Top-down Characterization of Core Histones using an Online 2D Nanocapillary LC System

Zhixin Tian, Rui Zhao, Nikola Tolić, Ronald J. Moore, David L. Stenoien, Errol W. Robinson, Richard D. Smith, Ljiljana Paša-Tolić
Pacific Northwest National Laboratory, Richland, WA

Pacific Northwest NATIONAL LABORATORY

Overview

- **A metal-free online reversed phase liquid chromatography (RPLC) – weak cation exchange-hydrophilic interaction liquid chromatography (WCX-HILIC) system was developed and coupled with Fourier transform mass spectrometry (FTMS) for comprehensive high-throughput high-sensitivity characterization of histones at the intact protein level.**
- **In a proof-of-principle experiment using HeLa core histones, ~500 histones were identified unambiguously in a single two-dimensional (2D) LC-MS/MS analysis (4 h gradient). Increasing the gradient to 10 h resulted in identification of 548 isoforms for H4 alone.**
- **26% identified histones were phosphorylated on average, and up to three phosphorylaton sites per isoform were confidently identified.**

Introduction

- Core histones are key chromatin proteins serving as spools to package and organize DNA into structural and manageable chromosomes[1]. Core histones are heavily modified through lysine acetylation, lysine or arginine methylation, serine or threonine phosphorylation[2].
- These covalent post-translational modifications (PTMs), or "histone code", play important epigenetic roles in all DNA-related cellular processes including transcription[3], replication[4], repair[5], and alternative splicing[6].
- Because all PTMs along the entire protein sequence function in concert[7,8], top-down MS is superior for elucidating the histone code compared to traditional antibody-based or even more recent bottom-up[9] and middle-down[10] MS approaches. However, because of the high complexity and large dynamic range of histone mixtures, high-resolution separation prior to MS is required to minimize ion suppression and under-sampling challenges.
- Hence, we developed a 2D LC-FTMS platform and used it to confidently identify >500 HeLa core histones in a single 2D LC-MS/MS analysis.
- False discovery rate (FDR) was evaluated by searching a Arabidopsis thaliana (ARATH) database.

CONTACT: Zhixin (Michael) Tian, Ph.D.
Biological Sciences Division, K8-98
Pacific Northwest National Laboratory
P.O. Box 999, Richland, WA 99352
E-mail: zhixin.tian@pnl.gov

Methods

- **Material:** HeLa core histones were purchased from Active Motif.
- **2D LC-MS/MS:** Histones were separated into individual families (H4, H2B, H2A, H3) in the first dimension RPLC using a column (600 mm × 200 µm i.d.) packed in house with Phenomenex (Torrance, CA) Jupiter C5 (5 µm particles; 300 Å pore size). Fractions were collected online (based on the UV monitoring). Each RPLC fraction was further separated in the second dimension WCX-HILIC using a column (50 cm×100 µm i.d.) packed in house with PolyLC (Columbia, MD) PolyCAT A (5 µm particles; 1000 Å pore size). The second dimension LC was coupled online with LTQ-Orbitrap Velos operated in high-resolution MS and MS/MS mode.
- **Protein identification:** Raw datasets were searched against annotated human TD database using ProSightPC 2.0 with methylation (mono-, di-, tri-), acetylation, and phosphorylation as dynamic modifications.

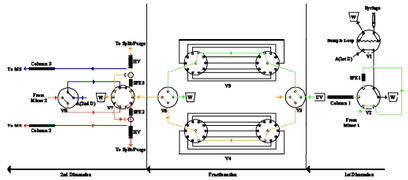

Figure 1. Schematic diagram of RPLC/WCX-HILIC 2D LC system. The green line indicates separation of the core histones in the first dimension (RPLC) and fractionation; the yellow line marks loading of collected fraction into the SPE column prior to the second dimension separation (WCX-HILIC); the red line denotes separation of one fraction in the second dimension; while the blue line designates equilibration of one of the analytical columns.

Results

Table 1. Total number of histones (phosphorylated histones) identified in various LC-MS/MS analyses.

	H4	H2B	H2A	H3	Total
RPLC only	63 (15)	37 (2)	46 (19)	23 (2)	169 (38)
WCX-HILIC only	40 (6)	131 (27)	38 (10)	22 (3)	231 (46)
2D LC	151 (16)	212 (56)	132 (57)	*	495 (129)

* in progress

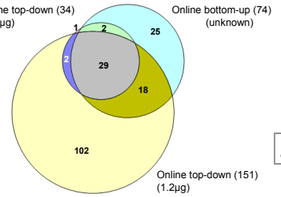

Figure 2. Comparison of H4 isoform identifications by three different methods: offline top-down[11], online bottom-up[12], and online top-down (this study).

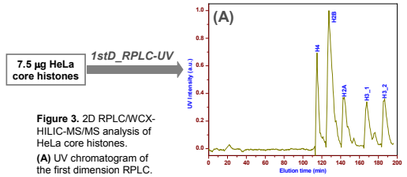

Figure 3. 2D RPLC/WCX-HILIC-MS/MS analysis of HeLa core histones.
(A) UV chromatogram of the first dimension RPLC.

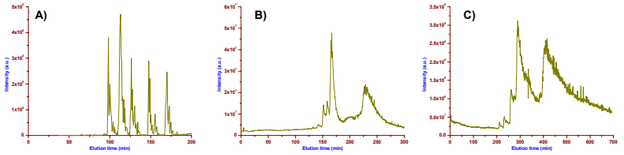

Figure 5. MS-only total-ion-chromatograms. **(A)** RPLC-MS/MS of core histones; **(B)** WCX-HILIC-MS/MS of core histones; **(C)** WCX-HILIC-MS/MS of RPLC fraction H4 with 10-h gradient.

2ndD_WCX/HILIC-MS/MS

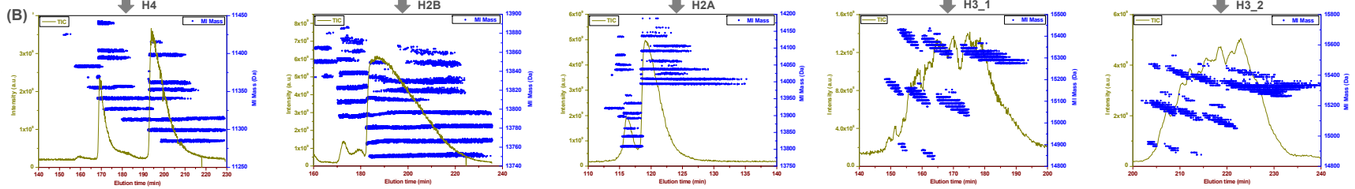

Figure 3. (B) MS-only total-ion-chromatograms with overlaid 2D displays (monoisotopic mass vs. elution time) generated form the WCX-HILIC-MS/MS of the RPLC fractions.

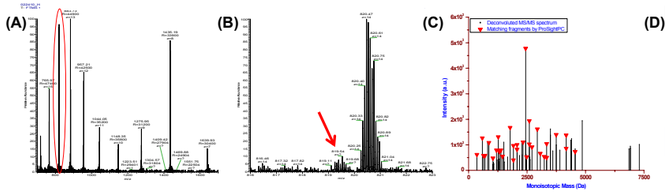

Figure 4. Identification of H4 isoform S1acK5acK8acK12acK16acK20me2.
(A), (B) parent ion spectrum;
(C) neutral mass ETD spectrum with matching fragments marked;
(D) assigned fragments with modified amino acid colored.

(D) >H4_HUMAN, P62805, Histone H4.. (Type: *predicted*, Signal Peptide: *false*, Propep: *false*)

c1 S-G-R-G-K-G-G-K-G-L-G-K-G-G-A-K-R-H-R-K-V-L-R-D-N-I-Q-G-I-T-I z73
c31 K-P-A-I-R-R-L-A-R-R-G-G-V-K-R-I-S-G-L-I-Y-E-E-T-R-G-V-L-K-V z43
c61 F-L-E-N-V-I-R-D-A-V-T-Y-T-E-H-A-K-R-K-T-V-T-A-M-D-V-V-Y-A-L z13
c91 K-R-Q-G-R-T-L-Y-G-F-G-G z1

ID/Gene	Length	Mass	Mass Diff	PPM Diff	C Ions	Z Ions	Total Ions	PDE Score	P Score	E-Value
27717772 227742	102	11467.4	-.1099	-9.581	24	2	26	86.7	2.33E-47	1.76E-40

Methylation (mono) Acetylation Phosphorylation Dimethylation Trimethylation

Table 2. Partial list of identified H4 isoforms from 2D LC-MS/MS analysis.

Family	Accession Number	Retention Time (min)	Theo. Mass (Da)	Obs. Mass (Da)	Histone Codes	P_Score	Norm. Int. (%)
H4	P62805	136.54	11383.4	11383.4	S1acK8acK12acK20me2	3.9E-50	0.715
H4	P62805	137.05	11398.4	11399.3	S1acK5acK8acK20me3	8.8E-14	0.043
H4	P62805	137.05	11380.4	11380.4	S1pR3me3K20me3	1.2E-13	0.101
H4	P62805	137.55	11363.4	11364.3	S1pR3me2	1.2E-05	2.445
H4	P62805	137.55	11383.4	11383.4	R3me2K8acK12acK16ac	7.3E-21	0.137
H4	P62805	137.55	11374.4	11375.2	R4me	3.4E-12	0.137
H4	P62805	137.55	11297.4	11296.4	K5ac	7.3E-09	3.439
H4	P62805	137.55	11311.4	11311.5	S1acK20me	1.3E-06	3.439
H4	P62805	137.55	11339.4	11339.4	S1acK16ac	2.4E-16	3.439
H4	P62805	137.55	11383.4	11383.3	S1acK5acK16acK20me2	5.1E-21	3.439
H4	P62805	137.55	11323.3	11324.2	S1pR3me	2.0E-12	3.439
H4	P62805	138.05	11339.4	11339.4	K5acK16ac	3.0E-13	0.487
H4	P62805	138.53	11386.4	11386.5	none	1.4E-04	0.223
H4	P62805	139.02	11341.4	11340.4	S1acK16acK20me2	8.3E-54	4.187
H4	P62805	143.02	11325.4	11326.3	S1acK20me2	6.1E-09	3.439
H4	P62805	143.52	11313.4	11312.5	S1acK12ac	5.8E-29	1.116
H4	P62805	143.52	11341.4	11340.5	S1acK12acK20me2	6.5E-61	1.436
H4	P62805	143.52	11356.4	11357.3	S1acK12acK20me3	1.5E-30	1.436
H4	P62805	143.52	11325.4	11325.4	R3me2K16ac	2.2E-06	0.487
H4	P62805	143.52	11327.4	11327.5	S1acK12acK20me	1.5E-13	0.699
H4	P62805	143.52	11327.4	11327.5	S1acR3meK16ac	2.6E-24	0.699
H4	P62805	144.53	11327.4	11326.5	K5acK16acK20me	6.1E-22	0.320
H4	P62805	144.53	11327.4	11326.5	S1acK5acK20me	2.3E-15	0.320
H4	P62805	145.03	11325.4	11326.4	K5acK20me2	9.0E-21	0.172

Table 3. Proteins identified using Arabidopsis thaliana (ARATH) database confirms low FDR (i.e., p<1E-2 was required for identification).

Core Histone	Accession Number	Family	Theo. Mass (Da)	Obs. Mass (Da)	PTMs	Matching Fragments	P_Score
H4	P59259	H4	11439.4	11439.3	S1acK8acK12acK16ac	14	7.4E-23
	P59259	H4	11397.4	11397.4	S1acK8acK12ac	18	3.1E-33
	P59259	H4	11397.4	11396.4	S1acK5acK12ac	18	4.5E-30
	P59259	H4	11439.4	11439.4	S1acK5acK12acK16ac	24	2.4E-45
	P59259	H4	11397.4	11397.4	S1acK5acK8ac	8	9.2E-07
	P59259	H4	11355.4	11355.4	S1acK5ac	13	1.3E-12
	P59259	H4	11397.4	11398.3	S1acK8acK16ac	19	2.3E-26
	P59259	H4	11355.4	11354.4	S1acK8ac	22	8.0E-42
	P59259	H4	11355.4	11354.4	K12acK16ac	6	1.2E-04
	P59259	H4	11439.4	11440.3	S1acK12acK16acK20ac	15	8.7E-28
	P59259	H4	11355.4	11356.2	S1acK12ac	16	8.1E-21
	P59259	H4	11313.4	11314.3	K5ac	16	1.3E-11
	P59259	H4	11313.4	11313.6	K20ac	13	1.6E-12
	P59259	H4	11397.4	11398.2	S1acK16acK20ac	16	1.4E-22
	Q6NR90	Q6NR90	11271.4	11270.4	none	16	2.7E-17
	Q6NR90	Q6NR90	11313.4	11314.4	S1ac	18	3.6E-40
	P59259	H4	11313.4	11313.4	K12ac	7	7.6E-09
H2B	Q8GYF3 Q4VDZ2	Q8GYF3 Q4VDZ2	13998.6 13220.1	13997.7 13219.7	none M1ac	0 3	1.8E-03 3.7E-04
H2A	Q9LPK9	Q9LPK9	13119.3	13118.9	none	0	2.6E-04

```
SeqA Name                    Len(aa)  SeqB Name                    Len(aa)  Score
==========================================================================================
1    sp|P62805|H4_HUMAN     103      2    sp|P59259|H4_ARATH     103      98
==========================================================================================
CLUSTAL 2.0.12 multiple sequence alignment

sp|P62805|H4_HUMAN      MSGRGKGGKGLGKGGAKRHRKVLRDNIQGITKPAIRRLARRGGVKRISGLIYEETRGVLK 60
sp|P59259|H4_ARATH      MSGRGKGGKGLGKGGAKRHRKVLRDNIQGITKPAIRRLARRGGVKRISGLIYEETRGVLK 60
                        ************************************************************

sp|P62805|H4_HUMAN      VFLENVIRDAVTYTEHAKRKTVTAMDVVYALKRQGRTLYGFGG 103
sp|P59259|H4_ARATH      IFLENVIRDAVTYTEHARRKTVTAMDVVYALKRQGRTLYGFGG 103
                        :***************:***********************
```

Figure 6. Sequence similarity between Human and ARATH H4.

>H4_HUMAN, P62805, Histone H4.. (Type: *predicted*, Signal Peptide: *false*, Propep: *false*)

c1 S-G-R-G-K-G-G-K-G-L-G-K-G-G-A-K-R-H-R-K-V-L-R-D-N-I-Q-G-I-T- z73
c31 K-P-A-I-R-R-L-A-R-R-G-G-V-K-R-I-S-G-L-I-Y-E-E-T-R-G-V-L-K-V z43
c61 F-L-E-N-V-I-R-D-A-V-T-Y-T-E-H-A-K-R-K-T-V-T-A-M-D-V-V-Y-A-L- z13
c91 K-R-Q-G-R-T-L-Y-G-F-G-G- z1

ID/Gene	Length	Mass	Mass Diff	PPM Diff	C Ions	Z Ions	Total Ions	PDE Score	P Score	E-Value
27717244 71192	102	11397.4	-.027	-2.3714	14	10	24	102	6.12E-47	4.62E-40

>H4_ARATH, P59259, Histone H4.. (Type: *predicted*, Signal Peptide: *false*, Propep: *false*)

c1 S-G-R-G-K-G-G-K-G-L-G-K-G-G-A-K-R-H-R-K-V-L-R-D-N-I-Q-G-I-T- z73
c31 K-P-A-I-R-R-L-A-R-R-G-G-V-K-R-I-S-G-L-I-Y-E-E-T-R-G-V-L-K-I z43
c61 F-L-E-N-V-I-R-D-A-V-T-Y-T-E-H-A-R-R-K-T-V-T-A-M-D-V-V-Y-A-L- z13
c91 K-R-Q-G-R-T-L-Y-G-F-G-G- z1

ID/Gene	Length	Mass	Mass Diff	PPM Diff	C Ions	Z Ions	Total Ions	PDE Score	P Score	E-Value
3351858 2957	102	11397.4	-.0382	-3.3541	11	7	18	67.3	3.13E-33	1.54E-27

Figure 7. Fragment ion assignments for two H4 isoforms identified in HUMAN and ARATH database searches. Mass difference between amino acids V+K and I+R equals 42, which is compensated by one extra acetylation.

Conclusions

- **A metal-free online 2D RPLC/WCX-HILIC system was developed for comprehensive high-throughput characterization of histones at the intact protein level.**
- **More than 500 histones were unambiguously identified (<1% FDR) in a single 24-h 2D LC-MS/MS analysis using 7.5 µg HeLa core histone mixture.**
- **The metal free platform enables sensitive analysis of phosphoproteins (e.g., 26% of identified histones on average were phosphorylated; 1- 3 phosphorylation sites were located per isoform).**
- **548 H4 histone isoforms were identified in a 10-h WCX-HILIC MS/MS analysis of H4 RPLC fraction, suggesting that >2000 histone isoforms could potentially be identified in a single 2D LC-MS/MS analysis.**

References

1. RD Kornberg, *Science* 184:868 (1974)
2. T Kouzarides, *Cell* 128:693 (2007)
3. A P Wolffe, *Science* 272:371 (1996).
4. KB Falbo and XT Shen, *Molecules and Cells* 28:149 (2009)
5. H van Attikum and SM Gasser, *Nature Reviews Molecular Cell Biology* 6:757 (2005)
6. RF Luco, et al., *Science* 327:996 (2010)
7. T Jenuwein and CD Allis, *Science* 293:1074 (2001)
8. P Tropberger and R Schneider, *Epigenetics* 5 (2), 112 (2010)
9. MB Trelle,et al. , *Journal of Proteome Research* 8:3439 (2009)
10. NL Young, et al., *Molecular & Cellular Proteomics* 8:2266 (2009)
11. JJ Pesavento, et al., *Journal of Biological Chemistry*, 283:14927 (2008)
12. D Phanstiel, et al., *Proceedings of the National Academy of Science, USA* 105, 4093 (2008)

Acknowledgements

We would like to thank Prof. Neil Kelleher for providing ProSightPC program and support. Portions of this work were supported by the U.S. Department of Energy (DOE) Office of Biological and Environmental Research, the NIH National Center for Research Resources (RR18522), the Environmental Molecular Science Laboratory's (EMSL) intramural program and the Pacific Northwest National Laboratory's (PNNL) Laboratory Directed Research and Development Program. Work was performed in the EMSL, a DOE national scientific user facility located on the campus of PNNL in Richland, Washington. PNNL is a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RLO1830.