[2], Richard Johnson[2], Olga Vitek[3], Michael MacCoss[2], Parag Mallick[1]

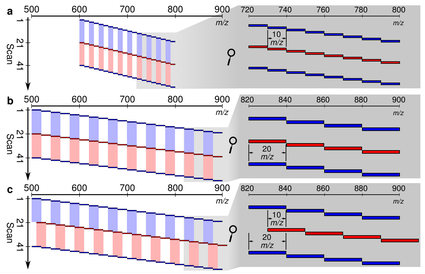
1 Canary Center for Early Cancer Detection, Stanford University
2 Department of Genome Sciences, University of Washington
3 Department of Statistics, Purdue University

Introduction and Background 1

Newly developed Data Independent Acquisition (DIA) methods have made significant progress towards obviating the tradeoff between proteome coverage and sensitivity, combining some of the advantages of targeted and data dependent methods [1]. However, one challenge involved in this method is the presence of interference from different precusors in the same isolation window (particularly modified forms of the same peptide, which cannot easily be distinguished when windows are even moderately wide). The MSX method developed by Egertson et al [2] attempted address this challenge by recording multiple small isolation windows in the same spectrum, however this method requires a Q Exactive instrument (or similar) to perform multiplexed fills and also reduces the number of collected ions (since ions from each window are injected in series). The method improves specificity but its associated tradeoffs are as yet unclear. Here we present (a) an acquisition scheme using wide but overlapping windows and (b) a computational demultiplexing algorithm that together provide increased precursor specificity and ~40% improved limit of quantitation in a yeast spike-in experiment at no cost to duty cycle, fill time, mass range, or resolution. Our method thus provides additional specificity "for free" relative to closely comparable nonoverlapping DIA acquisition schemes, making it a potentially valuable modification to a wide range of DIA acquisition schemes.

Overlapping Window Acquisition Method 2

Integrating together several ion fills from different windows is one method of collecting multiplexed spectra, but another possible method is to consecutively scan a series of isolation windows that partially overlap with one another. If we assume that the underlying signals change little across neighboring scans, then it should be possible to deduce which part of the isolation window a peak came from by observing which nearby scans it is present in (or absent from). This method has the advantage of not relying upon multifill capability and thus being applicable to nearly any spectrometer capable of acquiring DIA data. In addition, the ions from different parts of the window are collected in parallel rather than series, so that multiplexing does not compromise ion collection. Concretely, we have developed an acquisition method that combines elements of a 10 Th window DIA scheme and a 20 Th window DIA scheme. Our method is almost identical to a 20 Th cycling DIA scheme (panel b) with the single difference that every odd-numbered duty cycle is offset by 10 Th (panel c), allowing complementary information about the 10 Th regions of overlap to be collected. With proper decoding this method has the potential to provide the same specificity as if we had recorded from 10 Th windows (panel a), while retaining the wider mass range of a 20 Th scheme.



Deconvolution Strategy 3

In the overlap acquisition scheme, each scan of a 20 Th window can be approximated as a linear superposition of data from two 10 Th subwindows. Two full duty cycles represent 40 scans over independent combinations of the 41 windows. If we add one regularization row, this forms a linearly determined system of the form:

$$A\vec{x} = \vec{b}$$

where $\vec{b}$ represents the observed scans of the 20 Th windows, $\vec{x}$ represents the unknown 10 Th sub-windows, and A is a "design matrix" relating the two together. We can solve this equation subject to the constraint that ion counts are positive, namely:

$$\min_x |A\vec{x} - \vec{b}|^2 \quad \text{subject to} \quad x_i \geq 0$$

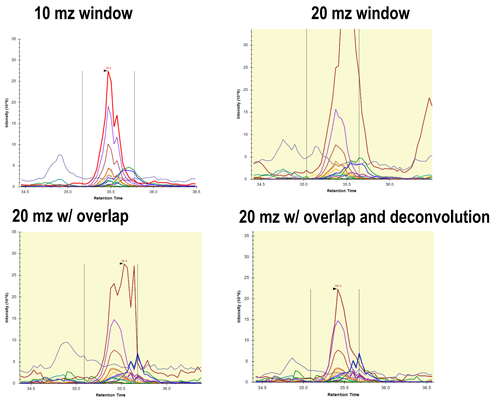
This constrained least-squares problem can be solved using an active-set method [2]. To improve running time, we focus only on the 7 scans closest to a scan of interest, approximating distant scan as non-contributing. Our algorithm has been integrated into skyline and runs in 7 minutes per 1000 transitions.

Spike-In Experimental Design 4

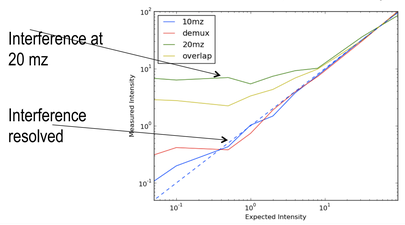
- 6 protein bovine mix in S. Cerevisiae background
- 9 point dilution curve from 96 fm to 0.05 fm
- 3 replicates of each condition
- Conditions randomized within replicates and concentrations:
  1. 10 mz windows from 500-700 Th, 2 sec duty cycle
  2. 20 mz windows from 500-900 Th, 2 sec duty cycle
  3. 20 mz overlapping windows, 500-900 Th, 2 sec duty cycle
- Thermo Q-Exactive
  - MS2 55 ms fill, 64 ms transient @ 16k res
  - MS1 55 ms fill, 128 ms transient @ 32k res
- 86 precursors / 82 tryptic peptides in 500 – 700 Th range
- Focus on the 21 consistently observable @ 96 fm

Example Results for Transition 5

Transition (VLVLDTDYK++, b2+)@ 2 fm



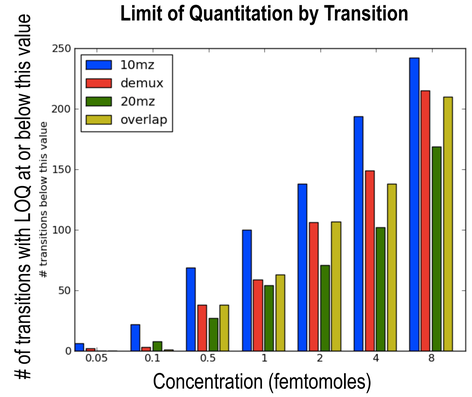
Dilution curves for all conditions, transition (VLVLDTDYK++, b2+)



Dataset Wide Results 6

Demultiplexing increases the total number of transitions that are observable down to a given LOQ. We define the LOQ of a transition as the lowest concentration down to which we see unbroken <40% deviation from the expected concentration. Expected concentration is defined by fitting a calibration curve in log space.

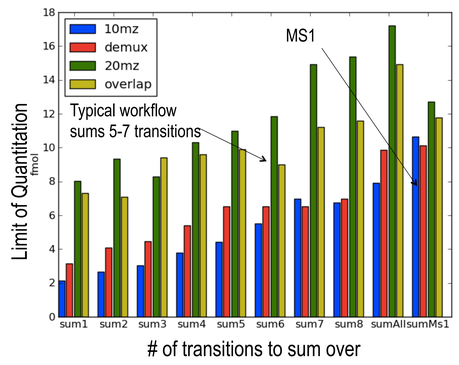
Limit of Quantitation by Transition



Demultiplexing improves peptide LOQ (as measured by sums over the best transitions on different metrics), allowing the overlap acquisition scheme to perform better than 20 mz and nearly as good as 10 mz.

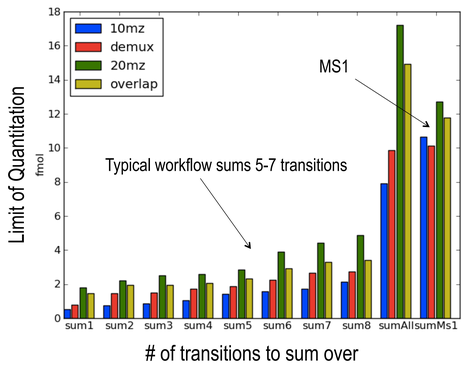
Limit of Quantitation by Peptide

Typical Transition Selection (by highest intensity)



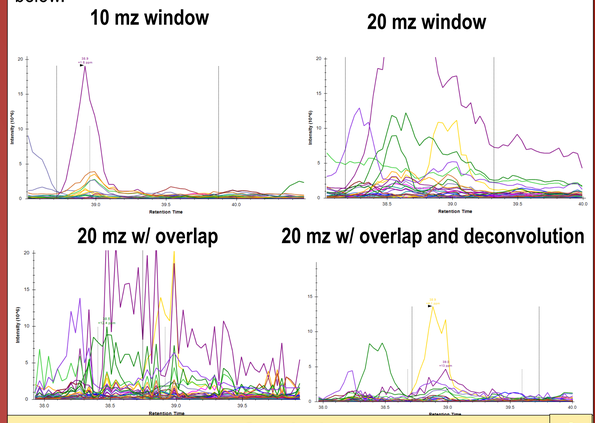
Limit of Quantitation by Peptide

Optimal Transition Selection (best-case analysis)



Results in Yeast Background 7

To show that overlap-based improvements generalize beyond artificial spike-in experiments, we also measured the yeast background in control replicates for each of the conditions. We found that of 61 peptides randomly selected from the 2,000 most abundant yeast peptides in the mass range, 16 (26%) had a major interference that was qualitatively removed by deconvolution, as shown for one example below.



Conclusions 8

- About 25% of the observable peptides in our experiment showed at least one major interference at 20 Th not present at 10 Th
- Deconvolution was effective in removing these interferences
- Removal of these interferences improved LOQ on both the transition and peptide levels, allowing us to achieve an LOQ (and specificity) nearly comparable to that of 10 Th windows, without compromising the duty cycle, mass range, resolution, and fill time of 20 Th windows.

Future Directions 9

- Look at improvements in identification in addition to quantification
- Applying the technique to a pathway or whole-proteome study.
- What is the right isolation window size, duty cycle, and scan parameters for optimizing the impact of overlap acquisition?
- Using deconvolution to detect post-translational modifications
- Deconvolution can be applied in principle on most major mass spectrometers -> demonstrate cross-instrument utility?
- What are the ultimate limits to deconvolution?

References

[1] Gillet LC, Navarro P, Tate S, Roest H, Selevsek N, Reiter L, Bonner R, Aebersold R, Mol Cell. Prot., 2012
[2] Egertson et al, ASMS Abstract, 2012
[3] Lawson and Hanson, 1974, Solving Least Squares Problem.s

Contact information

Dario Amodei
1501 South California Ave.
Palo Alto, CA, 94304
T: 415.505.5563
damodei@gmail.com

Jarrett Egertson
3720 15th Ave. NE
Seattle, WA, 98195
T: 858.337.6847
jegertso@uw.edu

http://www.canarycenter.stanford.edu
http://www.gs.washington.edu