Application of SWATH Proteomics to Mouse Biology

Yibo Wu[1,†], **Evan G. Williams**[1,†], and **Ruedi Aebersold**[1,2]

[1]Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Zurich CH-8093, Switzerland [2]Faculty of Science, University of Zurich, Zurich CH-8057, Switzerland

Abstract

The quantitative measurement of the proteome has been shown to yield new insights into physiology and cell biology that cannot be determined from the genome and transcriptome, because the quantitative relationship between transcriptome and proteome is complex. MS-based proteomics techniques, such as SWATH-MS, have recently advanced to the extent that they may be reliably applied by biologists that are not specialists in mass spectrometry. We provide here a standard protocol for preparation of tissue samples for input into the SWATH-MS analytical pipeline. This protocol is designed for high throughput processing for tissues with 5 mg of sample available for analysis. Studies with extremely limited amounts of tissue should consider PCT-SWATH. An experienced single user for this protocol should be able to process 48 samples a day for injection into the mass spectrometer, or up to 144 samples a week. The machine time necessary for running these samples with SWATH is approximately 1.5 hours per sample.

Keywords

Proteomics; Systems Biology; Mass Spectrometry

INTRODUCTION

In this protocol, we describe the sample preparation methods used for isolating proteins from tissues or cells. The isolated proteins are then fragmented into short peptides with trypsin, cleaned of salts and other contaminants that may affect the mass spectrometer, and then prepared for injection into a SWATH-MS workflow. The exact same peptide preparation may be used with a range of mass spectrometric methods and strategies. These and their performance profiles have been reviewed (1, 2). This protocol is focused on the analysis of the thus-generated samples by SWATH-MS due to the high degree of reproducibility and quantitative accuracy this method provides across sample cohorts consisting of hundreds of samples. Before generating samples, it is important to check the SWATHAtlas (3) to see if reference SWATH-MS libraries are already available—i.e. that the libraries were generated from same organism and the same (or highly similar) tissue. If no suitable library is available, it is recommended to fractionate one sample from each

experimental group and run them in a data-dependent acquisition (DDA) mode ("discovery, aka shotgun"; Basic Protocol 3). Essentially, the complete SWATH-MS protocol follows the four protocols in order as defined here. Basic Protocol 1 describes the extraction of total protein from tissue samples. Basic Protocol 2 describes the digestion of the total protein into peptides. Basic Protocol 3 describes the fractionation of peptide samples, which will be run in the MS in discovery mode to generate the search space for SWATH-MS. Basic Protocol 4 describes briefly data-independent acquisition (DIA) in SWATH-MS mode.

Note that that Basic Protocol 3 (sample fractionation and library preparation) is not necessary for every sample—it is recommended to only perform this once per condition, and indeed it may be completely skipped if suitable reference libraries are available (e.g. libraries were previously generated in the same tissue or organism, ideally on the same MS as is planned for the SWATH-MS). SWATH-MS has the advantage of selecting and quantifying the same subset of proteins in all MS runs, even across dozens or hundreds of examples, in contrast to DDA where different, sets of peptides are measured each time. With practice and at least two 24-sample centrifuges, it should be feasible to prepare Basic Protocols 1 and 2 for 48 samples in a single batch (a 3 day process), or 144 samples in a standard week. The timing for Basic Protocol 3 is variable depending on how many conditions are used, and whether the library is available. For Basic Protocol 4, roughly 100 samples can be completed per week assuming a 1 hour gradient (corresponding to approximately a 1.5 hour run time). It is not necessary to continue directly from one protocol stage to the next—the samples may be safely frozen and maintained in −80°C freezers for extended periods for total protein measurements.

Basic Protocol 1. PROTEIN PREPARATION PROTOCOL

Take note that alternative protein extraction protocols are possible and may be desirable or even necessary for particular project designs. This protocol is designed to be simple and rapid for the extraction and measurement of whole protein levels from whole tissue in ample quantities (5 mg). Modifications must be made if, for instance, the reader wishes to analyze the phosphoproteome or other post-translational modifications.

Materials

Reagents

Tissue sample

We recommend using between 5 and 50 mg starting tissue. This should be sufficient for tissues with low protein content (e.g. white adipose), as little as ~2 μg of peptide mixture is sufficient for a single SWATH-MS run. The tissue amounts do not need to be closely controlled and weighed. However, if taking a smaller amount of a large tissue, e.g. 5 mg from a kidney, it is recommended to homogenize the tissue beforehand to ensure that the sample taken is representative of the full tissue. While 2 μg of final peptide may be used to to run SWATH-MS, this protocol may be difficult for preparing very small amounts of starting tissue, largely due to the difficulty of proper homogenization. For small quantities (< 5 mg) pressure cycling technology (PCT)-SWATH is recommended (4).

You can typically expect at minimum 5–10% protein yield if isolating total protein from full tissues (e.g. 30 mg tissue will yield ~2 mg pure protein), though this will vary substantially between tissues and between experimenters.

Water (bidistilled is sufficient)

RIPA-M buffer (see recipe), freshly prepared

Urea-T buffer (see recipe), freshly prepared

IGEPAL CA-630 (or Nonidet P-40)

Sodium Chloride (NaCl)

Ethylenediaminetetraacetic Acid (EDTA)

Tris(hydroxymethyl)aminomethane (Tris, THAM)

Urea

Protease Inhibitor Cocktail (e.g. Sigma P8340 or cOmplete Protease Inhibitor Cocktail Tablets)

Equipment

Homogenizer (dounce pestle recommended, but a metallic bead homogenizer can also work)

Refrigerated centrifuge (to 4°C; 24+ capacity for 2 mL tubes and 20,000 g rotation speed recommended)

Vortex

Sonicator (recommended)

−80°C Freezer (if long term storage is necessary)

Protocol steps: Prepare buffers fresh the day they are to be used. Two buffers are necessary, with a recommended amount of 2 mL RIPA-M per sample, and 1 mL of Urea-T per sample (see below under the section Reagents and Solutions).

Prepare tissue sample

As little as 20 μg of protein is sufficient for the second step of this protocol, so this protocol prepares more protein than is strictly necessary. If saving tissue quantities are extremely precious, please consider PCT-SWATH (4).

1. Use at least 1 mL of RIPA-M buffer for every 50 mg of wet tissue.

    This ratio is not critical (i.e. it is not necessary to precisely weigh the tissue or adjust volume to keep consistent across samples). For tissues with high protease activity, particularly pancreas, a higher ratio of RIPA to tissue may be desirable, but excess will not affect the sample (e.g. 1 mL RIPA-M for 15 mg of wet tissue is fine).

2. Use the tissue homogenizer to thoroughly homogenize the tissue in the RIPA-M buffer while keeping the sample around 0°C. The homogenization speed and duration will depend highly upon tissue—e.g. livers may be fully homogenized at low speed for 30 seconds, while skeletal muscle and heart may need 2 minutes at a higher speed. Experiment with your sample type and your homogenizer to determine a speed to where a visual assessment of the homogenate does not observe any large particles. Take care to keep consistent across all your samples for an experiment.

3. Transfer the homogenate to a new tube of suitable volume (e.g. 1.5 mL in the example volumes), or separate into multiple tubes if necessary. Centrifuge the samples for 15 minutes at 20,000 g at 4°C.

4. Transfer the supernatant to a new tube and place on ice. (Note that an additional ½ volume will be added to this tube in Step 6.)

5. Resuspend the pellet from step 3 in at least 500 µL UREA-T buffer per 50 mg of wet tissue. Vortex the sample briefly and then sonicate for 5 minutes (e.g. 40 kHz in an ice water bath). Again, the ratio is not critical and higher amounts can be used (e.g. 500 µL of UREA-T for 20 mg of wet tissue).

6. Centrifuge these resuspended pellets for 15 minutes at 20,000 g at 4°C.

7. Collect the supernatant and mix with the supernatant from Step 3. Discard the pellet.

8. Quantify the protein (e.g. by bicinchoninic acid assay (BCA) or a NanoDrop machine).

9. Unless the quantity is very low, aliquot the samples into at least two tubes and store at −80°C. One tube can remain frozen for future analyses (e.g. Western blots) and the other may be used for peptide preparation.

You may stop at this step or continue to the peptide digestion (Basic Protocol 2).

Basic Protocol 2. PEPTIDE PREPARATION PROTOCOL

In order to run shotgun and SWATH-MS (and indeed all "bottom-up proteomics" techniques), the total extracted proteins must be digested into shorter peptides. In this protocol, trypsin is used to digest the protein into short amino acid chains, cleaved at lysine and arginine residues. The samples are then purified from any impurities such as lipids or salts that could affect theLC separation or MS ionization process. Take care that all liquid reagents, such as the water supply or acetone, are sufficiently pure for MS ("HPLC-grade"). Bidistilled water is not likely to be sufficiently clean, as mass spectrometers are sensitive to contaminations, e.g., salts and detergents.

This protocol is approximately a three day process and the volumes described are suggested for 50 µg of input protein. The lower and upper bounds of protein here will be limited by the capacity of the C18 columns used, and the amount of trypsin. This same protocol may be

followed for 20–300 μg of input protein—simply scale the volumes suggested here as appropriate.

Materials

Reagents

Protein samples (e.g., from Basic Protocol 1)

Water (HPLC-grade)

Ammonium Bicarbonate (NH_4HCO_3)

Acetonitrile (ACN) (HPLC-grade)

Acetone (HPLC-grade), cooled to −20°C

Methanol (HPLC-grade)

Urea

Potassium Hydroxide (KOH)

Dithioethreitol (DTT)

Indole-3-Acetic Acid (IAA)

Trypsin (sequencing-grade)

Formic Acid (FA)

Indexed Retention Time peptides (iRT; from Biognosys)

Equipment

Centrifuge

−20°C Freezer

Heated Shake Plate (up to 37°C)

Silica C18 Columns (e.g. MacroSpin Column from The Nest Group)

Heated, Vacuum Drying Centrifuge

Several buffers can be prepared in advance and stored in sealed flasks for extended periods at room temperature (i.e., months). However, take care—ACN is volatile and will evaporate out of the mixtures, leading to decreasing ACN:H_2O ratios if a flask is opened and used many times. This does not preclude using the same bottle for several days of experiments, nor preparing the bottles far in advance, but it is something to keep in mind. For stocks, it is recommended to prepare in advance:

1. 0.1 M NH_4HCO_3 in H_2O

2. A high-concentration ACN:H_2O solution (8:2) + 0.1% FA

3. A medium-concentration ACN:H_2O solution (5:5) + 0.1% FA

4. A low-concentration ACN:H_2O solution (2:98) + 0.1% FA

5. 1% FA solution in H_2O (1:99).

6. 0.1% FA solution in H_2O (1:999)

Day 1

1 Thaw samples on ice, then transfer 50 μg of protein to a new tube.

2 Add 6 volumes of cold HPLC-grade acetone (−20°C) to each sample to precipitate the protein. The next steps will be simpler if the acetone is < 1.0 mL in volume.

3 Leave the samples in a −20°C freezer and wait for a few hours (e.g. 4–24 hours; keep consistent within a study).

Day 2

4 Prepare three fresh reagents: (1) 8 M urea + 0.1 M NH_4HCO_3; (2) 36 mM DTT; (3) 160 mM IAA.

> IAA is light sensitive and should be kept and prepared in a low-light setting at all times and/or wrapped in aluminum foil.

5 Warm a shaking plate to 37°C.

6 Centrifuge the samples (from step 3) at 20,000 g for 10 minutes. Proteins should be well-fixed to the bottom of the tube. Remove acetone supernatant.

7 Add 90 μL of 8 M urea buffer (from step 4) for every 50 μg of protein and resuspend samples with a quick vortex.

8 Add 45 μL of freshly prepared 36 mM DTT buffer for every 50 μg of protein.

9 Vortex briefly, then incubate samples on the 37°C shaking plate for 30 minutes at 600 rpm. Remove samples and cool shaking plate to 25°C.

10 Reduce light in the room as much as possible, then add 45 μL of 160 mM IAA for every 50 μg of protein.

11 Vortex briefly, then incubate samples on the 25°C shaking plate for 45 minutes. Make sure that the samples are covered from light during this time (e.g. with aluminum foil).

12 Dilute samples with 0.1 M NH_4HCO_3 to a final urea concentration of 1.5 M (e.g. 300 μL for every 50 μg of protein). Samples can now be exposed to light.

13 Add sequencing-grade trypsin to the sample (at least 1 μg per 50 μg protein).

14 Warm shaking plate to 37°C, then place samples here for 16–24 hours at 1000 rpm. Take particular care at this step to keep trypsin digestion times consistent for all digestions of a particular study. Avoid going beyond 24 hours, as trypsin will start to self-digest, which can create large peaks on mass spectrometry runs, obscuring the desired data. Conversely, digestion times of less than 16 hours may not be complete.

Day 3

15 Activate the Silica C18 Columns with 450 μL of HPLC-grade methanol. (Note: double-check this with the protocol that comes with your C18 provider!)

16 Centrifuge for 3 minutes at 1000 g. Discard methanol.

17 Again, add 450 μL of HPLC-grade methanol.

18 Again, centrifuge for 3 minutes at 1000 g. Discard methanol.

19 Wash with 450 μL of ACN:H_2O **8:2** + 0.1% FA.

20 Centrifuge for 3 minutes at 1000 g. Discard flow through.

21 Again, wash with 450 μL of ACN:H_2O **8:2** + 0.1% FA.

22 Again, centrifuge for 3 minutes at 1000 g. Discard flow through.

23 Prepare with 450 μL of ACN:H_2O **2:98** + 0.1% FA.

24 Centrifuge for 3 minutes at 1000 g. Discard flow through.

25 Take samples from the shaking plate (step 14). It is recommended to add 0.1% FA to the sample (i.e. add 53 μL of 1% FA to the sample volume of 480 μL) and vortex, and check with pH paper to ensure the pH is acidic. FA is necessary to ensure the homogenous charge state of the peptide species and therefore consistent behavior upon binding to the C18 during cleanup and separation on the LC.

26 Centrifuge at 20,000 g for 10 minutes. There should be no precipitate at the bottom. If there is, be careful to not transfer it into the C18 column in step 27.

27 Take 450 μL from the digested peptide samples and load onto the C18 Columns. This will result in a loss of 15% of the peptide quantity (i.e. due to total volume of 530 μL). If maximum recovery is necessary, the entire sample may be loaded. Be careful not to bring contaminants which were centrifuged to the bottom of the tube.

28 Centrifuge at 1000 g for 3 minutes.

29 Reload the outflow onto the column.

30 Again, centrifuge for 3 minutes at 1000 g. The peptides should be trapped in the column. Discard flow through.

31 Wash columns with **0.1% FA** (1:999 FA:H_2O).

32 Centrifuge for 3 minutes at 1000 g. Discard flow through.

33 Discard the old collection tube, add a new (and final) collection tube.

34 Add 450 μL of ACN:H_2O **5:5** + 0.1% FA to elute the sample.

35 Centrifuge for 3 minutes at 1000 g. Reload flow through onto the column.

36 Again, centrifuge for 3 minutes at 1000 g. Discard <u>column</u>.

37 Dry samples in a vacuum centrifuge. A warmed vacuum centrifuge to 37–45°C will expedite this process.

> *If you do not plan on running your samples in the mass spectrometer immediately, stop at this step after the samples are dried, and freeze them at −80°C.*

38 On the day you expect to start the mass spectrometry runs, resuspend the dried samples with ACN:H$_2$O **2:98** + 0.1% FA to a target concentration of around 250–1000 ng/µL. (The peptide quantity at the end will probably be 25% to 75% the input protein quantity.)

39 Vortex and sonicate to resuspend the sample fully.

40 Centrifuge the samples at high speed (e.g. 20,000 g) for 10 minutes to pellet any contaminants that may remain.

41 Quantify the peptide concentrations on a NanoDrop spectrophotometer.

42 Transfer some of each sample to the mass spectrometer sample tubes—the best quantity and concentration depends upon which mass spectrometer will be used. For SWATH-MS on a SciEx TripleTOF 5600, it is recommended to load 5 to 10 µg at a concentration of 250–1000 ng/µL in 19 µL of final volume. It is recommended to transfer approximately even quantities across all samples, then dilute all samples to the same final concentration with ACN:H$_2$O **2:98** + 0.1% FA. However, do not dilute below 250 ng/µL. The quantification data will be normalized afterwards, but it is better to begin with as similar loadings as possible.

43 If possible, run a few samples on a less sensitive mass spectrometer to ensure general protein quality and to check for any contaminations that would block the machine for the SWATH mass spectrometry run.

44 Add 1 µL of the indexed retention time (iRT) peptides per 20 µL loaded in step 42 (i.e. 100 femtomoles of iRT peptides). This allows for correction across samples for small shifts in the retention time measured.

45 Samples are now ready for injection in the mass spectrometer in either shotgun mode (for generating the library; additional fractionations are recommended) or SWATH mode (for quantifying the peptides; no fractionations are necessary). Within the range certain machine allows, inject as much peptide as possible for the machine to ensure that sufficient quantities of lowly expressed peptides can be measured. Low amounts (e.g. 100 ng) can be run, but fewer proteins will be quantified. Note that high amounts (e.g. > 2 µg) may cause problems with certain machines.

Basic Protocol 3. DATA-DEPENDENT ACQUISTION AND SPECTRAL LIBRARY GENERATION PROTOCOL

Data-independent acquisition (DIA) approaches, such as SWATH-MS, have the benefit of quantifying the same subsets of proteins across all samples regardless of the study size. However, to achieve this desired consistency, SWATH-MS uses a targeted "peptide-centric" data analysis strategy that relies on spectral reference libraries as prior information for peptide identification (5). Therefore, to use the SWATH-MS method effectively, a peptide library must be developed first. The content of the library defines the search space for the subsequent SWATH-MS measurements. Essentially, this library is a list of query parameters (precursor m/z, fragment m/z and retention time, etc) of the peptides identified by the DDA ("discovery") method, which can be taken as a reference. Libraries can be made for each independent experiment, but it is not necessary—so long as a library has been previously generated in the same organism and the same (or similar) tissues. Quality and coverage of these libraries are crucial for the performance of SWATH-MS. If suitable libraries are not available from SWATHAtlas (3), it may be necessary to generate one. We have published a detailed protocol for library generation (6). Here, we summarize the key steps and typical MS settings.

To increase the proteome coverage for the DDA-generated library, samples should be fractionated e.g. by using isoelectric focusing by off-gel electrophoresis (OGE) or SDS-PAGE. Here, we show a protocol using OGE. The resulting fractions are then used to acquire high-quality fragment ion spectra in DDA mode, which is preferably obtained on the same type of instrument, which will be used for the SWATH-MS, e.g. a TripleTOF 5600+ mass spectrometer. However, an instrument with a beam-type collision cell or ion trap-type collision cell that functions in higher-energy collisional dissociation (HCD) mode can also be used because these instruments generate similar fragment ion spectra from the same peptides (7). It is important to note that such extensive spectral libraries require more stringent false discovery rate (FDR) control in the SWATH-MS data analysis, and it is not guaranteed that all peptides in these libraries can be detected in the SWATH-MS analysis. Conversely, a peptide that is not identified in the library cannot be identified in subsequent DIA runs. It is also recommended to add indexed retention time reference peptides (iRT) into all samples that are used for library generation, as this allows effective peptide retention time normalization (8).

If a new library needs to be generated, it is not necessary to fractionate and run DDA on every single sample—at most, only one sample per condition is necessary. In situations with many similar conditions even this may be unnecessarily redundant. Consider the following: if examining the proteome of a wildtype versus knockout study, generating the library exclusively in the knockout means that the knocked out protein will not be detected in *any* SWATH runs, as it will not be in the library. Conversely, a separate protein that is only expressed in the samples of the knockout mouse will not be detected if the library is generated exclusively with the control condition. This concern should be weighed against the increase in time necessary to generate libraries from each condition. Beyond this protocol, there are several other sources describing the generation of tissue libraries (9–11),

although each protocol has slight differences. The following protocol is designed to be relatively quick.

Reagents

Peptide mixture (e.g., from Basic Protocol 2)

Buffer A: 2% ACN, 0.1% FA solution in H_2O

Buffer B: 98% ACN solution with 0.1% FA

Equipment

AB Sciex 5600 TripleTOF mass spectrometer

Eksigent NanoLC Ultra 2D Plus HPLC (interfaced to the above)

1. For off-gel electrophoresis (OGE) fractionation, use 1–2 mg of peptide mixtures digested as described above. The peptide mixtures are the pool of digested peptides of protein extract from different conditions.

2. Load the peptide mixtures for OGE fractionation as previously described (12). Briefly, separate the peptide mixtures using a pH 3–10 IPG strip (Amersham Biosciences) and a 3100 OFFGEL Fractionator (Agilent Technologies) with collection in 24 wells.

3. Combine the 24 fractions into 10 fractions and purify with C18 columns. Evaporate all peptide samples to dryness and resolubilize in buffer A for MS analysis, exactly as in Day 3 of the preceding protocol (Protocol 2).

4. Analyze each fraction or each sample in a DDA/shotgun mode with the mass spectrometer you plan to use for SWATH (e.g. AB Sciex TripleTOF 5600), which should be interfaced to an HPLC (e.g. an Eksigent NanoLC Ultra 2D Plus) (6).

5. Load sample onto the C18 (Magic, 3 μm) packed (10 to 15 cm length of packing) emitter coupled with an analytical column (e.g. PicoFrit with a 75 μm diameter) with buffer A. The sample can be eluted gradually over 135 minutes with a variable linear gradient of 2% to 35% of buffer B, at a flow rate of 300 nL/min.

6. Set the standard DDA instrument parameters to select the 20 most intense precursors with charge states +2 to +5 for fragmentation. Acquire the MS2 spectra in the range of 50 to 2000 m/z. We use conservative MS2 accumulation time (150 msec) to get high quality spectra. Exclude precursor ions from reselection for 15 sec.

7. Transform the DDA files to mzXML files using ProteoWizard (13). This puts the data into an open, non-proprietary format. It is optimal to use fragment ion peak areas instead of peak height for centroiding.

8. Search the mzXML files against the canonical UniProt proteome database for the particular organism using database search engines (Comet, Sequest, Mascot,

Tandem, etc), and integrate the search results using the Trans-Proteome Pipeline (TPP) (14).

9. Set cysteine carboxymethylation as static modification, and methionine oxidation as variable modification. Others modifications, e.g. phosphorylation, can be set as required though this may require modifications to the protein preparation protocol to retain such modifications.

10. To control FDR of the peptide-spectrum matches (PSMs), generate protein sequence reversals or pseudo-reversals and append to the target database (15). The generation of decoy peptides is typically performed within the search engines.

11. Allow peptides with up to one missed cleavage site. It is possible to adjust later on and remove peptides with missed cleavages. It is recommended (though not necessary) to leave these alternative possibilities in with the library, and they can be removed at the end if necessary.

12. Set mass tolerance for precursor and fragment ions. Typically for an AB Sciex 5600 TripleTOF mass spectrometer, set mass tolerance to 25 parts per million (ppm) for precursor ions and 0.4 Da for fragmental ions.

13. Filter out fragments that are smaller than 350 m/z or bigger than 2,000 m/z.

14. Filter out fragments with m/z in the precursor SWATH window.

15. Combine the pepXML files using iProphet (16), and use the integrated pepXML file to generate the redundant spectral library containing all PSMs using SpectraST (9). It is important to estimate FDR at PSM, peptide and proteins levels using the MAYU software (6, 17). For large library generation, it is recommended to use 1% protein FDR.

16. Construct the consensus library using SpectraST. Retention time of peptides are aligned to reference values, e.g. iRT values.

17. Select the top 5 most abundant b and y fragment ions of each peptide to generate the assays for SWATH/MS targeted extraction. Please note that for library containing assays for C-terminally heavy isotope-labeled peptides, only y ions are included in the library. Decoy assays are appended to the target assay library for FDR estimation.

Basic Protocol 4. DATA-INDEPENDENT ACQUSITION AND TARGETED "peptide-centric" DATA EXTRACTION PROTOCOL

Discovery (DDA) proteomics achieves high proteome coverage if complex samples are fractionated and is still the most commonly used proteomics technique. However, the identification and quantification of peptides from DDA are biased toward proteins with higher abundance in the sample, and it suffers from inherently poor reproducibility when large number of samples are analyzed, as the same proteins are not necessarily quantified in each run—thus the overlap diminishes as more samples are analyzed. This is particularly the

case for complex, unfractionated samples. Targeted proteomics methods, such selected reaction monitoring (SRM), have been developed to increase the sensitivity and reproducibility of proteome measurement, but the comparatively low throughput of these methods (typically up to ~100 proteins per run (18)) limits its application in studies where broad subsets of the proteome need to be quantified. As a next generation quantitative proteomics technique, SWATH-MS has demonstrated substantial advantages in scope compared to SRM, and reliability compared to shotgun—essentially providing a middle way between these two techniques. Here we describe the general steps and settings of SWATH-MS measurement on a TripleTOF 5600 mass spectrometer.

Reagents

Samples

Buffer A: 2% ACN, 0.1% FA solution in H_2O

Buffer B: 98% ACN solution with 0.1% FA

UPS2 Proteomic Standard (optional, for batch effect control or approximation of absolute quantities; not suitable for human samples—for human samples, a non-human-protein derived control is necessary; see "Critical Parameters" in the next section).

Equipment

AB Sciex 5600 TripleTOF mass spectrometer

Eksigent NanoLC Ultra 2D Plus HPLC (interfaced to the above)

Protocol steps

1. Measure the samples on the MS in a randomized sequence to minimize the potential measurement biases coming along time during the runs, e.g. retention times may shift on the chromatography, or mass accuracy, instrument sensitivity may change slightly over time on the MS. If multiple batches are anticipated for the study, it is recommended to add a standard protein control to each sample. For instance, for non-human samples, the UPS2 Proteomic Standard from Sigma Aldrich may be used (the spike-in proteins must be not present in the samples naturally). Note that if a batch control is used, the proteins from the batch must be included in the proteomic library from Basic Protocol 3.

2. Load samples into the MS with buffer A, and elute from the column over 60 minutes using a continuously variable gradient of 2% to 35% of buffer B. To assess the quantitative reproducibility, it is recommended to include technical replicate injections of a representative pooled sample.

3. For SWATH-MS measurement, operate the MS in a looped ion product mode.

4. Construct a set of 64 overlapping windows covering the 400 to 1200 m/z precursor range. Variable windows can be set based on the number of precursor

within m/z regions using the Variable Window Acquisition feature in Analyst TF Software 1.7.

5. Collect the SWATH MS2 spectra from 100 to 2,000 m/z.

6. Set the collision energy according to the calculation for ions of +2 charge centered upon the window with a spread of 15 eV.

7. Use an accumulation time (dwell time) of 50 msec for fragment ion scans in high-sensitivity mode, and 150 ms for survey scans in high-resolution mode acquired at the beginning of each cycle, resulting in a duty time of ~3.4 sec. Please not this is not absolutely necessary, but some alternative analysis tools, e.g. DIAUmpire, require high quality MS1 spectra as well.

8. After data acquisition, convert the SWATH-MS.wiff files to .mzXML files using ProteoWizard (13).

9. Perform SWATH-MS targeted data extraction using OpenSWATH workflow (19). OpenSWATH applies a target-decoy scoring model to estimate the FDR using the mProphet algorithm (20, 21). Please note the below steps refer to using OpenSWATH in iPortal. The command line of OpenSWATH is OpenSwathWorkflow. Please refer to Röst et al. to run OpenSWATH (21).

10. Select the transformed mzXML files to be analyzed.

11. Select the generated spectral library (from Basic Protocol 3, or downloaded from SWATHAtlas).

12. Set the retention time window. By default 300 sec.

13. Select retention time alignment method in OpenSWATH, e.g. iRT realignment (8) or TRIC (22).

14. Extract fragment ion chromatograms according to the target-decoy assay library with a width of 0.05 m/z.

15. Set peptide FDR to 0.01. In OpenSWATH, peak groups are scored based on the elution profile of the fragment ions, similarity of elution time and relative intensities with the assay libraries, and the features of MS2 spectra extracted at the chromatographic peak apex. Peptide FDR is estimated according to the score distribution of target and decoy assays (2). FDR of 0.01 is permissive, so retain the FDR for each peptide in later spreadsheets—it may be desirable to remove peptides that only just cleared this threshold.

16. Start OpenSWATH analysis. PyProphet statistical models will be generated, and a data matrix will be output containing the intensities and quality scores for all peptides quantified, along with a list of which protein(s) the peptides correspond to. Cases of non-proteotypic peptides (e.g. when a peptide sequence can correspond to more than one protein) can be discard or not.

> *There are several protocols for compressing each peptide back into a single protein (23). For instance, the peptide with the topN highest*

average intensity is taken as the best measurement of the protein level. Alternately, principle component analysis may be performed on all peptides corresponding to a particular protein (24, 25), and the first principle component may be taken as the best approximation of the overall protein level. Further research will be necessary to determine a standard ideal technique for the final analysis.

REAGENTS AND SOLUTIONS

1   RIPA-M

1% NP-40

150 mM NaCl

1 mM EDTA

50 mM Tris

Protease Inhibitor Cocktail (amount to use is based on tissue quantity used)

Adjust to final pH of 7.5

2   Urea-T

50 mM Tris

75 mM NaCl

8 M Urea

Adjust to final pH of 8.1

COMMENTARY

Background Information

SWATH-MS is an emerging technology for next generation proteomics (26). Classical shotgun methods with isotopic or chemical labeling suffer from low data completeness and low reproducibility for protein quantification, when large number of samples are measured. Targeted proteomics strategy, e.g. selected reaction monitoring (SRM), allows sensitive and precise protein quantification for limited number of proteins per measurement. As an emerging technology, SWATH-MS permits the rapid and consistent quantification of thousands of proteins across large sample cohorts (27), and it does not require any isotopic or chemical labeling of the input samples or tissues. This four-stage protocol is designed for the rapid preparation and analysis of whole protein from tissues.

Critical Parameters

Take care to avoid contamination for the MS runs, which can occur either from using impure reagents or from insufficiently purified samples (e.g. in the C18 cleaning step). If running samples across several distinct batches for a continuous study, it is critical to both inject one or two identical samples in every batch of the experiment, and it is also recommended to include a loading control protein for every sample. For instance, a specific quantity of

bovine serum albumin may be included in a mouse or human sample, providing a known quantity of an exogenous protein that may be referenced against in every run. For the best results, it may be necessary to include an entire set of controls and standard curve for concentrations that will provide a known, consistent, and synthetic reference which can be compared across all samples and all batches. For instance, you may include the UPS2 Proteomic Standard from Sigma Aldrich into every sample—a known quantity of 48 proteins that should be standard across all runs. Take note that this standard will not work for human samples, as the UPS2 is human derived, and for other organisms, take note that certain of the digested product-peptides may be identical across species.

Troubleshooting

When performing quality check on the MS for your samples, there should be no striking, solitary peaks in the total ion chromatography (TIC). If there are, they should be investigated and will likely be a product of contamination. Such contaminations may or may not affect your sample run, and they may be difficult to trace down and uncover. If the final data yields expected positive control results, then troubleshooting may not be necessary.

Anticipated Results

The number of proteins quantified in a SWATH measurement is highly dependent on several parameters, including the tissue type and the analysis instrument. Due to the current dynamic range of instrument, a LC-MS analysis of blood plasma or urine will likely yield no more than a few hundred proteins, while analysis of complex tissues or cell lines may yield upwards of 20,000 peptides, which correspond to 4,000 unique proteins. In general, the larger and more diverse the sample set, the more peptides (and proteins) will be measured. The number of peptides identified per protein will vary widely depending on both the length of the protein and how many proteotypic peptides it generates. Proteins in families with high sequence similarity, e.g. olfactory receptors, may be difficult to properly detect and differentiate.

Time Considerations

Protein preparation takes approximately one day. Sixty samples can be easily prepared by an experienced technician, provided that both a motorized tissue homogenizer and two 30-sample-capacity refrigerated centrifuges are available.

Peptide preparation takes two or three days depending on how the user follows the protocol. Forty-eight samples per batch can be easily prepared by an experienced technician, and there is sufficient time for a user to stagger preparation, e.g. prepare 48 samples on "Day 1, Day 2, Day 3" and a further 48 samples on "Day 2, Day 3, Day 4". If many dozen samples will be prepared, the peptide digestion steps can be expedited by the use of a multichannel pipette and 96-well C18 plates (e.g. MiniSpin™ or MACROSpin provided by The Nest Group). This will require some minor modification of the volumes indicated in the protocol, but no fundamental differences—50 µg of peptides can be cleaned up through such 96-well plates.

The time for sample quality control and mass spectrometry runs depends largely on the usage of the mass spectrometry facility and machine settings, with waiting for available time

on the machines typically the longest waiting step. Each sample run—for quality control, for shotgun, and for SWATH—typically takes 60–90 minutes, but these values are adjustable and there is no single solution.

Acknowledgments

LITERATURE CITED

1. Aebersold R, Mann M. Mass-spectrometric exploration of proteome structure and function. Nature. 2016; 537:347–355. [PubMed: 27629641]

2. Gillet LC, Leitner A, Aebersold R. Mass Spectrometry Applied to Bottom-Up Proteomics: Entering the High-Throughput Era for Hypothesis Testing. Annu Rev Anal Chem (Palo Alto Calif). 2016; 9:449–472. [PubMed: 27049628]

3. Rosenberger G, Koh CC, Guo T, Rost HL, Kouvonen P, Collins BC, Heusel M, Liu Y, Caron E, Vichalkovski A, Faini M, Schubert OT, Faridi P, Ebhardt HA, Matondo M, Lam H, Bader SL, Campbell DS, Deutsch EW, Moritz RL, Tate S, Aebersold R. A repository of assays to quantify 10,000 human proteins by SWATH-MS. Sci Data. 2014; 1:140031. [PubMed: 25977788]

4. Shao S, Guo T, Koh CC, Gillessen S, Joerger M, Jochum W, Aebersold R. Minimal sample requirement for highly multiplexed protein quantification in cell lines and tissues by PCT-SWATH mass spectrometry. Proteomics. 2015; 15:3711–3721. [PubMed: 26287124]

5. Ting YS, Egertson JD, Payne SH, Kim S, MacLean B, Kall L, Aebersold R, Smith RD, Noble WS, MacCoss MJ. Peptide-Centric Proteome Analysis: An Alternative Strategy for the Analysis of Tandem Mass Spectrometry Data. Molecular & cellular proteomics : MCP. 2015; 14:2301–2307. [PubMed: 26217018]

6. Schubert OT, Gillet LC, Collins BC, Navarro P, Rosenberger G, Wolski WE, Lam H, Amodei D, Mallick P, MacLean B, Aebersold R. Building high-quality assay libraries for targeted analysis of SWATH MS data. Nat Protoc. 2015; 10:426–441. [PubMed: 25675208]

7. Toprak UH, Gillet LC, Maiolica A, Navarro P, Leitner A, Aebersold R. Conserved peptide fragmentation as a benchmarking tool for mass spectrometers and a discriminating feature for targeted proteomics. Molecular & cellular proteomics : MCP. 2014; 13:2056–2071. [PubMed: 24623587]

8. Escher C, Reiter L, MacLean B, Ossola R, Herzog F, Chilton J, MacCoss MJ, Rinner O. Using iRT, a normalized retention time for more targeted measurement of peptides. Proteomics. 2012; 12:1111–1121. [PubMed: 22577012]

9. Lam H, Deutsch EW, Eddes JS, Eng JK, King N, Stein SE, Aebersold R. Development and validation of a spectral library searching method for peptide identification from MS/MS. Proteomics. 2007; 7:655–667. [PubMed: 17295354]

10. Craig R, Cortens JC, Fenyo D, Beavis RC. Using annotated peptide mass spectrum libraries for protein identification. Journal of proteome research. 2006; 5:1843–1849. [PubMed: 16889405]

11. Frewen BE, Merrihew GE, Wu CC, Noble WS, MacCoss MJ. Analysis of peptide MS/MS spectra from large-scale proteomics experiments using spectrum libraries. Analytical chemistry. 2006; 78:5678–5684. [PubMed: 16906711]

12. Picotti P, Bodenmiller B, Mueller LN, Domon B, Aebersold R. Full dynamic range proteome analysis of S. cerevisiae by targeted proteomics. Cell. 2009; 138:795–806. [PubMed: 19664813]

13. Kessner D, Chambers M, Burke R, Agus D, Mallick P. ProteoWizard: open source software for rapid proteomics tools development. Bioinformatics. 2008; 24:2534–2536. [PubMed: 18606607]

14. Keller A, Eng J, Zhang N, Li XJ, Aebersold R. A uniform proteomics MS/MS analysis platform utilizing open XML file formats. Molecular systems biology. 2005; 1:2005 0017.

15. Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nature methods. 2007; 4:207–214. [PubMed: 17327847]

16. Shteynberg D, Deutsch EW, Lam H, Eng JK, Sun Z, Tasman N, Mendoza L, Moritz RL, Aebersold R, Nesvizhskii AI. iProphet: multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. Molecular & cellular proteomics : MCP. 2011; 10:M111 007690.

17. Reiter L, Claassen M, Schrimpf SP, Jovanovic M, Schmidt A, Buhmann JM, Hengartner MO, Aebersold R. Protein identification false discovery rates for very large proteomics data sets generated by tandem mass spectrometry. Molecular & cellular proteomics : MCP. 2009; 8:2405–2417. [PubMed: 19608599]

18. Wu Y, Williams EG, Dubuis S, Mottis A, Jovaisaite V, Houten SM, Argmann CA, Faridi P, Wolski W, Kutalik Z, Zamboni N, Auwerx J, Aebersold R. Multilayered genetics and omics dissection of mitochondrial activity in a mouse reference population. Cell. 2014:158. [PubMed: 24361105]

19. Rost HL, Rosenberger G, Navarro P, Gillet L, Miladinovic SM, Schubert OT, Wolski W, Collins BC, Malmstrom J, Malmstrom L, Aebersold R. OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. Nature biotechnology. 2014; 32:219–223.

20. Reiter L, Rinner O, Picotti P, Huttenhain R, Beck M, Brusniak MY, Hengartner MO, Aebersold R. mProphet: automated data processing and statistical validation for large-scale SRM experiments. Nature methods. 2011; 8:430–435. [PubMed: 21423193]

21. Röst AR, Hannes L, Olga Schubert T. Automated SWATH Data Analysis Using Targeted Extraction of Ion Chromatograms. bioRxiv. 2016

22. Rost HL, Liu Y, D'Agostino G, Zanella M, Navarro P, Rosenberger G, Collins BC, Gillet L, Testa G, Malmstrom L, Aebersold R. TRIC: an automated alignment strategy for reproducible protein quantification in targeted proteomics. Nat Methods. 2016; 13:777–783. [PubMed: 27479329]

23. Rosenberger G, Ludwig C, Rost HL, Aebersold R, Malmstrom L. aLFQ: an R-package for estimating absolute protein quantities from label-free LC-MS/MS proteomics data. Bioinformatics. 2014; 30:2511–2513. [PubMed: 24753486]

24. Choi M, Chang CY, Clough T, Broudy D, Killeen T, MacLean B, Vitek O. MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. Bioinformatics. 2014; 30:2524–2526. [PubMed: 24794931]

25. Teo G, Kim S, Tsou CC, Collins B, Gingras AC, Nesvizhskii AI, Choi H. mapDIA: Preprocessing and statistical analysis of quantitative proteomics data from data independent acquisition mass spectrometry. J Proteomics. 2015; 129:108–120. [PubMed: 26381204]

26. Gillet LC, Navarro P, Tate S, Rost H, Selevsek N, Reiter L, Bonner R, Aebersold R. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Molecular & cellular proteomics : MCP. 2012; 11:O111 016717.

27. Williams EG, Wu Y, Jha P, Dubuis S, Blattmann P, Argmann CA, Houten SM, Amariuta T, Wolski W, Zamboni N, Aebersold R, Auwerx J. Systems proteomics of liver mitochondria function. Science. 2016; 352:aad0189. [PubMed: 27284200]