Formaldehyde-Induced Histone Modifications in Vitro

ARTICLE *in* CHEMICAL RESEARCH IN TOXICOLOGY · AUGUST 2008

Impact Factor: 3.53 · DOI: 10.1021/tx8000576 · Source: PubMed

5 AUTHORS, INCLUDING:

Gunnar Boysen
University of Arkansas for Medical Sciences
69 PUBLICATIONS **1,228** CITATIONS
SEE PROFILE

Leonard Collins
University of North Carolina at Chapel Hill
31 PUBLICATIONS **1,490** CITATIONS
SEE PROFILE

James Swenberg
University of North Carolina at Chapel Hill
423 PUBLICATIONS **15,507** CITATIONS
SEE PROFILE

Formaldehyde-Induced Histone Modifications *in Vitro*

Kun Lu,[†,‡] Gunnar Boysen,[‡] Lina Gao,[‡] Leonard B. Collins,[‡] and James A. Swenberg*[,‡]

Curriculum in Applied Science & Engineering and Department of Environmental Science and Engineering, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599

Received February 13, 2008

Numerous experiments have demonstrated the genotoxic and mutagenic effects of formaldehyde, including DNA−protein cross-links (DPC). Histone was reported to be involved in the formation of DPC in which the ε-amino groups of lysine and exocyclic amino groups of DNA were thought to be cross-linked through multiple step reactions. Using mass spectrometry, the N-terminus of histone and lysine residues located in both the histone N-terminal tail and the globular fold domain were identified as binding sites for formaldehyde in the current study. The observation that only lysine residues without post-translational modification (PTM) can be attacked by formaldehyde indicates that PTM blocks the reaction between lysine and formaldehyde. Additionally, we found that formaldehyde-induced Schiff bases on lysine residues could inhibit the formation of PTM on histone, raising the possibility that formaldehyde might alter epigenetic regulation.

Introduction

Formaldehyde is produced worldwide on a large scale and is widely used in the manufacture of resins, particle board, plywood, leather goods, paper, pharmaceuticals, and other products (*1*). Therefore, occupational and environmental exposures to formaldehyde are quite common. Previous research has demonstrated that formaldehyde is genotoxic and mutagenic to mammalian cells, as well as to bacteria and lower eukaryotes (*1–6*). Exposure to formaldehyde results in the formation of DNA−protein cross-links (DPC[1]) as a primary genotoxic effect (*2–7*), and histones are reported to be cross-linked to DNA (*7, 8*). The formation of DPC, in which the ε-amino groups of lysine and exocyclic amino groups of DNA are thought to be involved (*9, 10*), proceeds in multiple-step reactions, with formaldehyde initiating an attack on the amino groups of protein, followed by DPC formation.

Core histones are located in the nucleosome, the fundamental repeating unit of chromatin, and contain a histone 3 (H3)/histone 4 (H4) tetramer flanked by 2 histone 2A (H2A)/histone 2B (H2B) dimers, around which 147 base pairs of DNA are wrapped (*11*). The highly conserved structure of histone includes a folded core and an unstructured tail. The histone core is a globular fold domain with a helix−loop−helix−loop−helix motif, which facilitates dimerization through a hand-shaking motion (*12*). However, the crystal structure reveals that histone tails do not have any defined conformation. Many conserved amino acid residues such as lysine are subject to a broad spectrum of post-translational modifications (PTM) including methylation and acetylation. These modifications impact biological actions such as gene expression by acting as markers for the specific recruitment of regulatory complexes and by changing the structure of the chromatin between heterochromatin and euchromatin, which is usually referred to as epigenetic regulation (*13*).

It has been shown that the lysine residues of histones are probably involved in DPC. However, the accessibility of amino acids, which is controlled not only by reactive ability but also by the conformation states of the residues (*14*), has not been evaluated. In addition, a wealth of post-translational modifications on histone raises the question whether the PTM on histone have any influence on the reaction with formaldehyde. Another very important question is whether formaldehyde-induced modifications impact the formation of PTM on histone lysine residues. In this investigation, histone 4 was chosen as our model histone protein due to its indispensable role in epigenetic regulation. Using mass spectrometry, we have now identified the residues of histone 4 that are accessible to formaldehyde binding. We have also demonstrated that PTM on lysine residues block the reaction between formaldehyde and the ε-amino groups of lysine residues. In addition, we found that formaldehyde-induced Schiff bases on lysine can inhibit the formation of PTM.

Materials and Methods

Chemicals and Reagents. Potassium phosphate, ammonium bicarbonate, glycine, trifluoroacetic acid, formic acid, acetonitrile, sodium cyanoborohydride ($NaCNBH_3$), acetyl coenzyme A trilithium salt (acetylCoA), and endoproteinase Asp-N were purchased from Sigma (St. Louis, MO). Modified sequencing grade trypsin and chymotrypsin were from Promega (San Luis Obispo, CA) and PrinSep (Adelphia, NJ), respectively. The histone 4 isolated from calf thymus was purchased from Roche Applied Science (Indianapolis, IN) and human recombinant histone 4 expressed in *E.coli*, active PCAF enzyme, and 5X histone acetylation (HAT) assay buffer (250 mM Tris-base, pH 8.0, 50% glycerol, 0.5 mM EDTA, and 5 mM dithiothreitol) were obtained from Millipore (Billerica, MA). 20% formaldehyde in water was procured from Tousimis (Rockville, MD),

* Corresponding author. James A. Swenberg, D.V.M., Ph.D., The University of North Carolina at Chapel Hill, Department of Environmental Sciences and Engineering, CB# 7431, Chapel Hill, NC, 27599. Phone: (919) 966-6139. Fax: (919) 966-6123. E-mail: jswenber@email.unc.edu.

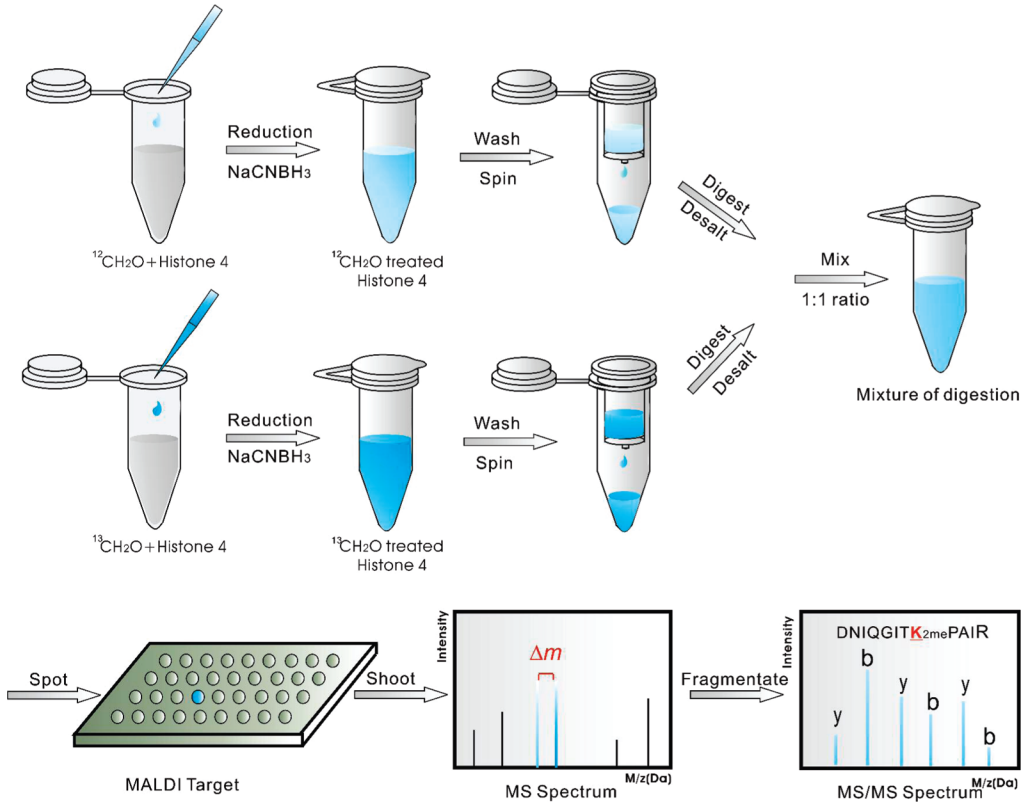
† Curriculum in Applied Science & Engineering.

‡ Department of Environmental Science and Engineering.

[1] Abbreviations: DPC, DNA−protein cross-links; H3, histone 3; H4, histone 4; H2A, histone 2A; H2B, histone 2B; PTM, post-translational modifications; acetylCoA, acetyl coenzyme A trilithium salt; HAT, histone acetylation; MWCO, molecular weight cutoff; ESI, electrospray ionization; MALDI-TOF MS, matrix-assisted laser desorption/ionization tandem time-of-flight mass spectrometry.

10.1021/tx8000576 CCC: $40.75
Published on Web 07/26/2008

Scheme 1. Analytical Approach Used in This Study



while stable isotope-labeled formaldehyde, $^{13}CH_2O$, was obtained from Cambridge Isotope Laboratories (Andover, MA). The peptides were synthesized by Genscript Corporation (Piscatway, NJ). All chemicals were used as received unless otherwise stated.

Experimental Methods. The analytical approach used in this study is illustrated in Scheme 1. First, the reaction of histone 4 with formaldehyde was performed, followed by reduction with $NaCNBH_3$. Twenty micrograms of histone 4 was dissolved in 50 μL of 10 mM potassium phosphate buffer (pH 7.2). Formaldehyde then was added to a final concentration of either 5 mM or 50 μM. After 10 min of reaction between formaldehyde and histone, reduction was performed with $NaCNBH_3$ at a final concentration of 5 mM. The reaction solution was further incubated for 3 h. Excessive formaldehyde and other small chemicals were removed using Millipore Microcon YM-3 spin columns with a 3000 molecular weight cutoff (MWCO). Two identical reactions were prepared using either native formaldehyde ($^{12}CH_2O$) or ^{13}C labeled formaldehyde ($^{13}CH_2O$). After proteolysis, equimolar amounts of $^{12}CH_2O$-treated and $^{13}CH_2O$-treated histone 4 desalted digestion solutions were mixed for further mass spectrometry analyses. In addition, the resultant histone 4 from the formaldehyde and $NaCNBH_3$ treatment was further treated separately with formaldehyde or formaldehyde plus equimolar glycine. In these reactions, the final concentration of formaldehyde or glycine varied from 5 mM to 100 mM, and the reaction time ranged from 3 h to 2 weeks.

Digestion by Enzymes. Untreated histone 4 and formaldehyde-treated histone 4 were digested with trypsin, chymotrypsin, and endoproteinase Asp-N, individually. For the trypsin digestion, 20 μg of histone was dissolved in 50 mM NH_4HCO_3 (pH 7.8), and 0.4 μg of trypsin was added. The solution was incubated for 3 h at 37 °C. For chymotrypsin cleavage, 0.4 μg of chymotrypsin was mixed with 20 μg of histone dissolved in 50 mM Tris HCl (pH 8.0) and 1 mM $CaCl_2$. The digestion mixture was incubated for 3 h at 30 °C. Endoproteinase Asp-N digestion was performed in 100 mM NH_4HCO_3 (pH 8.5) at a 1:50 enzyme to substrate ratio for 18 h at 37 °C.

Histone Acetylation Assay. Five micrograms of synthetic histone 4 N-terminal peptide (amino acid 1−23) and 500 ng of active PCAF enzyme were dissolved in 1X HAT assay buffer. AcetylCoA was added to a final concentration of 0.2 mM, and the solutions were incubated for 1 h at 30 °C with shaking. The resultant reaction mixtures were analyzed by liquid chromatography−mass spectrometry. In parallel, 10 μg of histone 4 synthetic peptide was treated with 50 mM formaldehyde at room temperature for 3 h. The formaldehyde-modified peptide was dried using a speed vacuum. Then the formaldehyde-modified peptide was incubated with active PCAF enzyme for the HAT assay, followed by mass spectrometry analysis.

Liquid Chromatography−Mass Spectrometry (LC-MS). LC-MS analyses were performed on an ion trap mass spectrometer LCQ-Deca (Thermo Electron, Waltham, MA) operating in full scan as well as dependent scan mode. Analytes were separated by reverse phase chromatography using a 250 mm × 2.5 mm analytical column from Grace Vydac (Hesperia, CA). The mobile phase consisted of 0.1% formic acid in water (solvent A) and acetonitrile (solvent B). A linear gradient was started from 5% acetonitrile to 55% in 15 min. The flow rate of the mobile phase was set as 200 μL/min. An electrospray ionization (ESI) source was used to analyze peptide samples. For the fragmentation of precursor ions, the normalized collision energy varied from 30% to 35%, depending on the structures of peptides. The activation time was set at 30 ms.

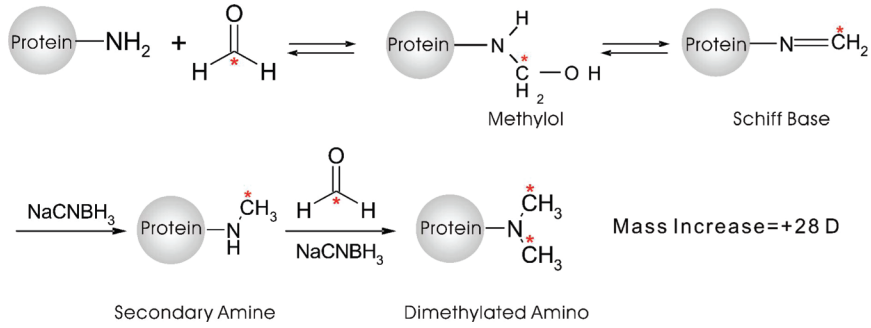
Matrix-Assisted Laser Desorption/Ionization Tandem Time-of-Flight Mass Spectrometry (MALDI-TOF MS). MALDI-TOF mass spectrometric analyses were performed on a 4700 Proteomics Analyzer (Applied Biosystems, Framingham, MA) operating in reflectron mode. Most of the tandem mass spectrometric experiments were performed on the MALDI-TOF/

Table 1. Peptide Fragments of Histone 4 Observed by Mass Spectrometry

no.	peptide fragments	calculated $(M + H)^+$ (Da)	measured $(M + H)^+$ (Da)	$\Delta M/M$ (ppm)
1[a]	$_{96}$TLYGFGG$_{102}$	714.34	714.35	19.4
2	$_{60}$VFLENVIR$_{67}$	989.57	989.58	10.3
3	$_{46}$ISGLIYEETR$_{55}$	1180.62	1180.66	36.2
4	$_{68}$DAVTYTEHAK$_{77}$	1134.54	1134.55	9.07
5	$_{80}$TVTAMDVVYALK$_{91}$	1310.70	1310.71	6.71
6	$_{24}$DNIQGITKPAIR$_{35}$	1325.75	1325.77	14.5
7[b]	$_{91}$KRQGRTLV$_{98}$	1021.59	1021.63	40.5
8	$_{89}$ALKRQGRTL$_{97}$	1042.64	1042.69	42.4
9	$_{50}$IYEETRGVL$_{58}$	1079.57	1079.61	42.7
10	$_{89}$ALKRQGRTLY$_{98}$	1205.71	1205.76	43.7
11	$_{38}$ARRGGVKRISGL$_{49}$	1269.78	1269.84	45.9
12	$_{62}$LENVIRDAVTY$_{72}$	1292.68	1292.74	46.4
13	$_{38}$ARRGGVKRISGLIY$_{51}$	1545.93	1546.01	53.4
14	$_{59}$KVFLENVIRDAVTY$_{72}$	1666.91	1667.01	57.8
15	$_{23}$RDNIQGITKPAIRRL$_{37}$	1751.03	1751.12	48.9
16	$_{73}$TEHAKRKTVTAMDVVY$_{88}$	1848.96	1849.07	59.7
17	$_{38}$ARRGGVKRISGLIYEETRG$_{58}$	2330.34	2330.50	71.1
18[c]	$_{68}$DAVTYTEHAKRKTVTAM$_{84}$	1921.98	1922.03	28.7
19	$_{85}$DVVYALKRQGRTLYGFGG$_{102}$	2000.07	2000.11	19.7
20	$_{1}$SGRGKGGKGLGKGGAKRHRKVLR$_{23}$	2360.43	2360.43	15.5
21[d]	$_{1}$S$_{ac}$GRGKGGKGLGKGGAKRHRK$_{2me}$VLR$_{23}$	2430.47	2430.54	28.8
22	$_{1}$S$_{ac}$GRGKGGKGLGKGGAK$_{ac}$RHRK$_{2me}$VLR$_{23}$	2472.48	2472.55	28.3

[a] Peptides 1−6 were obtained after trypsin digestion. [b] Peptides 7−17 were yielded by chymotrypsin cleavage. [c] Peptides 18−22 were observed after Asp-N proteolysis. [d] Peptides 21−22 were obtained from histone 4 with PTM.

Scheme 2. Formation of a Schiff Base and Its Reduction to a Dimethylated Group by $NaCNBH_3$



TOF mass spectrometer using air as the collision gas at a medium pressure setting (4e-007 torr) and a laser intensity of 5400 ABI units (Nd:YAG laser, 355 nm wavelength, 3−7 ns pulse, >12 μJ pulse energy.). The matrix applied for peptide analysis was α-cyano-4-hydroxy-cinnamic acid. The MS-Digest or MS-Fit program (Protein Prospector, University of California at San Francisco) was used to automatically assign the peptides. The MS/MS fragment ion spectra were manually matched to the predicted peptide fragmentation generated by Data Explorer software.

Results

Characterization of Histone 4. Two different types of histone 4 were used in this study. Histone 4 with PTM was isolated from calf thymus tissues, while unmodified human recombinant H4 was purified after expression in *E. coli* cells. The sequences of the both proteins were identical. Prior to the formaldehyde reaction, the purity and molecular weight of histone 4 were examined by LC-MS. Histone 4 was digested with trypsin and chymotrypsin to improve the coverage for peptide mapping, yielding 6 and 11 detectable fragments for trypsin and chymotrypsin proteolysis, respectively, as listed in Table 1.

By combining the peptides from independent trypsin and chymotrypsin digestion, all the residues located in the histone fold domain could be covered. However, neither trypsin nor chymotrypsin was able to provide any information about the histone N-terminal tail. To obtain this data, endoproteinase Asp-N was utilized to produce N-terminal peptides. This protease cleaves proteins at the amino side of aspartic acid. In theory, this will result in 4 nonoverlapping peptides in complete digestion. The largest predicted fragment at 5004.91 Da is over the cutoff of the MALDI-TOF mass spectrometer operating at reflectron mode; therefore, only 3 peptides were observed, as listed in Table 1. In addition, after the Asp-N digestion of histone 4 with PTM, two abundant peptides with PTM were detected, as shown in Table 1.

Evaluation of the Accessibility of Lysine Residues. The reaction between formaldehyde and amine involves a nucleophilic attack of amine on the carbonyl group of formaldehyde, followed by rapid proton transfer resulting in methylol groups. Subsequently, labile Schiff bases are produced after dehydration of methylol. These two reactions are both reversible; thus, they can not be easily detected. Therefore, a reduction approach with sodium cyanoborohydride was chosen to reduce the Schiff base to a detectable dimethylamino structure. The reaction mechanism is shown in Scheme 2. First, a Schiff base formed from formaldehyde attack is quickly reduced to a secondary amine, which is relatively more reactive than a primary amine. The secondary amine then reacts with another formaldehyde molecule and is further reduced to form a dimethylamino group (*15*).

In addition to $^{12}CH_2O$, stable isotope-labeled formaldehyde, $^{13}CH_2O$, was used to treat histone 4. Mixing equal ratios of

Table 2. Formaldehyde-Modified Histone 4 Peptide Fragments Observed by MALDI-TOF Mass Spectrometry[a]

no.	observed doublets		sequence	Δ*M* (Da)	Δ*m* (Da)	*n*
1[b]	1353.85	1355.86	$_{24}$DNIQGIT**K**$_{2me}$PAIR$_{35}$	+28	2	1
2	1414.96	1416.96	$_{56}$GVL**K**$_{2me}$VFLENVIR$_{67}$	+28	2	1
3	1651.07	1655.08	$_{79}$**K**$_{2me}$TVTAMDVVYAL**K**$_{2me}$R$_{92}$	+56	4	2
4[c]	1297.87	1299.87	$_{38}$ARRGGV**K**$_{2me}$RISGL$_{49}$	+28	2	1
5	1905.10	1909.11	$_{73}$TEHA**K**$_{2me}$R**K**$_{2me}$TVTAMDVVY$_{88}$	+56	4	2
6[d]	2528.83	2540.86	$_{1}$S$_{2me}$GRG**K**$_{2me}$GG**K**$_{2me}$GLG**K**$_{2me}$GGA**K**$_{2me}$RHR**K**$_{2me}$VLR$_{23}$	+168	12	6

[a] ΔM, mass increase of peptide upon formaldehyde treatment; Δ*m*, mass difference between the doublets; *n*, the number of reactive lysine residue and N-terminal amino group. [b] Peptides 1−3 were obtained after trypsin digestion. [c] Peptides 4−5 were yielded by chymotrypsin cleavage. [d] Peptide 6 were observed after Asp-N proteolysis.

$^{12}CH_2O$- and $^{13}CH_2O$- peptides results in the formation of doublets in a mass spectrum. The mass difference between a doublet allows us to calculate how many formaldehyde molecules were incorporated into each peptide. For instance, if the mass difference between the peptide pair equals 2, this indicates that 2 formaldehyde units reacted with the residue, and consequently, 1 lysine residue is reactive with formaldehyde. Thus, the general formula is $n = \Delta m/2$, where n is the number of reactive lysines, and Δm corresponds to the mass difference between the doublet.

The accessibility of each lysine residue by formaldehyde was confirmed by either MALDI-TOF or LC-MS. A good match of mass increase after formaldehyde treatment or MS/MS fragmentation pattern of formaldehyde-modified peptides permitted the identification of the exact reactive sites with high confidence. The formaldehyde-modified histones were digested with 3 different enzymes, and the resulting peptides were analyzed by MALDI-TOF mass spectrometry to determine the mass increases of individual peptide sequences upon formaldehyde treatment (Table 2).

There were 5 detectable tryptic peptides by MALDI-TOF mass spectrometry, with three of them having the corresponding doublets when ^{13}C labeled formaldehyde was used, as listed in Table 2. Figure 1A shows a typical doublet ($\Delta m = 2$ Da) arising from the application of $^{12}CH_2O$ and $^{13}CH_2O$. It is very straightforward to assign the peptide at 1353.85 Da as the dimethylated product of original tryptic peptide at 1325.77 Da ($_{24}$DNIQGITKPAIR$_{35}$) after observing the 28 Da mass increase and a doublet separated by 2 Da. The specific sequence and exact reaction sites were confirmed by MS/MS sequencing with MALDI TOF/TOF, and the MS/MS fragmentation pattern of the precursor ion of 1353.82 Da is given in Figure 1B.

According to the doublets resulting from ^{13}C labeled formaldehyde, the fragments at 1414.96 and 1651.07 Da were identified as having 1 and 2 lysine residues accessible to formaldehyde. Their structures were determined to be $_{56}$GVL**K**$_{2me}$VFLENVIR$_{67}$ and $_{79}$**K**$_{2me}$TVTAMDVVYAL**K**$_{2me}$R$_{92}$, respectively. It is interesting to note that two expected tryptic peptides at 989.57 and 1310.70 Da disappeared after the reaction with formaldehyde. However, no direct formaldehyde-induced precursor ions could be attributed to them. Increased cleavage resistance from the dimethylation at K59 and K91 after $NaCNBH_3$ reduction could be responsible for their absence.

Two unique fragments at 1297.87 and 1905.10 Da (shown in Figure 1C) were identified as $_{38}$ARRGGV**K**$_{2me}$RISGL$_{49}$ and $_{73}$TEHA**K**$_{2me}$R**K**$_{2me}$TVTAMDVVY$_{88}$ after chymotrypsin proteolysis, suggesting that K44 and K77 are two additional residues in the fold domain that can be attacked by formaldehyde.

The accessibility of lysine residues along the N-terminal tail was evaluated after endoproteinase Asp-N cleavage, which liberated the N-terminal peptide consisting of the first 23 amino acids. According to the mass difference ($\Delta m = 12$ Da) between the resulting doublet, 12 formaldehyde units were incorporated into the 5 lysine residues and the N-terminus, resulting in 168 Da mass increases ($\Delta M = 168$ Da) compared with the untreated N-terminal peptide (shown in Figure 1D). Therefore, all the lysine residues located in the globular fold domain and the N-terminal tail coupled with the end amino group of the N-terminus are accessible to formaldehyde, providing a total of 12 potential reactive sites. The reactive sites identified by mass spectrometry are visualized in a structural model, as shown in Figure 1E.

In addition to the high concentration formaldehyde used in this study, a physiological level of formaldehyde (50 μM) was also used to treat histone 4. The binding sites we identified were consistent with those found based on the experiment using the higher formaldehyde treatment (5 mM) (data not shown here). The concentration of formaldehyde only influenced the percentage of modified histone, as shown in Figure 2. Figure 2A shows the molecular weight of unmodified histone 4 (11236 Da), which was determined after deconvoluting the peaks from ESI-MS. Figure 2B gives the deconvoluted spectrum for the 5 mM formaldehyde-treated histone sample. The most intense peak was attributed to modified histone, and its molecular weight was determined to be 11574 Da. The mass increase (about 338 Da) resulted from the formation of dimethyl groups on 11 lysine residues as well as the N-terminal amino group. Figure 2C shows the deconvoluted mass spectrum of histone treated with 50 μM formaldehyde. Two major peaks were found and assigned as unmodified histone 4 (11236 Da) and formaldehyde-modified histone 4 (11574 Da). Compared with Figure 2B, the percentage of modified histone is much lower than that of the higher formaldehyde-treated sample.

In theory, other amino acids including arginine, glutamine, tyrosine, histidine, cysteine, and asparagines could also potentially react with formaldehyde (*14*). However, neither the formaldehyde nor the formaldehyde-glycine treatment described in the experimental section yielded any detectable modifications on these residues using the current analytical approach. Remarkably, we found that a 1 h reaction at 37 °C with 5 mM formaldehyde can induce obvious intermolecular cross-links, but no such cross-links could be observed even with a 2 week, 100 mM formaldehyde treatment once these lysine residues of histone were blocked by dimethylation (data not shown). These results may further highlight the higher reactivity and importance of lysine in forming intermolecular cross-links.

Influence of PTM on the Reaction. Prior to formaldehyde treatment, H4 with PTM was cleaved by endoproteinase Asp-N, and the resultant abundant N-terminal peptides are shown in Figure 3 (control panel). The peptide fragment at 2430.54 Da was assigned to the peptide $_{1}$SGRGKGGKGLGKGGAKRHRKVLR$_{23}$ with dimethylation on K20 and one acetylation on the N-terminus, while the peptide observed at 2472.55 Da had an additional acetylation on lysine, resulting in 42 Da mass increases compared with that of the fragment at 2430.54

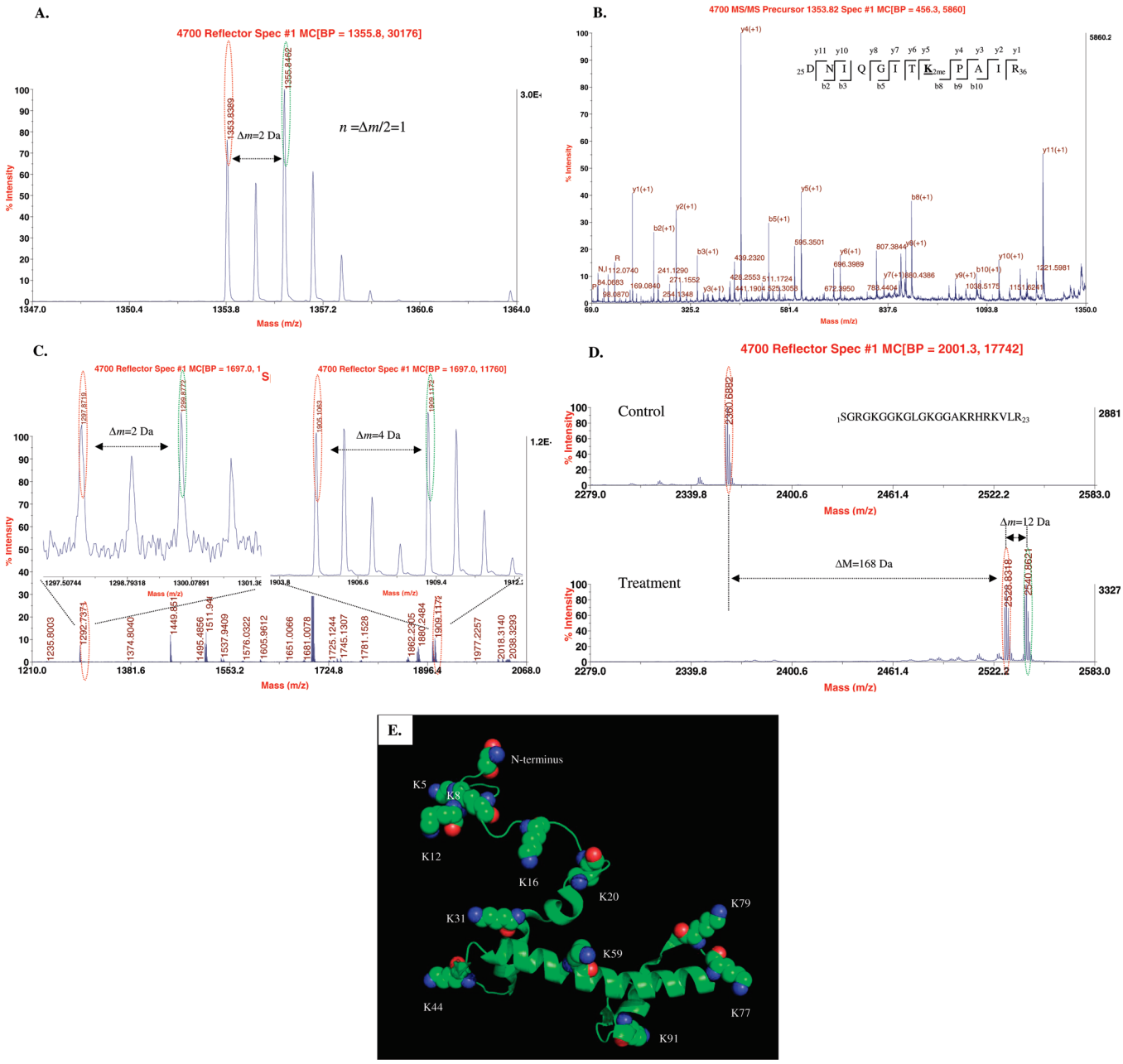


Figure 1. Mass spectra of formaldehyde-modified histone 4 peptides. (A) The formation of a doublet separated by 2 Da after native formaldehyde and ^{13}C labeled formaldehyde treatment. (B) MALDI-TOF/TOF MS/MS spectrum of the precursor ion at 1353.82 Da with the identified structure as $_{24}$DNIQGIT**K**$_{2me}$PAIR$_{35}$. (C) MALDI-TOF MS spectrum of two doublets, 1297.87−1299.87 and 1905.10−1909.11, obtained after chymotrypsin proteolysis. (D) MALDI-TOF mass spectrum of unmodified N-terminal peptide observed at 2360.68 Da after Asp-N cleavage (control panel) and formaldehyde-modified N-terminal doublet separated by 12 Da after Asp-N digestion (treatment panel). (E) The representation of reactive sites including lysine and the N-terminus of histone 4. The potential reactive residues are shown with a sphere model, constructed on the basis of a 1.9 Å crystal structure (PDB Code: 1KX5) and rendered with PyMOL (DeLano Scientific LLC, Palo Alto, CA).

Da. These assignments are consistent with previous identifications of histone acetylation isomers (*16*).

Figure 3 (treatment panel) shows the mass spectrum of formaldehyde-treated histone 4 after Asp-N digestion. Two relatively abundant fragments were observed at 2542.71 and 2556.68 Da corresponding to the formaldehyde-modified products of untreated peptides at 2430.54 and 2472.55 Da (ΔM equals 112 and 84 Da, respectively). Although unmodified histone offers 5 lysine residues and the N-terminus as potential reactive sites along the N-terminal tail, the number of reactive lysine residues was determined to be 4 and 3 for the fragments at 2542.71 and 2556.68 Da, respectively. The observation that only the lysine residues void of PTM could be attacked by formaldehyde clearly shows that PTM inhibits the reaction between formaldehyde and lysine.

Impact of Formaldehyde-Induced Modifications on PTM Formation. Figure 4A shows the ESI spectrum of a synthetic H4 N-terminal peptide, which was used as the substrate of histone acetyltransferases for the histone acetylation assay. The precursor ion at 1181.45 *m*/*z* is the doubly charged form of H4 N-terminal peptide. Figure 4B shows the ESI mass spectrum of histone acetyltransferase (PCAF)-treated H4 N-terminal peptide, which was obtained after incubating H4 N-terminal peptide with PCAF enzyme in the presence of acetylCoA for 1 h at 30 °C. The doubly charged ions at 1202.24 *m*/*z* and 1223.84 *m*/*z* are monoacetylated (*m*/*z*

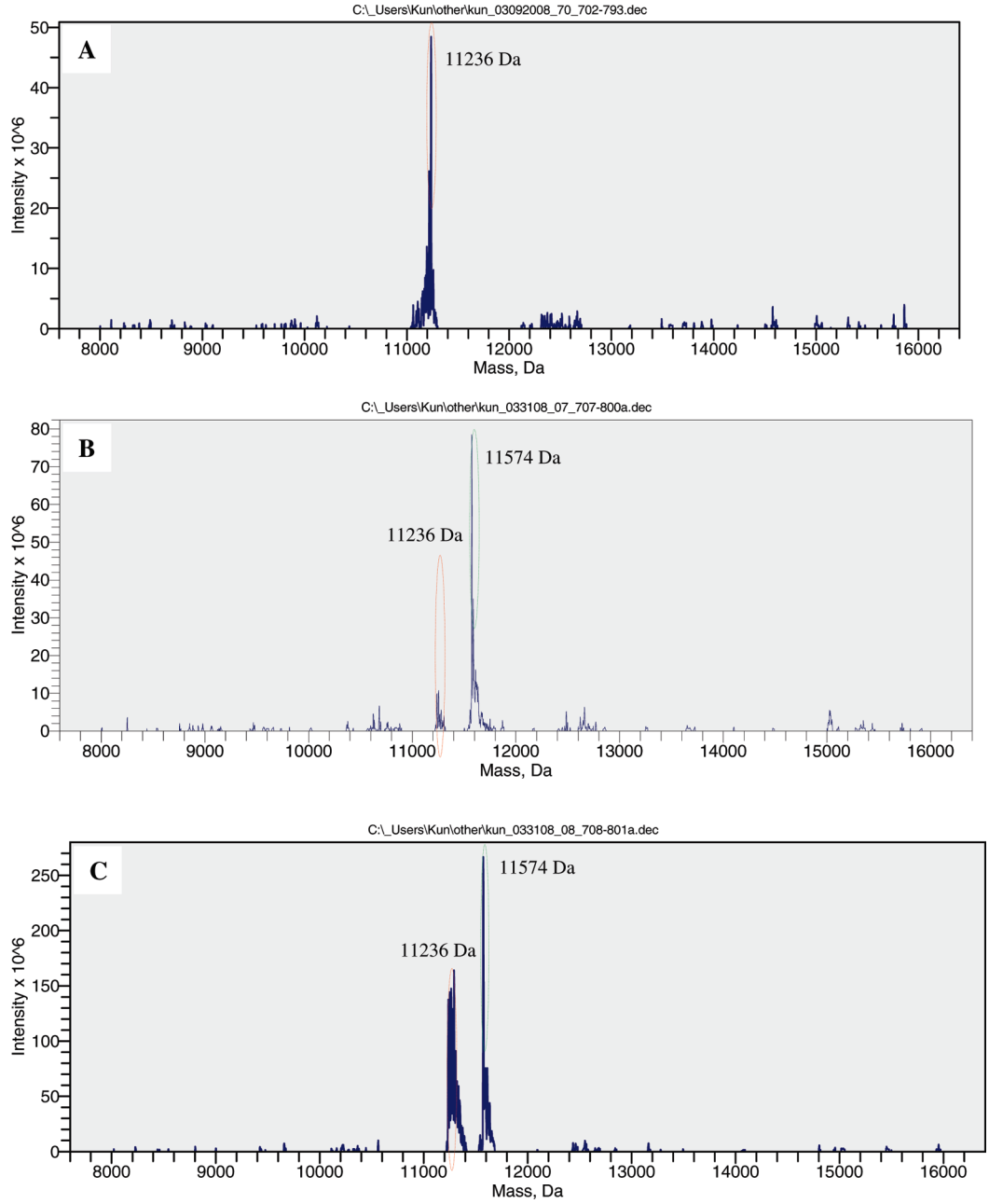


Figure 2. The molecular weight of untreated histone 4 and formaldehyde-treated histone 4. (A) The mass spectrum of histone 4 after deconvoluting major peaks from ESI-MS, showing the molecular weight of histone 4 equal to 11236 Da. (B) The deconvoluted mass spectrum for the 5 mM formaldehyde-treated histone 4 sample. (C) The deconvoluted mass spectrum for the 50 μM formaldehyde-treated histone 4 sample.

increase = 21 Da) and diacetylated (*m*/*z* increase = 42 Da) H4 N-terminal peptides.

Figure 4C shows the ESI mass spectrum of a formaldehyde-modified H4 N-terminal tail. The *m*/*z* shift is about 36 Da for a doubly charged ion, which is attributed to the formation of 6 Schiff bases on lysine residues as well as the N-terminus. Figure 4D shows the ESI mass spectrum of a formaldehyde-modified H4 N-terminal peptide after a histone acetyltransferase treatment. It was generated by incubating formaldehyde-modified H4 N-terminal peptide with PCAF enzyme in the presence of acetylCoA for 1 h at 30 °C. Compared with the formaldehyde-modified H4 N-terminal peptide shown in Figure 3C, no mass increase upon further histone acetyltransferase treatment could be observed, which clearly shows that formaldehyde-induced Schiff bases inhibit the formation of PTM *in vitro*.

Discussion

One of the goals of this study was to evaluate the reactivity of histone residues toward formaldehyde using mass spectrometry. We have unambiguously demonstrated that all the lysine residues located in both the globular fold domain and the unstructured N-terminal tail are accessible to formaldehyde. We also found that PTM on histone residues inhibited the attack of formaldehyde. Additionally, we demonstrated that formaldehyde-induced lysine modifications inhibited the formation of PTM *in vitro*.

One question may be raised by the observation that PTM on residues inhibits the reaction between formaldehyde and lysine residues. Histones are subject to a variety of post-translational modifications, including methylation, acetylation, and phosphorylation. The accessibility of DNA for many important biological events such as transcription, replication, recombination, and

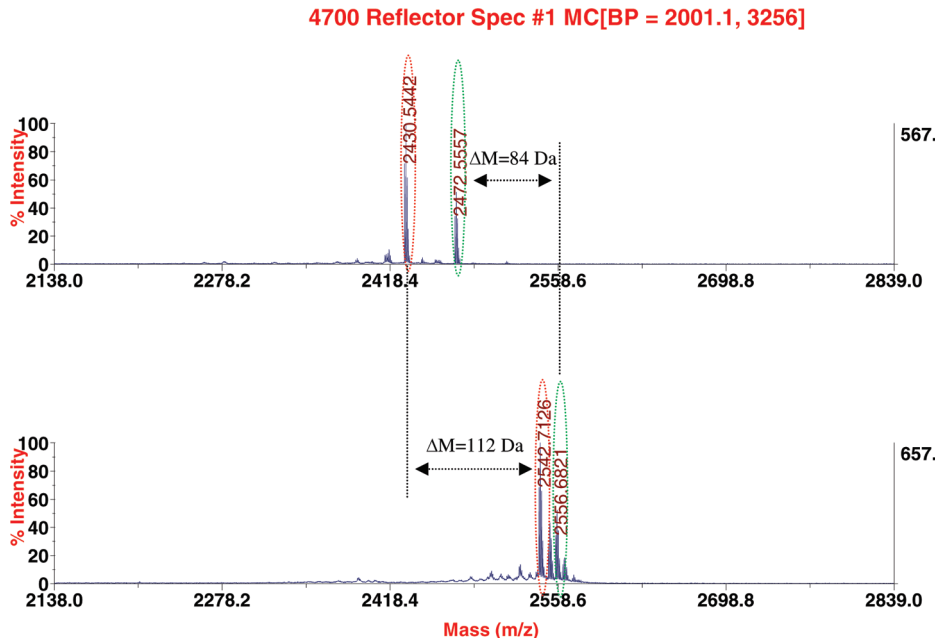


Figure 3. MALDI-TOF MS spectrum of two abundant histone 4 N-terminal peptides with PTM observed at 2430.54 and 2472.55 Da after Asp-N cleavage prior to formaldehyde treatment (control panel) and MS spectrum of two formaldehyde-modified N-terminal peptides observed at 2542.71 and 2556.68 Da after Asp-N digestion (treatment panel).

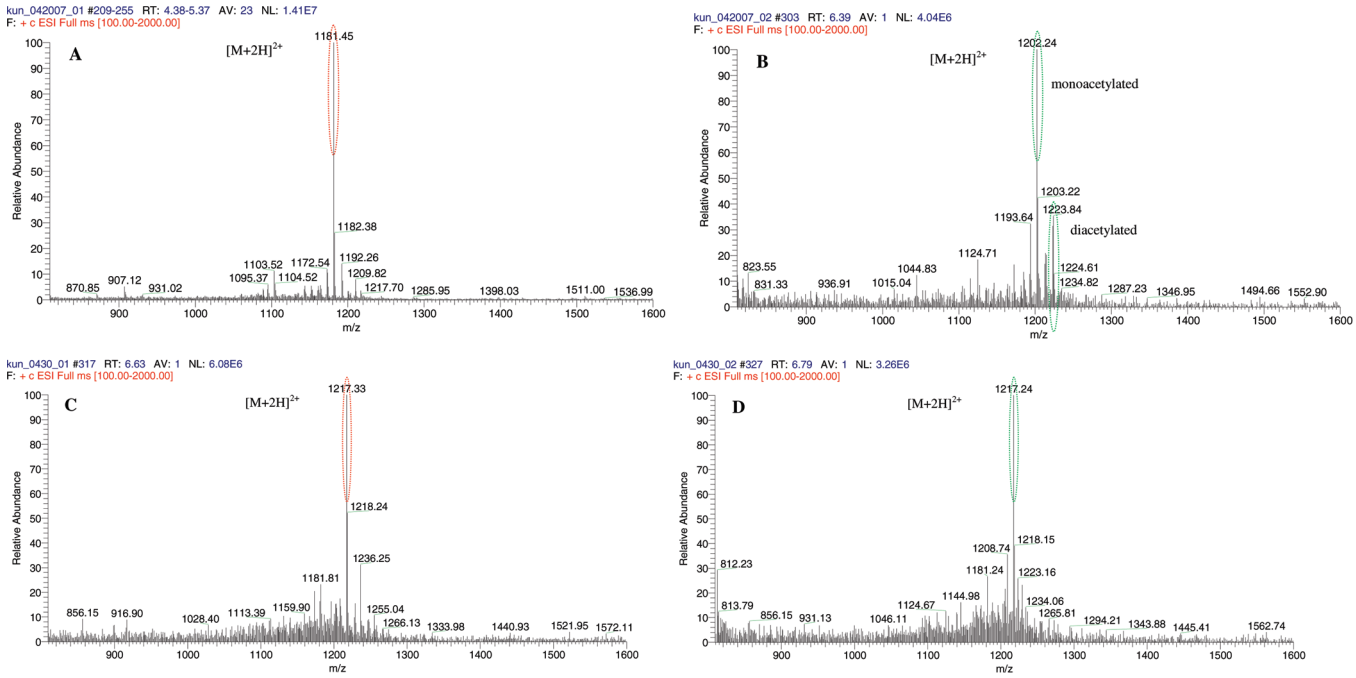


Figure 4. (A) ESI mass spectrum of a synthetic histone 4 N-terminal peptide (amino acid 1−23). A doubly charged ion is shown in the spectrum. (B) ESI mass spectrum of the PCAF enzyme-treated histone 4 N-terminal peptide. The fragments at 1202.24 *m*/*z* and 1223.84 *m*/*z* are monoacetylated and diacetylated histone 4 N-terminal peptides. (C) ESI mass spectrum of the formaldehyde-modified histone 4 N-terminal peptide. The 36 *m*/*z* increase is attributed to the formation of 6 Schiff bases on lysine and the N-terminus. (D) ESI mass spectrum of the formaldehyde-modified histone 4 N-terminal peptide treated by PCAF enzyme. The precursor ion at 1217.24 *m*/*z* presents the same mass with the formaldehyde-modified histone 4 N-terminal peptide.

repair is regulated by these modifications through forming either open or condensed chromatin (*17*). As far as the histone 4 tail is concerned, K5, K8, K12, K16, and the end amino group of S1 can be acetylated, whereas methylation occurs on K20, which can be mono-, di-, or trimethylated on the side chain (*16*). Are the lysine residues still accessible to formaldehyde considering the widespread occurrence of PTM on histone? It is well known that histone acetylation is catalyzed by histone acetyltransferases and that histone deacetylation is realized by histone deacetylases, making this process completely reversible (*18*). Lysine methylation was once thought to be an irreversible process; however, enzymes identified recently are capable of demethylating histone at specific sites (*19*). Although lysine residues are protected by PTM, formaldehyde could still attack lysine residues provided the PTM is removed. Our finding highlights the importance of less modified states of histone or nascent histone, which usually is only diacetylated in the cytoplasm at K5 and K12 by B-type histone acetyltransferase (*16*).

The second question is how the modifications caused by formaldehyde evolve. There are three different structures for such modifications. The reaction between formaldehyde and lysine includes the quick formation of methylol groups, followed

by the formation of Schiff bases from partial dehydration of methylol groups. The final step may yield intramolecular or intermolecular cross-links, depending on the local physical and chemical environment of reactive groups. Although methylol and Schiff bases are involved in reversible reactions, the existence of all three structures was confirmed in previous research using mass spectrometry or NMR (*14*). Therefore, these distinct structures have potential biological impact. As we have shown, Schiff bases induced by formaldehyde inhibit the formation of PTM on lysine *in vitro*. Clearly, additional experiments need to be done in order to track the exact structure occurring on individual lysine molecules and the corresponding biological influence *in vivo*.

Although previous studies have demonstrated the genotoxic and mutagenic effects of formaldehyde, the fact that all lysine residues are the targets of formaldehyde raises the possibility of an additional mode of action for its toxicity. Formaldehyde could alter epigenetic regulation in which histone modifications occurring on lysine play a central role. We have clearly shown that formaldehyde-induced Schiff bases inhibit the formation of PTM on lysine. Therefore, these formaldehyde-induced lysine adducts on histone may impair the PTM pattern and possibly disturb the subsequent recruitment of specific proteins highly associated with the PTM pattern, triggering a series of abnormal cascade effects. Additionally, the balance between histone acetylation and deacetylation could be disturbed by the attachment of formaldehyde on lysine residues. The balance between histone acetylation and deacetylation is important for normal cell growth, and an imbalance of acetylation in promoter regions may induce the deregulation of gene expression. Accumulating evidence has linked imbalances between acetylation and deacetylation to carcinogenesis and cancer progression (*20–22*). However, our finding that formaldehyde-induced Schiff bases inhibit the formation of PTM is based on a simplified *in vitro* model. Furthermore, the Schiff bases are generally reversible in nature. Therefore, additional experiments in cells or tissues will be needed to demonstrate that such effects occur in biology.

In conclusion, identification of reactive sites on histone is an initial step in understanding the mechanisms of formaldehyde toxicity and carcinogenicity, with many questions remaining to be elucidated. What is the structure and fate of formaldehyde-induced lysine adducts? What is the biological impact of formaldehyde-induced modification on each lysine residue? Where are the exact sites for the formation of DPC? What are the toxicological mechanisms of DPC? How are DPC repaired? Further *in vitro* and *in vivo* experiments should be carried out in order to shed light on these intriguing questions.

Acknowledgment. We thank Dr. Carol Parker and Nely Dichev from the Michael Hooker Proteomics Center at University of North Carolina at Chapel Hill for their help with the peptide analyses. This work was supported in part by NIH grant P30-ES10126 and a grant from the Formaldehyde Council, Inc.

References

(1) International Agency for Research on Cancer. (1995) Wood Dust and Formaldehyde. *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. 62*, 217−245.
(2) Speit, G., Schutz, P., and Merk, O. (2000) Induction and repair of formaldehyde-induced DNA-protein crosslinks in repair-deficient human cell lines. *Mutagenesis 15*, 85–90.
(3) Craft, T. R., Bermudez, E., and Skopek, T. R. (1987) Formaldehyde mutagenesis and formation of DNA-protein crosslinks in human lymphoblasts in vitro. *Mutat. Res. 176*, 147–155.
(4) Merk, O., and Speit, G. (1998) Significance of formaldehyde-induced DNA-protein crosslinks for mutagenesis. *Environ. Mol. Mutagen 32*, 260–268.
(5) Shaham, J., Bomstein,Y., Meltzer, A., Kaufman, Z., Palma, E. and Ribak, J. (1996) DNA-protein crosslinks, a biomarker of exposure to formaldehyde—*in vitro* and *in vivo* studies. *Carcinogenesis 17*, 121−125.
(6) Shaham, J., Bomstein, Y., Gurvich, R., Rashkovsky, M., and Kaufman, Z. (2003) DNA-protein crosslinks and p53 protein expression in relation to occupational exposure to formaldehyde. *Occup. Environ. Med. 60*, 403–409.
(7) Quievryn, G., and Zhitkovich, A. (2000) Loss of DNA-protein crosslinks from formaldehyde-exposed cells occurs through spontaneous hydrolysis and an active repair process linked to proteosome function. *Carcinogenesis 21*, 1573–1580.
(8) Solomon, M. J., and Varshavsky, A. (1985) Formaldehyde-mediated DNA-protein cross linking: A probe for in vivo chromatin structures. *Proc. Natl. Acca. Sci. U.S.A. 82*, 6470–6474.
(9) Siomin, Y. A., Simonov, V. V., and Poverenny, A. M. (1973) The reaction of formaldehyde with deoxynucleotides and DNA in the presence of amino acids and lysine-rich histone. *Biochim. Biophys. Acta 331*, 27–32.
(10) Jackson, V. (1999) Formaldehyde crosslinking for studying nucleosomal dynamics. *Methods 17*, 125–139.
(11) Davery, C. A., Sargent, D. F., Lunger, 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.
(12) Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature 389*, 251–260.
(13) Santos-Rosa, H., and Caldas, C. (2005) Chromatin modifier enzymes, the histone code and cancer. *Eur. J. Cancer 41*, 2381–2402.
(14) Metz, B., Kersten, G. F. A., Hoogerhout, P., Brugghe, H. F., Timmermans, H. A. M., de Jong, A., Meiring, H., Hove, J. T., Hennink, W. E., Crommelin, D. J. A., and JisKoot, W. (2004) Identification of formaldehyde-induced modifications in proteins: reaction with model peptides. *J. Biol. Chem. 279*, 6235–6243.
(15) Means, G. E., and Feeney, R. E. (1995) Reductive alkylation of protein. *Anal. Biochem. 224*, 1–16.
(16) Zhang, K., Williams, K. E., Huang, L., Yau, P., Siino, J. S., Bradbury, E. M., Jones, P. R., Minch, M. J., and Burlingame, A. L. (2002) Histone acetylation and deacetylation: identification of acetylation and methylation sites of Hela histone 4 by mass spectrometry. *Mol. Cell. Proteomics 1*, 500–508.
(17) Cosgrove, M. S., and Wolberger, C. (2005) How does the histone code work? *Biochem. Cell Biol. 83*, 468–476.
(18) Peterson, C. L., and Laniel, M. A. (2004) Histones and histone modifications. *Curr. Biol. 14*, R546–551.
(19) Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M. E., Borchers, C. H., Tempst, P., and Zhang, Y. (2006) Histone demethylation by a family of JmjC domain-containing proteins. *Nature 439*, 811–816.
(20) Santos-Reboucas, C. B., and Pimentel, M. M. (2007) Implication of abnormal epigenetic patterns for human diseases. *Eur. J. Hum. Genet. 15*, 10–17.
(21) Ducasse, M., and Brown, M. A. (2006) Epigentic aberrations and cancer. *Mol. Cancer 5*, 60–70.
(22) Agalioti, T., Chen, G., and Thanos, D. (2002) Deciphering the transcriptional histone acetylation code for a human gene. *Cell 111*, 381–392.

TX8000576