Whole cell, label free protein quantitation with data independent acquisition: Quantitation at the MS2 level

Peter McQueen¹, Vic Spicer², John Schellenberg³, Oleg Krokhin¹,⁴, Richard Sparling³, David Levin⁵ and John A. Wilkins¹,⁴,⁶

¹ Manitoba Centre for Proteomics and Systems Biology, Winnipeg, Manitoba, Canada
² Department of Physics and Astronomy, University of Manitoba, Winnipeg, Manitoba, Canada
³ Department of Microbiology, University of Manitoba, Winnipeg, Manitoba, Canada
⁴ Department of Internal Medicine, University of Manitoba, Winnipeg, Manitoba, Canada
⁵ Department of Biosystems Engineering, University of Manitoba, Winnipeg, Manitoba, Canada
⁶ Health Sciences Centre, Winnipeg, Manitoba, Canada

Label free quantitation by measurement of peptide fragment signal intensity (MS2 quantitation) is a technique that has seen limited use due to the stochastic nature of data dependent acquisition (DDA). However, data independent acquisition has the potential to make large scale MS2 quantitation a more viable technique. In this study we used an implementation of data independent acquisition—SWATH—to perform label free protein quantitation in a model bacterium Clostridium stercorarium. Four tryptic digests analyzed by SWATH were probed by an ion library containing information on peptide mass and retention time obtained from DDA experiments. Application of this ion library to SWATH data quantified 1030 proteins with at least two peptides quantified (~40% of predicted proteins in the C. stercorarium genome) in each replicate. Quantitative results obtained were very consistent between biological replicates ($R^2 \approx 0.960$). Protein quantitation by summation of peptide fragment signal intensities was also highly consistent between biological replicates ($R^2 \approx 0.930$), indicating that this approach may have increased viability compared to recent applications in label free protein quantitation. SWATH based quantitation was able to consistently detect differences in relative protein quantity and it provided coverage for a number of proteins that were missed in some samples by DDA analysis.

Keywords:
Bottom-up proteomics / Data dependent acquisition / Data independent acquisition / Label free quantitation / Relative protein quantitation / SWATH / Technology

Additional supporting information may be found in the online version of this article at the publisher’s web-site

1 Introduction

MS-based peptide sequencing has become the key method for protein identification and quantification. Peptide mixtures obtained from proteolytic digestions are analyzed in succession by MS and MS/MS in order to obtain parent mass and amino acid sequence information, respectively [1]. Experimental data are aligned against theoretical peptide fragmentation databases to identify proteins [2, 3]. Protein identification is the first necessary step in all quantitation procedures. Information from peptide peak intensities in MS or MS/MS modes can also be used to quantify proteins and give relative protein expression measurements.

Quantitation techniques in proteomics are usually classified by the type of acquisition used to extract peak intensities (MS or MS/MS) or by the use of stable isotopes [4].
Relative quantitation methods such as ICAT [5], SILAC [6], or iTRAQ [7], use isotope-based labeling in order to quantify proteins. The ICAT and SILAC methods employ MS-based quantitation, while iTRAQ uses specific reporter ions in MS/MS spectra to assign relative peptide abundances. SRM uses isotopically labeled reference peptides to perform absolute quantitation of proteins through the measurement of specific fragments isolated by tandem MS [8]. SRM does not perform protein–peptide identification, but uses peptide sequence information obtained from prior experiments as building blocks for a targeted acquisition method.

Label free quantitation of cellular proteins is quickly becoming the predominant method for analysis of complex proteomes [9] because it requires fewer steps in sample processing, costs less than other quantitative methods, and is broadly applicable [10]. This type of quantitation is typically performed either by measurement of extracted ion chromatograms of peptide signal intensities (MS1 quantitation) [11] or by the measurement of corresponding peptide fragment intensities (MS2 quantitation). An example of MS1 quantitation is intensity based absolute quantitation which sums the intensity of all peptides belonging to a specific protein then divides by the number of theoretically observable peptides to provide an estimate of protein abundance [12]. Normalized spectral index is a recent technique for label free quantitation, which combines aspects of peptide and spectral counting with fragment ion measurement [13]. Spectral counting is a common technique for MS2 label free quantitation, where the number of MS2 spectra identified for a specific protein is taken as an indicator of relative abundance [14].

The majority of quantitation techniques reported to date are based on data dependent acquisition (DDA) methods [15] where, after an initial scan, the most abundant peptides are selected for fragmentation. Several groups have experimented with summed MS2 ion intensities to provide information on protein quantity [16–18]. However, the stochastic nature of selection can result in different proteins being identified, even when an identical sample is analyzed multiple times in succession [19]. This approach is also biased toward fragmenting the most abundant peptides and cannot provide consistent sampling throughout the full profile of an eluting peptide peak—both of which are key conditions for accurate quantitation. Recent improvements in acquisition speed of mass spectrometers [20] may resolve this: a 2014 report by Krey et al. [16] demonstrated that MS2 quantitation based on summed fragment intensity using an Orbitrap Velos instrument “nearly matched” MS1 intensity based absolute quantitation quantitation data. Recent time-of-flight mass spectrometers offer up to 50 Hz acquisition rate in DDA MS2 mode [21], which might also demonstrate superior MS2 quantitation.

These advances have allowed for the development of data independent acquisition (DIA) methods that eliminate the variability associated with ion selection during DDA analysis of peptide mixtures. In DIA, multiple peptides are isolated and then fragmented simultaneously, as opposed to fragmenting single peptides in DDA. The resulting mass spectrum in each case consists of overlapping fragmentation spectra of many different peptides. Examples of DIA methods reported to date include PACIFIC [22], MS$^3$ [23], and SWATH [24]. These methods differ mainly in the number of peptides selected for simultaneous fragmentation.

SWATH’s default configuration uses 25 m/z blocks of ions that are isolated in the mass spectrometer and fragmented simultaneously. This process is repeated across the entire m/z range (typically 400–1200 m/z) in order to obtain fragmentation information on as many peptides as possible. The sampling speed of recent mass spectrometers is sufficient for multiple acquisitions across the chromatographic profile of an individual peak, thus providing more consistent quantitative information than DDA. The result is multiple SWATH “blocks” containing information on ions isolated across the entire LC–MS run. In theory, the results of these experiments should contain sequencing and quantitative information on all peptides in a given sample. Peptide signals (intensity of selected fragment ions for a particular 25 m/z block) can be extracted using ion libraries obtained from previous DDA experiments. SWATH provides a permanent record of all peptide fragmentation information; this record can be reanalyzed using different settings to optimize quantitation or to monitor species, not identified in the original DDA acquisition.

Identification and quantitation procedures in SWATH are separate, unlike traditional MS1 or MS2 spectral counting quantitation based on DDA. The ion library (the list of m/z and retention times for the parent and fragment ions) is created based on preliminary DDA measurements and may (or may not) contain information on the species fragmented during SWATH acquisitions. This is particularly critical when samples with significant variations in proteins content are analyzed. Inclusion of the whole repertoire of proteins into the ion library might require preliminary DDA runs for all samples to be compared or extensive 2D-LC MS DDA acquisition.

The limited number of SWATH applications reported to date have targeted small populations of proteins [25–27]. We questioned the potential of SWATH to provide a proteome-wide snap shot of protein expression for a particular organism. This was felt to be an attractive method relative to other quantitative techniques in that SWATH should have higher reproducibility between replicates than DDA. Furthermore, the capability introduces the potential to compare large numbers of different conditions with minimal sample preparation and method development.

The intent of this study was to evaluate the potential of SWATH as a method for the rapid, relative quantitation of large numbers of proteins in a single analysis. We have used a combination of DDA and SWATH in order to perform high-throughput relative quantitative analysis in a model organism Clostridium stercorarium [28]. SWATH quantitation was evaluated in terms of reproducibility of protein signal intensities between biological and technical replicates and relative protein signal intensity ratios across different growth conditions. The limit of reproducibility in DDA acquisitions
(MS2 quantitation based on fragment signal summation) was also determined by comparison of SWATH and DDA protein quantitation results.

2 Materials and methods

2.1 Culturing of C. stercorarium

Clostridium stercorarium DSM 8532 was cultured on 1191 medium [29] to midexponential phase at 60°C using either 0.2% xylose or 0.2% cellobiose as the primary carbon source. Cells from each culture were collected by centrifugation at 5000 g, washed three times with PBS (8.00 g/L NaCl, 0.20 g/L KCl, 1.44 g/L Na₂HPO₄, 0.24 g/L KH₂PO₄, 0.24 g/L KH₂PO₄, pH 7.5) and then frozen at −80°C until needed.

2.2 Filter assisted sample preparation for cell lysis and protein digestion

The filter assisted sample preparation method was used to generate tryptic digests for subsequent LC–MS acquisitions [30]. Cell pellets (~50 μL) were suspended in 500 μL of SDT buffer (100 mM Tris, 100 mM DTT, 4% SDS, pH 8.5) and heated at 95°C for 10 min. To ensure complete cell lysis, samples were sonicated using three 12 s pulses, with cooling on ice for 1 min in between each pulse. Cell lysates were frozen at −80°C until processed for analysis. A total of 200 μL of cell lysate was added to a 50 mL 10 mM Tris/MWCO Millipore (Billerica, MA) centrifugation filter already containing 12 mL of UA buffer (100 mM Tris, 8 M urea, pH 8.5). Samples were centrifuged at 4000 g until an equal volume of buffer was left on each filter. This washing procedure was repeated twice in order to remove the majority of SDS. An equal volume of 100 mM iodoacetamide solution was added to each sample and left at room temperature in the dark for 45 min. Samples were washed twice with 12 mL of 100 mM ammonium bicarbonate to remove excess urea. Protein concentration was determined by the BCA assay [31]. Sequencing grade trypsin (Promega, Madison, WI) was added to each vial at 1:100 enzyme:substrate ratio and incubated overnight at room temperature. Peptides were collected by the addition of 1 mL of 500 mM NaCl and centrifugation at 4000 g into a clean 50 mL tube. Final peptide concentration was determined by nano drop UV absorbance spectrometer (Thermo Fisher, Rockford, IL) at 280 nm. Peptide samples were desalted by RP-HPLC, lyophilized and resuspended in 0.1% formic acid and spiked with a six peptide standard mixture [32] before subsequent LC–MS analyses in DDA and SWATH acquisition modes.

2.3 LC–MS/MS analysis

A Triple TOF 5600 mass spectrometer (ABSciex, Mississauga, ON) coupled to a nano-flow Tempo LC system (Eksigent, Dublin CA) was used for the analysis. Samples (10 μL) were injected via a 300 μm × 5 mm PepMap100 trap column (Thermo Fisher, Rockford IL) and separated on a 100 μm × 200 mm analytical column packed with 5 μm Luna C18 (Phenomenex, Torrance CA). Both eluents A (water) and B (acetonitrile) contained 0.1% formic acid as ion-pairing modifier, samples were separated using a 0.5–30% B gradient over 105 min (0.28% acetonitrile/min) followed by 5 min of washing (90% acetonitrile) and a 10 min equilibration (0.5% acetonitrile) step. Either 2 or 0.5 μg of digest was injected for DDA or SWATH analyses, respectively.

Each cycle of data dependent acquisition included a 250 ms MS scan (400–1600 m/z) and up to 40 MS/MS (100 ms each, 100–1600 m/z) for ions with charge state from +2 to +5 and an intensity of at least 300 counts per second. Selected ions and their isotopes were dynamically excluded from further fragmentation for 12 s. Raw spectra files were converted to searchable Mascot Generic File format carrying MS/MS acquisition information. Peptide identifications were performed using a customized version of the X!Tandem algorithm [33] (complete carbamidomethyl Cys modification, maximum of one missed cleavage, mass accuracy of ±10 ppm and 0.05 Da for parent and fragment ions, respectively). False positive rates were computed internally by X!Tandem. Retention times for all identified peptides were assigned to each nonredundant species as the intensity weighted time average for the two most intense matching MS/MS spectra.

Each cycle of SWATH analysis consisted of a 250 ms MS scan and a 100 ms MS/MS scan in 25 m/z blocks in 400–1250 m/z range: a total of 34 SWATH blocks collected for each scan. Precursor selection windows had an overlap of 1 Da with each adjacent window to ensure complete isotope coverage between SWATH blocks. Collision energy was set to optimum energy for a +2 ion at the center of each SWATH block with a 15 eV collision energy spread. The mass spectrometer was always operated in high sensitivity mode.

2.4 Label free MS2 quantitation

Protein level DDA expression values were extracted from X!Tandem XML files given as the “sumI” variable listed for each protein. More specifically, this value can be found in X!Tandem XML reports under the “sumI” field of the “protein” declaration. The sumI value is simply the summation of all fragment intensities obtained from collisionally induced dissociation spectra for each peptide belonging to a particular protein.

2.5 Label free quantitation with data independent acquisition

The ion library used for SWATH quantitation was constructed with an in-house algorithm that extracted fragment
mass to charge ratios directly from X!Tandem XML files for each of the four DDA runs and combined them into a single library, under the assumption that peptides seen in any DDA run could potentially be detected in all SWATH runs. Peptides for inclusion in the ion library were only taken from proteins that had at least two nonredundant peptides identified and with a protein expectation value \( \log(e) \leq -3 \) [34].

For every peptide in the library, the most intense CID spectrum in the DDA run collection was selected to provide the SWATH transitions, with its parent \( m/z \) and charge values as the “Q1” and “prec_z” column entries, respectively. The “confidence” column for each peptide was computed as \( 0.99 \cdot 10^{-\text{expectation-value}} \). The peptide’s amino acid sequence was used to compute all possible singly and doubly charged b-ion and y-ion fragments, giving a series of “Q3” entries, each having “frg_type,” “frg_z,” and “frg_nr” columns. Observed CID fragment intensities were integrated across a ± 20 PPM window from each computed Q3 transition value, yielding the “relative_intensity” column value; any transition with integrated value greater than zero was included in the final ion library. The retention times of these peptides were averaged across the four runs. This nonredundant, averaged, collection was then formatted to a tab-delimited table of parent and fragment transitions to drive SWATH quantitation with PeakView (ABSciex, Missisauga ON). The settings used by PeakView to perform SWATH quantitation were as follows: (1) mass accuracy 50 ppm (i.e. ± 25 ppm from ion library mass), (2) retention time 5 min (i.e. ± 2.5 min from ion library retention time) (3) use six peptides with six transitions required for each peptide, (4) 1% false discovery rate (5) only use peptides with 99% confidence in identification or higher. Peak area outputs from PeakView were further organized into tab delimited text files containing log2 signal intensities between biological replicates and different growth states for both DDA and SWATH signal intensities. All further data analysis and graph generation was performed with the R programming language.

3 Results and discussion

3.1 Generating an ion library for SWATH quantitation

Protein quantitation by SWATH was performed postacquisition using an ion library based on information extracted from data dependent acquisition. The ion-library encompassed all proteins detected in each sample, in order to maximize the number of possible proteins quantified by SWATH. In our study four replicates of \( C. \) stercorarium were grown on xylose or cellobiose (two biological replicates each), then digested and analyzed by 1D LC–MS/MS. The final library contained 191 972 transitions spanning 15 075 peptides belonging to 1309 proteins. Identifications between biological replicates were very reproducible with ~90% overlap between samples (Fig. 2). A total of 998 and 980 proteins were identified in both replicates for xylose and cellobiose conditions, respectively, with a false positive rate of 0.40–0.43%. This result illustrates advances in the performance of bottom-up proteomic analysis by DDA—previous studies on the reproducibility of protein identifications gave overlaps of between 70–80% between technical replicates [19, 35, 36]. This increase in reproducibility between replicates is likely a product of increasing the numbers of peptides that can be analyzed by MS/MS in a single scan cycle.

The DDA derived information on peptide transitions and retention times was used to construct an ion library. Peptides were only included if their corresponding protein identification had at least two nonredundant peptides and a protein-level expectation value of \( \log(e) \leq -3 \). The final library contained 191 972 transitions spanning 15 075 peptides belonging to 1309 proteins. The overlap in potential peptide transitions was found to be very low, finding only 250 transition collisions for 222 peptides (out of 191 972 transitions in the original ion library). Further analysis was conducted assuming that the low number of transition collisions had
an insignificant effect on quantitation. PeakView transition filtering constraints of retention time ± 2.5 min and mass ± 23 ppm was applied to the four SWATH run collection (Fig. 1), resulting in a peptide level intensity report containing 4704 peptide entries for 1207 proteins (see the Supporting Information). If proteins with two or more peptides are used for SWATH identification this gives 1030 proteins quantified by SWATH. Thus, with only 4 × ~2 h SWATH runs it was possible to quantify ~40% of predicted C. stercorarium open reading frames (GenBank Accession: CP004044.1) under two different growth conditions.

3.2 Quantifying the C. stercorarium proteome using MS/MS signal intensities in DDA and SWATH modes

The reproducibility of SWATH quantitation was examined by calculating the coefficient of determination \( R^2 \) between log2 protein signal intensities across biological replicates (Fig. 3). For further evaluation we only used proteins quantified by SWATH with two or more peptides. The majority of proteins with only one peptide quantified had poor reproducibility between replicates and were discarded from analysis (data not shown). The Triple TOF 5600 provides very high MS/MS acquisition rates (for this study the acquisition rate was set to 40 MS/MS per cycle) giving consistent identification outputs between replicate runs. The higher sampling rate minimizes the stochastic nature of parent ion selection; in combination with the high reproducibility of MS/MS acquisition, MS2 quantitation signals are more stable. The \( R^2 \) value for DDA quantitation was 0.929 (Fig. 3). The \( R^2 \) value for MS2 quantitation by DDA between technical replicates was 0.926 (data not shown). The \( R^2 \) value for SWATH quantitation was 0.960 for combined xylose and cellobiose datasets (Fig. 3). Two technical replicates of C. stercorarium analyzed by SWATH had a similar \( R^2 \) value of 0.978. The similar reproducibility for biological and technical replicates suggests that the higher sampling rate is minimizing the stochastic nature of parent ion selection and provides more consistent identifications between replicates.

The dynamic range for SWATH quantitation was slightly higher than for DDA quantitation. The dynamic range for DDA was roughly four orders of magnitude \((7.7 \times 10^1-1.3 \times 10^6\) or \(12.9-27.0\) in \( \log_2 \) units) while SWATH covered nearly five orders of magnitude \((1.3 \times 10^1-1.2 \times 10^5\), \(10.4-26.9\) in \( \log_2 \) units). Additionally, the distribution of protein signal intensities between SWATH replicates over this range is nearly linear (Fig. 3), whereas the deviation of DDA protein signal intensities between replicates is greater at lower intensities—likely the result of inconsistent parent ion selection for low abundance species. The DDA peak selection criteria are based on ion intensity, giving multiple MS/MS acquisitions for abundant components and thus a complete profile of the corresponding peptide peak. Conversely, low abundance peptides are fragmented only once or twice at random MS2 fragmentation intensities from the peptide’s chromatographic peak, yielding inconsistent peak profiles across multiple runs. The increased variation in protein signal intensity for these low abundance proteins is likely the result of MS2 signal noise contributing to the extracted peptide intensities.

While correlation between biological replicate protein signal intensities was similar for SWATH and DDA (~0.960 and 0.929, respectively), the nature of DDA meant not every protein was detected in every run. Of the 1030 proteins quantified by SWATH, 88 were identified by DDA in only three replicates, 79 were identified in only two replicates, and 25 were only identified in a single replicate (192 proteins total), supporting the notion that proteins inconsistently detected by DDA may indeed be present in all four replicate samples. The average log2 DDA signal intensity for proteins identified in three replicates was 16.4, proteins in two replicates had an average signal intensity of 16.7 and proteins only identified in a single replicate had an average signal intensity of 15.7. The average SWATH log2 signal intensities for the same proteins were 18.2, 17.9, and 17.0, indicating that SWATH was able to better quantify the population of proteins with inconsistent DDA results at an increased signal intensity of ~2–3 fold over DDA. To put these numbers in perspective, in the

![Figure 2. Venn diagrams showing overlap in protein identifications (≥2 peptides, log2(≤) ≤ −3) between biological replicates of C. stercorarium cultured on either xylose (top) or cellobiose (bottom).](image-url)
four original DDA acquisitions, 157 proteins were identified in a single replicate, 101 were identified in two replicates, 124 were identified in three replicates, and 913 proteins were identified in all four replicates (≥2 peptides, protein expectation value $\log(e) \leq -3$).

### 3.3 Relative protein quantitation in *C. stercorarium* cultured under different conditions

The purpose of most relative quantitation studies is to identify those proteins that are altered under different biological conditions [37]. The general strategy is to focus on those proteins which display the greatest variation in abundance ratios relative to the assumed normal frequency distribution of ratios for the entire protein population. However, the actual observed ratio for any protein is a combination of the reproducibility of the measurements and of the actual physiological changes. If we make the assumption that the variation between biological replicates is the result of technical variation we can make an estimate of this by determining the reproducibility of multiple replicates. It can then be assumed that any variation in ratios beyond this value between different biological states is the result of the biological response.

We calculated the relative protein expression ratios between biological replicates under each growth condition based on SWATH and DDA derived quantitation. This provided measures of the expected technical variation. The SDs between biological replicate ratios for SWATH data were 0.415 and 0.452 for xylose and cellobiose replicates, respectively. Similarly the SD for ratios calculated across biological replicates using DDA signal intensity was 0.640 and 0.683 for the same growth conditions. Estimations of the variation between “cross-states” (i.e. ratios of the signal intensities for the same proteins in cells grown on xylose or cellobiose) were then determined. In contrast to the biological replicates, the SDs of the frequency distributions of the cross-state comparisons displayed a much higher values i.e. 1.06 and 0.996 for SWATH data and 1.27 and 1.21 for DDA data (Fig. 4). These results also demonstrate how the SWATH analysis provided more consistent quantitation than DDA.

Measuring changes in relative protein expression has provided information on many different biological processes. However, most studies only focus on those proteins that exhibit significant changes relative to the population as a whole. This approach could overlook more subtle but biologically significant global changes in protein expression. Assuming normal probability distributions of biological replicate ratios, ~90–95% of ratios are within 1.5 SDs of the mean (Fig. 4). If the premise is that this variation is from culture conditions and sample processing, we can define the point where biological variation between different growth states becomes significant. Expressing this as a log₂ ratio, differences above 0.68 for SWATH quantitation or 1.0 for DDA quantitation would represent significant changes relative to those expected for technical variation. Based on this value 198 out of 1030 proteins (~20% of all proteins quantified) displayed statistically significant changes in protein expression under the two growth conditions in both biological replicate pairs where at least two peptides were quantified. Increasing the number of biological replicates analyzed per condition would offer a more precise estimate of the cut-off value for significant biological variation. These results were independently verified by 1D iTRAQ [38] showing similar changes in protein expression with many proteins related to carbohydrate metabolism found to be differentially regulated between these two growth conditions (unpublished data). Collectively, the data suggest that there are extensive changes in protein expression in response to different growth conditions.
3.4 Protein quantitation using ion libraries derived from 2D LC–MS experiments

The number of proteins quantified by SWATH is limited by the number of proteins identified by DDA (i.e., proteins included in the ion library), and also by the large protein dynamic range present in complex biological lysates [39]. Analyzing the original samples by 2D LC–MS/MS can easily increase the number of proteins identified by DDA and generate an extended ion library. 2D LC–MS/MS acquisition using pH 10—pH 2 RP–RP separation scheme [40] identified 1563 proteins (including all 1309 from 1D acquisitions) and 15,279 peptides according to the same criteria (two peptides per protein, protein score log(e) ≥ −3). Since second dimension separation in 2D LC–MS/MS was performed using shorter gradients, constructing the ion library from these identification data required a peptide retention time realignment. This was achieved by using a standard mixture of peptides previously used in our lab [31]. However, the 2D library was only able to quantify 955 proteins in the four SWATH replicates (only including proteins with ≥2 peptides quantified, results not shown). Of these 955 proteins, 74 were not found by using the ion library based on 1D data dependent acquisitions. The small increase in the number of uniquely quantified proteins suggests that limitations in the amount of the sample subjected to SWATH acquisitions might be a limiting factor to achieving deeper DIA coverage. This also suggests that more substantial gains can be made in the number of proteins quantified by SWATH through the application of 2D LC–MS/MS before SWATH analysis. A reproducible first dimension fractionation should reduce noise and transition overlap in each SWATH block, while also permitting injection of larger quantities of protein digests.

4 Concluding remarks

Data independent SWATH acquisition can be used for the rapid simultaneous quantitation of a large number of proteins in whole cell digests. Using 2 h SWATH acquisitions we were able to quantify 1030 proteins (~40% of predicted C. stercorarium open reading frames) in four replicates under two different growth conditions. The protein quantitation by SWATH demonstrated good reproducibility between biological replicates and the capability to detect the regulation of protein expression in the bacterium grown on different substrates. We also illustrated that the amount of the injected digest and the large dynamic range of protein abundances limit the number of quantified proteins, rather than the size of the ion library used to interrogate the DIA outputs. This suggests that the number of proteins accessible in SWATH could be increased through 2D fractionation of peptide digests prior to SWATH analysis—this will reduce sample complexity and further improve quantitation by elimination of overlapping transitions and an overall noise reduction. However, analysis by 2D LC–MS/MS may be unnecessary if the quantitation targets are high abundance proteins easily detected via 1D DIA analysis as these were reproducibly quantified by SWATH. The static, additive nature of the ion library permits future analyses of this organism from only SWATH acquisitions. Once DDA acquisition is performed, the fragmentation patterns and chromatographic properties of peptides are transferred to the ion library. Potentially, this collection could continue to be updated through deeper fractionation or enrichment of samples until all possible proteins are detected. Furthermore, we found that quantitation of the most abundant proteins using the MS2 signal on a Triple TOF 5600 can provide information comparable to SWATH. SWATH was
able to recover quantitative data for proteins that were not identified in all replicates by DDA and quantified them more reproducibly.

This work was supported in part by grants from Genome Canada (Sparling, Levin, Wilkins and Krokhin) and the Natural Sciences and Engineering Research Council of Canada (Krokhin).

The authors have declared no conflict of interest.

5 References

[1] Aebersold, R., Goodlett, D. R., Mass spectrometry in proteomics. Chem. Rev. 2001, 101, 269–295.
[2] Craig, R., Cortens, J. C., Fenyo, D., Beavis, R. C., Using annotated peptide mass spectrum libraries for protein identification. J. Proteome Res. 2006, 5, 1843–1849.
[3] Nesvizhskii, A. I., Protein identification by tandem mass spectrometry and sequence database searching. Methods Mol. Biol. 2007, 367, 87–119.
[4] Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., Kuster, B., Quantitative mass spectrometry in proteomics: a critical review. Anal. Bioanal. Chem. 2007, 389, 1017–1031.
[5] Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F. et al., Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat. Biotechnol. 1999, 17, 994–999.
[6] Stahl-Zeng, J., Lange, V., Ossola, R., Eckhardt, K. et al., High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites. Mol. Cell Proteomics 2007, 6, 1809–1817.
[7] Wiese, S., Reidegeld, K. A., Meyer, H. E., Warscheid, B., Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. Proteomics 2007, 7, 340–350.
[8] Lange, V., Picotti, P., Domon, B., Aebersold, R., Selected reaction monitoring for quantitative proteomics: a tutorial. Mol. Syst. Biol. 2008, 4, 222–236.
[9] Evans, C., Noirel, J., Ow, S. Y., Salim, M. et al., An insight into iTRAQ: where do we stand now? Anal. Bioanal. Chem. 2012, 404, 1011–1027.
[10] Wong, J. W., Cagney, G., An overview of label-free quantitation methods in proteomics by mass spectrometry. Methods Mol. Biol. 2010, 604, 273–283.
[11] Schilling, B., Rardin, M. J., MacLean, B. X., Zawadzka, A. M. et al., Platform-independent and label-free quantitation of proteomic data using MS1 extracted ion chromatograms in skyline: application to protein acetylation and phosphorylation. Mol. Cell. Proteomics 2012, 11, 202–214.
[12] Nagaraj, N., Wisniewski, J. R., Geiger, T., Cox, J. et al., Deep proteome and transcriptome mapping of a human cancer cell line. Mol. Syst. Biol. 2011, 7, 548–556.
[13] Griffin, N. M., Yu, J., Long, F., Oh, P. et al., Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis. Nat. Biotechnol. 2010, 28, 83–89.
[14] Lundgren, D. H., Hwang, S. I., Wu, L., Han, D. K., Role of spectral counting in quantitative proteomics. Expert Rev. Proteomics 2010, 7, 39–53.
[15] Stahl, D. C., Swiderek, K. M., Davis, M. T., Lee, T. D., Data-controlled automation of liquid chromatography/tandem mass spectrometry analysis of peptide mixtures. J. Am. Soc. Mass Spectrom 1996, 7, 532–540.
[16] Krey, J. F., Wilmarth, P. A., Shin, J. B., Klimek, J. et al., Accurate label-free protein quantitation with high- and low-resolution mass spectrometers. J. Proteome Res. 2014, 13, 1034–1044.
[17] Spinelli, K. J., Klimek, J. E., Wilmarth, P. A., Shin, J. B. et al., Distinct energy metabolism of auditory and vestibular sensory epithelia revealed by quantitative mass spectrometry using MS2 intensity. Proc. Natl. Acad Sci USA 2012, 109, E268–E277.
[18] Shin, J. B., Streijger, F., Beynon, A., Peters, T. et al., Hair bundles are specialized for ATP delivery via creatine kinase. Neuron 2007, 53, 371–386.
[19] Liu, H., Sadygov, R. G., Yates, J. R., 3rd, A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal. Chem. 2004, 76, 4193–4201.
[20] Holcapek, M., Jirasko, R., Lisa, M., Recent developments in liquid chromatography-mass spectrometry and related techniques. J. Chromatogr. A 2012, 1259, 3–15.
[21] Andrews, G. L., Simons, B. L., Young, J. B., Hawkridge, A. M., Muddiman, D. C., Performance characteristics of a new hybrid quadrupole time-of-flight tandem mass spectrometer (TripleTOF 5600). Anal. Chem. 2011, 83, 5442–5446.
[22] Panchaud, A., Scherl, A., Shaffer, S. A., von Haller, P. D. et al., Precursor acquisition independent from ion count: how to dive deeper into the proteomics ocean. Anal. Chem. 2009, 81, E681–E688.
[23] Silva, J. C., Denny, R., Dorschel, C., Gorenstein, M. V. et al., Simultaneous qualitative and quantitative analysis of the Escherichia coli proteome: a sweet tale. Mol. Cell. Proteomics 2006, 5, 589–607.
[24] Gillet, L. C., Navarro, P., Tate, S., Rost, H. et al., Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Mol. Cell. Proteomics 2012, 11, O111 016717.
[25] Moran, D., Cross, T., Brown, L. M., Colligan, R. M., Dunbar, D., Data-independent acquisition (MSE) with ion mobility provides a systematic method for analysis of a bacteriophage structural proteome. J. Virol. Methods 2014, 195, 9–17.
[26] Heim, S., Dobritzsch, D., Rodiger, A., Agne, B., Baginsky, S., Protein identification and quantification by data-independent acquisition and multi-parallel collision-induced dissociation mass spectrometry (MSIE) in the chloroplast stroma proteome. J. Proteomics 2014, 98, 79–89.
[27] Liu, Y., Huttenhain, R., Surinova, S., Gillet, L. C. et al., Quantitative measurements of N-linked glycoproteins in human plasma by SWATH-MS. Proteomics 2012, 13, 1247–1256.
[28] Madden, R. H., Isolation and characterization of Clostridium stercorarium sp. nov., cellulolytic thermophile. Int. J. Syst. Bacteriol. 1983, 33, 837–840.
[29] Sparling, R., Islam, R., Cicek, N., Carere, C. et al., Formate synthesis by Clostridium thermocellum during anaerobic fermentation. *Can. J. Microbiol.* 2006, 52, 681–688.

[30] Wisniewski, J. R., Zougman, A., Nagaraj, N., Mann, M., Universal sample preparation method for proteome analysis. *Nat. Methods* 2009, 6, 359–362.

[31] Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K. et al., Measurement of protein using bicinchoninic acid. *Anal. Biochem.* 1985, 150, 76–85.

[32] Krokhin, O. V., Spicer, V., Peptide retention standards and hydrophobicity indexes in reversed-phase high-performance liquid chromatography of peptides. *Anal. Chem.* 2009, 81, 9522–9530.

[33] Craig, R., Cortens, J. P., Beavis, R. C., The use of proteotypic peptide libraries for protein identification. *Rapid Commun. Mass Spectrom.* 2005, 19, 1844–1850.

[34] Fenyo, D., Beavis, R. C., A method for assessing the statistical significance of mass spectrometry-based protein identifications using general scoring schemes. *Anal. Chem.* 2003, 75, 768–774.

[35] Nilsson, T., Mann, M., Aebersold, R., Yates, J. R., 3rd et al., Mass spectrometry in high-throughput proteomics: ready for the big time. *Nat. Methods* 2010, 7, 681–685.

[36] Bateman, N. W., Goulding, S. P., Shulman, N. J., Gadok, A. K. et al., Maximizing peptide identification events in proteomic workflows using data-dependent acquisition (DDA). *Mol. Cell. Proteomics* 2014, 13, 329–338.

[37] Unwin, R. D., Evans, C. A., Whetten, A. D., Relative quantification in proteomics: new approaches for biochemistry. *Trends Biochem. Sci.* 2006, 31, 473–484.

[38] Aggarwal, K., Choe, L. H., Lee, K. H., Shotgun proteomics using the iTRAQ isobaric tags. *Brief Funct. Genomic Proteomic* 2006, 5, 112–120.

[39] Corthals, G. L., Wasinger, V. C., Hochstrasser, D. F., Sanchez, J. C., The dynamic range of protein expression: a challenge for proteomic research. *Electrophoresis* 2000, 21, 1104–1115.

[40] Dwivedi, R. C., Spicer, V., Harder, M., Antonovici, M. et al., Practical implementation of 2D HPLC scheme with accurate peptide retention prediction in both dimensions for high-throughput bottom-up proteomics. *Anal. Chem.* 2008, 80, 7036–7042.