Title

Isolation of Histone from Sorghum Leaf Tissue for Top Down Mass Spectrometry Profiling of Potential Epigenetic Markers.

Authors

Zhou, Mowei

Abdali, Shadan H

Dilworth, David

et al.

Publication Date

2021

DOI

10.3791/61707

Peer reviewed

Isolation of Histone from Sorghum Leaf Tissue for Top Down Mass Spectrometry Profiling of Potential Epigenetic Markers

AUTHORS AND AFFILIATIONS:

Mowei Zhou[1], Shadan H. Abdali[1], David Dilworth[2], Lifeng Liu[2], Benjamin Cole[2], Neha Malhan[1], Amir Ahkami[1], Tanya E. Winkler[1], Joy Hollingsworth[3], Julie Sievert[3], Jeff Dahlberg[3], Robert Hutmacher[4,5], Mary Madera[6], Judith A. Owiti[6], Kim Hixson[1], Peggy G. Lemaux[6], Christer Jansson[1], Ljiljana Paša-Tolić[1]

[1]Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA, USA

[2]DOE-Joint Genome Institute, Lawrence Berkeley Laboratory, Berkeley, CA, USA

[3]Kearney Agricultural Research and Extension Center, University of California Agriculture and Natural Resources, Parlier, CA, USA

[4]West Side Research and Extension Center, University of California, Five Points, CA, USA

[5]Department of Plant Sciences, University of California, Davis, USA

[6]Department of Plant and Microbial Biology, University of California, Berkeley CA, USA

Corresponding author:

Ljiljana Paša-Tolić (Ljiljana.PasaTolic@pnnl.gov)

Key Words

Drought, Epigenetic, histone clipping, post-translational modifications, proteomics, sorghum, top down mass spectrometry

Summary

The protocol has been developed to effectively extract intact histones from sorghum leaf materials for profiling of histone post-translational modifications that can serve as potential epigenetic markers to aid engineering drought resistant crops.

Abstract

Histones belong to a family of highly conserved proteins in eukaryotes. They pack DNA into nucleosomes as functional units of chromatin. Post-

translational modifications (PTMs) of histones, which are highly dynamic and can be added or removed by enzymes, play critical roles in regulating gene expression. In plants, epigenetic factors including histone PTMs are related to their adaptive responses to the environment. Understanding the molecular mechanisms of epigenetic control can bring unprecedented opportunities for innovative bioengineering solutions. Herein, we describe a protocol to isolate the nuclei and purify histones from sorghum leaf tissue. The extracted histones can be analyzed as their intact forms by top-down mass spectrometry (MS) coupled to online reversed-phase (RP) liquid chromatography (LC). Combinations and stoichiometry of multiple PTMs on the same histone proteoform can be readily identified. In addition, histone tail clipping can be detected using the top-down LC-MS workflow thus yielding the global PTM profile of core histones (H4, H2A, H2B, H3). We have applied this protocol previously to profile histone PTMs from sorghum leaf tissue collected from a large-scale field study, aimed at identifying epigenetic markers of drought resistance. The protocol could potentially be adapted and optimized for chromatin immunoprecipitation – sequencing (ChIP-seq), or for studying histone PTMs in similar plants.

Introduction

The increasing severity and frequency of drought is expected to affect productivity of cereal crops.[1, 2] Sorghum is a cereal food and energy crop known for its exceptional ability to withstand water-limiting conditions.[3, 4] We are pursuing mechanistic understanding of the interplay between drought stress, plant development, and epigenetics of sorghum [*Sorghum bicolor* (L.) Moench] plants. Our previous work has demonstrated strong connections between plant and rhizosphere microbiome in drought acclimation and responses at the molecular level.[5–7] This research will pave the way for utilizing epigenetic engineering in adapting crops to future climate scenarios. As part of the efforts in understanding epigenetics, we aim to study protein markers that impact gene expression within the plant organism.

Histones belong to a highly conserved family of proteins in eukaryotes that pack DNA into nucleosomes as fundamental units of chromatin. Post-translational modifications (PTMs) of histones are dynamically regulated to control chromatin structure and influence gene expression. Like other epigenetic factors including DNA methylation, histone PTMs play important roles in many biological processes.[8, 9] Antibody-based assays such as Western blots have widely been used to identify and quantify histone PTMs. In addition, the interaction of histone PTMs and DNA can be effectively

probed by Chromatin immunoprecipitation – sequencing (ChIP-seq).[10] In ChIP-seq, chromatin with specific targeted histone PTM is enriched by antibodies against that specific PTM. Then the DNA fragments can be released from the enriched chromatin and sequenced. Regions of genes that interact with the targeted histone PTM are revealed. However, all these experiments heavily rely on high quality antibodies. For some histone variants/homologs or combinations of PTMs, development of robust antibodies can be extremely challenging (especially for multiple PTMs). In addition, antibodies can only be developed if the targeted histone PTM is known.[11] Therefore, alternative methods for untargeted, global profiling of histone PTMs are necessary.

Mass spectrometry (MS) is a complementary method to characterize histone PTMs, including unknown PTMs for which antibodies are not available.[11, 12] The well-established “bottom-up” MS workflow uses proteases to digest proteins into small peptides prior to liquid chromatography (LC) separation and MS detection. Because histones have large numbers of basic residues (lysine and arginine), the trypsin digestion (protease specific to lysine and arginine) in the standard bottom-up workflow cuts the proteins into very short peptides. The short peptides are technically difficult to analyze by standard LC-MS, and do not preserve the information about the connectivity and stoichiometry of multiple PTMs. The use of other enzymes or chemical labeling to block lysines generates longer peptides that are more suitable for characterization of histone PTMs.[13, 14]

Alternatively, the digestion step can be completely omitted. In this “top-down" approach, intact protein ions are introduced into the MS by electrospray ionization (ESI) after online LC separation, yielding ions of the intact histone proteoforms. In addition, ions (i.e. proteoforms) of interest can be isolated and fragmented in the mass spectrometer to yield the sequence ions for identification and PTM localization. Hence, top-down MS has the advantage to preserve the proteoform-level information and capture the connectivity of multiple PTMs and terminal truncations on the same proteoform.[15, 16] Top-down experiments can also provide quantitative information and offer insights of biomarkers at the intact protein level.[17] Herein, we describe a protocol to extract histone from sorghum leaf and analyze the intact histones by top-down LC-MS.

Protocol

1. Preparing sorghum leaf material

NOTE: The sorghum plants were grown in soil in the field in Parlier, CA. Leaf tissue was collected by tearing off the third and fourth fully emerged leaf

from the primary tiller. More details of field condition, sample growth, and collection can be found in published report.[18] The example data shown in Figure 1 and Figure 2 were from sorghum leaf collected at week 2 after planting. Although variation of yield is expected, we believe this protocol is generally agnostic to specific sample conditions. We have successfully used the same protocol for sorghum plant leaf tissue collected from 2, 3, 5, 8, 9, and 10 weeks after planting.

1.1 Collect sorghum leaves from plants into 50 mL centrifuge tubes and immediately freeze the tube in liquid nitrogen.

1.2 Grind leaves with liquid nitrogen and immediately transfer to a centrifuge tube.

1.3 Store the ground leaf at -80 °C until use. Take about 4 g of cryo-ground leaf powder for histone analysis of each sample.

2. Preparing buffers and materials (3-4 hours)

NOTE: The high concentration stock solutions can be made ahead of time and stored until use. But all working buffers must be made fresh on the day of the extraction (by dilution from stock and mixing with other contents) and to be placed on ice during the process. The whole experiment should be performed at 4 °C unless recommended otherwise.

2.1 Prepare 2.5M Sucrose by dissolving 42.8g sucrose (342.30g/mol) in 15 mL of sterile water on heat plate in a glass container with continuous stirring. Bring up the volume to 50mL once the sucrose has dissolved completely. Store the sucrose in 4 ◦C until used.

2.2 Prepare 1M of Tris pH 8 by dissolving 1.576 g Tris HCl in 10 mL H_2O in a 15 mL centrifuge tube. Adjust pH with NaOH to 8 and check with pH paper. Store it at 4 ◦C until used.

2.3 Prepare 1M of Dithiothreitol (DTT) by weighting 231 mg of DTT (154.25 g/mol) and dissolving it in 1.5 mL sterile water. DTT must be made fresh, or use stored frozen aliquots.

2.4 (Optional) Prepare the additional inhibitors by mixing three different salts. Prepare 18.38 mg of Sodium Orthovanadate (183.91 g/mol) in 1 mL of sterile water, then prepare separately Sodium Butyrate by adding 11.008 mg of Sodium Butyrate (110.09 g/mol) in 1 mL of Sterile water. Prepare the final salt by adding 4.199 mg of Sodium Floride (41.99 g/mol) in 1 mL of water. Mix the three salts solutions together with equal volume as stock solution for "additional inhibitors" (33 mM of each of the three chemicals).

NOTE: Sodium vanadate polymerizes at concentrations higher than 0.1 mM under neutral pH. It is advised to activate sodium vanadate to depolymerize it for maximum efficacy following published protocols.[19] Alternatively, activated sodium vanadate is commercially available. Herein we did not intentionally activate the sodium vanadate so the efficacy may have been reduced. We have not yet tested activated sodium vanadate for this protocol.

2.5 Prepare 1 M of $MgCl_2$ by dissolving 0.952 g of anhydrous magnesium chloride (95.2 g/mol) in 10 mL of H_2O in a 15 mL centrifuge tube. Store 1M of $MgCl_2$ at 4 °C until used

2.6 Prepare 10% (v/v) Triton X-100 by mixing 53.5g Triton X-100 with 35 mL of sterile water, bring up to 50 mL with water and store it at room temperature.

2.7 Prepare 5% Guanidine buffer pH7 (referred as "Gdn buffer") that will be used to condition the resin at least overnight – prepare 0.1M Potassium Hydrogen phosphate dibasic (K_2HPO_4) by weighting 870 mg of K_2HPO_4 in 50 mL of sterile water and store at 4 °C.

2.8 Weigh 0.7 g of guanidine hydrochloride and dissolve in 0.1M K_2HPO_4 to a final volume of 14mL. Adjust pH to 7 by checking with pH paper.

2.9 Soak the dry weak cation exchange (WCX) resin in 5% Guanidine buffer pH 7 overnight. Remove the supernatant and refill with fresh 5% Gdn buffer and soak it again overnight to let the resin fully equilibrate (till the supernatant has the same pH as the original buffer).

2.10 Before starting the experiment in the next section, mix the reagents to make EB1, EB2A, and EB2B buffer based on Table 1. Add all inhibitors and DTT fresh just before use.

Table 1. Composition for extraction buffers (EBs).

Reagents	Stock concentration	**EB1**	**EB2A**	**EB2B**
		Volume (mL)	Volume (mL)	Volume (mL)
Sucrose	2.5M	4.4	1.25	0.5
Tris HCl pH8	1M	0.25	0.125	0.05
DTT	1M	0.125	0.0625	0.025
H_2O		20.225	9.6875	4.375
protease inhibitor (PI) tablet		0.5 pill	0.5 pill	0.5 pill
Additional inhibitors (Optional)	33mM	0.25	0.125	0.05
$MgCl_2$	1M		0.125	0.05
Triton X100	10%		1.25	
Overall Volume		25 mL	12.5 mL	5 mL

2.11 Make the Nuclei Lysis Buffer (NLB) based on Table 2. NLB has to be prepared in advance and stored at 4 °C. Add PI tablets fresh just before use at 1X (0.5 tablet per 5 mL). See Table 2 for specific volumes.

Table 2. Composition for the nuclei lysis buffer (NLB).

NLB	Stock concentration	Volume (mL)
NaCl	5M	0.4
Tris HCl pH8	1M	0.05
Triton X100	10%	0.5
EDTA	0.5M	0.2
H_2O		3.85
PI tablets		0.5 pill
Additional inhibitors (optional)	33mM	0.05

Overall Volume		5 mL

3. Nuclei Isolation Procedure

NOTE: It is recommended to perform steps 3.1-3.3 of the first day (2-3h), save the nuclei in NLB buffer at -80 °C, and resume the following day (or later) for protein purification (4h). The nuclei isolation steps in this protocol was adapted from a sorghum ChIP-seq protocol being used at the Joint Genome Institute. Additional washes and sucrose gradient separation may be required to ensure nuclei purity for ChIP-seq applications.

3.1 Filtration of Debris (~0.5h)

3.1.1 Weigh ground leaf powder ~4 g, ensuring it remains frozen by placing on dry ice or liquid Nitrogen until ready to use.

3.1.2 Add protease inhibitor tablets to EB1 to a final concentration of 0.2X (0.5 tablet for 25 mL per sample)

NOTE: It is recommended to use a miniature plastic pestle or a pipet tip to pre-crush tablets in a microcentrifuge tube prior to adding to buffers to aid in dissolution of the tablet in the buffer. To prevent material loss, add the PI tablet and sonicate the buffer to dissolve the tablet.

3.1.3 Add 20 mL of EB1 into the frozen ground leaf powder, gently vortex and mix them until the powder if completely suspended. Keep gently mixing them for ~10min.

3.1.4 Filter through mesh 100, rinsing the filtered material twice with 2 mL of EB1 each time.

NOTE: Both the filtrate and the filtered debris should be green. If tracking using a microscope, you should be able to see intact nuclei and intact chloroplasts in the filtrate at this point. The majority of large debris should be absent/depleted. Mix dyes such as methylene blue with sample. Nuclei are easily observable as ~3-5 µm diameter dark blue/aquamarine spheres when visualized using a using a 20X, 40X and/or 100X objective. Relative to nuclei, chloroplasts are similar in size, but greenish in color and often more oval in shape. Vacuoles are also similar to nuclei in size and shape, but they will not readily take up the Methylene blue dye.

3.1.5 Centrifuge the combined filtrate at 3000g for 10 minutes at 4 °C in a swinging bucket rotor to pellet debris and large subcellular organelles, including nuclei and chloroplasts.

NOTE: It is recommended to prepare EB2A during this spin (see step 3.2.1).

3.1.6 Decant the supernatant, being careful to not disturb the pellet.

NOTE: As no detergent has yet been added, the pellet should remain intense green and the supernatant should be, at most, pale green/yellow.

3.2 Lysis of non-target organelles (~0.5h)

3.2.1 Prepare EB2A by adding protease inhibitors to a final concentration of 0.4X (0.5 tablet per 12.5mL EB2A).

3.2.2 Resuspend pellet from step 3.1.6 in 5 mL of EB2A and incubate on ice for 10 minutes with gentle mixing.

NOTE: The detergent concentration needs to be optimized to preferentially lyse intact cells and chloroplasts but not nuclei. The amount required can vary among organisms. It is recommended to check for lysis of chloroplasts and retention of intact nuclei under microscope.

3.2.3 Centrifuge at 2100g for 15 minutes at 4 °C in a swinging bucket rotor to pellet debris and nuclei.

NOTE: At this stage, the supernatant should be intensely green, and the pellet should be much less green than observed at previous stages due to the lysis of chloroplasts and chlorophyll release into the cytosol.

3.2.3 Decant the supernatant, being careful to not disturb the pellet.

3.3 Isolation of Nuclei from remaining cytoplasmic contaminants (~0.5h)

3.3.1 Prepare EB2B by adding protease inhibitors to a final concentration of 1X (0.5 tablet per 5 mL EB2B).

3.3.2 Resuspend crude nuclear pellet from step 3.2.3 in 2 mL of EB2B.

NOTE: EB2B does not contain TritonX-100, so no additional lysis should occur at this point.

3.3.3 Centrifuge at 2100g for 15 minutes at 4 °C in a swinging bucket rotor to pellet debris and nuclei.

NOTE: Small organelles and cytoplasmic components should not pellet, so they should remain in the supernatant.

3.3.4 Decant the supernatant, being careful to not disturb the pellet.

3.3.5 Resuspend the pellet using 250 µL of NLB (add 0.5 protease inhibitor tablet fresh for 5 mL).

NOTE: The goal is to resuspend the nuclei in a minimum amount of NLB without significant material loss. Because NLB is very viscous and the pellets contain a large amount of insoluble debris, it is very difficult to pipet and tends to cling to the inside of pipet tips. For this reason, it is recommended to reuse the same pipet tip whenever possible. If you are concerned with residual material in a pipet tip, simply hang the pipet from a shelf or rack for ~1 minute to allow gravity to collect material at the opening of the tip. Do not aggressively pipet to resuspend the pellets. Instead, use the pipet tip as a stir rod until the pelleted material can be aspirated into the pipet tip. i.e. it is perfectly fine for large pellet clumps to persist at this stage so long as it can be drawn into a pipet tip.

3.3.6 Vortex 15 seconds at max to homogenize and partially resuspended material. Sonicate for 5 min at 4 °C, then store at -80 °C.

NOTE: For subsequent steps, keep in mind that the total amount of NLB added is 250µL, but the total apparent volume of the sample can be up to twice as much due to insoluble debris. The sample is frozen and thawed to assist in lysis of nuclei.

3.4 Nuclei lysis and histone extraction (~4h)

3.4.1 Add 750 µL 5% Gdn buffer to the thawed sample. Sonicate for 15 minutes at 4 °C.

3.4.2 Transfer sample into a single 2 mL tube and spin 10,000 *x g* for 10 minutes at 4 °C.

NOTE: The supernatant will likely look green. The following chromatography steps should remove most of the pigments from the protein.

3.4.3 While waiting on step 3.4.1 and 3.4.2, prepare the column for ion exchange chromatography clean up. Rinse the chromatography column with 2 mL acetonitrile and 4 mL water to minimize contamination on surface. Then load 200~300 µL WCX resin (pre-conditioned with 5% Gdn buffer) onto the chromatography column. Let the resin settle. Wash four times with 1 mL 5% Gdn buffer. Keep the tube and column on ice for the rest of the purification steps.

3.4.4 Put the chromatography column on a 2 mL collection tube. Load the supernatant from step 3.4.2 slowly onto the resin bed without disrupting the resin (try to slowly drop from the side of the tubes). Let the solution flow through by gravity. As the solution is flowing through, load the eluent back to the top of the column 6-8 times to maximum binding to the resin. Then discard the eluent.

3.4.5 Load 2 mL 5% Gdn buffer to wash non-histone proteins off the column. Discard the eluent.

3.4.6 Elute histones with 1 mL 20% Gdn buffer. Collect the eluent which contains histone proteins.

3.4.7 Use 3k molecular weight cut off (MWCO) spin filter (0.5 mL) to desalt the eluent from step 3.4.6. Before use, load 500 µL wash solvent (0.2% formic acid in 3% ACN) and spin it down twice to clean the filter.

NOTE: It is recommended to start washing the MWCO filter while performing the resin chromatography steps to save time. The following spin filter steps take ~ 3-4 hr.

3.4.8 First load 500 µL of histone sample, spin at 14 kg for ~25 min to reduce volume down to ~100 µL. Then load another 400 µL of sample and spin at 14 kg again for ~25min. Load the final 100 µ L of sample, rinse the sample tube with 300 µL wash solvent and load the solvent into the filter. Spin at 14 kg again for ~25 min.

3.4.9 Load 400 µL w ash solvent, spin at 14 kg for ~25 min to reduce volume to ~100 µL or less. Each cycle reduces salt by 1/5. Repeat for another three cycles to bring guanidine concentration to ~0.01%. Reverse the filter into a clean collection tube and spin at 1 kg for 2 min. Save the purified histone sample at -20 C or -80 C for analysis.

NOTE: It is recommended to spin longer (30-40min) at the last step to minimize sample volume in order to obtain higher concentration. The volume should be able to go down to 50-70 µL.

4. Mass spectrometry of purified histones

4.1 Liquid chromatography mass spectrometry (LC-MS) data acquisition

4.1.1 Estimate protein concentration by Bicinchoninic Acid (BCA) assay following manufacturer's protocol.

NOTE: BCA can only give an estimate of total protein concentration, but not the quality of histone purification. If MS instrumentation is not readily available for checking the quality of histone purification, Western blot can be used. Reversed-phase LC coupled with 210 nm ultraviolet absorbance detection as described in our previous report can be also used.[20] The chromatogram can be compared with a known standard for checking sample quality. However, different organism can have different elution profiles. Therefore, using histone standards from similar organisms is highly recommended.

4.1.2 Connect a C18 reversed phase (RP) analytical column (e.g., 3 µm 300 Å, column inner diameter 75 µm, outer diameter 360 µm, length 70 cm) and a C18 trap column (e.g. 3.6 µm, column inner diameter 150 µm, outer diameter 360 µm, length 5cm) to a dual-pump nanoflow liquid chromatography system (e.g., Waters NanoAcquity). The binary solvents are A: 0.1% formic acid in water, and B: 0.1% formic acid in acetonitrile.

NOTE: The dual pump LC includes a wash pump and a gradient pump. Both pumps go through two stages in each analysis – a trapping stage followed by then an analytical stage. In the trapping stage, the wash pump flows into the trap column and the gradient pump flows into the analytical column. In the analytical stage, the trap column is coupled with the analytical column, and the gradient pump flows into both columns. The wash pump goes to the waste then.

4.1.3 Trapping stage: set up the LC method to first load 1-2 µg of histone sample onto the trap column. Desalt the sample by the wash pump at 3 µL/min 5% solvent B for 10 min. Set the analytical pump at 0.3 µL/min 5% solvent B for equilibration.

4.1.4 Analytical stage: set the gradient pump (0.3 µL/min) to start at from 5% B and ramp to 30% at 15min. Then increase to 41% B at 100 min before a high organic wash up to 95% B at the end.

NOTE: The gradient can be optimized depending on the different retention profiles on individual columns. Typically, full-length histones elute around 30-40% B on the specified LC conditions. Longer gradients can be used to increase the numbers of MS2 spectra to capture more histone proteoforms.

4.1.5 Set up data-dependent acquisition method on a high-resolution MS (e.g. Thermo Orbitrap Fusion Lumos or similar) with electron transfer dissociation (ETD) capability. Use the intact protein mode and perform all necessary calibrations suggested by the manufacturer. Critical parameters (when using Thermo Orbitrap Fusion Lumos) are as follows:
MS1: scan range 600-2000 m/z; resolution 120k; 4 microscans; AGC target 1E6; max injection 50 ms.
MS2: resolution 120k; 1 microscan; AGC target 1E6; data dependent MS/MS: alternating ETD (25 ms reaction time, max injection time 500 ms) and higher-energy collisional dissociation (HCD, 28% normalized collision energy with ±5% stepped energy, max injection time 100 ms); isolation window of 0.6 Da; priority on highest charge states.
Dynamic exclusion: 120 s, ±0.7 Da. Exclude charge states lower than 5 and undetermined charge states.

NOTE: It is recommended to run a few injections of peptide or histone standards on new columns to equilibrate and check the system, before running the actual samples. For running large number of samples, add short blanks or washes in between samples to minimize carry over. Let the columns equilibrate for 15-20 min at the starting condition (5% solvent B) before the next sample.
Longer LC gradients and higher max injection time for MS2 can improve the spectral quality for identifying more histone proteoforms.

4.2 LC-MS data processing and proteoform identification

4.2.1 Obtain the (sorghum) protein database in FASTA format from JGI (https://genome.jgi.doe.gov) or UniProt (https://www.uniprot.org/).

4.2.2 Use MSConvert[21] (http://proteowizard.sourceforge.net/tools.shtml) to convert the instrument raw data files (*.raw) into mzML format.

4.2.3 Download TopPIC suite[22] (http://proteomics.informatics.iupui.edu/software/toppic/) for data processing. The program can be run in either command line or through the graphical interface.

4.2.4 Use TopFD in the TopPIC suite to deconvolute the spectra from the mzML file from step 4.2.2. Default parameters can be used. But the "precursor window" (-w) need to be reduced to 1 m/z because a narrow isolation window is used.

4.2.5 Use TopPIC in the TopPIC suite to identify proteoforms. Most of the default parameters can be used. Set the spectrum and proteoform cutoff type to FDR (false discovery rate), and set the cutoff value to 0.01 (1% FDR) or as desired. Set the "proteoform error tolerance" to 5 (Dalton). Load the FASTA file from step 4.2.1 and the "*_ms2.msalign" file from step 4.2.4. Then start the search.

NOTE: The "proteoform error tolerance" setting will combine proteoforms with similar masses (± 5 Da) as one. This helps reduce redundancy in the proteoform counts. However, it should be used with caution because large tolerance will merge proteoforms with small or no mass differences. This parameter is only available in TopPIC version 1.3 or later.

4.2.6 The identified proteoforms can be examined in the "*_proteoform.csv" file or visualized using the Topview module under the "*_html" folder of the output.

4.2.7 The proteoforms list generated from the steps above using TopPIC annotates the histone PTMs as mass shifts. In order to localize individual PTMs, a modification file list must be included. Detailed description can be found in the TopPIC manual. Alternatively, proceed to the next step to perform a complementary data analysis using the Informed-Proteomics package[23] (https://github.com/PNNL-Comp-Mass-Spec/Informed-Proteomics).

4.2.8 Follow the instructions and use the PbfGen module to convert the instrument raw data to PBF file. Then deconvolute the MS1 data using ProMex module to output a ms1ft file (feature list, each feature represents a unique combination of mass and retention time).

4.2.9 Create a focused FASTA for Informed-Proteomics using the identified protein list from TopPIC in step 4.2.6.

NOTE: Searching the entire genome using Informed-Proteomics with large number of variable PTMs can be extremely slow, and may cause crashes. Therefore, it is recommended to reduce the size of FASTA by only including the target proteins.

4.2.10 Create a targeted modification list to search for histone PTMs following the format in the example file. The common PTMs to include are: Lysine acetylation, lysine mono-methylation, lysine di-methylation, lysine tri-methylation, serine/threonine/tyrosine phosphorylation, protein N-terminal acetylation, methionine/cysteine oxidation. For sorghum, protein N-terminal mono-methylation, di-methylation, and trimethylation should be added.

NOTE: Informed-Proteomics only looks for PTMs specified in the list. If unspecified PTMs are present, the proteoform may not be identified, or misidentified to other proteoforms. However, the PTM list should be kept as short as possible to minimize the search time.

4.2.11 Execute the MSPathFinder module to identify proteoforms using the files from step 4.2.8, the focused FASTA from step 4.2.9, and the modification list from step 4.2.10. Default parameters can be used.

4.2.12 The results can be visualized in LcMsSpectator by loading all the result files.

NOTE: Other bioinformatics tools are available for processing and visualizing top-down data, each with its own strengths.[24–28] Sorghum and many other organisms have limited known information regarding histone PTMs in the database. We recommend using TopPIC first to identify mass shifts from PTMs. This analysis can readily discover both known and unknown PTMs. Then the detected PTMs can be searched in a targeted fashion either by specifying a PTM list in TopPIC, or with other complementary tools.

Representative Results

Following the protocol, the histones can be extracted and identified using the LC-MS analysis. The raw data and processed results are available at MassIVE (https://massive.ucsd.edu/) via accession: MSV000085770. Based on the TopPIC results from the representative sample (available also from MassIVE), we identified 303 histone proteoforms (106 H2A, 72 H2B, 103 H3, and 22 H4 proteoforms). Co-purified ribosomal proteoforms have also been detected,

typically eluting early in the LC. They usually consist of ~20% of the identified proteoforms, but do not overlap with the histone proteoforms eluting in the later stage of the LC gradient. The results can be easily visualized with the latest TopPIC or Informed-Proteomics packages. For demonstration, we will focus on the data visualization using the Informed-Proteomics package, which can be used to directly load raw MS files and manually examine proteoform identifications. Please note that the two software packages use different algorithm and parameters. The reported numbers of proteoforms will not be identical. We recommend reporting the proteoform counts from TopPIC because it is more conservative, and it does consider unknown PTMs. Informed-Proteomics package has integrated data processing and visualization for easy manual validation. For organisms with well-annotated PTMs, we recommend ProSightPC[24] for best site localization. Combining the results using multiple tools can increase the the number of and the confidence of proteoform identifications.

After processing the data with Informed-Proteomics, the LC-MS feature map can be visualized in LcMsSpectator, which displays the deconvoluted protein masses against the LC retention time. By clicking on the identified proteoforms in the software, the associated feature will be highlighted with a small green rectangle in the feature map. Major histone proteins should be seen in specific regions of the map, which indicates the success of the experiment. Figure 1a shows a representative LC-MS feature map of intact histones. Full-length histone proteoforms are highlighted in the dashed boxes. Most proteoforms detected can be confidently identified using MS^2 data.

Figure 1b shows the zoom in of the region with H2A and H2B proteoforms. Most of them have N-terminal modifications of 42 Da. This nominal mass corresponds to either trimethylation (42.05 Da) or acetylation (42.01 Da), which are commonly seen for histones. Their accurate masses differ only by 0.04 Da, and are difficult to differentiate at the intact protein level (~2 ppm). In high resolution MS^2 spectra, the two PTMs can be easily differentiated and confirmed because of the lower mass of the fragments.[29] In addition, H2A and H2B histones have multiple homologs with very similar sequences as noted by the different UniProt accession numbers in Figure 1b. Again, high resolution LC-MS analysis can readily identify and differentiate them. Two types of H2As were identified for sorghum histones. The 16 kDa H2A histones in Figure 1b have extended terminal tails in the non-conserved regions of histones. Another group of H2A histones without the extended tails (14 kDa) can be seen in Figure 1c.

For H4 histones, N-terminal acetylation was identified as major PTM. Additional lysine acetylations and methionine oxidations can be also

observed simply by examining the mass differences of the features in Figure 1d. We also observed an unknown modification of 112.9 Da in addition to the N-terminal acetylation (the feature above “3Ac” in Figure 1d). This is likely some unknown adducts from reagent used in the preparation. We have previously detected sulfate ion adducts on H4, which may be attributed to residual salts combined with high basicity of histone proteins. For H3, two protein sequences were identified H3.3 and H3.2 (Figure 1e). Although these two protein sequences differ at only 4 residues (32, 42, 88, and 91), they can still be easily distinguished in LC-MS based on the separation in both dimensions, mass and retention time. H3 proteins are heavily modified by varying degrees of methylation and acetylation. The high degree of modification can be easily visualized by the dense, parallel lines in the feature map, which are 14 Da apart. However, three methylation groups (14*3 Da) have the equal nominal mass to one acetylation (42 Da). Because these PTMs cannot be easily resolved at intact protein level, they are referred to as “methyl equivalents” (i.e. multiples of 14 Da; one acetylation equals three methyl equivalents). In Figure 1e, H3 proteoforms are labeled in the form of methyl equivalents based on their intact mass. Due to limited resolution of the RPLC separation, many different H3 proteoforms are likely co-eluting and fragmented in the same spectrum. The method presented here will only identify the most abundant combinations of methylation and acetylation as illustrated in Figure 2. For more comprehensive characterization of H3, more targeted analysis is still required.[30, 31]

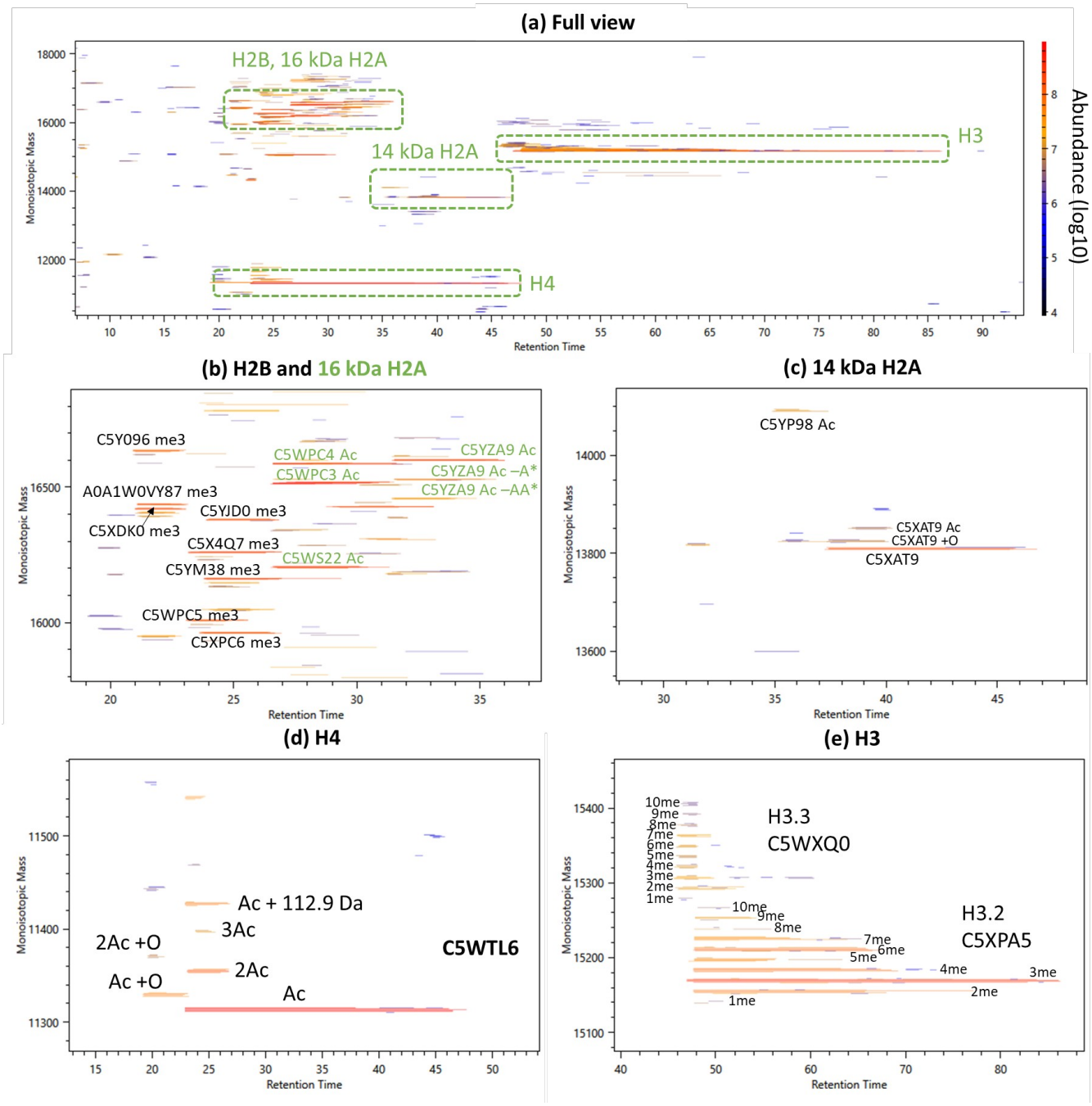

Figure 1: LC-MS feature map on intact histones extracted from sorghum leaves. The figure shows LC retention time (in min) vs. molecular mass for all detected proteoforms. The log abundance is shown by the color scale next to the top map (log 10 abundance). (a) The major histone peaks are labeled by the dashed boxes. Most the features outside the boxes are truncated histones and ribosomal proteins. Zoom-in views for each group of histones: (b) H2B and 16kDa H2A, (c) H3, (d) 14 kDa H2A, and (e) H3. The

UniProt accession numbers are noted alongside each feature, followed by detected PTMs. “Ac”, “me”, “+O” indicate acetylation, methylation, and oxidation, respectively. In (b), two truncated H2A C5YZA9 proteoforms are labeled, which had one or two C-terminal alanine clipped (shown as -A*, and -AA*).

A representative example of proteoform identification is shown in Figure 2 using MSPathfinder and visualized in LcMsSpectator. The fragmentation spectrum in Figure 2a was generated using ETD, which yields c and z type ions along the protein backbone. HCD of the same precursor can be used to validate the identification, but HCD generally provides limited sequence coverage.[20] The precursor ions in the previous and next MS1 spectra are shown in Figure 2b-c, with their matched isotope peaks highlighted in purple. The sequence coverage map in Figure 2d can help localize any possible PTMs. A high-confidence identification should have most of the fragments matched, precursor ion matched, and good sequence coverage to help localize PTMs. In this example, an H3.2 proteoform was identified with two PTMs – di-methylation on K9 and methylation on K27. Following the same method, other proteoforms with different PTMs and terminal truncations can be manually validated.

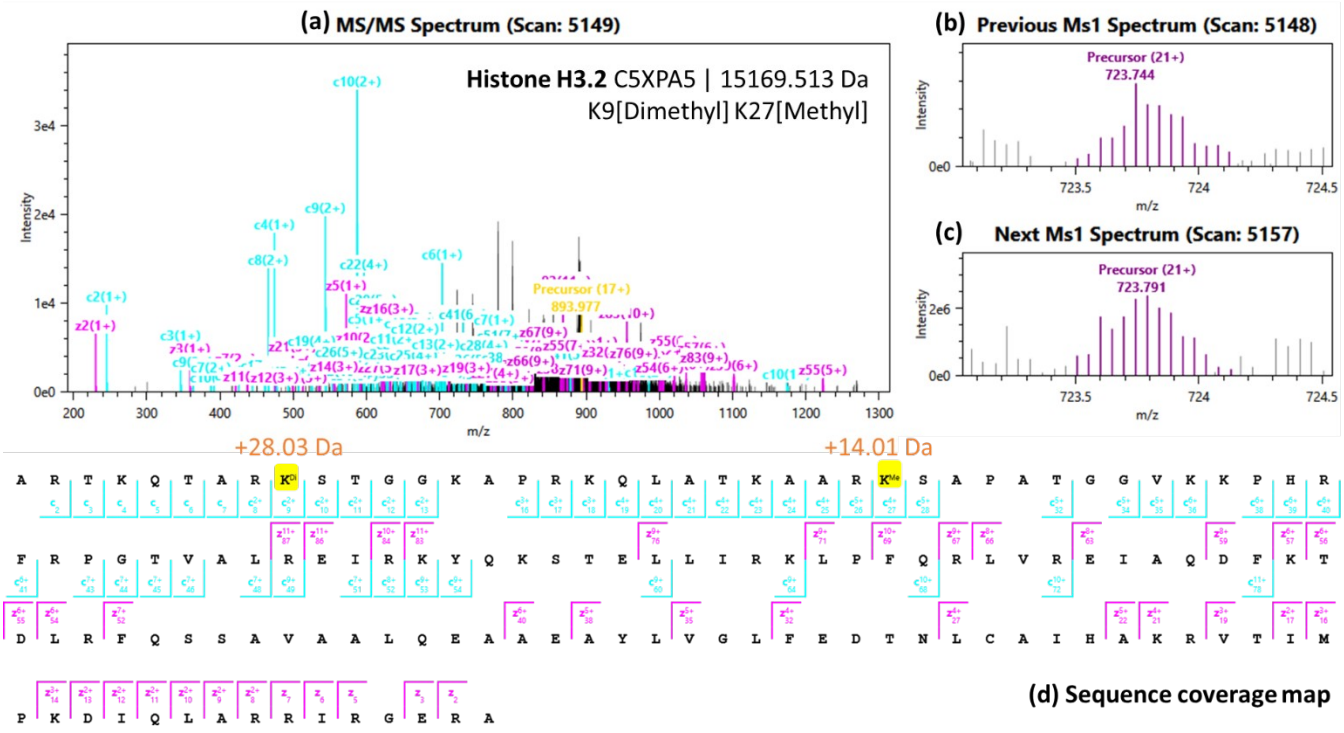

Figure 2: Representative example of an identified histone H3.2 proteoform, with its (a) ETD spectrum, (b) precursor ion in the previous MS1 spectrum, (c) precursor ion in the next MS1 spectrum, and (d) sequence coverage map. The c ions from the N-terminus are labeled in cyan, and the z ions from the C-terminus are in pink. Two PTMs were identified and highlighted in yellow in (d) with their mass shifts annotated.

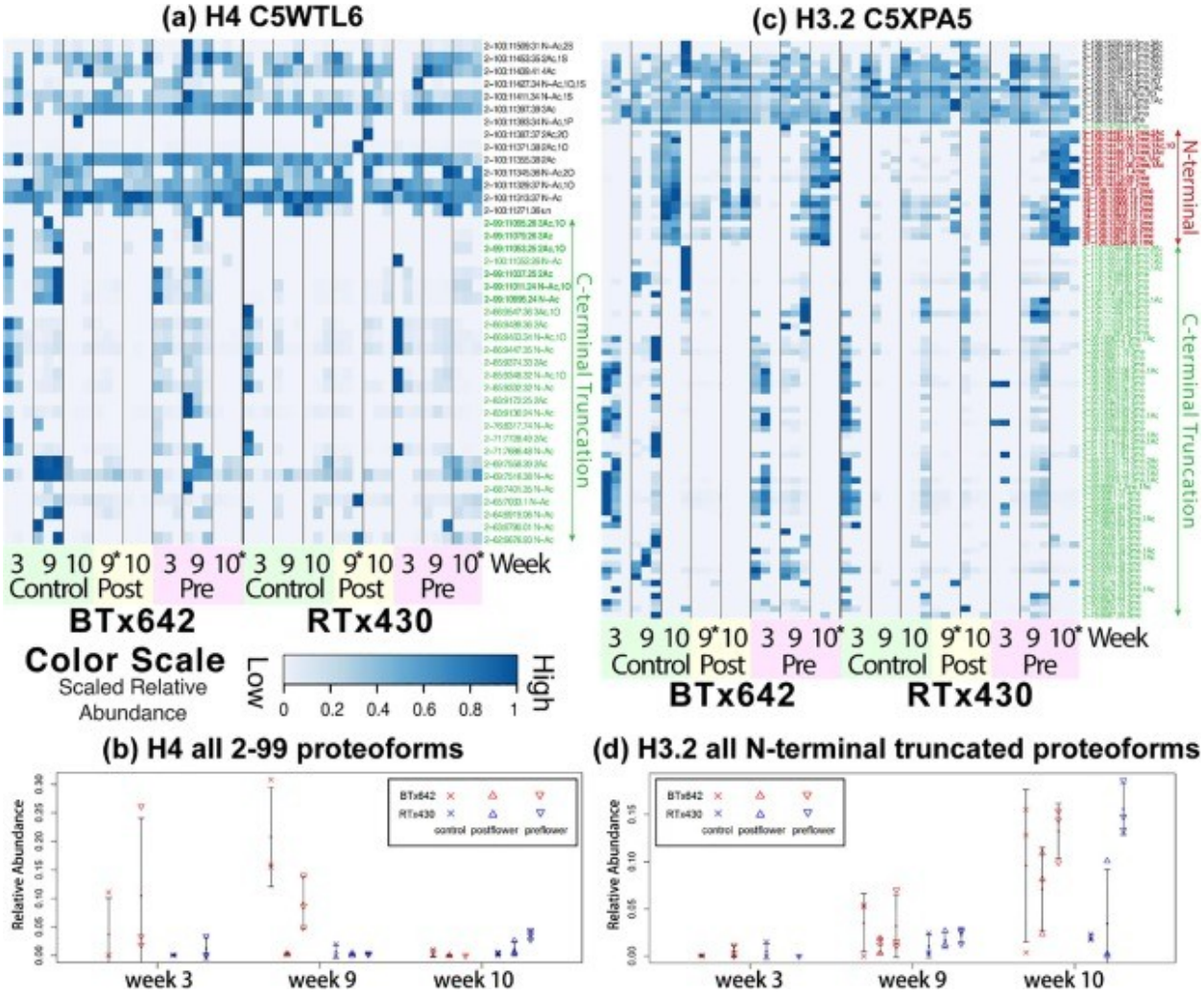

Figure 3: Quantitative comparison of histone proteoforms. (a) Heatmap of histone H4 proteoforms across different samples. For each proteoform, the abundance extracted from top-down MS data was normalized to the sum of all identified H4 proteoforms in each analysis, yielding the "relative abundance". The values were then scaled to the maximum of each row to better show the changes in low abundance proteoforms. The scaled relative abundance is denoted in the color key at the bottom of the heatmap. Growth conditions are noted on the horizontal axis (Pre: pre-flowering drought, Post: post-flowering drought). Three replicates are grouped together, and are separated by black vertical stripes from other conditions. For samples labeled with asterisks only technical replicates were acquired. Proteoforms are represented on the vertical axis, in the format "starting residue – ending residue: mass; putative modification". (b) Relative abundance plot of the truncated H4 proteoforms 2–99 (proteoforms highlighted in bold in (a) are summed) at different conditions. The key to the symbols is shown in the legend in the top right corner. Filled

dots in the middle of the error bars are the average values. (c) Heatmap of H3.2 proteoforms and (d) abundance plot for all identified N-terminal truncated H3.2 are shown in the same format as those for H4. Proteoforms smaller than 8 kDa in (c) were omitted for simplicity. The N-terminal and C-terminal truncated H3.2 proteoforms showed different responses across the growth conditions. See reference[29] for details. Reprinted with permission from ELSEVIER.

Quantitative comparison of the detected histone proteoforms can reveal potential epigenetic markers. We have applied this protocol previously to 48 sorghum samples collected from field (“additional inhibitors” in Table 1-2 were not used in this study).[29] Two different genotypes of sorghum were compared in response to pre-flowering or post-flowering droughts. By comparing the relative abundance of the proteoforms, we discovered some interesting changes of truncated histone proteoforms that are specific to sample conditions as shown in Figure 3. C-terminal truncation of H4 was observed only in week 3 and 9 for some of the samples (Figure 3a-b). For H3.2, N-terminal truncated proteoforms were generally more abundant in week 10 (Figure 3c-d). In contrast, C-terminal truncated H3.2 tend to be seen in earlier time points (Figure 3c). More importantly, the two genotypes did not respond in the exact same way. The H4 C-terminal truncated proteoforms were significantly more abundant in BTx642 than in RTx430 (Figure 3b). Such data reveals potential epigenetic markers of plant development and stress tolerance that can be further tested with other techniques.

Discussion

The presented protocol describes how to extract histones from sorghum leaf (or more generally plant leaf) samples. The average histone yield is expected to be 2-20 µg per 4-5 g sorghum leaf material. The materials are sufficiently pure for the downstream histone analysis by LC-MS (mostly histones with ~20% ribosomal protein contamination). Lower yield may be obtained due to sample variations, or potential mishandling/failures throughout the protocol. Maintaining the integrity of the nuclei before the nuclei lysis step is critical, therefore aggressive vortexing and pipetting should be avoided before adding NLB. In addition, loss of nuclei may occur when removing the supernatants from the pellets. Care must be taken to not disrupt the pellets when pipetting. The Triton X-100 concentration of 1% was optimized to selectively lyse the non-targeted organelles but not the nuclei (step 3.2). Optimal detergent concentration for other tissue or organisms may be different and need to be experimentally determined. Color change of the

supernatant during the filtration process could indicate potential issues such as inefficient release of chloroplast or insufficient grinding of leaf. If possible, use a microscope to check for lysis of chloroplasts and retention of intact nuclei after each step to further optimize the protocol (especially if modifying the protocol for other tissues or plants). This protocol has only been tested with sorghum leaf tissue. It does not work for sorghum root tissue likely due to interference from soil. Application to other plant leaf tissues has not been tested and application to different plants may need additional optimization. For adapting the nuclei isolation protocol for ChIP-seq applications, an additional sucrose gradient density separation after step 3.3.4 (before using NLB) is advised to reduce cytoplasmic contamination. Because of the extensive clean-up steps, small amounts of residual non-nuclei materials are not expected to cause significant interference for histone analysis in LC-MS and can be left with the pellet.

Several initial trials failed when using commercial tablets of phosphatase inhibitors (PhosSTOP, Roche). The supernatant in step 3.1.6 appeared to be intense green when the tablets were used in the extraction buffer. The final extract showed low number of identified histones. We suspect the proprietary ingredients in the tablets may have caused nuclei lysis before step 3.4, reducing the overall histone yield. Another possible reason for failure is the incompatibility of the ingredients in the histone purification step with the ion exchange resin (step 3.4). We have used this protocol to consistently extract high purity histones for subsequent LC-MS over 150 samples. On average we were able to obtain higher yield without using the “additional inhibitors” (unpublished data). Therefore, it is advised to cautiously test new inhibitors when modifying or adapting this protocol for other purposes. If phosphorylation is not of interest, the phosphatase inhibitors can be omitted in the extraction buffers.

The steps in 3.4 can take 3-4 hours or more. It is recommended to break the protocol in two days – freeze the nuclei pellet from step 3.3 and perform the purification on day two (or later). The freeze-thaw cycle may partially help the nuclei lysis. The MWCO filter steps (3.4.7) can be very time consuming but can be easily scaled up by preparing multiple samples in parallel. Do not add the protease inhibitor tablets in step 3.4. Many commercial tablets contain polymers (e.g. polyethene glycol) as fillers, which will interfere with LC-MS analysis. At this step, the most other proteins should have been removed or denatured, so enzyme inhibitors are not critical. However, it is still necessary to keep the samples at 4 °C or frozen to minimize degradation.

Following this protocol, histones can be successfully extracted from sorghum leaves. Histone PTMs can be characterized with LC-MS. The method can be potentially applied to large scale studies for comparing histone PTMs

between different biological samples (e.g. different genotypes, plants grown under different conditions, etc.) as shown by the example data in Figure 3. However, the data processing still requires extensive manual analysis for confidently assigning proteoforms, especially for unexpected (or novel) PTMs. New developments in bioinformatics tools are anticipated to automate the workflow and significantly increase the throughput for large scale studies. Another limitation is that top-down MS method currently cannot easily differentiate many proteoforms of hyper-modified H3 (e.g., multiple sites of mono/di/tri-metlylation and acetylation). The single dimension reversed-phase LC cannot fully separate the different H3 proteoforms. Therefore, the MS2 spectra of H3 will typically contain fragments from multiple proteoforms and cannot be easily and confidently deconvoluted. Combining top-down with bottom-up or middle-down methods[30, 32, 33] can be especially beneficial for characterization of histone H3. Alternatively, multi-dimensional separation can be considered to improve the depth of top-down MS.[34–36]

Histone PTM profiling by LC-MS enables discovery of novel epigenetic markers for designing chromatin modifiers and improve the resilience of plants to severe environmental conditions. A pilot study using sorghum from two cultivars and grown under drought conditions in the field indicated that selective histone terminal clipping in leaf may be related to drought acclimation and plant development.[29] The identified histone markers may serve as targets by complementary techniques such as ChIP-seq. Comprehensive understanding of epigenetic factors gained from these complementary techniques would be indispensable for engineering innovative solutions to crops in response to environmental changes.

ACKNOWLEDGMENTS

We thank Ronald Moore and Tom Fillmore for helping with mass spectrometry experiments, and Matthew Monroe for data deposition. This research was funded by grants from US Department of Energy (DOE) Biological and Environmental Research through the Epigenetic Control of Drought Response in Sorghum (EPICON) project under award number DE-SC0014081, from the US Department of Agriculture (USDA; CRIS 2030-21430-008-00D), and through the Joint BioEnergy Institute (JBEI), a facility sponsored by DOE (Contract DE-AC02-05CH11231) between Lawrence Berkeley National Laboratory and DOE. The research was performed using Environmental Molecular Sciences Laboratory (EMSL) (grid.436923.9), a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research.

References

1. Farooq, M., Wahid, A., Kobayashi, N., Fujita, D., Basra, S.M.A. Plant drought stress: Effects, mechanisms and management. *Sustainable Agriculture*. 153–188, doi: 10.1007/978-90-481-2666-8_12 (2009).

2. Dai, A. Drought under global warming: A review. *Wiley Interdisciplinary Reviews: Climate Change*. **2** (1), 45–65, doi: 10.1002/wcc.81 (2011).

3. Rooney, W.L., Blumenthal, J., Bean, B., Mullet, J.E. Designing sorghum as a dedicated bioenergy feedstock. *Biofuels, Bioproducts and Biorefining*. **1** (2), 147–157, doi: 10.1002/bbb.15 (2007).

4. Mullet, J.E., Klein, R.R., Klein, P.E. Sorghum bicolor - An important species for comparative grass genomics and a source of beneficial genes for agriculture. *Current Opinion in Plant Biology*. **5** (2), 118–121, doi: 10.1016/S1369-5266(02)00232-7 (2002).

5. Xu, L. *et al.* Drought delays development of the sorghum root microbiome and enriches for monoderm bacteria. *Proceedings of the National Academy of Sciences*. **115** (18), E4284–E4293, doi: 10.1073/pnas.1717308115 (2018).

6. Gao, C. *et al.* Strong succession in arbuscular mycorrhizal fungal communities. *ISME Journal*. **13** (1), 214–226, doi: 10.1038/s41396-018-0264-0 (2019).

7. Gao, C. *et al.* Fungal community assembly in drought-stressed sorghum shows stochasticity, selection, and universal ecological dynamics. *Nature Communications*. **11** (1), doi: 10.1038/s41467-019-13913-9 (2020).

8. Bannister, A.J., Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Research*. **21** (3), 381–395, doi: 10.1038/cr.2011.22 (2011).

9. Yuan, L., Liu, X., Luo, M., Yang, S., Wu, K. Involvement of histone modifications in plant abiotic stress responses. *Journal of Integrative Plant Biology*. **55** (10), 892–901, doi: 10.1111/jipb.12060 (2013).

10. Park, P.J. ChIP-seq: Advantages and challenges of a maturing technology. *Nature Reviews Genetics*. **10** (10), 669–680, doi: 10.1038/nrg2641 (2009).

11. Huang, H., Lin, S., Garcia, B.A., Zhao, Y. Quantitative proteomic analysis of histone modifications. *Chemical Reviews*. **115** (6), 2376–2418, doi: 10.1021/cr500491u (2015).

12. Moradian, A., Kalli, A., Sweredoski, M.J., Hess, S. The top-down, middle-down, and bottom-up mass spectrometry approaches for characterization of histone variants and their post-translational

modifications. *Proteomics*. **14** (4–5), 489–497, doi: 10.1002/pmic.201300256 (2014).

13. Sidoli, S., Garcia, B.A. Characterization of individual histone posttranslational modifications and their combinatorial patterns by mass spectrometry-based proteomics strategies. *Methods in Molecular Biology*. **1528**, 121–148, doi: 10.1007/978-1-4939-6630-1_8 (2017).

14. Maile, T.M. *et al.* Mass Spectrometric Quantification of Histone Post-translational Modifications by a Hybrid Chemical Labeling Method. *Molecular & Cellular Proteomics*. **14** (4), 1148–1158, doi: 10.1074/mcp.o114.046573 (2015).

15. Dang, X. *et al.* The first pilot project of the consortium for top-down proteomics: A status report. *Proteomics*. **14** (10), 1130–1140, doi: 10.1002/pmic.201300438 (2014).

16. Schaffer, L. V. *et al.* Identification and Quantification of Proteoforms by Mass Spectrometry. *Proteomics*. **19** (10), 1800361, doi: 10.1002/pmic.201800361 (2019).

17. Cupp-Sutton, K.A., Wu, S. High-throughput quantitative top-down proteomics. *Molecular Omics*. doi: 10.1039/C9MO00154A (2020).

18. Varoquaux, N. *et al.* Transcriptomic analysis of field-droughted sorghum from seedling to maturity reveals biotic and metabolic responses. *Proceedings of the National Academy of Sciences*. **116** (52), 27124 LP – 27132, doi: 10.1073/pnas.1907500116 (2019).

19. Gordon, J.A. Use of vanadate as protein-phosphotyrosine phosphatase inhibitor. *Methods in Enzymology*. **201** (C), 477–482, doi: 10.1016/0076-6879(91)01043-2 (1991).

20. Zhou, M. *et al.* Profiling changes in histone post-translational modifications by top-down mass spectrometry. *Methods in Molecular Biology*. **1507**, 153–168, doi: 10.1007/978-1-4939-6518-2_12 (2017).

21. Chambers, M.C. *et al.* A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology*. **30** (10), 918–920, doi: 10.1038/nbt.2377 (2012).

22. Kou, Q., Xun, L., Liu, X. TopPIC: a software tool for top-down mass spectrometry-based proteoform identification and characterization. *Bioinformatics*. **32** (22), btw398, doi: 10.1093/bioinformatics/btw398 (2016).

23. Park, J. *et al.* Informed-Proteomics: Open-source software package for top-down proteomics. *Nature Methods*. **14** (9), 909–914, doi: 10.1038/nmeth.4388 (2017).

24. LeDuc, R.D., Fellers, R.T., Early, B.P., Greer, J.B., Thomas, P.M., Kelleher,

N.L. The C-Score: A Bayesian Framework to Sharply Improve Proteoform Scoring in High-Throughput Top Down Proteomics. *Journal of Proteome Research*. **13** (7), 3231–3240, doi: 10.1021/pr401277r (2014).

25. Fornelli, L. *et al.* Advancing Top-down Analysis of the Human Proteome Using a Benchtop Quadrupole-Orbitrap Mass Spectrometer. *Journal of Proteome Research*. **16** (2), 609–618, doi: 10.1021/acs.jproteome.6b00698 (2017).

26. Sun, R.-X. *et al.* pTop 1.0: A High-Accuracy and High-Efficiency Search Engine for Intact Protein Identification. *Analytical Chemistry*. **88** (6), 3082–3090, doi: 10.1021/acs.analchem.5b03963 (2016).

27. Xiao, K., Yu, F., Tian, Z. Top-down protein identification using isotopic envelope fingerprinting. *Journal of Proteomics*. **152**, 41–47, doi: 10.1016/j.jprot.2016.10.010 (2017).

28. Cai, W. *et al.* MASH Suite Pro: A Comprehensive Software Tool for Top-Down Proteomics. *Molecular & Cellular Proteomics*. **15** (2), 703–714, doi: 10.1074/mcp.o115.054387 (2016).

29. Zhou, M. *et al.* Top-down mass spectrometry of histone modifications in sorghum reveals potential epigenetic markers for drought acclimation. *Methods*. doi: 10.1016/j.ymeth.2019.10.007 (2019).

30. Garcia, B.A., Pesavento, J.J., Mizzen, C.A., Kelleher, N.L. Pervasive combinatorial modification of histone H3 in human cells. *Nature Methods*. **4** (6), 487–489, doi: 10.1038/nmeth1052 (2007).

31. Zheng, Y. *et al.* Unabridged Analysis of Human Histone H3 by Differential Top-Down Mass Spectrometry Reveals Hypermethylated Proteoforms from MMSET/NSD2 Overexpression. *Molecular & Cellular Proteomics*. **15** (3), 776–790, doi: 10.1074/mcp.m115.053819 (2016).

32. Garcia, B.A. *et al.* Chemical derivatization of histones for facilitated analysis by mass spectrometry. *Nature Protocols*. **2** (4), 933–938, doi: 10.1038/nprot.2007.106 (2007).

33. Holt, M. V., Wang, T., Young, N.L. One-Pot Quantitative Top- and Middle-Down Analysis of GluC-Digested Histone H4. *Journal of the American Society for Mass Spectrometry*. **30** (12), 2514–2525, doi: 10.1007/s13361-019-02219-1 (2019).

34. Tian, Z. *et al.* Enhanced top-down characterization of histone post-translational modifications. *Genome Biology*. **13** (10), R86, doi: 10.1186/gb-2012-13-10-R86 (2012).

35. Wang, Z., Ma, H., Smith, K., Wu, S. Two-dimensional separation using high-pH and low-pH reversed phase liquid chromatography for top-down proteomics. *International Journal of Mass Spectrometry*. **427**, 43–51, doi: 10.1016/j.ijms.2017.09.001 (2018).

36. Gargano, A.F.G. *et al.* Increasing the Separation Capacity of Intact Histone Proteoforms Chromatography Coupling Online Weak Cation Exchange-HILIC to Reversed Phase LC UVPD-HRMS. *Journal of Proteome Research*. **17** (11), 3791–3800, doi: 10.1021/acs.jproteome.8b00458 (2018).