ELSEVIER

Journal of Chromatography B, 783 (2003) 173–179

JOURNAL OF CHROMATOGRAPHY B

www.elsevier.com/locate/chromb

Analysis of core histones by liquid chromatography–mass spectrometry and peptide mapping

Kangling Zhang[a,*], Hui Tang[b]

[a]*UCR MS facility, Department of Chemistry, University of California at Riverside, Riverside, CA 92521, USA*
[b]*Department of Chemistry, University of the Pacific, Stockton, CA 95211, USA*

Received 29 March 2002; received in revised form 23 August 2002; accepted 23 August 2002

Abstract

Histone acetylation and methylation are processes that are generally considered to play crucial roles in the chromatin-based regulatory mechanism. Characterization of the histones as well as their modification sites has become increasingly important. In this paper, the use of LC–MS and peptide mapping methods to analyze chicken core histones and identify the modification sites is reported. Microbore C_4 HPLC separated the core histones into H2A, H2B, H3 and H4 using HFBA as the ion-pairing agent. The four subclasses of histones and their putative acetylated or methylated isoforms were identified by LC–MS simultaneously. MALDI-TOF and tandem mass spectrometry provided peptide mapping of the modification sites of the histones through trypsin digestion of the HPLC eluents. This approach is straightforward and prospective for further application in the field of chromatin research.

Keywords: Peptide mapping; Histones

1. Introduction

Core histone octamers (H2A, H2B, H3 and H4) wrapped around by DNA constitutes nucleosomes, the fundamental structural units of chromatin [1,2]. The N-termini of histones, also called histone tail, contain multiple basic lysine residues that are targets for post-transcriptional modifications—mainly acetylation and methylation [3]. Histone acetylation and methylation are processes that are generally considered to play crucial roles in the chromatin-based regulatory mechanism [4,5]. Histone isolation, identification and the determination of its modification are essential steps prior to its study [6]. SDS–PAGE has been previously used to separate the core histones into subclasses; separation of the acetylated and/or methylated isoforms of each histone subclass has not yet been addressed although a differential migration for acetylated histones has been observed in electrophoresis [7]. AUT gel electrophoresis can be used to separate individual acetylation isoforms of histone H3 and H4, however it is time consuming (2 full days) and acetylation isoforms cannot be differentiated from methylation isoforms [8–10]. Furthermore, different degrees of methylation of histones (mono, di and tri) cannot be separated on the gel. Reversed-phase, ion-exchange or hydrophilic-interaction liquid chromatography

*Corresponding author. Tel.: +1-909-787-5287; fax: +1-909-787-4713.
E-mail address: kangling@citrus.ucr.edu (K. Zhang).

1570-0232/02/$ – see front matter
PII: S1570-0232(02)00631-1

have proved to be effective in isolating individual isoforms of core histones [6,11–14], however, the identification of each fraction still relies on AUT PAGE gel electrophoresis and microsequencing.

LC–MS, MALDI-TOF and tandem MS have emerged as irreplaceable technologies in protein identification and peptide mapping for modification sites [15,16]. In this report, we describe the application of this technology to the identification of histones and their modified isoforms. Histones were separated by microbore C_4 reversed-phase HPLC. The effluent stream was split 1 to 10 with a T-splitter allowing 10% to enter the electrospray mass spectrometer while 90% was collected and subjected to enzymatic digestion and characterization by MALDI-TOF or tandem MS when sequence analysis is necessary. Fig. 1 summarizes this process. The LC–MS and peptide mapping methodology developed in this report is a new approach to identifying histone modification sites. The directness, high speed and sensitivity of LC–MS allows its application to the area of chromatin research which has recently become very popular.

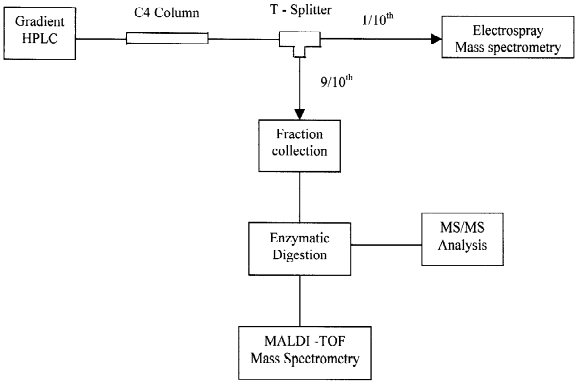

Fig. 1. Schematic of using LC–MS and peptide mapping by MS to identify chicken core histones. Histones were first separated by microbore C_4 column HPLC. 1/10th of the HPLC fraction from a T-splitter was introduced to the electrospray source, and the remainder was subjected to enzymatic digestion and MALDI-TOF or tandem MS analysis.

2. Experimental

2.1. Isolation of core histones from chicken blood

Histone preparation from chicken blood was performed as described previously [17–19]. Briefly, fresh blood from young adult (3–4 months old) chickens (Foster Farms, Livingston, CA, USA) was mixed with anticoagulant (16 m*M* citric acid, 89 m*M* trisodium citrate, 16 m*M* NaH_2PO_4, 0.13 *M* glucose, using 140 ml anticoagulant per liter blood) on ice and the erythrocytes separated by repeated centrifugation (5000 *g*, 10 min) and washed (0.14 *M* NaCl, 10 m*M* trisodium citrate, 0.1 m*M* PMSF). The fine pellet was lysed in Triton X-100 sucrose buffer (0.25 *M* sucrose, 0.010 *M* $MgCl_2$, 0.5 m*M* PMSF, 0.05 *M* Tris–HCl, pH 7.50 and 0.5% Triton X-100) overnight and washed (the washing buffer is same as the lysis buffer without detergent) to whiteness and stored at 3 °C to release the nuclei. The washed pellet were suspended in 2 l of 0.2 *M* sulfuric acid and stirred at 4 °C for 4 h. The supernatant after centrifugation at 2000 *g* for 15 min was saved and the precipitate was re-extracted with 1 l of 0.2 *M* sulfuric acid. The extracts were pooled and dialyzed against distilled water and lyophilized to afford 0.5 g core histones per liter of blood.

2.2. LC–MS analysis of core histones

LC–MS was conducted using an Applied Biosystems (Foster City, CA, USA) Model 140 B HPLC solvent delivery system coupled to a Mariner TOF mass spectrometer equipped with an electrospray ion source (Applied Biosystems). The histone samples (10 μl containing 7.5 μg histones) were injected onto a Vydac (Western Analytical Products, Murrieta, CA, USA) C_4 microbore column (150×2.0 mm, 5 μm particle size) linked with a C_{18} guard column (2.1 mm I.D. cartridge, 5 μm particle size). The histones were eluted by a linear gradient from 38% B to 55% within 120 min at a flow-rate of 50 μl/min. Mobile phase A was water, 0.10% (v/v) HFBA, and mobile phase B was 0.08% HFBA in acetonitrile. Approximately 10% of the eluent flow after the T-splitter was introduced into the electrospray ionization (ESI) source of the mass spectrometer. The remaining eluent was introduced to the Beckman (Palo Alto, CA, USA) Model 163 UV detector set at 214 nm wavelength and fractions were collected and dried to a certain concentration by a speedvac for further enzymatic digestion and subsequent MALDI-

TOF or tandem MS analysis. The mass spectrometer was run in the positive ion mode. Mass spectrometric parameters were optimized to obtain the maximum sensitivity using a tuning solution [1 pmol/μl of gramicidin S in methanol–water–0.1% acetic acid (50:49.9:0.1, v/v)]. The capillary voltage was 2200 V, the orifice voltage was 60 V and the source temperature was 150 °C. Ultra-high pure nitrogen gas was used as nebulizing (1.0 l/min) and auxiliary gases (0.3 l/min). The mass spectrometer was also calibrated by the above tuning solution. Full-scan data were recorded by scanning from 500 to 2500 m/z in 5 s.

2.3. Enzymatic digestion and MALDI-TOF mass spectrometry

Each dried HPLC fraction was redissolved in 20 μl of 25 m*M* ammonium bicarbonate containing 0.1 μg of trypsin (Sigma). The solution was incubated at 37 °C overnight and then dried to completeness by a Speedvac. The digested peptides were dissolved in 10 μl of 0.1% trifluroacetic acid (TFA) and 1 μl of the solution was mixed with α-cyano-4-hydroxy-cinamic matrix solution (Hewlett-Packard) and exposed to the MALDI analysis. Monoisotopic masses of all peptides were measured by MALDI using a Voyager DE-STR Biospectrometry Workstation (Perspective Biosystem, Foster City, CA, USA) with delayed extraction operated in the reflectron mode. All MALDI spectra were externally calibrated by using a peptide standard mixture consisting of angiotensin II (monoisotopic mass 1046.5423, Sigma), bradykinin (monoisotopic mass 1060.5692, Sigma), LHRH (monoisotopic mass 1182.5809, Sigma), bombesin (monoisotopic mass 1619.8229, Sigma), α-MSH (monoisotopic mass 1664.8008, Sigma) and adrenocorticotropic hormone fragment 18–39 (monoisotopic mass 2465.1989, Sigma). From 100 to 256 shots were accumulated to give an acceptable spectrum.

3. Results and discussion

3.1. LC–MS analysis of core histones

Reversed-phase HPLC has been successfully used for core histone separation using HFBA, an alternative to TFA as ion-pair agent [11]. We applied the same HPLC conditions to LC–MS based on the strategy shown in Fig. 1. Histones were separated by HPLC using a microbore C_4 column and a linear narrow gradient increasing buffer B from 38% to 55% within 120 min (mobile phase A was water, 0.10% (v/v) HFBA, and mobile phase B was 0.08% HFBA in acetonitrile). The four histones H2B, H2A, H4 and H3 elute sequentially as seen in Fig. 2 which shows the total ion current (TIC) diagram.

Figs. 3–7 display the electrospray spectra for all the major peaks shown in the TIC labeled H2B, H2A, H4, H3a and H3b, respectively. The inserts are the reconstructed spectra using the deconvolution software from which the molecular mass of each histone can be read out. As shown in the inserted reconstructed spectrum in Fig. 3, peak of m/z 13 791.1 is assigned as H2B in a good agreement with the calculated monoisotopic molar mass (13 791.0) of histone H2B. Beside the main peak, a series of peaks with smaller intensities at m/z 13 805.2, 13 818.6, 13 832.5 were observed. It demonstrated H2B has either one site to be mono-, di- and trimethylated (or monoacetylated because monoacteylated protein has the same molecular mass as the trimethylated protein) or three sites to be monomethylated. Peptide mapping by enzymatic digestion and analysis by tandem MS supported that lysine 30 is mono-, di- and trimethylated (author's unpublished data).

In Fig. 4, the peak at m/z 13 850.8 was assigned to H2A whose calculated monoisotopic molar mass is 13 851.0. Peaks of m/z 13 892.88 and m/z 13 934.38 are the one- and two-site acetylated isoforms. Peptide mapping by enzymatic digestion and

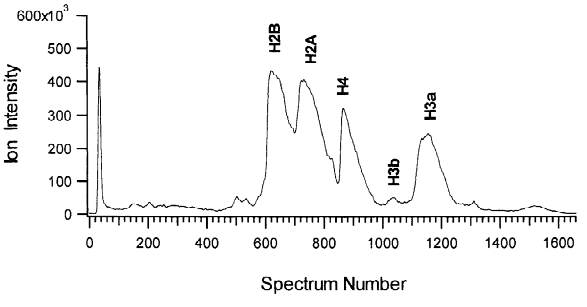

Fig. 2. TIC chromatogram of LC–MS of core histones.

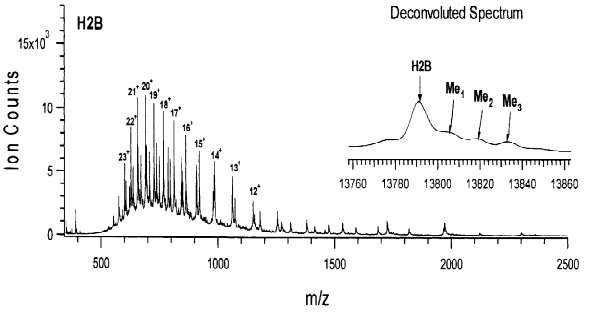

Fig. 3. ESI mass spectrum of histone H2B corresponding to the peak 1 in the TIC diagram of Fig. 2. Accumulated scans covering the full area of each peak were averaged for a spectrum with a better signal-to-noise ratio. The insert spectra were reconstructed from the charge distribution profile of the measured electrospray spectra by the instrument inherent software (same for the following figures). Me_1, Me_2 and Me_3, mono-, di- and trimethylation.

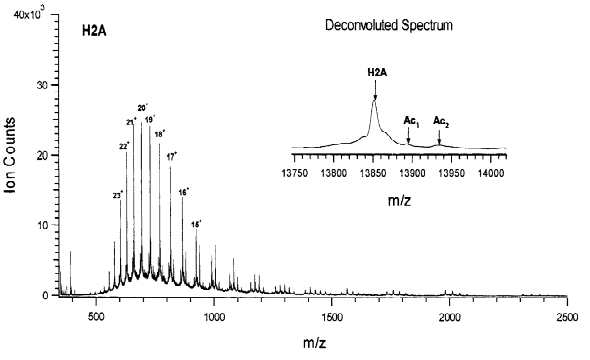

Fig. 4. ESI mass spectrum of histone H2A corresponding to the peak 2 in TIC of Fig. 2. Ac_1 and Ac_2, mono- and diacetylated isoforms.

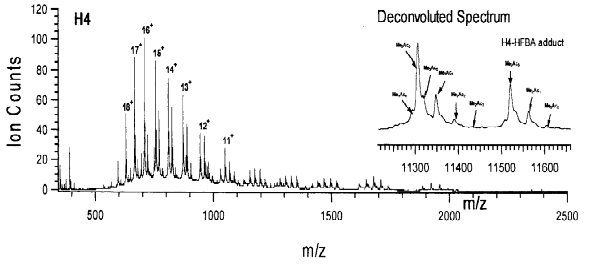

Fig. 5. ESI mass spectrum of histone H4. In the insert, the convoluted spectrum, six peaks at m/z of 11 306.2, 11 292.2, 11 320.5, 11 348.3, 11 380.4 and 11 432.5 were observed. Me_1Ac_0, Me_2Ac_0 and Me_3Ac_0, mono-, di- and trimethylated and nonacetylated H4 isoforms. Me2Ac1, Me2 Ac2 and Me2Ac3, mono-, di- and triacetylated as well as dimethylated H4 isoforms. H4 and HFBA adduct was observed at m/z 11 520.

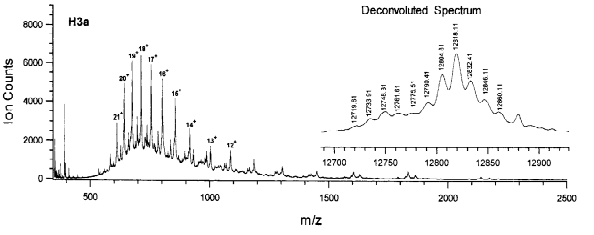

Fig. 6. ESI mass spectrum of histone H3a. In the insert, the convoluted spectrum, a group of peaks differing by 14 Da between each pair of neighboring peaks showed that multiple methylation and/or acetylation isoforms existed.

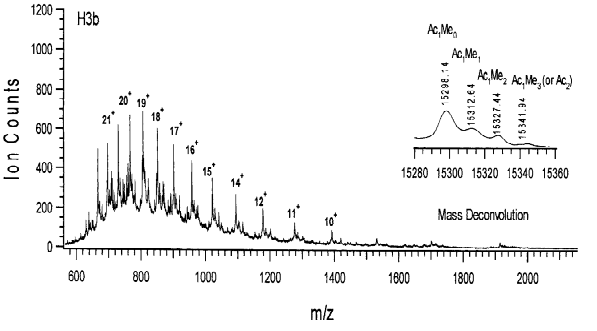

Fig. 7. ESI mass spectrum of histone H3b. In the insert, the convoluted spectrum, four peaks are observed at m/z of 15 298.1, 15 312.6, 15 327.4 and 15 341.9 corresponding to multiply methylated or acetylated H3 isoforms.

MALDI post-source decay (MALDI-PSD) (see Fig. 8) analysis showed that of these two isoforms of H2A are one in which lysine 9 is acetylated and the other in which lysine 4 and 9 are both acetylated.

In Fig. 5, the major peak of m/z 11 306.2 is the unacetylated histone H4 (note that the first serine amino group is acetylated) while lysine 20 is dimethylated. Mono- (peak at m/z of 11 292.2) and trimethylated (peak at m/z of 11 320.06) H4 only constitute a small percentage. Peaks of m/z 11 348.3, m/z 11 390.4 and m/z 11 432.5 are the mono-, di- and triacetylated histone H4 isoforms, respectively, in which dimethylated lysine 20 dominates. The peak of m/z 11 520.0 is the H4-HFBA adduct accompanied by the same pattern of acetylated isoforms. Peptide mapping by trypsin digestion and MALDI-TOF MS analysis reveals that lysine 16, 12 and 8 are acetylated in the order of 16–12–8 (see Table 1 and details in next section). MS–MS analysis has confirmed this unique arrangement of acetylation pattern

Table 1
MALDI-TOF analysis of HPLC fractions containing H2a, H2b, H3a, H3b and H4

HPLC peaks	Histones	Measured masses	Calculated masses	Peptide sequences	
1	H2B	664.30	664.36	87–92	STITSR
		703.33	703.37	73–79	IAGEASR
		816.39	816.45	93–99	EIQTAVR
		901.43	901.49	80–86	LAHYNKR
		1743.60	1743.81	58–72	AMGIMNSFVNDIFER
2	H2A	850.48	850.52	82–86	HLQLAIR
		944.33	944.52	21–29	AGLQFPVGR
		1692.90	1692.90	82–95	HLQLAIRNDEELNK
		2104.10	2104.18	82–99	HLQLAIRNDEELNKLLGK
3	H4	515.33	515.33	36–39	RLAR
		530.31	530.26	13–17	$GGAK_{ac}R$[a]
		543.41	543.36	20–23	$K_{me2}VLR$
		927.54	927.55	9–17	$GLGK_{ac}GGAK_{ac}R$
		989.64	989.57	60–67	VFLENVIR
		1180.70	1180.61	46–55	ISGLIYEETR
		1211.70	1211.71	6–17	$GGK_{ac}GLGK_{ac}GGAK_{ac}R$
		1325.80	1325.75	24–35	DNIQGITKPAIR
		1438.80	1438.84	4–17	$GK_{ac}GGK_{ac}GLGK_{ac}GGAK_{ac}R$
		1466.80	1466.76	80–92	TVTAMDVVYALKR
4	H3a	537.31	537.32	37–40	KPHR
		594.25	594.33	53–56	RYQK
		630.27	630.36	130–134	IRGER
		715.33	715.40	123–128	DIQLAR
		831.33	831.49	57–63	STELLIR
		844.36	844.50	116–122	RVTIMPK
		1032.39	1032.59	41–49	YRPGTVALR
		1447.80 $+14.02n_{1-6}$	1447.85	27–40	$K_{me1-3}SAPATGGVK_{me1-3}KPHR$
5	H3b[b]	704.39 $+14.02n_{0-1}$	704.40	3–8	$TK_{me0-1}QTR$
		901.54 $+14.02n_{0-6}$	901.52	9–17	$K_{me0-3}STGGK_{me0-3}APR$
		1028.57	1028.62	18–25	KQLATKacAAR
		1070.59	1070.63	18–25	$K_{ac}QLATKacAAR$
		1335.70	1335.69	73–83	$EIAQDFK_{me0-2}TDLR$

[a] K_{ac} and K_{me} represent the acetylated and methylated lysines and numbers 0–3 represent mono-, di- and trimethylted lysines.
[b] Only masses not observed in H3a are shown here.

and this pattern is consistent with that observed in human HeLa cells [20]. Data also showed that dimethylation of lysine 20 occurs in all isoforms of histone H4 [20].

From the reconstructed spectrum in Fig. 6, we observed a series of peaks from the smallest m/z 12 719.61 to the highest m/z 12 860.11 separated by 14 Da between each pair of neighboring peaks. These probably originate from multiple methylation of lysine residues 27 and 36 and 79. Since the detected mass for the peak was around 2500 Da smaller than the calculated isotopic molar mass of histone H3, it may be either a variant of H3 short of an N-terminal about 24 amino acid length ($MH^+ = 2554.6$) or the N-terminal 1–24 was lost during the process of protein isolation. As seen from the X-ray structure of human nucleosome, the N-terminal 1–45 of H3 extends out of the chromatin assembly mark-

edly [2] and thus the N-terminal 1–24 is more likely subjected to be cleaved by an undefined protease or mechanically sheared when DNA is detached from core histones by acid precipitation to extract core histones.

In Fig. 7, the electrospray mass spectrum of HPLC fraction containing histone H3b. After deconvolution (see the insert spectrum), a series of four peaks of mass values separated by 42 Da at m/z of 15 298.1, 15 312.6, 15 327.4 and 15 341.9 was observed. These correspond to multiple acetylation or methylation of histone H3. Peptide mapping demonstrated that lysine 4 is monomethylated, lysines 9, 14, 27, 36 and 79 are from mono- to trimethylated, lysines 14, 18 and 23 are acetylated [19].

3.2. MALDI-TOF MS analysis of histones and peptide mapping

A total of 90% of the HPLC effluent was collected manually following the UV peaks; the UV profile is not shown here. Five major fractions were collected and analyzed by MALDI-TOF MS following trypsin digestion. Histone identification was made by matching the measured masses of the peptides with the theoretical masses of the enzymatic products of a specific histone (Table 1). The identified H2A, H2B, H4, H3a and H3b corresponding to the major peaks on the TIC of LC–MS are labeled on top of each peak (Fig. 2).

Four types of acetylated isoforms of histone H4 were identified directly by MALDI-TOF MS. As shown in Table 1, masses 530, 927, 1211 and 1438 correspond to one-site acetylation at lysine 16, two-site acetylation at lysines 12 and 16, three-site acetylation at lysines 8, 12 and 16, and four-site acetylation at lysines 5, 8, 12 and 16. The acetylation pattern is exactly the same as observed in the sodium butyrate treated Hela S3 cells for which a ''zip'' mechanism was proposed by Zhang et al. [20]. It has not been feasible to identify the modification isoforms of histone H2A, H2B and H3 using trypsin digestion and MALDI-TOF MS. Because avian erythrocyte histones are only modified at low levels [21] and because the N-termini of histones contain multiple lysine (acetylated, methylated and unmodified) and arginine residues, they are easily cleaved by trypsin into various fragments ($M_w < 500$) which cannot be detected by MALDI-TOF MS due to matrix interference. A strategy to use a low trypsin-to-protein ratio and short incubation time has proved practicable for identification of histone H3 modifications. Using this strategy, lysines 27, 36 and 79 have been identified to be methylated in histone H3a. Besides the modification sites observed in H3b, lysines 4, 9 and 14 in histone H3b have been determined to be methylated and lysines 14, 18 and 23 are acetylated. The details can be found in Ref. [19]. H2A and H2B were treated with V8 which cleaved the protein at the C-terminal of amino acids D and E. Two peptides with the same sequence in H2A have been isolated by HPLC and analyzed by

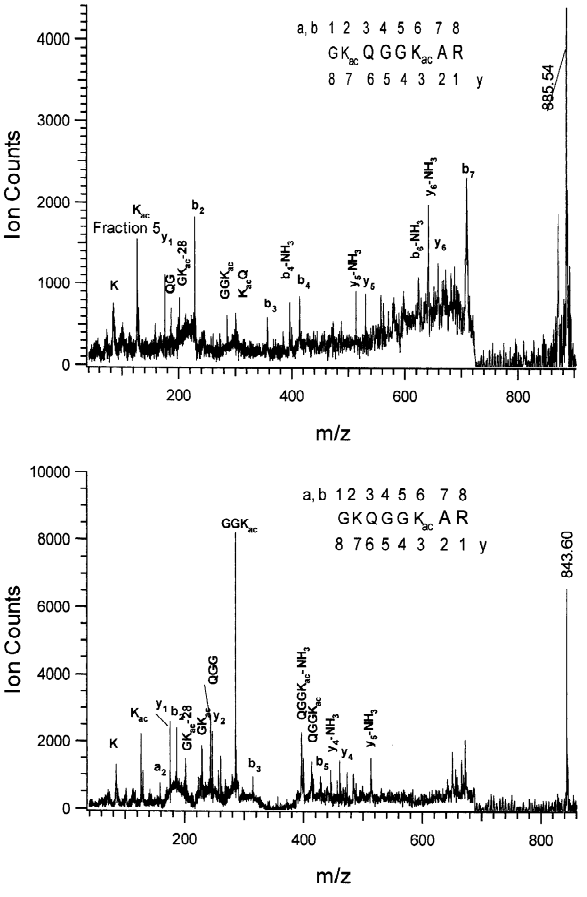

Fig. 8. MALDI-PSD spectra of precursor ions at m/z 843 and 857. Sequence analysis shows lysine 9 in the peptide GKQGG^{9}K$_{ac}$AR (m/z 843) is acetylated and both lysine 4 and 9 in the peptide G^{4}K$_{ac}$QGG^{9}K$_{ac}$AR (m/z 857) are acetylated. The fragmentation schematics are shown in the spectra.

MS. Peptide mapping by tandem mass MS showed that in one of them, lysine 9 is acetylated, and in the other, both lysine 4 and lysine 9 are acetylated (Fig. 8). Similarly, lysine 30 in H2B was identified to be mono-, di- and tri-methylated (author's unpublished data), which is consistent with the observation by LC–MS that three isoforms of H2B, among which 14 Da mass difference between one and the other, were detected. Based on the peptide mapping results, acetylation for histone H4, H2A and H3 is in the order from the histone structural portion to the end of the tail portion i.e. from lysine 16→12→8→5 in histone H4, from lysine 9→4 in histone H2A and from lysine 23→18→14 (?) in histone H3. Therefore, the ''zip'' model may be applied in all histones except H2B in which no acetylation sites have been observed.

4. Conclusion

LC–MS in combination with MALDI-TOF MS proved capable of identifying histones and their acetylation and methylation isoforms. With the advent of a mass spectrometer with high sensitivity and resolution such as a quadruple TOF instrument coupled with a nano-flow HPLC, the critical application of this methodology to small amounts of core histones isolated from biological fluids or tissues and the determination of in vivo sites of histone acetylation and methylation will become possible [22].

5. Nomenclature

MALDI-TOF	matrix assisted laser desorption/ionization-time of flight
LC–MS	liquid chromatography–mass spectrometry
HFBA	heptafluorobutyric acid
HAT	histone acetyltransferase
HDAC	histone deacetylase
AUT	acetic–urea–Triton X-100
PMSF	phenylmethylsulfonyl fluoride.

Acknowledgements

Mass spectral data were provided by the UCSF Mass Spectrometry Facility (A.L. Burlingame, Director) supported by the Biomedical Research Technology Program of the National Center for Research Resources, NIH NCRR BRTP RR01614. Support from Professor A.L. Burlingame is greatly acknowledged.

References

[1] V. Ramakrishnan, Curr. Opin. Struc. Biol 4 (1994) 44.
[2] K. Luger, A. Wmäder, R.K. Richmond, D.F. Sargent, T.J. Richmond, Nature 389 (1997) 251.
[3] B.D. Strahl, C.D. Allis, Nature 403 (2000) 41.
[4] V.G. Allfrey, R. Faulkner, A.E. Mirsky, Proc. Natl. Acad. Sci. USA 51 (1964) 786.
[5] T. Jenuwein, C.D. Allis, Science 293 (2001) 1074.
[6] K.W. Marvin, P. Yau, E.M. Bradbury, J. Biol. Chem. 265 (1990) 19839.
[7] E.I. Georgieva, R. Sendra, Anal. Biochem. 269 (1987) 399.
[8] S. Panyim, R. Chalkley, Biochemistry 8 (1969) 3972.
[9] B.A. Moss, Biochim. Biophys. Acta 359 (1974) 205.
[10] S. Panyim, R. Chalkley, Arch. Biochem. Biophys. 130 (1969) 337.
[11] M. McCroskey, V.E. Groppi, J.D. Pearson, Anal. Biochem. 163 (1987) 427.
[12] M. Couppez, D. Belaïche, Arch. Biochem. Biophys. 325 (1996) 29.
[13] H. Linder, B. Sarg, C. Meraner, W. Helliger, J. Chromatogr. A 743 (1996) 137.
[14] N. Zh, M.J.B. Louis, C. Mahadevan, J. Chromatogr. A 763 (1997) 65.
[15] B. Kuster, M. Mann, Curr. Opin. Struc. Biol. 8 (1998) 393.
[16] K.F. Medzihradszky, G.W. Adams, A.L. Burlingame, J. Am. Soc. Mass Spectrom. 7 (1996) 1.
[17] J.M. Sun, H.Y. Chen, M. Moniwa, S. Samuel, J.R. Davie, Biochemistry 38 (1999) 5939.
[18] M.J. Hendzel, J.R. Davie, J. Biol. Chem. 264 (1989) 19208.
[19] K. Zhang, H. Tang, L. Huang, J.W. Blankenship, P.R. Jones, F. Xiang, P.M. Yau, A.L. Burlingame, Anal. Biochem. 306 (2002) 259.
[20] K. Zhang, L. Huang, K. Williams, E.M. Bradbury, P.M. Yau, P.R. Jones, A.L. Burlingame, Mol. Cell Proteomics 1 (2002) 500.
[21] J.H. Waterborg, J. Biol. Chem. 265 (1990) 17157.
[22] J.C. Rice, C.D. Allis, Curr. Opin. Cell Biol. 13 (2001) 263.