Title

Sequential Windowed Acquisition of Reporter Masses for Quantitation-First Proteomics

Journal

Journal of Proteome Research, 18(4)

ISSN

1535-3893

Authors

Barshop, William D
Rayatpisheh, Shima
Kim, Hee Jong
et al.

Publication Date

2019-04-05

DOI

10.1021/acs.jproteome.8b00884

Peer reviewed

HHS Public Access

Author manuscript

J Proteome Res. Author manuscript; available in PMC 2020 April 05.

Published in final edited form as:

J Proteome Res. 2019 April 05; 18(4): 1893–1901. doi:10.1021/acs.jproteome.8b00884.

Sequential Windowed Acquisition of Reporter Masses for Quantitation-First Proteomics

William D. Barshop[1], **Shima Rayatpisheh**[1], **Hee Jong Kim**[1], **James A. Wohlschlegel**[*,1]

[1]Department of Biological Chemistry, David Geffen School of Medicine, UCLA, Los Angeles, CA, USA

Abstract

The standard approach for proteomic data acquisition of isobaric tagged samples by mass spectrometry is Data Dependent Acquisition (DDA). This semi-stochastic, identification-first paradigm generates a wealth of peptide-level data without regard to relative abundance. We introduce a data acquisition concept called Sequential Windowed Acquisition of Reporter Masses (SWARM). This approach performs quantitation-first thereby allowing subsequent acquisition decisions to be predicated on user-defined patterns of reporter ion intensities. The efficacy of this approach is validated through experiments with both synthetic mixtures of *Escherichia coli* ribosomes spiked into human cell lysates at known ratios, and the quantitative evaluation of the human proteome's response to the inhibition of Cullin-based protein ubiquitination via the small molecule MLN4924. We find SWARM informed PRM acquisitions display effective acquisition biasing toward analytes displaying quantitative characteristics of interest, resulting in an improvement in the detection of differentially abundant analytes. The SWARM concept provides a flexible platform for further development of new acquisition methods.

Keywords

TMT; isobaric tags; isobaric; quantitative proteomics; DIA; SWATH; SWARM; iTRAQ

Introduction:

Comparative proteomic experiments have become increasingly dependent on the use of isobaric mass tags to facilitate the multiplexing of samples into single acquisitions[1]. As a bottom-up proteomic method, protein samples from multiple biological samples are digested with trypsin, and chemically tagged via reaction with either an amino or sulfhydryl reactive reagent. These tags are isobaric while intact but produce isotopically labeled reporter ions of different masses upon fragmentation. Under this regime, peptide ions are detected as intact analytes (MS1) and can be quantified after a single round of gas-phase purification through a linear ion trap or quadrupole followed by fragmentation and detection of reporter ion relative intensities (MS2). On newer instruments, multiple fragmentation products from the MS2 level may be subsequently copurified in a linear ion trap by synchronous-precursor-selection (SPS) and further fragmented to the MS3 level to produce reporter ion signals[2].

[*]Corresponding Author: jwohl@ucla.edu.

In bottom-up proteomic mass spectrometry, peptide ion targets are selected in real-time by the instrument through Data Dependent Acquisition (DDA). After intact analytes have been detected in an MS1 survey scan covering the *m/z* range of interest, putative peptide targets are ranked by signal intensity and sequentially isolated and fragmented before repeating the cycle with another MS1 survey scan. This acquisition methodology facilitates the identification of large numbers of peptides but does so in a way that is indiscriminate to quantitative trends in the isobaric tag reporter ion intensities. This results in frequent sampling of zero-fold change analytes that are likely of low biological interest. The glut of peptide and protein identifications and quantitative data can lead to thorough and deep characterization of a proteome but comes at the cost of increasing the stringency of multiple hypothesis testing correction during statistical tests for differential protein abundance[3].

To address these issues, we introduce the concept of a quantitation-first data acquisition cycle for isobaric tagged samples (Figure 1). This Data Independent Acquisition-like cycle (DIA), termed Sequential Windowed Acquisition of Reporter Masses (SWARM), leverages fragmentation of small regularly placed isolation windows tiled across the target *m/z* space. Unlike traditional DIA approaches, SWARM scans only acquire data for the reporter ion mass range, providing data about the relative abundance of isolated and fragmented analytes while lacking any identification information. These SWARM scans allow users to bias acquisition toward analytes exhibiting user-defined differences in reporter ion intensities. In this work, we utilize SWARM scanning to generate quantitative maps of isobarically tagged protein mixtures. These quantitative maps are subsequently filtered to create target lists for subsequent parallel reaction monitoring (PRM) experiments focused on the identification of the high fold change peptides.

Experimental Procedures:

An implementation of the SWARM scan cycle on an Orbitrap Fusion Lumos mass spectrometer:

Intact full (MS1) scans were interlaced between SWARM (MS2) scan cycles covering from 500 *m/z* to 900 *m/z* (Figure 1). Each SWARM isolation window was purified by the quadrupole with a width of 1.5 *m/z* and placed in a DIA-like design tiled uniformly across the *m/z* space. Each cycle contained 268 windows. The ribosome experiments contained an additional SWARM scan cycle offset by a half-window width. To reduce the number of available target windows for the PRM, the data from the second SWARM windows is dropped from consideration, while the MLN4924 SWARM acquisition omits this second scan cycle entirely. SWARM scans were acquired in the linear ion trap, using a high HCD collision energy of 60%, the normal scan rate for the ribosome experiments and the rapid scan rate for the MLN4924 experiments, a 1e4 AGC target, 5ms maximum injection time, and covering only the *m/z* range 125–132. Each window was organized as a target in a tMS2 experiment, with a default z of 2 for collision energy calculation purposes.

All raw datasets acquired and analyzed here are available within ProteomeXchange on the MassIVE data repository, under the identifier PXD011544.

SWARM Quantitative Map Data Processing and target selection:

SWARM datasets were imported into a Skyline document containing small molecule targets placed at the center of each SWARM window[4]. For each of these small molecule targets representing a SWARM window, extracted ion chromatograms for each reporter mass of the Thermo TMT6Plex kit were generated with a 0.7 *m/z* extraction tolerance (±0.35 Th). Intensity chromatograms were exported from Skyline and further processed by means of an in-house Python (v. 2.7) script. Briefly, the chromatograms were imported into a Pandas dataframe, and time indexed by the retention time of the first scan in the SWARM cycle. For each scan, SWARM reporter ion data was dropped if all reporter peak areas reported by Skyline fell below 15000. Further, individual reporter ion peak areas below 10000 were censored as NA values. The remaining reporter ion channel data was median equalized across all SWARM windows, and fold change values were calculated for every scan as the ratio of the average for the three condition replicates in each experimental condition and ultimately log2 transformed. These log2 relative fold change values were plotted across retention time as quantitative maps that enable visualization of the biological condition reporter ion fold changes throughout the sampled *m/z* space. These maps are constructed without the identities of analytes that contribute to the detected reporter ion fold changes. The maps were filtered such that only scans exhibiting fold changes that exceed a user specified threshold are retained, referred to here as the trigger condition. These high fold change SWARM scans are designated as trigger scans, as they represent *m/z*-retention time segments of the chromatographic separation from which we would desire to identify the analyte of interest. For the *E. coli* ribosome spike-in experiments, a 2-fold change minimum between the average of the triplicate conditions was designated as the trigger condition, while a 2.25-fold change threshold was used for the MLN4924 experiment. PRM target windows were generated, centered on and seeded by the triggering scan(s), and extending for 30–60 seconds depending on how many sequential scans in that window met the trigger condition.

Testing Reporter Ion DIA/SWARM Scan Rates:

For all scan rate testing experiments, ionization voltage was set to zero. This effectively forces scans to hit the maximum injection time limit during ion accumulation. MS2 scans were set to only cover the region 122–135 *m/z*, and built as part of a DIA cycle with quadrupole isolation windows of 2 *m/z* spanning from 500 to 900 *m/z*. Under these conditions, we varied the scan resolution of the orbitrap or the scan rate of the linear ion trap. Concomitantly, we varied the maximum injection time of the scans. For each combination of scan duration/resolution and maximum injection time, three one-minute datasets were acquired. The scan rate for each run was calculated across the full run, and the triplicates averaged and plotted. A small random wiggle was added to the data to enhance visualization of overlapping datapoints.

Nanoflow Chromatography and peptide ionization:

Coupled with the mass spectrometric analysis, samples were separated on a 25cm C18 reversed phase column (75uM ID × 25cm), packed with 1.9uM ReproSil-Pur beads with 120A pores (Dr. Maisch GmbH). Analytical columns were connected to an Ultimate 3000

ProFlow nanoflow UHPLC (Thermo Fisher Scientific). Each acquisition was run on a 140min chromatographic gradient with water with 3% DMSO and 0.1% formic acid as buffer A, and acetonitrile with 3% DMSO and 0.1% formic acid as buffer B. Briefly, gradients began at 400nl/min flow rate, at 1% organic. Over the first 5 minutes, the flow rate drops to 200nl/min, and the percent organic increases to 5.5%B, with a Dionex curve value of 4. Until 128min, a linear gradient to 27.5%B was executed, followed by a linear increase to 35%B at 135min. Over one minute, the percent organic increases to 80%B and held for two minutes, before dropping back to 1%B over a half minute and held until the end of the chromatographic run. Ionization was carried out via electrospray ionization through a Nimbus ionization source (Phoenix S&T), with ionization voltage set to 2.2kV.

PRM Follow-Up Data Acquisition:

The PRM acquisitions were collected using orbitrap MS2 scans at 15,000 resolution at 400 *m/z* after CID fragmentation at 35% collision energy, with a 10ms activation time and an AGC target of 5e4 and a maximum injection time of 150ms. The data from the MS2 scans were used to target 10 fragments per scan for SPS-MS3 (Synchronous Precursor Selection) occurring from 400–1500 *m/z*, and omitting targets occurring in a window from −18 *m/z* to +5 *m/z* from the PRM target/SWARM window center. MS3 scans were generated using HCD fragmentation at 60% collision energy and detecting the reporter ion intensities in the linear ion trap with a rapid scan from 125 to 132 *m/z*, an AGC target of 1e4 and 150ms maximum injection time. Quadrupole isolation window was set to 1.5 *m/z* to match the placement of the SWARM window responsible for triggering the PRM. MS1 precursor scans covering 495–905 *m/z* were acquired between each PRM cycle at a resolution of 120,000 with a maximum injection time of 50ms and an AGC target of 4e5.

Paired DDA Data Acquisition:

DDA acquisitions were carried out utilizing the manufacturer's default SPS-MS3 acquisition method but matched to the same precursor range (500–900 *m/z)* sampled in the SWARM acquisitions. Briefly, Orbitrap MS1 scans were acquired at 120,000 resolution in a 3 second cycle time, followed by selection of precursors in the 500–900 *m/z* space, with monoisotopic precursor selection set to peptide, an intensity threshold of 5e3, charge states from +2 through +7, and a 30 second dynamic exclusion. These scans were carried out with an AGC target of 4e5 and a maximum injection time of 50ms. Selected targets were subjected to CID fragmentation at 35% collision energy, with a 10ms activation time. The AGC target was set to 1e4, with a maximum injection time of 35ms and scanned in the linear ion trap at the turbo scan rate. Ten fragments were selected for SPS-MS3 after filtering for those from 400–1500 *m/z*, and not occurring in a window from −18 *m/z* to +5 *m/z* from the precursor and excluding isobaric tag losses from TMT fragmentation. The quantitation scans were obtained in the orbitrap, after HCD fragmentation at 65% collision energy, at a resolution of 50,000 with a maximum injection time of 105ms and an AGC target of 1e5.

E. coli ribosome sample preparation:

A single 15cm plate of HEK293 cells was harvested by scraping in cold PBS and pelleted at 800g. The cell pellet was lysed in 8M Urea with 100mM Tris-HCl at pH 8.0, with 1mM of AEBSF, DTT, leupeptin and pepstatin A. Pierce Universal Nuclease for Cell Lysis was

added to the lysate at 1:10000 by volume, followed by centrifugation at 20,000g to clarify lysate and recovery of the soluble fraction. The soluble lysate was precipitated in TCA, washed in ice-cold acetone, and the precipitate resuspended in 1M Urea, 50mM HEPES pH 8.5. Protein concentration was measured by BCA, and 300ug taken for digestion and TMT labeling, described below.

The protein content was split equally between all six channel labels in the 6-plex kit, yielding a constant protein background. Separately, 120ug of purified *E. coli* ribosome (New England Biosciences, cat #P0763S) was suspended in 1M Urea, 50mM HEPES and digested and TMT labeled in the same manner described. The final mixing of each TMT labeled sample allowed for defined ratios to be controlled for the *E. coli* ribosomes. Each TMT channel contained 8.3ug HEK293 lysate, while the *E. coli* ribosomes were added in at either a 1:1:1:5:5:5 ratio (.2ug or 1ug per channel) or 1:1:1:10:10:10 ratio (.1ug or 1ug per channel). After the final desalting, the sample was resuspended in 50uL, and 1uL used for each injection.

MLN4924 sample preparation:

Six 10cm plates of HeLa cells were cultured, with three of them treated for 24 hours before harvest with 1uM final concentration of MLN4924 (Chemietek, cat #CT-M4924), and the remaining three treated for an equivalent time with a DMSO vehicle control. Cells were harvested by scraping in cold PBS and pelleted at 800g. Each cell pellet was lysed in 8M Urea, 100mM Tris-HCl at pH 8.0, with 1mM of AEBSF, DTT, leupeptin and pepstatin A and 1:10000 Universal Nuclease (Pierce). After 30 minutes of rotation, samples were centrifuged at max speed to remove the cellular debris. Proteins were precipitated in 20% TCA followed by 3 cold acetone washes. Samples were resuspended in 1M urea, 50mM HEPES and protein content quantified and normalized by BCA, with 50ug from each sample taken for digestion and TMT labeling, as described below.

Proteomic Sample Digestion and TMT labeling:

For each sample, protein was reduced via addition of TCEP to a final concentration of 5mM and incubated for 20 minutes at room temperature, Iodoacetamide, to alkylate cysteine residues, was added to 10mM, followed by an additional 20-minute incubation in the dark. Endopeptidase Lys-C was added to each sample at a 1:100 ratio of enzyme to substrate, followed by a 4-hour incubation at 37°C in the dark. After the incubation, a 100mM stock of $CaCl_2$ was used to bring the solution's working concentration of $CaCl_2$ to 1mM. Trypsin protease was added to each sample at a 1:20 enzyme to substrate ratio, and the samples incubated overnight at 37°C in the dark. Finally, digestion was quenched by the addition of formic acid to a final concentration of 5%, and each sample centrifuged at maximum speed for 5 minutes to pellet any insoluble components, and the soluble fraction moved into new tubes. Samples were desalted by binding to C18 tips, washed twice in 200uL of 5% formic acid, and eluted in 60% ACN with 5% formic acid.

Desalted, digested samples were dried by SpeedVac, before being resuspended in TMT labeling buffer of 200mM HEPES pH 8.5 and 30% ACN. Samples were briefly sonicated, while isobaric labels from Thermo's TMTsixplex™ kit (Cat. #90066) were brought to room

temperature and resuspended in 41uL of anhydrous ACN per 800ug of TMT label. 5uL of the appropriate TMT label was added to each sample, mixed thoroughly, and incubated at room temperature for 1 hour. The labeling reaction was quenched with 6uL of 5% hydroxylamine, 200mM HEPES per 50ug of protein in each sample. After 15 minutes at room temperature, formic acid was added to a final concentration of 5%. Each TMT labeled sample was kept separately prior to appropriate mixing for its experiment, and a final round of C18 tip desalting, as described above. The final labeled, mixed, and desalted samples were resuspended in 5% formic acid prior to chromatographic and mass spectrometric analysis.

Proteomic Database Search:

All DDA and PRM data were searched via the Andromeda search engine provided within the MaxQuant platform (v 1.6.2.10). For each analysis, database searching was performed against the EMBL-EBI Human Reference Proteome, with the *E. coli* reference proteome additionally appended when appropriate. For data acquired using linear ion trap MS2 scans, fragment mass tolerance was set to 0.5Da, while Orbitrap MS2 scans were searched with 20ppm fragment mass tolerances. DDA acquisitions were searched with a first search tolerance of 20ppm and a main search tolerance of 4.5ppm. The PRM searches were allowed a 6Da window for both searches, as their scans are targeted to the center of the triggering SWARM window and therefore do not provide accurate precursor *m/z*. Carbamidomethylation was specified as a fixed modification, and Trypsin/P digestion of the database specified allowing for a maximum of 2 missed cleavage sites and using the “specific” digestion mode. Search results were filtered by MaxQuant’s provided reversed-sequence target-decoy FDR calculation at both a PSM and Protein level FDR of 1%[5–6].

Reporter Ion Quantification:

For each experiment, quantitative scans were filtered to require precursor isolation specificities of at least 70%. TMT reporter ion signals measured in the linear ion trap were extracted with a ±0.35 *m/z* tolerance, or ±.0035 *m/z* when measured in the Orbitrap. Reporter ion intensities for the DDA experiments, which utilize the orbitrap for MS3 quantitative scans, were generated by MaxQuant and read from the evidence table output of each runs’ respective analysis. For the PRM experiments, we extracted reporter ion intensities directly from the vendor ‘.raw’ files for every MS3 scan, and associated those values with the identifications made available in the MaxQuant evidence tables for each identified MS2 scan. Direct scan data and metadata access from the vendor data was provided by the RawDiag package (v 0.0.10) in R (v 3.5.0)[7]. Reporter ion intensities were calculated as the sum of the signal detectable within the extracted tolerance window relative to the position of the monoisotopic mass of each reporter ion.

When reporter ion intensities were extracted through RawDiag, we also calculated the precursor isolation purity for the associated MS2 scan from the appropriate MS1 master scan. To calculate precursor isolation purity, high resolution MS1 scans were converted from vendor format into mzML via msconvert (v. 3.0.11252) with vendor peak picking turned on[8]. The resulting mzML files were read into R via the MSnbase package (v. 2.6.2)[9]. For each PSM reported by MaxQuant, signal contained within the relevant MS2 scan’s precursor

isolation window was extracted with a 15ppm tolerance around the monoisotopic mass of the identified peptide and from any C13 isotope that would fall within the isolation window at the given charge state of the analyte. This extracted signal was divided by the sum of all signal within the extraction window to represent the precursor isolation purity. The R script responsible for the TMT signal extraction and integration with MaxQuant results for the SWARM+PRM samples has been included as part of the ProteomeXchange submission (PXD011544).

Statistical Analysis of TMT Intensities:

Reporter ion intensity outputs were used to provide relative quantitative values to the R package MSstatsTMT (v. 0.4.3), after normalizing each run's TMT channels by equalization of median intensities calculated from the subset of all human derived PSMs. Protein intensities for each experiment and channel were calculated via log transformation of the summation of TMT reporter ion intensities across the protein's filtered PSMs. The requirement of a minimum of two uniquely mapping quantitative features for a protein to be carried into the comparative analysis was imposed through MSstatsTMT's "removeProtein_with1Feature" and "useUniquePeptide" options. Fold changes were calculated between biological conditions, and p-values for differential abundance testing generated by means of the *t*-test implementation provided. The Benjamini-Hochberg method was employed to adjust p-values for multiple hypothesis testing. For the MLN4924 experiments, Cullin-associated genes were annotated by a list of 549 Cullin-Associated genes aggregated by a prior study, after mapping to 551 Uniprot Swissprot-reviewed protein accessions[10–11].

For each TMT acquisition, the full table of analysis outputs and intensities are available in Supplemental Tables 2–7.

Results and Discussion:

For each sample, SWARM scans were acquired from 500–900 *m/z* in 1.5 *m/z* windows (Figure 1A). These scans were collected in the linear ion trap, where we can leverage the benefit of a small ion capacity of the trap for rapid fill times and simultaneously exploit the linear ion trap's scan duration proportionality to the width of the *m/z* range of the scan. Under these conditions, the linear ion trap is capable of scanning in excess of 65Hz on a DIA cycle when scans are limited to the TMT reporter *m/z* range while the Orbitrap scans at a rate of approximately 37.5Hz (Supporting Information Figure S1–2). The SWARM data was imported into Skyline, and TMT reporter extracted ion chromatograms generated for each precursor window (Figure 1B)12. Reporter ion channels were filtered to remove low-intensity regions of the chromatography, and log transformed fold change ratios were calculated for the two conditions. The reporter ion relative intensities and sample condition fold change values that are calculated for each SWARM precursor window and resampled during each SWARM cycle represent a relative quantitative map of the sample. After applying fold change filters between the sample condition triplicate sets, we are able to generate a map of only the high fold change scans (Trigger Map), which seed target *m/z* and retention times for follow-up PRM acquisition (Figure 1C). Each sample was additionally

acquired by DDA, utilizing the vendor default SPS-MS3 quantitative method for TMT samples (Experimental Procedures).

We first tested the SWARM methodology using a model protein mixture in which equal amounts of HEK293 whole cell lysate were labeled across all six channels using 6-plex TMT isobaric tags. Purified *E. coli* ribosomes were labeled and spiked into each channel at ratios of either 1:1:1:5:5:5 or 1:1:1:10:10:10. These samples were acquired by SWARM and a follow-up PRM (SWARM+PRM) targeting windows with at least two-fold change with SPS-MS3 for quantitation (Figure 2, Supporting Information Figure S3). The datasets were searched using the Andromeda search algorithm, as part of MaxQuant (v. 1.6.2.10) against the EMBL Human and *Escherichia coli* reference proteome databases concatenated together[5–6]. Precursor mass tolerance for DDA searches was set to 20 ppm, while a 6 Da window was allowed for the PRM scans which contain no assigned target precursor to provide an accurate *m/z*. This necessity stems from the targeting of the PRM isolation windows at the same *m/z* of the SWARM window which produced the triggering scan(s) instead of targeting the accurate mass of a single precursor as in DDA.

As expected, the PRM runs identify many fewer confident peptide identifications when compared to the DDA acquisitions (Figure 3A, Supporting Information Table 1). However, we observed more similar counts of confident *E. coli* peptide identifications between the PRM and DDA runs, with slightly higher counts in the DDA runs (Figure 3B). Concomitantly, we observe a dramatic increase in the number of *E. coli* peptide-spectrum matches (PSMs) in the PRM datasets corresponding to ribosome-derived peptides (Figure 3C). This increase in the number of PSMs comes with the benefit of increasing the quantitative sampling of these *E. coli* peptides. Of the 58 Swiss-Prot validated, *E. coli* ribosomal proteins on Uniprot annotated with the gene ontology term “Structural Constituent of Ribosome [3735]”, confident identifications are slightly more numerous in the DDA runs than from the SWARM+PRM acquisitions (Supporting Information Table 1)13. However, at the protein level, the wealth of data produced by unbiased data dependent sampling of both the human and *E. coli* proteins can come at the cost of harsh correction when adjusting for multiple hypothesis testing.

For the quantitative comparisons, proteins with an adjusted p-value less than or equal to 0.05 with a minimum 2-fold change were considered significantly differentially abundant. The focused acquisition of the SWARM+PRM approach yields more sensitive comparisons, producing a list of 34 significantly enriched *E. coli* proteins in the 5:1 ribosome sample (Figure 4A, Supporting Information Table 2) while no *E. coli* proteins were deemed significantly enriched in the DDA acquisition under these conditions (Figure 4B, Supporting Information Table 3). At more extreme fold changes, the advantage of SWARM+PRM is less pronounced although still evident. In the 10:1 ratio ribosome sample, SWARM+PRM is able to detect 32 *E. coli* proteins as differentially enriched while DDA offers a close 27 significant proteins (Figure 4C–D, Supporting Information Tables 4–5). In these samples, SWARM+PRM offers an effective edge in the detection of differentially abundant *E. coli* ribosomal proteins at lower adjusted p-value thresholds (Supporting Information Figure S5). The t-statistic depends partially on the magnitude of the difference of means (i.e. magnitude of protein fold-change between conditions), the dispersion, and number of the measurements

of each group. Further, the resulting p-values are subject to adjustment for multiple hypothesis testing correction. We observe that the DDA datasets produce protein fold-change estimates that appear to be modestly more accurate than SWARM+PRM, with more *E. coli* proteins nearer to the anticipated fold-changes (Figure 4). Simultaneously, the distribution of standard deviations for the log-scaled protein intensity measurements in each condition reveals a slightly lower dispersion in the replicate channel measurements during the DDA runs when compared to the SWARM+PRM runs for the targeted *E. coli* proteins (Supporting Information Figure S6). With similar dispersion and fold-change measurements, results suggest that the differentiating factor for these experiments is the focusing of acquisition on a small population of analytes during SWARM+PRM. This leads to a modestly sized population of analyte targets when compared to the large number of comparisons tested when implementing DDA. Overall, in these synthetic samples, SWARM +PRM provides effective biasing of instrument acquisition focus toward *E. coli* ribosomes, substantially increasing *E. coli* PSMs while maintaining controlled false positive rates during differential abundance testing.

We then applied SWARM+PRM toward surveying proteins whose abundance is altered upon the reduction of Cullin RING Ligase (CRL) activity via inhibition of the Nedd8-activating enzyme (NAE) via the small molecule MLN4924[14]. Proteins dysregulated upon CRL inhibition would be candidates for association with CRL E3 ubiquitin ligase complexes. Triplicate cultures of HeLa cells were treated with either 1uM MLN4924 or DMSO for 24 hours. The six cultures were lysed, and the soluble protein content of the whole cell lysate digested and labeled with TMT 6-plex isobaric tags. These samples were analyzed using SWARM+PRM design similar to the ribosome samples, but targeting SWARM windows at retention times exhibiting at least 2.25-fold change in either direction between the DMSO and MLN4924 treated triplicate channels (Supporting Information Figure S4A–B). A corresponding DDA acquisition was also acquired under the same parameters as in the ribosome experiments. Quantified proteins were considered significantly differentially abundant at a maximum adjusted p-value of 0.05, and a minimum fold change of 1.75 in either direction. We additionally compared the identified proteins to a list of previously annotated Cullin-associated genes. The SWARM+PRM results showed an improvement in the detection of differentially abundant proteins after MLN4924 treatment compared to the DDA acquisition (Figure 5A–B)10. We identified a population of 71 proteins meeting these thresholds from the SWARM+PRM experiment, where no proteins were deemed significant from the analysis of the DDA data (Supporting Information Tables 6–7). Of the proteins classified as significantly differentially abundant, 3 were previously annotated as being Cullin-associated, including the well characterized Cullin target β-catenin[15].

The choice of fold-change threshold for the scheduling of the follow-up PRM impacts which segments of the retention time and *m/z* space are interrogated. By extension, this alters the cycle time duration of the PRM throughout the run, and the overall targeting selectivity through the number of active windows. To visualize the impact of lowering the fold-change trigger threshold on the scheduling of PRM windows, we calculated window activity upon the variation of the trigger threshold from 1.25-fold through 3-fold change in 0.25-fold increments within the MLN4924 experiment (Supporting Information Figure S7). As the trigger threshold is lowered, the number of active PRM windows is increased across the run,

leading to an increase in the cycle time and reduction in the selectivity of PRM window activity. By a 1.25-fold change trigger threshold, portions of the run are at or near saturation of active windows, which turns the PRM run into a DIA acquisition. In order to maintain effective focusing of the machine acquisition, and an acceptable acquisition cycle time, a sufficiently high fold-change trigger threshold must be imposed. Accordingly, the selectivity of the SWARM approach may be compromised at low fold-change thresholds, making the detection of lower fold-change analytes more difficult.

Conclusions:

We have explored the feasibility and utility of the inchoate implementation of the SWARM cycle, a novel approach to data acquisition for samples tagged by isobaric mass tags. This quantitation-first approach allows for acquisition behaviors to be predicated on the quantitative results of the SWARM scans. These experiments demonstrate the utility of the SWARM scan cycle as an effective way to bias acquisition toward higher fold change targets in subsequent PRM runs. By focusing acquisition, the SWARM methodology can impactfully reduce the burden of multiple hypothesis testing correction as we demonstrated using complex mixtures of ribosomes spiked into whole cell lysates (Figure 4) and in cultured cells treated with the small molecule MLN4924 (Figure 5). This increased capacity to detect statistically significant differences between the composition of protein samples is critical for maximizing the biological meaningful conclusions that can be gleaned from these datasets. Further, we expect that this benefit to multiple hypothesis testing correction will be particularly relevant to peptide-level quantitation where the number of hypotheses tested scales with the number of identified peptides.

A comparison of the current implementation of the DDA and SWARM+PRM approaches reveals that the role of DDA in providing deep proteome characterization remains firm. While SWARM+PRM provides a novel way of focusing machine acquisition toward analytes displaying user-definable quantitative characteristics, the approach may have limited utility when characterization of more modest protein fold-changes are of biological interest. Furthermore, DDA provides a more comprehensive set of quantitative values proteome-wide, while SWARM informed acquisitions narrowly focus inquiry toward a subset of the proteome. Currently, SWARM on the Orbitrap Lumos necessitates use of the lower resolution linear ion trap for reporter ion scanning in order to facilitate sufficiently small and numerous quadrupole isolation windows at a duty cycle compatible with modern nanoflow chromatographic separations of chemically complex mixtures. This mass analyzer choice effectively constrains the approach to isobaric tags with at least unit mass separations. We hope that newer generations of high-resolution mass analyzers will enable speeds and sensitivities capable of SWARM scanning higher multiplex isobaric tag kits with sub-integer mass separations.

While the current incarnation of our approach necessitates a second acquisition, and therefore highly reproducible chromatography, we envision that a major advantage of future implementations of SWARM will be the ability to incorporate real-time decision making into trigger scan events in order to enable a wide range of flexible acquisition strategies. The use of PRM runs forces quantitative sampling of analytes across the full chromatographic

peak, despite the low signal-to-noise ratio found at the peak edges. Initially, the capacity to trigger a data-dependent MS2 scan on any SWARM reporter scan with a fold-change exceeding a user-defined threshold which would remove the necessity to perform follow-up PRM acquisitions as is done in the current workflow. Additionally, algorithmic determination of fold-change trigger thresholds in response to empirical fold-change distributions observed during a run may help to properly focus machine acquisition. Longer term implementations could include real-time decisions based on thermal stability performed in the context of thermal proteome profiling experiments or changes in signaling kinetics measured in phosphopeptide profiling analyses. We expect the adaptability of SWARM to different proteomic methodologies will be a key asset in its future development.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

We thank Dr. Julian Whitelegge for discussions and advice in the earliest stages of this project. This research was supported by National Institutes of Health Grants GM089778 (J.A.W.), and GM112763 (J.A.W.). W.D.B. was supported by the Ruth L. Kirschstein National Research Service Award GM007185 from the National Institutes of Health.

References:

1. Thompson A; Schafer J; Kuhn K; Kienle S; Schwarz J; Schmidt G; Neumann T; Johnstone R; Mohammed AK; Hamon C, Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal Chem 2003, 75 (8), 1895–904. [PubMed: 12713048]
2. McAlister GC; Nusinow DP; Jedrychowski MP; Wuhr M; Huttlin EL; Erickson BK; Rad R; Haas W; Gygi SP, MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. Anal Chem 2014, 86 (14), 7150–8. [PubMed: 24927332]
3. Pascovici D; Handler DC; Wu JX; Haynes PA, Multiple testing corrections in quantitative proteomics: A useful but blunt tool. Proteomics 2016, 16 (18), 2448–53. [PubMed: 27461997]
4. MacLean B; Tomazela DM; Shulman N; Chambers M; Finney GL; Frewen B; Kern R; Tabb DL; Liebler DC; MacCoss MJ, Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics 2010, 26 (7), 966–8. [PubMed: 20147306]
5. Cox J; Mann M, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nature biotechnology 2008, 26 (12), 1367–72.
6. Cox J; Neuhauser N; Michalski A; Scheltema RA; Olsen JV; Mann M, Andromeda: a peptide search engine integrated into the MaxQuant environment. Journal of proteome research 2011, 10 (4), 1794–805. [PubMed: 21254760]
7. Trachsel C; Panse C; Kockmann T; Wolski WE; Grossmann J; Schlapbach R, rawDiag: An R Package Supporting Rational LC-MS Method Optimization for Bottom-up Proteomics. Journal of proteome research 2018, 17 (8), 2908–2914. [PubMed: 29978702]
8. Kessner D; Chambers M; Burke R; Agus D; Mallick P, ProteoWizard: open source software for rapid proteomics tools development. Bioinformatics 2008, 24 (21), 2534–6. [PubMed: 18606607]
9. Gatto L; Lilley KS, MSnbase-an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation. Bioinformatics 2012, 28 (2), 288–9. [PubMed: 22113085]

10. Liao H; Liu XJ; Blank JL; Bouck DC; Bernard H; Garcia K; Lightcap ES, Quantitative proteomic analysis of cellular protein modulation upon inhibition of the NEDD8-activating enzyme by MLN4924. Molecular & cellular proteomics : MCP 2011, 10 (11), M111 009183.

11. The UniProt, C., UniProt: the universal protein knowledgebase. Nucleic acids research 2017, 45 (D1), D158–D169. [PubMed: 27899622]

12. Schilling B; Rardin MJ; MacLean BX; Zawadzka AM; Frewen BE; Cusack MP; Sorensen DJ; Bereman MS; Jing E; Wu CC; Verdin E; Kahn CR; Maccoss MJ; Gibson BW, Platform-independent and label-free quantitation of proteomic data using MS1 extracted ion chromatograms in skyline: application to protein acetylation and phosphorylation. Molecular & cellular proteomics : MCP 2012, 11 (5), 202–14. [PubMed: 22454539]

13. UniProt Consortium, T., UniProt: the universal protein knowledgebase. Nucleic acids research 2018, 46 (5), 2699. [PubMed: 29425356]

14. Soucy TA; Smith PG; Milhollen MA; Berger AJ; Gavin JM; Adhikari S; Brownell JE; Burke KE; Cardin DP; Critchley S; Cullis CA; Doucette A; Garnsey JJ; Gaulin JL; Gershman RE; Lublinsky AR; McDonald A; Mizutani H; Narayanan U; Olhava EJ; Peluso S; Rezaei M; Sintchak MD; Talreja T; Thomas MP; Traore T; Vyskocil S; Weatherhead GS; Yu J; Zhang J; Dick LR; Claiborne CF; Rolfe M; Bolen JB; Langston SP, An inhibitor of NEDD8-activating enzyme as a new approach to treat cancer. Nature 2009, 458 (7239), 732–6. [PubMed: 19360080]

15. Aberle H; Bauer A; Stappert J; Kispert A; Kemler R, beta-catenin is a target for the ubiquitin-proteasome pathway. EMBO J 1997, 16 (13), 3797–804. [PubMed: 9233789]

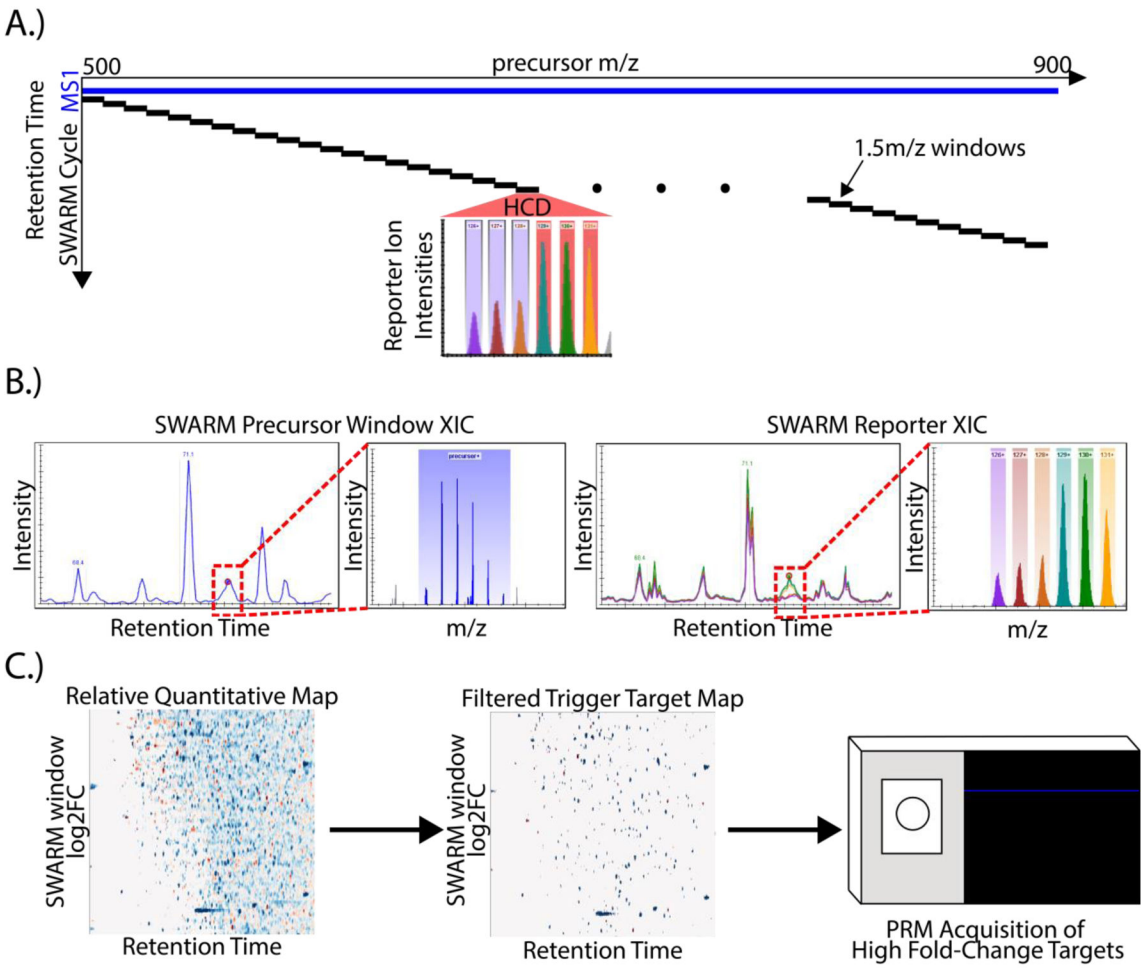

Figure 1.

A.) An overview schematic of the Sequential Windowed Acquisition of Reporter Masses (SWARM) scan cycle. This DIA-like cycle is comprised of small regularly spaced isolation windows in which HCD fragmentation is carried out and isobaric tagging reporter ions detected.

B.) An example SWARM window MS1 extracted ion chromatogram across retention time and the corresponding TMT reporter ion signals extracted from the MS2 linear ion trap scans. An example of a set of high fold change SWARM scans are highlighted.

C.) The general workflow for data acquisition under the SWARM+PRM paradigm. First, SWARM scanning is used to produce a relative fold change quantitation map of the sample. The map is filtered to retain only high fold change analytes and used to construct a set of PRM target *m/z* values and retention times. In a follow-up PRM acquisition, quantitation and identification scans are produced.

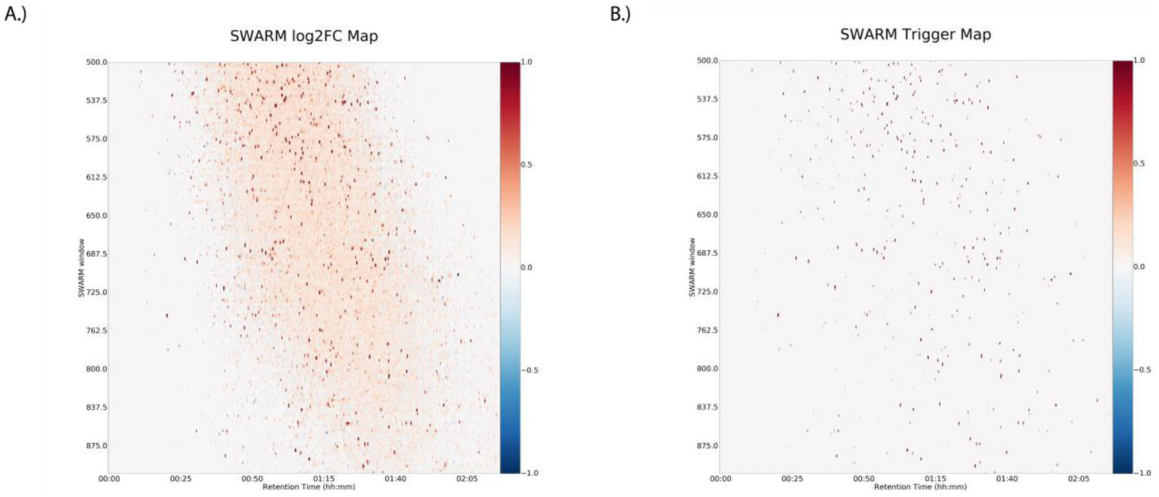

Figure 2.

A.) The SWARM quantitative map for the *E. coli* ribosome 5:1 sample shows the log2(FC) between the two conditions across chromatographic time for each SWARM window's *m/z*. B.) The quantitative map was filtered for a minimum 2-fold change to produce the trigger map. From these scan events, PRMs were scheduled for targeted reacquisition.

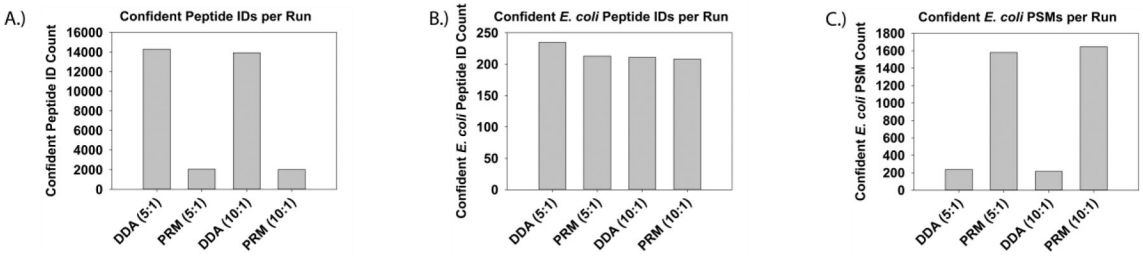

Figure 3.

A.) Peptide counts passing the 1% FDR filter for the *E. coli* ribosome spike-in experiments.

B.) Confident *E. coli* peptide counts passing the FDR filter.

C.) Confident *E. coli* PSM counts passing the FDR filter.

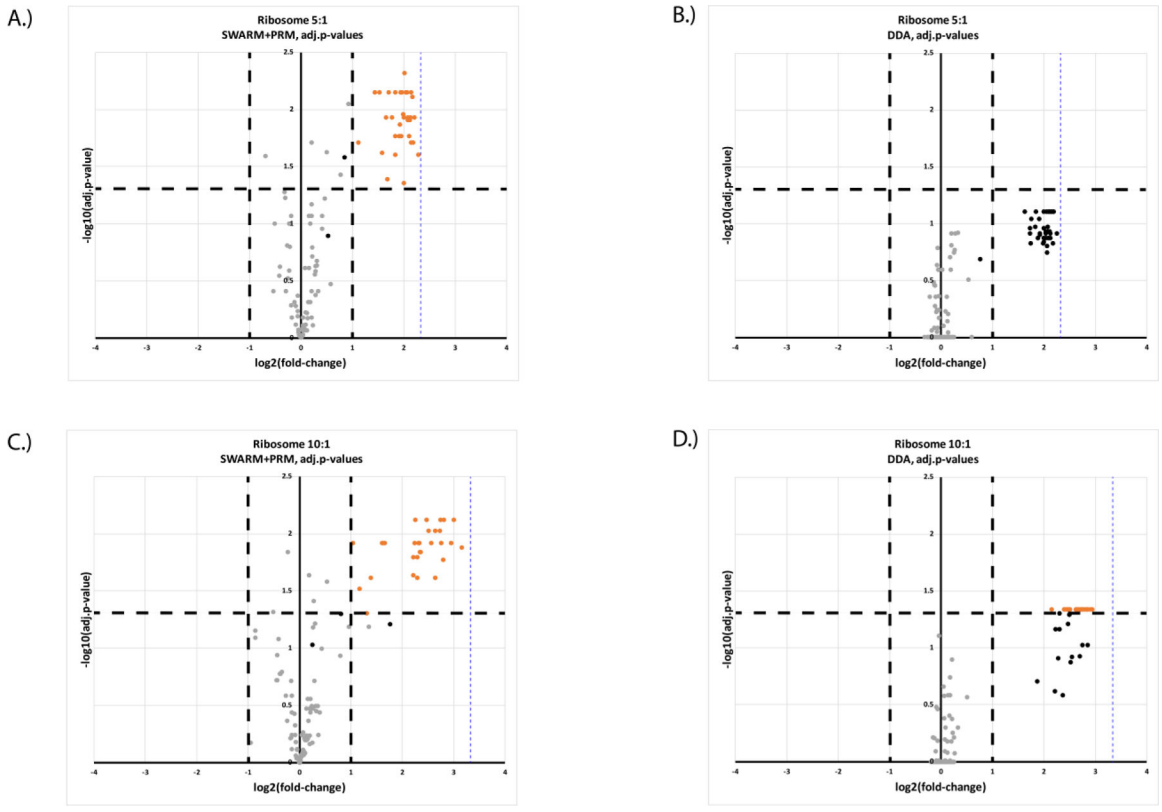

Figure 4.

A.) Volcano plot for the SWARM+PRM acquisition of the *E. coli* 5:1 spike-in sample. A total of 33 proteins were significantly differentially enriched (adj. p-value <=0.05) above a 2-fold change threshold in either direction. True positive, high fold change, significant *E. coli* proteins are shown in orange. False negative *E. coli* proteins are denoted in black, while true negative human proteins are gray. Vertical dashed line indicates true 5-fold change.

B.) Volcano plot for the DDA acquisition of the *E. coli* 5:1 sample. No proteins were deemed significantly differentially under our criteria. Vertical dashed line indicates true 5-fold change.

C.) Volcano plot for the SWARM+PRM acquisition of the *E. coli* 10:1 sample. 32 *E. coli* proteins were deemed to be significantly enriched in this analysis. Vertical dashed line indicates true 10-fold change.

D.) Volcano plot for the DDA acquisition of the *E. coli* 10:1 sample. 27 *E. coli* proteins were deemed to be significantly enriched. Vertical dashed line indicates true 10-fold change.

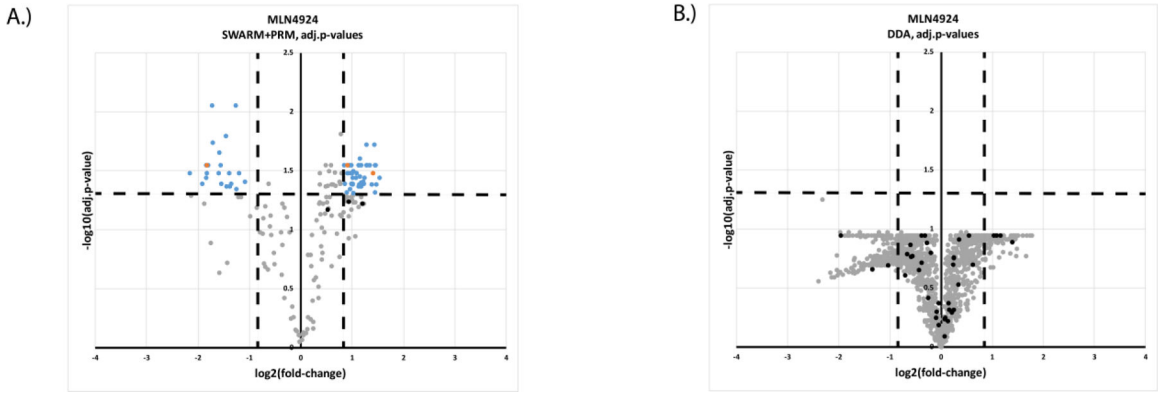

Figure 5.

A.) Volcano plot for the SWARM+PRM acquisition of the MLN4924 sample. 71 Proteins remained significant below an adjusted p-value of 0.05 and above 1.75-fold change. Three Cullin associated proteins which are enriched upon MLN4924 treatment are included in this population. High fold change, significant Cullin-associated proteins are shown in orange and significant non-associated proteins in blue. Non-significant Cullin-associated proteins are denoted in black, while non-significant non- associated proteins are gray.

B.) Volcano plot for the DDA acquisition of the MLN4924 sample. No proteins were deemed significantly differentially regulated under our criteria.