

Chasing tails: Cathepsin-L improves structural analysis of histones by HX-MS

Malvina Papanastasiou[#], James Mullahoo, Katherine C DeRuff, Besnik Bajrami, Ioannis Karageorgos¹,

Stephen E Johnston, Ryan Peckner, Samuel A Myers, Steven A Carr, Jacob D Jaffe

The Broad Institute of MIT and Harvard, Cambridge, MA

¹Biomolecular Measurement Division, National Institute of Standards and Technology, Gaithersburg, MD;

Institute for Bioscience and Biotechnology Research, Rockville, MD.

[#] To whom correspondence should be addressed:

Malvina Papanastasiou

Broad Institute of MIT & Harvard

415 Main St, Cambridge

MA 02142

Tel.: (617) 714-8497

Email: malpap@broadinstitute.org

Keywords: hydrogen/deuterium exchange, mass spectrometry, histones, cathepsin-L, mononucleosomes

Running title: Cathepsin-L for histone tail analysis by HX-MS

Abbreviations:

EM: electron microscopy

ETD: electron transfer dissociation

HX-MS: hydrogen/deuterium exchange mass spectrometry

NMR: nuclear magnetic resonance

PBS: phosphate buffered saline

PTMs: post-translational modifications

TCEP: tris(2-carboxyethyl)-phosphine hydrochloride

UPLC: ultra-high performance liquid chromatography

ABSTRACT

The *N*-terminal regions (tails) of histone proteins are dynamic elements that protrude from the nucleosome and are involved in many aspects of chromatin organization. Their epigenetic role is well-established, and post-translational modifications present on these regions contribute to transcriptional regulation. Considering their biological significance, relatively few structural details have been established for histone tails, mainly due to their inherently disordered nature. While hydrogen/deuterium exchange mass spectrometry (HX-MS) is well-suited for the analysis of dynamic structures, it has seldom been employed in this context, presumably due to the poor *N*-terminal coverage provided by pepsin. Inspired from histone-clipping events, we profiled the activity of cathepsin-L under HX-MS quench conditions and characterized its specificity employing the four core histones (H2A, H2B, H3 and H4). Cathepsin-L demonstrated cleavage patterns that were substrate- and pH-dependent. Cathepsin-L generated overlapping *N*-terminal peptides about 20 amino acids long for H2A, H3 and H4 proving its suitability for the analysis of histone tails dynamics. We developed a comprehensive HX-MS method in combination with pepsin and obtained full sequence coverage for all histones. We employed our method to analyze histones H3 and H4. We observe rapid deuterium exchange of the *N*-terminal tails and cooperative unfolding (EX1 kinetics) in the histone-fold domains of histone monomers in-solution. Overall, this novel strategy opens new avenues for investigating the dynamic properties of histones that are not apparent from the crystal structures, providing insights into the structural basis of the histone code.

INTRODUCTION

Histones are highly conserved proteins that are integral components of the nucleosome, the repeating unit of chromatin. In the 'canonical nucleosome structure', 147 DNA base pairs wrap around a histone octamer that consists of a core H3-H4 tetramer flanked by two H2A-H2B heterodimers (1). Interactions within the octamer core and with the surrounding DNA promote a compact and rigid nucleosome structure bordered by flexible *N*- and *C*-terminal tails. An extensive array of covalent post-translational modifications (PTMs) decorating the 'tails' combinatorially regulate gene transcription, replication, and DNA repair, commonly known as the 'histone-code' (2). These PTMs are pivotal in modulating interactions of effector proteins with the nucleosomal DNA and with neighboring nucleosomes to modulate the formation of higher order chromatin structure (3). Decoding the information hidden in this extensive modification network and knowing the structure and dynamics of the tails is crucial in understanding regulation of gene expression.

Histones share a highly similar structure known as the histone-fold composed of the helix-loop-helix motif (4) complemented by highly basic, dynamic, flanking termini that extend outside the nucleosome boundaries. Given their disordered nature, parts of the histone tails are absent in high-resolution X-ray nucleosome structures reported to date (1, 5–7), although extension of the DNA sequence has resulted in improved electron density clarity of the tails (8). NMR spectroscopy has provided complete coverage of either all histone tails in single nucleosome particles (9) or of individual tails in condensed nucleosomal arrays (10–14). However, size limitations (15) prohibit the analysis of samples of increasing complexity (i.e. as would be the case of a histone/reader complex), a space in which cryo-EM has already shown great promise (16). Cryo-EM has been employed successfully for elucidating structures of nucleosomes alone or in complex with chromatin modifiers (17–19) and for deciphering the organization of higher-order chromatin (for a recent review see (20)). Finally, computational studies including coarse-grained models (21, 22) and molecular dynamics simulations (23–25) on the tails have provided additional information to the role these have on internucleosomal interactions and chromatin compaction.

Despite the wide applicability of hydrogen-deuterium exchange-mass spectrometry (HX-MS) towards the study of unstructured regions within folded proteins (26–28), there are only a few reports employing bottom-up (peptide level) HX-MS to the analysis of histones to-date (29–36) and these mainly

focus on features of the histone-fold domains. Histone dynamics of H3 variants and H4 have been reported in the context of tetramers, nucleosomes and nucleosomal arrays (29–31). Further, HX-MS has been employed to study interactions between histone dimers H2A-H2B and H3-H4 in complex with chaperones that facilitate chromatin assembly (32–36). In all these studies, pepsin, the most common enzyme for generating peptide fragments under HX quench conditions (37) was employed and for most histones provided poor coverage of the tails. Neprrosin was introduced recently as an alternative enzyme for histone studies producing *N*-terminal tail peptides 1-38 for H3 and 1-32 for H4 in acidic conditions, (38) demonstrating its potential applicability for future HX-MS studies. Other novel HX-MS proteases that have emerged over the past few years (39–45) have yet to be tested to this class of proteins; their reported cleavage preferences however do not seem to render them appropriate for generating measurable size peptides (7-25 amino acids) given the sequence of the tails. Recently, an alternative approach employing electron transfer dissociation (ETD) in combination with HX-MS top-down and middle-down approaches for this class of proteins was demonstrated (46). Despite the high-resolution histone maps obtained however, the widespread implementation of ETD in HX-MS studies lags behind typical bottom-up approaches, as the analysis, although seemingly straightforward, presents technical challenges (47–49).

In this study, we sought to investigate the suitability of cathepsin-L, a protease that has been shown to proteolytically process the amino-terminal tail of H3 *in vivo* (50), and to enhance coverage of histone tails in bottom-up HX-MS studies. Cathepsin-L, is one of the 11 cysteine cathepsin proteases that share a common catalytic mechanism and strong sequence similarity with the papain super-family of proteases. Cathepsins are synthesized as propeptides and are processed in the mature form either autocatalytically or by other proteases (51). Their role as degradative enzymes of the lysosome is well established (52), although their observed localization in the nucleus and in tumor microenvironments suggests a variety of extended functions, i.e. in cell development and differentiation (50), as well as in cancer progression (53). Cysteine cathepsins exhibit broad specificity and cleave various proteins *in vitro* at acidic pH (52). Cathepsin-L has been shown to generate a 'primary' cleavage site at H3 Ala21 and additional downstream residues (50, 54), an event of potential epigenetic importance, commonly referred to as 'histone clipping' (55). Additionally, cathepsin-L has been shown to generate multiple cleavage sites downstream of H2A Ala21 *in-vitro* (56). Taken together, these data suggest that cathepsin-L may be applicable to the generation

of histone *N*-terminal peptides of appropriate size for HX-MS studies. Inspired by these reports, we investigated the activity of cathepsin-L at HX-MS quench conditions for all core histones. We show enhanced *N*-terminal coverage for most of the proteins, and developed a method to investigate in-solution dynamics of monomeric H3 and H4 by HX-MS.

EXPERIMENTAL PROCEDURES

Materials

Human recombinant histones (H2A3; Q7L7L0 (M2502S), H2B2E; Q16778 (M2505S), H3.1; P68431 (M2503S) and H4; P62805 (M2504S)) were purchased from New England BioLabs Inc (Ipswich, MA). Mononucleosomes, recombinant human (16-0009) containing H2A1B (P04908), H2B1K (O60814), H3.1 (P68431), and H4 (P62805) were purchased from EpiCypher (Durham, NC). Human recombinant cathepsin-L was purchased either from PromoKine (purity >90%; PK-RP577-1135-5, PromoCell GmbH) or from BioVision (purity >90%; 1135-5, Milpitas, CA). Dulbecco's phosphate buffered saline (PBS; D8537), cathepsin-L inhibitor (RKLLW-NH₂, ≥97%; SCP0110), deuterium oxide (99.9 atom % D; 151882) and guanidine hydrochloride (≥98%; 50950) were purchased from Sigma-Aldrich (Saint Louis, MO). Tris(2-carboxyethyl)-phosphine hydrochloride (TCEP·HCl; 20491) and immobilized pepsin on cross-linked agarose beads (6%; 20343) were purchased from Thermo Scientific (Rockford, IL). Acetonitrile (ACN, 99.9%; A998) and formic acid (FA, >99.5% purity; A117) were from Fisher (Fair Lawn, NJ) and water from J.T.Baker (JT4218-3; Center Valley, PA). Leucine enkephalin (MS Leucine Enkephalin Kit; 700002456) used for lock mass in the MS instrument was from Waters (Milford, MA). SDS-PAGE gels (4-15%, Mini-PROTEAN TGX Precast Protein Gels) were from Bio-Rad (4561086; Hercules, CA) and staining reagent (EZBlue™ Coomassie Brilliant Blue G-250) was from Sigma-Aldrich (G1041; St Louis, MO).

In vitro enzyme activity analysis

Individual histones (1 uL of 1 ug/uL, PBS pH 7.5) were diluted with PBS (9 uL, pH ranging from 7.5 to 2.5, TCEP 5 mM), and cathepsin-L (1 uL of 0.5 ug/uL, 25 mM sodium acetate, pH 5.5, TCEP 5 mM) was immediately added to the samples. The enzymatic digestions were allowed to proceed for 2.5 min either on ice (H3) or at room temperature (H4, H2A3 and H2B.1) and were quenched with SCP0110 (1 uL of 240

uM in PBS, pH 2.5). Samples were subsequently subjected to electrophoresis under denaturing conditions and were stained with EZBlue™ following standard protocol.

Intact Protein Analysis and Deconvolution

Intact histones (2uL, 50 ng/uL in 5% FA) were flow injected and analyzed on a Q Exactive Plus Orbitrap MS (Thermo Fisher Scientific, Bremen) using a standard m/z range of 400-2000. MS parameters were: spray voltage 2 kV, source temperature 250 °C, AGC target 3e6, maximum injection time (IT) 200 ms and resolution (at m/z 200) 140,000. Data were acquired in profile mode and ESI mass spectra were deconvoluted using Intact software (Protein Metrics Inc., San Carlos CA).

In-solution digestions and analysis using a Q Exactive Plus MS

Individual histones (1.5 ug, PBS pH 7.5) were diluted with PBS (pH 2.5) and cathepsin-L (0.15 ug, 25 mM sodium acetate, pH 2.5) was added to the samples. For cathepsin-L activity experiments, digestions were performed in the presence of PBS containing TCEP, guanidine HCl, or both (at pH 2.5) at various concentrations (Results & Discussion). The enzymatic digestions were allowed to proceed for 10 min at either 37 °C or 0 °C and were quenched by boiling samples at 95 °C for 5 min. Samples were desalted on Empore C18 silica bead Stagetips (57) following standard protocol (58). Elution mixtures were lyophilized and reconstituted in 5% formic acid prior to nLC-MS/MS analysis. Peptides were injected using a Proxeon Easy nLC 1000 and analyzed on a Q Exactive Plus Orbitrap MS (Thermo Fisher Scientific, Bremen). Briefly, 2 uL of samples (1 ug) were loaded onto a microcapillary column (360 uM OD x 75 um ID) with an integrated electrospray emitter tip (10 um, New Objective) packed with 20 cm of ReproSil-Pur C18-AQ 1.9 um beads (Dr. Maisch, GmbH). The column was heated to 50 °C using a column heater (Phoenix S&T). Peptides were separated using a flow rate of 200 nl/min and the following gradient: 3% B to 40% B in 30 min, to 90% B in 4 min, keep at 90% B for 6 min and return to initial conditions in 1 min. Samples were analyzed in data-dependent analysis (DDA) mode using a Top-12 method. Ion source parameters were: spray voltage 2 kV, source temperature 250 °C. Full MS scans were acquired in the m/z range 300-2000, with an AGC target 3e6, maximum IT 20 ms and resolution (at m/z 200) 70,000. MS/MS parameters were as follows: AGC target 1e5, maximum IT 50 ms, loop count 10, isolation window 1.6 m/z , isolation offset 0.3 m/z , NCE 27,

resolution (at m/z 200) 17,500 and fixed first mass 100 m/z ; unassigned and singly charged ions were excluded from MS/MS. Data were acquired in centroid mode in both MS and MS/MS scans. Peptides were identified using SpectrumMill Proteomics Workbench (prerelease version B.06.01.202, Agilent Technologies). A non-specific enzyme search was performed against a fasta file containing sequences of human histones, pepsin, cathepsin and contaminant proteins identified upon tryptic digestion of the samples (52 entries). Peptide and fragment tolerances were at ± 20 ppm, minimum matched peak intensity 40% and peptide false discovery rates were calculated to be less than 1% using the target-decoy approach (59). No fixed or variable modifications were included in the search. Spectra with a score higher than 4 were accepted.

HX sample preparation and analysis using a QToF-MS.

For hydrogen exchange experiments, individual histones (3 μL , 0.4 $\mu\text{g}/\mu\text{L}$ in PBS; 10 mM Na_2HPO_4 , 1.8 mM KH_2PO_4 , 2.7 mM KCl and 147 mM NaCl, pH 7.35) were equilibrated on ice and were mixed with ice-chilled deuterated PBS (47 μL , final D_2O content during reaction 94% v/v), prepared by two cycles of lyophilization and reconstitution in D_2O . Samples were quenched at 10, 100, 1000 and 10000 s with TCEP (5 μL of 55 mM containing 13.8% FA) to a final pH 2.4 and were snap-frozen in liquid nitrogen. A single time point (10 s) was prepared similarly using mononucleosomes (3 μL of 0.4 $\mu\text{g}/\mu\text{L}$ total protein content). Non-deuterated samples (0 sec) were prepared in PBS and full deuteration (FD) controls were prepared by incubating proteins as described above for 10 min at 95 °C. Samples were prepared and analyzed in duplicates. Prior to LC-MS analysis, samples were thawed using a mini centrifuge and were mixed with ice-chilled cathepsin-L (5 μL , 100 ng/ μL in 5 mM TCEP, pH 2.5). Samples were incubated for 5 min at 3 °C using an Eppendorf™ Thermomixer™ R (Thermo Scientific, Rockford, IL) and were immediately injected into a nanoAcquity UPLC with HX technology (Waters). Samples were digested online for 1 min at 300 $\mu\text{L}/\text{min}$ using the Enzymate™ BEH Pepsin Column (130 Å, 2.1 x 30 mm, 5 μm , 186007233; Waters) at 0 °C and were subsequently loaded on an Acquity UPLC R BEH C18 VanGuard pre-column (130 Å, 1.7 μm , 2.1 x 5 mm, 186003975; Waters) using FA (0.23% (v/v)). Peptides were separated on an Acquity UPLC R BEH C18 analytical column (130 Å, 1.7 μm , 1 x 100 mm, 186002346; Waters) at 40 $\mu\text{L}/\text{min}$ using solvents A (0.23% v/v FA) and B (0.23% v/v FA in ACN). The following gradient was applied: 3% B to 10% B in 0.5

min, to 40% B in 7 min, to 60% B in 1.2 min, to 97%B in 0.3min; kept at 97% B for 0.6 min and returned to initial conditions in 0.4 min. Samples were analyzed on a Synapt XEVO G2-XS QToF MS (MassLynx™ 4.1, SCN 916; Waters). Spectra were collected in Resolution mode (m/z 350-1500) using the following parameters: capillary voltage 3 kV, cone voltage 40 V, ion source block temperature 80 °C, cone gas flow 100 L/h, desolvation gas flow 800 L/h at 150 °C, and nebulizer gas flow 6 bar. Step Wave values were: DC Offset, 15V; Wave Height 1, 5 V; Wave Height 2, 30 V; Wave Velocity 1, 150 m/s and Wave Velocity 2, 150 m/s. Leucine enkephalin (2 ng/uL in 50% [v/v] ACN and 0.23% [v/v] FA) was infused for lock mass spray at 4 uL/min; three spectra of 0.5 sec were acquired every 20 sec. For peptide identification, individual histones (1 ug each) were run in data-dependent acquisition (DDA) mode upon digestion with cathepsin-L, pepsin or a serial combination of both (Results & Discussion). MS survey scan time was 0.1 sec and the 5 most intense multiply-charged ions with >10000 intensity/sec were chosen for MS/MS. Fragmentation scans were acquired for m/z 100-1600 with a total ion current target of 15000 intensity/sec for up to 0.8 sec. A ramped collision energy profile of 18-40 eV was applied and fragmented precursor ions were excluded from further MS/MS for 8 sec. The lock mass spray and UPLC parameters were as described above. Peptides were identified using PLGS (Waters, Milford MA). For all histones, a non-specific enzyme search was performed with parameters as described above for SpectrumMill Proteomics Workbench identification. Spectra were recalibrated using Leucine-Enkephalin (556.2771 Da) for lock mass.

Analysis of deuterium uptake.

Peptide lists were subsequently loaded to DynamX 3.0 (Waters) and deuterium (D) uptake curves of individual peptides were calculated based on the assigned centroid m/z value of the isotopic cluster envelopes. D-uptake values were normalized using experimental values obtained from the analysis of FD controls divided by the deuterium fraction (0.94) in the solvent during the exchange and the theoretical maximum number of exchangeable amides (60). For peptides displaying bimodal distributions, data for each deuteration time point were imported into HX-Express2 (61) and processed using double binomial fittings. Long peptides (>21 residues) that presented convoluted mass envelopes and required triple binomial fitting were excluded from analysis.

RESULTS AND DISCUSSION

Cathepsin-L activity profiling at pH 2.5.

Cathepsin-L displays optimal activity at slightly acidic pH (~5.5), though activity in the cell nucleus at pH ~7-8 has also been reported (50). To test whether cathepsin-L is functional at pH 2.5 (required during HX-MS quench conditions (37)), we digested the four core histones individually over a pH gradient ranging from basic to acidic to test the enzyme's activity (Figure 1A). Cathepsin-L demonstrated cleavage patterns that were substrate and pH dependent, and interestingly, for all substrates exhibited optimum activity in the lower acidic range (~3.5-2.5). For H3.1, H2A and H2B, several sub-bands were detected across the entire pH range tested, although activity against H3.1 was reduced at pH 2.5. Digestions at higher pH values yielded fewer bands indicating lower enzyme activity, in accordance with similar *in vitro* experiments of cathepsin-L (73.3% sequence similarity) for H3.1 (50). In contrast, cathepsin-L cleaved H4 exclusively at pH values <5.5 with highest intensity fragments detected at pH 2.5. These results confirm previous reports showing that cathepsin-L has maximal activity at acidic pH and that cathepsin-L affects substrate kinetics in a pH-dependent manner (52).

We next employed nanoLC-MS/MS to identify peptides generated by cathepsin-L for the four histones (Supplemental Figure S1) and investigated cleavage preferences. Prior to digestion, intact histones were analyzed individually by ESI-MS, confirming that the major proteins species present had complete and unmodified sequences (~11-15 kDa, Supplemental Figure S2). Histones were digested in triplicate in a slightly reducing environment that enhances the enzyme's activity (Figure 1B). The enzyme's reproducibility was assessed by the number of common peptides identified within individual histones across replicates. Therefore, for H2A 57% of peptides was identified in three replicates, for H2B 53%, for H3.1 55% and for H4 65% (Supplemental Table S1). Of note, another ~20% of peptides was identified in two replicates for individual histones. The moderate reproducibility observed is attributed to notable secondary exopeptidase activity detected (Supplemental Figure S3A), concordant with cysteine cathepsins that are known to contain critical structural elements for both endopeptidase and exopeptidase activities (52). The median pairwise Pearson's correlation of peptide intensities detected in at least two replicates was 0.93 (Supplemental Figure S3B). The "primary" cleavage sites (derived from peptides accounting for ~80% of total ion current) are depicted in Supplemental Figure S3C. For H2A3, cleavage at G44 is in line with

cathepsin-L *in vitro* proteolysis experiments of H2A2A (98% homologous to H2A3) at pH 6 reported previously (56). For H3.1, the “primary” cleavage sites lie towards the C-terminus of the protein. However, cleavage sites at the proteins’ *N*-terminus agree with H3.1 fragments generated *in vivo* by cathepsin-L (50). For H4, proteolysis using cathepsin-L has not yet been reported in the literature. Our studies indicate that R23 is the most abundant amongst “primary” cleavage sites, with a few more detected across the protein’s length.

We then scanned for optimal concentrations of reducing agent (TCEP) and denaturant (guanidine hydrochloride), typically added to the samples during the quenching step (0 °C, pH 2.5) and prior to proteolysis to improve digestion efficiency (37). Our results indicate that a slightly reducing environment (5 mM) and no denaturant result in optimal Cathepsin-L activity (Fig. 1B-C). This is consistent with maintenance of a reduced active site Cys while leaving structural disulfides intact.

To next determine amino acid cleavage preferences of cathepsin-L at pH 2.5, we combined all histone peptides identified above and discarded redundant cleavage sites (194 peptides remaining). We aligned peptides (P5 to P4’ (62)) and loaded them to the iceLogo web application (63). To account for the high fraction of positively charged amino acids, as a reference set we employed the four histones and cleaved consecutive amino acids across their entire sequences to generate 9 amino acid peptide sequences (455 peptides). Comparison of the frequency percentage of an amino acid at locations P5 to P4’ between the experimental and reference set revealed a preference for valine, leucine and isoleucine in P2 (positive values in Figure 2), in agreement with previous studies on cathepsin-L cleavage of hemoglobin (64). In H3, the P2 residue (L20) has been shown to be essential for recognition by cathepsin-L, whereas substitution of P1 and P3 has no significant impact on the proteolytic activity of the enzyme (65).

Enhanced histone *N*-terminus coverage by cathepsin-L/Pepsin under HX-MS conditions

Next, we assessed the ability of cathepsin-L to generate peptides from the *N*-terminal regions (<50 amino acids) of histones that have been difficult to access using pepsin in earlier HX studies (29, 30, 33, 34, 36). Although there is no well-defined ideal peptide fragment length in HX-MS studies, these range typically between ~7-15 residues (66). In addition, generation of overlapping peptides is highly desirable for improved structural resolution (67). We compared peptide maps generated using cathepsin-L

proteolysis, online pepsin digestion and in-tandem cathepsin-L/online pepsin digestion under HX-MS conditions (cathepsin-L:histone ratio 1:2, 3 °C) (Supplemental Table S2). For H3, pepsin produced long peptides (~50 amino acids), whereas cathepsin-L generated similar size peptides but also additional overlapping ones (1-21 and 1-22), improving resolution in this region (Figure 3, Table 1, Supplemental Table 2). Peptide 1-21 is the shortest generated H3 peptide (at pH 2.5) reported in the literature to date (neprosin produced peptide 1-32 at similar pH (38)) and contains five amino acids that can be post-translationally modified (K4, K9, S10, K14 and K18), emphasizing the challenging nature of the H3 tail. For H4, cathepsin-L generated peptides ~23 amino acids long and several complementary ones starting at D24, corroborating R23 as a “primary” cleavage site. Pepsin cleaved H4 at D24 generating peptides 1-24 and 25-49, partially demonstrated in a previous HX study of the H3-H4 dimer (29). Combined use of both cathepsin-L and pepsin almost doubled the number of H4 peptides and overlapping peptides and was therefore preferred despite a slight increase observed in the average length that was the result of the larger peptides generated by pepsin. For H2A each protease generated several peptides with few of them being common amongst cathepsin-L and pepsin; five overlapping peptides with overhanging C-terminal residues were detected up to residue G28, providing the highest resolution among histones. Finally, pepsin proved better for H2B giving a shorter peptide (1-15) and its complementary one (16-40), whereas cathepsin-L produced peptides that were all >37 amino acids in length. The use of both proteases in this case however resulted in shorter average peptide lengths and greater sequence overlap. Given the highly basic content of the *N*-termini, detected peptides were highly charged and therefore excellent candidates for HX studies employing ETD for single-residue resolution (68). In contrast to the tails, both proteases generated many overlapping peptides for the remaining protein sequences (beyond amino acid 50), emphasizing the contrast in the amino acid composition between the tails and the core histone-fold regions. Using cathepsin-L, we were able to obtain full sequence coverage for H2A and H4, and ~85% coverage for H2B and H3. Overall, the use of cathepsin-L/online pepsin digestion in-tandem proved more efficient compared to each protease alone, evident by the high number of peptides identified and improved resolution. Of note, the average peptide length was not impacted by the use of single or dual protease digestion and was in the range of 18-25 amino acids. We further tested the developed digestion scheme in intact mononucleosomes containing the histone octamer (H2A1B, H2B1K, H3 and H4). Overlapping *N*-terminal peptides for H2A (1-

24, 1-28), H3 (1-21, 1-23) and H4 (1-23) were detected and full coverage was obtained for all histones (with the exception of H3, 93%), indicating that our method may be deployed successfully in a mononucleosomal context (Supplemental Table S2).

Flexible H3 and H4 tails in monomers

During DNA replication, newly synthesized histones are recognized by chaperones that coordinate their nuclear import and deposit them onto DNA to form nucleosomes. Most histone chaperones have a preference for binding and importing histone heterodimers (69), and HX structural information has been focused mainly on the analysis of heterodimers (34), tetramers (29) and nucleosomes (30, 31) or on heterodimer-chaperone interactions (32–36), disregarding monomeric states. Recent evidence, however, has shown that H3 and H4 are predominantly monomeric in the cytosol and can be rapidly imported into the nucleus bound tightly to importin- β proteins (70). Obtaining structural information at the monomer level, therefore, is crucial for delineating structural elements governing histone-chaperone interactions. We employed our method using dual cathepsin-L/pepsin digestion to probe conformational dynamics of the H3 and H4 monomers. The monomeric states of the proteins were confirmed using size-exclusion chromatography (Supplemental Figure S4). Proteins were incubated in D₂O for different periods of time (10 s, 100 s, 1000 s and 10000 s) and exchange was performed on ice to slow HX rates (71). Upon quenching, proteins were digested for a total of 5 min using cathepsin-L followed by online pepsin digestion and analyzed by LC-MS (Figure 4A). The D-uptake of individual profiles was subsequently calculated; for peptides with bimodal isotope distributions, the fraction of each population was modeled further (Figure 4B). In total, 36 peptides were generated for H3 (87% coverage) and 56 for H4 (100% coverage) monomers and a subset of those was analyzed for their D-content (23 for H3 and 30 for H4), depicted in the form of a heat map for individual time points (Figure 5A and 5B) and onto the histone structures (Figure 5C and 5D). D-recoveries were 83% for H3 and 80% for H4 (average values, calculated as described in (60)). D-uptake graphs for individual peptides are shown in Supplemental Figure S5.

Both monomers demonstrated highly flexible *N*-termini, evident by extensive deuteration at the earliest time point. For H3, all amide protons exchanged with deuterons (peptides 1-21 and 1-22) indicating that this region is unstructured. To our knowledge this is the first time that dynamics of the H3 *N*-terminus

has been captured at this peptide resolution by bottom-up approaches. In previous HX-MS studies of H3 and variants, poor coverage of the *N*-terminus was obtained using pepsin (29–31, 34). In the present study, the higher coverage obtained was made possible by use of cathepsin-L. Exchange profiles for the longer overlapping peptides (1-47, 1-48 and 1-49) extending to the first residues of helix α N (46-57) were obtained for the monomers. Deuterium exchange indicated some slight increase in protection (~88% at 10 s vs. 100% for 1-1-22 of H3), suggesting participation of amide protons in H-bonds in the 23-49 region.

For H4, the *N*-terminal tail was characterized by peptides 1-23, 1-24 and few longer (1-38, 1-48, 1-49) peptides as well as their *N*-terminal flanking ones (24-38, 24-49, 25-49) that contain segments of helix α 1 and loop L1. For peptides 1-23 and 1-24, D-uptake was ~80% throughout the time-course of the experiment. This could be due to some helical content that has been observed in molecular dynamics studies (73) and corroborated further by circular dichroism for 1-23 of unmodified H4 (17% helical content) (74). Further, previous HX studies of the H2A-H2B dimer have shown that ionic strength plays a critical role in secondary structure formation with lower salt concentrations (60 mM NaCl) leading to partial unfolding and higher salt concentrations enhancing H-bond formation (33). These findings suggest that the tail of the H4 monomer may indeed form a stable secondary structure at our experimental conditions (147 mM NaCl). In the absence of salt, the monomeric H4 tail showed rapid exchange at the earliest time point corroborating further the dependence of structure formation as a function of ionic strength; high salt concentration that was required to maintain the H3-H4 tetrameric complex resulted in reduced deuterium uptake (46). Further, in previous studies of the H3-H4 heterodimer and heterotetramer, H4 peptide 1-24 (pepsin generated) showed rapid exchange at the earliest time point (29, 34), suggesting that the chemical environment also plays a major role in the structural conformation of the tails (34).

Extensive EX1 kinetics in the histone folds of H3 and H4 monomers

Two mechanisms have been proposed to describe HX kinetics of a protein under native conditions, termed EX2 and EX1 (76, 77). In EX2, the HX chemical rate is slower than the protein's structural refolding rate and the protein goes through many unfolding and refolding events prior to a successful exchange event; in EX1, the HX rate is faster than the refolding rate and exchange will occur on multiple exposed residues prior to refolding. EX2 kinetics are presented with a single isotopic distribution whereas EX1 with

two distinct mass envelopes, with the lower mass envelope representing the 'closed' state and the higher mass envelope representing the 'open' state of a protein (Figure 4B). Dynamics of the histone folds (4) were dominated by bimodal isotopic distributions (Supplemental Figure S6), indicating exchange kinetics in the EX1 regime that occur when the rate of refolding is slower than the rate of chemical exchange (77). As a result, for regions in which we observed EX1 kinetics, we included a second heat map for each protein, with the fast-exchanging regions corresponding to the 'open' and the slow-exchanging regions corresponding to the 'closed' state (Figure 5A and 5B) and mapped exchange profiles for both conformations on respective structures (Figure 5C and 5D) (1). To avoid erroneous assignment of EX1 kinetics, blanks were injected to preclude carryover between samples (78) and the StepWave ion guide that has shown to cause artificial bimodal patterns for highly charged peptides (79) was optimized. For all time points we further calculated the relative population of each state in solution (61) (Supplemental Table S3, Supplemental Figure S7). For both histones, relative populations changed over time, suggesting EX1 kinetics rather than two distinct but static conformations in solution that would be indicated if the relative population remained constant (61).

The histone fold is a conserved structural motif, composed of three α -helices ($\alpha 1$, $\alpha 2$, $\alpha 3$) connected via two loops (L1, L2) that promotes histone dimer formation (4); in H3, there is an extra helix preceding the histone fold termed αN (1). Bimodal distributions for H3 residues 48-125 were detected, entailing helix αN and the histone fold motif (denoted as $\alpha 1$ -L1- $\alpha 2$ -L2- $\alpha 3$ in Figure 5A). Exceptions to this were the central and carboxy-terminal part of $\alpha 3$ and the C-terminal of the protein that demonstrated EX2 kinetics. Helix αN (residues 48-61) was mostly rigidly packed (~ 0.8 in 'closed' state at 10 s), evident by the low D-uptake over the time course of the experiment ($< 10\%$); the remaining fraction (~ 0.2) was unfolded. Helix αN has previously been shown to adopt its state in a context-dependent manner, appearing unfolded in (H3-H4)₂ heterotetramers (29, 34, 46) but stably folded in nucleosomes (1, 30, 46), nucleosomal arrays (31) and in complex with chaperones (34). $\alpha 1$ -L1 indicated dynamic regions ($\sim 30\%$ overall D-uptake in the time course of the experiment) in contrast to $\alpha 2$ that showed the least extent of deuterium incorporation in the histone fold. Segment $\alpha 2$ -L2- $\alpha 3$ that promotes tetramer formation of the H3-H4 histone pairs (1), is mainly unfolded (> 0.75 of population, $> 89\%$ D-uptake). EX1 kinetics for this region have been observed previously in the (H3-H4)₂ tetramer at high ionic strength and have been explained by the transient loss of tetramerization,

further supported by the abolishment of the biphasic behavior upon complex formation of the H3-H4 dimer with DAXX that stabilized this region (34). Further, this region in monomeric H3 and in an H3-H4 dimer complex with ASF1 has been identified to interact with chaperone NASP (80). Taken together, our data suggest that in physiological concentrations of salt and in the absence of H4, this region presents a high degree of flexibility that may be required to accommodate the breadth of the histone chaperone network. For H4, peptides within the histone fold exhibited extensive EX1 kinetic profiles with $\alpha 1$ and $\alpha 3$ helices occupying mainly an 'open' (>0.7 at 10 s) and helix $\alpha 2$ mainly a 'closed' state (~ 0.8 at 10 s). For $\alpha 1$ and $\alpha 3$, the 'closed' population exchanged $>40\%$ deuterium at 10 s indicating partially folded elements (Supplemental Figure S5B). This contrasts with $\alpha 2$, where D-uptake was $<20\%$ in the time course of the experiment indicating a tightly folded helix (Supplemental Figure S7). EX1 behavior of the $\alpha 2$ helix has been observed in the (H3-H4)₂ tetramer, but not when H4 forms a tetramer complex with the H3 variant CENP-A (29). CENP-A appears to rigidify helix $\alpha 2$ and minimize exchange. Further, cooperative unfolding detected previously for both H2A and H2B in the H2A-H2B heterodimer was also localized at helical regions of the histone fold. Similarly to the H3-H4 dimer with DAXX (34), bimodal distributions were no longer observed upon complex formation of the H2A-H2B dimer with Nap1 (33).

Overall, EX1 kinetics seem to be a feature of certain regions of histones, irrespective of whether these are in a monomeric state, engaged in dimers, tetramers or in complexes with chaperones. Differences in experimental conditions (such as temperature and salt concentrations, affecting H-bond formation) render comparison of prior studies to our current work difficult with respect to establishment of the exact boundaries of regions that exhibit EX1 exchange kinetics. However, the EX1 kinetics detected in our study for monomeric histones are extensive and expand the entire protein sequence (omitting the *N*-terminal tails), suggesting high conformational flexibility, in contrast to histones engaged in complexes where EX1 behavior is confined to short regions. In light of recent evidence showing that H3 and H4 monomers associate with importin-beta proteins and are imported into the nucleus in a monomeric state expands their interaction network further (70). We presume that a high degree of flexibility is a feature of histones required to counterbalance the breadth of chaperones that regulate their nuclear import and assembly into nucleosomes and provide exciting directions for future studies.

In summary, cathepsin-L in tandem with pepsin provided a high-resolution peptide map of histones suitable for HX-MS, including their tails that have thus far eluded crystallographic characterization. The introduction of cathepsin-L as a new tool to facilitate dual-protease techniques could be highly impactful for the study of important biological molecules beyond histones. Such cases may include protein classes known to be challenging for HX-MS, such as antibodies (81), or proteins which have been identified as cathepsin's substrates (i.e. elastins, collagens and proteoglycans) (52). For example, the fusion viral glycoprotein F has been reported to be activated through cleavage by cathepsin-L in cells (82) but was presented with moderate coverage in an HX-MS study employing pepsin only (83). We think that the advance we have shown here for histones will encourage the HX-MS community to consider cathepsin-L when probing challenging proteins, and therefore the biggest potential impact is to the HX-MS community rather than the narrow histone structural biology one.

Acknowledgements

This work was supported by a grant from the National Institute of Health (NIH) to J.D.J. at the Broad Institute of MIT & Harvard (U54-HG008097). The authors would like to acknowledge Leland Mayne (University of Pennsylvania) and Miklos Guttman (University of Washington) for discussions with respect to the analysis of bimodal distributions.

Data availability

The original mass spectra have been deposited in the public proteomics repository MassIVE and are accessible at <ftp://MSV000083315@massive.ucsd.edu>

When providing the dataset password: histones. If requested, also provide the username:

MSV000083315. This data will be made public upon acceptance of the manuscript.

Footnotes

Author contributions: M.P., S.A.M. designed research; M.P., J.M., K.C.D., B.B., I.K., performed research; B.B., S.E.J., contributed new reagents/analytic tools; M.P., J.M., K.C.D., I.K., R.P., S.A.M., analyzed data; M.P., S.A.M., S.A.C., J.D.J. wrote the paper; S.A.C., J.D.J, supervised research

Conflict of interest

The authors do not declare any conflict of interest.

REFERENCES

1. Luger, K., Rechsteiner, T. J., Flaus, A. J., Wayne, M. M. Y., and Richmond, T. J. (1997) Characterization of nucleosome core particles containing histone proteins made in bacteria. *J. Mol. Biol.* 272, 301–311
2. Strahl, B. D., and Allis, C. D. (2000) The language of covalent histone modifications. *Nature* 403, 41–45
3. Pepenella, S., Murphy, K. J., and Hayes, J. J. (2014) Intra- and inter-nucleosome interactions of the core histone tail domains in higher-order chromatin structure. *Chromosoma* 123, 3–13
4. Arents, G., Burlingame, R. W., Wang, B. C., Love, W. E., and Moudrianakis, E. N. (1991) The nucleosomal core histone octamer at 3.1 Å resolution: a tripartite protein assembly and a left-handed superhelix. *Proc. Natl. Acad. Sci. U. S. A.* 88, 10148–10152
5. White, C. L., Suto, R. K., and Luger, K. (2001) Structure of the yeast nucleosome core particle reveals fundamental changes in internucleosome interactions. *EMBO J.* 20, 5207–5218
6. Tachiwana, H., Kagawa, W., Osakabe, A., Kawaguchi, K., Shiga, T., Hayashi-Takanaka, Y., Kimura, H., and Kurumizaka, H. (2010) Structural basis of instability of the nucleosome containing a testis-specific histone variant, human H3T. *Proc. Natl. Acad. Sci. U. S. A.* 107, 10454–10459
7. Aguilar-Gurrieri, C., Larabi, A., Vinayachandran, V., Patel, N. A., Yen, K., Reja, R., Ebong, I. O., Schoehn, G., Robinson, C. V., Pugh, B. F., and Panne, D. (2016) Structural evidence for Nap1-dependent H2A-H2B deposition and nucleosome assembly. *EMBO J.* 35, 1465–1482
8. Davey, C. A., Sargent, D. F., Luger, K., Maeder, A. W., and Richmond, T. J. (2002) Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J. Mol. Biol.* 319, 1097–1113
9. Zhou, B.-R., Feng, H., Ghirlando, R., Kato, H., Gruschus, J., and Bai, Y. (2012) Histone H4 K16Q mutation, an acetylation mimic, causes structural disorder of its N-terminal basic patch in the nucleosome. *J. Mol. Biol.* 421, 30–37
10. Kato, H., Gruschus, J., Ghirlando, R., Tjandra, N., and Bai, Y. (2009) Characterization of the N-terminal tail domain of histone H3 in condensed nucleosome arrays by hydrogen exchange and NMR. *J. Am. Chem. Soc.* 131, 15104–15105
11. Gao, M., Nadaud, P. S., Bernier, M. W., North, J. A., Hammel, P. C., Poirier, M. G., and Jaroniec, C. P. (2013) Histone H3 and H4 N-terminal tails in nucleosome arrays at cellular concentrations probed by magic angle spinning NMR spectroscopy. *J. Am. Chem. Soc.* 135, 15278–15281
12. Hilliard, P. R., Jr, Smith, R. M., and Rill, R. L. (1986) Natural abundance carbon-13 nuclear magnetic resonance studies of histone and DNA dynamics in nucleosome cores. *J. Biol. Chem.* 261, 5992–5998
13. Smith, R. M., and Rill, R. L. (1989) Mobile histone tails in nucleosomes. Assignments of mobile segments and investigations of their role in chromatin folding. *J. Biol. Chem.* 264, 10574–10581
14. Cary, P. D., Moss, T., and Bradbury, E. M. (1978) High-resolution proton-magnetic-resonance studies of chromatin core particles. *Eur. J. Biochem.* 89, 475–482
15. Frueh, D. P., Goodrich, A. C., Mishra, S. H., and Nichols, S. R. (2013) NMR methods for structural studies of large monomeric and multimeric proteins. *Curr. Opin. Struct. Biol.* 23, 734–739
16. Kühlbrandt, W. (2014) Biochemistry. The resolution revolution. *Science* 343, 1443–1444

17. Chaban, Y., Ezeokonkwo, C., Chung, W.-H., Zhang, F., Kornberg, R. D., Maier-Davis, B., Lorch, Y., and Asturias, F. J. (2008) Structure of a RSC-nucleosome complex and insights into chromatin remodeling. *Nat. Struct. Mol. Biol.* 15, 1272–1277
18. Wilson, M. D., Benlekbir, S., Fradet-Turcotte, A., Sherker, A., Julien, J.-P., McEwan, A., Noordermeer, S. M., Sicheri, F., Rubinstein, J. L., and Durocher, D. (2016) The structural basis of modified nucleosome recognition by 53BP1. *Nature* 536, 100–103
19. Zocco, M., Marasovic, M., Pisacane, P., Bilokapic, S., and Halic, M. (2016) The Chp1 chromodomain binds the H3K9me tail and the nucleosome core to assemble heterochromatin. *Cell Discov* 2, 16004
20. Wilson, M. D., and Costa, A. (2017) Cryo-electron microscopy of chromatin biology. *Acta Crystallogr D Struct Biol* 73, 541–548
21. Kenzaki, H., and Takada, S. (2015) Partial Unwrapping and Histone Tail Dynamics in Nucleosome Revealed by Coarse-Grained Molecular Simulations. *PLoS Comput. Biol.* 11, e1004443
22. Arya, G., and Schlick, T. (2006) Role of histone tails in chromatin folding revealed by a mesoscopic oligonucleosome model. *Proc. Natl. Acad. Sci. U. S. A.* 103, 16236–16241
23. Biswas, M., Voltz, K., Smith, J. C., and Langowski, J. (2011) Role of histone tails in structural stability of the nucleosome. *PLoS Comput. Biol.* 7, e1002279
24. Erler, J., Zhang, R., Petridis, L., Cheng, X., Smith, J. C., and Langowski, J. (2014) The role of histone tails in the nucleosome: a computational study. *Biophys. J.* 107, 2911–2922
25. Shaytan, A. K., Armeev, G. A., Goncarencu, A., Zhurkin, V. B., Landsman, D., and Panchenko, A. R. (2016) Coupling between Histone Conformations and DNA Geometry in Nucleosomes on a Microsecond Timescale: Atomistic Insights into Nucleosome Functions. *J. Mol. Biol.* 428, 221–237
26. Papanastasiou, M., Koutsogiannaki, S., Sarigiannis, Y., Geisbrecht, B. V., Ricklin, D., and Lambris, J. D. (2017) Structural Implications for the Formation and Function of the Complement Effector Protein iC3b. *J. Immunol.* 198, 3326–3335
27. Li, X., Tao, Y., Murphy, J. W., Scherer, A. N., Lam, T. T., Marshall, A. G., Koleske, A. J., and Boggon, T. J. (2017) The repeat region of cortactin is intrinsically disordered in solution. *Sci. Rep.* 7, 16696
28. Killoran, R. C., Sowole, M. A., Halim, M. A., Konermann, L., and Choy, W.-Y. (2016) Conformational characterization of the intrinsically disordered protein Chibby: Interplay between structural elements in target recognition. *Protein Sci.* 25, 1420–1429
29. Black, B. E., Foltz, D. R., Chakravarthy, S., Luger, K., Woods, V. L., Jr, and Cleveland, D. W. (2004) Structural determinants for generating centromeric chromatin. *Nature* 430, 578–582
30. Black, B. E., Brock, M. A., Bédard, S., Woods, V. L., Jr, and Cleveland, D. W. (2007) An epigenetic mark generated by the incorporation of CENP-A into centromeric nucleosomes. *Proc. Natl. Acad. Sci. U. S. A.* 104, 5008–5013
31. Panchenko, T., Sorensen, T. C., Woodcock, C. L., Kan, Z.-Y., Wood, S., Resch, M. G., Luger, K., Englander, S. W., Hansen, J. C., and Black, B. E. (2011) Replacement of histone H3 with CENP-A directs global nucleosome array condensation and loosening of nucleosome superhelical termini. *Proc. Natl. Acad. Sci. U. S. A.* 108, 16588–16593
32. Bassett, E. A., DeNizio, J., Barnhart-Dailey, M. C., Panchenko, T., Sekulic, N., Rogers, D. J., Foltz, D. R., and Black, B. E. (2012) HJURP uses distinct CENP-A surfaces to recognize and to stabilize CENP-A/histone H4 for centromere assembly. *Dev. Cell* 22, 749–762
33. D’Arcy, S., Martin, K. W., Panchenko, T., Chen, X., Bergeron, S., Stargell, L. A., Black, B. E., and

- Luger, K. (2013) Chaperone Nap1 shields histone surfaces used in a nucleosome and can put H2A-H2B in an unconventional tetrameric form. *Mol. Cell* 51, 662–677
34. DeNizio, J. E., Elsässer, S. J., and Black, B. E. (2014) DAXX co-folds with H3.3/H4 using high local stability conferred by the H3.3 variant recognition residues. *Nucleic Acids Res.* 42, 4318–4331
 35. Liu, W. H., Roemer, S. C., Zhou, Y., Shen, Z.-J., Dennehey, B. K., Balsbaugh, J. L., Liddle, J. C., Nemkov, T., Ahn, N. G., Hansen, K. C., Tyler, J. K., and Churchill, M. E. (2016) The Cac1 subunit of histone chaperone CAF-1 organizes CAF-1-H3/H4 architecture and tetramerizes histones. *Elife* 5,
 36. Mattioli, F., Gu, Y., Balsbaugh, J. L., Ahn, N. G., and Luger, K. (2017) The Cac2 subunit is essential for productive histone binding and nucleosome assembly in CAF-1. *Sci. Rep.* 7, 46274
 37. Engen, J. R., and Wales, T. E. (2015) Analytical Aspects of Hydrogen Exchange Mass Spectrometry. *Annu. Rev. Anal. Chem.* 8, 127–148
 38. Schröder, C. U., Lee, L., Rey, M., Sarpe, V., Man, P., Sharma, S., Zabrouskov, V., Larsen, B., and Schriemer, D. C. (2017) Neprosin, a Selective Prolyl Endoprotease for Bottom-up Proteomics and Histone Mapping. *Mol. Cell. Proteomics* 16, 1162–1171
 39. Cravello, L., Lascoux, D., and Forest, E. (2003) Use of different proteases working in acidic conditions to improve sequence coverage and resolution in hydrogen/deuterium exchange of large proteins. *Rapid Commun. Mass Spectrom.* 17, 2387–2393
 40. Zhang, H.-M., Kazazic, S., Schaub, T. M., Tipton, J. D., Emmett, M. R., and Marshall, A. G. (2008) Enhanced digestion efficiency, peptide ionization efficiency, and sequence resolution for protein hydrogen/deuterium exchange monitored by Fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* 80, 9034–9041
 41. Rey, M., Man, P., Brandolin, G., Forest, E., and Pelosi, L. (2009) Recombinant immobilized rhizopuspepsin as a new tool for protein digestion in hydrogen/deuterium exchange mass spectrometry. *Rapid Commun. Mass Spectrom.* 23, 3431–3438
 42. Rey, M., Yang, M., Burns, K. M., Yu, Y., Lees-Miller, S. P., and Schriemer, D. C. (2013) Nepenthesin from Monkey Cups for Hydrogen/Deuterium Exchange Mass Spectrometry. *Mol. Cell. Proteomics* 12, 464–472
 43. Kadek, A., Mrazek, H., Halada, P., Rey, M., Schriemer, D. C., and Man, P. (2014) Aspartic protease nepenthesin-1 as a tool for digestion in hydrogen/deuterium exchange mass spectrometry. *Anal. Chem.* 86, 4287–4294
 44. Yang, M., Hoepfner, M., Rey, M., Kadek, A., Man, P., and Schriemer, D. C. (2015) Recombinant Nepenthesin II for Hydrogen/Deuterium Exchange Mass Spectrometry. *Anal. Chem.* 87, 6681–6687
 45. Tsiatsiani, L., Akeroyd, M., Olsthoorn, M., and Heck, A. J. R. (2017) *Aspergillus niger* Prolyl Endoprotease for Hydrogen-Deuterium Exchange Mass Spectrometry and Protein Structural Studies. *Anal. Chem.* 89, 7966–7973
 46. Karch, K. R., Coradin, M., Zandarashvili, L., Kan, Z.-Y., Gerace, M., Englander, S. W., Black, B. E., and Garcia, B. A. (2018) Hydrogen-Deuterium Exchange Coupled to Top- and Middle-Down Mass Spectrometry Reveals Histone Tail Dynamics before and after Nucleosome Assembly. *Structure*,
 47. Wei, H., Mo, J., Tao, L., Russell, R. J., Tymiak, A. A., Chen, G., Iacob, R. E., and Engen, J. R. (2014) Hydrogen/deuterium exchange mass spectrometry for probing higher order structure of protein therapeutics: methodology and applications. *Drug Discov. Today* 19, 95–102
 48. Iacob, R. E., and Engen, J. R. (2012) Hydrogen exchange mass spectrometry: are we out of the quicksand? *J. Am. Soc. Mass Spectrom.* 23, 1003–1010

49. Landgraf, R. R., Chalmers, M. J., and Griffin, P. R. (2012) Automated hydrogen/deuterium exchange electron transfer dissociation high resolution mass spectrometry measured at single-amide resolution. *J. Am. Soc. Mass Spectrom.* 23, 301–309
50. Duncan, E. M., Muratore-Schroeder, T. L., Cook, R. G., Garcia, B. A., Shabanowitz, J., Hunt, D. F., and Allis, C. D. (2008) Cathepsin L Proteolytically Processes Histone H3 During Mouse Embryonic Stem Cell Differentiation. *Cell* 135, 284–294
51. Turk, B., Turk, D., and Turk, V. (2000) Lysosomal cysteine proteases: more than scavengers. *Biochim. Biophys. Acta* 1477, 98–111
52. Turk, V., Stoka, V., Vasiljeva, O., Renko, M., Sun, T., Turk, B., and Turk, D. (2012) Cysteine cathepsins: From structure, function and regulation to new frontiers. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1824, 68–88
53. Olson, O. C., and Joyce, J. A. (2015) Cysteine cathepsin proteases: regulators of cancer progression and therapeutic response. *Nat. Rev. Cancer* 15, 712–729
54. Duarte, L. F., Young, A. R. J., Wang, Z., Wu, H.-A., Panda, T., Kou, Y., Kapoor, A., Hasson, D., Mills, N. R., Ma'ayan, A., Narita, M., and Bernstein, E. (2014) Histone H3.3 and its proteolytically processed form drive a cellular senescence programme. *Nat. Commun.* 5, 5210
55. Dhaenens, M., Glibert, P., Meert, P., Vossaert, L., and Deforce, D. (2015) Histone proteolysis: A proposal for categorization into “clipping” and “degradation.” *Bioessays* 37, 70–79
56. Anderson, L. C., Karch, K. R., Ugrin, S. A., Coradin, M., English, A. M., Sidoli, S., Shabanowitz, J., Garcia, B. A., and Hunt, D. F. (2016) Analyses of Histone Proteoforms Using Front-end Electron Transfer Dissociation-enabled Orbitrap Instruments. *Mol. Cell. Proteomics* 15, 975–988
57. Rappsilber, J., Mann, M., and Ishihama, Y. (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* 2, 1896–1906
58. Mertins, P., Qiao, J. W., Patel, J., Udeshi, N. D., Clauser, K. R., Mani, D. R., Burgess, M. W., Gillette, M. A., Jaffe, J. D., and Carr, S. A. (2013) Integrated proteomic analysis of post-translational modifications by serial enrichment. *Nat. Methods* 10, 634–637
59. Elias, J. E., and Gygi, S. P. (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* 4, 207–214
60. Mayne, L. (2016) Chapter thirteen - Hydrogen exchange mass spectrometry. *Methods Enzymol.* 566, 335–356
61. Guttman, M., Weis, D. D., Engen, J. R., and Lee, K. K. (2013) Analysis of overlapped and noisy hydrogen/deuterium exchange mass spectra. *J. Am. Soc. Mass Spectrom.* 24, 1906–1912
62. Schechter, I., and Berger, A. (1967) Pi, and Pi' are commonly used nomenclature for amino acid residues of a substrate (or an inhibitor). *Biochem. Biophys. Res. Commun.* 27, 157
63. Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J., and Gevaert, K. (2009) Improved visualization of protein consensus sequences by iceLogo. *Nat. Methods* 6, 786–787
64. Lowther, J., Robinson, M. W., Donnelly, S. M., Xu, W., Stack, C. M., Matthews, J. M., and Dalton, J. P. (2009) The importance of pH in regulating the function of the *Fasciola hepatica* cathepsin L1 cysteine protease. *PLoS Negl. Trop. Dis.* 3, e369
65. Adams-Cioaba, M. A., Krupa, J. C., Xu, C., Mort, J. S., and Min, J. (2011) Structural basis for the recognition and cleavage of histone H3 by cathepsin L. *Nat. Commun.* 2, 197

66. Rand, K. D., Zehl, M., and Jørgensen, T. J. D. (2014) Measuring the hydrogen/deuterium exchange of proteins at high spatial resolution by mass spectrometry: overcoming gas-phase hydrogen/deuterium scrambling. *Acc. Chem. Res.* 47, 3018–3027
67. Kan, Z.-Y., Walters, B. T., Mayne, L., and Englander, S. W. (2013) Protein hydrogen exchange at residue resolution by proteolytic fragmentation mass spectrometry analysis. *Proc. Natl. Acad. Sci. U. S. A.* 110, 16438–16443
68. Rand, K. D., Zehl, M., Jensen, O. N., and Jørgensen, T. J. D. (2009) Protein hydrogen exchange measured at single-residue resolution by electron transfer dissociation mass spectrometry. *Anal. Chem.* 81, 5577–5584
69. Keck, K. M., and Pemberton, L. F. (2012) Histone chaperones link histone nuclear import and chromatin assembly. *Biochim. Biophys. Acta* 1819, 277–289
70. Apta-Smith, M. J., Hernandez- Fernaund, J. R., and Bowman, A. J. (2018) Evidence for the nuclear import of histones H3.1 and H4 as monomers. *EMBO J.*, e98714
71. Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1993) Primary structure effects on peptide group hydrogen exchange. *Proteins* 17, 75–86
72. Ferreira, H., Somers, J., Webster, R., Flaus, A., and Owen-Hughes, T. (2007) Histone Tails and the H3 α N Helix Regulate Nucleosome Mobility and Stability. *Mol. Cell. Biol.* 27, 4037–4048
73. Lins, R. D., and Røthlisberger, U. (2006) Influence of Long-Range Electrostatic Treatments on the Folding of the N-Terminal H4 Histone Tail Peptide. *J. Chem. Theory Comput.* 2, 246–250
74. Wang, X., Moore, S. C., Laszckzak, M., and Ausió, J. (2000) Acetylation Increases the α -Helical Content of the Histone Tails of the Nucleosome. *J. Biol. Chem.* 275, 35013–35020
75. Dorigo, B., Schalch, T., Bystricky, K., and Richmond, T. J. (2003) Chromatin fiber folding: requirement for the histone H4 N-terminal tail. *J. Mol. Biol.* 327, 85–96
76. Englander, S. W., and Kallenbach, N. R. (1983) Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Q. Rev. Biophys.* 16, 521–655
77. Englander, S. W. (2006) Hydrogen exchange and mass spectrometry: A historical perspective. *J. Am. Soc. Mass Spectrom.* 17, 1481–1489
78. Fang, J., Rand, K. D., Beuning, P. J., and Engen, J. R. (2011) False EX1 signatures caused by sample carryover during HX MS analyses. *Int. J. Mass Spectrom.* 302, 19–25
79. Guttman, M., Wales, T. E., Whittington, D., Engen, J. R., Brown, J. M., and Lee, K. K. (2016) Tuning a High Transmission Ion Guide to Prevent Gas-Phase Proton Exchange During H/D Exchange MS Analysis. *J. Am. Soc. Mass Spectrom.* 27, 662–668
80. Bowman, A., Lercher, L., Singh, H. R., Zinne, D., Timinszky, G., Carlomagno, T., and Ladurner, A. G. (2016) The histone chaperone sNASP binds a conserved peptide motif within the globular core of histone H3 through its TPR repeats. *Nucleic Acids Res.* 44, 3105–3117
81. Hudgens, J. W., Gallagher, E. S., Karageorgos, I., Anderson, K. W., Filliben, J. J., Huang, R. Y.-C., Chen, G., Bou-Assaf, G. M., Espada, A., Chalmers, M. J., Harguindey, E., Zhang, H.-M., Walters, B. T., Zhang, J., Venable, J., Steckler, C., Park, I., Brock, A., Lu, X., Pandey, R., Chandramohan, A., Anand, G. S., Nirudodhi, S. N., Sperry, J. B., Rouse, J. C., Carroll, J. A., Rand, K. D., Leurs, U., Weis, D. D., Al-Naqshabandi, M. A., Hageman, T. S., Deredge, D., Wintrode, P. L., Papanastasiou, M., Lambris, J. D., Li, S., and Urata, S. (2019) Interlaboratory Comparison of Hydrogen-Deuterium Exchange Mass Spectrometry Measurements of the Fab Fragment of NISTmAb. *Anal. Chem.* 91, 7336–7345

82. Pagar, C. T., and Dutch, R. E. (2005) Cathepsin L is involved in proteolytic processing of the Hendra virus fusion protein. *J. Virol.* 79, 12714–12720
83. Wong, J. J. W., Young, T. A., Zhang, J., Liu, S., Leser, G. P., Komives, E. A., Lamb, R. A., Zhou, Z. H., Salafsky, J., and Jardetzky, T. S. (2017) Monomeric ephrinB2 binding induces allosteric changes in Nipah virus G that precede its full activation. *Nat. Commun.* 8, 781

Table 1: Summary results of peptide identification under HX-MS conditions of histones following digestion with Cathepsin-L, online pepsin and a combination of both.

		Unique Peptides (#)				Sequence Coverage (%)				Average Length				Redundancy			
		H2A3	H2B2E	H3.1	H4	H2A3	H2B2E	H3.1	H4	H2A3	H2B2E	H3.1	H4	H2A3	H2B2E	H3.1	H4
N-terminal (50 AAs)	Cathepsin-L	12	5	4	6	100	100	100	100	31.3	28.4	34.8	28.2	7.5	3.1	2.8	3.4
	Pepsin	9	7	2	5	100	100	100	100	33.3	34.1	47.5	39.2	6.0	5.7	2.0	3.9
	Cathepsin-L/Pepsin	14	12	5	12	100	100	100	100	30.1	31.8	37.4	29.4	9.6	8.3	3.8	7.1
Total Protein Length	Cathepsin-L	41	43	15	40	100	85.6	85.2	100	27.7	24.1	22	27.2	8.8	8.6	2.5	10.6
	Pepsin	57	91	54	50	100	100	98.5	100	23.9	24.1	18.3	23	10.6	17.9	7.3	11.3
	Cathepsin-L/Pepsin	80	112	60	88	100	100	98.5	100	24.7	24	18.6	23.1	15.3	21.6	8.3	19.9

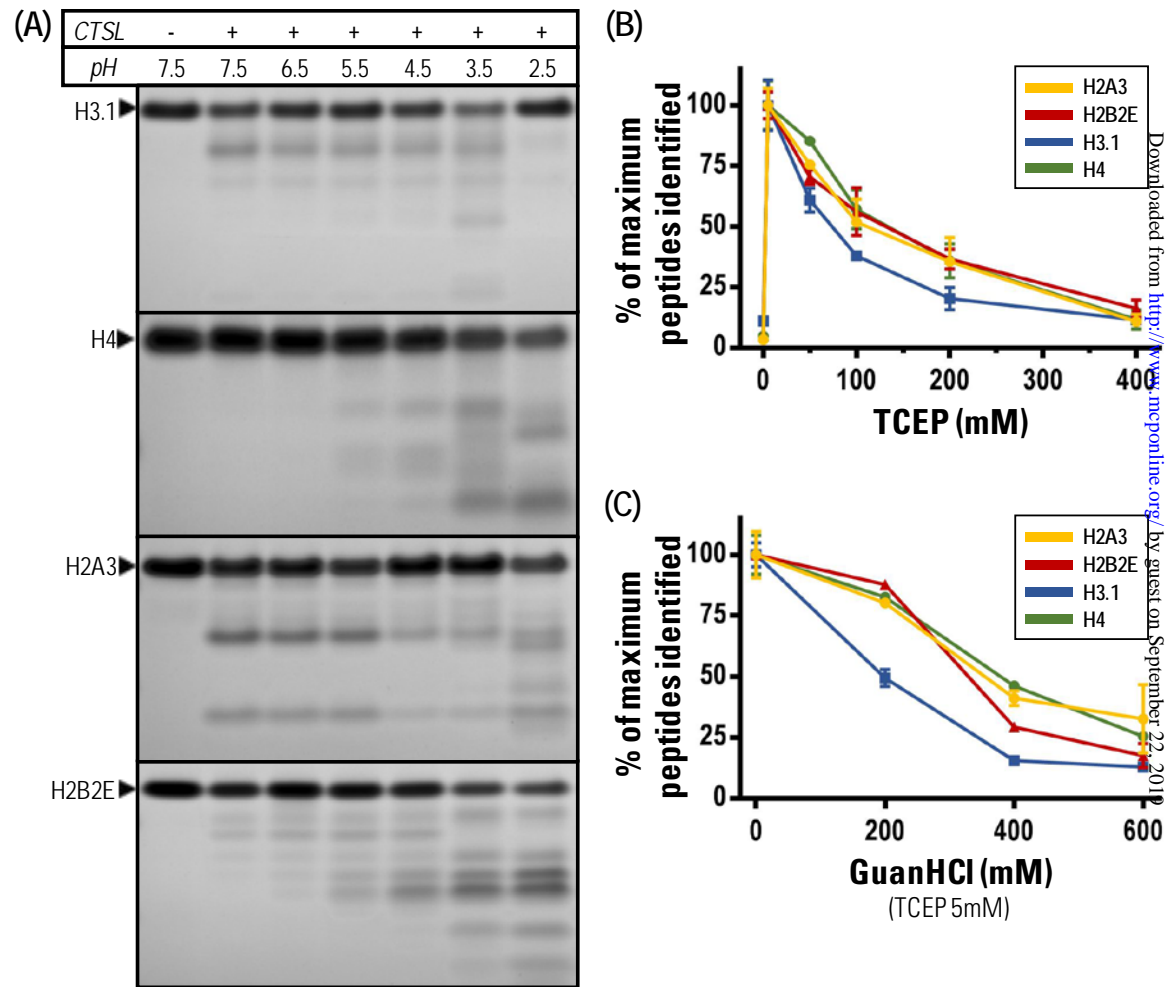
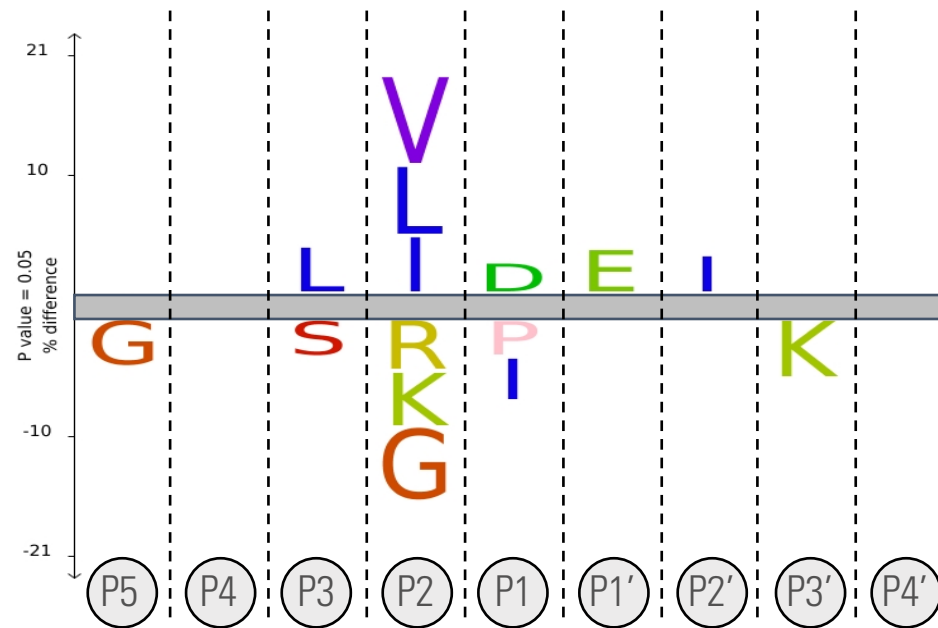


Figure 1. Cathepsin-L activity profiling.

(A) Effect of pH on cathepsin-L activity. The enzyme was incubated with recombinant, purified histones at different pH, either on ice (H3.1) or at RT (H4, H2A3 and H2B2E) (2.5 min, 1:2 (w/w) enzyme:substrate). Histones alone were used for reference (first lane). The arrows to the left indicate the bands corresponding to intact histones.

(B, C) Effect of (B) TCEP and (C) guanidine HCl on peptide identification (%). Following digestion (10 min, 0 °C, 1:10 (w/w) enzyme:substrate), samples were desalted and analyzed by an easyLC-Q Exactive Plus Orbitrap MS (Thermo Scientific, Bremen). Error bars represent the standard deviation of the mean of triplicate measurements for TCEP and duplicate measurements for guanidine HCl. Peptide (%) values were obtained by normalizing to the highest number of peptides identified within each histone (Supplemental Table 1).



Downloaded from <http://www.mcponline.org/> by guest on September 22, 2019

Figure 2. Specificity of cathepsin-L derived from histone digestion at pH 2.5.

To build the IceLogo, we employed peptides identified from all four histones (TCEP 5 mM, 37 oC, 10 min) and discarded redundant cleavage sites. Peptide sequences were aligned and loaded to the IceLogo web application (62). Peptides originating from histone sequences were used as a reference set. Subsite nomenclature is according to Schechter & Berger (61). Over-represented amino acids around the cleavage site (P1) in the experimental set are indicated with positive values whereas underrepresented amino acids with negative values. Residues are colored based on their hydrophobicity properties.

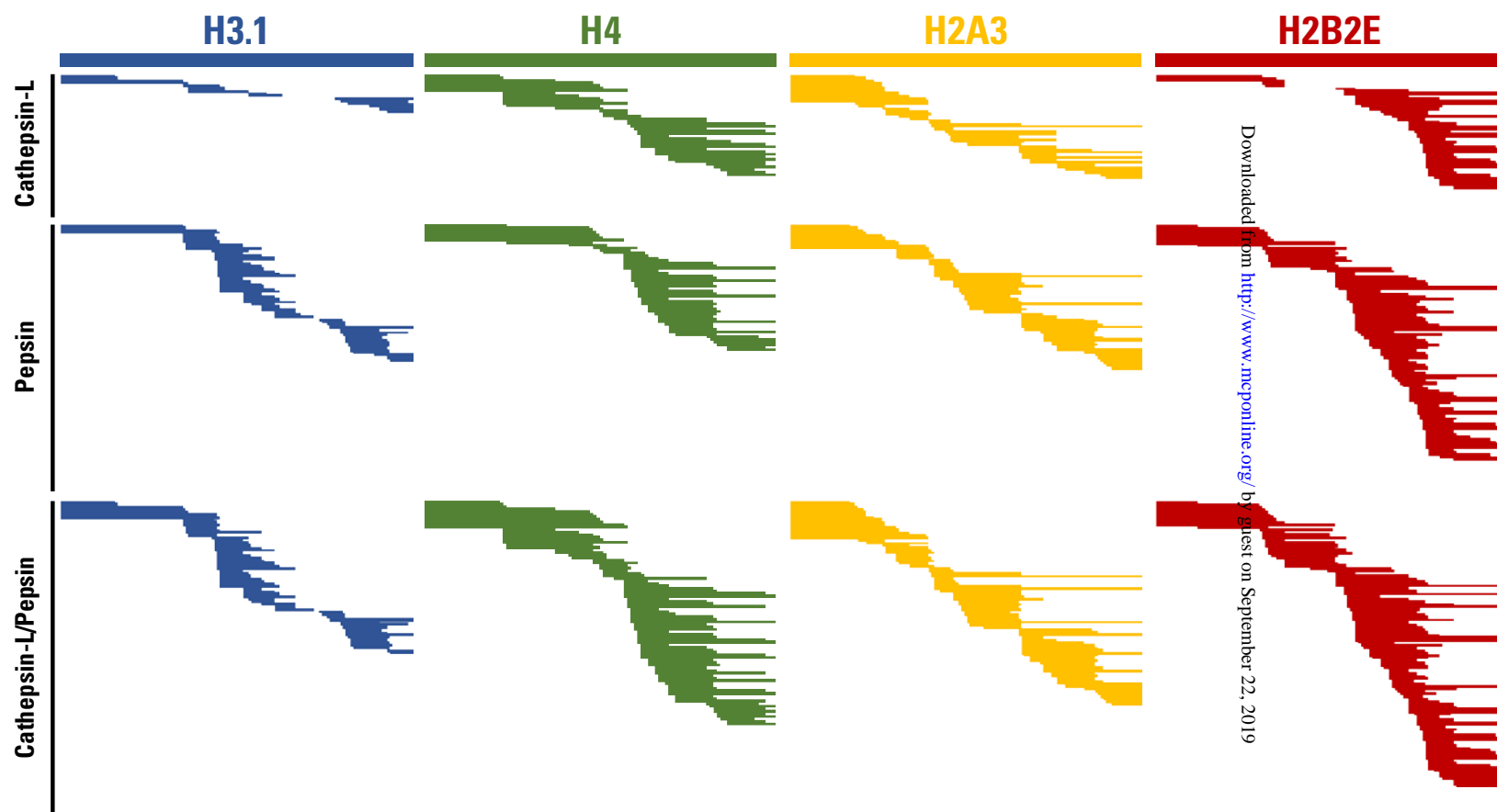


Figure 3. Peptide maps following digestion with cathepsin-L, online pepsin and a combination of both.

Each line represents a unique peptide identified following digestion with cathepsin-L (5 min, enzyme: protein ratio 1:2, 3 °C), online pepsin (1 min, 300 uL/min) and a combination of both using above parameters. All digestions occurred in a slightly reducing environment (TCEP 5 mM). Peptides were desalted online and analyzed by a nanoAcquity-Synapt XEVO G2-XS MS (Waters). Upon identification, peptides were further processed and validated in DynamX 3.0 (Waters) (Supplemental Table 2).

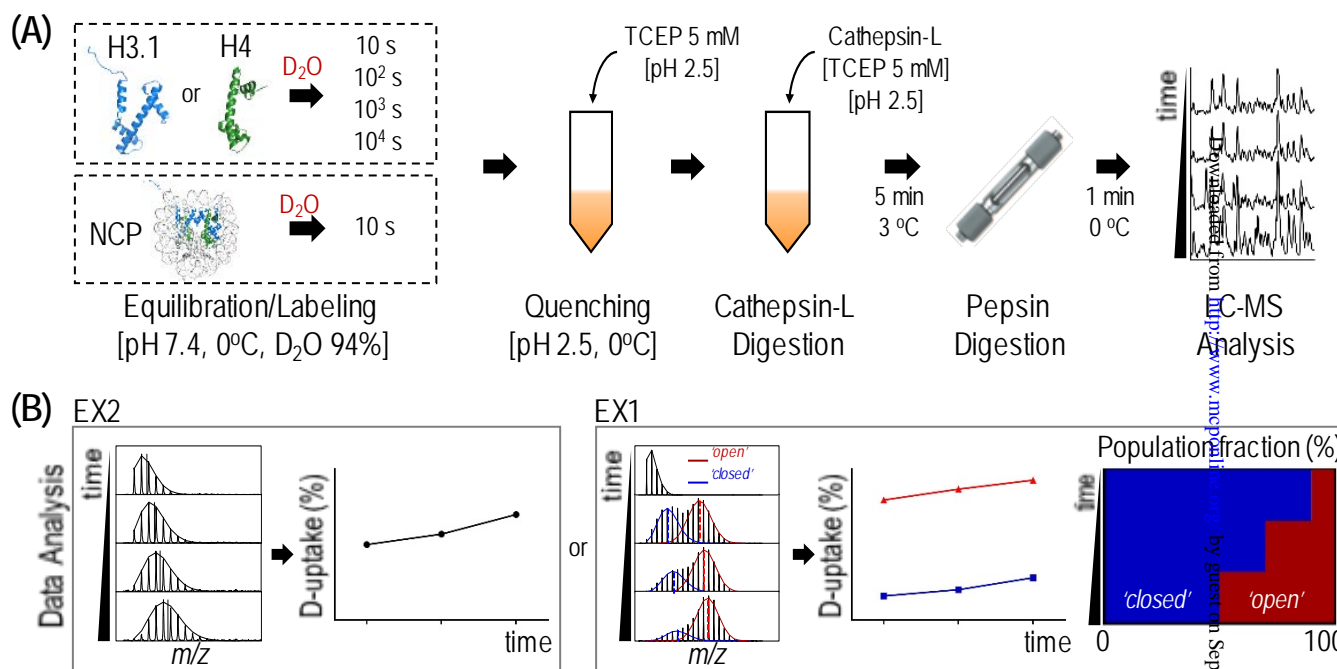
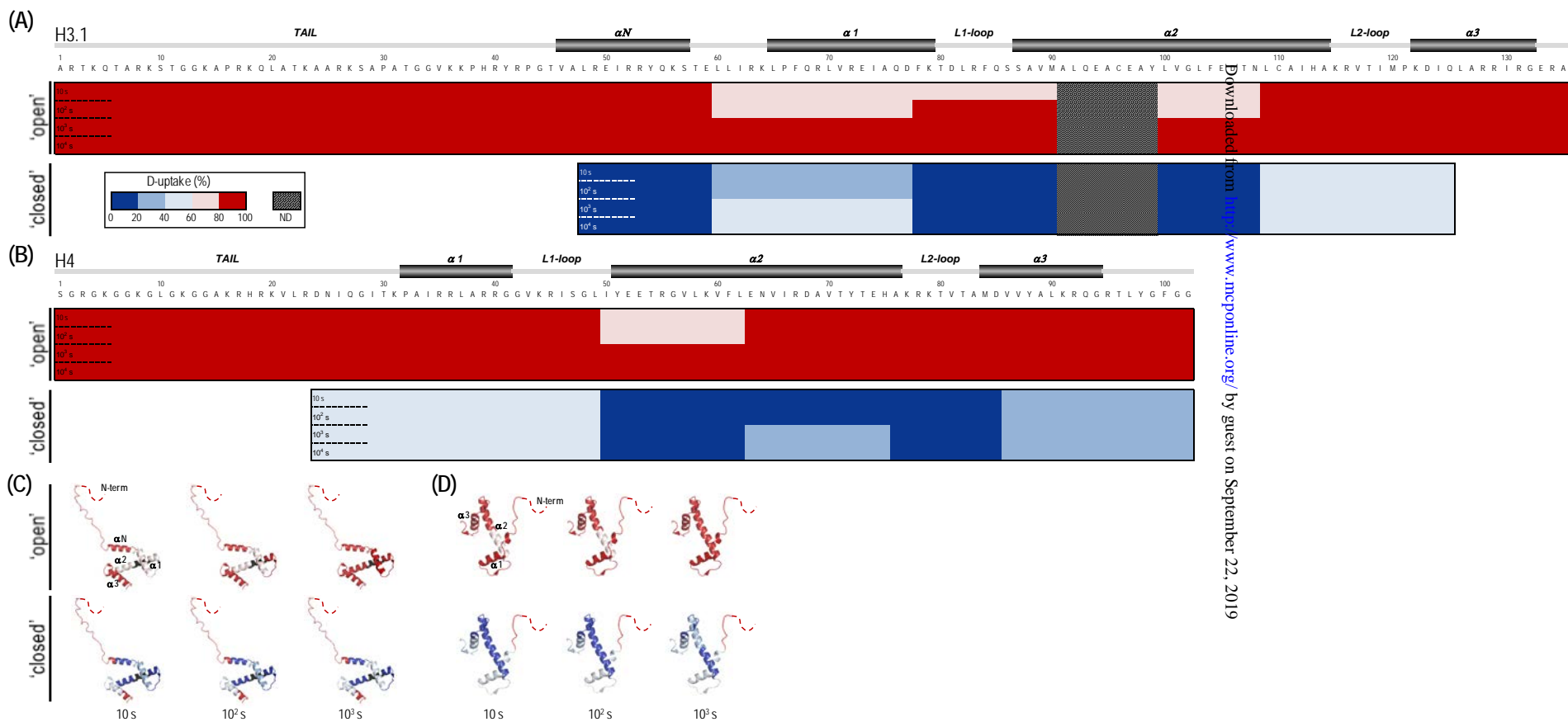


Figure 4. HX-MS workflow.

(A) Experimental scheme for deuterium exchange of histone monomers (H3.1, H4) and mononucleosomes (NCP). Samples were prepared and analyzed in duplicates.

(B) Cartoon representation of D-uptake data analysis. Isotopic distributions (EX2) were processed in DynamX 3.0 (Waters) and the D-uptake (centroid depicted with a black line) of peptides over time was calculated. Bimodal isotopic distributions (EX1) were processed in HX-Express2 (60) and the D-uptake was calculated (red and blue dashed lines correspond to a centroid value calculated for each distribution) along with the fraction of each population in solution (the size of each bar corresponds to the 'area under the curve' for each binomial distribution observed). D-uptake values from EX2 and EX1 were normalized to full deuteration controls as described in Mayne (59).



Downloaded from www.mcponline.org/ by guest on September 22, 2019

Figure 5. HX-MS analysis of the H3 and H4 monomers.

(A-B) Heat maps of the H-to-D exchange for H3.1 (A) and H4 (B) monomers. The annotation of histone structural elements is shown at the top (Luger *et al*, 1997). Heat map color-coding is based on the deuterium uptake (%) calculated upon normalizing D-uptake values to respective values obtained using fully deuterated control samples as described in Mayne, 2016. Regions with bimodal isotopic envelopes detected for the monomers are depicted with two heat maps, one representing the 'open' conformation and one the 'closed' conformation. Individual time points of exchange are 10, 10², 10³ and 10⁴ s. Individual D-uptake values and profiles at the peptide level are given in Supplemental Table 3 and Supplemental Figure 5.

(C-D) Average deuterium incorporation for monomeric H3.1 (C) and H4 (D) for the 'open' and 'closed' conformations. Values are mapped onto mononucleosomal structures of H3.1 and H4 (PDB: 1AOI). Color coding follows that of the heat maps and dashed lines at the N-termini of the proteins indicate regions not observed in the crystal structure.