

N^ε-Formylation of lysine is a widespread post-translational modification of nuclear proteins occurring at residues involved in regulation of chromatin function

Jacek R. Wiśniewski*, Alexandre Zougman and Matthias Mann

Department of Proteomics and Signal Transduction, Max-Planck Institute for Biochemistry, Am Klopferspitz 18, D-82152 Martinsried, Germany

Received July 12, 2007; Revised November 5, 2007; Accepted November 9, 2007

ABSTRACT

Post-translational modification of histones and other chromosomal proteins regulates chromatin conformation and gene activity. Methylation and acetylation of lysyl residues are among the most frequently described modifications in these proteins. Whereas these modifications have been studied in detail, very little is known about a recently discovered chemical modification, the N^ε-lysine formylation, in histones and other nuclear proteins. Here we mapped, for the first time, the sites of lysine formylation in histones and several other nuclear proteins. We found that core and linker histones are formylated at multiple lysyl residues located both in the tails and globular domains of histones. In core histones, formylation was found at lysyl residues known to be involved in organization of nucleosomal particles that are frequently acetylated and methylated. In linker histones and high mobility group proteins, multiple formylation sites were mapped to residues with important role in DNA binding. N^ε-lysine formylation in chromosomal proteins is relatively abundant, suggesting that it may interfere with epigenetic mechanisms governing chromatin function, which could lead to deregulation of the cell and disease.

Mapping of post-translational modifications of histones has been an object of many recent studies (1–5). In depth identification of the modifications, their dynamics and comparative analyses of organismal differences, are important for understanding mechanisms of chromatin regulation. Proteins involved in DNA binding, regulation of chromatin condensation and DNA repair, appear to be

modulated by sequential and cooperative interplay of a variety post-translational modifications including acetylation, methylation and phosphorylation (6,7). In addition to core histones (H2A, H2B, H3 and H4), proteins such as linker histone H1, high mobility group (HMG) and other chromosomal proteins can also influence the regulation of chromatin function. We and others recently showed that linker histones carry post-translational modification features similar to core histones (8,9). Analyses of post-translational modifications of mammalian linker histones revealed multiple sites of acetylation, methylation, phosphorylation and formylation. The latter modification was observed at residues that are known to be primarily involved in DNA binding and that were also found to be acetylated or monomethylated. As the observed modifications were restricted to only a few sites in H1, occurred in different cell culture models and tissues and with different extraction procedures, we suggested that formylation reflects a specific post-translational modification and is not an artifact resulting from extraction and analysis procedures (9).

Supporting this argument, more recently, Jiang *et al.* (10) have demonstrated that lysine formylation can arise from oxidative damage of DNA. They showed that 3'formylphosphate generated by oxidation of deoxyribose reacts with the N^ε-moiety of lysine. Due to the spatial proximity to DNA, histones are the proteins primarily affected by this modification. The frequency of formylation in cultured human lymphoblastoid cells TK-6 has been estimated to be 0.1% of all lysines in acid soluble chromatin proteins. Furthermore, in histones this value can increase several-fold upon oxidative stress (10). Thus, formylated lysines can reach the abundance levels of more prominent histone modifications. Due to the chemical similarity of N^ε-acetylation and N^ε-methylation to N^ε-formylation (Figure 1A), this modification may have functional consequences for histones.

*To whom correspondence should be addressed. Tel: +49 89 8578 2205; Fax: +49 89 8578 2219; Email: jwisniew@biochem.mpg.de

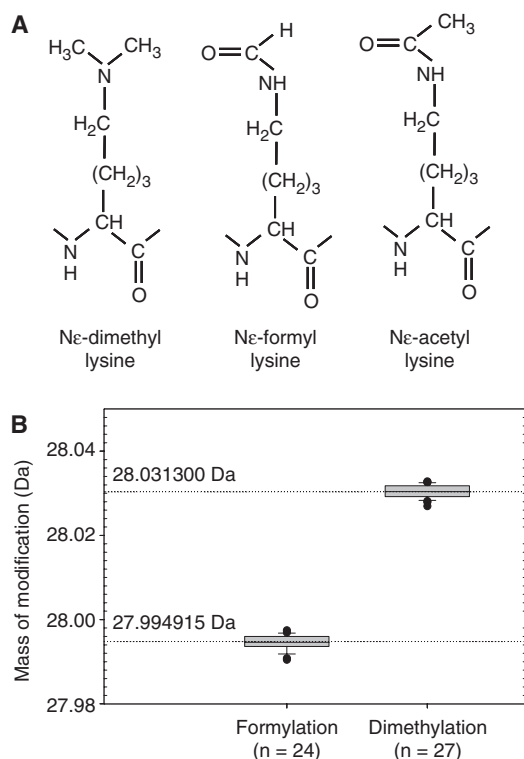


Figure 1. (A) Acetylated, dimethylated and formylated lysyl residues show similar structures. (B) Complete mass separation of the otherwise identical dimethylated and formylated peptides. Boxplot of mass deviations from theoretical masses for formylation and dimethylation in histone H3 measured using the LTQ-Orbitrap instrument. The box represents the interquartile range of 25–75%, the center line represents median, and ‘whiskers’ (small horizontal lines) extend to the most extreme ratio up to 1.5-fold interquartile range. The mass differences of dimethylated or formylated peptides are 28.0313 and 27.9949 Da, respectively, versus unmodified peptide. Mass deviations for dimethylation are taken from Supplementary Table S2.

High-resolution mass spectrometry has become an indispensable tool in elucidation of known and novel modification sites (11). In particular, the mass accuracy achievable with modern instruments such as the linear ion trap orbitrap mass spectrometer is at the low p.p.m. or sub-p.p.m. level. This accuracy makes it possible to unambiguously distinguish modifications that lead to the same nominal mass shift. We have recently used these mass spectrometric capabilities to investigate modifications in human linker histone H1. Among several other modifications, H1 proved to be formylated and we mapped the sites of this modification (9). In this study, we describe sites of formylation in histones and other nuclear proteins isolated by different extraction methods from a variety of cultured cell types as well as from human and mouse tissues. Our data show that the lysyl residues affected by this modification are also frequently found to be acetylated and/or methylated.

MATERIALS AND METHODS

Tissue

Human MCF7, HeLa and A549 cells were cultured in DMEM supplemented with 10% FBS and 1%

penicillin/streptomycin. Cells were washed with PBS, pH 7.5 and stored at -80°C .

Extraction of nuclear proteins from mouse liver

Livers were dissected from mice of the inbred strain C57BL/6 that were sacrificed by decapitation. Nuclei were isolated according to the method of Blobel and Potter (12) with some modifications. Briefly, one liver was homogenized in 10 ml of 0.25 M sucrose in buffer A: 10 mM Mops/NaOH, pH 7.0, 10 mM NaCl, 5 mM MgCl_2 , 1 mM CaCl_2 , 1 mM DTT containing Complete Protease Inhibitor Cocktail (Roche Diagnostics, Mannheim, Germany). The homogenate was loaded onto the top of a 1.8 M sucrose layer and centrifuged at 28 000 g in 4°C for 30 min. The pellet containing nuclei was resuspended in 1 ml of 0.25 M sucrose in Buffer A. Nuclei were extracted with 1.4 M ammonium sulfate and the extracts were dialyzed against water.

Histone H1 extraction and purification

Frozen tissues were extracted with HClO_4 as described previously (9). Briefly, ~ 200 – 400 mg of each tissue was blended in 0.6–1.2 ml of 5% (v/v) HClO_4 using an Ultra Turbax blender (IKA, Staufen, Germany) at the maximum speed of approx 25 000 r.p.m. for 30 s. The suspension was subjected to three freezing (-20°C), thawing and vortexing (10 s) cycles. The cultured cells were extracted in the same way but without initial homogenization. The final suspension was centrifuged at 15 000 g for 10 min. Pellets were re-extracted with 0.2 M H_2SO_4 . Both supernatants were precipitated with ice cold 100% (v/v) CCl_3COOH was added to a final concentration of 33% (v/v). The proteins were precipitated for 30 min on ice and collected by centrifugation at 15 000 g for 10 min. The pellets were washed with 0.2% HCl in acetone, then twice with pure acetone and vacuum-dried. The protein pellets were solubilized in 0.1% (v/v) CF_3COOH in water and were separated on C_{18} reverse phase column as described previously (9).

Digestion and LC-MS/MS analysis

Proteins were resolved on 4–12% gradient Bis-Tris gels (NuPAGE, Novex, Invitrogen, Carlsbad, CA, USA) and stained using Colloidal Blue Stain Kit (Novex). Gel bands containing ~ 0.5 – 2 μg protein were trypsinized as described previously (13). The resulting peptide mixtures were desalted using in-house made C_{18} STAGE tips (14), vacuum-dried and reconstituted in 0.5% acetic acid prior to analysis. In all steps any use of formic acid or formaldehyde was avoided.

The LC-MS/MS setup was similar to that described before (15). The samples were injected onto an in-house made fused-silica capillary column (15 cm, inner diameter 75 μm , packed with 3 μm ReproSil-Pur C18-AQ media (Dr Maisch GmbH, Ammerbuch-Entringen, Germany), using the Agilent 1100 nanoflow system (Agilent Technologies, Palo Alto, CA, USA). The LC setup was connected to either an LTQ Orbitrap or LTQ-FT mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) equipped with a nanoelectrospray ion source

(Proxeon Biosystems, Odense, Denmark). The peptides were separated with 100 or 200 min gradients from 5 to 40% acetonitrile in 0.5% acetic acid. The data-dependent acquisition mode was employed. Survey MS scans were acquired either in the orbitrap with the resolution set to a value of 60 000 or in the FT with the resolution set to 100 000. Up to 5 most intense ions per scan were fragmented and analyzed in the linear trap. For accurate mass measurements of the survey MS scans the lock-mass option was used (16). Additional accurate mass measurements of the fragmentation products were performed in the orbitrap with the resolution set to 15 000.

The data were searched for post-translational modifications against the International Protein Index (IPI) database with the aid of the MASCOT (Matrix Science, London, UK) search engine (17). All relevant spectra were verified manually.

RESULTS

With the exception of the thiol-moiety of cysteine, the ϵ -amine of lysine is the strongest nucleophile in proteins and can undergo a variety of chemical reactions. These include the reaction with formaldehyde that results in formylation, and therefore procedures in which formaldehyde is used such as silver staining of gels, for example, have to be avoided (18). In our study, three different protein extraction methods were used to rule out the possibility that any of the reagents used were contaminated with formic acid or formaldehyde or could, directly or indirectly, generate N^ϵ -formylation of lysines. For this reason histones were extracted either using 0.2 M sulfuric acid or from density gradient purified nuclei with ammonium sulfate. Linker histones were also extracted together with high mobility group proteins with 5% perchloric acid. To further rule out that formylation could be caused by particular cell culture conditions, our analyses of core histones were performed using three different cultured cell types: MCF7 breast cancer cells, HeLa cervical cancer cells and A549 lung cancer cells. Mouse livers also were used as a source of histones. Linker histones were prepared from MCF7 cells and from human cancer tissues.

Identification of N^ϵ formylated lysines

Mass spectrometry (MS) is a powerful tool for the mapping of post-translational modifications in proteins, but the quality of the obtained data is highly dependent on the resolving power and mass-accuracy of the type of mass spectrometer employed. High accuracy MS instrumentation makes it possible to differentiate between moieties of the same nominal mass—thus helping to discern modifications such as trimethylation and acetylation. The presence of formylation on endogenous proteins was first suggested during in-depth profiling of H1 modifications with the LTQ-Orbitrap instrument (19) when some of the identified ‘dimethylated’ peptides had unusually high deviations from their theoretical masses (9). Dimethylation and formylation are modifications of the same nominal mass of 28 Da, but they differ by

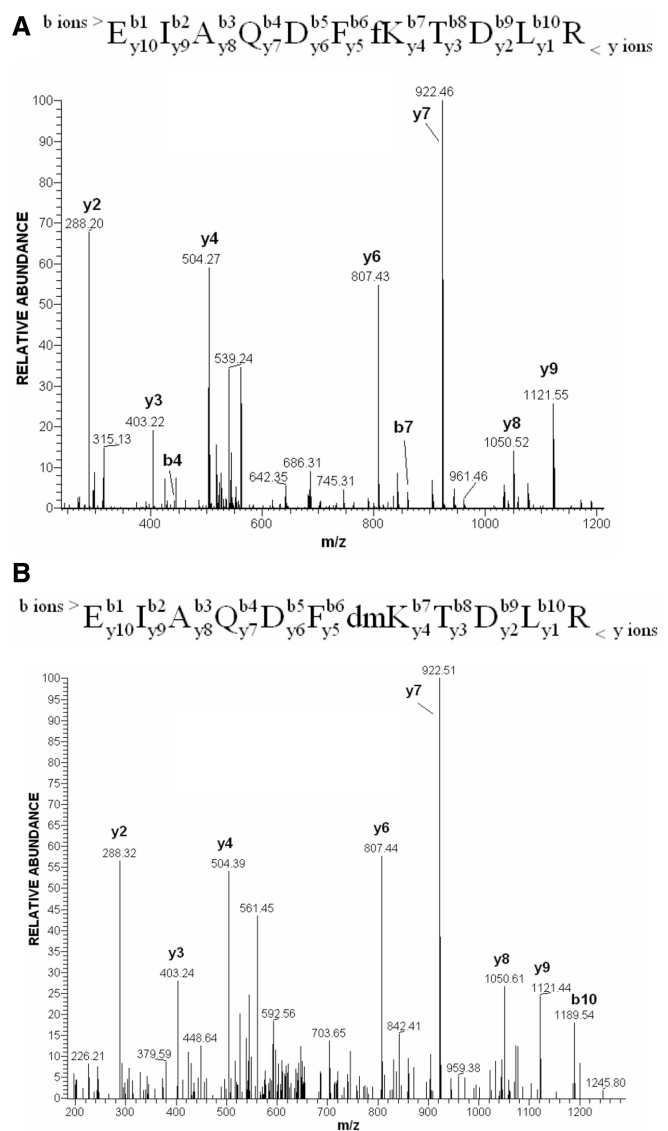


Figure 2. The MS/MS fragmentation spectra of (A) *EIAQDF/KTDLR* formylated and (B) *EIAQDFdmKTDLR* dimethylated peptides of Histone H3. The major groups of ions produced on cleavage of peptide bonds by low energy collision fragmentation are *b* ions extending from the N-terminus of a peptide, and *y* ions extending from its C-terminal part. The comprehensive introduction to the peptide sequencing by mass spectrometry is given by Steen and Mann (33).

0.0364 Da (Figure 1A). High-resolution mass spectrometers capable of mass measurement within a low p.p.m. window can easily distinguish these mass differences. Figure 1B shows the measured distribution of mass deviations of formylated and dimethylated peptides of histone H3 observed in different runs. The graph clearly demonstrates that these two modifications can be discriminated with high confidence by the accurate mass measurement. Fragmentation signatures of a histone H3 peptide that was identified as carrying either formylation or dimethylation are shown in Figure 2. Lysine-80 in the peptide 74–84 is the modification site. The fragmentation

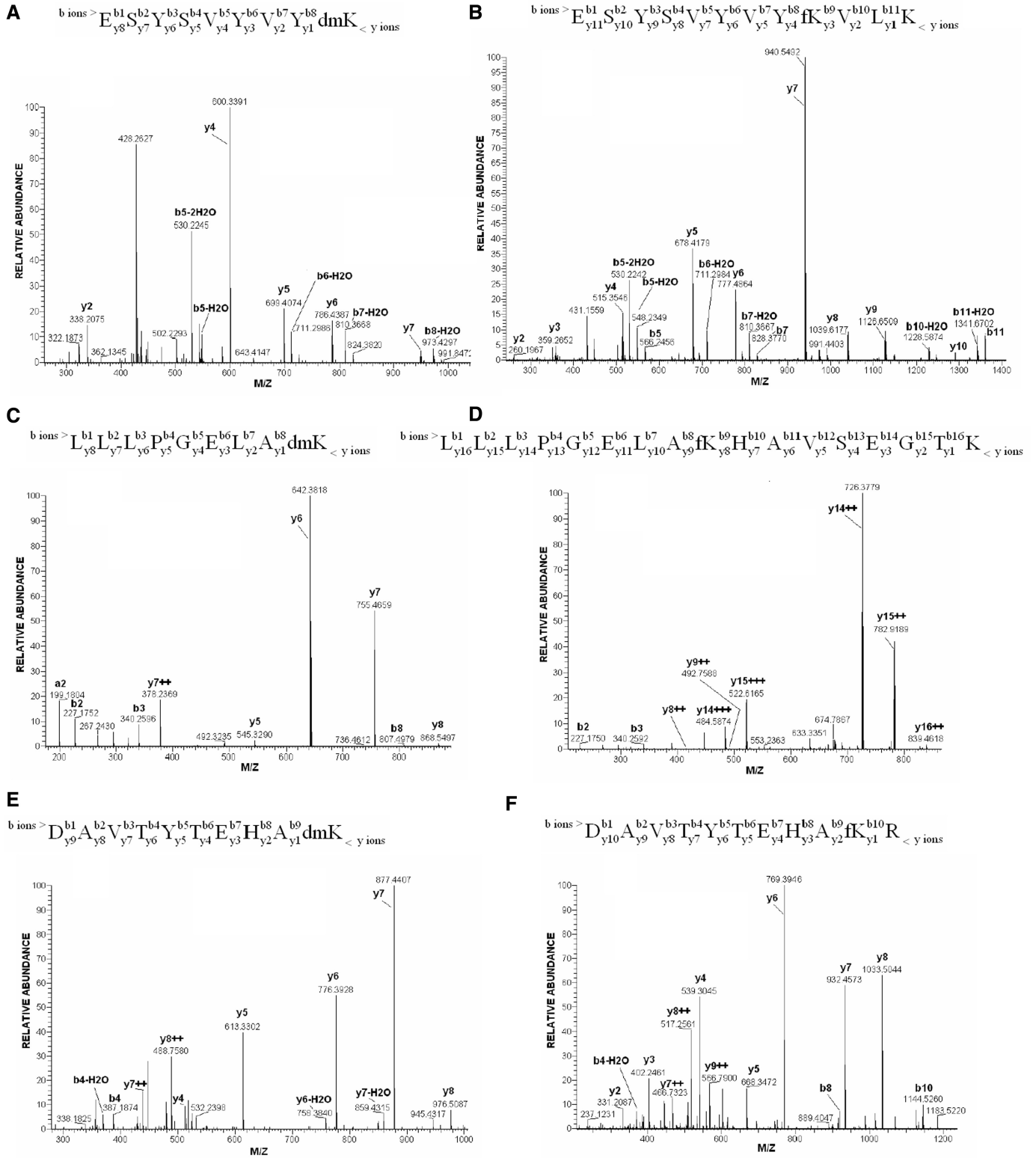


Figure 3. High accuracy MS/MS spectra of peptides dimethylated (A, C, E) and formylated (B, D, F) at the same residue, respectively. (A and B) Modifications at K-108 in histone H2B. (C and D) Modifications at K-43 in histone H2B. (E and F) Modifications at K-77 in histone H4. In contrast to dimethylation, formylation was observed mainly on peptides that were not cut at the modified lysyl residue. An example of the MS and MS/MS data interpretation: after the MS scan, the *ESYSVYVYKVLK* peptide (B) was identified as carrying an extra mass of 27.995 Da that corresponds to the delta mass of formylation. Peptide fragmentation analysis reveals that, starting from y_4 ion (*KVLK*), the formyl group is added to y -ion series. The observations, additionally supported by the presence of b_{10} (*ESYSVYVYKVL*) and b_{11} (*ESYSVYVYKVL*) ions with the mass shift indicative of formylation, unambiguously position the modification at the middle lysyl residue—*ESYSVYVYKVLK*. Average absolute mass accuracy of all assigned peaks was better than 2 p.p.m.

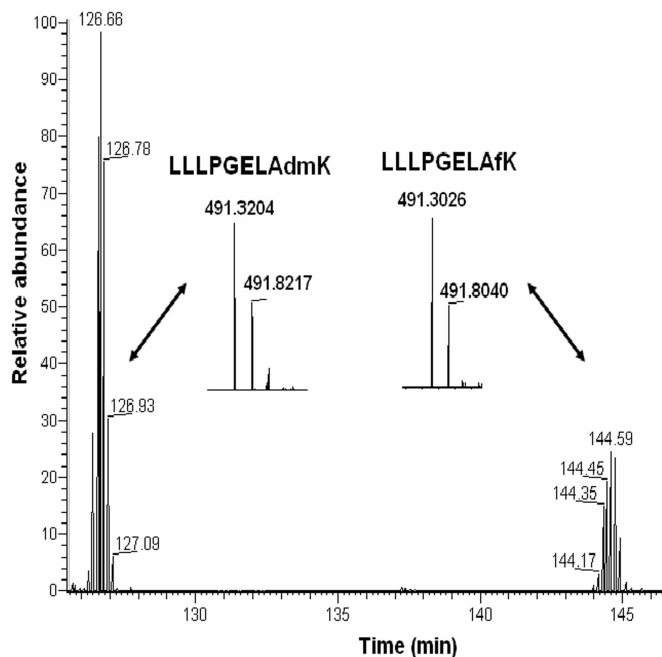


Figure 4. Dimethylated (dm) and formylated (f) forms of the peptide LLLPGELA*K (from histone H2B; K-108) resolved by chromatography. Inserts show the mass envelopes of dimethylated and formylated peptides.

spectra appear almost identical when acquired in the low accuracy linear ion trap detection mode, and, without the knowledge gained from the high accuracy peptide mass measurement would likely have been misidentified. A lysine to arginine mutation could in principle also be the cause of a nominal mass increase of 28 Da (28.00615 Da). However, the difference of 0.01123 Da between this modification and formylation is easily discriminated by high accuracy mass spectrometry. In Figure 3, examples of high accuracy fragmentation spectra of peptides dimethylated and formylated at the same residue are shown. We have observed that trypsin cuts only weakly at the N-site of formylated lysine, what is in contrast to its activity on dimethylated lysine. Therefore identification of both modifications at the same residue was often possible only on peptides with different lengths. For this reason, it is difficult to make conclusions on the ratio of the abundances of dimethylation and formylation comparing the intensities of the relevant peptides in the chromatogram. Only in seldom cases, when a dimethylated peptide was miscleaved at the C-site of the dimethylated residue, a comparison of the ratios of both modifications at the same site was possible (Figure 4). For the peptide LLLPGELA*K from histone H2B (HeLa cells) formylation appears to be less abundant than dimethylation (Figure 4). Interestingly, the formylated peptide shows increased retention in comparison to its dimethylated version (Figure 4).

In this study, we isolated histones and other nuclear proteins from different cell types and tissues. These protein mixtures were digested with trypsin and online analyzed on the LTQ-Orbitrap as described in Materials

and Methods section. Only the formylation sites observed at least in two analyses are presented in Table S1. In total we identified 165 peptides containing *N*^ε-formyl lysine using high accuracy mass spectrometry (Supplementary Table S1). They represent 47 unique formylation sites in core and linker histones, high mobility group proteins and two other proteins (Table 1).

Core histones

Our analysis allowed identification of 19 *N*^ε-lysine formylation sites in core histones that were located in their N-terminal tails and globular domains. Two sites of formylation were found in the tail of histone H3 and three in its globular domain (Table 1). The sites at K-18 and K-23 that are located in the tail of H3 are known to occur in both acetylated and methylated forms (3). In the core of H3 we identified formylation at K-64, K-79 and K-122. Methylation or dimethylation of these residues have been previously reported (20).

In histone H4, we identified one formylation site in the tail (K-12) and five sites in its globular domain. This site is functionally important because acetylated K-12 is selectively targeted by the BRD2 bromodomain and activates transcription from the cyclin E promoter (21). In histone H2A, three sites of formylation at K-36, K-95 and K-118 were mapped to the globular domain of the histone, whereas in histone H2B a site in the tail K-5 and four sites in the core of this histone were mapped at K-34, K-44, K-47 and K109. Lysine H2B-K-5 is known to be acetylated.

These data show that formylation often occurs at residues that can also be methylated and acetylated. Clearly, formylation of these sites prevents subsequent methylation and acetylation. Since many of those modifications are important in chromatin regulation, formylation could interfere with this process.

In the nucleosomal particles some of the histone lysyl moieties are involved in direct interactions with DNA (22). These include residues H3-K64, H4-K79, H2B-K34 that we identified in a formylated state and that create main chain interactions with DNA. Additionally, we measured formylation on H4-K77, which sits in the major groove of DNA (22). Thus, formylation of lysyl residues may also impair organization of the nucleosomes.

Linker histones

Analysis of linker histones revealed formylated lysines distributed over all three structural domains of these proteins: the N-terminal tail, the central domain and the C-terminal tail. For the most abundant H1 variant in human cells, the H1.4 protein, we identified 12 sites of formylation (Table 1). Seven of these sites are located in the central domain and three of them, K-64, K-85 and K-97 (H1.4 residue numbering), at residues involved in DNA binding (23,24). Since formylation, similar to acetylation, can interfere with DNA binding, it is possible that these modifications can significantly change the binding properties of H1. It is interesting in this context that nine of the formylated sites were previously identified as acetylated.

Table 1. Formylation sites identified in human and mouse tissues

Protein	Peptide	Site	Domain	Occurrence of formylation (times)						
				MCF7 (3) ^a	MCF7 (2)	HeLa (3)	A543 (3)	Mouse liver (2)	Human cancer ^b (3)	
				Extraction method						
				HClO ₄	H ₂ SO ₄	H ₂ SO ₄	H ₂ SO ₄	Salt	HClO ₄	
H1.2	Ac-SETAPAAPAAAPAEfKAPVK	K-17	N-tail	2					2	
H1.4	Ac-SETAPAAPAAAPAEfKTPVK	K-17	N-tail	2					2	
H1.2–H1.4	fKASGPPVSELITK	K-34	GH1	3						
H1.5	fKATGPPVSELITK	K-37	GH1	2						
H1.2–H1.4	ASGPPVSELITfKAVAASK	K-46	GH1	1						2
H1.1–H1.4	SGVSLAALfKK	K-63	GH1	3						
H1.1–H1.4	fKALAAAGYDVEK	K-64	GH1	3						
H1.4	ERSGVSLAALfK	K-64	GH1							3
H1.5	fKALAAGGYDVEK	K-67	GH1	2						
H1.1–H1.4	ALAAAGYDVeFk	K-75	GH1					2		
H1.1–H1.5	LGLfKSLVSK	K-85	GH1	2						3
H1.1–H1.4	SLVSfKGTLVQTK	K-90	GH1	2						
H1.1–H1.4	GTLVQTFKGTGASGSFK	K-97	GH1	3						
H1.4	fKAASGEAKPK	K-110	C-tail						2	
H1.4	fKATGAATPK	K-140	C-tail	1					1	
H1.3	fKVAGAATPK	K-141	C-tail	2						
H1.3; H1.4	fKPAAAAGAK	K-160	C-tail						2	
H1.2	fKPAAAATVK	K-160	C-tail	2						
H2A	fKGNYAER	K-36	Core		1				1	
H2A	NDEELNfKLLGK	K-95	Core		1	2	2		1	
H2A	VTIAQGGVLPNIQAVLLPfKK	K-118	Core		2	2	2			
H2B	PEPAfKSAPAPK	K-5	N-tail			1			2	
H2B	fKESYSVYVYK	K-34	Core		2	2	2		1	
H2B	ESYSVYVYfKVLK	K-43	Core			1			1	
H2B	VLfKQVHPDTGISSK	K-45	Core			2				
H2B	LLLPGELAfK	K-107	Core		2	3	1		2	
H3	fKQLATaKAAAR, fKQLATK	K-18	N-tail		2	2	2		2	
H3	QLATfKAAAR	K-23	N-tail						2	
H3	fKLPfQR	K-64	Core		1	3	2			
H3	EIAQDFfKTDLR	K-79	Core		2	3	2		1	
H3	VTIMPFKDIQLAR	K-122	Core			2	1		1	
H4	GLGfKGGAAKR	K-12	N-tail			2	1		1	
H4	DNIQGfTKPAIR	K-31	N-tail		2	3	2		2	
H4	GVLfKVFLENVIR	K-59	Core						2	
H4	DAVTYTEHAfK	K-77	Core		1	2	1			
H4	fKTVTAMDVVYALK	K-79	Core			3			1	
H4	TVTAMDVVYALfKR	K-91	Core			2			1	
HMGB1	GfKFEDoxMAK	K-59	HMG box A	2						
HMGB1	FfKDPNAPK	K-90		2					2	
HMGB1	IfKGEHPGLSIGDVAK	K-114	HMG box B	1					1	
HMGB1	YEfKDIAAAYR	K-157	HMG box B	2						
HMGN2	LSAfKPAAPKPEPK	K-31	NBD1 ^c	2						
HMGA1	fKQPPVSPGTALVGSQK	K-31	DBD2 ^d	2						
Lamin A/C	LfKDLEALLNSK	K-135				2			1	
Calgizzarin	Ac-AfKISSPTETER	K-2		2						
Calgizzarin	NQfKDPGVLDLDR	K-55		2						
Calgizzarin	YAGfKDGNYNTLSK	K-27		2						

^aNumber of independent experiments.^bSingle samples of breast ductal, colon, and ovarian cancer were analyzed.^cNBD, nucleosomes-binding domain.^dDBD, DNA-binding domain.

'Shaded areas' indicate that HClO₄ extracts do not contain core histones and lamin A/C; in sulfuric acid extracts only protein bands containing core histones were analyzed; in salt extracts of mouse liver HMGN1 and HMGA1 were not identified; and in cancer samples only fractions containing linker histones were analyzed.

HMG proteins

Formylation sites were found in each on the dominant members of the high mobility group protein families HMGA, HMGB and HMGN. Formylation of K-31 in HMGA1 is located in one of the three DNA-binding

domains of this protein. HMGN2 is the most abundant of the nucleosome-binding proteins and we found formylation in its nucleosome-binding domain. In the HMGB1 protein, four different formylation sites were mapped. Three sites are located in the HMG-box domains (K-59,

K-114 and K-157 and in the linker between the domains (K-90). K-114 plays an important role in interaction with DNA, because mutation of this residue results in significant loss of DNA-binding affinity and supercoiling activity of the HMG-box B-domain (25).

Other proteins

Calgizzarin, also known as S100A11, is a calcium-binding protein implicated in a variety of biologic functions such as proliferation and differentiation, as well as in cancer. For a long time cellular localization of this protein was obscure, but recently calgizzarin was identified in nuclear fractions from cultured cells (26). We observed three formylation sites in this protein. Lamins are components of the nuclear lamina, a fibrous layer on the nucleoplasmic side of the inner nuclear membrane, which is thought to provide a framework for the nuclear envelope and may also interact with chromatin. Identification of formylation of K-135 in lamin A/C suggests that this residue may be located in the vicinity of DNA.

DISCUSSION

Recently, formylation has been recognized as a novel post-translational modification of lysine that occurs in histones (9,10). The two previous studies had mapped none (10) or three sites in a single protein (9). In this study, we have considerably extended our knowledge of the occurrence of this post-translational modification in histones and other nuclear proteins by proteomic mapping of 48 formylated sites in 10 different proteins. We took advantage of the high-mass accuracy of modern MS instrumentation that allowed unambiguous discrimination between dimethylation (28.0313 Da) and formylation (27.9949 Da). In addition, both modifications can be distinguished by different LC-retention times. We observed that formylation can affect lysyl residues important for interaction with DNA and also those modified by acetylation and methylation. Thus, formylation can impair the function of histones by interfering with the histone code-mediated formation and stabilization of binding sites for regulatory proteins such as transcription factors and regulators of chromatin activity.

Histone H1 was found to be the most frequently formylated among histones. Whereas in four core histones 19 formylation sites were identified altogether, in the most abundant linker histone variant H1.4 thirteen formyl-lysines were mapped. The higher number of H1 formylation sites correlates with the high degree of H1 formylation observed previously (10). The abundance of formylation in all three structural domains of H1 may reflect the fact that, similarly to the DNA-binding domain, both N-terminal and C-terminal tails are in close contact with DNA.

Since histones and some other chromosomal proteins have slow turnover rates (27–29), lysine formylation could accumulate with age and contribute to deregulation of chromatin function. Furthermore, lysine formylation is stimulated by oxidative stress (10) and, therefore, this modification may be involved in development of diseases

linked to this stress, including cancer. In fact, lysine formylation may be involved in loss of protein function in the organism in an analogous way to other chemical modifications, such as glycation in diabetes (30) or oxidative damage in aging (31,32).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Piotr Ziółkowski (The Wrocław Medical University, Poland) for the kind gift of human tissue biopsies, and Sonja Krüger for technical assistance. This work was supported by the Max-Planck Society for the Advancement of Science and by HEROIC, an Integrated Project funded by the European Union under the 6th Framework Programme (LSHG-CT-2005-018883). Funding to pay the Open Access publication charges for this article was provided by HEROIC.

Conflict of interest statement. None declared.

REFERENCES

1. Chu, F., Nusinow, D.A., Chalkley, R.J., Plath, K., Panning, B. and Burlingame, A.L. (2006) Mapping post-translational modifications of the histone variant MacroH2A1 using tandem mass spectrometry. *Mol. Cell. Proteomics*, **5**, 194–203.
2. Medzihradszky, K.F., Zhang, X., Chalkley, R.J., Guan, S., McFarland, M.A., Chalmers, M.J., Marshall, A.G., Diaz, R.L., Allis, C.D. *et al.* (2004) Characterization of Tetrahymena histone H2B variants and posttranslational populations by electron capture dissociation (ECD) Fourier transform ion cyclotron mass spectrometry (FT-ICR MS). *Mol. Cell. Proteomics*, **3**, 872–886.
3. Hake, S.B., Garcia, B.A., Duncan, E.M., Kauer, M., Dellaire, G., Shabanowitz, J., Bazett-Jones, D.P., Allis, C.D. and Hunt, D.F. (2006) Expression patterns and post-translational modifications associated with mammalian histone H3 variants. *J. Biol. Chem.*, **281**, 559–568.
4. Garcia, B.A., Barber, C.M., Hake, S.B., Ptak, C., Turner, F.B., Busby, S.A., Shabanowitz, J., Moran, R.G., Allis, C.D. *et al.* (2005) Modifications of human histone H3 variants during mitosis. *Biochemistry*, **44**, 13202–13213.
5. Bonenfant, D., Coulot, M., Towbin, H., Schindler, P. and van Oostrum, J. (2006) Characterization of histone H2A and H2B variants and their post-translational modifications by mass spectrometry. *Mol. Cell. Proteomics*, **5**, 541–552.
6. Kouzarides, T. (2007) Chromatin modifications and their function. *Cell*, **128**, 693–705.
7. Santos-Rosa, H. and Caldas, C. (2005) Chromatin modifier enzymes, the histone code and cancer. *Eur. J. Cancer*, **41**, 2381–2402.
8. Garcia, B.A., Joshi, S., Thomas, C.E., Chitta, R.K., Diaz, R.L., Busby, S.A., Andrews, P.C., Ogorzalek-Loo, R.R., Shabanowitz, J. *et al.* (2006) Comprehensive phosphoprotein analysis of linker histone H1 from Tetrahymena thermophila. *Mol. Cell. Proteomics*, **5**, 1593–1609.
9. Wiśniewski, J.R., Zougman, A., Krüger, S. and Mann, M. (2007) Mass spectrometric mapping of linker histone H1 variants reveals multiple acetylations, methylations, and phosphorylation as well as differences between cell culture and tissue. *Mol. Cell. Proteomics*, **6**, 72–87.
10. Jiang, T., Zhou, X., Taghizadeh, K., Dong, M. and Dedon, P.C. (2007) N-formylation of lysine in histone proteins as a secondary modification arising from oxidative DNA damage. *Proc. Natl Acad. Sci. USA*, **104**, 60–65.

11. Mann, M. and Jensen, O.N. (2003) Proteomic analysis of post-translational modifications. *Nat. Biotechnol.*, **21**, 255–261.
12. Blobel, G. and Potter, V.R. (1966) Nuclei from rat liver: isolation method that combines purity with high yield. *Science*, **154**, 1662–1665.
13. Shevchenko, A., Wilm, M., Vorm, O. and Mann, M. (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.*, **68**, 850–858.
14. Rappsilber, J., Ishihama, Y. and Mann, M. (2003) Stop and go extraction tips for matrix-assisted laser desorption/ionization, nano-electrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.*, **75**, 663–670.
15. Olsen, J.V. and Mann, M. (2004) Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. *Proc. Natl Acad. Sci. USA*, **101**, 13417–13422.
16. Olsen, J.V., de Godoy, L.M., Li, G., Macek, B., Mortensen, P., Pesch, R., Makarov, A., Lange, O., Horning, S. *et al.* (2005) Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics*, **12**, 2010–2021.
17. Perkins, D.N., Pappin, D.J., Creasy, D.M. and Cottrell, J.S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, **20**, 3551–3567.
18. Oses-Prieto, J.A., Zhang, X. and Burlingame, A.L. (2007) Formation of epsilon-formyllysine on silver-stained proteins: implications for assignment of isobaric dimethylation sites by tandem mass spectrometry. *Mol. Cell. Proteomics*, **6**, 181–192.
19. Makarov, A., Denisov, E., Kholomeev, A., Balschun, W., Lange, O., Strupat, K. and Horning, S. (2006) Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal. Chem.*, **78**, 2113–2120.
20. Garcia, B.A., Hake, S.B., Diaz, R.L., Kauer, M., Morris, S.A., Recht, J., Shabanowitz, J., Mishra, N., Strahl, B.D. *et al.* (2007) Organismal differences in post-translational modifications in histones H3 and H4. *J. Biol. Chem.*, **282**, 7641–7655.
21. Kanno, T., Kanno, Y., Siegel, R.M., Jang, M.K., Lenardo, M.J. and Ozato, K. (2004) Selective recognition of acetylated histones by bromodomain proteins visualized in living cells. *Mol. Cell*, **13**, 33–43.
22. Luger, K. and Richmond, T.J. (1998) The histone tails of the nucleosome. *Curr. Opin. Struct. Biol.*, **8**, 33–40.
23. Goytisolo, F.A., Gerchman, S.E., Yu, X., Rees, C., Graziano, V., Ramakrishnan, V. and Thomas, J.O. (1996) Identification of two DNA-binding sites on the globular domain of histone H5. *EMBO J.*, **15**, 3421–3429.
24. Brown, D.T., Izard, T. and Misteli, T. (2006) Mapping the interaction surface of linker histone H1(0) with the nucleosome of native chromatin in vivo. *Nat. Struct. Mol. Biol.*, **13**, 250–255.
25. Štros, M. and Muselikova, E. (2000) A role of basic residues and the putative intercalating phenylalanine of the HMG-1 box B in DNA supercoiling and binding to four-way DNA junctions. *J. Biol. Chem.*, **275**, 35699–35707.
26. Sakaguchi, M., Miyazaki, M., Kondo, T. and Namba, M. (2001) Up-regulation of S100C in normal human fibroblasts in the process of aging in vitro. *Exp. Gerontol.*, **36**, 1317–1325.
27. Gurley, L.R. and Hardin, J.M. (1969) The metabolism of histone fractions. II. Conservation and turnover of histone fractions in mammalian cells. *Arch. Biochem. Biophys.*, **130**, 1–6.
28. Hancock, R. (1969) Conservation of histones in chromatin during growth and mitosis in vitro. *J. Mol. Biol.*, **40**, 457–466.
29. Ghidelli, S., Claus, P., Thies, G. and Wiśniewski, J.R. (1997) High mobility group proteins cHMG1a, cHMG1b, and cHMG1 are distinctly distributed in chromosomes and differentially expressed during ecdysone dependent cell differentiation. *Chromosoma*, **105**, 369–379.
30. Vlassara, H. and Palace, M.R. (2002) Diabetes and advanced glycation endproducts. *J. Int. Med.*, **251**, 87–101.
31. Johnson, F.B., Sinclair, D.A. and Guarente, L. (1999) Molecular biology of aging. *Cell*, **96**, 291–302.
32. Huang, H. and Manton, K.G. (2004) The role of oxidative damage in mitochondria during aging: a review. *Front Biosci.*, **9**, 1100–1117.
33. Steen, H. and Mann, M. (2004) The ABC's (and XYZ's) of peptide sequencing. *Nat. Rev. Mol. Cell. Biol.*, **5**, 699–711.