Published in final edited form as:

Proteomics. 2010 October ; 10(20): 3610–3620. doi:10.1002/pmic.201000367.

Two-dimensional liquid chromatography system for online top-down mass spectrometry

Zhixin Tian, Rui Zhao, Nikola Tolić, Ronald J. Moore, David L. Stenoien, Errol W. Robinson, Richard D. Smith, and Ljiljana Paša-Tolić

Pacific Northwest National Laboratory, Richland, WA, USA

Abstract

An online metal-free weak cation exchange-hydrophilic interaction LC/RPLC system has been developed for sensitive, high-throughput top-down MS. Here, we report results for analyzing PTMs of core histones, with a focus on histone H4, using this system. With just ~24 µg on-column of core histones (H4, H2B, H2A, and H3) purified from human fibroblasts, 41 H4 isoforms were identified, with the type and location of PTMs unambiguously mapped for 20 of these variants. Compared to corresponding offline studies reported previously, the online weak cation exchange-hydrophilic interaction LC/RPLC platform offers significant improvement in sensitivity, with several orders of magnitude reduction in sample requirements and a reduction in the overall analysis time. To the best of our knowledge, this study represents the first online 2-D LC-MS/MS characterization of core histone mixture at the intact protein level.

Keywords

2-D chromatography; ESI-MS/MS; Histone; PTM; Technology; Top-down

1 Introduction

LC-MS/MS has become an effective and frequently used analytical platform for proteome characterization over the past decade [1,2]. High-resolution separations prior to MS minimize ion suppression and under-sampling challenges associated with the analysis of highly complex proteomes, which often span several orders of dynamic range in protein abundance. For this reason, developments and applications of multidimensional (MD) LC (MDLC) have accompanied and facilitated advances in proteomics, especially top-down (TD) proteomics where proteins are separated and fragmented directly (*i.e.* without enzymatic digestion as in the bottom-up approach). One advantage of TD proteomics is the characterization of multiple (*i.e.* combinatorial) PTMs dispersed along the primary protein sequence. Hence, differentiating between multiple protein isoforms, intractable with bottom-up proteomic approaches, becomes possible.

Historically, several types of MDLC systems, including heart-cutting MDLC [3], directly coupled-column MDLC [4], and column-switching MDLC [5] have been used. Heart-cutting MDLC has only seen a limited use in proteomics, since typically only a single desired range of the eluates from the first dimension is chosen (according to foreknowledge of the sample's composition) to be transferred to the second dimension for further separation

Correspondence: Dr. Ljiljana Paša-Tolić, EMSL, MSIN: K8-98, Pacific Northwest National Laboratory, P.O. Box 999, Richland, WA 99352, USA, ljiljana.pasatolic@pnl.gov, Fax: 11-509-371-6546.

The authors have declared no conflict of interest.

and characterization. Both directly coupled-column MDLC and column-switching MDLC are comprehensive MD separations with high peak capacity, where every component of a sample is subjected to separation in both dimensions. Multidimensional protein identification technology (MudPIT) has become the most prominent directly coupled-column MDLC for proteomic research [6].

In MudPIT, a sample is fractionated in the first dimension *via* strong cation exchange chromatography and further separated in the second dimension using RPLC. This configuration is relatively simple, but the first dimension can only be run in a stepwise fashion, which limits resolution. In addition, possible combinations of separation modes are restricted by the limited combinations of mobile phases that could be used for effective orthogonal separation. It is possible to utilize MudPIT-like strategies with a continuous gradient through some additional considerations in methodology [7].

In column-switching MDLC, fractions from the first dimension are online transferred, or stored, in a series of loops, for the second dimension where the separation is accomplished using two or more columns. This setup allows greater flexibility in terms of combinations of separation modes, and it will be the basis of the work presented here. With this setup, Unger *et al.* has resolved about 1000 peaks in 96 min in the analysis of proteins and peptides with a molecular weight below 20 kDa [8].

Characterization of histone PTMs and their dynamics has been the most notable application of TD MS to date, primarily due to their relatively low mass and high abundance. There is great interest in characterizing the many histone isoforms as the specific combination of PTMs form the so-called histone code, which regulates many aspects of chromatin function ranging from transcriptional regulation to chromosome separation during mitosis. Due to the high complexity of histone mixtures, there is a need for separation of both core histone families (H4, H2B, H2A, and H3) and protein isoforms within each family. RPLC has been regularly used to resolve the four histone families as well as some H2A and H3 variants, although H2A and H4 often co-elute [9,10].

RPLC, however, cannot separate many of the post-translationally modified isoforms (*e.g.* methylated, acetylated, and phosphorylated) from the unmodified proteins and from each other [11]. (This limited resolution in RPLC mode can actually be advantageous in that all of the variants with different PTMs for each core histone are present in the same fraction allowing for direct comparison of their relative abundance.) Weak cation exchange-hydrophilic interaction LC (WCX-HILIC) has proven to be an excellent complementary orthogonal mode to RPLC and has been successfully used to separate acetylated isoforms of histone H4 [12], methylated isoforms of histone H4 [13], phosphorylated isoforms of histone H1 [14], as well as sequence variants of histone H1 [15].

WCX-HILIC, a mixed-mode chromatography introduced by Alpert in 1990 [16], features the simultaneous presence of dominant hydrophilic interaction (due to the use of a higher percentage of organic solvent in the mobile phase than traditional WCX, the mechanism for separation of differentially methylated isoforms) and electrostatic interaction (due to an ionic stationary phase, the mechanism for separation of differentially acetylated isoforms) [17] between the stationary phase and the analyte. Because WCX-HILIC normally uses a salt (NaClO₄) gradient that is not compatible with MS, fractions have to be collected, recovered, and desalted before they are individually introduced into the mass spectrometer using direct infusion ESI.

Using RPLC/HILIC approach, the Kelleher laboratory identified 42 histone H4 isoforms by offline infusion of 93 HILIC fractions from 150 µg of histone H4 collected from multiple RPLC runs of crude HeLa S3 histone proteins [18]. While this is certainly a landmark work

in the area of TD MS, such fraction collection and infusion steps are extremely time consuming, labor intensive, and require a large amount of starting material due to the sample losses during the offline transfer, recovery, and desalting processes. Insufficient sensitivity and relatively low throughput significantly limit the applicability of this methodology for biological applications.

To overcome these limitations of offline MDLC and enable sensitive characterization of protein isoforms using TD MS, a new online 2-D WCX-HILIC/RPLC LC platform was developed. RPLC was used for the second dimension due to its compatibility with MS, and WCX-HILIC was implemented as the first dimension. Fractions from the first dimension were desalted and transferred to the second dimension online to achieve higher sensitivity and throughput. The 2-D LC platform was coupled with Fourier-transform (FT) MS and tested by analyzing histones purified from human fibroblasts. Using ~24 µg of core histone mixture, 41 H4 isoforms were identified, and the type and locations of PTMs were unambiguously mapped for 20 of these variants. Compared to corresponding offline studies reported previously, online WCX-HILIC/RPLC platform offered higher sensitivity and throughput for characterization of histones.

2 Materials and methods

The experimental workflow begins with a core histone sample (details following) separated in the first dimension using WCX-HILIC. Online fractions from this separation were collected in a series of 20 capillary loops with the start and stop time of each fraction guided by the chromatographic profile generated by an integrated UV detector. Each of the 20 WCX-HILIC fractions was then further separated in the second dimension using RPLC coupled by ESI to an LTQ Orbitrap mass spectrometer (Thermo Electron, San Jose, CA, USA). Mass spectra were acquired in high-resolution MS and high-resolution MS/MS (CID) mode. A total of 20 LC-MS/MS data sets were acquired, one for each of the 20 fractions collected from the first dimension separation. LC-MS/MS data sets were then subjected to database search using ProSightPC 2.0 (Thermo Electron) [19] for modified protein identification. Protein identifications from all fractions were combined into a final list of uniquely identified proteins (see Supporting Information).

The schematic diagram of the metal-free 2-D WCX-HILIC/RPLC system is shown in Fig. 1. All valves from V1 to V8 (except V4 and V5) were Cheminert 5000 psi nano-volume valves with polyether ether ketone (PEEK) stators and rotors, 1/32" PEEK nanovolume fittings, and 100 µm bores (Valco Instruments, Houston, TX, USA). Column selectors, valves V4 and V5, will be discussed in detail in the following section. The entire LC path exposed to analyte, including a silica capillary sample loop, was free of metal to ensure enhanced sensitivity for phosphoproteins by avoiding analyte losses associated with metal surface interactions [20].

2.1 Materials and chemicals

All flexible fused-silica capillary tubing (TSP, 360 µm od and 75, 150, or 200 µm id) was purchased from Polymicro Technologies (Phoenix, AZ, USA). ACN (HPLC grade) and isopropanol alcohol (IPA, HPLC grade) were purchased from Fisher Scientific (Pittsburgh, PA, USA); triethyl-ammonium phosphate buffer solution (1 M in H₂O) was purchased from Fluka (Milwaukee, WI, USA); NaClO₄ (minimum 99%), TFA (≥99%), and glacial acetic acid (HOAc, 99.99%) were purchased from Sigma (Bellefonte, PA, USA).

2.2 Extraction of core histones

Primary neonatal Normal Human Dermal Fibroblasts were purchased from Lonza (Walkersville, MD, USA) and grown in Fibroblast Growth Media-2 (Lonza). Core histones were purified from confluent Normal Human Dermal Fibroblasts cultures using a histone purification kit from Active Motif (Carlsbad, CA, USA). Briefly, cultured cells were centrifuged, washed with serum-free media, and homogenized in the extraction buffer; the cell extracts were then micro-centrifuged separating the crude histones in the supernatant from other proteins. After neutralization, the crude histones were further purified using a resin column to give the final core histone mixtures including H4, H2B, H2A, and H3.

2.3 WCX-HILIC of core histone mixture

The core histone mixture was first separated on a PolyLC (Columbia, MD, USA) PolyCAT A (5- μ m particles; 1000 Å pore size) column (800 mm $\times$ 200 μ m id) packed in house. The separation was carried out under constant pressure (*i.e.* 3000 psi) using two ISCO (Lincoln, NE, USA) Model 100 DM 10 000 psi syringe pumps (with a Series D Pump Controller). Mobile phase A consisted of 15 mM triethylammonium phosphate containing 70% v/v ACN, and mobile phase B consisted of mobile phase A plus 0.68 M NaClO₄. An aqueous solution of the core histone mixture (1.2 μ g/ μ L) was loaded into the silica capillary sample loop (20 μ L, 112 cm $\times$ 150 μ m id) using a 100- μ L Hamilton (Reno, NV, USA) syringe. The sample was then transferred to the first dimension SPE column (SPE1) using Buffer A at 1000 psi with a flow rate of $\sim$ 5 μ L/min for 10 min. After loading, the separation gradient was started. An exponential gradient was generated by adding mobile phase B (3000 psi) to a stirred mixer (2.5 mL of 100% mobile phase A initially), with the split flow rate controlled at $\sim$ 10 μ L/min. The column flow rate was $\sim$ 1 μ L/min at the beginning of the gradient. An online UV detector (SPEC-TRA100, Thermo Separation Products, 214 nm) was used to monitor the elution profile of proteins. Protein fractions were collected based on chromatographic peaks into the storage capillary as described below. Upstream of Mixer 1, three Cheminert 15 000 psi UHPLC nanovolume valves (not shown in Fig. 1), stainless stator, and Valco E3 rotor were used to control buffer refill and selection. All buffers were subjected to online degassing using Phenomenex (Torrance, CA, USA) DEGASSEX DG-4400.

2.4 Online fractionation

Two Cheminert column selector valves were used in parallel to alternate between silica capillary tubes (56 cm $\times$ 150 μ m id) to individually store fractions. Each end of the fraction storage capillary was connected to one port of a stream selection valve. The volume of eluant stored in each fraction storage capillary is fully customizable and is mainly determined by the column flow and fractionation rate. Each selector comes with an electric control panel for simple and fast port selection. All the selection valves have PEEK stators and Valco E rotors. Each pair of column selector valves has ten fraction storage capillaries, and here two such systems were used to store a total of 20 fractions. Fractions were collected alternatively between fractions collection systems. This arrangement permitted concurrent fraction collection and initiation of the second RPLC dimension.

2.5 RPLC-MS/MS

Each stored fraction from the first dimension WCX-HILIC was further separated in the second dimension RPLC using a 600 mm $\times$ 75 μ m id column packed in house with a Phenomenex Jupiter C5 (5 μ m particles; 300 Å pore size). The separation was carried out under constant pressure of 4000 psi using two ISCO Model 100 DM 10 000 psi syringe pumps (with Series D Pump Controller). Mobile phase A consisted of 20% ACN aqueous

solution with 5% IPA, 0.6% HOAc, and 0.01% TFA; mobile phase B consisted of 45% ACN, 45% IPA, and identical percentages of HOAc and TFA as Buffer A.

A Cheminert ten-port Nanovolume injection valve was used to alternatively equilibrate/load and run the analytical gradient between two capillary RPLC columns to increase sample throughput in the second dimension. Once the first dimension fraction were collected, RPLC Buffer A (1000 psi, about 5 $\mu\text{L}/\text{min}$) was turned on (valve not shown) for 10 min to push and load the stored fraction into the SPE column of the second dimension. The SPE columns (SPE2 and SPE3, 5 cm $\times$ 150 μm id) were packed with the same stationary phase as the RPLC column with online sol–gel frits (5 mm) at both ends [21]. As the first dimension fractions have a high percentage of organic solvent (70% ACN) and cannot be loaded into the second dimension hydrophobic SPE column directly, a tee was used to split the buffer A stream into two sub-streams with one stream (about 1 $\mu\text{L}/\text{min}$, 75 μm id) pushing the stored fraction out and the other stream (about 4 $\mu\text{L}/\text{min}$, 150 μm id) diluting the fraction before it enters the SPE column. Once one fraction was loaded, RPLC Buffer B was switched on to initiate the analytical gradient and separate the loaded proteins through the capillary RPLC column. High-resolution ESI-MS/MS acquisition using a LTQ Orbitrap was started simultaneously with the gradient and continued throughout the 180-min RPLC separation. ESI was performed by connecting the end of the RPLC column to a 20 μm id chemically etched capillary emitter [22] with a PEEK union. The high voltage required for ESI was applied through a metal union coupled in the split/ purge line, which was not exposed directly to the analyte, so there is no contact between the sample and the metal union. This design has been demonstrated to improve sensitivity and limit of detection for phosphoprotein analysis.

The temperature of the MS inlet capillary was set at 300°C to reduce TFA adducts often seen for core histones, and especially pronounced for H4. Both the MS and MS/MS acquisitions had a resolving power setting of 60 000. Automatic Gain Control target of 1E6 and 3E5 were used for FTMS full scans and MSⁿ scans, respectively. Single microscans were used for both FTMS full scan and MSⁿ scans. Fragmentation of the top five most abundant protein ions, which were isolated in a ± 1.5 m/z window, was performed by CID with normalized collision energy of 35% and an activation time of 30 ms. Dynamic exclusion was implemented with an exclusion duration of 200 s and an exclusion list size of 500. A charge state greater than 3 for the precursor ion was required to trigger MS/MS data acquisition in order to increase the number of tandem MS scans obtained for intact protein precursor ions.

2.6 Protein identification

Intact protein RPLC-FTMS mass spectra were first processed using in-house developed software (ICR-2LS available for download at <http://ncrr.pnl.gov/software/>) to deisotope and deconvolute both the parent and fragment ion spectra and to derive monoisotopic masses for both the intact proteins and their fragments. Protein isoforms and PTMs were identified by searching each acquired data set against annotated TD human database [23] (117 059 basic sequences, and 7 563 274 protein forms) using ProSightPC 2.0. The Thrash algorithm [24] option was chosen to calculate neutral mass for both precursor and fragment ions. The mass and retention time tolerance were set to 0.50 m/z and 2.0 min, respectively. The minimum S/N , minimum RL, maximum charge, and maximum mass were set to 1.0, 0.9, 40, and 25 kDa, respectively. The minimum number of fragments was set to 6 and minimum intact mass to 5000 Da. After filtering, the protein database search in the “Absolute Mass” mode was carried out for each tandem mass spectrum.

All candidate protein isoforms in the database with monoisotopic masses within ± 1 Da of the precursor observed in the FTMS acquisition were retrieved and all experimental

fragment ions derived from this precursor ion were then matched to *in silico* fragments of each candidate protein isoform individually with 10 ppm mass tolerance. Dynamic PTMs considered methylation (mono-, di-, and tri-), acetylation, and phosphorylation. The best match for each precursor ion (based on P score) was imported into the Repository Report to generate an identification list for each raw data set. All identification lists were then combined and grouped by theoretical mass, accession number, and PTMs in Microsoft Access. The identification with the lowest P score in each group was chosen as the final identification.

3 Results

3.1 Analysis of core histones using RPLC-FTMS

To compare 2-D *versus* 1-D separations, 2.4 µg of the core histone sample was initially analyzed using RPLC only (second dimension in Fig. 1). The total ion current chromatogram and plot of the neutral mass *versus* RPLC elution time are shown in Fig. 2. Clearly, the RPLC separation alone was able to resolve the four core histone families. The acquired 1-D data set was then subjected to database search using ProSightPC as described in the experimental section and the final list of identified H4 isoforms is shown in Table 1. Detailed information regarding the amino acid sequence for each isoform, identified b/y ion cleavage locations, and identified PTMs highlighted in different colors are provided as Supporting Information.

In total, 23 histone H4 isoforms were identified confidently with *P* scores below 1E^{-4} utilizing the search criteria previously specified. Of these 23 isoforms, 13 H4 isoforms were unique (*i.e.* the amino acid sequence and the type, number, and location of all PTM(s) were assigned unambiguously with Number of Best Hits as 1). If there were insufficient fragments to uniquely locate PTM(s), one representative isoform was selected and the total number of all other possible isoforms was reported. Therefore, the number of best hits for the remaining 20 isoforms is greater than 1. It should be noted that the current ProSightPC protein database is built from UniProt, previously Swiss-Prot, databases and there is no specific relationship between accession number and sequence (*i.e.* two sequences with mutation of amino acids may share the same accession number). In this study, all identified H4 isoforms have the same accession number, but some isoforms have a mutation of amino acid 77 from alanine to proline. Amino acid sequences for each identified H4 isoform can also be found in the Supporting Information.

3.2 Analysis of core histones using 2-D WCX-HILIC/ RPLC-FTMS

Twenty-four microgram of core histones were initially separated using WCX-HILIC (first dimension in Fig. 1). Protein elution was monitored with an online UV detector (214 nm) placed right after the column. The UV signal was captured and recorded (Fig. 3) using in-house software embedded in ICR-2LS and 20 fractions were collected for subsequent RPLC separation. Each fraction from the first dimension WCX-HILIC was then separated and analyzed using RPLC coupled online a LTQ Orbitrap for high-resolution MS and high-resolution MS/MS acquisition. A total of 20 LC MS/MS data sets were acquired, one for each of the first dimension fractions. These data sets were then processed and subjected to ProSightPC search to identify (modified) proteins.

Following the second dimension separation and MS analyses, fractions 9, 12, 8, and 17 contained mainly H2A, H2B, H3, and H4 histones, respectively (see Fig. 3). The total ion current chromatograms (MS only) and plots of monoisotopic mass (MS only) *versus* elution time for these fractions are shown in Fig. 4. In these examples, each fraction from the first dimension was further resolved into 2–3 groups of isoforms in the second dimension.

Besides the four fractions mentioned above, fractions 2 (H2B), 5 (H2A, H4), 13 (H2B), 15 (H4), 18 (H4), 19 (H4), and 20 (H4) also contained histone proteins. The remaining fractions consisted mainly of contaminant proteins (basically abundant proteins not completely removed in histone enrichment step). For example, fraction 10 contained mainly 60S ribosomal protein L35, while fraction 15 contained mainly ubiquitin.

The complete list of H4 isoforms identified from all fractions (and sorted by *P* score) is shown in Table 2. The detailed information (including fragment name, *m/z*, monoisotopic and theoretical mass, and mass measurement error) for the matching b and y ions for the top-ranked isoform (aS1-aK16-2 mK20-H4) is given in Table 3 to illustrate high data quality. The amino acid sequence with PTMs and identified b/y ions, are shown in Fig. 5C. The corresponding mass spectra (parent ion and tandem MS) are shown in Fig. 5A and B. The same information, as shown in Table 3 and Fig. 5C, for all the other histones listed in Table 2 is provided in the Supporting Information.

4 Discussion

Histone modifications (acetylation, methylation, phosphorylation, *etc.*) play an important role in transcriptional and epigenetic regulation, DNA repair processes, DNA synthesis, and cell division. Distinct patterns of PTMs result in a histone code that guide chromatin transitions during these cellular processes. Because of their relatively small mass (<25 kDa) and high abundance, histones are ideal targets for TD MS studies and have been extensively analyzed using various MS approaches over the last decade. In this work, 41 H4 isoforms were identified with high confidence (*i.e.* with *P* scores below 1E^{-4}) from 24 μg of core histones (an equivalent of approximately 6 μg of H4) and the type and location of PTMs was unambiguously mapped for 20 of these variants. This represents a significant improvement in sensitivity when compared to the corresponding offline system in terms of the total number of identifications *per* unit of sample. Namely, in a previously published work, 42 histone H4 isoforms were identified by offline infusion of HILIC fractions from 150 μg of histone H4 collected from multiple RPLC runs of crude HeLa S3 histone proteins [18]. Fraction collection and infusion steps are time consuming, labor intensive, and require a large amount of starting material due to sample losses during the offline transfer, recovery, and desalting processes. Insufficient sensitivity and relatively low throughput significantly limit the usefulness of this methodology for biological applications. A new online 2-D WCX-HILIC/RPLC LC platform described herein overcomes the limitations of offline MDLC and enables sensitive characterization of protein isoforms using TD MS.

Compared to 1-D RPLC (Table 1), 2-D WCX-HILIC/RPLC (Table 2) yielded approximately a two-fold increase in the number of identified H4 isoforms. For instance, isoform aS1-aK12-aK16-mK20-H4 was identified in the 2-D LC-FTMS but not in the 1-D LC-FTMS analysis primarily because it was well separated from other much more abundant isoforms (see Fig. 6). With the additional dimension of separation, this isoform appeared as the most abundant peak in the spectrum (Fig. 6D), and was chosen for fragmentation by data-dependent acquisition, which led to confident identification. In the case of 1-D separation, this particular isoform co-eluted with more abundant isoforms and hence was not chosen for fragmentation due to the limited time window for MS/MS (*i.e.* under-sampling). Thus in essence, adding a second stage of separation extended the attainable dynamic range and reduced the under-sampling problem. Depletion of the most abundant proteins using antibodies prior to MD separation, which has been a regular practice for bottom-up proteomics, will further boost the dynamic range and will be incorporated in our future work.

Basically, the conventional data-dependent strategy, where precursor selection is typically based on the ion intensity, does not work effectively for intact protein MS/MS due to the presence of multiple charge states of one (or few) highly abundant (and typically related) proteins at any particular elution time. Hence, smarter MS/MS strategies, such as selection of a precursor ion based on charge state instead of intensity (*e.g.* fragmenting a single charge state *per* protein), will be required for improved TD MS.

Furthermore, the fact that the elution pattern in the first dimension WCX-HILIC is not governed exclusively by core histone family or by acetylation status, as is true in the corresponding offline 2-D separation scheme (RPLC-HILIC), complicates the downstream second dimension RPLC separation and LC-MS data analysis. However, this dilemma will disappear if the salt gradient WCX-HILIC is replaced with a pH gradient WCX-HILIC. Recently, Young *et al.* reported a novel approach for the high-throughput characterization of histones that uses a salt-free pH gradient WCX-HILIC, which can be directly coupled to an ion trap acquiring data-dependent ETD tandem mass spectra [25]. The 2-D LC platform described herein is highly versatile and can be easily adapted for reversed operation. In fact, we are currently investigating the possibility of applying this mode of chromatography (*i.e.* RPLC followed by WCX/HILIC) to first separate major histone families, and then differentially modified forms within each family to ease data analysis and interpretation and improve our ability to perform comparative LC-MS measurements.

In summary, an on-line metal-free 2-D WCX-HILIC/ RPLC system has been set up for sensitive high-throughput protein characterization using TD MS and applied for the characterization of core histones purified from human fibroblasts. Simultaneous separation of both core histones and their post-translationally modified isoforms presents a great challenge for any 1-D separation. Addition of a second orthogonal separation mode greatly increased the dynamic range, and enabled identification of low abundant isoforms, which otherwise are often missed due to insufficient intensity, an increased opportunity for ionization suppression and under-sampling issues. When coupled with FTMS, 2-D WCX-HILIC/RPLC platform enabled confident identification (*i.e.* with P scores below 10^{-4}) of 41 H4 isoforms starting from 24 μg of core histones (an equivalent of approximately 6 μg of H4). This represents significant improvement in sensitivity when compared to the corresponding offline system considering the total number of identifications *per* unit of sample.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank Professor Neil L. Kelleher for providing the ProSightPC and Dr. Paul Thomas for help with running the program. Portions of this work were supported by the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL) Intramural Research and Capability Development Program, the U.S. Department of Energy (DOE) Office of Biological and Environmental Research, and the NIH National Center for Research Resources (grant RR018522). The research was performed using EMSL, a national scientific user facility sponsored by the Department of Energy's Office of Biological and Environmental Research and located at Pacific Northwest National Laboratory in Richland, Washington. PNNL is a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RLO 1830.

Abbreviations

FTMS	Fourier-transform MS
HILIC	hydrophilic interaction LC

HOAC	glacial acetic acid
IPA	isopropanol alcohol
MD	multidimensional
MudPIT	multidimensional protein identification technology
PEEK	polyether ether ketone
TD	top-down
WCX	weak cation exchange

References

1. McLafferty FW, Breuker K, Jin M, Han X, et al. Top-down MS, a powerful complement to the high capabilities of proteolysis proteomics. *FEBS J.* 2007; 274:6256–6268. [PubMed: 18021240]
2. Gstaiger M, Aebersold R. Applying mass spectrometry-based proteomics to genetics, genomics and network biology. *Nat Rev Genet.* 2009; 10:617–627. [PubMed: 19687803]
3. Cortes, HJ. *Multidimensional Chromatography: Techniques and Applications.* Marcel Dekker; New York: 1990.
4. McDonald WH, Ohi R, Miyamoto DT, Mitchison TJ, Yates JR. Comparison of three directly coupled HPLC MS/MS strategies for identification of proteins from complex mixtures: single-dimension LC-MS/MS, 2-phase MudPIT, and 3-phase MudPIT. *Int J Mass Spectrom.* 2002; 219:245–251.
5. Holland LA, Jorgenson JW. Separation of nanoliter samples of biological amines by a comprehensive two-dimensional microcolumn liquid chromatography system. *Anal Chem.* 1995; 67:3275–3283. [PubMed: 8686888]
6. Washburn MP, Wolters D, Yates JR. 3rd, Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol.* 2001; 19:242–247. [PubMed: 11231557]
7. Shen Y, Jacobs JM, Camp DG II, Fang R, et al. Ultra-high-efficiency strong cation exchange LC/RPLC/MS/MS for high dynamic range characterization of the human plasma proteome. *Anal Chem.* 2004; 76:1134–1144. [PubMed: 14961748]
8. Wagner K, Miliotis T, Marko-Varga G, Bischoff R, Unger KK. An automated on-line multidimensional HPLC system for protein and peptide mapping with integrated sample preparation. *Anal Chem.* 2002; 74:809–820. [PubMed: 11866061]
9. Garcia BA, Siuti N, Thomas CE, Mizzen CA, Kelleher NL. Characterization of neurohistone variants and post-translational modifications by electron capture dissociation mass spectrometry. *Int J Mass Spectrom.* 2007; 259:184–196.
10. Lindner H, Helliger W, Puschendorf B. Histone separation by high-performance liquid-chromatography on C-4 reverse-phase columns. *Anal Biochem.* 1986; 158:424–430. [PubMed: 3812983]
11. Lindner H, Wesierska-Gadek J, Helliger W, Puschendorf B, Sauermann G. Identification of ADP-ribosylated histones by the combined use of high-performance liquid chromatography and electrophoresis. *J Chromatogr.* 1989; 472:243–249. [PubMed: 2760138]
12. Lindner H, Sarg B, Meraner C, Helliger W. Separation of acetylated core histones by hydrophilic-interaction liquid chromatography. *J Chromatogr A.* 1996; 743:137–144. [PubMed: 8817877]
13. Sarg B, Koutzamani E, Helliger W, Rundquist I, Lindner HH. Postsynthetic trimethylation of histone H4 at lysine 20 in mammalian tissues is associated with aging. *J Biol Chem.* 2002; 277:39195–39201. [PubMed: 12154089]
14. Lindner H, Sarg B, Helliger W. Application of hydrophilic-interaction liquid chromatography to the separation of phosphorylated H1 histones. *J Chromatogr A.* 1997; 782:55–62. [PubMed: 9440922]

15. Lindner H, Sarg B, Hoertnagl B, Helliger W. The microheterogeneity of the mammalian H1(0) histone. Evidence for an age-dependent deamidation. *J Biol Chem.* 1998; 273:13324–13330. [PubMed: 9582379]
16. Alpert AJ. Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. *J Chromatogr.* 1990; 499:177–196. [PubMed: 2324207]
17. Mant CT, Hodges RS. Mixed-mode hydrophilic interaction/cation-exchange chromatography (HILIC/CEX) of peptides and proteins. *J Sep Sci.* 2008; 31:2754–2773. [PubMed: 18668504]
18. Pesavento JJ, Bullock CR, LeDuc RD, Mizzen CA, Kelleher NL. Combinatorial modification of human histone H4 quantitated by two-dimensional liquid chromatography coupled with top down mass spectrometry. *J Biol Chem.* 2008; 283:14927–14937. [PubMed: 18381279]
19. Boyne MT, Garcia BA, Li M, Zamdborg L, et al. Tandem mass spectrometry with ultrahigh mass accuracy clarifies peptide identification by database retrieval. *J Proteome Res.* 2009; 8:374–379. [PubMed: 19053528]
20. Zhao R, Ding SJ, Shen Y, Camp DG 2nd, et al. Automated metal-free multiple-column nanoLC for improved phosphopeptide analysis sensitivity and throughput. *J Chromatogr B Analyt Technol Biomed Life Sci.* 2009; 877:663–670.
21. B'Hymer C, Sutton RMC, Sutton KL, Caruso JA. The use of sol-gel frits to minimize suction effects in capillary electrophoresis-nebulizer interfaces for plasma spectrometry. *Anal Commun.* 1999; 36:349–353.
22. Kelly RT, Page JS, Luo Q, Moore RJ, et al. Chemically etched open tubular and monolithic emitters for nanoelectrospray ionization mass spectrometry. *Anal Chem.* 2006; 78:7796–7801. [PubMed: 17105173]
23. Pesavento JJ, Kim YB, Taylor GK, Kelleher NL. Shotgun annotation of histone modifications: a new approach for streamlined characterization of proteins by top down mass spectrometry. *J Am Chem Soc.* 2004; 126:3386–3387. [PubMed: 15025441]
24. Horn DM, Zubarev RA, McLafferty FW. Automated reduction and interpretation of high resolution electrospray mass spectra of large molecules. *J Am Soc Mass Spectrom.* 2000; 11:320–332. [PubMed: 10757168]
25. Young NL, DiMaggio PA, Plazas-Mayorca MD, Baliban RC, et al. High throughput characterization of combinatorial histone codes. *Mol Cell Proteomics.* 2009; 8:2266–2284. [PubMed: 19654425]

Figure 1.

Schematic diagram of the online metal-free WCX-HILIC/RPLC system. Two Cheminert 5000 psi 6-port metal-free Nanovolume injection valves (V1 and V2) were used in the first dimension to facilitate sample injection to the sample loop, loading to the SPE column (SPE1), and HILIC separation (column 1). The eluate from the first dimension was monitored with an online UV detector before it was directed into the storage and then the second dimension separation using fractionation valves (V3, V4, V5, and V6). In the second dimension, a same type valve (but 10-port instead of 6-port) (V7) was used to control loading of the first dimension fraction onto the SPE columns (SPE2 and SPE3), separation with two RPLC columns (columns 2 and 3), and equilibration of the RPLC columns. Mixers 1 and 2, initially filled with Buffer A, were used to generate elution gradient by adding Buffer B with a certain split flow rate. Two metal unions were coupled in the split/purge line to apply the high voltage (marked as HV) for ESI. W, waste; D, dimension; MS, mass spectrometer.

Figure 2. MS-only TIC (bottom panel) and 2-D display (mono-isotopic mass (including both parent and fragment spectra) *versus* elution time) (top panel) reconstructed from RPLC-FTMS analysis of 2.4 μ g core histone mixture.

Figure 3. UV chromatogram acquired during the separation of 24 µg core histones using WCX-HILIC. Twenty fractions for the 20 peaks as numbered were collected for subsequent RPLC-MS/MS based protein identification *via* protein database search using ProSightPC. (a.u., arbitrary units.)

Figure 4. Representative MS-only TIC chromatograms (top panel) and 2-D displays for H2A, H2B, H3, and H4 reconstructed from RPLC-FTMS analyses of WCX-HILIC fractions 9, 12, 8, and 17 as indicated in Fig. 3.

Figure 5. aS1-aK16-2 mK20-H4 was identified with high confidence using ProSightPC. (A) survey mass spectrum. (B) Tandem mass spectrum of precursor ion m/z 946.70 ($z=12$) with matching b and y fragment ions marked with solid triangles. (C) Protein sequence with PTMs and matching b and y ions marked. The detailed information for the matching b and y fragment ions are given in Table 3.

Figure 6. Mass spectra of aS1-aK12-aK16-mK20-H4 in 1-D (A, B) and 2-D (C, D) separation. See text for details.

Table 1

Human histone H4 isoforms identified in RPLC-FTMS analysis

Theoretical mass (Da)	Observed mass (Da)	Mass difference (Da)	Number of observed fragments	Number of matching fragments	Number of best hits	PTMs	P score
11341.4	11341.4	0.0	318	45	1	aS1, aK16, 2mK20	1.30E-52
11299.4	11299.3	-0.1	366	47	1	aS1, 2mK20	1.40E-51
11285.4	11284.4	-1.0	366	33	1	aS1, mK20	9.72E-33
11327.4	11326.4	-1.0	321	27	1	aS1, aK16, mK20	2.08E-26
11341.4	11340.4	-1.0	228	25	4	aS1, mR3, aK8, mK20	1.67E-25
11382.4	11382.3	-0.1	318	20	1	aS1, aK16, 3mK20	2.47E-18
11283.4	11284.4	1.0	259	17	1	2mK20	3.32E-16
11369.4	11368.4	-1.0	130	14	20	mR3, aK5, aK8, aK16	7.30E-16
11383.4	11383.4	0.0	217	16	3	aS1, aK12, aK16, 2mK20	7.85E-16
11325.4	11326.3	0.9	318	17	1	aS1, 2mK20	7.43E-15
11337.3	11338.2	0.9	147	13	5	pS1, mR3, mK20	3.69E-14
11397.4	11398.3	0.9	67	10	15	aK5, aK8, aK12, aK16	2.67E-13
11339.4	11340.4	1.0	321	16	1	aS1, aK16	4.37E-13
11398.4	11398.2	-0.2	131	12	6	aS1, aK5, aK8, 3mK20	4.43E-13
11412.4	11413.4	1.0	71	10	10	mR3, aK5, aK8, aK12, 3mK20	4.80E-13
11340.4	11341.3	0.9	271	15	2	aS1, 3mK20	1.46E-12
11425.4	11425.5	0.1	36	8	5	aS1, aK5, aK8, aK16, 2mK20	4.62E-12
11352.4	11353.4	1.0	147	11	1	pS1, 3mK20	2.27E-11
11297.4	11298.4	1.0	228	12	1	aS1	3.68E-11
11402.4	11402.4	-0.1	124	9	8	aK6	2.80E-09
11314.4	11314.4	0.0	83	8	1	aS1, 3mK20	7.73E-09
11355.4	11355.3	-0.1	83	6	1	aS1, aK12, aK16	2.63E-06
11300.4	11300.3	-0.1	184	6	1	mmR3, 3mK20	9.08E-04

a = acetyl, m = methyl, 2m = dimethyl, mm = omega-N, omega-N'-dimethyl (arginine), 3m = trimethyl, p = phospho.

Table 2

Human histone H4 isoforms identified in WCX-HILIC/RPLC-FTMS analysis

Theoretical mass (Da)	Observed mass (Da)	Mass difference (Da)	Number of observed fragments	Number of matching fragments	Number of best hits	PTMs	P score
11341.4	11341.4	0.0	201	42	1	as1, aK16, 2mK20	2.51E-56
11299.4	11299.3	-0.1	238	37	1	as1, 2mK20	2.22E-45
11327.4	11326.5	-0.9	183	24	4	as1, aK16, mK20	1.43E-28
11313.4	11313.4	0.0	210	20	1	as1, aK16	8.04E-21
11314.4	11313.5	-1.0	204	22	5	as1, 3mK20	1.26E-20
11285.4	11285.3	-0.1	130	16	1	as1, mK20	2.97E-19
11355.4	11355.4	0.0	192	17	1	as1, aK12, aK16	6.81E-18
11325.4	11325.4	-0.1	201	16	1	as1, 2mK20	6.49E-17
11383.4	11383.5	0.1	69	12	10	as1, aK8, aK12, 2mK20	3.71E-16
11297.4	11297.4	0.0	82	11	1	as1	8.82E-15
11384.4	11385.3	0.9	119	13	4	as1, mmR3, aK5, 3mK20	1.72E-14
11354.4	11354.4	0.0	178	13	4	mR3, aK12, 3mK20	6.34E-13
11351.3	11352.2	0.9	198	14	2	as1, PY51	1.24E-12
11286.4	11287.4	1.0	61	9	1	mR3, 3mK20	3.92E-12
11300.4	11300.3	-0.1	136	11	1	mmR3, 3mK20	9.30E-12
11339.4	11340.3	0.9	143	11	4	as1, aK8	2.30E-11
11326.4	11327.3	0.9	82	9	5	mmR3, 3mK20	4.13E-11
11283.4	11284.3	0.9	238	12	1	2mR3	7.95E-11
11356.4	11356.4	0.0	103	9	1	as1, aK16, 3mK20	4.68E-10
11298.4	11298.4	0.0	227	11	1	3mK20	2.78E-09
11369.4	11370.4	1.0	152	9	1	as1, aK12, aK16, mK20	1.65E-08
11312.4	11313.3	0.9	208	11	1	mR3, 3mK20	1.88E-08
11340.4	11341.3	0.9	114	8	2	as1, 3mK20	2.02E-08
11396.4	11397.3	0.9	116	8	3	as1, mR3, aK8, 3mK20	2.42E-08
11402.4	11402.3	-0.1	74	7	14	aK6	3.05E-08
11398.4	11399.3	0.9	82	7	10	as1, aK5, aK8, 3mK20	6.83E-08
11352.4	11353.3	0.9	82	6	2	3mK20, pS47	3.82E-07
11416.4	11416.3	-0.1	110	7	2	aM1, mK21	3.97E-07

Theoretical mass (Da)	Observed mass (Da)	Mass difference (Da)	Number of observed fragments	Number of matching fragments	Number of best hits	PTMs	P score
11311.4	11311.3	-0.1	130	7	1	aS1, mK20	4.50E-07
11439.4	11439.3	-0.1	171	8	1	aS1, aK5, aK8, aK12, aK16	5.88E-07
11397.4	11397.4	0.0	171	8	1	aK5, aK8, aK12, aK16	5.88E-07
11394.4	11395.4	1.0	126	7	10	aK16, 3mK20, pS47	9.64E-07
11353.4	11354.4	1.0	79	6	5	aS1, mmR3, 2mK20	1.77E-06
11309.3	11310.3	1.0	207	8	2	pS47	1.89E-06
11411.4	11412.3	0.9	110	6	30	aS1, aK8, aK12, aK16, mK20	6.44E-06
11415.3	11414.4	-0.9	113	6	1	pS47, pY51	1.30E-05
11395.4	11396.4	1.0	114	6	3	aK8, aK12, aK16, mK20	1.46E-05
11342.4	11343.4	1.0	141	6	5	mmR3, aK16, 3mK20	1.82E-05
11328.4	11329.4	1.0	157	6	1	aS1, mR3, 3mK20	5.89E-05
11408.4	11408.4	0.0	171	6	14	pS1, mR3, aK8, 3mK20	6.69E-05
11412.4	11413.2	0.8	163	6	1	aS1, mR3, aK5, aK8, 3mK20	7.10E-05

a =acetyl, m =methyl, 2m =dimethyl, mm =omega-N, omega-N'-dimethyl (arginine), 3m =trimethyl, p =phospho.

Table 3

The detailed information for the matching b and y fragment ions as marked in Fig. 5C

Name	m/z	Mass monoisotopic (Da)	Theoretical mass (Da)	Error (ppm)	Name	m/z	Mass monoisotopic (Da)	Theoretical mass (Da)	Error (ppm)
B3	343.17	342.17	342.17	2.98	Y23	1252.17	2502.33	2502.31	9.01
B8	770.42	769.42	769.42	-3.60	Y24	1316.21	2630.40	2630.40	0.89
B17	818.98	1635.94	1635.93	4.43	Y28	781.68	3122.70	3122.69	1.97
B25	671.64	2682.53	2682.54	-2.82	Y30	1118.60	3352.77	3352.78	-4.33
B35	753.04	3760.17	3760.20	-8.90	Y31	1172.95	3515.83	3515.85	-6.04
B39	852.31	4256.52	4256.53	-2.22	Y32	1206.64	3616.91	3616.90	4.43
B40	736.44	4412.61	4412.63	-4.32	Y33	1239.67	3716.00	3715.96	8.59
B52	813.20	5685.36	5685.35	2.23	Y34	1263.34	3787.00	3787.00	-0.35
B53	970.07	5814.38	5814.39	-1.53	Y39	1097.09	4384.33	4384.32	0.24
B62	854.51	6828.02	6828.03	-1.93	Y42	1194.39	4773.51	4773.52	-1.06
B68	840.38	7554.35	7554.40	-6.62	Y43	1219.16	4872.60	4872.59	2.86
B70	966.56	7724.45	7724.50	-6.71	Y47	1318.48	5269.88	5269.86	5.25
Y11	606.32	1201.63	1210.62	5.78	Y49	922.17	5527.00	5527.01	-1.88
Y13	726.91	1451.81	1451.80	6.19	Y49	1106.41	5527.00	5527.01	-1.55
Y14	762.43	1522.85	1522.84	6.17	Y50	1132.21	5655.99	5656.05	-9.62
Y15	843.96	1685.91	1685.90	3.48	Y57	1287.30	6341.45	6431.46	-1.82
Y16	893.50	1784.98	1784.97	5.93	Y63	1013.13	7084.84	7084.87	-4.29
Y18	1000.54	1999.06	1999.07	-1.06	Y67	1084.04	7581.22	7581.19	4.11
Y19	1066.05	2130.09	2130.11	-5.44	Y77	1083.37	8658.92	8658.86	6.60
Y20	1101.57	2201.13	2201.14	-6.18	Y90	1022.68	10216.78	10216.80	2.12
Y21	1152.09	2302.17	2302.19	-7.76	Y94	962.10	10572.00	10572.00	1.85
Y22	1201.63	2401.25	2401.26	-3.50					