TH
20
OPEN ACCESS
ANNIVERSARY

OXFORD

Direct histone proteoform profiling of the unannotated, endangered coral *Acropora cervicornis*

Cassandra N. Fuller[1], Sabrina Mansoor[2], Kevin Jeanne Dit Fouque[1], Lilian Valadares Tose[1],
Javier Rodriguez-Casariego[2,3], Mariangela Kosmopoulou[4], Detlev Suckau[5],
Francisca N. de Luna Vitorino[6], Benjamin A. Garcia[6], Jose M. Eirin-Lopez [2],
Francisco Fernandez-Lima [1,7,*]

[1]Department of Chemistry and Biochemistry, Florida International University, Miami, FL 33199, United States
[2]Environmental Epigenetics Laboratory, Institute of Environment, Florida International University, Miami, FL 33199, United States
[3]Department of Marine Biology and Ecology, Rosenstiel School, University of Miami, Miami, FL 33124, United States
[4]Fasmatech Science & Technology, TESPA Lefkippos, NCSR Demokritos, Agia Paraskevi, Athens 15341, Greece
[5]Bruker Daltonics GmbH & Co. KG, Bremen 28359, Germany
[6]Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, MO 63110, United States
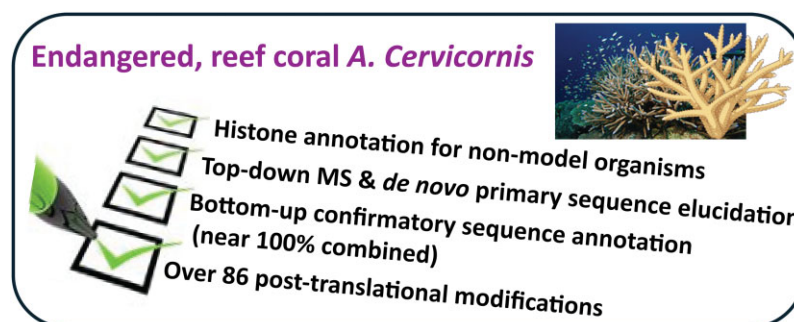[7]Biomolecular Sciences Institute, Florida International University, Miami, FL 33199, United States
*To whom correspondence should be addressed. Email: fernandf@fiu.edu

Abstract

Epigenetic modifications directly regulate the patterns of gene expression by altering DNA accessibility and chromatin structure. A knowledge gap is presented by the need to directly measure these modifications, especially for unannotated organisms with unknown primary histone sequences. In the present work, we developed and applied a novel workflow for identifying and annotating histone proteoforms directly from mass spectrometry-based measurements for the endangered Caribbean coral *Acropora cervicornis*. Combining high-accuracy *de novo* top-down and bottom-up analysis based on tandem liquid chromatography, trapped ion mobility spectrometry, non-ergodic electron-based fragmentation, and high-resolution mass spectrometry, near complete primary sequence (up to 99%) and over 86 post-translational modification annotations were obtained from pull-down histone fractions. In the absence of reliable genome annotations, H2A, H2B, and H4 histone sequences and the annotation of the post-translational modifications of the stressed *A. cervicornis* coral allow for a better understanding of chromatin remodeling and new strategies for targeting intervention and restoration of endangered reef corals.

Graphical abstract

Endangered, reef coral *A. Cervicornis*

- Histone annotation for non-model organisms
- Top-down MS & *de novo* primary sequence elucidation
- Bottom-up confirmatory sequence annotation (near 100% combined)
- Over 86 post-translational modifications

Introduction

Epigenetic modifications stand out due to their responsiveness to environmental signals and reversibility as potential tools for phenotype manipulation and enhancement in endangered species threatened by climate change. Corals have evolved remarkable phenotypic plasticity [1], involving genetic and non-genetic mechanisms that allow these organisms to acclimate and maximize their survival in rapidly changing environments [2–4]. The endangered staghorn coral (*Acropora cervicornis*) is one of the main builders of the structural base of

Caribbean coral reefs, providing extensive habitats for marine organisms and contributing to coastal protection [5]. Coral reefs in general, and *A. cervicornis* in particular, are uniquely susceptible to rapid variations in environmental conditions resulting in a dramatic decline of their populations under current global climate change conditions [6, 7], despite extensive restoration efforts.

Epigenetic reprogramming via DNA methylation has been the main focus of attention as a target for intervention [8], leaving other critical mechanisms (notably those

intimately involved in chromatin remodeling) and their inter-actions understudied [9]. Histone protein variations and their post-translational modifications (PTMs) represent a critical epigenetic mechanism modulating gene function and DNA metabolism through structural changes at the chromatin level [10, 11]. Histones are a diverse family of basic proteins that bind to the DNA, forming the nucleosome, a highly dynamic chromatin fundamental subunit [12]. The affinity of histones for DNA is critically influenced by chemical PTMs in the amino-terminal tails (e.g. acetylation [ac], ubiquitination [ub], phosphorylation [ph], and methylation [me$_{1–3}$]), facilitating the activation/repression of different genome regions in re-sponse to environmental signals [10, 13]. Despite its funda-mental role in epigenetic regulation, studies addressing the epigenetics of chromatin structural components (e.g. histone variants and PTMs) in corals are scarce [9, 14]. The purifi-cation of high-quality histones and the lack of specific anti-bodies remain the main challenges hampering coral epigenetic studies. The lack of appropriate reference information on his-tone variant sequences and occurring chemical modifications in proteins constitutes a strong limiting factor. Consequently, a major knowledge gap related to coral's chromatin composi-tion and dynamics remains to be addressed.

Mass spectrometry (MS) methods have shown potential for the analysis and identification of histones, especially due to continuous improvements in the mass resolving power and mass accuracy over the last decades [15]. Traditional bottom-up MS proteomics is the most widely used method for histone PTM analysis [16–18], which has benefitted largely from the commercial integration of non-ergodic electron-based disso-ciation methods (e.g. electron transfer dissociation, ETD; and electron capture dissociation, ECD) [19], and more recently, from the mobility separation of positional isomers before tan-dem MS/MS [20–22]. A major trade-off from histone bottom-up analysis is the need for derivatization methods to produce longer and more informative proteolytic peptides [23–26], in addition to the possibility of losing information about the protein during digestion. Middle-down MS proteomics ap-proaches using enzymes that only partially digest the histones have shown promise in the characterization of histone tails with varying PTMs [27]; nevertheless, the high isomeric con-tent requires the use of complementary pre-separation and electron/UV-based fragmentation for more informative se-quencing and PTM localization [28–32]. Top-down MS pro-teomics has been effective for the annotation of histone pro-teoforms [33–44]; however, it requires pre-fractionation steps and ultra-high resolution to account for histone chemical di-versity and complexity [45–47]. Regioisomers ubiquitously co-exist in cells and will be ideally resolved before the MS step due to challenges associated with disentanglement of tandem MS/MS from more than two concurrent variants [48], despite recent efforts using fragment correlation MS [49]. Top "dou-ble down" MS includes mobility separation steps to achieve isomeric separation and reduce the spectral complexity lead-ing to better sequence coverage and characterization of his-tone proteoforms [36]. While sub-stoichiometric histone mod-ifications render epigenetic studies more difficult than general proteome analyses from the sensitivity perspective, an added challenge arises from the characterization of non-annotated species.

In this work, a new workflow capable of direct protein sequence generation and proteoform PTM annotation based on liquid chromatography (online and offline), trapped ion mobility spectrometry (TIMS), ECD fragmentation, and high-resolution MS is described and applied to the unannotated, en-dangered Caribbean coral *Acropora cervicornis*. The method combines histone pull-down extraction followed by liquid chromatography (online and offline), top-down mobility- and mass-precursor ion isolation before non-ergodic electron-based fragmentation and high-resolution molecular fragment detection. The *de novo* processing of the top-down data, com-bined with parallel bottom-up sequence confirmation, allows for direct full primary sequence determination and PTM an-notation of histone proteoforms. When applied to the endan-gered staghorn coral *Acropora cervicornis*, a high diversity of PTMs based on mono-, di-, and tri-methylation, oxidation, and acetylation was detected. The *de novo* top-down analysis provided high sequence coverage (∼90%) for all the proteo-forms, and when combined with bottom-up confirmation, full sequence coverages were obtained. Data showed that the high diversity of proteoforms (e.g. isomeric species) and depth of the analysis requires the use of complementary separations, not accessible using traditional proteomic analysis.

Materials and methods

Endangered coral histone extraction

Three coral colonies (*Acropora cervicornis* genets) were ob-tained from the Coral Reef Foundation nursery located at Tav-ernier, Florida Keys (N 24.982715°, W -80.436286°) and were propagated under CRF's permit #FKNMS-2019-012-A2. His-tones were isolated from 12 g of powdered coral fragments, pooled from propagated colonies that were exposed to nu-trient stress. Briefly, flash-frozen coral fragments were man-ually powdered in liquid nitrogen using a mortar and pes-tle. Coral tissue was separated from the calcium carbonate skeleton by suspending powdered coral (∼250 mg/mL) in pre-chilled 1× Pre-lysis buffer (EpiQuik™ Total Histone Extrac-tion Kit, EpigenTek, Farmingdale, NY) on ice for 10 min [50]. The tissue slurry was transferred to a 2 mL Dounce homog-enizer and the EpiQuik™ Total Histone Extraction Kit was used (see details in Supplementary Fig. S1). The supernatant fraction containing the acid-soluble proteins was subjected to acetone precipitation overnight [51]. The coral histone protein concentration and quality were assessed using the Qubit Pro-tein Assay Kit (Invitrogen: Carlsbad, CA) and sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE), re-spectively. Samples were resuspended in molecular-grade, ster-ile water and stored at −80°C.

Top-down MS coral proteoform analysis

The analysis consisted of liquid chromatography separation (online and offline) coupled with TIMS and tandem ECD-MS/MS. Details on the experimental LC-TIMS-q-ECD-TOF MS/MS platform based on a Bruker Maxis Impact II ToF MS (Bruker Daltonics Inc., Billerica, MA) can be found elsewhere (Supplementary Fig. S2) [22]. Liquid chromatography was carried out on a Dionex Ultimate 3000 LC system equipped with an XBridge peptide BEH C$_{18}$ column (4.6 × 250 mm × 5 μm) maintained at a temperature of 50°C. The mobile phase consisted of (A) 5% acetonitrile (ACN) with 0.2% tri-fluoroacetic acid (TFA) and (B) 95% ACN with 0.2% TFA with the following gradient: (i) 0–5 min with 5% B, (ii) 5–6 min to 30% B, (iii) 6–25 min to 40% B, (iv) 25–100 min to 55% B, (v) 100–102 min to 100% B, (vi) 102–107 min hold

100% B, (vii) 107–108 min to 5% B, and (viii) 108–113 min hold 5% B. The flow rate was kept at 0.8 mL/min for the duration of the 113 min gradient and each injection consisted of 100 μL of 2 mg/mL total histones. For offline LC analysis, fractions were collected, dried, and resuspended in a 1:1 methanol:water buffer with 0.1% formic acid.

An Apolo II ESI source operated with a 0.4 ml/min flow rate in positive mode was used for online LC analysis. The ESI source nitrogen nebulizer, dry gas, and temperature were 3 bar, 10 L/min, and 225°C with capillary and end plate voltages of 4500 and 500 V, respectively. A custom-built pulled tip nanoESI source, with similar performance to commercial nanoESI sources, was used for direct infusion analysis (offline LC) at 600–1200 V mounted on an XYZ stage. TIMS fundamentals and principles of operation have been previously described [52–54]. Briefly, mobility experiments were carried out using nitrogen (N_2) as the buffer gas at ambient temperature with a gas velocity defined by the TIMS funnel pressure differences ($P_1 = 2.5$ mbar and $P_2 = 0.68$ mbar). Trapped ions were radially confined using a rf voltage of 250 Vpp at 880 kHz. Deflector and base voltages of 135 and 60 V and a mobility ramp of −150 to −15 V over 100 ms were used. A custom-built 19 mm long electro-magnetostatic cell (EMS from e-MSion Inc., Corvallis, OR) was used [22, 29]. A 2.5 A EMS filament current was used and the collision cell was operated with high-purity argon (oxygen-free). While the experiments described were performed on a custom-built tims-qTOF MS platform, alternative capabilities are available using the integration of front-end IMS analyzers and ExD capabilities using third-party MS companies. Additional information regarding MS instrument operation can be found in Supplementary Fig. S3 in the supporting information.

Histone proteoform derivatization for bottom-up MS

Bottom-up MS histone samples were derivatized using propionic anhydride, as previously described (Supplementary Fig. S4) [21, 24, 25, 55, 56]. Briefly, dried extracted histones were reconstituted in 100 mM NH_4CO_3 (pH 8) to 1 μg/μL. Propionylation reagent (1:3 v/v propionic anhydride in ACN) was added to the samples (1:4 v/v), followed quickly by NH_4OH to maintain a pH of 8. The reaction was incubated at room temperature for 15 min, then repeated before drying the samples in a vacuum centrifuge. Derivatized samples were reconstituted in 100 mM NH_4CO_3 to a final concentration of 1 μg/μL and enzymatically digested using trypsin (1:10 wt/wt) [56] at room temperature overnight. The enzymatic activity was ceased by freezing the samples at −80°C for at least 1 h. Proteolytic digests were thawed and dried using a vacuum centrifuge, then resuspended to 1 μg/μL using 100 mM NH_4CO_3. The derivatization procedure was repeated to label the newly generated peptide N-terminals. Samples were desalted using homemade C_{18} stage-tips [26]. The stage-tips were quenched using ACN, followed by equilibration with 0.1% TFA. The pH of the samples was adjusted using glacial acetic acid (pH < 4), samples were loaded onto C_{18} stage-tips, washed with 0.1% TFA, and eluted using 0.5% acetic acid in 75% ACN. Samples were dried using a vacuum centrifuge, resuspended in the mobile phase A, and spiked with custom histone-like QC peptides to monitor the analytical reproducibility of the bottom-up MS analysis.

Bottom-up MS histone analysis

A nanoLC system (nanoElute 2) fitted with a C_{18} column (15 cm × 150 μm i.d., 1.5 μm, Bruker PepSeq column) kept at 50°C was connected to a commercial timsTOF Pro2 mass spectrometer (Bruker Daltonics, MA). A sample injection consisted of 1 μL of 200 ng/μL derivatized histone digest spiked with 25 ng/μL QC peptides. The nanoLC separation consisted of A (0.1% FA in water) and B (0.1% FA in 100% ACN) mobile phases with the following gradient: (i) 0 min 2% B, (ii) 0–60 min to 35%, (iii) 60–69 min to 95% B, and (iv) 70–78 to 2% B. A nanoESI source (Captive Spray, Bruker Daltonics MA) operated at 1200 V with a flow rate of 500 nL/min was used. Tandem CID MS/MS was performed on mobility and m/z selected precursor ions (PASEF mode) over the 100–1700 m/z and 0.60–1.80 $1/K_0$ range; CID collision energy was stepped as a function of the m/z and mobility [24, 26]. Additional information regarding MS instrument parameters can be found in Supplementary Fig. S3 in the supporting information.

Data analysis

Top-down LC-TIMS-MS and TIMS-q-ECD MS/MS data were analyzed using DataAnalysis software (DA, v6.1, Bruker Daltonics) and OmniScape 2025 software (OSc, Bruker Daltonics). A Tuning Mix standard (G24221A, Agilent Technologies, Santa Clara, CA) was used for external mobility and m/z calibration (see instrument parameters in Supplementary Fig. S3). Tandem q-ECD MS/MS spectra were processed for *de novo* sequencing with a 15 ppm mass error; the charge, neutral monoisotopic mass, and noise curve offset were determined for each data set. The UniProtKB [57] reference proteomes and Swiss-Prot databases were used with no species bias for the MS-BLAST [58] sequence similarity search. The high abundance of PTMs in the endogenous coral samples relative to unmodified proteins (only search considered in the *de novo* and MS-BLAST algorithms) required added manual annotation and confirmation of the proposed proteoforms, which was performed using a mass error of 10 ppm (Supplementary Figs S5–S49). For example, annotated MS/MS spectra of the more abundant proteoforms can be found in Supplementary Figs S50–S55 in the supporting information; note that all fragment assignments were manually curated.

Bottom-up LC-TIMS-q-CID MS/MS data were analyzed using a script (Supplementary Fig. S56) in DataAnalysis software (DA, v6.1, Bruker Daltonics) for the extraction and quantification of a predetermined peptide target list with varying PTMs ($me_{1–3}$, ac, and ox). The derivatization and MS analysis reproducibility in the bottom-up workflow were evaluated using the internal, custom QC synthetic peptide standard (GVKFRGSTGGKAPRGKAPATSGMVGPHR), which, when subjected to propionylation, digestion, and desalting procedures, as described in Supplementary Fig. S4, results in the targeted screening for QC1 (GSTGGKAPR, 471.75^{2+}) and QC2 (GKAPATSGMVGPHR, 739.38^{2+}). The hexakis(2,2-difluoroethoxy) phosphazine, hexakis(2,2,3,3-tetrafluoropropoxy) phosphazine, and hexakis(1H, 1H, 7H-dodecafluoroheptoxy) phosphazine standards (Apollo Scientific Ltd, UK) were used for external mobility and m/z calibration (see instrument parameters in Supplementary Fig. S3). All reported peptides and PTMs (Supplementary Figs S57–S77) were manually curated using the precursor mass (± 0.01 Da)

and isotopic profile, reported peptide mobility profiles (including positional isomers, RSD < 2%) [21], and MS/MS product ion assignment (<10 ppm).

Proteoform sequence confirmation was performed and evaluated based on the sequence coverage (%SC), sequence validation percentage (%SVP) [59], intensity coverage (%IC), and MS score (MS):

$$\%SC = \frac{number\ of\ amino\ acids\ confirmed}{total\ number\ of\ amino\ acids} \times 100\%, \quad (1)$$

$$\%SVP = \frac{continuous\ amino\ acids\ confirmed\ from\ terminals\ (within\ tolerance)}{total\ number\ of\ amino\ acids} \times 100\%$$
$$(2)$$

$$\%IC = \frac{intensity\ of\ matched\ fragment\ peaks}{intensity\ of\ all\ peaks} \times 100\%, (3)$$

$$MS\ score = \%SC \times \%IC. \quad (4)$$

The relative abundance of top-down isomeric proteoforms was calculated based on the intensity of reporter ions specific to each isomer and normalized to the area of the integrated MS^1 spectra using Equation 5:

$$Relative\ abundance = \frac{intensity\ of\ selected\ isomer\ reporter\ ion\ (s)}{combined\ intensity\ of\ all\ isomeric\ reporter\ ions} \times MS^1\ area$$
$$. (5)$$

Results

Coral tissue histone pull-down and extraction

The coral tissue histone extraction (Fig. 1A) resulted in ∼5 μg protein mass per 100 mg of powdered coral fragments (tissue + skeleton). Several bands were observed in the SDS–PAGE analysis; the most abundant bands suggested the presence of histone families. At a lower abundance, other acid-soluble nuclear proteins were also observed (Supplementary Fig. S78A).

Coral histone primary sequences and PTM annotation

The analysis of the pooled *A. cervicornis* histones using LC-TIMS-MS confirmed several histone families (see Supplementary Fig. S78). Inspection of the LC-derived 2D IMS-MS plots allowed for the separation of the histone variants based on their retention time (Fig. 1B). The endogenous histone variants showed a high degree of PTMs; different from traditional top-down methods, a better MS identification was achieved due to the mobility separation of the charge state distribution (CSD) of interest from potential mass interferences (e.g. co-eluting proteins) and reduced chemical noise. Two proteins with a ∼11.3 kDa mass at retention times (RT) of 32.0 min (11.250–11.432 kDa mass range) and 36.9 min (11.234–11.442 kDa mass range) were flagged for further analysis as potential H4 variants (Fig. 1B); in addition to the mass, the 36.9 min RT LC band is consistent with the observation of the human (HeLa) H4 histone (Supplementary Fig. S78B). Four proteins with a ∼13.3–13.5 kDa mass at RT 30.1 min (13.401–13.470 kDa mass range), 33.8 min (13.355–13.464 kDa mass range), 35.2 min (13.421–13.484 kDa mass range), and 45.9 min (13.408–13.540 kDa mass range) were also flagged as potential H2A/H2B variants (Supplementary Fig. S78C). The CSD and deconvoluted MS showed distinctive mass shifts corresponding to multiple known PTMs (e.g.

$me_{1–3}$ with a +14.02, +28.03, and +42.05 Da, ac with +42.01 Da, and ox with +15.99 Da per modified amino acid and/or N-term; Supplementary Fig. S79). Additionally, mass shifts were observed in the 33.8 min and 45.9 min RT bands, which did not correspond to common PTMs (+28.01 and +26.27 Da, respectively).

De novo sequencing of H4 histone variants. Tandem ECD MS/MS on mobility and mass-isolated precursor ions showed abundant fragmentation for the most abundant charge states: +10 to +12 and +11 to +13 for 32.0 min and 36.9 min RT LC bands, respectively (Fig. 1B). The *de novo* sequencing of the ECD-MS/MS spectra generated a list of AA strings: ∼150 and ∼300 (Supplementary Fig. S80) for the 32.0 and 36.9 min RT LC bands, respectively. The MS-BLAST database [58] search of the AA strings resulted in the identification of high-scoring pairs (HSPs; query sequences producing significant local alignments to a database sequence) matching H4 sequences from various organisms (Fig. 2 and Supplementary Fig. S80). High-scoring pairs are assigned a score by MS-BLAST based on the length and AA matches of the query (unknown) sequence to a sequence from the database being searched (e.g. SwissProt). Ambiguities such as unclear AA ordering or AAs with the same or similar masses (i.e. I/L and K/Q, respectively) in either the query or database sequences may also bias the HSP score for a given AA string. The highest total score (sum of HSP scores as determined by MS-BLAST) was for *Penaeus vannamei* (a.k.a *Litopenaeus vannamei*) or *Whiteleg shrimp* (Protein ID P83865) with 372; the ambiguity of the 92–100 protein segment partially biases the score. The next high-scoring H4 proteins were *Dendronephthya klunzingeri* (Q6LAF1) and *Acropora formosa* (P35059) with 355 and less ambiguity in the segments. Considering that *A. formosa* shares the genus and ancestry with *A. cervicornis*, this H4 sequence was used as the starting point for further analysis.

De novo sequencing of H2A and H2B histone variants. Tandem ECD MS/MS was performed on the most abundant charge states for each RT fraction: +14 to +16 for the 30.1, 33.8, and 35.2 min RT bands and + 12 to +14 for the 45.9 min RT band (Supplementary Fig. S81). *De novo* sequencing generated a list of amino acid (AA) strings: ∼500, ∼400, ∼500, and ∼600 for the 30.1, 33.8, 35.2, and 45.9 min RT LC bands, respectively (Supplementary Figs S82A–S85A). The MS-BLAST database [58] search resulted in HSPs from either H2B (30.1, 33.8, and 35.2 min RT band) or H2A (45.9 min RT band) sequences from various organisms (Supplementary Figs S82B–S85B). The highest-scoring protein matched for the H2A fraction was *A. formosa* (Protein ID P35061), a closely related coral from the same genus with a total score of 228 (Supplementary Fig. S85C). The same case was seen for the 33.8 min and 35.2 min H2B fractions (Protein ID P35067), which had total matching scores of 393 and 148, respectively (Supplementary Figs S83C and S84C).

Acropora cervicornis H4 variants

Combining the information at the protein level (e.g. potential PTMs), the *de novo* results, and the manually annotated PTMs, a candidate proteoform annotation list was established for each RT band (see the proposed sequences in Fig. 2). For example, the N-terminal amino acids and PTMs were sequenced manually and suggested that the N-term was observed acetylated for most proteoforms in both H4 LC bands (based on the +42.01 Da mass shifts of the first 3–5 AAs
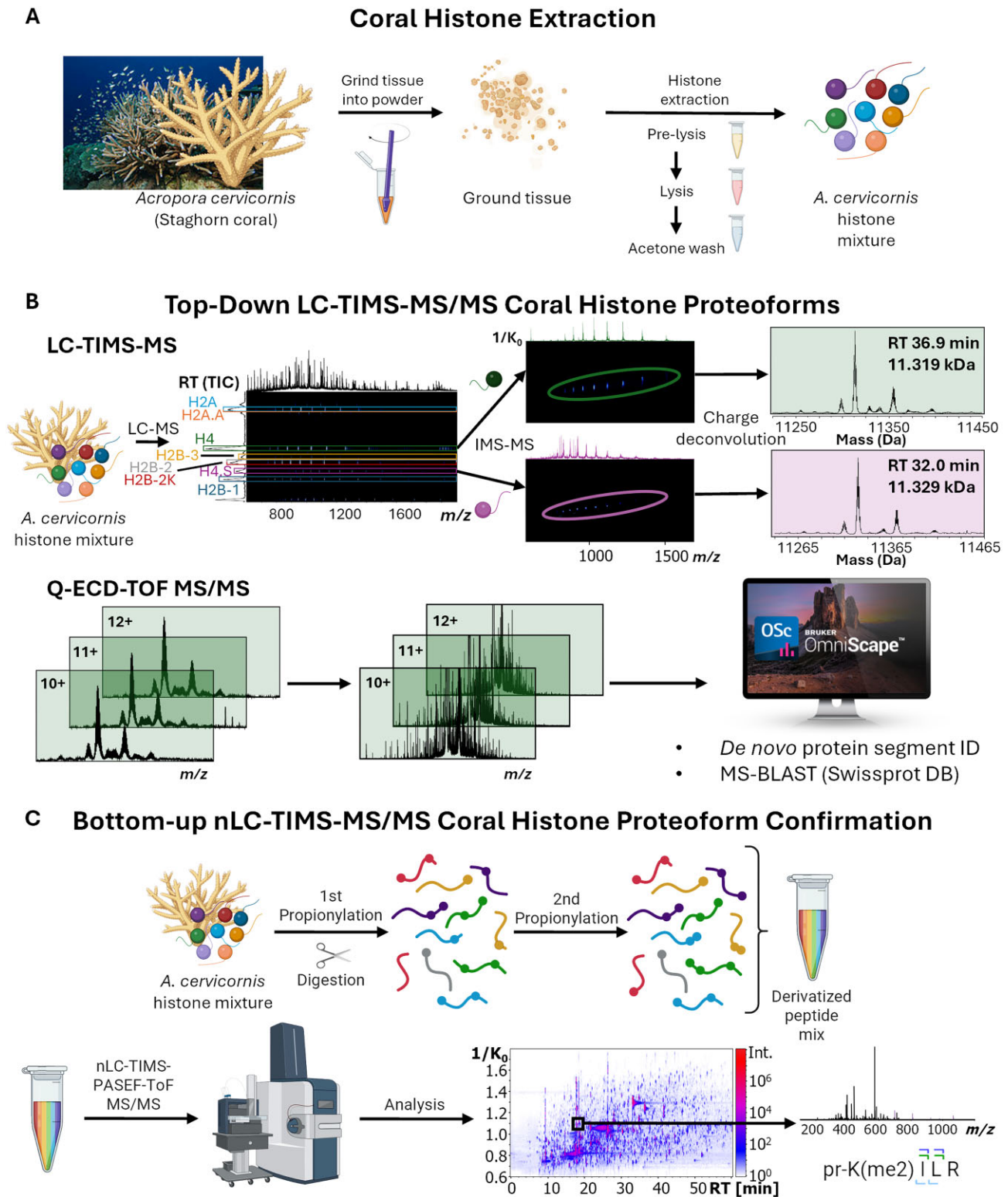
Figure 1. Direct protein sequence generation and proteoform PTM annotation based on liquid chromatography (online and offline), TIMS, ECD fragmentation, and high-resolution MS for *A. cervicornis*. (**A**) Histone extraction from *A. cervicornis* coral. (**B**) Top-down identification of histone variants with subsequent ECD-MS/MS and *de novo* sequencing. (**C**) Bottom-up nLC-TIMS-MS/MS peptide confirmation.
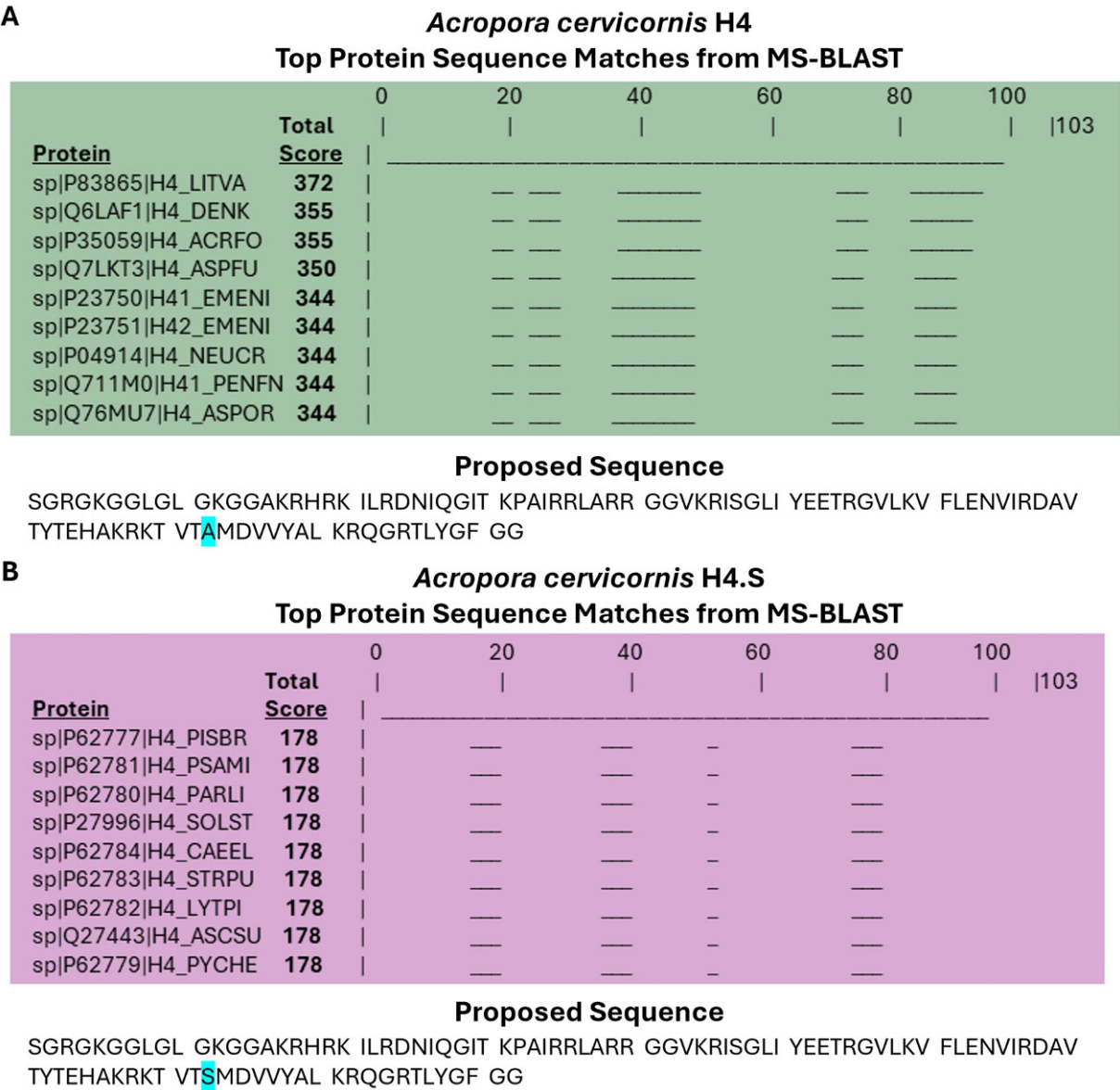
A

Acropora cervicornis H4
Top Protein Sequence Matches from MS-BLAST

		0	20	40	60	80	100	
Protein	**Total Score**							|103
sp|P83865|H4_LITVA	372							
sp|Q6LAF1|H4_DENK	355							
sp|P35059|H4_ACRFO	355							
sp|Q7LKT3|H4_ASPFU	350							
sp|P23750|H41_EMENI	344							
sp|P23751|H42_EMENI	344							
sp|P04914|H4_NEUCR	344							
sp|Q711M0|H41_PENFN	344							
sp|Q76MU7|H4_ASPOR	344							

Proposed Sequence

SGRGKGGLGL GKGGAKRHRK ILRDNIQGIT KPAIRRLARR GGVKRISGLI YEETRGVLKV FLENVIRDAV
TYTEHAKRKT VTAMDVVYAL KRQGRTLYGF GG

B

Acropora cervicornis H4.S
Top Protein Sequence Matches from MS-BLAST

		0	20	40	60	80	100	
Protein	**Total Score**							|103
sp|P62777|H4_PISBR	178							
sp|P62781|H4_PSAMI	178							
sp|P62780|H4_PARLI	178							
sp|P27996|H4_SOLST	178							
sp|P62784|H4_CAEEL	178							
sp|P62783|H4_STRPU	178							
sp|P62782|H4_LYTPI	178							
sp|Q27443|H4_ASCSU	178							
sp|P62779|H4_PYCHE	178							

Proposed Sequence

SGRGKGGLGL GKGGAKRHRK ILRDNIQGIT KPAIRRLARR GGVKRISGLI YEETRGVLKV FLENVIRDAV
TYTEHAKRKT VTSMDVVYAL KRQGRTLYGF GG

Figure 2. MS-BLAST results for *A. cervicornis* 32.0 and 36.9 min fractions. (**A**) *A. cervicornis* H4 and (**B**) H4.S top 9 matching proteins found in the SwissProt database with total scores and mapped protein segments, and proposed sequences for each fraction. The single amino acid variation A83S is highlighted.

corresponding to the c_3 and c_{3-5}, b_3 fragments in the 32.0 min and 36.9 min spectra, respectively). The most abundant mass signal of the 32.0 and 36.9 min RT LC bands corresponded to N-acSK20me$_2$, confirmed by the additional mass shift observed from mass fragments of the first 23 AAs. (e.g. mass shifts of +28.03 observed in $c_{20,21,23}$ and c_{20-23}, a_{21} fragments, respectively). The difference in the sequences from the two LC bands was derived from discrepancies toward the C-terminal fragment assignment and resulted from a single amino acid variation of A83 to S83 (A83S); this AA position is commonly seen mutated in the H4 histones (see H4 sequences listed in Supplementary Figs S80 and S86). The H4 *A. cervicornis* variant proposed (RT 36.9 min LC band) has a 100% homology with the H4 *A. formosa* sequence and no less than 92% with the other listed H4 sequences. The second H4 variant (RT 32.0 min LC band), now referred to as H4.S *A. cervicornis* has a 99% homology with the H4 *A. formosa* sequence. An

all-versus-all comparison of the established H4 *A. cervicornis* sequence and the 9 selected matches from MS-BLAST showed that <12% of the AA positions in H4 are prone to mutations between species/variants.

The H4 *A. cervicornis* top-down analysis resulted in the annotation of 28 proteoforms with me$_{0-2}$ and ac$_{0-4}$ and a high sequence coverage range (80%–95%), sequence validation percentage (up to 99%), intensity coverage range of 6%–70%, and MS score range of 5.5–66.9 (see relative abundances in Fig. 3A and MS/MS statistics in Supplementary Figs S5–S18 and Supplementary Table S1). The first six mass signals in the MS (RT 36.9 min LC band) correspond to unmodified, K20me$_1$, K20me$_2$, N-acS, N-acSK20me$_1$, and N-acSK20me$_2$ proteoforms. The next mass signals correspond to a mass shift of 2ac (N-acSK12ac and N-acSK16ac), 2ac + me$_1$ (N-acSK12acK20me$_1$ and N-acSK16acK20me$_1$), and 2ac + me$_2$ (N-acSK12acK20me$_2$
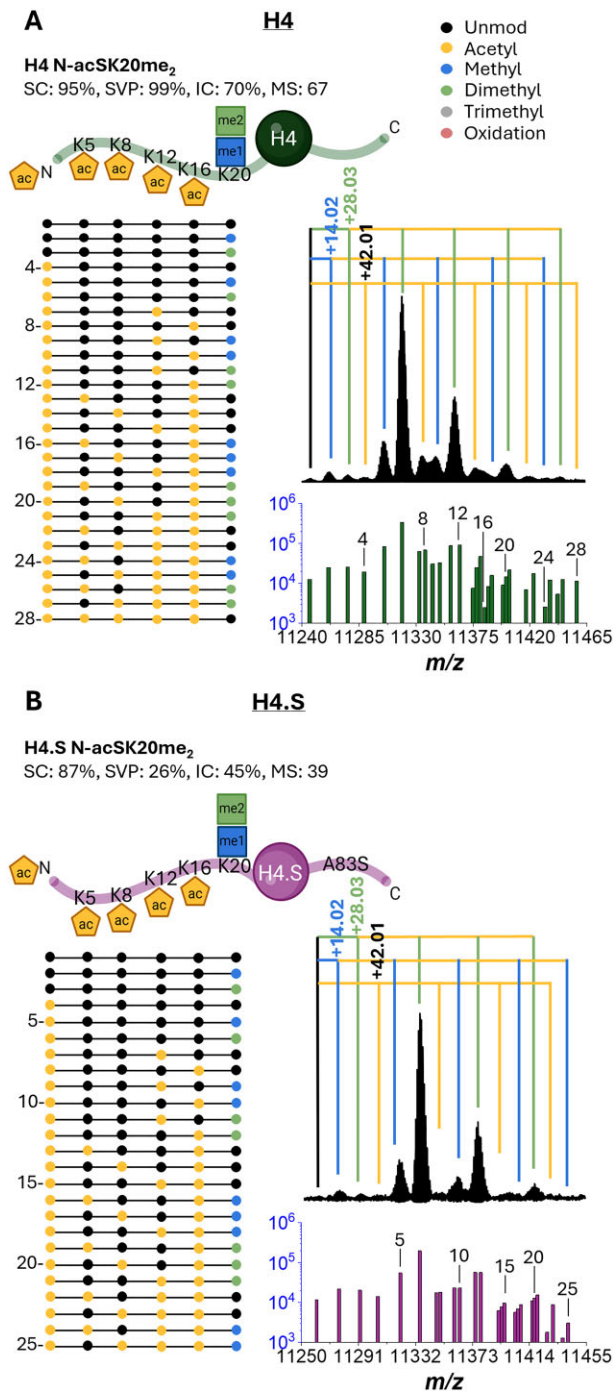
and N-acSK16acK20me$_2$); the ratio of K12ac to K16ac in these proteoforms was nearly equal. The next mass signals correspond to 3ac (N-acSK5acK16ac, N-acSK8acK16ac, and N-acSK12acK16ac), 3ac + me$_1$ (N-acSK5acK16acK20me$_1$, N-acSK8acK16acK20me$_1$, and N-acSK12acK16acK20me$_1$), and 3ac + me$_2$ (N-acSK5acK16acK20me$_2$, N-acSK8acK16acK20me$_2$, and N-acSK12acK16acK20me$_2$). The following three mass signals correspond to 4ac (N-acSK5acK12acK16ac and N-acSK8acK12acK16ac), 4ac + me$_1$ (N-acSK5acK12acK16acK20me$_1$ and N-acSK8acK12acK16acK20me$_1$), and 4ac + me$_2$ (N-acSK5acK12acK16acK20me$_2$ and N-acSK8acK12acK16acK20me$_2$). The next mass signal corresponded to 5ac (N-acSK5acK8acK12acK16ac), observed at lower levels than the other acetylated proteoforms.

The H4.S *A. cervicornis* top-down analysis resulted in the annotation of 25 proteoforms with me$_{0-3}$ and ac$_{0-4}$ and a high sequence coverage (70%–87%), sequence validation percentage up to 30%, intensity coverage range of 6%–45%, and MS score range of 4.6–39.2 (see relative abundances in Fig. 3B and MS/MS statistics in Supplementary Figs S19–S31 and Supplementary Table S1). The first six mass signals in the MS (RT 32.0 min LC band) correspond to unmodified, K20me$_1$, K20me$_2$, N-acS, N-acSK20me$_1$, and N-acSK20me$_2$. The next mass signals correspond to a mass shift of 2ac (N-acSK12ac and N-acSK16ac), 2ac + me1 (N-acSK12acK20me$_1$ and N-acSK16acK20me$_1$), and 2ac + me2 (N-acSK12acK20me$_2$ and N-acSK16acK20me$_2$). The next mass signals correspond to a mass shift of 3ac (N-acSK5acK16ac, N-acSK8acK16ac, and N-acSK12acK16ac), 3ac + me1 (N-acSK5acK16acK20me$_1$, N-acSK8acK16acK20me$_1$, and N-acSK12acK16acK20me$_1$), and 3ac + me2 (N-acSK5acK16acK20me$_2$, N-acSK8acK16acK20me$_2$, and N-acSK12acK16acK20me$_2$). The next mass signals correspond to 4ac (N-acSK5acK12acK16ac and N-acSK8acK12acK16ac) and 4ac + me$_1$ (N-acSK5acK12acK16acK20me$_1$ and N-acSK8acK12acK16acK20me$_1$).

Acropora cervicornis H2A variants

The H2A N-terminal AAs and PTMs were sequenced manually, leading to the suggestion that the N-terminal was primarily observed to be acetylated based on the +42.01 Da mass shift of the first 3–5 AAs corresponding to the c_{3-5}, $b_{3,5}$, and $a_{3,5}$ fragment ions (Fig. 4A). In the case of the H2A band (45.9 min RT), a mass shift between the two most abundant MS peaks of +26.27 Da was observed (Supplementary Fig. S79F). This same mass shift was observed between other peaks in the distribution and is not a common mass shift among known PTMs. Upon further inspection, the mass shift was found to be derived from a C-terminal single amino acid variation of P123 to A123 (P123A), indicating the presence of two coeluting H2A variants (referred to as H2A and H2A.A for the variants containing P123 and A123, respectively, from this point forward). In particular, the presence of reporter ions a_{124}, $c_{123-124}$, and y_3 corresponding to the P123 in the H2A N-acS proteoform (92.74% SC, 84.00% SVP, 65.77% IC, 59.94 MS) and ions a_{124}, b_{124}, c_{124}, and y_3 corresponding to the A123 in the H2A.A N-acS proteoform (89.52% SC, 72.80% SVP, 39.92% IC, 34.45 MS) was used to confirm the presence of both variants. The C-terminal of H2A proteins commonly mutates between and within species, as suggested
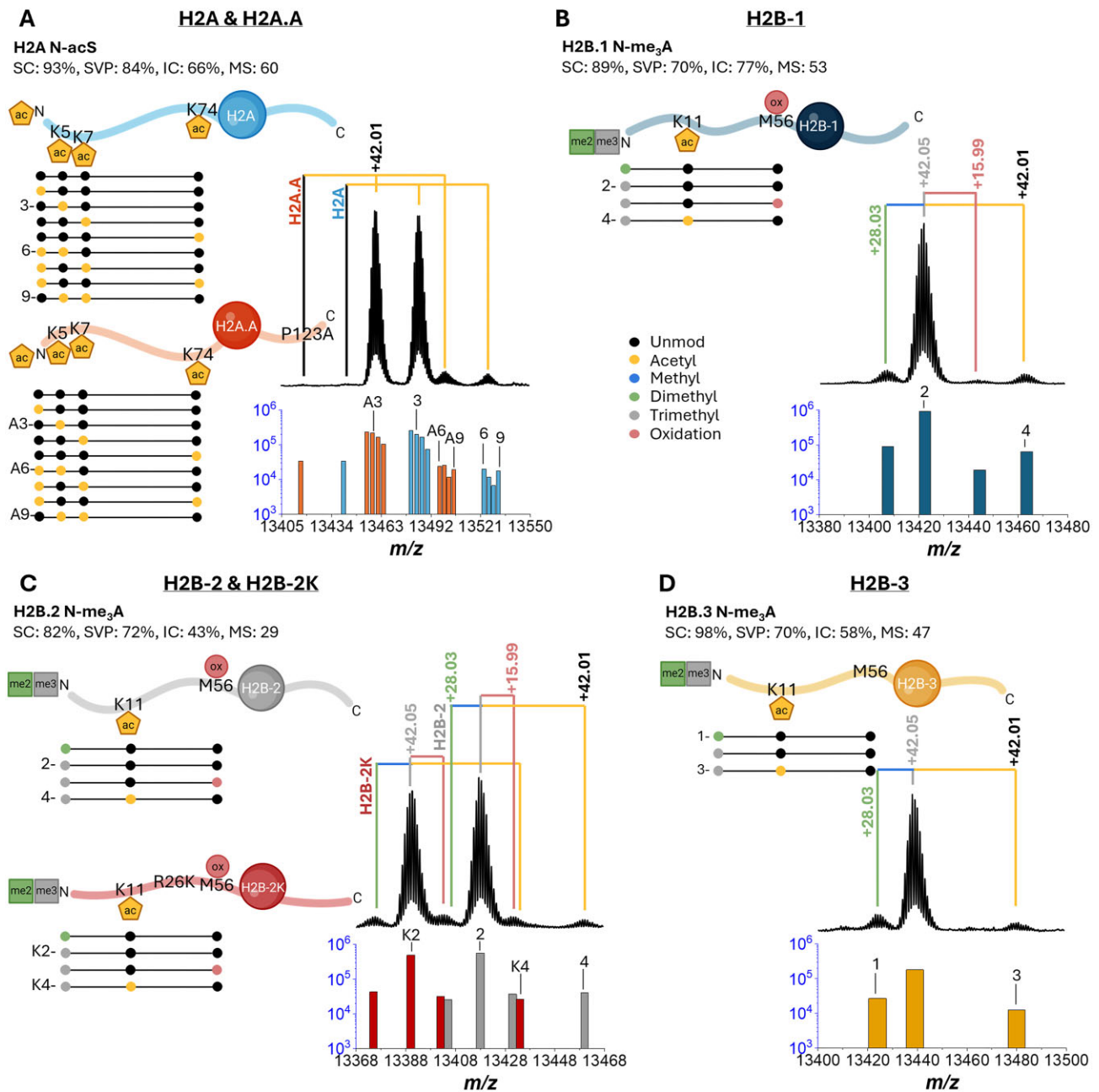
Figure 3. *Acropora cervicornis* histone H4 detected variants and corresponding proteoforms. (**A**) H4 and (**B**) H4.S variants with PTM positions shown with specific PTM combinations below (dotted lines; left). The deconvoluted MS1 spectra, labeled with mass shifts corresponding to the observed PTM(s) (branched lines, colored according to the PTM represented by the observed mass shift from the unmodified mass), and bar plot (right), showing the relative abundances of each of the observed proteoforms with numbered labels corresponding to the PTM combinations shown (dotted lines). Details on sequence coverage at the protein and peptide levels are provided in supplementary data.

Figure 4. *Acropora cervicornis* histone H2A and H2B detected variants and corresponding proteoforms. (**A**) H2A & H2A.A, (**B**) H2B-1, (**C**) H2B-2 & H2B-2K, and (**D**) H2B-3 variants with PTM positions shown with specific PTM combinations below (dotted lines; left). The deconvoluted MS1 spectra, labeled with mass shifts corresponding to the observed PTM(s) (branched lines, colored according to the PTM represented by the observed mass shift from the unmodified mass), and bar plot (right), showing the relative abundances of each of the observed proteoforms with numbered labels corresponding to the PTM combinations shown (dotted lines). Details on sequence coverage at the protein and peptide levels are provided in supplementary data.

by the matching H2A sequences found by the *de novo* MS-BLAST search (Supplementary Fig. S85D). The most abundant signal from both variants corresponded to N-acS, as previously determined by N-terminal fragments from the first 3–5 AAs and sequence coverages of ∼93% and ∼90% achieved for variants H2A and H2A.A, respectively (Fig. 4A). The proposed H2A *A. cervicornis* variants both demonstrate a 93% homology match to the H2A *A. formosa* sequence (top MS-BLAST result); however, homology ranges from 70% to 90% to the remaining 9 top matching BLAST sequences (Supplementary Fig. S85D). An all-versus-all comparison of

the proposed sequences to the top 10 MS-BLAST results suggests that up to 40% of the AA positions in H2A are prone to mutations or insertions/deletions between species and/or variants.

The H2A/H2A.A *A. cervicornis* combined top-down and bottom-up analysis resulted in the annotation of 18 proteoforms (9 per variant) with a high sequence coverage (68%–93% and 67%–90%), sequence validation percentage ranges of 38%–84% and 44%–73%, intensity coverage ranges of 13%–66% and 8%–40%, and MS score ranges of 9.5–60.0 and 4.9–34.4 for H2A and H2A.A, respectively

(see relative abundances in Fig. 4A and MS/MS statistics in Supplementary Figs S32–S41 and Supplementary Table S1). The H2A.A and H2A unmodified mass signals are followed by two much higher (two orders of magnitude) signals corresponding to H2A.A and then H2A 1ac (N-acS, K5ac, K7ac, and K74ac) with consistent ratios of each proteoform between variants. The next mass signals correspond to 2ac proteoforms for each variant (N-acSK5ac, N-acSK7ac, N-acSK74ac, and K5acK7ac). The ratios of each 2ac proteoform between variants are mostly consistent, except for the N-acSK7ac proteoform being slightly higher than N-acSK5ac in the H2A.A versus the H2A variant.

Acropora cervicornis H2B variants

Three separate LC peaks were identified as H2B variants (30.1, 33.8, and 35.2 min RT bands). For each fraction, the *de novo* sequence strings were combined with the manual mass shift annotations detected in the CSDs at the protein level to determine proteoform candidate lists. In each case, the N-terminal AAs and PTMs were sequenced manually.

The 30.1 min RT band (referred to as H2B-1 from this point forward) was largely observed to contain trimethylation at the N-terminal based on the +42.05 Da mass shift of the first 3–6 AAs corresponding to c_{4-6}, $b_{5,6}$, and a_{4-6} and the presence of the z_{120} and y_{121} C-terminal ions, which confirm the K3 to be largely unmodified (Fig. 4B). Sequence coverage higher than 89% was achieved for the H2B-1 variant with the most abundant proteoform being N-me$_3$A. The H2B-1 *A. cervicornis* histone was the only fraction not found to have high sequence homology to *A. formosa* in the MS-BLAST search of the *de novo* sequence strings (Supplementary Fig. S82). Homology of the protein compared to the top 10 MS-BLAST results ranged from 70% to 83% with the highest match being to *Asterias rubens* (common European starfish, Protein ID P02286). An all-versus-all comparison of the proposed sequence to the top 10 BLAST results suggested ∼36% of AAs in the H2B sequence may be prone to mutations, insertions, or deletions between species/variants (Supplementary Fig. S82D).

The 33.8 min RT band (referred to as H2B-2 from this point forward) was observed to be largely trimethylated at the N-terminal based on a +42.05 Da mass shift of the first 3–6 AAs corresponding to the c_{4-6} and b_6 N-terminal ions and the presence of z_{120} and y_{121} C-terminal ions, indicating that K3 was largely unmodified. The CSD of the H2B-2 fraction showed a distinctive mass shift between the two most abundant peaks of +28.01 Da (Supplementary Fig. S79D). Upon inspection, this mass shift was not a result of dimethylation (+28.03 Da), but rather a single amino acid variation in the N-terminal tail of R26 to K26 (R26K), indicating two coeluting variants (referred to as H2B-2 and H2B-2K for the variants containing R26 and K26, respectively, from this point forward). The N-terminal of H2B proteins is highly variable between and within species, as shown by the matching H2B sequences found by the *de novo* MS-BLAST search (Supplementary Figs S82D–S84D). The most abundant signal from both variants corresponded to N-me$_3$A proteoforms, as determined by the previously stated N-terminal fragments and sequence coverages of ∼82% and ∼79%, for H2B-2 and H2B-2K, respectively (Fig. 4C). The signal-to-noise ratio (S/N) of each charge state for this fraction was lower than the other observed H2A/B fractions, reducing the number/quality of fragment ions observed, and leading to

the slightly lower sequence coverages obtained. The proposed H2B-2 *A. cervicornis* variants demonstrate 87% and 86% (H2B-2 and H2B-2K, respectively) homology matches to the H2B *A. formosa* sequence (top MS-BLAST result). Shared homology between the other 9 top BLAST matches ranges from 72% to 81% and 71% to 80% for H2B-2 and H2B-2K, respectively (Supplementary Fig. S83D). An all-versus-all comparison of the proposed sequences to the top 10 MS-BLAST results suggested that ∼39% of AAs in the H2B sequence may be prone to mutations, insertions, or deletions between species/variants (Supplementary Fig. S83D).

The 35.2 min RT band (referred to as H2B-3 from this point forward) was also observed to primarily be trimethylated at the N-terminal based on a +42.05 Da mass shift of the first 3–6 AAs corresponding to c_3, $b_{4,6}$, and $a_{4,6}$ ions, as well as z_{120} and y_{121} C-terminal ions confirming that the K3 is unmodified (Fig. 4D). The highest sequence coverage (nearly 98%) was achieved for H2B-3 with the most abundant proteoform being N-me$_3$A. This protein was observed with the fewest proteoform peaks in the MS spectrum (Supplementary Fig. S79E). The H2B-3 *A. cervicornis* histone variant had an 84% homology match to the H2B *A. formosa* protein sequence and a homology match range of 77%–80% to the remaining top 9 MS-BLAST result sequences. An all-versus-all comparison of the proposed sequence to the top 10 MS-BLAST results suggested ∼30% of AAs in the H2B sequence may be prone to mutations/insertions/deletions between species, with most of these changes observed in the N-terminal sequences (Supplementary Fig. S84D).

A total of 15 H2B variant proteoforms were annotated. The H2B-1 *A. cervicornis* combined top-down and bottom-up analysis resulted in the annotation of 4 proteoforms with a sequence coverage range of 74%–90%, sequence validation percentage range of 20%–70%, intensity coverage range of 12%–77%, and MS score range of 7.1–53.0 (see relative abundances in Fig. 4B and MS/MS statistics in Supplementary Figs S42 and S43, and Supplementary Table S1). The four MS peaks correspond to the unmodified, N-me$_2$A, N-me$_3$A, N-me$_3$AM56ox, and N-me$_3$AK1ac proteoforms. The H2B-2/2K *A. cervicornis* combined top-down and bottom-up analysis resulted in the annotation of 8 proteoforms (4 each) with sequence coverage ranges of 63%–82% and 60%–79%, sequence validation percentage ranges of 25%–72% and 26%–44%, intensity coverage ranges of 19%–43% and 8%–3%, and MS score ranges of 8.6–28.9 and 3.8–19.3 for H2B-2 and H2B-2K, respectively (see relative abundances in Fig. 4C and MS/MS statistics in Supplementary Figs S44–S47 and Supplementary Table S1). The eight annotated MS peaks correspond to the H2B-2K N-me$_2$A, H2B-2K N-me$_3$A, H2B-2K N-me$_3$AM56ox, H2B-2 N-me$_2$A, H2B-2 N-me$_3$A, H2B-2 N-me$_3$AM56ox, H2B-2K N-me$_3$AK11ac, and H2B-2 N-me$_3$AK11ac proteoforms. No unmodified H2B-2/2K was confirmed to be observed. The H2B-3 *A. cervicornis* combined top-down and bottom-up analysis resulted in the annotation of 3 proteoforms with a sequence coverage range of 85%–97%, sequence validation percentage range of 25%–70%, intensity coverage range of 17%–58%, and MS score range of 10.8–47.2 (see relative abundances in Fig. 4D and MS/MS statistics in Supplementary Figs S48 and S49, and Supplementary Table S1). The three MS peaks correspond to the N-me$_2$A, N-me$_3$A, and N-me$_3$AK11ac proteoforms. No unmodified H2B-3 could be confirmed to be observed.

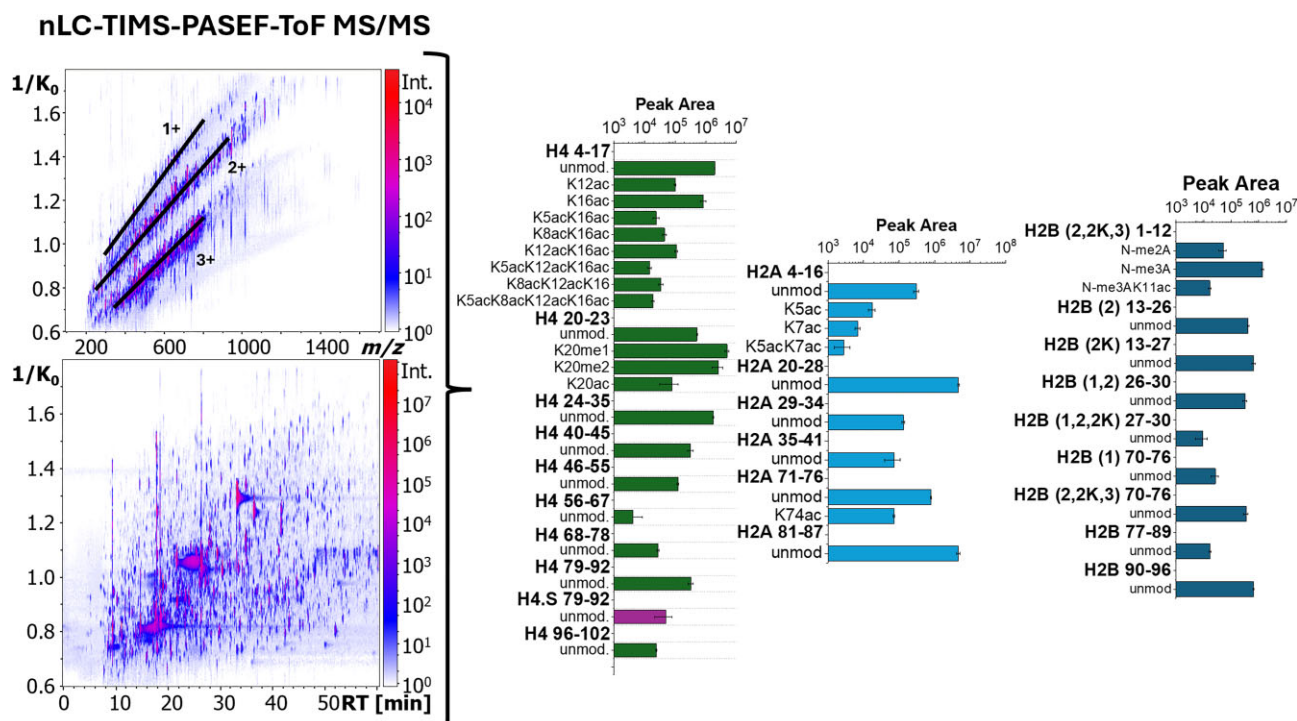## nLC-TIMS-PASEF-ToF MS/MS

Figure 5. Bottom-up confirmation of histone peptides using nLC-TIMS-PASEF-ToF MS/MS. Detected peptides for H4, H2A, and H2B variants with relative abundances.

Bottom-up histone proteoform confirmation

Bottom-up nLC-TIMS-PASEF-ToF MS/MS analysis provided extended confirmation of H4, H2A, and H2B sequences based on LC RT, mobility ($1/K_0$), and tandem MS/MS fragmentation (Fig. 1C and Supplementary Table S2).

H4 *A. cervicornis* bottom-up confirmation. Over 21 peptides were identified and confirmed corresponding to: (i) sequence-confirming H4 24–35, 40–45, 46–55, 56–67, 68–78, and 96–102; (ii) variant-confirming H4 79–92 and H4.S 79–92; and (iii) PTM-confirming H4 4–17 and 20–23 peptides (Supplementary Figs S57–S77). The H4 4–17 peptide was detected with ac_{0-4} and a sequence coverage range of 46%–92%, intensity coverage range of 4%–32%, and MS score range of 2.5–29.5 (see relative abundances in Fig. 5 and Supplementary Fig. S87 and MS/MS statistics in Supplementary Figs S57–S77 and Supplementary Table S2). The H4 4–17 peptide was detected in the unmodified, 1ac (K12ac and K16ac), 2ac (K5acK16ac, K8acK16ac, and K12acK16ac), 3ac (K5acK12acK16ac and K8acK12acK16ac), and 4ac (K5acK8acK12acK16ac) forms. Additionally, H4 20–23 was detected with me_{0-2} and ac_{0-1} and a sequence coverage range of 33%–100%, intensity coverage range up to 51%, and MS score range of 0.6–51.4 (see relative abundances in Fig. 5 and Supplementary Fig. S87, and MS/MS statistics in Supplementary Figs S57–S77 and Supplementary Table S2). The H4 20–23 peptide was detected in the unmodified, K20me1, K20me2, and K20ac forms.

H2A *A. cervicornis* bottom-up confirmation. A total of 10 H2A peptides were identified and confirmed. The H2A and H2A.A unmodified peptides 4–16, 20–28, 29–34, 35–41, 71–76, and 81–87 were able to confirm shared sequence segments between the two variants with sequence coverage ranging from 20% to 100%, intensity coverage of up to 77%, and MS scores of 0.3–71.3 (see peptide abundances in

Fig. 5, and Supplementary Fig. S88 and MS/MS statistics in Supplementary Table S2). The H2A/H2A.A 4–16 peptide was detected with up to two acetylations at positions K5 (average 55% SC, 15% IC, 9.2 MS score), K7 (average 61% SC, 16% IC, 10.1 MS score), or both (no MS/MS available; see relative abundances in Fig. 5 and Supplementary Fig. S88). Additionally, acetylation was observed at position K74 in the H2A/H2A.A 71–76 peptide (average 27% SC, 49% IC, 14.8 MS score). The C-terminal 88–125 peptide was not observed for either variant.

H2B *A. cervicornis* bottom-up confirmation. A total of eleven H2B peptides were identified and confirmed. The common H2B peptides 77–89 and 90–96 were observed in the unmodified forms with averaged sequence coverages of 66% and 50%, intensity coverages of 16% and 67%, and MS scores of 10.9 and 33.8, respectively (see peptide abundances in Fig. 5, and Supplementary Fig. S89 and MS/MS statistics in Supplementary Table S2). The H2B-2/2K/3 1–12 was observed in the N-me$_2$A (average 15% SC, 13% IC, 2.8 MS score), N-me$_3$A (average 48% SC, 9% IC, 5.9 MS score), and N-me$_3$AK11ac (average 33% SC, 54% IC, 15.4 MS score) forms (see relative abundances in Fig. 5 and Supplementary Fig. S89). The H2B-2 13–26 peptide was observed in the unmodified form (average 66% SC, 17% IC, 12.1 MS score), as was the H2B-2K 13–27 peptide (average 69% SC, 20% IC, 14.5 MS score). The H2B-1/2 26–30 (average 41% SC, 2% IC, 0.7 MS score) and H2B-1/2/2K 27–30 (average 33% SC, 4% IC, 1.3 MS score) peptides were observed as unmodified, as well. The H2B-1 70–76 (average 61% SC, 63% IC, 38.9 MS score) and H2B-2/2K/3 70–76 (average 77% SC, 50% IC, 34.2 MS score) peptides were also observed only as unmodified. The N-terminal peptide for H2B-1 was not observed; peptides in the range of 31–69 and the C-terminal 97–122 common peptide were also not observed.

Discussion

Direct analysis of histone proteoforms based on liquid chromatography (online and offline), mobility separation, and ECD MS/MS resulted in a high-confidence, direct alternative to homology modeling to annotate *A. cervicornis* histone proteoforms. The high confidence of the method relies on improved top-down precursor ion isolation (i.e. LC, IMS, and MS) and non-ergodic fragmentation (ECD MS/MS) leading to high primary sequence coverage (over 89%) based on *de novo* sequence prediction and database screening; when complemented with bottom-up approaches, near full sequence coverage was obtained. This resulted in a method capable of effective characterization of histone protein variants by generating the primary sequence for unknown histones, with the added benefit of detecting the PTMs at the proteoform level. When applied to the non-model organism, *A. cervicornis*, with little to no genomic reference information, this is the first report of the H4, H2A, and H2B core histone sequences and PTMs observed under stressed conditions. The method was effectively applied to the most abundant MS1 signals corresponding to H4, H2A, and H2B histones. Histones H1 and H3 were observed at the MS1 level at lower intensities and may require sample enrichment alternatives, which will be described in a follow-up publication.

The *A. cervicornis* histone proteoforms showed a high diversity of PTMs, including acetylation, oxidation, mono-, di-, and tri-methylation, as well as positional isomers, not distinguishable using traditional bottom-up MS proteomics workflows. Two H4 (53 proteoforms) were identified with up to ~95% sequence coverage, two H2A variants (18 proteoforms) were identified and sequenced with up to ~93% sequence coverage, and four different H2B *A. cervicornis* variants (15 proteoforms) were sequenced with up to nearly 98% coverage with little to no genomic reference material (see full table of proteoforms in Supplementary Table S1). The sequence coverages were reported after manual curation, and an average fragment false discovery rate of ~16% was calculated by comparing the number of total fragments identified by OSc to the number of final curated fragments. Additionally, sequence validation percentage is biased toward the N- and C-terminals, occasionally resulting in an SVP cutoff when terminal ions are absent or there is a break in sequence coverage, despite a high total sequence coverage and an average fragmentation efficiency of ~47%. Inspection of the H4 PTM patterns agrees with previously reported acetylation patterns ("zip model" [60]), where the H4 4–17 peptide acetylation decreases in abundance from K16 to K5. Unique to this approach is the direct identification of histone sequence segments and their validation based on MS and MS/MS annotation and sequence homology. A current caveat on proteoform discovery is the lack of PTMs in the existing data processing algorithms, which is accentuated in the case of endogenous samples where the most abundant mass signals deviate from the unmodified protein, which in some cases was not present at all. Results showed that in cases where the unmodified protein was observed, it was typically two orders of magnitude lower than the most abundant mass signals. The top-down MS proteoform annotation showed good agreement with the bottom-up MS results—all the H4 and H4.S sequence-confirming and variant-confirming peptides were detected, and several H2A and H2B sequence and PTM-confirming peptides were observed (see Fig. 5 and Supplementary Figs S87–S89). More-over, complete proteoform annotation was achieved using top-down MS proteomics, partially due to the reduced chemical noise provided by the complementary mobility separation. The signal-to-noise in IMS-MS experiments was ~3× higher than that of the LC-MS (see Supplementary Fig. S90 for LC-MS and IMS-MS S/N comparisons). While histones are presumed to be partially conserved (typically <12% AA variation for H4 and up to 40% for H2A/H2B), the histone variability across species can benefit from these methods to account for AA mutations, insertions/deletions, and PTMs. The described method can be expanded beyond histones. Moreover, when applied to corals, inspection of the histone annotations and diversity demonstrates high chemical diversity in proteins that are traditionally considered conserved. The effect of amino acid variations and PTMs on the resistance and adaptability of corals to the environment can be addressed at a more detailed and systematic epigenetic level. Even within a single species, a population of environmentally stressed organisms can exhibit several variants and proteoforms of the same protein as mutations, genetic splicing, and epigenetic mechanisms act in favor of survival. Special to the method is the ability to predict and annotate any proteoform, particularly when applied to histone families highly prone to PTMs. When applied to epigenetic studies of non-annotated, endangered species (e.g. corals), in the context of climate change and phenotypic responses, this method can support intervention strategies to increase species resilience for successful conservation and restoration efforts [3].

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

Funding

Data availability

The data underlying this article are available in the ProteomeXchange consortium and can be accessed with the unique identifier PXD059506.

The protein sequence data reported in this paper will appear in the UniProt Knowledgebase under the accession numbers C0HMF0 (H4 & H4.S), C0HMF1 (H2A & H2A.A), C0HMF2 (H2B.1), C0HMF3 (H2B.2 & H2B.2K), and C0HMF4 (H2B.3).

References

1. Torda G, Donelson JM, Aranda M *et al.* Rapid adaptive responses to climate change in corals. *Nature Clim Change* 2017;**7**:627–36. https://doi.org/10.1038/nclimate3374
2. van Oppen MJH, Gates RD, Blackall LL *et al.* Shifting paradigms in restoration of the world's coral reefs. *Global Change Biol* 2017;**23**:3437–48. https://doi.org/10.1111/gcb.13647
3. National Academies of Sciences, Engineering, and Medicine. *A Research Review of Interventions to Increase the Persistence and Resilience of Coral Reefs*. Washington, DC: The National Academies Press, 2019.
4. Hackerott S, Martell HA, Eirin-Lopez JM. Coral environmental memory: causes, mechanisms, and consequences for future reefs. *Trends Ecol Evol* 2021;**36**:1011–23. https://doi.org/10.1016/j.tree.2021.06.014
5. Precht W, Bruckner A, Aronson R *et al.* Endangered acroporid corals of the Caribbean. *Coral Reefs* 2002;**21**:41–2. https://doi.org/10.1007/s00338-001-0209-2
6. Miller M, Bourque A, Bohnsack J. An analysis of the loss of acroporid corals at Looe Key, Florida, USA: 1983–2000. *Coral Reefs* 2002;**21**:179–82. https://doi.org/10.1007/s00338-002-0228-7
7. Eddy TD, Lam VWY, Reygondeau G *et al.* Global decline in capacity of coral reefs to provide ecosystem services. *One Earth* 2021;**4**:1278–85. https://doi.org/10.1016/j.oneear.2021.08.016
8. Eirin-Lopez JM, Putnam HM. Marine environmental epigenetics. *Annu Rev Mar Sci* 2019;**11**:335–68. https://doi.org/10.1146/annurev-marine-010318-095114
9. Rodriguez-Casariego JA, Ladd MC, Shantz AA *et al.* Coral epigenetic responses to nutrient stress: histone H2A.X phosphorylation dynamics and DNA methylation in the staghorn coral *Acropora cervicornis. Ecol Evol* 2018;**8**:12193–207. https://doi.org/10.1002/ece3.4678
10. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. *Cell Res* 2011;**21**:381–95. https://doi.org/10.1038/cr.2011.22
11. Zink L-M, Hake SB. Histone variants: nuclear function and disease. *Curr Opin Genet Dev* 2016;**37**:82–9.
12. Zlatanova J, Bishop TC, Victor JM *et al.* The nucleosome family: dynamic and growing. *Structure* 2009;**17**:160–71. https://doi.org/10.1016/j.str.2008.12.016
13. Zhou X, Cain CE, Myrthil M *et al.* Epigenetic modifications are associated with inter-species gene expression variation in primates. *Genome Biol* 2014;**15**:547. https://doi.org/10.1186/s13059-014-0547-3
14. Nawaz K, Cziesielski MJ, Mariappan KG *et al.* Histone modifications and DNA methylation act cooperatively in regulating symbiosis genes in the sea anemone Aiptasia. *BMC Biol* 2022;**20**:265. https://doi.org/10.1186/s12915-022-01469-y
15. Moradian A, Kalli A, Sweredoski MJ *et al.* The top-down, middle-down, and bottom-up mass spectrometry approaches for characterization of histone variants and their post-translational modifications. *Proteomics* 2014;**14**:489–97. https://doi.org/10.1002/pmic.201300256
16. Zhang K, Tang H, Huang L *et al.* Identification of acetylation and methylation sites of histone H3 from chicken erythrocytes by high-accuracy matrix-assisted laser desorption ionization-time-of-flight, matrix-assisted laser desorption ionization-postsource decay, and nanoelectrospray ionization tandem mass spectrometry. *Anal Biochem* 2002;**306**:259–69.
17. Zhang L, Eugeni EE, Parthyn MR *et al.* Identification of novel histone post-translational modifications by peptide mass fingerprinting. *Chromosoma* 2003;**112**:77–86. https://doi.org/10.1007/s00412-003-0244-6
18. McKittrick E, Gafken PR, Ahmag K *et al.* Histone H3.3 is enriched in covalent modifications associated with active chromatin. *Proc Natl Acad Sci USA* 2004;**101**:1525–30. https://doi.org/10.1073/pnas.0308092100
19. Lermyte F, Valkenborg D, Loo JA *et al.* Radical solutions: principles and application of electron-based dissociation in mass spectrometry-based analysis of protein structure. *Mass Spectrom Rev* 2018;**37**:750–71. https://doi.org/10.1002/mas.21560
20. Bohrer BC, Merenbloom SI, Koeniger SL *et al.* Biomolecule analysis by ion mobility spectrometry. *Annual Rev Anal Chem* 2008;**1**:293–327. https://doi.org/10.1146/annurev.anchem.1.031207.113001
21. Fuller CN, Jeanne Dit Fouque K, Valadares Tose L *et al.* Online, bottom-up characterization of histone H4 4–17 isomers. *Anal Chem* 2024;**96**:17165–73. https://doi.org/10.1021/acs.analchem.4c02549
22. Jeanne Dit Fouque K, Wellmann M, Leyva Bombuse D *et al.* Effective discrimination of gas-phase peptide conformers using TIMS-ECD-ToF MS/MS. *Anal Methods* 2021;**13**:5216–23. https://doi.org/10.1039/D1AY01461G
23. Karch KR, Sidoli S, Garcia BA Identification and quantification of histone PTMs using high-resolution mass spectrometry. *Methods Enzymol* 2016;**574**:3–29.
24. Fuller CN, Valadares Tose L, Vitorino FNL *et al.* Bottom-up histone post-translational modification analysis using liquid chromatography, trapped ion mobility spectrometry, and tandem mass spectrometry. *J Proteome Res* 2024;**23**:3867–76. https://doi.org/10.1021/acs.jproteome.4c00177
25. Garcia BA, Ueberheide BM, Busby SA *et al.* Chemical derivatization of histones for facilitated analysis by mass spectrometry. *Nat Protoc* 2007;**2**:933–8. https://doi.org/10.1038/nprot.2007.106
26. Fernandez-Rojas M, Fuller CN, Valadares Tose L *et al.* Histone modification screening using liquid chromatography, trapped ion mobility spectrometry, and time-of-flight mass spectrometry. *J Vis Exp* 2024. https://doi.org/10.3791/65589
27. Jiang T, Hoover ME, Holt MV *et al.* Middle-down characterization of the cell cycle dependence of histone H4 posttranslational modifications and proteoforms. *Proteomics* 2018;**18**:1700442. https://doi.org/10.1002/pmic.201700442
28. Miller SA, Fouque JD, R K. *et al.* Trapped ion mobility spectrometry, ultraviolet photodissociation, and time-of-flight mass spectrometry for gas-phase peptide isobars/isomers/conformers

discrimination. *J Am Soc Mass Spectrom* 2022;**33**:1267–75. https://doi.org/10.1021/jasms.2c00091

29. Jeanne Dit Fouque K, Kaplan D, Voinov VG *et al*. Proteoform differentiation using tandem trapped ion mobility, electron capture dissociation, and ToF mass spectrometry. *Anal Chem* 2021;**93**:9575–82. https://doi.org/10.1021/acs.analchem.1c01735

30. Garabedian A, Baird MA, Porter J *et al*. Linear and differential ion mobility separations of middle-down proteoforms. *Anal Chem* 2018;**90**:2918–25. https://doi.org/10.1021/acs.analchem.7b05224

31. Berthias F, Cooper-Shepherd DA, Holck FHV *et al*. Full separation and sequencing of isomeric proteoforms in the middle-down mass range using cyclic ion mobility and electron capture dissociation. *Anal Chem* 2023;**95**:11141–8. https://doi.org/10.1021/acs.analchem.3c02120

32. Shliaha PV, Gorshkov VA, Kovalchuk SI *et al*. Middle-down proteomic analyses with ion mobility separations of endogenous isomeric proteoforms. *Anal Chem* 2020;**92**:2364–68. https://doi.org/10.1021/acs.analchem.9b05011

33. Smith LM, Kelleher NL, Linial M *et al*. Proteoform: a single term describing protein complexity. *Nat Methods* 2013;**10**:186–7. https://doi.org/10.1038/nmeth.2369

34. Donnelly DP, Rawlins CM, DeHart CJ *et al*. Best practices and benchmarks for intact protein analysis for top-down mass spectrometry. *Nat Methods* 2019;**16**:587–94. https://doi.org/10.1038/s41592-019-0457-0

35. Anderson LC, Karch KR, Ugrin SA *et al*. Analyses of histone proteoforms using front-end electron transfer dissociation-enabled orbitrap instruments. *Mol Cell Proteomics* 2016;**15**:975–88.

36. Jeanne Dit Fouque K, Miller SA, Pham K *et al*. Top-"double-down" mass spectrometry of histone H4 proteoforms: tandem ultraviolet-photon and mobility/mass-selected electron capture dissociations. *Anal Chem* 2022;**94**:15377–85. https://doi.org/10.1021/acs.analchem.2c03147

37. Wang Q, Fang F, Wang Q *et al*. Capillary zone electrophoresis-high field asymmetric ion mobility spectrometry-tandem mass spectrometry for top-down characterization of histone proteoforms. *Proteomics* 2024;**24**:2200389. https://doi.org/10.1002/pmic.202200389

38. Berthias F, Bilgin N, Mecinović J *et al*. Top-down ion mobility/mass spectrometry reveals enzyme specificity: separation and sequencing of isomeric proteoforms. *Proteomics* 2024;**24**:2200471. https://doi.org/10.1002/pmic.202200471

39. Walker JN, Lam R, Brodbelt JS. Enhanced characterization of histones using 193 nm ultraviolet photodissociation and proton transfer charge reduction. *Anal Chem* 2023;**95**:5985–93. https://doi.org/10.1021/acs.analchem.2c05765

40. Dang X, Singh A, Spetman BD *et al*. Label-free relative quantitation of isobaric and isomeric Human histone H2A and H2B variants by fourier transform ion cyclotron resonance top-down MS/MS. *J Proteome Res* 2016;**15**:3196–203. https://doi.org/10.1021/acs.jproteome.6b00414

41. Schachner LF, Jooß K, Morgan MA *et al*. Decoding the protein composition of whole nucleosomes with Nuc-MS. *Nat Methods* 2021;**18**:303–8. https://doi.org/10.1038/s41592-020-01052-9

42. Zhang H, Cui W, Wen J *et al*. Native electrospray and electron-capture dissociation FTICR mass spectrometry for top-down studies of protein assemblies. *Anal Chem* 2011;**83**:5598–606. https://doi.org/10.1021/ac200695d

43. Roberts DS, Loo JA, Tsybin YO *et al*. Top-down proteomics. *Nat Rev Methods Primers* 2024;**4**:38. https://doi.org/10.1038/s43586-024-00318-2

44. Li H, Nguyen HH, Ogorzalek Loo RR *et al*. An integrated native mass spectrometry and top-down proteomics method that connects sequence to structure and function of macromolecular

complexes. *Nature Chem* 2018;**10**:139–48. https://doi.org/10.1038/nchem.2908

45. Young NL, DiMaggio PA, Plazas-Mayorca MD *et al*. High throughput characterization of combinatorial histone codes. *Mol Cell Proteomics* 2009;**8**:2266–84.

46. Sidoli S, Schwämmle V, Ruminowicz C *et al*. Middle-down hybrid chromatography/tandem mass spectrometry workflow for characterization of combinatorial post-translational modifications in histones. *Proteomics* 2014;**14**:2200–11. https://doi.org/10.1002/pmic.201400084

47. Chen B, Brown KA, Lin Z *et al*. Top-down proteomics: ready for prime time? *Anal Chem* 2018;**90**:110–27. https://doi.org/10.1021/acs.analchem.7b04747

48. Xuan Y, Creese AJ, Horner JA *et al*. High-field asymmetric waveform ion mobility spectrometry (FAIMS) coupled with high-resolution electron transfer dissociation mass spectrometry for the analysis of isobaric phosphopeptides. *Rapid Comm Mass Spectrometry* 2009;**23**:1963–9. https://doi.org/10.1002/rcm.4101

49. Li Y, Cavet G, Zare RN *et al*. Fragment correlation mass spectrometry: determining the structures of biopolymers in a complex mixture without isolating individual components. *Proc Natl Acad Sci USA* 2024;**121**:e2409676121. https://doi.org/10.1073/pnas.2409676121

50. Rodríguez-Casariego JA, Mercado-Molina AE, Lemos LS *et al*. Multi-omic characterization of mechanisms contributing to rapid phenotypic plasticity in the coral *Acropora cervicornis* under divergent environments. *Coral Reefs* 2023;**43**:53–66.

51. Rivera-Casas C, Gonzalez-Romero R, Garduño RA *et al*. Molecular and biochemical methods useful for the epigenetic characterization of chromatin-associated proteins in bivalve molluscs. *Front Physiol* 2017;**8**:490. https://doi.org/10.3389/fphys.2017.00490

52. Fernandez-Lima F, Kaplan DA, Suetering J *et al*. Gas-phase separation using a trapped ion mobility spectrometer. *Int J Ion Mobil Spectrom* 2011;**14**:93–8. https://doi.org/10.1007/s12127-011-0067-8

53. Michelmann K, Silveira JA, Ridgeway ME *et al*. Fundamentals of trapped ion mobility spectrometry. *J Am Soc Mass Spectrom* 2015;**26**:14–24. https://doi.org/10.1007/s13361-014-0999-4

54. Silveira JA, Ridgeway ME, Park MA. High resolution trapped ion mobility spectrometery of peptides. *Anal Chem* 2014;**86**:5624–7. https://doi.org/10.1021/ac501261h

55. Bhanu NV, Sidoli S, Garcia BA. A workflow for ultra-rapid analysis of histone post-translational modifications with direct-injection mass spectrometry. *Bio Protoc* 2020;**10**:e3756. https://doi.org/10.21769/BioProtoc.3756

56. Sidoli S, Bhanu NV, Karch KR *et al*. Complete workflow for analysis of histone post-translational modifications using bottom-up mass spectrometry: from histone extraction to data analysis. *J Vis Exp* 2016;**111**:e54112.

57. The UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* 2021;**49**:D480–9. https://doi.org/10.1093/nar/gkaa1100

58. Shevchenko A, Sunyaev S, Loboda A *et al*. Charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time-of-flight mass spectrometry and BLAST homology searching. *Anal Chem* 2001;**73**:1917–26. https://doi.org/10.1021/ac0013709

59. Resemann A, Jabs W, Wiechmann A *et al*. Full validation of therapeutic antibody sequences by middle-up mass measurements and middle-down protein sequencing. *MAbs* 2016;**8**:318–30. https://doi.org/10.1080/19420862.2015.1128607

60. Zhang K, Williams KE, Huang L *et al*. Histone acetylation and deacetylation: identification of acetylation and methylation sites of HeLa histone H4 by mass spectrometry. *Mol Cell Proteomics* 2002;**1**:500–8.