Multiplexed MS/MS for improved data-independent acquisition

Jarrett D Egertson1, Andreas Kuehn2, Gennifer E Merrihew1, Nicholas W Bateman3, Brendan X MacLean1, Ying S Ting1, Jesse D Canterbury4, Donald M Marsh1, Markus Kellmann2, Vlad Zabrouskov4, Christine C Wu3 & Michael J MacCoss1

In mass spectrometry-based proteomics, data-independent acquisition (DIA) strategies can acquire a single data set useful for both identification and quantification of detectable peptides in a complex mixture. However, DIA data are noisy owing to a typical five- to tenfold reduction in precursor selectivity compared to data obtained with data-dependent acquisition or selected reaction monitoring. We demonstrate a multiplexing strategy, MSX, for DIA analysis that increases precursor selectivity fivefold.

In discovery proteomics experiments, tandem mass spectrometry (MS/MS) data are usually collected on peptides through automated data-dependent acquisition (DDA). With DDA, mass information on intact peptides in a full-scan mass spectrum (MS1) is used to decide which subset of peptides will be targeted for acquisition of fragmentation (MS/MS) spectra necessary for sequence identification1. Although DDA is a powerful and versatile strategy, it suffers from several fundamental limitations. For one, the number of peptides sampled is limited by the MS/MS sampling speed regardless of the dynamic range and peak capacity of the mass analyzer. A single MS1 spectrum can contain over 100 molecular species, of which only a handful are analyzed by MS/MS before the next full scan. This sampling of peptides for MS/MS analysis follows a random sampling model2 biased toward high-abundance peptides. In a complex protein digest, as many as 84% of peptides may remain unsampled3, and as many as 30% of the sampled peptides can vary between replicate analyses of the same sample2. DDA may also compromise the sensitivity of MS/MS because the MS1 spectrum typically contains greater chemical background interference than does an MS/MS spectrum. A peptide above the MS/MS detection limit can go unselected for fragmentation because the precursor is masked by background interference. Additionally, the MS/MS spectra acquired by DDA are rarely sampled at the optimal portion of the peptide elution profile4, and as many as 15–20% of the sampled MS/MS spectra from a complex mixture are chimeric (that is, they contain two or more coeluting molecular species within the isolation window)5.

An alternative to DDA is DIA. In DIA, MS/MS scans are collected systematically and independently of precursor information. This approach has seen many variations such as collecting fragmentation data without precursor-ion selection6, using ion mobility–collision-induced dissociation–time-of-flight mass spectrometry7, using wide isolation windows8 and using narrow isolation windows combined with many injections9. However, until recently, all of these methods have been using a database searching strategy to assign peptide sequences to the fragmentation spectra. Because of the increased complexity of these MS/MS spectra, peptide identification using traditional database search strategies is often less effective for these data than with DDA data.

Recently, alternative strategies have been reported10,11 that involve querying target peptides against DIA data, as opposed to trying to assign peptide sequences to every chimeric MS/MS spectrum acquired. Fourier transform–all reaction monitoring (FT-ARM)10 and SWATH11 are similar to a targeted analysis using selected reaction monitoring (SRM) and are fundamentally incomparable with a discovery experiment that uses a database search engine to qualitatively profile the peptide content in a mixture. The appeal of these approaches is that any peptide precursor- and product-ion data within the limit of detection of the instrument can be extracted from the data. The relative fragment-ion intensities, peptide-precursor isotope peaks and retention time of the extracted ion chromatograms are used to confirm the identity of the target molecular species just as with a targeted SRM experiment. Unlike in SRM, different hypotheses can be tested on the data without the need to perform another mass spectrometry experiment. However, with current mass spectrometers approaching acquisition speeds of 10 Hz, a 20-m/z-wide precursor isolation window is required to sample a 400-m/z range every 2 s. This large isolation width is undesirable because of substantially increased fragment-ion interference (Supplementary Figs. 1–3). For example, peptides and their modified forms (such as oxidized methionine) are difficult to distinguish if they are isolated in the same window owing to their similar fragmentation patterns (Supplementary Fig. 1).

We present a multiplexing strategy, MSX, in which five separate 4-m/z isolation windows are analyzed per spectrum. These spectra are demultiplexed into the five separate 4-m/z isolation windows using a strategy with similarities to Hadamard multiplexing12, resulting in data with the sampling frequency of a DIA approach.

1Department of Genome Sciences, University of Washington School of Medicine, Seattle, Washington, USA. 2Thermo Fisher Scientific (Bremen) GmbH, Bremen, Germany. 3Department of Cell Biology and Physiology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA. 4Thermo Fisher Scientific, San Jose, California, USA. Correspondence should be addressed to M.J.M. (maccoss@uw.edu).

Received 29 August 2012; accepted 3 May 2013; published online 23 June 2013; doi:10.1038/nmeth.2528
Figure 1 | Multiplexed data-independent acquisition with MSX. A common implementation of DIA is to use a repeated cycle of wide-isolation window MS/MS scans to cover a mass range. In this example, the 500- to 900-m/z range is covered with 40 scans each sampling a single 10-m/z-wide window. In MSX, each scan isolates five 4-m/z-wide windows before fragment-ion mass analysis. The five windows isolated in each scan are chosen randomly from the set of 100 possible nonoverlapping windows covering the 500- to 900-m/z range. Each mixed MSX spectrum is demultiplexed into the five component spectra corresponding to the isolated windows.

using 20 20-m/z-wide windows but the selectivity of an approach using 100 4-m/z-wide windows. Demultiplexing improves precursor selectivity by narrowing the range of potential precursors for an MS/MS spectrum from a 20-m/z window down to a 4-m/z window (Supplementary Figs. 4 and 5) and generating the unmixed fragment-ion spectrum with signal from only the 4-m/z window. We have implemented this demultiplexing approach in the open-source Skyline software tool13 (http://skyline.maccosslab.org/), which provides a useful interface for the visualization and analysis of these data. For a detailed discussion on precursor selectivity, the design of MSX methods and comparisons to other data acquisition techniques, refer to the Supplementary Note.

We collected MSX liquid chromatography–MS/MS data on a Saccharomyces cerevisiae lysate using five 4-m/z-wide isolation windows per scan on a Q-Exactive mass spectrometer (Thermo Scientific). Owing to the multiple fills per mass analysis, this multiplexing technique is best suited for instrument in which isolation and collisional activation of peptides is fast relative to mass analysis. We randomly selected 5 of the 100 possible 4-m/z isolation windows in the range of 500–900 m/z to be analyzed in each multiplexed scan (Fig. 1 and Online Methods). To analyze these spectra, we modified Skyline to detect MSX spectra and demultiplex them automatically upon import (available in v.1.3). We generated a Skyline document containing peptides with spectra in the US National Institute of Standards and Technology S. cerevisiae Q-TOF and Ion Trap spectral libraries (5/24/2011 builds)14. We analyzed the spectra with and without demultiplexing.

Extracted data for the example peptide GPLVLEYETYR without demultiplexing contained many intense fragment-ion peaks present throughout the gradient, representing fragments from other peptides (Fig. 2a). Demultiplexing removed the majority of these interfering peaks because they did not originate from precursors in the same isolation window as GPLVLEYETYR (Fig. 2b). Interference was also present in many of the fragment-ion chromatograms that overlapped in elution time with GPLVLEYETYR (Fig. 2c). This interference had a characteristic spike shape because it originated from precursors in isolation windows different from that of the target peptide. Owing to the random sampling of isolation windows for each scan (Online Methods, Fig. 1 and Supplementary Video 1), the window containing the interfering precursor was not isolated in consecutive scans containing the target isolation window. Demultiplexing removed the interfering signal while retaining the signal originating from the target isolation window (Fig. 2d), resulting in a higher dot-product similarity (0.96 versus 0.94 for demultiplexed and non-demultiplexed fragments, respectively) to a DDA spectrum for GPLVLEYETYR acquired with a 2-m/z-wide isolation window (Supplementary Fig. 6).

To test the performance of this MSX method for peptide quantification, we spiked a commercial six–bovine-protein digest into a complex matrix (S. cerevisiae lysate, soluble fraction) in amounts ranging from 50 attomoles to 100 femtomoles (fmol) on-column. We collected MS1 and MSX data simultaneously on each spike-in sample by acquiring MSX scans (resolving power: 17,500) with MS1 scans (resolving power: 35,000) interlaced every ten scans. We quantified five of the proteins using a total of 36 peptides (Fig. 3, Supplementary Data and Supplementary Table 1).

The lower limit of detection for the 36 peptides we quantified averaged 8.66 and 4.98 fmol for MSX and MS1, respectively. All peptides showed a linear response above the limit of detection, with R² values of the regression lines averaging 0.95 and
0.98 for MSX and MS1, respectively. Although the MSX method was less sensitive than the MS1 method on average ($P = 0.016$, paired $t$-test, $n = 37$), the results are notable given that the MSX data provide structural selectivity information in addition to quantification. The slight reduction in sensitivity of MSX relative to MS1 is not unexpected in the absence of chemical noise. MS1 is expected to be more sensitive in a simple mixture because in MS/MS the ion beam is split into multiple fragmented products, each of which has a fraction of the original intensity. Of the 36 peptides, 7 suffered from interference in the MS1 signal, resulting in an average 3.4-fold improvement in sensitivity ranging from 1.3- to 8.3-fold by MSX quantification. Acquiring MS1 and MSX data simultaneously combines the high sensitivity of MS1 with the structural selectivity of MSX while theoretically improving quantitative precision by combining measurements from the precursor (MS1) and fragment-ion peaks (MSX) for quantitation. The improved selectivity of MSX resulted in greater improvements in the interpretation of DIA data, is a practical technique for global, highly selective, and reproducible relative quantitation of peptides in a proteomics experiment.

A commercial six-protein digest was spiked into a soluble S. cerevisiae lysate in amounts ranging from 50 attomoles to 100 fmol on-column. MSX data were acquired with an MS1 scan interleaved every ten scans. (a) Signal intensity for each spike-in point (normalized to two background peptides) for the peptide LVNELTEFAK with (right) and without (left) log scaling of the $x$ and $y$ axes. The lower limits of detection were 0.41 fmol and 1.02 fmol for MSX- and MS1-based quantitation, respectively. The slopes of the regression lines were $0.030 \pm 0.004$ (95% confidence interval) and $0.071 \pm 0.002$ for MSX and MS1, respectively. (b,c) MS1 (b) and MSX (c) signal at 1.02 fmol. M, M + 1 and M + 2 represent precursor isotopes.

**METHODS**

Methods and any associated references are available in the online version of the paper.

**ACKNOWLEDGMENTS**

The authors acknowledge financial support from US National Institutes of Health grants R01 GM103551, P41 GM103533 and F31 AG037265.

**AUTHOR CONTRIBUTIONS**

A.K., C.C.W., J.D.C., J.D.E., M.J.M., M.K., V.Z. and Y.S.T. designed experiments. A.K., C.C.W., J.D.E., M.J.M., N.W.B. and V.Z. interpreted results. J.D.E. and M.J.M. wrote the manuscript. A.K. created a firmware modification for the Q-Exactive. A.K. and M.K. provided preliminary data. G.E.M., J.D.E. and N.W.B. performed experiments. B.X.M., D.M.M. and J.D.E. wrote software.

**COMPETING FINANCIAL INTERESTS**

The authors declare competing financial interests: details are available in the online version of the paper.

Reprints and permissions information is available online at [http://www.nature.com/reprints/index.html](http://www.nature.com/reprints/index.html).

1. Stahl, D.C., Swiderek, K.M., Davis, M.T. & Lee, T.D. *J. Am. Soc. Mass Spectrom.* 7, 532–540 (1996).
2. Liu, H., Sadygov, R.G. & Yates, J.R. *Anal. Chem.* 76, 4193–4201 (2004).
3. Michalski, A., Cox, J. & Mann, M. *J. Proteome Res.* 10, 1785–1793 (2011).
4. Wenner, B.R. & Lynn, B.C. *J. Am. Soc. Mass Spectrom.* 15, 150–157 (2004).
5. Hoopmann, M.R., Finney, G.L. & MacCoss, M.J. *Anal. Chem.* 79, 5620–5632 (2007).
6. Purvine, S., Eppel, J.T., Yi, E.C. & Goodlett, D.R. *Proteomics* 3, 847–850 (2003).
7. Myung, S. et al. *Anal. Chem.* 75, 5137–5145 (2003).
8. Venable, J.D., Dong, M.D., Wohlschlegel, J., Dillin, A. & Yates, J.R. *Nat. Methods* 1, 39–45 (2004).
9. Panchaud, A. et al. *Anal. Chem.* 81, 6481–6488 (2009).
10. Weisbrod, C.R., Eng, J.K., Hoopmann, M.R., Baker, T. & Bruce, J.E. *J. Proteome Res.* 11, 1621–1632 (2012).
11. Gillet, L.C. et al. *Mol. Cell Proteomics* 11, 0111.016717 (2012).
12. Williams, E.R., Loh, S.Y., McLafferty, F.W. & Cody, R.B. *Anal. Chem.* 62, 698–703 (1990).
13. MacLean, B. et al. *Bioinformatics* 26, 966–968 (2010).
14. US National Institute of Standards and Technology. *NIST* Peptide Tandem MS Mass Spectral Libraries (eds. Stein, S.E. & Rudnick, P.A.) (http://peptide.nist.gov/) (2009; 24 May 2011 builds).
15. Yost, R.A. & Enke, C.G. *Anal. Chem.* 51, 1251–1264 (1979).
ONLINE METHODS

Implementation of multiplexed MS/MS on a Q-Exactive mass spectrometer. In a multiplexed MS/MS (MSX) scan, fragments from multiple precursor isolation windows are stored together before mass analysis in the Orbitrap. All of the MSX data in this manuscript were collected using five 4-m/z-wide isolation windows per MS/MS scan, although different combinations of isolation windows per scan and isolation width can be used. The five windows isolated for each MS/MS scan are chosen randomly from a list of all 4 m/z isolation windows to be analyzed in the experiment. In these experiments, the list is all 100 nonoverlapping 4-m/z-wide windows between 500 and 900 m/z (Supplementary Table 2). The window width and position are shifted slightly from their integer values to reduce the likelihood of placing window edges where peptides are most likely to occur (Supplementary Fig. 8). For example, the first window is centered at 502.4783 m/z rather than 502 m/z and has a width of 4.002 m/z. This optimization is due to the fact that peptide masses are distributed across the mass range every ~1.0005 m/z in ‘allowable regions’ interspersed with ‘forbidden zones’ where peptide masses do not occur. Because the allowable regions are spaced apart by 1.0005 m/z on average, the isolation window width is set to be a multiple of this number. Additionally, windows are shifted so that the edges fall in forbidden zones. All window edges are calculated using the equation 0.25 + 1.00045475 × N, where N is an integer. The offset value of 0.25 was determined to be optimal according to analysis of the distribution of peptide precursor masses (+1, +2, +3 charged) in the Bibliospec spectral library. For window edges to be optimally placed, they must be a member of the series 0.25 + 1.00045475 × N. For these data, the edges of the windows are N × 4 × 1.00045475 + 0.25, where N is every integer in the range (inclusive) of 125–225.

Each scan is built by removing five random isolation windows from the isolation list. This process is repeated for each scan until the isolation list is empty (20 scans in this case), at which point the list is repopulated. By sampling from the list without replacement, each isolation window is sampled every 20 scans (~3.5 s) on average and is guaranteed to be sampled at least every 40 scans. Because the data presented in this manuscript have been collected, an additional optimization to avoid the selection of the same pairs of isolation windows in nearby scans has been implemented. If selecting five isolation windows randomly from the isolation list results in pairing two windows together that have been paired in the same scan recently, the random selection is repeated until this is no longer the case or too much time has passed. With the 5 × 4-m/z multiplexing scheme, it is typical to not see the same two isolation window pairs observed within 200 scans (~35 s) of each other.

Scans are defined before an experiment using this method in Skyline (v.1.3) and exported as an inclusion list containing a sequence of 5,000 isolation windows (1,000 scans). During analysis, the Q-Exactive (Thermo Fisher Scientific) loops through this list sequentially, choosing five isolation windows for each MSX scan. This capability is implemented in a firmware modification that keeps the mass list from being sorted by the instrument computer before data acquisition.

When an MSX scan is acquired, each isolation window is isolated, fragmented and stored serially in time before mass analysis in the Orbitrap. The fill times for all isolation windows in a given scan are the same. If the fill times for each window were different, it would be impossible to calculate the ion current (charges per second) for a fragment-ion population because the fill time for the precursor(s) of that fragment-ion population would be unknown.

For implementation of automated gain control (AGC), an MS1 scan is acquired every tenth scan. For each MSX scan, the most recent MS1 scan is analyzed to determine the summed ion current for all five precursor isolation windows in the scan. The summed ion current is determined by summing up the total MS1 extracted ion current for each of the five isolation windows. The target number of ions is divided by this summed ion current to determine the fill time (the same for each window). With this approach, the total number of ions in the trap is fixed, but the number of ions from any of the five windows will vary. The maximum fill time for an isolation window is set to 20 ms, meaning that the maximum total fill time for a scan is 100 ms (5 windows × 20 ms). MS1 and MSX scans are acquired with resolving power 35,000 and 17,500 respectively. The AGC target for MS1 scans is set to 10⁶ ions and 10⁵ ions per isolation window for MS1 and MSX scans, respectively.

Demultiplexing by non-negative least-squares optimization. Owing to the random selection of isolation windows for each scan, any two spectra with an isolation window in common will not share any other isolation windows. This characteristic helps when demultiplexing these spectra because overlapping fragment-ion information is more likely to come from the single isolation window the spectra have in common than from any of the others. In other words, spectra can be demultiplexed by leveraging information (the isolation windows used and fragment-ion intensities) from neighboring spectra. To do this, a system of equations is built describing each observed multiplexed spectrum as a linear combination of multiple unobserved single-precursor component spectra (Fig. 1).

The system of equations is represented by matrix multiplication: B = A × X. Matrix B contains the fragment-ion intensities for the multiplexed spectra: each column is a fragment-ion m/z, and each row is a spectrum. Matrix A contains the isolation-window pattern for each spectrum: each column is an isolation window, and each row is a spectrum. Isolation windows have the value 1 if they are present in a spectrum and 0 if they are absent. The matrix X is an unknown matrix containing the demultiplexed component spectra. Each row is a demultiplexed spectrum for an isolation window, and each column is a demultiplexed fragment-ion intensity. The system of equations is solved by non-negative least-squares. For each spectrum to be demultiplexed, this system of equations is built containing 140 consecutive MSX spectra with the spectrum to be demultiplexed at the center of this window. For the five-isolation-window experiment, 140 spectra are used, but for other methods, the number of spectra in the window is (no. of isolation windows total / isolation windows per scan) × (2 + isolation windows per scan). Setting the number of spectra this way guarantees that the system of equations is not underdetermined. Using neighboring spectra to demultiplex a spectrum has the negative consequence of reducing signal intensity due to the averaging of the demultiplexed spectra with nearby spectra similar to a boxcar smooth. So that this reduction is minimized, each row of matrices B and A is multiplied by the Savitzky-Golay smoothing coefficient (5 wide, second order) for that row.
Finally, each demultiplexed peak is normalized such that the sum of the intensities from each demultiplexed component is equivalent to the intensity of the original observed peak. This demultiplexing happens automatically upon import into Skyline.

Evaluation of demultiplexing. A soluble S. cerevisiae lysate was reduced, alkylated, digested for 1 h with trypsin, and cleaned by dual-mode solid-phase extraction (Oasis MCX cartridges, Waters Corporation). LC-MS data (described above) and DDA data were acquired on the Q-Exactive on 1.2 µg of sample per injection. A 40-cm 75 µM fused-silica column packed with reversed-phase C12 Jupiter resin (Phenomenex) was used to separate the sample across a 90-min linear acetonitrile gradient from 0 to 25% buffer B. Chromatography was performed using an EASY-nLC II (Thermo Fisher Scientific) system set to a flow rate of 250 nl/min. Buffer A was 2% ACN, 0.1% formic acid and 97.9% water. Buffer B was 99.9% ACN and 0.1% formic acid. The data were analyzed using a modified version of Skyline that processed the data with and without demultiplexing enabled. The DDA data were acquired using a top-five method with a 75-s dynamic exclusion list and 3.3 × 10^5 trigger-intensity threshold. MS1 spectra had a resolving power of 70,000 at 200 m/z with an AGC target of 10^6 ions and a maximum injection time of 250 ms. MS/MS spectra were acquired with HCD fragmentation, a normalized collision energy of 30.0, a 2-m/z-wide isolation window, a resolving power of 17,500 at 200 m/z, an AGC target of 2 × 10^5 ions and a maximum injection time of 120 ms.

Spike-in experiment to evaluate quantitative performance. An equimolar six-protein digest (Bruker-Michrom) was spiked into a complex matrix (soluble S. cerevisiae lysate digest) over four orders of magnitude to test the quantitative performance of MSX. The proteins used were β-lactoglobulin (NCBI GI: 2194089), lactoperoxidase (GI: 129823), carbonic anhydrase (Experimental Protein Index (IPI): IPI00716246), glutamate dehydrogenase (GI: 118533), α-casein (IPI00706094 (S1) and IPI00698843 (S2)) and serum albumin (GI: 1351907). The yeast sample was reduced, alkylated, digested for 1 h with trypsin, and cleaned by dual-mode solid-phase extraction (Oasis MCX cartridges, Waters) after digestion. Prior to the spike-in experiments, the complex matrix was run four times to condition the liquid chromatography column. Spike-in experiments with 0.6 µg of the complex matrix loaded on column with 0.05, 0.12, 0.31, 0.77, 1.92, 4.8, 12, 30 and 75 fmol of the bovine protein digest spiked in were run followed by blanks of the complex matrix and buffer A and then by spike-ins of 0.07, 0.16, 0.41, 1.02, 2.6, 6.4, 16, 40 and 100 fmol. Samples were run using the same chromatography conditions and column type as in the previous section “Evaluation of demultiplexing”, except using C18 Aqua resin (Phenomenex) instead of the C12 Jupiter reversed-phase resin.

Thirty-six peptides from five of the spiked-in proteins and not present in the background matrix were quantified (Fig. 3, Supplementary Data and Supplementary Table 1). Peptides were quantified using the area under the curve of the M, M + 1 and M + 2 peaks for MS1, and a manually curated subset of the b- and y-ion series for MSX (Supplementary Table 1). The signal from each peptide was normalized by the signal from two highly abundant peptides (DNSQVFGVAR and ESTLHLVLR) from the background matrix. The lower limit of detection was determined for each peptide by manual inspection. For MSX data, the lower limit of detection was the lowest abundance where at least three transitions coelute and have similar intensity ratios to that seen at higher abundance. For MS1 data, the lower limit of detection was the lowest abundance where at least the M and M + 1 isotope peaks coelute and have the same rank order as at higher abundances. The sensitivity of the two techniques was compared using a paired t-test on the log-transformed lower limits of detection for the 36 peptides (37 precursors) quantified. Regression lines were fit to the MS1 and MSX data for each peptide using all points at or above the lower limit of detection.

Software. The ability to design MSX methods and analyze the resulting data (including demultiplexing) has been implemented as part of the free, open-source Skyline software tool, which can be downloaded at http://skyline.maccosslab.org/. Skyline itself is part of the larger ProteoWizard project. The ProteoWizard source code (including Skyline) can be perused in a web browser at http://sourceforge.net/p/proteowizard/ or viewed from the SVN repository at https://svn.code.sf.net/p/proteowizard/code/trunk/pwiz/. The source code is available under an Apache 2.0 open-source license.

Supplementary Data

Supplementary Table 1

16. Frahm, J.L., Howard, B.E., Heber, S. & Muddiman, D.C. J. Mass Spectrom. 41, 281–288 (2006).
17. Gay, S., Binz, P.A., Hochstrasser, D.F. & Appel, R.D. Electrophoresis 20, 3527–3534 (1999).
18. Frewen, B.E., Merrihew, G.E., Wu, C.C., Noble, W.S. & MacCoss, M.J. Anal. Chem. 78, 5678–5684 (2006).
19. Lawson, C.L. & Hanson, R.J. Solving Least Squares Problems (Society for Industrial and Applied Mathematics, Philadelphia, 1995).
20. Chambers, M.C. et al. Nat. Biotechnol. 30, 918–920 (2012).