

# Application of SWATH Proteomics to Mouse Biology

Yibo Wu,<sup>1,3</sup> Evan G. Williams,<sup>1,3</sup> and Ruedi Aebersold<sup>1,2</sup>

<sup>1</sup>Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Zurich, Switzerland

<sup>2</sup>Faculty of Science, University of Zurich, Zurich, Switzerland

<sup>3</sup>These authors contributed equally to this work.

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 point at which they may be reliably applied by biologists who are not specialists in mass spectrometry. Here we provide standard protocols for preparation of tissue samples for input into the SWATH-MS analytical pipeline. These protocols are designed for high-throughput processing of tissues with  $\geq 5$  mg of sample available for analysis. Studies with extremely limited amounts of tissue should consider PCT-SWATH. An experienced single user should be able to process 48 samples per 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 hr per sample. Data acquisition protocols are also provided. © 2017 by John Wiley & Sons, Inc.

Keywords: proteomics • systems biology • mass spectrometry

## How to cite this article:

Wu, Y., Williams, E. G., & Aebersold, R. (2017). Application of SWATH proteomics to mouse biology. *Current Protocols in Mouse Biology*, 7, 130–143. doi: 10.1002/cpmo.28

## INTRODUCTION

In this article, we describe 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 strategies and their performance profiles have been reviewed (Aebersold & Mann, 2016; Gillet, Leitner, & Aebersold, 2016). This article is focused on the analysis of samples generated according to such preparation strategies by SWATH-MS. This method provides a high degree of reproducibility and quantitative accuracy across cohorts consisting of hundreds of samples. Before generating samples, it is important to check the SWATHAtlas (Rosenberger et al., 2014a) to see if reference SWATH-MS libraries are already available—i.e., that the libraries were generated from the 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, a.k.a shotgun”; Basic Protocol 3). Essentially, the complete SWATH-MS procedure 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, and the measurement of these samples on the MS in discovery mode to generate the search space for SWATH-MS. Basic Protocol 4 briefly describes 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 perform this only once per condition, and indeed it may be completely skipped if suitable reference libraries are available (e.g., if libraries were previously generated in the same tissue or organism, ideally on the same MS instrument to be used for the SWATH-MS). Furthermore, this step may not be necessary in the future as improvements continue in software solutions, such as DIA-Umpire (Tsou et al., 2015), which can directly generate the SWATH database search space using only DIA data. 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 via Basic Protocols 1 and 2, up to 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-hr gradient (corresponding to approximately a 1.5-hr 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  $-80^{\circ}\text{C}$  freezers for extended periods for total protein measurements.

## PROTEIN PREPARATION FOR SWATH-MS

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 ( $\geq 5$  mg). Modifications must be made if, for instance, the reader wishes to analyze the phosphoproteome or other post-translational modifications.

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

*NOTE:* Double-distilled water is adequate for this protocol.

### Materials

- RIPA-M buffer (see recipe), freshly prepared
- Urea-T buffer (see recipe), freshly prepared
- Tissue sample
- Protein quantification assay kit: e.g., bicinchoninic acid (BCA) or Bradford
- Homogenizer (Dounce pestle recommended, but a metallic bead homogenizer can also work)
- Refrigerated centrifuge (temperature range,  $\leq 4^{\circ}\text{C}$ ; capacity,  $\geq 24$  2-ml tubes; rotation speed,  $\geq 20,000 \times g$ )
- Bath sonicator (recommended)
- Vortex mixer

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-M 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).*

## BASIC PROTOCOL 1

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 or weighed. However, if taking a smaller amount of a large tissue sample, 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 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 (Shao et al., 2015).

2. Use a 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 the tissue—e.g., livers may be fully homogenized at low speed for 30 sec, while skeletal muscle and heart may need 2 min at a higher speed. Experiment with the sample type and homogenizer to determine a speed at which no large particles are visible in the homogenate. Take care to keep homogenization conditions consistent across all the samples for an experiment.*

3. Transfer the homogenate to a new tube of suitable volume (e.g., 1.5 ml for the example volumes in this protocol), or separate into multiple tubes if necessary. Centrifuge the samples for 15 min at 20,000 × g, 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 using a bath-type sonicator for 5 min (e.g., 40 kHz in an ice water bath).

*Again, the ratio is not critical and higher amounts of buffer can be used (e.g., 500 µl of UREA-T for 20 mg of wet tissue).*

6. Centrifuge these resuspended pellets for 15 min at 20,000 × g, 4°C.
7. Collect the supernatant and mix with the supernatant from step 3. Discard the pellet.
8. Measure protein concentration in the supernatant from step 7, e.g., by bicinchoninic acid assay (BCA).
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 as described in Basic Protocol 2.*

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

## PEPTIDE PREPARATION FOR SWATH-MS

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. Next, impurities are removed, such as lipids or salts that could affect the LC separation or MS ionization process. Take care that all liquid reagents, such as the water supply and acetone, are sufficiently pure for MS (“HPLC-grade”). Double-distilled water is not likely to be sufficiently clean, as mass spectrometers are sensitive to contaminants, e.g., salts and detergents.

This protocol is approximately a 3-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, as well as the amount of trypsin. This same protocol may be followed for 20 to 300  $\mu\text{g}$  of input protein—simply scale the volumes suggested here appropriately.

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<sub>2</sub>O 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 the following in advance:

1. 0.1 M NH<sub>4</sub>HCO<sub>3</sub> in H<sub>2</sub>O
2. A high-concentration ACN:H<sub>2</sub>O solution (8:2) + 0.1% TFA
3. A medium-concentration ACN:H<sub>2</sub>O solution (5:5) + 0.1% TFA
4. A low-concentration ACN:H<sub>2</sub>O solution (2:98) + 0.1% FA
5. 1% FA solution in H<sub>2</sub>O (1:99).
6. 0.1% TFA solution in H<sub>2</sub>O (1:999)

*NOTE:* Use HPLC-grade water for all recipes and protocol steps.

### **Materials**

Protein samples (e.g., from Basic Protocol 1)  
Acetone (HPLC-grade), cooled to  $-20^{\circ}\text{C}$   
Urea in 0.1 M ammonium bicarbonate (NH<sub>4</sub>HCO<sub>3</sub>)  
Ammonium bicarbonate (NH<sub>4</sub>HCO<sub>3</sub>)  
Dithiothreitol (DTT)  
Iodoacetamide (IAA)  
Trypsin (sequencing-grade)  
Methanol (HPLC-grade)  
Acetonitrile (ACN; HPLC-grade)  
Formic acid (FA; 98% to 100%; see annotation to step 24, below)  
Trifluoroacetic acid (TFA)  
Indexed Retention Time peptides (iRT; from Biognosys)

Heated shaker plate ( $37^{\circ}\text{C}$ )  
Refrigerated centrifuge (temperature range,  $\leq 4^{\circ}\text{C}$ ; capacity,  $\geq 24$  2-ml tubes;  
rotation speed,  $\geq 20,000 \times g$ )  
Silica C18 columns (e.g., MacroSpin Columns from The Nest Group)  
Vacuum evaporator (e.g., SpeedVac)  
Vials for mass spectrometry  
Mass spectrometer  
Vortex mixer  
NanoDrop microspectrophotometer  
Bath sonicator

### **Day 1**

1. Thaw samples on ice, then transfer 50  $\mu\text{g}$  of protein to a new tube.
2. Add 6 volumes of cold HPLC-grade acetone ( $-20^{\circ}\text{C}$ ) 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  $-20^{\circ}\text{C}$  freezer and wait for a few hours (e.g., 4 to 24 hr; keep consistent within a study).

## Day 2

4. Prepare three fresh reagents:

8 M urea in 0.1 M  $\text{NH}_4\text{HCO}_3$   
36 mM DTT  
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) 10 min at 20,000 × g, 4°C. Proteins should be well-fixed to the bottom of the tube. Remove acetone supernatant.
7. Add 90 µl of 8 M urea buffer (see step 4) for every 50 µg of protein and resuspend samples with a quick vortexing.
8. Add 45 µl of freshly prepared 36 mM DTT buffer (see step 4) for every 50 µg of protein.
9. Vortex briefly, then incubate samples on the 37°C shaking plate (see step 5) for 30 min 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 (see step 4) for every 50 µg of protein.
11. Vortex briefly, then incubate samples on the 25°C shaking plate for 45 min. Make sure that the samples are covered to protect from light during this time (e.g., with aluminum foil).
12. Dilute samples with 0.1 M  $\text{NH}_4\text{HCO}_3$  (see step 4) 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 on it for 16 to 24 hr at 600 rpm.

*Take particular care at this step to keep trypsin digestion times consistent for all digestions of a particular study. Avoid going beyond 24 hr, as trypsin will start to self-digest, which can create large peaks on mass spectrometry runs, obscuring the desired data. Conversely, a digestion time of less than 16 hr may not be sufficient.*

## Day 3

15. Activate the silica C18 columns with 450 µl of HPLC-grade methanol.

**IMPORTANT NOTE:** *Double-check this against the protocol that comes with your C18 column.*

16. Centrifuge 3 min at 1000 × g, room temperature. Discard flowthrough.
17. Repeat steps 15 and 16 twice.
18. Add 450 µl of ACN:H<sub>2</sub>O **8:2** + 0.1% TFA.
19. Centrifuge for 3 min at 1000 × g, room temperature. Discard flowthrough.
20. Repeat steps 18 and 19 twice.
21. Add 450 µl of 0.1% TFA.

22. Centrifuge for 3 min at  $1000 \times g$ , room temperature. Discard flowthrough.
23. Repeat steps 21 and 22 twice.
24. Take samples from the shaking plate (step 14).

*It is recommended to add 0.1% FA to the sample (i.e., add 53  $\mu\text{l}$  of 1% v/v FA to the sample volume of 480  $\mu\text{l}$ ), vortex, and check with pH paper to ensure that 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.*
25. Centrifuge 10 min at  $20,000 \times g$ ,  $4^\circ\text{C}$ .

*There should be no precipitate at the bottom. If there is, be careful not to transfer it into the C18 column in step 26.*
26. Take 450  $\mu\text{l}$  from each digested peptide sample and load onto a prepared C18 column.

*This will result in a loss of around 15% of the peptide quantity (i.e., due to total volume of around 530  $\mu\text{l}$ ). If maximum recovery is necessary, the entire sample may be loaded. Be careful not to load contaminants that were centrifuged to the bottom of the tube in step 25.*
27. Centrifuge 3 min at  $1000 \times g$ , room temperature.
28. Reload the outflow onto the column.
29. Again, centrifuge 3 min at  $1000 \times g$ , room temperature. The peptides should be trapped in the column. Discard flowthrough.
30. Add 450  $\mu\text{l}$  of 0.1% TFA.
31. Centrifuge for 3 min at  $1000 \times g$ , room temperature. Discard flowthrough.
32. Discard the old collection tube and replace with a new collection tube.
33. Repeat steps 31 and 32.
34. Add 450  $\mu\text{l}$  of ACN:H<sub>2</sub>O **5:5** + 0.1% TFA to elute the sample.
35. Centrifuge for 3 min at  $1000 \times g$ , room temperature.
36. Transfer column to a new collection tube and repeat steps 34 and 35. Combine flowthrough from the two steps of elution. Discard column.
37. Dry samples in a vacuum evaporator (e.g., SpeedVac).

*Warming the evaporator to  $37^\circ$  to  $45^\circ\text{C}$  will expedite this process.*

*If you do not plan on measuring your samples in the mass spectrometer immediately, stop at this step after the samples are dried, and freeze them at  $-80^\circ\text{C}$ .*
38. On the day that you expect to start the mass spectrometry measurement, resuspend the dried samples with ACN:H<sub>2</sub>O **2:98** + 0.1% FA to a target concentration of around 250 to 1000 ng/ $\mu\text{l}$ .

*The peptide quantity at the end will probably be 25% to 75% of the input protein quantity.*
39. Vortex and sonicate to resuspend the sample fully.
40. Centrifuge the samples 10 min at high speed (e.g.,  $20,000 \times g$ ),  $4^\circ\text{C}$ , to pellet any contaminants that may remain.

41. Measure the peptide concentrations on a NanoDrop spectrophotometer according to the manufacturer's instructions.
42. Transfer some of each sample to mass spectrometer sample tubes—the best quantity and concentration depend upon which mass spectrometer will be used.

*For SWATH-MS measurement on a Sciex TripleTOF 5600 mass spectrometer, it is recommended to transfer 5 to 10  $\mu\text{g}$  of peptide to a MS vial at a concentration of 250 to 1000 ng/ $\mu\text{l}$  in 19  $\mu\text{l}$  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<sub>2</sub>O 2:98 + 0.1% FA. However, do not dilute below 250 ng/ $\mu\text{l}$ . The quantification data will be normalized afterwards, but it is better to begin with as similar loadings as possible.*

43. Add 1  $\mu\text{l}$  of indexed retention time (iRT) peptides per 20  $\mu\text{l}$  loaded in step 42 (i.e., 100 femtomoles of iRT peptides).

*This allows for correction across samples for small shifts in the measured retention time.*

44. 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 measurement.
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 that the particular 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 analyzed, but fewer proteins will be quantified. Note that high amounts (e.g., >2  $\mu\text{g}$ ) may cause problems with certain machines.*

### **BASIC PROTOCOL 3**

## **DATA-DEPENDENT ACQUISITION 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. 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 (Ting et al., 2015). Therefore, to use the SWATH-MS method effectively, a peptide library is generally developed first. Alternatively, systems are being developed that allow untargeted analysis of DIA data without the need for a spectral library, for instance by DIA-Umpire (Tsou et al., 2015). 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 (Rosenberger et al., 2014a), it may be necessary to generate one. We have published a detailed protocol for library generation (Schubert et al., 2015). Here, we summarize the key steps and typical MS settings.

To increase the proteome coverage for the DDA-generated library, samples are preferably fractionated, e.g., by using isoelectric focusing with 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 are preferably obtained on the same type of instrument that 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 (Toprak et al., 2014). 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 by peptide-centric query strategies. It is also recommended to add indexed retention time reference peptides (iRT) to all samples that are used for library generation, as this allows effective peptide retention time normalization (Escher et al., 2012).

If a new library needs to be generated, it is not necessary to fractionate and perform DDA runs on every single sample—at most, only one sample per condition is necessary. In situations with many similar conditions, even this may be redundant. Consider the following: if examining the proteome of a wild-type 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 (Craig, Cortens, Fenyó, & Beavis, 2006; Frewen, Merrihew, Wu, Noble, & MacCoss, 2006; Lam et al., 2007), although each protocol has slight differences. The following protocol is designed to be relatively quick.

### **Materials**

Peptide mixture (e.g., from Basic Protocol 2)  
Buffer A: 2% ACN, 0.1% FA solution in H<sub>2</sub>O  
Buffer B: 98% ACN solution with 0.1% FA  
  
pH 3-10 IPG strip (Amersham Biosciences)  
3100 OFFGEL Fractionator (Agilent Technologies)  
Sciex 5600 TripleTOF mass spectrometer  
Eksigent NanoLC Ultra 2D Plus HPLC (interfaced to the mass spectrometer)  
C18 column (Magic, 3 μm; from New Objective)  
Analytical column (e.g., PicoFrit; Thermo Fisher Scientific; with 75-μm diameter)

1. For off-gel electrophoresis (OGE) fractionation, use 1 to 2 mg of peptide mixture digested as described in Basic Protocol 2.

*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 (Picotti, Bodenmiller, Mueller, Domon, & Aebersold, 2009). 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 on Day 3 of Basic 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., Sciex TripleTOF 5600), which should be interfaced to an HPLC (e.g., an Eksigent NanoLC Ultra 2D Plus; Schubert et al., 2015).
5. Load sample onto the C18 (Magic, 3  $\mu\text{m}$ ) packed (10 to 15 cm length of packing) emitter coupled with an analytical column (e.g., PicoFrit with a 75- $\mu\text{m}$  diameter) with buffer A. Elute the sample gradually over 135 min with a variable linear gradient of 2% to 35% 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 (100 to 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 (Kessner, Chambers, Burke, Agus, & Mallick, 2008).

*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; Keller, Eng, Zhang, Li, & Aebersold, 2005).
9. Set cysteine carboxymethylation as static modification and methionine oxidation as variable modification. Other modifications, e.g., phosphorylation, can be set as required, though this may require adaptations 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 (Elias & Gygi, 2007).

*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 a Sciex 5600 TripleTOF mass spectrometer, set mass tolerance to 25 parts per million (ppm) for precursor ions and 0.05 Da for fragment ions.
13. Filter out fragments that are smaller than 350  $m/z$  or bigger than 2000  $m/z$ .
14. Filter out fragments with  $m/z$  in the SWATH precursor isolation window.
15. Combine the pepXML files using iProphet (Shteynberg et al., 2011), and use the integrated pepXML file to generate the redundant spectral library containing all PSMs using SpectraST (Lam et al., 2007). It is important to estimate FDR at PSM, peptide, and proteins levels using the MAYU software (Reiter et al., 2009; Schubert et al., 2015). 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 five most abundant b and y fragment ions of each peptide to generate the assays for SWATH/MS targeted extraction. Please note that for a 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.

## **DATA-INDEPENDENT ACQUISITION AND TARGETED “PEPTIDE-CENTRIC” DATA EXTRACTION PROTOCOL**

**BASIC  
PROTOCOL 4**

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 (Wu et al., 2014), but the comparatively low throughput of these methods (typically up to ~100 proteins per run limits their 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.

### **Materials**

Samples (e.g., from Basic Protocol 2)

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)

Buffer A: 2% ACN, 0.1% FA solution in H<sub>2</sub>O

Buffer B: 98% ACN solution with 0.1% FA

Sciex 5600 TripleTOF mass spectrometer

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

Spectral library (from Basic Protocol 3, or downloaded from SWATHAtlas)

1. Measure the samples on the MS in a randomized sequence to minimize the potential measurement biases that arise with time during the runs, e.g., retention times may shift on the chromatography, or mass accuracy or 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 min 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.

**SWATH  
Proteomics for  
Mouse Biology**

**139**

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

- Collect the SWATH MS2 spectra from 100 to 2000  $m/z$ .
- Set the collision energy according to the calculation for ions of +2 charge centered upon the window with a spread of 15 eV.
- Use an accumulation time (dwell time) of 50 msec for fragment ion scans in high-sensitivity mode, and 150 msec for survey scans in high-resolution mode acquired at the beginning of each cycle, resulting in a duty time of  $\sim 3.4$  sec.

*This is not absolutely necessary, but some alternative analysis tools, e.g., DIA-Umpire, require high-quality MS1 spectra as well.*

- After data acquisition, convert the SWATH-MS.wiff files to .mzXML files using ProteoWizard (Kessner et al., 2008).
- Perform SWATH-MS targeted data extraction using the OpenSWATH workflow (Röst et al., 2014; Röst, Aebersold, & Schubert, 2016a). OpenSWATH applies a target-decoy scoring model to estimate the FDR using the mProphet algorithm (Reiter et al., 2011).

*Please note that the steps below refer to using OpenSWATH in iPortal. The command line of OpenSWATH is OpenSwathWorkflow. Refer to Röst et al. (2016a) to run OpenSWATH.*

- Select the transformed .mzXML files to be analyzed.
- Select the generated spectral library (from Basic Protocol 3, or downloaded from SWATHAtlas).
- Set the retention time window.

*By default, this is 300 sec.*

- Select retention time alignment method in OpenSWATH, e.g., iRT realignment (Escher et al., 2012) or TRIC (Röst et al., 2016b).
- Extract fragment-ion chromatograms according to the target-decoy assay library with a width of 0.05  $m/z$ .
- Set peptide FDR to 0.01.

*In OpenSWATH, peak groups are scored based on 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 (Gillet et al., 2016). 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.*

- 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 proteins corresponding to each peptide. Cases of non-proteotypic peptides (e.g., when a peptide sequence can correspond to more than one protein) can be discarded or not.

*There are several protocols for compressing each peptide back into a single protein (Rosenberger, Ludwig, Röst, Aebersold, & Malmstrom, 2014b). For instance, the peptide with the topN highest average intensity is taken as the best measurement of the protein level. Alternately, principal component analysis may be performed on all peptides corre-*

sponding to a particular protein (Choi et al., 2014; Teo et al., 2015) and the first principal 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

Use HPLC-grade water for all recipes and protocol steps.

### *RIPA-M buffer*

1% (v/v) NP-40  
150 mM NaCl  
1 mM EDTA  
50 mM Tris-HCl, pH 7.5  
Protease inhibitor cocktail (amount to use is based on tissue quantity used)  
Prepare fresh just before use

### *Urea-T buffer*

50 mM Tris·Cl, pH 8.1  
75 mM NaCl  
8 M urea  
Prepare fresh just before use

## COMMENTARY

### Background Information

SWATH-MS is an emerging technology for next-generation proteomics (Gillet et al., 2012). 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. A targeted proteomics strategy, e.g., selected reaction monitoring (SRM), allows sensitive and precise protein quantification for limited numbers of proteins per measurement. As an emerging technology, SWATH-MS permits the rapid and consistent quantification of thousands of proteins across large sample cohorts (Williams et al., 2016), 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 of the MS runs, which can occur either from using impure reagents or from insufficiently purified samples (e.g., in the C18 cleaning step). If measuring samples across several distinct batches for a continuous study, it is critical to 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 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 that can be compared across all samples and all batches. For instance, one may include the UPS2 Proteomic Standard from Sigma-Aldrich in every sample—a known quantity of 48 proteins that should be standardized across all runs. Take note that this standard will not work for human samples, as the UPS2 is human derived; for other organisms, take note that certain of the digested product-peptides may be identical across species.

### Troubleshooting

When performing a quality check on the MS for 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 the sample run, and they may be difficult to track 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 the instruments, an 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 several tens of thousands of peptides, which correspond to thousands of unique proteins. In general, the larger and more diverse the sample set, the more peptides (and proteins) can be quantified. 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 differentiate.

### Time Considerations

Protein preparation takes approximately 1 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 2 or 3 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 dozens of samples are to be prepared, the peptide digestion steps can be expedited by the use of a multichannel pipettor 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 measurements 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 to 90 min, but these values are adjustable and there is no single solution.

### Acknowledgments

We acknowledge Ludovic Gillet and Alexander Leitner for insightful discussion. Evan Williams was supported by an NIH F32 Ruth Kirchstein Fellowship (F32GM119190). Prof. Aebersold was supported by the ERC [Proteomics4D (AdvG grant 670821 and Proteomics v3.0; AdvG 233226)], the

SNSF (31003 A\_166435), and SystemsX (PhosphonetX).

### Literature Cited

- Aebersold, R., & Mann, M. (2016). Mass-spectrometric exploration of proteome structure and function. *Nature*, *537*, 347–355. doi: 10.1038/nature19949.
- Choi, M., Chang, C. Y., Clough, T., Broudy, D., Killeen, T., MacLean, B., & Vitek, O. (2014). MSstats: An R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics*, *30*, 2524–2526. doi: 10.1093/bioinformatics/btu305.
- Craig, R., Cortens, J. C., Fenyo, D., & Beavis, R. C. (2006). Using annotated peptide mass spectrum libraries for protein identification. *Journal of Proteome Research*, *5*, 1843–1849. doi: 10.1021/pr0602085.
- Elias, J. E., & Gygi, S. P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature Methods*, *4*, 207–214. doi: 10.1038/nmeth1019.
- Escher, C., Reiter, L., MacLean, B., Ossola, R., Herzog, F., Chilton, J., . . . Rinner, O. (2012). Using iRT, a normalized retention time for more targeted measurement of peptides. *Proteomics*, *12*, 1111–1121. doi: 10.1002/pmic.201100463.
- Frewen, B. E., Merrihew, G. E., Wu, C. C., Noble, W. S., & MacCoss, M. J. (2006). Analysis of peptide MS/MS spectra from large-scale proteomics experiments using spectrum libraries. *Analytical Chemistry*, *78*, 5678–5684. doi: 10.1021/ac060279n.
- Gillet, L. C., Leitner, A., & Aebersold, R. (2016). Mass spectrometry applied to bottom-up proteomics: Entering the high-throughput era for hypothesis testing. *Annual Review of Analytical Chemistry (Palo Alto, Calif.)*, *9*, 449–472. doi: 10.1146/annurev-anchem-071015-041535.
- Gillet, L. C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., . . . Aebersold, R. (2012). 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*, *11*, O111 016717. doi: 10.1074/mcp.O111.016717.
- Keller, A., Eng, J., Zhang, N., Li, X. J., & Aebersold, R. (2005). A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Molecular Systems Biology*, *1*, 2005 0017. doi: 10.1038/msb4100024.
- Kessner, D., Chambers, M., Burke, R., Agus, D., & Mallick, P. (2008). ProteoWizard: Open source software for rapid proteomics tools development. *Bioinformatics*, *24*, 2534–2536. doi: 10.1093/bioinformatics/btn323.
- Lam, H., Deutsch, E. W., Eddes, J. S., Eng, J. K., King, N., Stein, S. E., & Aebersold, R. (2007). Development and validation of a spectral library searching method for peptide identification from MS/MS. *Proteomics*, *7*, 655–667. doi: 10.1002/pmic.200600625.

- Picotti, P., Bodenmiller, B., Mueller, L. N., Domon, B., & Aebersold, R. (2009). Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. *Cell*, *138*, 795–806. doi: 10.1016/j.cell.2009.05.051.
- Reiter, L., Claassen, M., Schrimpf, S. P., Jovanovic, M., Schmidt, A., Buhmann, J. M., ... Aebersold, R. (2009). Protein identification false discovery rates for very large proteomics data sets generated by tandem mass spectrometry. *Molecular & Cellular Proteomics*, *8*, 2405–2417. doi: 10.1074/mcp.M900317-MCP200.
- Reiter, L., Rinner, O., Picotti, P., Huttenhain, R., Beck, M., Brusniak, M. Y., ... Aebersold, R. (2011). mProphet: Automated data processing and statistical validation for large-scale SRM experiments. *Nature Methods*, *8*, 430–435. doi: 10.1038/nmeth.1584.
- Rosenberger, G., Koh, C. C., Guo, T., Röst, H. L., Kouvonon, P., Collins, B. C., ... Aebersold, R. (2014a). A repository of assays to quantify 10,000 human proteins by SWATH-MS. *Scientific Data*, *1*, 140031. doi: 10.1038/sdata.2014.31.
- Rosenberger, G., Ludwig, C., Röst, H. L., Aebersold, R., & Malmstrom, L. (2014b). aLFQ: An R-package for estimating absolute protein quantities from label-free LC-MS/MS proteomics data. *Bioinformatics*, *30*, 2511–2513. doi: 10.1093/bioinformatics/btu200.
- Röst, H. L., Aebersold, R., & Schubert, O. T. (2016a). Automated SWATH data analysis using targeted extraction of ion chromatograms. *Methods in Molecular Biology*, *1550*, 289–307. doi: 10.1007/978-1-4939-6747-6\_20.
- Röst, H. L., Liu, Y., D'Agostino, G., Zanella, M., Navarro, P., Rosenberger, G., ... Aebersold, R. (2016b). TRIC: An automated alignment strategy for reproducible protein quantification in targeted proteomics. *Nature Methods*, *13*, 777–783. doi: 10.1038/nmeth.3954.
- Röst, H. L., Rosenberger, G., Navarro, P., Gillet, L., Miladinovic, S. M., Schubert, O. T., ... Aebersold, R. (2014). OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nature Biotechnology*, *32*, 219–223. doi: 10.1038/nbt.2841.
- Schubert, O. T., Gillet, L. C., Collins, B. C., Navarro, P., Rosenberger, G., Wolski, W. E., ... Aebersold, R. (2015). Building high-quality assay libraries for targeted analysis of SWATH MS data. *Nature Protocols*, *10*, 426–441. doi: 10.1038/nprot.2015.015.
- Shao, S., Guo, T., Koh, C. C., Gillissen, S., Joergler, M., Jochum, W., & Aebersold, R. (2015). Minimal sample requirement for highly multiplexed protein quantification in cell lines and tissues by PCT-SWATH mass spectrometry. *Proteomics*, *15*, 3711–3721. doi: 10.1002/pmic.201500161.
- Shteynberg, D., Deutsch, E. W., Lam, H., Eng, J. K., Sun, Z., Tasman, N., ... Nesvizhskii, A. I. (2011). iProphet: Multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. *Molecular & Cellular Proteomics*, *10*, M111 007690. doi: 10.1074/mcp.M111.007690.
- Teo, G., Kim, S., Tsou, C. C., Collins, B., Gingras, A. C., Nesvizhskii, A. I., & Choi, H. (2015). mapDIA: Preprocessing and statistical analysis of quantitative proteomics data from data independent acquisition mass spectrometry. *Journal of Proteomics*, *129*, 108–120. doi: 10.1016/j.jprot.2015.09.013.
- Ting, Y. S., Egertson, J. D., Payne, S. H., Kim, S., MacLean, B., Kall, L., ... MacCoss, M. J. (2015). Peptide-centric proteome analysis: An alternative strategy for the analysis of tandem mass spectrometry data. *Molecular & Cellular Proteomics*, *14*, 2301–2307. doi: 10.1074/mcp.O114.047035.
- Toprak, U. H., Gillet, L. C., Maiolica, A., Navarro, P., Leitner, A., & Aebersold, R. (2014). Conserved peptide fragmentation as a benchmarking tool for mass spectrometers and a discriminating feature for targeted proteomics. *Molecular & Cellular Proteomics*, *13*, 2056–2071. doi: 10.1074/mcp.O113.036475.
- Tsou, C. C., Avtonomov, D., Larsen, B., Tucholska, M., Choi, H., Gingras, A. C., ... Nesvizhskii, A. I. (2015). DIA-Umpire: Comprehensive computational framework for data-independent acquisition proteomics. *Nature Methods*, *12*, 258–264. doi: 10.1038/nmeth.3255.
- Williams, E. G., Wu, Y., Jha, P., Dubuis, S., Blattmann, P., Argmann, C. A., ... Auwerx, J. (2016). Systems proteomics of liver mitochondria function. *Science*, *352*, aad0189. doi: 10.1126/science.352.6282.247-b.
- Wu, Y., Williams, E. G., Dubuis, S., Mottis, A., Jovaisaite, V., Houten, S. M., ... Aebersold, R. (2014). Multilayered genetics and omics dissection of mitochondrial activity in a mouse reference population. *Cell*, *158*, 1415–1430. doi: 10.1016/j.cell.2014.07.039.