Supporting Information For:

An Improved Top-Down Mass Spectrometry Characterization of *Chlamydomonas reinhardtii* Histones and Their Post-Translational Modifications

Sarah R. Rommelfanger\(^1\), Mowei Zhou\(^3\), Henna Shaghasi\(^4\), Shin-Cheng Tzeng\(^1\), Bradley S. Evans\(^1\), Ljiljana Paša-Tolić\(^3\), James G. Umen\(^1\),\(^2\), James J. Pesavento\(^4\)*

*Corresponding author*

\(^1\)Donald Danforth Plant Science Center, St. Louis, MO 63132, USA

\(^2\)Washington University in St. Louis, St. Louis, MO 63130, USA

\(^3\)Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA 99354, USA

\(^4\)Saint Mary’s College of California, Moraga, CA 94575, USA

**Address reprint requests to:**

Jim Pesavento
Saint Mary’s College of California
Department of Biology
1928 St. Mary’s Rd.
Moraga, CA, 94608
Ph: 925-631-4430
Email: jjp6@stmarys-ca.edu

The contents of this file ‘Supplemental File 1’ include:
- Expanded TDMS methodology with sample calculation
- Supplemental methods for Figures S1 and S2
- Figure S1
- Figure S2
- Figure S3
- Figure S4
- Figure S5
- Figure S6
- Figure S7
Supplemental Methods

Expanded description of TDMS data analysis workflow, with example calculation

The following section provides additional information on the data analysis and workflow for characterizing and quantitating Chlamydomonas histone proteoforms. We have chosen oxidized monoacetylated histone H4 to serve as an example.

Step 1.

**Figure S0a. Summing all mass spectra containing oxidized histone H4 from the LC-MS/MS run.** Oxidized histone H4 has an elution window that does not overlap with that of unoxidized histone H4, therefore we analyze each separately. In the example above, we sum spectra that include masses close to where histone H4 elutes, usually up to +400 Da above the theoretical unmodified mass. Acetylation is found to have a significant influence on elution time. The summed scans are exported as a "*.raw" file.

Step 2.

**Figure S0b. Importing summed spectra into FreeStyle™ for deconvolution.** The .raw file exported from Step 1 is imported into FreeStyle and deconvoluted. We then export the monoisotopic masses identified within this elution window with their relative intensities as a .csv file for quantitating global abundance values in later steps.
Step 3. Conversion of LCMS .raw file into .mzML and .MSALIGN files

The MS/MS data from the LCMS .raw file is then converted into an .MSALIGN text file that includes precursor masses, scan numbers, fragment ion masses, and their intensities. The .raw file is loaded into MSConvert and processed into an .mzML file using the default parameters and no filters. The .mzML file is then converted to an .MSALIGN file, usually through FLASHDeconv with the command-line script:

```
FLASHDeconv -out Y -max_charge 30 -max_mass 50000 -promex_out 1 -topfd_out 1 -in X
```

(where ‘Y’ is the output path and ‘X’ is the file path for the input .mzML file)

Step 4.

![Image](C:/Users/im/Desktop/Auxiliary Figures/TDFG2020/111/1/SF_11960230_targeted.png)

**Figure S0c. Searching for precursor masses of interest using the SMC (Search MS and Combine MS/MS) Python script.** The .MSALIGN file is imported into SMC. The targeted precursor mass of 11432 is searched between scan numbers 4176-4351 (as determined in Step 1) with a mass tolerance of 2 Da. After clicking ‘Run’, the program generates a .txt file saved in the same directory as SMC, and a pop-up with the same information for quick access to copy and paste these values. In this case, the program found that scans 4249, 4255 and 4262 have MS/MS data on the 11432 Da precursor mass.

Step 5.
First, the presence of Nα-ac and M84ox is confirmed by seeing good alignment of fragment masses to manually-entered reference proteoforms containing these expected modifications. Then, separate reference proteoforms are created also containing acetylation isomers at K5, K8, K12, K16 and K20, and alignment of fragment masses are assessed for each isomer. In this way, TDValidator quickly shows the presence of isomers by color-coding unique fragment ions in the MS/MS spectrum to its respective reference proteoform. The bottom panel here shows an unmodified c5 ion (K5 unacetylated) in black, and an acetylated c5 ion (K5ac) in blue, at approximately equal relative abundances. The summed MS/MS spectra are then exported as a .csv file containing m/z values and their respective intensities.

Step 6. Calculating fragment ion relative ratios (FIRRs) from summed MS/MS data
The summed MS/MS data contains the m/z and intensity values only. Therefore, manual analysis must include a target list of known fragment ion m/z that report on each isomeric proteoform. To calculate the FIRRs for each proteoform from the summed spectra, the abundance of the monoisotopic peaks from z = 1+ and/or 2+ fragment ions is used. For example, the monoisotopic ion intensity of the c51+ ion of unacetylated K5 is found to be 2463.234, while the c51+ ion of K5ac is found to be 2603.766. Thus, the FIRR for this pair is 51% acetylated and 49% unacetylated.
To expedite the data analysis process, we added another feature to SMC that takes a proteoform search list (includes fragment ion masses that report on site/type of PTM defined by the user) and the TDValidator .csv output from Step 5. More information about this feature can be found in the SMC user guide (www.github.com/pesavent/SMC). In the image shown above, a spreadsheet program is used to generate FIRRs for each pair of ions found within the TDValidator .csv file. Once all the FIRRs are generated, the average of each FIRR that reports on a single modified amino acid is taken. For example, the FIRRs that all report on K5ac (from ions c5, c6 and c7) are averaged to generate a final FIRR of 47.3% of K5ac. This process is repeated for each modifiable site (K8, K12 and K16), accounting for the contribution of intensities from other isomers present (i.e., K5ac FIRR must be subtracted from K8ac FIRR). In some cases, the FIRRs were excluded because of low abundance and/or overlapping isotopic distributions (e.g., the strikethrough values for the c11+ and c15+ ions above). By the c16 ion, all of the mass difference is accounted for, indicating K16ac is the last detectable acetylation site in this dataset.

Once the FIRRs for each proteoform are calculated, the values are multiplied by the intact protein intensity relative ratios (PIRRs) that were generated by deconvoluting the summed LCMS spectra (Step 2).

### Table S0a. Calculating the global abundance of monoacetylated histone H4 proteoforms. The global abundance of the monoacetylated from for this particular LCMS run is found to be 23.1%, which is used to calculate the global abundance of each proteoform.
Step 8.

Table S0b. Averaging monoacetylated histone H4 proteoforms across multiple LCMS runs. For three biological replicates of asynchronous wild type Chlamydomonas (strain 21gr), the global abundance values are averaged, and the standard deviation of a population for each is calculated.

| Monoacetyl (n=3) | Abundance (%) | Std. Dev (%) |
|-----------------|---------------|--------------|
| K5ac            | 9.0           | 1.4          |
| K8ac            | 1.1           | 0.8          |
| K12ac           | 6.8           | 1.5          |
| K16ac           | 3.6           | 1.2          |

Table S0c. Final format of the data. This data is represented in Table 2 in the main text of the manuscript.

DAPI-staining and microscopy of enriched nuclei

Integrity of the isolated nuclei was assessed by staining with 4',6-diamidino-2-phenylindole (DAPI) and visualization using epifluorescence microscopy. 12.5 µL aliquots of nuclei-enriched pellets (~10% of sample) were reserved after the nuclei isolation protocol and washing, and before freezing and subsequent histone extraction. These nuclei aliquots were mixed with 500 µL of NIB containing 5 µg/mL of DAPI, incubated at room temperature for 10 minutes, then cells were concentrated by centrifuging for 10-30 seconds at <2,000 x g in a benchtop mini-centrifuge and discarding most of the supernatant, leaving approximately 20 µL of liquid in which to resuspend the nuclei. Paraformaldehyde (PFA) fixed whole cells, and cells that were frozen and macerated but not detergent-lysed, were also stained with DAPI in this manner.

Cells were transferred to a microscope slide and imaged using a Leica DMI 6000 B (Leica, Germany) microscope using a 40 x objective lens. DIC optics were used for light images. Epifluorescent images were taken using a metal halide excitation source (Sutter Instruments), a Leica A4 filter cube (a/b excitation, c dichroic, d/e emission) for detection of DAPI, and TX2 filter cube (a/b excitation, c dichroic, d/e emission) for chlorophyll, and detection using a Roper Coolsnap HQ2 CCD camera. Diameters of cells and nuclei were calculated relative to the size of the 10 µm overlay on each image.

SDS-PAGE, H3 immunoblot, and silver stain

Yield and purity of histone extracts were estimated by anti-histone H3 immunoblots and silver staining of SDS-PAGE gels, respectively. One percent (1%) of each water-soluble histone extract was loaded per lane, alongside standard dilutions of calf thymus histones (Sigma, H9250), in a 15% SDS-PAGE gel with 6% stacking gel, and separated at 80 V for 30 minutes, then 120 V for 1 hour. For the H3 immunoblot, the proteins were transferred from the SDS-PAGE gel to a polyvinylidene difluoride (PVDF) membrane at 50 V for one hour at 4 °C in a wet transfer chamber containing transfer buffer (25 mM Tris-HCl (pH 7.6), 192 mM glycine, 20% v/v methanol, 0.03% w/v sodium dodecyl sulfate (SDS)). Membrane was blocked for 30 minutes at room temperature in a solution of 5% w/v powdered fat-free milk in Tris-buffered saline with 0.1% Tween-20 (TBST), washed 3x 10 minutes in TBST, incubated overnight at 4 °C with the histone H3 polyclonal primary antibody (Invitrogen, PA5-16183) at 1:10,000 dilution in 5% milk in TBST, washed again 3x 10 minutes in TBST, incubated 1 hour at room temperature with the polyclonal HRP-conjugated goat-anti-rabbit secondary antibody (Invitrogen, PI-31460) at 1:5,000 dilution in 5% milk in TBST, and washed again 3x 10 minutes in TBST. Immobilon Forte Western HRP chemiluminescent
substrate (Millipore, WBLUF0100) was added to the membrane for 90 seconds, then membrane was imaged with a BioRad GelDoc XR+ Imaging System. H3 signal was quantified using the BioRad ImageLab 6.1 software, and approximate histone yield per sample was calculated using a standard curve equation based on the H3 signal from the calf thymus histone dilution lanes.

Total protein was visualized by silver staining to roughly assess the proportion of non-histone proteins in each extract, and to verify the presence of strong bands for histone-sized proteins in each extract. SDS-PAGE gels were silver-stained according to the “Short silver nitrate staining” protocol detailed in Chevallet et al. 2006, Nature Protocols, with the modification of increasing the ethanol concentration in the fixing solution from 30% v/v to 40% v/v ethanol.

Chevallet, M., Luche, S. & Rabilloud, T. Silver staining of proteins in polyacrylamide gels. Nat Protoc 1, 1852–1858 (2006). https://doi.org/10.1038/nprot.2006.288

Supplemental Figures

**Figure S1.** Overlay of light-microscopy and fluorescence microscopy of DAPI-stained samples of (A) Chlamydomonas PFA-fixed whole cells, (B) Chlamydomonas cells subjected to one pre-maceration detergent treatment with NIBA + 5% Triton, and (C) Chlamydomonas cells subjected to entire nuclei enrichment protocol including pre-maceration detergent treatment, frozen tissue grinding, post-maceration detergent treatment, and washing. Greyscale channel is differential interference contrast (DIC) light microscopy, red channel is chlorophyll autofluorescence, and blue channel is DAPI-stained chromatin signal. The nuclei-enriched sample (C) shows DAPI-signal-emitting particles of the same size as nuclei seen inside whole cells (A), suggesting that our nuclei-enriched pellets indeed contain intact isolated nuclei. The presence in panels B and C of nuclear DAPI signal inside a whole-cell-sized particle that lacks chlorophyll signal suggests that the detergent treatment partially, and the full nuclei-enrichment procedure fully, abolishes chlorophyll signal by damaging the outer cell wall allowing the plasma membrane and chloroplast membrane to be solubilized by the detergent. The presence in panel B of seemingly intact whole cells containing both DAPI-signal and chlorophyll autofluorescence signal suggests that the detergent treatment alone is insufficient to damage the cell wall and plasma membrane sufficiently that the nuclei become accessible for nuclei extraction. The larger clusters of DAPI-stained particles in
panel C are likely aggregates of nuclei adhered together via strands of chromatin leaking out of the nuclei.

**Figure S2.** Immunoblot (upper panel) and silver stain (lower panel) of Chlamydomonas histone extracts separated by SDS-PAGE. Lanes A and B contain biological replicates of Chlamydomonas histone extracts, with $1 \times 10^7$ cell equivalents (1% of sample) per lane. Numbered lanes indicate the nanograms per lane for reference dilutions of calf thymus histones (Sigma, H9250) and a protein size marker (lane M, Gold Bio, BLUEstain2 protein ladder, P008). Identically loaded 15% SDS-PAGE gels were subject to either anti-histone H3 immunoblot (upper panels) to quantify histone signal intensity for estimation of yield, or silver stain (lower panels) to visualize total protein for assessment of extract complexity. Linear regression equation based on calf thymus histone H3-blot band intensities was used to calculate estimated total histone mass in Chlamydomonas histone extracts to normalize loading to TDMS. Chlamydomonas histone bands were identified based on similar migration to calf thymus histones. Pairs of panels shown are from the same gel, with white space indicating lanes cropped out.
Figure S3. (A) Summed MS1 spectrum of histone H4 from Chlamydomonas showing multiply-modified proteoforms, their monoisotopic masses, mass shifts (Δm), and corresponding relative abundance. (B) The c₃ ion (left) and c₁₃ ion (right) in the ETD fragmentation data for H4 show ~30% H4R₃me₁ and ~70% unmodified H4R₃ due to co-fragmentation of nearly isobaric acH4R₃me₁K₇₉me₁M₈₄ox and acH4K₇₉me₁M₈₄ox₂.
**Figure S4.** Multiple sequence alignment (MSA) of Chlamydomonas H2A protein sequences, gene IDs, monoisotopic masses (without N-terminal methionine), and standardized histone protein nomenclature. Below the MSA are descriptive characteristics of regions within H2A’s alignment that report on conservation (histogram of physiochemical properties), quality (mutation likelihood), consensus (percentage of modal residue per amino acid position), and occupancy (presence of an amino acid at that position). For more information visit https://www.ebi.ac.uk/Tools/msa/clustalo/.
Figure S5. H2A reproducibility and graphical fragmentation maps (GFMs) of ΔNα-ac and C-terminally truncated H2A.v proteoforms. H2A.v (left column) and H2A.0-4 (right column) MS from 4 biological replicates (rows). The top two spectra are from Figure 4 in the main text. Boxed isotopes correspond to H2A proteoforms that were fragmented by ETD and are numbered to match the GFM below (intact H2A.Z (#3), H2A.Z C-terminal truncations (#1, #2), and removal of N-terminal acetylation on H2A.1 (#4)).
Figure S6. (A) MS spectrum for monoubiquitylated canonical H2B variants. For reference, the 25050.64 Da species was confirmed to be H2B.5ub1, while the 25211.67 Da is expected to be H2B.14ub1 (see main text Figure 5A, right MS). (B) Summed HCD spectra of precursor masses ranging from 25050 to 25096 Da. Fragment ions in red align to the sequence of H2B.5, while fragment ions in blue align to the sequence of Chlamydomonas ubiquitin.
**A. Table of unoxidized H3 proteoforms identified**

| Feature | Monoisotopic mass (Daltons) | Mass shift from H3.1 | PTM combinations on proteoform (low confidence assignments shown as mass shifts and residue range) | Percent total feature intensity |
|---------|-----------------------------|---------------------|------------------------------------------------------------------------------------------------|------------------------------|
| a       | 15,168.49                   | - 1 Da              | unmodified H3.1                                                                                   | 7.26%                        |
| b       | 15,182.51                   | + 13 Da             | K4/9me1 (other sites possible at lower abundance)                                                  | 23.90%                       |
| c       | 15,198.54                   | + 29 Da             | K4me1 + 14 Da (K35 - Q86)                                                                         | 19.61%                       |
| d       | 15,212.52                   | + 43 Da             | K4/9me1 + 28 Da (K27 - M119)                                                                     | 30.67%                       |
| e       | 15,225.52                   | + 56 Da             | K4me1 + 42 Da (T28 - Y40)                                                                         | 13.78%                       |
| f       | 15,239.54                   | + 70 Da             | K4me1 + 56 Da (P37 - F57)                                                                        | 1.58%                        |
| g       | 15,253.53                   | + 84 Da             | K4me1 + 71 Da (L20 - I118)                                                                        | 0.37%                        |
| h       | 15,267.53                   | + 98 Da             | K4me1 + 84 Da (T28 - T117)                                                                       | 0.55%                        |
| i       | 15,281.52                   | + 112 Da            | K4me1 + 98 Da (K23 - L99)                                                                        | 0.26%                        |
| j       | 15,295.52                   | + 126 Da            | K4me1 + 112 Da (A23 - R115)                                                                      | 0.12%                        |
| k       | 15,308.54                   | + 139 Da            | K4me1 + 125 Da (Q54 - F77)                                                                        | 0.19%                        |
| l       | 15,323.52                   | + 154 Da            | K4me1 + 140 Da (K23 - T44)                                                                        | 0.04%                        |
| m       | 15,338.52                   | + 169 Da            | K4me1 + 155 Da (K23 - D105)                                                                       | 0.05%                        |
| n       | 15,350.55                   | + 181 Da            | K4me1 + 167 Da (A25 - F83)                                                                        | 0.09%                        |
| o       | 15,365.51                   | + 196 Da            | K4me1 + K18ac + 140 Da (Q19 - L91)                                                                | 0.11%                        |
| p       | 15,380.56                   | + 211 Da            | K9me1 + 197 Da (K23 - H118)                                                                       | 0.02%                        |
| q       | 15,394.57                   | + 225 Da            | K4/9me1 + 211 Da (G32 - F103)                                                                     | 0.15%                        |
| r       | 15,407.57                   | + 238 Da            | K4me3 + K9ac + K14ac + K18ac + 70 Da (Q19 - I118)                                                | 0.20%                        |
| s       | 15,423.58                   | + 254 Da            | K4me3 + K9ac + K14ac + 128 Da (Q23 - T79)                                                         | 0.15%                        |
| t       | 15,438.58                   | + 269 Da            | K4/9me1 + 255 Da unlocalized                                                                     | 0.27%                        |
| u       | 15,452.59                   | + 283 Da            | K4me3 + K9ac + K14ac + K18ac + 115 Da (Q19 - F77)                                                | 0.34%                        |
| v       | 15,466.61                   | + 297 Da            | K4me3 + 174 Da (Q19 - I118) or K4me1 + 283 Da (Q19 - I118)                                       | 0.28%                        |

**B. Summed MS1 spectra of all unoxidized H3 proteoforms**

![Figure S7](image_url)  

**Figure S7.** (A) Unoxidized H3 proteoforms detected in a single LCMS run with high-mass H3 proteoforms targeted for inclusion. Feature letters correspond to the features illustrated in Figure 6. PTM combinations on proteoform lists high-confidence PTM assignments, along with the remaining mass shifts that could not be assigned to individual residues, and the range of residues on which these mass shifts may occur. Percent total feature intensities are calculated as the intensity of features corresponding to the listed precursor monoisotopic mass as a percentage of the total intensity of all unoxidized H3 features from this single LCMS run. (B) Summed MS1 spectrum for all LC scans containing unoxidized H3 precursor masses, with feature letters oriented above the most abundant mass (not listed). The proteoform labels, monoisotopic masses, and percent intensities listed here are identical to those presented in Figure 6. Fragment ions for most precursor masses showed a predominance of K4me1, while proteoforms r, s, u, and v (bolded and underlined), contained predominantly K4me3 fragments.