In Vivo Posttranslational Modifications of the High Mobility Group A1a Proteins in Breast Cancer Cells of Differing Metastatic Potential†

Dale D. Edberg,§ James E. Bruce,‡ William F. Siems,‡ and Raymond Reeves*,§

School of Molecular Biosciences and Department of Chemistry, Washington State University, Pullman, Washington 99164

Received January 22, 2004; Revised Manuscript Received May 18, 2004

ABSTRACT: The high mobility group (HMG) proteins are important modulators of chromatin structure and gene transcription. Overexpression of HMGA1 proteins in vivo induces neoplastic transformation and promotes a highly metastatic cellular phenotype. This study focuses on characterization of HMGA1a in vivo posttranslational modification (PTM) patterns found in a nonmetastatic and two metastatic lines of MCF-7 human breast cancer cells of differing tumorigenic potential. PTM types and the amino acids on which they occur were identified by matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry. Mass analysis was restricted to MALDI peaks having less than ±150 parts per million (ppm) error, thereby holding our analysis to a more stringent criterion than previously published work with the HMG proteins. Validation of MALDI-TOF MS analysis was accomplished utilizing electrospray ionization tandem mass spectrometry (ESI MS/MS) and manual analysis of ion fragmentation spectra. Patterns and sites of PTMs identified in this study suggest that HMGA1a proteins, like the histones, exhibit a biochemical modification "code" that relates to cellular function. For example, both increased levels of acetylation and a previously unidentified dimethylation of both lysine and arginine residues were found on HMGA1a proteins from metastatic cells compared to proteins found in their nonmetastatic precursors. Additionally, the types of modification present on lysine-45 (e.g., unmodified, acetylation, or dimethylation) varied, depending on the metastatic potential of cells. These findings suggest that examination of the PTM patterns on HMGA1 proteins may provide valuable information concerning the physiological and phenotypic state of mammalian cells.

High mobility group (HMG) proteins are nonhistone DNA binding proteins that are implicated in regulating expression of numerous mammalian genes (*1*−*4*). The human HMGA1[1] subfamily of HMG proteins consists of two major protein isoforms, HMGA1a (107 amino acids (aa), 11.5 kDa) and HMGA1b (96 aa, 10.4 kDa), which are produced by translation of alternatively spliced mRNA coded for by the *HMGA1* gene at chromosomal locus 6p21 (*5*). The HMGA1c (179 aa, 19.7 kDa) protein, another member of the family, is produced by a 67 nucleotide splice deletion from the *HMGA1* mRNA sequence, resulting in a frame shift in the transcript, which codes for a larger C-terminal tail region (*6*). In addition to transcription, HMGA1 proteins are implicated in a number of other of cellular processes including regulation of cell growth, apoptosis, chromatin remodeling, and tumor progression (*3*, *4*). Given these biological effects, it is not surprising that elevated levels of HMGA1 protein have been demonstrated to be a diagnostic feature of many types of cancer (*3*, *4*, *7*). Nevertheless, the molecular modifications found on HMGA1 proteins in cancer cells at different stages of neoplastic progression have, so far, received little attention.

In addition to overexpression, elevated levels of HMGA1 protein in cancer cells have also been directly correlated with increasing tumor malignancy (*3*, *4*, *7*). For example, increasing HMGA1a levels have been demonstrated to directly correlate with the metastatic potential of breast cancer cells (*8*−*10*). Recently, HMGA1 has been classified as an oncoprotein (*11*, *12*) with overexpression both causing neoplastic transformation of normal cells and promoting malignant metastatic progression of tumor cells (*12*, *13*). Consistent with their role as transcription factors that can either positively or negatively regulate gene expression (*3*), breast cancer cells overexpressing HMGA1 proteins exhibit changes in gene expression profiles compared to their normal precursors (*10*, *12*, *14*). The elevated levels of HMGA1 proteins in aggressive mammary epithelial cancers has been shown, for example, to directly inhibit transcription of the

† This work was supported, in part, by NIH Grant GM-46352 and WSU Cancer Research Development Fund Grant 17A-0165 (both to R.R.).

* Corresponding author. Mailing address: School of Molecular Biosciences, Washington State University, Fulmer Syn. Rm. 664, Pullman, WA, 99164-4660. Tel: 509-335-1948. Fax: 509-335-9688. E-mail: reevesr@mail.wsu.edu.

§ School of Molecular Biosciences.

‡ Department of Chemistry.

[1] Abbreviations: calcd mass, peptide mass calculated from the amino acid sequence plus biochemical modifications; ESI MS/MS, electrospray ionization tandem mass spectrometry; HA, hemagglutinin tag; HA7C, transgenic HA-HMGA1a overexpressing MCF-7/Tet off cell line; HA8A, transgenic HA-HMGA1b overexpressing MCF-7/Tet off cell line; HMGA1, high mobility group A1 proteins; MALDI-TOF MS, matrix-assisted laser desorption ionization time-of-flight mass spectrometry; *m*/*z*, mass-to-charge ratio; obsd mass, peptide mass observed by MALDI-TOF MS; PKC, protein kinase C; PTM, posttranslational modification; rh, recombinant human; RP-HPLC, reverse-phase high-performance liquid chromatography; SAP, shrimp alkaline phosphatase; SDS, sodium dodecyl sulfate; TCA, trichloroacetic acid; TFA, trifluoroacetic acid.

10.1021/bi049833i CCC: $27.50
Published on Web 08/20/2004

tumor suppressor gene BRCA1 as a result of its promoter being negatively regulated by HMGA1 protein binding (*15*). Likewise, transcriptional expression of many other genes coding for proteins that are linked to metastatic progression, including several integrins, cyclins, extracellular metalloproteases, and Rho proteins (*12*), as well as genes involved in activating the Ras/ERK pathway (*14*), is significantly modulated by overexpression of HMGA1 proteins in vivo. While these findings demonstrate that HMGA1 proteins are involved in the tumorigenic process via their role as transcription factors, no comprehensive determination of the constellation of biochemical modifications present on the HMGA1 proteins themselves in cancer cells that likely affect their gene regulatory function(s) has yet been reported.

For over a decade, posttranslational modifications (PTMs) of HMGA1 proteins have been studied. Prior research has shown that HMGA1 proteins are phosphorylated and acetylated in vivo and can be ADP-ribosylated in vitro (*16*). The HMGA1 proteins are phosphorylated by cell cycle dependent 2 (cdc2) kinase in the G2/M phase of the cell cycle (*17*). In addition, many other kinases have also been found to phosphorylate mammalian HMGA1 proteins both in vitro and in vivo (*3*, *4*). Importantly, phosphorylation is known to affect HMGA1 protein function. In vivo and in vitro PKC phosphorylation of HMGA1 proteins inhibits their binding to both naked DNA and nucleosome-containing DNA, thereby affecting the stability of protein−DNA complexes (*18*, *19*). Another kinase, casein kinase II (CK II), phosphorylates the acidic tail of HMGA1 proteins at serine residues 102 and 103 both in vitro and in vivo (*20*, *21*), linking intracellular signaling to phosphorylation events occurring in the nucleus (*22*). Interestingly, hyperphosphorylation of HMGA1 proteins has been demonstrated to occur in the early stages of apoptosis and is subsequently followed by a hypophosphorylation, thus implicating HMGA1 biochemical modifications in the apoptotic process (*23*, *24*). These findings indicate that phosphorylation of HMGA1 proteins is linked to many different cellular events, including transcriptional regulation, cell cycle progression, cell signaling, and cell death.

In addition to phosphorylation, other PTMs, such as acetylation, have been implicated in modulating HMGA1 protein function. For example, HMGA1 proteins can be acetylated by CREB binding protein (CBP) and the p300/CBP associated factor (P/CAF) at lysine residues 65 and 71, respectively (*25*, *26*). Acetylation of lysine 71 by P/CAF has been suggested to be associated with positive transcriptional regulation of the β interferon (IFN-β) gene promoter by HMGA1 proteins, while acetylation of lysine 65 by CBP has been associated with negative regulation (*25*, *26*). Additionally, in earlier work, we demonstrated that acetylation occurs at multiple sites on the HMGA1 proteins in vivo in MCF-7 cells (*18*), but no assignments of these modifications to specific lysine residues were made. Likewise, in this previous study, another important in vivo modification of HMGA1a was discovered, methylation, but again the precise amino acid residues modified were not identified.

Methylation of HMGA1 proteins has been the least studied of the PTMs, and the importance of this type of modification is just beginning to be understood. Recently, however, identification of a methylation site, arginine 25, on HMGA1a in human leukemia and prostate cancer cell lines was established (*27*). Although the biological function of monomethylation on arginine 25 is unknown, it has been correlated with cells undergoing programmed cell death (*27*).

In the present study, we investigated the in vivo PTMs of the HMGA1a protein by a highly sensitive method, MALDI-TOF mass spectrometry (MALDI-TOF MS), which requires small quantities of protein (nanogram to microgram). Three different MCF-7 human breast epithelial cancer cell lines were used to characterize the PTMs of the HMGA1a protein as they relate to different stages of neoplastic progression. One of the MCF-7 cell lines was nonmetastatic and noninvasive, whereas the other two derivative MCF-7 cell lines exhibited different degrees of metastatic potential as a result of experimentally induced overexpression of transgenic HMGA1 proteins (*12*). A significant technical advance employed in the present studies is that all of the HMGA1a proteins, and not just selected HMGA1a subfractions (as has been done in the past), were isolated from each cell line and analyzed by mass spectrometry. Our studies therefore revealed a much higher level of in vivo HMGA1a PTMs than has previously been reported. Importantly, a new type of modification for HMGA1a proteins, dimethylation of arginine and lysine residues, was identified that correlated with the metastatic potential of cells. Likewise, a differential modification of lysine 45 was observed that varied between unmodified, acetylated, and dimethylated forms and was also correlated with metastatic potential. These findings provide strong support for the idea that the complex patterns of PTMs found on individual HMGA1a proteins in vivo are both dynamic and cell-type-specific and also suggest that a malignancy-dependent biochemical "code" may exist for the HMGA1 proteins.

MATERIALS AND METHODS

Protein Isolation, Purification, and Detection. Recombinant human (rh) HMGA1 proteins were extracted from cells with dilute (5%) trichloroacetic acid (TCA) and purified by reverse-phase high-performance liquid chromatography (RP-HPLC) techniques as described by Reeves and Nissen (*28*). Briefly, following acid extraction, the HMGA1 proteins were purified utilizing a RP-HPLC C4 Microsorb analytical column with a 12−25%, 72-min linear gradient from 0.2% trifluoroacetic acid (TFA) in H_2O to 0.2% TFA in acetonitrile. Subsequently, HMGA1a protein fractions were further purified using a RP-HPLC C18 Dynamax analytical column with a 23%, 80-min linear gradient from 0.2% TFA in H_2O to 0.2% TFA in acetonitrile. Protein purity of isolated fractions was assessed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS−PAGE) following standard protocols (*28*) and confirmed by MALDI-TOF MS. In vivo modified HMGA1 proteins were isolated from the MCF-7 cell lines by following the above conditions.

Cell Lines and Cell Culture Methods. The human breast epithelial cancer cell line MCF-7/Tet off was purchased from Clontech (Palo Alto, CA, catalog no. C30071). MCF-7/Tet off cells were stably transfected with either HMGA1a or HMGA1b cDNA-containing expression vectors the transcription of which was controlled by a tetracycline-regulated promoter (*12*). A hemagglutinin (HA) peptide tag within the parental expression vector was linked in-frame to the N-terminal end of the HMGA1a and HMGA1b proteins. Therefore, the MCF-7/Tet off cell lines that contain stably

integrated transgenes coding for these HA-tagged proteins are called, respectively, HA7C and HA8A (i.e., 7C is the HMGA1a cDNA clone designation and 8A is the HMGA1b cDNA clone designation) (*29*). All cell lines were grown in Dulbecco's modified Eagle media (GibcoBRL, Grand Island, NY) and supplemented with 10% fetal bovine serum (Atlanta Biologicals, Norcross, GA), 2 mM L-glutamine, 100 U/mL penicillin G, and 100 mg/mL streptomycin sulfate (Sigma Co., St. Louis, MO). Stable integration of the transgenic HMGA1 cDNA expression constructs within the transgenic HA7C and HA8A cell lines was maintained by addition of either 100 mg/mL hygromycin (Calbiochem, San Diego, CA) or 50 mg/mL zeocin (Invitrogen, Carlsbad, CA) to the culture medium, respectively. The cells were harvested when confluent, frozen, and stored at −70 °C until needed for analysis.

Identification of In Vivo Modified HMGA1 Proteins Isolated from MCF-7 Cells. Western blot analysis was used to confirm the presence various modified forms of HMGA1 proteins extracted from both the parental and transgenic MCF-7 cell lines and from RP-HPLC purified protein fractions following published protocols (*28*). Specifically, the in vivo modified HMGA1 proteins present in various chromatographic fractions were identified by Western blot analyses using polyclonal rabbit anti-HMGA1 antiserum (MR19) raised against purified rhHMGA1a proteins following published protocols (*28*). Cultured cells (approximately $(2-3) \times 10^6$ cells per dish) were harvested for analysis in SDS−PAGE loading buffer. Samples were transferred to microcentrifuge tubes, sonicated for 1 min to lyse the cells and shear the DNA, boiled for 5 min to denature the proteins, and centrifuged at $15\,000 \times g$ for 2 min to pellet any particulate matter. RP-HPLC fractions were dried down, redissolved in distilled and deionized (dd) H_2O, and combined at a 1:1 volume ratio of protein solution and SDS−PAGE loading buffer for gel analysis. The purity of and concentration (within the nanogram to microgram range) of endogenous, in vivo modified HMGA1a proteins was determined by electrophoretic separation of the proteins on SDS−PAGE gels using known concentrations of pure recombinant human HMGA1 proteins as reference standards (data not shown).

Enzyme Digestion of the HMGA1a Proteins. Approximately 0.5 mg of purified native HMGA1 protein from each of the three experimental cells lines was separated into two equal fractions and lyophilized. One sample was dephosphorylated with shrimp alkaline phosphatase (SAP, USB, Cleveland, OH) with the reactions (3 μg of in vivo modified HMGA1a, 20 units of SAP) carried out at 37 °C for 10 h in the reaction buffer supplied by the manufacturer. Dephosphorylation reactions were terminated by addition of perchloric acid to a concentration of 5%, and the dephosphorylated HMGA1a proteins were purified from the SAP by the standard TCA precipitation procedure (*28*). Dephosphorylated HMGA1a samples were lyophilized, reconstituted, and divided into two fractions. Both artificially dephosphorylated and native in vivo modified HMGA1a proteins were digested with sequencing grade proteinase Arg-C (Sigma Co., St. Louis, MO) or sequencing grade trypsin (Promega, Madison, WI) (50:1 mass ratio of protein to enzyme) in 50 mM ammonium bicarbonate (pH 8.0) at 37 °C. The length of digestion was empirically determined to yield optimal peptide fragmentation of HMGA1a proteins for MALDI-TOF MS analysis. Tryptic reactions were carried out for 4 h, while Arg-C reactions were digested for 10 h. Partial tryptic digestions were necessary since complete cleavage of the HMGA1a protein results in small peptide fragments, making them difficult to analyze. Enzymatic reactions were terminated by adding TFA to a final volume of 2%. The solution was lyophilized, and the peptides were redissolved in ddH_2O.

MALDI-TOF Mass Spectrometry. MALDI-TOF mass spectrometry was performed in the linear mode using a PerSpective Biosystems Voyager DE-RP instrument (Framingham, MA) according to published protocols (*30*, *31*). All samples were analyzed in positive ion mode using a standard 337 nm nitrogen laser, and the TOF was measured over 256 laser pulses and averaged into a single spectrum. Each spectrum was calibrated either externally or internally using standards purchased from PE Biosystems (Foster City, CA), as described below. A range of ±150 ppm error was set as the criterion for the mass peaks to be analyzed based upon the calibration data. Furthermore, mass values were highly reproducible for all the digests in which multiple spectra were generated (on the average less than ±0.5 Da). Full-length rhHMGA1 proteins and purified in vivo modified HMGA1 proteins were mixed with a saturated matrix solution consisting of 3,5-dimethoxy-4-hydroxy-cinnamic acid, 50% acetonitrile/0.2% TFA. The HMGA1a protein digestions were mixed with α-cyano-4-hydroxy-cinnamic acid (α-CN) matrix solution and calibration mixture (Cal Mix) 3 and analyzed by MALDI-TOF MS. Both matrixes were purchased from Sigma Chemical Co. (St. Louis, MO). To maximize peak detection, three different scans were performed on every phosphorylated and dephosphorylated in vivo modified HMGA1a digested sample.

MALDI-TOF MS Calibration. Prior to analysis, the MALDI-TOF mass spectrometer was calibrated using a combination of Cal Mix 1, 2, and 3 (PE Biosystems) to determine the best parameters for data analysis and margin of error between the average calculated and the observed masses. The instrument settings of noise filter "one" and Gaussian smooth "23" gave the best peak definition over the range of analysis (400−5000 *m*/*z*) needed for the rhHMGA1a peptide samples. Internal calibrations were conducted on all individual peaks and different combinations of peaks present in the calibration mixtures. The 379.35 *m*/*z* α-CN matrix dimer and the 5734.59 *m*/*z* bovine insulin peaks gave the best ppm errors for both the Cal Mix 1−3 spectrum and the spectrum of rhHMGA1a tryptic digestions combined with Cal Mix 1−3 (see Figure 1 and Table 1). These conditions were chosen for the first "rough" internal calibration. Further fine-tune calibration of the spectra was accomplished by calibrating on multiple unmodified HMGA1a tryptic or Arg-C peptides. From analyzing these data (Table 1), we determined that the stringent conditions from −150 ppm through +150 ppm error would eliminate peaks resulting from possible extraneous contaminations. In conclusion, every MALDI-TOF mass spectrum of enzyme-digested HMGA1a proteins presented in this paper was analyzed exactly the same way: noise removal of one, Gaussian smooth 23, and internal calibration with the α-CN matrix dimer (379.35 *m*/*z*) and bovine insulin (5734.59 *m*/*z*) mass peaks as rough internal reference standards. Fine-tune internal calibrations were conducted on unmodified HMGA1a peptide mass peaks.

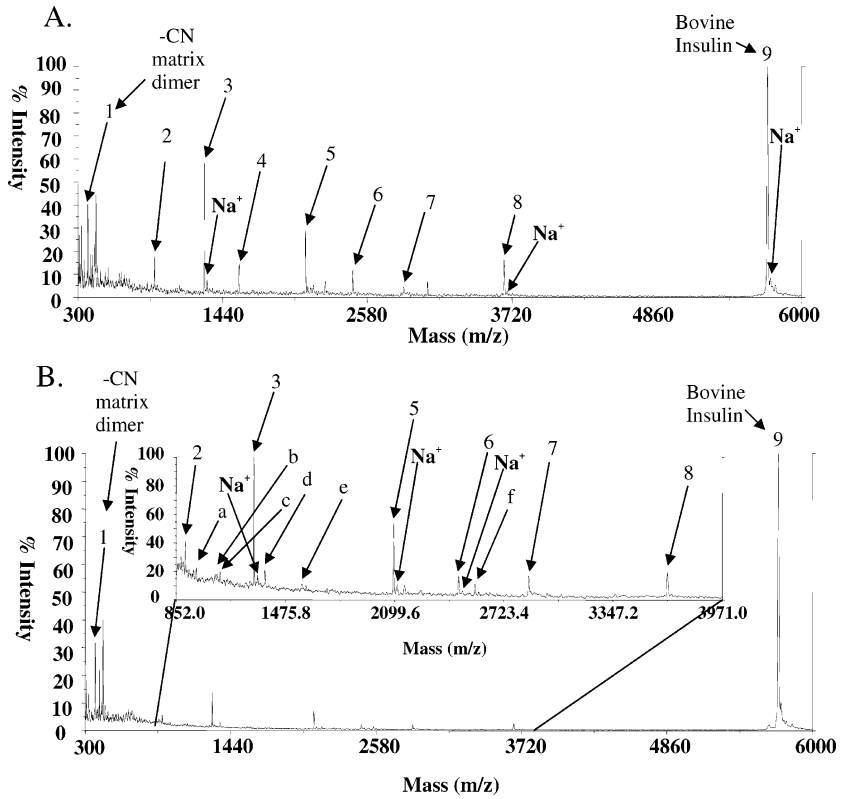

FIGURE 1: Calibration scans conducted on MALDI-TOF mass spectrometer with calibration mixtures and rhHMGA1a: (A) spectrum of calibration mixtures 1−3; (B) spectrum of calibration mixtures 1−3 combined with rhHMGA1a trypsin-digested protein. Both data sets were optimized with conditions as follows: Gaussian smooth 23, noise filter 1, and calibrated on the matrix dimer peak of 379.35 *m*/*z* and the bovine insulin peak of 5734.59 *m*/*z*. Masses of peaks and identified HMGA1a peptide fragments indicated by letters are recorded in Table 1. Calibration mixture peaks are indicated by numbers. Sodium adducts are indicated by "Na^+".

Data Analysis. Mass-to-charge ratios (*m*/*z*) for proteins and peptides were analyzed using Data Explorer, version 5.1 (PE Biosystems). The strategies of the subsequent analyses were as follows. The peak masses were copied from the Data Explorer program and pasted into an Excel (Microsoft, Redmond, WA) spreadsheet for further analysis. Known peptide peaks resulting from autodigestion of trypsin (http://prospector.ucsf.edu/ucsfhtml4.0/misc/trypsin.htm) were removed from each tryptic spectrum prior to peptide analysis. In addition, the peak masses from the MALDI-TOF spectrum being analyzed were pasted into an Excel spreadsheet on both the vertical axis and horizontal axis to produce a two-dimensional matrix. Each cell on the spreadsheet matrix contained the formula (= B$1 − $A2) that calculated mass differences between adjacent spectrum peaks. Mass differences from adjacent peaks that fell between 21.5 and 22.5 mass range, corresponding to possible sodium adducts, were removed from the peptide analysis. Additionally, human keratin peptides that contaminated the preparations were accounted for by using the EXPASY FindPept tool (www.expasy.ch), potential keratin masses being removed from further analysis. Peak masses remaining after removal of the possible trypsin peptide peaks, sodium adducts, and contaminating human keratin peaks were further analyzed using the Protein Prospector MS-digest program (www.prospector.ucsf.edu), EXPASY FindMod tool (www.expasy.ch), and manual mass calculations. The average mass values used for the analysis of potential PTMs including methylation, acetylation, and phosphorylation were 14.03, 42.08, and 79.98, respectively. Only peaks that corresponded to our data range criterion of ±150 ppm were further analyzed.

Electrospray Tandem Mass Spectrometry. HMGA1 peptides derived from proteolytic digestions were separated using an Agilent 1100 capillary LC system with a 60-cm capillary column (150 μm i.d. × 360 μm o.d., Polymicro Technologies, Pheonix, AZ) packed with 5-μm C_{18} particles (PoroS 20R2, Applied Biosystems, Foster City, CA). Peptides were eluted at a flow rate of 1.8 μL/min using water/0.1% acetic acid/0.01% TFA and acetonitrile/0.1% acetic acid/0.01% TFA with a linear gradient from 10% to 60% acetonitrile. The capillary column flow was infused into a Thermo-Finnigan LCQ Deca XP ion trap mass spectrometer that was operated to optimize the duty-cycle length with the quality of data acquired. This optimization resulted in a duty cycle consisting of a single full MS scan followed by three MS/MS scans on the three most intense precursor masses (determined by Xcaliber mass spectrometer software in real time) from the single parent full ion scan. Dynamic mass exclusion windows of 3 min were used, and the MS spectra for all samples were measured with an overall mass/charge (*m*/*z*) window of 400−2000.

SEQUEST Analysis of Peptides. Resulting tandem mass spectra of selected parental ions were analyzed by SEQUEST (Bioworks 2.0 Thermo Electron GmbH, Bremen, Germany) (*31*−*36*). SEQUEST analyzes experimental spectra compared to predicted idealized mass spectra generated from a database of protein sequences. These idealized spectra are weighted largely with b and y fragment ions (i.e., fragmentation at the peptide bond from the N- and C-termini, respectively,

Table 1: Calibration of MALDI-TOF Mass Spectrometer of Data from Figure 1[a]

A. Calibration Mixtures 1−3 Were Combined with Matrix and Scanned with the MALDI-TOF MS

peaks	MALDI	calcd	diff. from expect	ppm	source
1	379.3	379.35	0.05	131.80	α-CN dimmer
2	904.9	905.05	0.15	165.74	des-Arg-bradykinin
3	1297.5	1297.51	0.01	7.71	angiotensin I
4	1571.5	1571.61	0.11	69.99	Glu-fibrinopeptide B
5	2094.4	2094.46	0.06	28.65	ACTH (clip 1−17)
6	2467.0	2466.72	−0.28	−113.51	ACTH (clip 18−39)
7	2867.4	2867.80	0.40	139.48	insulin bovine (+2)
8	3660.6	3660.19	−0.41	−112.02	ACTH (clip 7−38)
9	5734.6	5734.59	−0.01	−1.74	insulin bovine (+1)

B. Combined Calibration Mixtures 1−3 Were Mixed with rhHMGA1a Tryptic Digestion Products and Matrix and Then Scanned with the MALDI-TOF MS

peaks	MALDI	calcd	diff. from expect	ppm	HMGA1a trypic peptides
a	983.9	984.09	0.19	193.07	EPSEVPTPK
b	1045.5	1045.56	0.06	57.39	trypsin peak LSSPATLNSR
c	1115.1	1115.28	0.18	161.39	TTTTPGRKPR
d	1385.6	1385.53	−0.07	−50.52	DGTEKRGRGRPR
e	1594.9	1594.85	−0.05	−31.35	KQPPVSPGTALVGSQK
f	2560.0	2559.62	−0.38	−148.46	KQPPVSPGTALVG SQKEPSEVPTPK

Reference Masses for Spectrum B

peaks	MALDI	calcd	diff. from expect	ppm	internal calibrant peaks
1	379.4	379.35	−0.05	−131.80	α-CN dimer
9	5734.6	5734.59	−0.01	−1.74	insulin bovine (+1)

[a] Peak masses of the calibration mixtures were supplied by the manufacturer. From these tables, it was determined that the mass spectrometer produced accurate results through a range of ±150 ppm error.

at the amide linkages). The peptide mass tolerance was 1.0, the fragment ion tolerance was 0.0, and trypsin and Arg-C enzyme rules were applied during SEQUEST analysis. Manual analysis of spectra from singly through triply charged parental peptides was performed to verify SEQUEST analysis (*37*).

RESULTS

Diverse HMGA1a Protein Populations within Different Lines of Human MCF-7 Breast Cancer Cells. Three different human breast mammary epithelial MCF-7 cancer cell lines that phenotypically exhibit different degrees of malignancy were utilized to characterize in vivo HMGA1a protein PTMs. Cells of the parental MCF-7/Tet off line express only very low levels of endogenous HMGA1 proteins and, although they are neoplastically transformed and immortal, exhibit many phenotypic characteristics of normal mammary epithelial cells. For example, both in vitro and in vivo studies have demonstrated that, like normal cells, the MCF-7/Tet off cells do not readily grow in soft agarose nor do they form primary or metastatic tumors when injected into immune-compromised nude mice (*12*). In contrast, cells of the HA7C cell line (derived from these parental cells by stable transfection) that overexpress a transgenic HMGA1a isoform protein grow readily in soft agarose and also exhibit a moderately metastatic and invasive phenotype when injected into nude mice (*12*). On the other hand, cells of the HA8A cell line (also derived from these parental cells) that overexpress a transgenic HMGA1b isoform protein not only

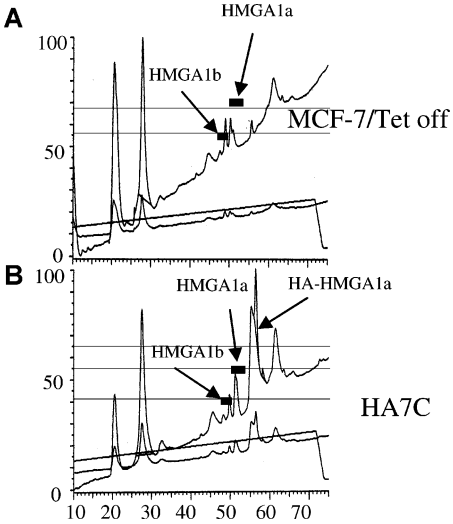

FIGURE 2: RP-HPLC profiles for MCF-7/Tet off and HA7C (HA-HMGA1a overexpressing) cells: (A) RP-HPLC profile of proteins extracted from MCF-7/Tet off cells (5×10^7) using TCA; (B) RP-HPLC profile of proteins extracted from HA7C cells (1.5×10^7) using TCA. Profiles are from C4 RP-HPLC scans of proteins isolated from the MCF-7 cell lines. The gradient was from 12% to 23% acetonitrile/0.2% TFA over 72 min and is indicated on the spectrum by the straight line. The lower protein profile on each scan is from the actual UV absorbance spectrum at 214 nm, and the upper protein profile on each scan is the absorbance normalized to 100% to allow for easier identification of fractions. Fractions that contain the HMGA1 proteins are indicated by solid bars. Note the predominance of the overexpressed transgenic HA-tagged HMGA1a and the increased expression of the endogenous HMGA1a protein in panel B.

grow readily in soft agarose but are also highly aggressive, metastatic, and invasive when injected into nude mice (*12*). For example, HA8A cell line transplants rapidly develop both primary and metastatic tumors in nude mice (∼1−2 months) whereas transplants of HA7C cells form primary tumors at a much lower frequency and take a considerably longer time to develop metastatic foci (*12*).

For our experiments, the endogenous HMGA1a proteins and the transgenic HA-tagged proteins were isolated from each of the cell lines and purified using RP-HPLC. Figure 2A shows a typical chromatographic profile of acid-soluble proteins isolated from the parental MCF-7/Tet off cell line and eluted from a C4 RP-HPLC column, indicating the peaks containing the endogenous native HMGA1a and HMGA1b isoform proteins. Figure 2B shows the RP-HPLC profile of acid-soluble proteins isolated from the HA7C cells that overexpress the transgenic HA-HMGA1a isoform protein. Similar profiles are also obtained with protein extracts from HA8A cells except that, in this case, the overexpressed transgenic HA-HMGA1b isoform protein elutes at a slightly different position on the chromatograph (data not shown). In all cases, however, the chromatographic fractions containing the endogenous native HMGA1 proteins (indicated by the black bars in Figure 2) were collected, and the identity of the proteins was further verified by SDS−PAGE and Western blotting analysis employing anti-HMGA1 specific antibodies (Figure 3). As shown in Figure 3A, SDS−PAGE analyses indicated that the native HMGA1a and HMGA1b proteins eluted from the C4 RP-HPLC column next to each other, as well as in overlapping chromatographic fractions. Therefore, to achieve better separation of these endogenous isoform proteins from each other, the HMGA1a fractions

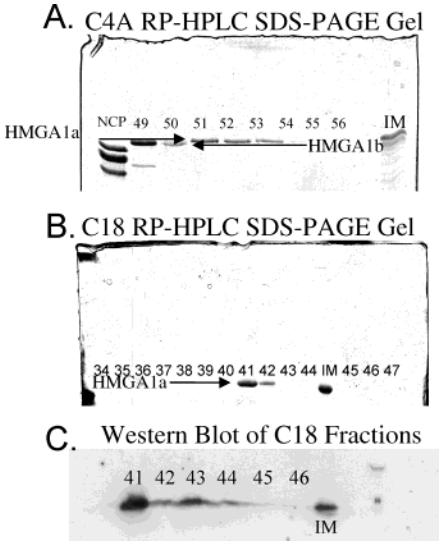

FIGURE 3: Analysis of RP-HPLC fractions by SDS−PAGE and Western blotting. Panels A and B show Coomassie Blue stained SDS-gels, whereas panel C shows the results of a Western blot using an anti-HMGA1 specific antibody. Panels B and C are two separate gels covering the same range of fractions from the C18 RP-HPLC column. In panel A, fractions collected from HA7C cells were isolated by running over a C4 RP-HPLC column and analyzed by SDS−PAGE. Fractions 49 and 50 show the presence of HMGA1b proteins, whereas fractions 51−54 contain HMGA1a proteins. The HMGA1a protein in fractions 51−54 were combined and subsequently eluted through a C18 RP-HPLC column to complete protein purification. In panel B, fractions collected from C18 RP-HPLC were analyzed by SDS−PAGE, and fractions 41−43 were found to contain highly purified (>95% pure) HMGA1a protein. Panel C shows Western Blot analysis confirming the presence of HMGA1a protein in fractions 41−46 eluted from the C18 RP-HPLC column. Fractions 41−45 were collected and pooled for analysis by mass spectrometry. The Western blot film was intentionally overexposed to identify all chromatographic fractions that contained HMGA1a proteins so as to minimize any potential loss of PTM proteins eluting in fractions with relatively low protein concentrations. The key is as follows: IM, purified recombinant human HMGA1a (rhHMGA1a) marker protein; NCP, marker nucleosome core particle histones isolated from chicken erythrocytes.

isolated by elution from the C4 RP-HPLC column were subsequently combined and further purified by a second fractionation on a C18 RP-HPLC column (Figure 3B). The HMGA1a-containing C18 RP-HPLC fractions were assessed by SDS−PAGE analysis (Figure 3B) and Western blotting procedures (Figure 3C), and the highly purified (>95%) HMGA1a positive fractions were combined for analysis by mass spectrometry.

When the HMGA1a Western blot-positive fractions from the C18 column were analyzed by MALDI-TOF MS, they were found to contain a number of the posttranslationally modified forms of HMGA1a proteins (Figure 4A). A broad mass range (11 545−12 300) for the ions was observed for the RP-HPLC fractionated full-length HMGA1a proteins isolated from the parental MCF-7/Tet off cell line (Figure 4A), corresponding to the many diverse posttranslationally modified forms of the HMGA1a protein present in these cells ranging from unmodified to extensively modified species.

In Vivo PTMs on Full-Length HMGA1a Proteins. MALDI-TOF MS analysis of purified in vivo modified HMGA1a proteins revealed a complex population of PTMs (Figure 4A). For example, both unmodified (11 545.05 Da) and a singly

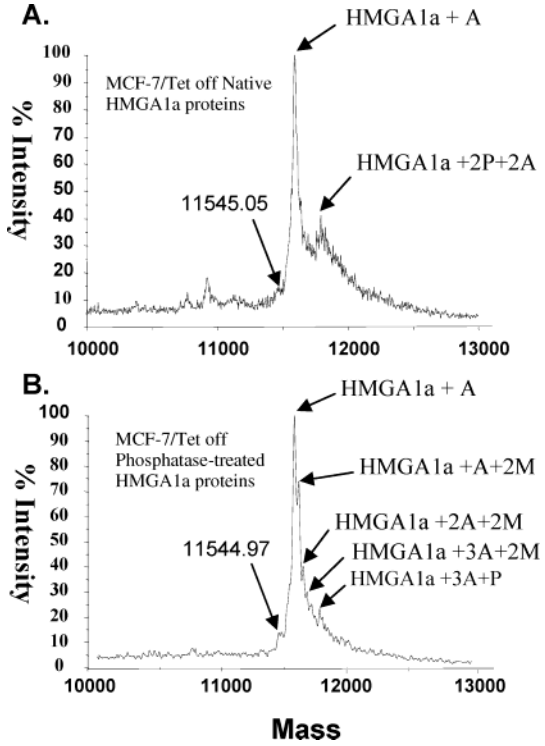

FIGURE 4: MALDI-TOF MS analysis of phosphorylated and dephosphorylated in vivo modified HMGA1a proteins from MCF-7/Tet off cells. Panel A shows the spectrum of highly purified (>95% pure) full-length HMGA1a proteins isolated from the MCF-7/Tet off cell line. The completely unmodified form of the protein (indicated by the peak at 11.54 kDa) does not contain the N-terminal acetylation commonly seen on most in vivo HMGA1 proteins. The largest peak corresponds to the HMGA1a protein with a single acetylation, most likely N-terminal acetylation. The large shoulder or "hill" formation observed on the spectrum corresponds to the diverse population of posttranslationally modified HMGA1a proteins present in the cell line. A few of the modifications have been tentatively identified by subtracting the mass of the unmodified peak from that of the modified peak and correlating mass differences to the mass of various known modifying groups. Panel B shows the spectrum of HMGA1a proteins isolated from MCF-7/Tet off cells that have been dephosphorylated in vitro by digestion with shrimp alkaline phosphatase. Note the shift to lower masses of the population of proteins found in the shoulder of the spectrum following phosphatase treatment. Some of the easily identified (due to unambiguous mass differences) modifications present on these dephosphorylated HMGA1a proteins are indicated in the figure.

acetylated form of the HMGA1a protein were detected on the spectrum, as well as what appeared to be more highly modified protein forms that contained both acetate and phosphate moieties. To investigate this possibility, protein samples were dephosphorylated by SAP treatment and analyzed by MALDI-TOF MS to visualize the PTMs that occur in addition to phosphorylation (Figure 4B). The spectrum of protein masses following dephosphorylation exhibited a marked shift toward the less modified forms, demonstrating that a significant amount of in vivo HMGA1a phosphorylation occurs in the parental MCF-7/Tet off cells. By measurement of mass differences between identified peaks, some of the PTMs present on full-length HMGA1a proteins were tentatively assigned (Figure 4B). This point is illustrated by the peak that appears to be corresponding to an acetylated and dimethylated form of HMGA1a that is present in the dephosphorylated HMGA1a sample (Figure 4B) but is absent from the endogenous in vivo modified HMGA1a sample

Table 2: MALDI-TOF MS Peaks of Arg-C-Digested HMGA1A Proteins from HA7C Cells[a]

A. Unmodified HMGA1a Arg-C Peptide Peaks From Figure 5A,B

peak	obsd mass	calcd mass	ppm error	location	sequence
a	1243.4	1243.45	40.21	73−83	KTTTTPGRKPR
b	1556.6	1556.81	134.89	58−72	GRPKGSKNKGAAKTR
c	2676.8	2676.81	3.74	84−106	GRPKKLEKEEEEGISQESSEEEQ
d	2969.4	2969.41	3.37	30−57	KQPPVSPGTALVGSQKEPSEVPTPKRPR
e	3058.3	3058.29	−3.27	81−106	KPRGRPKKLEKEEEEGISQESSEEEQ
f	3182.8	3182.65	−47.13	30−59	KQPPVSPGTALVGSQKEPSEVPTPKRPRGR
g	3217.4	3217.48	24.86	(1−29)	SESSSKSSQPLASKQEKDGTEKRGRGRPR
h	3650.1	3649.2	−246.63	24−57	GRGRPRKQPPVSPGTALVGSQKEPSEVPTPKRPR
i	3901.2	3901.24	10.25	73−106	KTTTTPGRKPRGRPKKLEKEEEEGISQESSEEEQ
j	4507.1	4507.2	22.19	30−72	KQPPVSPGTALVGSQKEP···KRPRGRPKGSKNKGAAKTR
k	5187.0	5186.99	−1.93	24−72	GRGRPRKQPPVSPGTALV···KRPRGRPKGSKNKGAAKTR

B. Modified Native HMGA1a Arg-C Peptide Peaks from Figure 5B

peak	obsd mass	calcd mass	ppm error	location	mod	sequence
1	1271.2	1271.51	243.80	73−83	2M	KTTTTPGRKPR
2	1299.6	1299.56	−30.78	73−83	A+M	KTTTTPGRKPR
3	1591.2	1591.43	144.52	73−84	4P+2M	KTTTTPGRKPR
4	1678.6	1678.87	160.82	58−72	P+A	GRPKGSKNKGAAKTR
5	1734.8	1734.98	103.75	58−72	P+2A+M	GRPKGSKNKGAAKTR
6	2931.3	2931.36	20.47	58−83	P+A+2M	GRPKGSKNKGAAKTRKTTTTPGRKPR
7	3409.5	3409.68	52.79	(1−29)	P+2A+2M	SESSSKSSQPLASKQEKDGTEKRGRGRPR
8	3513.7	3513.53	−48.38	(1−29)	3P+A+M	SESSSKSSQPLASKQEKDGTEKRGRGRPR
9	3915.3	3915.27	−7.66	73−106	M	KTTTTPGRKPRGRPKKLEKEEEEGISQESSEEEQ
10	3985.0	3985.40	100.37	73−106	2A	KTTTTPGRKPRGRPKKLEKEEEEGISQESSEEEQ

C. Modified Transgenic HA-HMGA1a Arg-C Peptide Peaks from Figure 5C

peak	obsd mass	calcd mass	ppm error	location	mod	sequence
I	1270.9	1271.51	479.74	73−83	2M	KTTTTPGRKPR
II	1403.4	1403.49	64.13	(1−12)[b]	HA tag	GYPDVPDQASRR[b]
III	1961.0	1961.10	50.99	(1−18)[b]	HA tag	GYPDVPDQASRRASVSGR[b]
IV	3772.7	3772.9	53.01	(1−15)[b]	HA+2P+A+2M	ASVSGREF[b]SESSSKSSQ...QEKDGTEKRGR
V	3980.7	3980.99	72.85	73−106	P	KTTTTPGRKPRGRPKKLEK...GISQESSEEEQ
VI	4061.3	4061.19	−27.09	73−106	2P	KTTTTPGRKPRGRPKKLEK...GISQESSEEEQ

[a] Data from Figure 5. [b] Sequence from the transgenic hemagglutinin tag.

(Figure 4A). This indicates that the acetylated and dimethylated HMGA1a is also likely phosphorylated in vivo.

MALDI-TOF MS of HMGA1a Peptides. The MALDI-TOF MS spectra of the peptides derived from enzymatic digestion of in vivo modified HMGA1a also revealed complex PTM patterns on the proteins. The spectrum obtained from Arg-C-digested, unmodified recombinant human HMGA1a protein (Figure 5A) contained fewer mass peaks than the spectra from both the Arg-C-digested native HMGA1a isolated from HA7C cells (Figure 5B) and the transgenic HA-tagged HMGA1a isolated from these same cells (Figure 5C). The increased complexity of the MS spectra shown in Figure 5B,C compared to that in Figure 5A is due to in vivo secondary biochemical modifications of many of the peptides derived from both the native and transgenic HMGA1a proteins. Importantly, however, all of the spectra contain a number of matching unmodified HMGA1a peptide mass peaks (Table 2; peaks a, c, and i), thus confirming the presence of unmodified HMGA1a peptides in proteolytic digests of both the native and transgenic HMGA1a protein populations. For example, mass analysis of peak a observed in the spectra of the recombinant (Figure 5A), native (Figure 5B), and transgenic (Figure 5C) protein reveals that it is the unmodified peptide fragment derived from amino acids 73−83 from the HMGA1a sequence (Table 2). This peptide is variously modified in vivo, giving rise to different peptide masses in the spectra of Arg-C-digested HMGA1a proteins from HA7C cells. As shown in Table 2 (peaks a, 2, 3), peptide 73−83 can be unmodified, acetylated, and methylated or have four phosphate groups plus two methyl groups attached. Peak 1 (Figure 5B) is a potential dimethylated form of peptide 73−83, but the ppm error falls outside the ±150 ppm selection criteria range (Table 2), so it was excluded from further analysis even though the mass difference between the calculated and observed values for this peak is only 0.31 Da. Likewise, peaks 4 and h (Table 2) were omitted from the final analysis since the ppm error for these masses was also outside our strict criterion for inclusion. By analyzing masses from the spectra of in vivo HMGA1a protein digestions, as shown in Figure 5, we were able to identify many of the types of PTMs on peptides of HMGA1a isolated from the three different MCF-7 cell lines. Interestingly, the PTM patterns on the transgenic HA-HMGA1a protein were found to be very similar to those present on the native endogenous HMGA1 proteins (unpublished observations), and therefore, additional data on it are not presented because the results would be redundant.

Confirmation of Phosphorylated HMGA1a Peptides. To verify that peptides derived from native proteins are indeed phosphorylated, we compared the spectra derived from in vivo modified HMGA1a peptides to corresponding spectra of in vivo modified samples that were enzymatically dephosphorylated prior to proteolytic Arg-C digestion (Figure 6 and Table 3). Confirmation of the presence of phosphate groups was necessary since phosphorylated peptides exhibit ionization suppression effects due to the negative charges associated with phosphate groups (*38*). A comparison of the spectra of Arg-C-derived peptides of HMGA1a proteins from

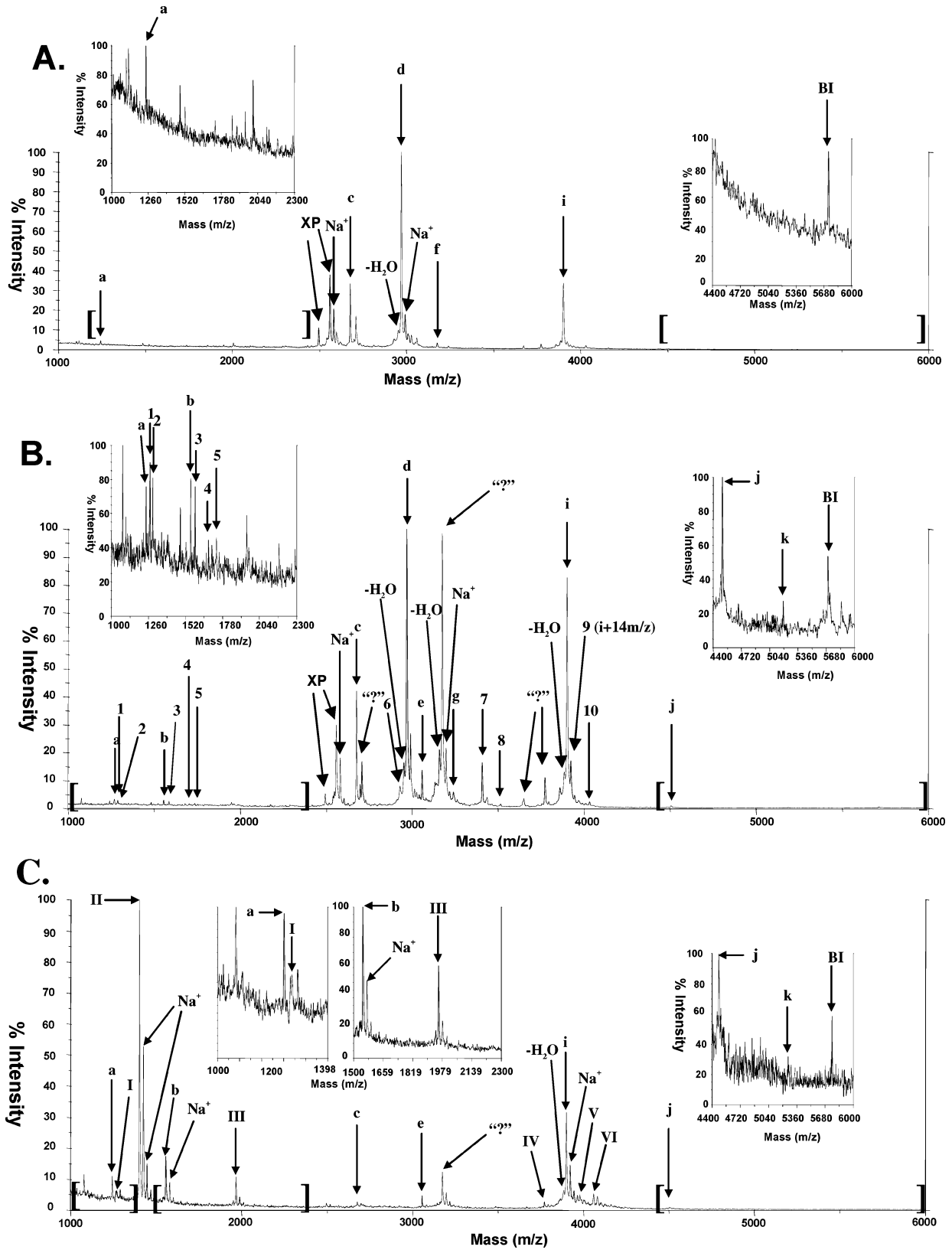

FIGURE 5: MALDI-TOF mass spectrum of Arg-C-digested HMGA1a protein from HA7C cells. Arg-C digest peaks corresponding to unmodified HMGA1a peptides are labeled with letters, and peaks corresponding to in vivo modified peaks are labeled with numbers. Enlarged insert regions are indicated with brackets on the main spectrum scan. The inserts are magnified segments of the main MALDI-TOF spectrum to better visualize minor mass peaks. Panel A shows the spectrum of unmodified recombinant HMGA1a protein. Panel B shows the spectrum of in vivo modified HMGA1a protein from HA7C cells. Panel C shows the spectrum of in vivo modified transgenic HA-tagged HMGA1a protein from HA7C cells. Data on positively identified peaks, indicated by letters or numbers, are also presented in Table 2. Peaks labeled with "XP" result from a nonstandard Arg-C cleavage of a C-terminal arginine before a proline residue (*27*) and are therefore not included in our analyses. Sodium adducts are indicated by "Na^+", and water losses are indicated by "$-H_2O$". The bovine insulin peak from calibration mixture 3 is illustrated by "BI" on the spectrum.

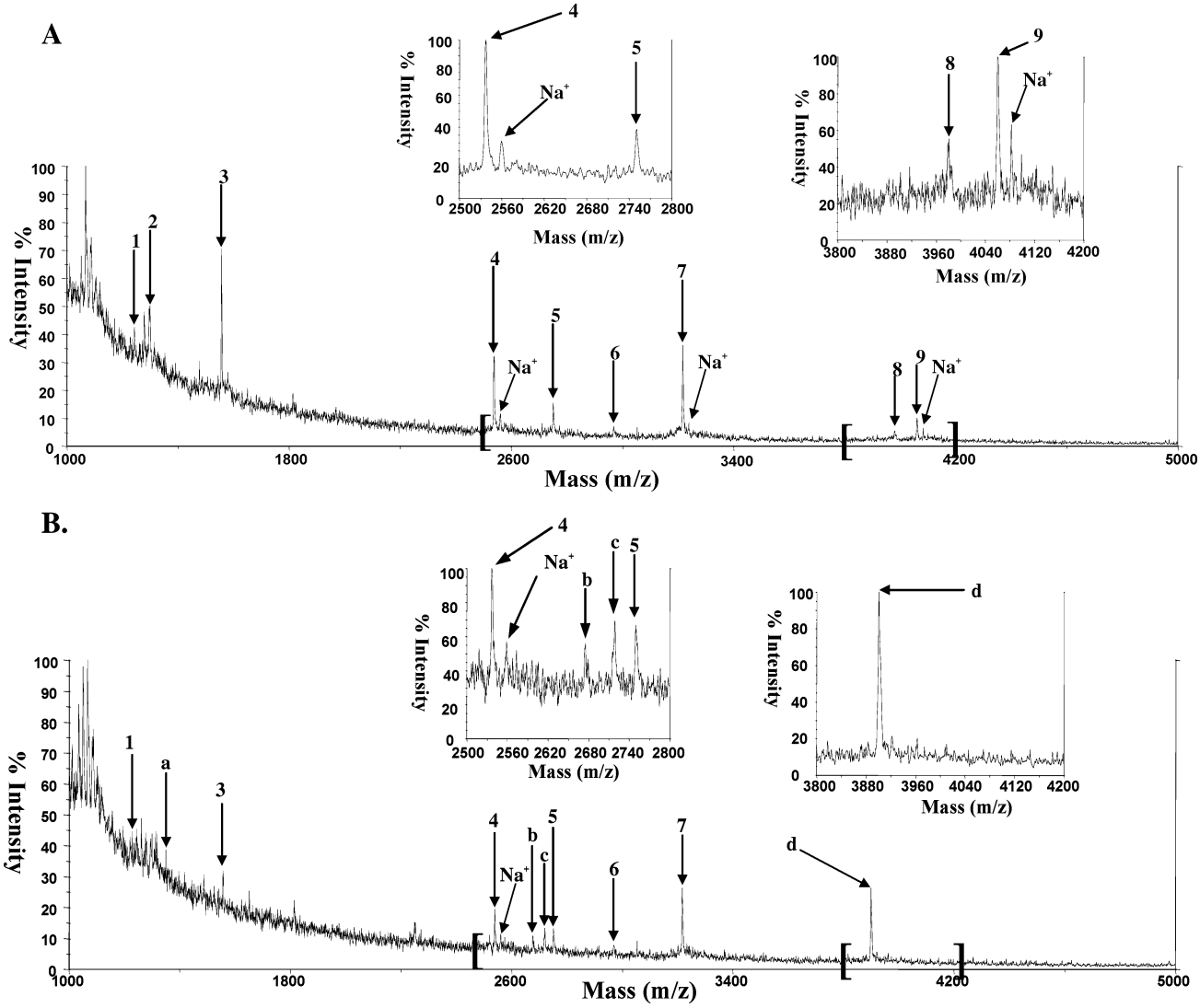

FIGURE 6: MALDI-TOF MS of Arg-C-digested HMGA1a proteins (phosphorylated and dephosphorylated) from MCF-7/Tet off cells. Insert regions are indicated with brackets on the main spectrum scan. The inserts are magnified segments of the main MALDI-TOF spectrum to better visualize minor mass peaks. Panel A shows Arg-C peptide fragments of HMGA1a isolated from MCF-7/Tet off cells. Mass peaks indicated with numbers are listed in Table 3, and sodium adducts are identified by "Na^+". Panel B shows dephosphorylated Arg-C peptide fragments of HMGA1a isolated from MCF-7/Tet off cells. Peak d is the dephosphorylated fragment derived from native peaks 8 and 9 shown in panel A. New peaks (e.g., b and c) are also present in the dephosphorylated HMGA1a Arg-C sample, indicating that these peptides may be phosphorylated in the MCF-7/Tet off Arg-C sample, although the actual native phosphorylated peptides corresponding to these peaks are not evident in the particular spectrum show in panel A but were observed in other spectra (data not shown).

parental MCF-7/Tet off cells, before (Figure 6A) and after (Figure 6B) SAP treatment, revealed the expected presence of phosphopeptides. Peak d in Figure 6B, for example, corresponds to the unmodified C-terminal end of HMGA1a from amino acids 73−106 (Table 3). This nonphosphorylated peptide was not detected in the spectrum of the native MCF-7/Tet off HMGA1a peptide sample (Figure 6A), whereas signals were detected in the spectrum for the mono- and diphosphorylated forms of this peptide (peaks 8 and 9 Figure 6A, Table 3). The disappearance of peaks 8 and 9 and the appearance of peak d in these spectra confirm that the HMGA1a proteins isolated from MCF-7/Tet off cells are phosphorylated on either serine or threonine residues within the 73−106 peptide sequence. Furthermore, peaks 8 and 9 demonstrate that at least two different posttranslationally modified HMGA1a proteins were isolated from the MCF-7/Tet off cells, one form containing a single phosphate moiety on 73−106 (i.e., peak 8) and another form containing two phosphate moieties (i.e., peak 9) (Table 3). Of equal importance, we observed the appearance of peaks b and c in the spectrum of HMGA1a proteins that had been dephosphorylated prior to Arg-C digestion. Peak b corresponds to HMGA1a peptide 84−106 in an unmodified form, and peak c corresponds to the same peptide containing a single acetyl group (Table 3). The detection of these peaks in the spectrum of the dephosphorylated HMGA1a sample (Figure 6B) but not the spectrum of the native protein sample (Figure 6A) is a strong indication that these peptides were originally phosphorylated in vivo. Interestingly, the 84−106 peptide derived from Arg-C digestion is not detected in the native protein sample prior to SAP treatment (Figure 6A). A likely explanation for this is that the dephosphorylated HMGA1a protein is cleaved more efficiently by the Arg-C protease than is the phosphorylated native protein, thus resulting in new mass peaks appearing in the spectrum. Furthermore, it is important to note that from just these two spectra we achieved 100% peptide sequence coverage of the HMGA1a protein isolated from the parental MCF-7/Tet off cells (Table 3).

Table 3: MALDI-TOF MS Peak Identification of Arg-C Peptides from Native and SAP-Treated HMGA1a Proteins Isolated From MCF-7/Tet off Cells[a]

Native HMGA1a						
peak	obsd mass	calcd mass	ppm error	sequence	location	mod
1	1243.4	1243.45	40.21	KTTTTPGRKPR	73−83	
2	1299.5	1299.56	46.17	KTTTTPGRKPR	73−83	A+M
3	1556.8	1556.81	6.42	GRPKGSKNKGAAKTR	58−72	
4	2537.6	2537.70	39.41	SESSSKSSQPLASKQEKDGTEKR	(1−23)	A
5	2750.7	2750.94	87.24	SESSSKSSQPLASKQEKDGTEKRGR	(1−25)	A
6	2969.1	2969.41	104.40	KQPPVSPGTALVGSQKEPSEVPTPKRPR	30−57	
7	3217.4	3217.48	24.86	SESSSKSSQPLASKQEKDGTEKRGRGRPR	(1−29)	A
8	3980.7	3981.22	130.61	KTTTTPGRKPRGRPKKLEKEEEEGISQESSEEEQ	73−106	P
9	4060.8	4061.20	98.49	KTTTTPGRKPRGRPKKLEKEEEEGISQESSEEEQ	73−106	2P
SAP-Treated HMGA1a						
peak	obsd mass	calcd mass	ppm error	sequence	location	mod
1	1243.4	1243.45	40.21	KTTTTPGRKPR	73−83	
a	1313.5	1313.59	68.51	KTTTTPGRKPR	73−83	A+2M
3	1556.8	1556.81	6.42	GRPKGSKNKGAAKTR	58−72	
4	2537.7	2537.74	0.23	SESSSKSSQPLASKQEKDGTEKR	(1−23)	A
b	2676.7	2676.8	37.36	GRPKKLEKEEEEGISQESSEEEQ	84−106	
c	2718.9	2718.89	−3.68	GRPKKLEKEEEEGISQESSEEEQ	84−106	A
5	2751.1	2750.98	−43.62	SESSSKSSQPLASKQEKDGTEKRGR	(1−25)	A
6	2969.4	2969.41	3.37	KQPPVSPGTALVGSQKEPSEVPTPKRPR	30−57	
7	3217.6	3217.53	−21.76	SESSSKSSQPLASKQEKDGTEKRGRGRPR	(1−29)	A
d	3901.1	3901.24	35.89	KTTTTPGRKPRGRPKKLEKEEEEGISQESSEEEQ	73−106	

[a] Data from Figure 6.

Likewise, the results illustrated in Figure 6 show how analysis of partial or incomplete peptide cleavage patterns by an enzyme can be used to confirm the presence of a particular PTM. For example, the sequences corresponding to peptides 1−23, 1−25, and 1−29 derived from Arg-C digestion were all detected containing a single acetyl group in both the native and SAP-treated samples (Table 3). Furthermore, these results also demonstrate why it was important to take at least three independent spectra for each experimental sample. The peak corresponding to HMGA1a peptide 73−83 containing an acetyl and a methyl group (i.e., A + M) was identified in the MCF-7/Tet off sample (peak 2, Figure 6A; Table 3) but was not listed as being present in the sample that was dephosphorylated with SAP (Figure 6B; Table 3). The reason for this is that not every MALDI-TOF MS spectrum will detect exactly the same constellation of peptides fragments in every scan with the same degree of mass accuracy. Although the residue 73−83 A + M peptide mass peak was present in the dephosphorylated SAP-treated sample, its ppm error was greater than our inclusion criteria of ±150 ppm, and it was therefore excluded from our analyses. Similarly, peak a shown in the SAP-treated sample (Figure 6B) corresponds to peptide 73−83 that contains one acetyl and two methyl groups (Table 3). However, this modified peptide is not listed in the non-SAP-treated native MCF-7/Tet off sample (Table 3) because, again, its mass error in the spectrum of that particular sample falls outside of our selection criteria used to identify peaks.

Types and Sites of in Vivo PTMs of HMGA1a. By analysis of the MALDI-TOF MS spectra of peptide fragments derived from incomplete tryptic or Arg-C proteolytic cleavages, combined with dephosphorylation digestions, we were able to construct detailed overlapping peptide maps that allowed us to identify many of the sites of in vivo PTMs on HMGA1a proteins isolated from the MCF-7/Tet off, the HA7C, and the HA8A cell lines (Figures 7 and 8). Furthermore, in addition to the identification of actual sites of in vivo secondary modifications, systematic analyses of these peptide maps exposed the dynamic nature of the PTMs on the HMGA1a proteins in vivo. For example, as illustrated in Figure 7, the population of N-terminal HMGA1a peptide fragments from HA7C cells exhibited a variety of PTMs ranging from a completely unmodified peptide through, in one case, a single peptide that contained three phosphates, an acetyl, and a methyl group and, in another case, a peptide that contained two phosphates, two acetyl, and two methyl groups. The N-terminus of HMGA1a is usually acetylated in vivo (*18*). Therefore, when a single acetyl group was found on an amino-terminal HMGA1a peptide fragment, it was tentatively assigned to the N-terminal serine residue. It is noteworthy, however, that in all three of the cell lines examined in this study, we also found N-terminal peptide fragments with no modification on this serine residue.

The in vivo, unmodified N-terminus was demonstrated by both the MALDI-TOF MS spectra of full-length proteins (Figure 4) and the detection of unmodified N-terminal peptide fragments after proteolytic digestion of native HMGA1a proteins (Figure 7). Furthermore, most proteolytic enzymes will not cleave at an N- or C-terminal amino acid that is posttranslationally modified (*39*, *40*), thereby ruling out acetylation and methylation occurring on the terminal ends of these peptide fragments. Therefore, to assign PTMs to specific amino acid residues, each segment of the peptide maps (Figure 7) had to be critically analyzed employing these considerations as guidelines. For example, we identified and analyzed more than 15 different N-terminal HMGA1a peptide fragments, spanning amino acids 1−30, exhibiting different levels of biochemical modifications on proteins isolated from the HA7C cell line (Figure 7; unpublished data). Among these, a peptide fragment corresponding to amino acid residues 6−17 was found that contained a single acetyl group, and since this region contains only one *internal* lysine at site 14, acetylation was assigned to this residue. Additional sites of acetylation in HMGA1a peptides 14−26 (P + 2A + 2M) and 6−30 (P + 3A + M) were similarly identified, thus allowing for assignment of acetyl groups to HMGA1a lysine residues 14, 17, and 22 (Figure 7). Indeed, similar types of analysis of all the multiple overlapping

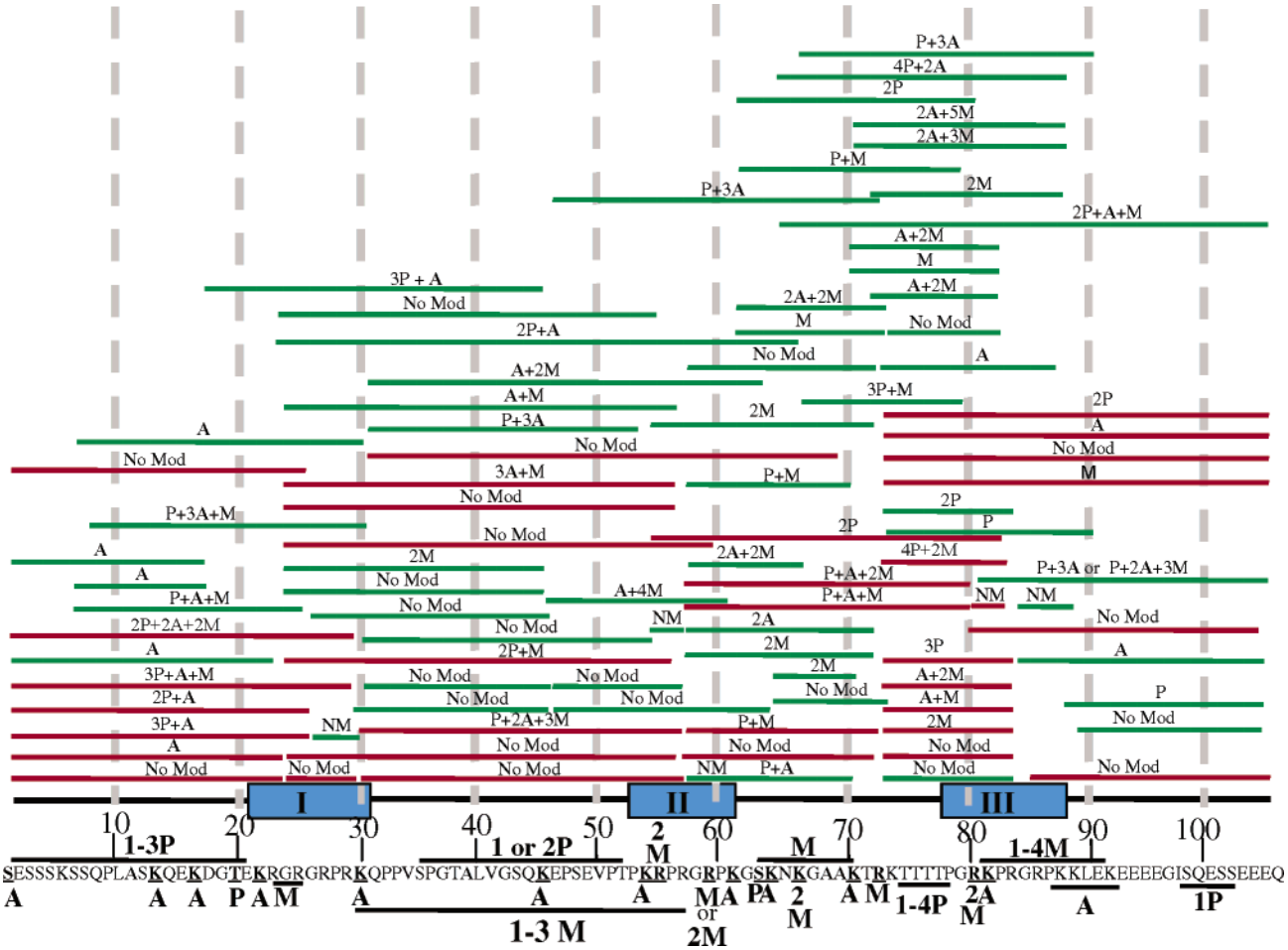

FIGURE 7: Overlapping peptide fragment map of HMGA1a proteins isolated from HA7C cells. The peptides were identified from multiple spectra, and the peak masses used to generate this peptide map are taken from Table 2 and other MALDI-TOF spectra (data not shown). The three AT-hook DNA-binding domains of the HMGA1a protein are designated by blue boxes and are labeled I, II, and III. The modified residues are designated by bold letters with the type of modification labeled by P, A, or M indicating phosphate, acetyl, or methyl groups, respectively. Unmodified peptides are designated by either NM (signifying no modification) or No Mod. Peptides resulting from partial Arg-C digestions are indicated by red bars and those originating from partial tryptic digestions by green bars. The types of PTMs detected are indicated above the peptide fragment bars. The black bars above or below the amino acid sequence denote peptide regions containing the indicated modifications but where no specific residue assignments could be made.

peptide fragments derived from each of the different areas of the HMGA1a protein (e.g., the N- and C-termini, as well as various internal peptide regions) enabled us to define many of the in vivo posttranslationally modified amino acid residues in proteins isolated from all three cell types (Figures 7 and 8). Nevertheless, due to the complexity of the in vivo patterns involved, some of the biochemical modifications identified by MALDI-TOF MS could not be unambiguously assigned to individual amino acid residues. These indeterminate modification sites for HMGA1a proteins isolated are indicated by black bars above or below the protein sequence in Figure 7 with the type of modifications that occur within the region indicated in bold letters.

Similar types of overlapping peptide mapping analyses for identification of PTMs were also performed on MALDI-TOF MS spectra of HMGA1a proteins isolated from both MCF-7/Tet off and HA8A cells (data not shown). Figure 8 summarizes, in diagrammatic form, the results of these PTM analyses for HMGA1a proteins isolated from MCF-7/Tet off, HA7C, and HA8A cells. From this figure, it is evident that both the types and patterns of in vivo PTMs found on the HMGA1a proteins isolated from each of these cell lines are very complex. When the biological importance of this figure is assessed, however, it is important to keep in mind that not all (or, in some cases, any) of these PTMs are likely to actually be present on any given individual HMGA1a protein inside a cell. Even though a PTM might be assigned to a specific amino acid residue in the figure, this assignment was based on analyses of peptide fragments derived from a heterogeneous population of variously modified full-length HMGA1a molecules present in each of the different cell lines. It is therefore likely that some of the HMGA1a proteins within the population contained the particular assigned modification while others did not. Consistent with this interpretation, it is also evident from this figure that although many of the PTMs are common to HMGA1a proteins in three cell lines, others are found on proteins in only two of the three lines, while a few (e.g., certain dimethylations indicated by shaded symbols above the stick diagrams in Figure 8) are observed on proteins in only one of the cell lines.

Identification of in Vivo HMGA1a PTMs by ESI-MS/MS. Given the number and complexity of PTMs observed on native HMGA1a proteins by MALDI-TOF MS, we deemed it necessary to confirm their existence by an independent analytical method. For this purpose, we employed electrospray ionization tandem mass spectrometry (ESI MS/MS)

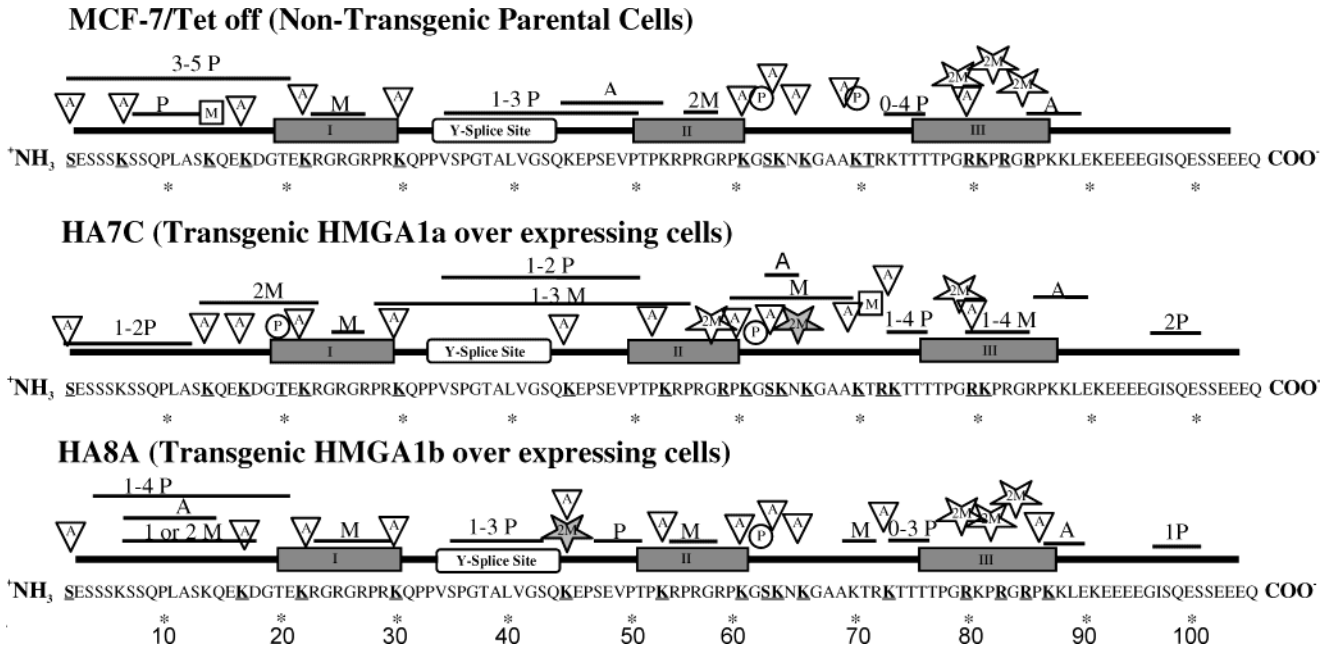

FIGURE 8: In vivo PTMs detected by MALDI-TOF MS on HMGA1a proteins isolated from MCF-7/Tet off, HA7C, and HA8A cells. Stick diagrams of the HMGA1a protein with the three DNA-binding (AT-hook) domains shown as shaded boxes (labeled I, II, and III) and the position of the Y splice site indicated by the clear box. Specific in vivo modified residues on proteins isolated from three different experimental cell lines are indicated in bold, underlined letters in the HMGA1a amino acid sequences beneath the diagrams, and the types of modifications are indicated by symbols (phosphorylations by circles, acetylations by triangles, methylations by squares, and dimethylations by stars) above the stick diagrams. Amino acid resides with unique modifications that are only found in one cell line have gray shaded symbols. Regions of the HMGA1a proteins containing PTMs identified by MALDI-TOF MS but without exact residue assignments are indicated by bars above the diagrams with the number and types of modifications shown over the bars. Every 10th amino acid residue of the HMGA1a sequence is denoted by ∗.

to determine the types and locations of PTMs on tryptic and Arg-C-derived peptides of HMGA1a proteins isolated from HA8A cells (Figure 9 and Table 4). ESI MS/MS spectra of peptides derived from HMGA1a proteins isolated from HA8A cells were analyzed using a combination of both the SEQUEST program (*33*−*35*) and manual calculations (*37*). Manual analysis was critical for these studies because it allowed for identification of ion fragments that the SEQUEST software does not account for, such as internal cleavages, water losses from serine and threonine residues, phosphate loss due to ionization, and other fragmentations such as a and z ions. In particular, the internal cleavages and phosphate and water losses detected in the spectra allowed us to verify that the parental peptides originally identified by the SEQUEST program were indeed correct matches (Figure 9).

Importantly, these ESI MS/MS analyses (Figure 9 and Table 4) not only confirmed many of the site-specific HMGA1a modifications that were previously found on HMGA1a peptide fragments by MALDI-TOF MS analysis (Figure 8) but also localized some of the PTMs to specific amino acid residues (Table 4) that MALDI-TOF analyses alone were unable to achieve. For example (Table 4), on one HMGA1a N-terminal peptide fragment two phosphates were identified (on Ser_1 and Ser_3), whereas on another peptide fragment covering this same region four phosphate groups were identified (on Ser_5, Ser_7, Ser_{12}, and Thr_{20}), none of which had been definitively assigned by MALDI-TOF MS. Of equal significance, when a multiply modified N-terminal peptide fragment (amino acid residues 1−25) was analyzed by ESI MS/MS, approximately 98% of the major peaks were positively identified (Figure 9). As shown by the inserted panel in Figure 9, this N-terminal fragment contains two phosphates (on Ser_1 and Ser_3) and two acetyl moieties (on Lys_{14} and Lys_{22}), the latter two modifications having previously been localized on these same amino acid residues by MALDI-TOF MS (Figure 8). These and other ESI MS/MS analyses therefore validate and extend the types and sites of in vivo PTMs identified on the HMGA1a by MALDI-TOF MS.

PTMs and Cellular Phenotypes. HMGA1a proteins from the three experimental cell lines were analyzed for unique sites and types of PTMs to determine whether any biochemical modification patterns could be identified that might be related to the differing tumor-forming and metastatic potential of the lines. These analyses detected a previously unknown PTM on HMGA1a proteins, an arginine dimethylation, that occurs at one or more arginine residues within the third AT-hook of HMGA1a isolated from all three cell lines (e.g., on residue Arg_{80} in both the MCF-7/Tet off and HA7C cells and on residues Arg_{83} and Arg_{85} in HA8A cells; Figure 8 and Table 5). In addition to these arginine dimethylations, dimethylations were also found on specific arginine and lysine residues in other regions of proteins isolated from both the metastatic HA7C (e.g., Arg_{59} and Lys_{66}) and HA8A (residues Lys_{45}) cells but not on proteins isolated from the nonmetastatic parental MCF-7/Tet off cells (Figure 8 and Table 5). Thus, a distinctive feature of the endogenous HMGA1a proteins isolated from the metastatic HA7C and HA8A cell lines is that they are more highly methylated than proteins from the nonmetastatic MCF-7/Tet off parental line from which they were originally derived.

HMGA1a proteins isolated from the metastatic HA7C and HA8A cell lines are also more highly acetylated than proteins

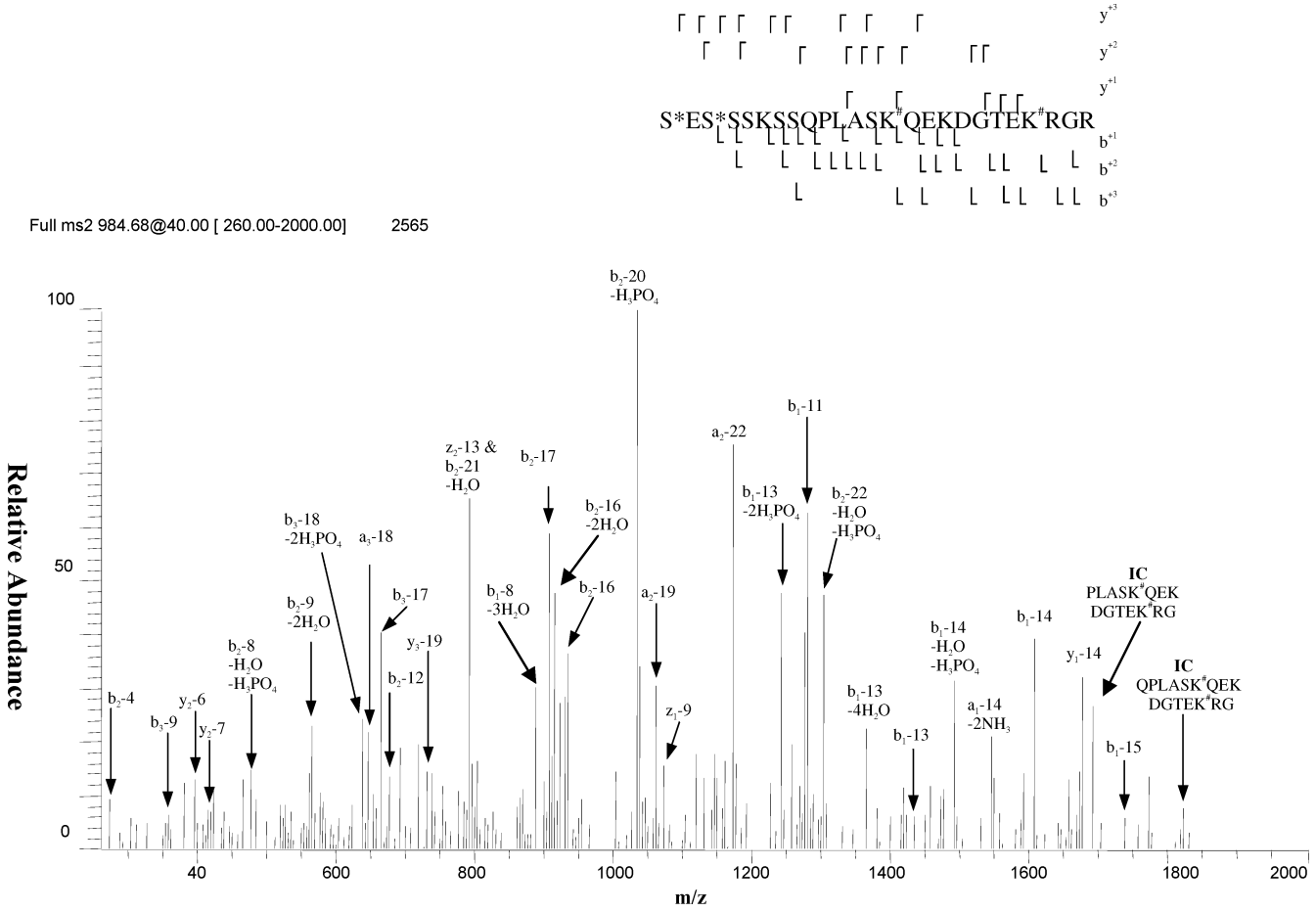

FIGURE 9: MS/MS ion fragmentation spectrum of a posttranslationally modified HMGA1a N-terminal parental peptide (1−25 amino acids) from the HA7C cell line. The HMGA1a parental peptide (mass of 2954.04) corresponding to amino acids 1−25 was selected for further analysis by ion fragmentation by the Xcaliber software. The fragmentation pattern shown above was first analyzed by SEQUEST software to select the b and y fragment ions (i.e., fragmentation of the peptide bond at the amide linkages from the N- and C-termini, respectively) and to match these ions to idealized mass spectra generated from a database of protein sequences. The idealized match, shown in the insert in the upper right-hand corner, displays the posttranslational modifications on the sequence. Serine residues that are phosphorylated are indicated with ∗, and lysine residues that are acetylated are indicated with #. The b and y ions that were identified from the fragmentation of the parental peptide are indicated with the symbols ⌊ and ⌈, respectively. The b^{+1}, b^{+2}, and b^{+3} and y^{+1}, y^{+2}, and y^{+3} ions arise from loss of charge as a result of fragmentation of triply charged parental peptides. The original spectrum resulting from the fragmentation was further analyzed by manual mass calculations to verify the b and y ions identified by SEQUEST and to subsequently identify peaks that were not assigned by this program. Ions that corresponded to water losses ($-H_2O$), phosphate losses ($-H_2PO_4$), ammonia losses ($-NH_3$), and internal cleavages (IC) could not be identified by SEQUEST and were therefore identified by manual mass analysis. Water is known to be lost from serine and threonine residues and phosphate moieties are known to be removed from serine/threonine residues under these fragmentation conditions.

from nonmetastatic MCF-7/Tet off cells (Figure 8 and Table 4). Although acetylation was detected on specific lysine residues in all of the AT-hook DNA binding motifs of HMGA1a proteins isolated from all three cell types, it is of considerable interest that the peptide region between the end of AT-hook II and the beginning of AT-hook III is more acetylated than in other parts of the protein, particularly in cells exhibiting more malignant phenotypes (Figure 8). For example, acetylation of Lys_{73}, located between AT-hooks II and III of HMGA1a proteins, was only observed in the more metastatic cell types (HA7C and HA8A) and not on proteins isolated from nonmetastatic MCF-7/Tet off cells. Additionally, Lys_{45} (which is located between AT-Hooks I and II and is not modified in the nonmetastatic MCF-7/Tet off cells) is acetylated in the moderately metastatic HA7C cells and *either* acetylated or dimethylated in proteins isolated from highly metastatic HA8A cells (Figure 8 and Table 4). These data demonstrate that HMGA1a proteins isolated from metastatic MCF-7 cells are more highly acetylated and methylated than proteins isolated from nonmetastatic MCF-7 cells and, likewise, indicate that these modifications appear to be residue-specific and dependent on the cell type.

DISCUSSION

Higher Degree of PTM Complexity Identified for in Vivo Modified HMGA1a Proteins. A significant difference between the present study and all previous reports is that attempts were made to characterize the posttranslational modifications found on the entire heterogeneous population of endogenous HMGA1a proteins present within closely related MCF-7 mammary epithelial cells that exhibit markedly different phenotypic characteristics. The present study is also the first to employ a combination of both MALDI-TOF MS and ESI MS/MS techniques to identify both the types and sites of PTMs found on HMGA1 proteins in vivo. From these analyses, a more complete and detailed picture of the PTM patterns found on native HMGA1 has emerged, one that indicates that in vivo biochemical modifications

Table 4: Electrospray MS/MS Analysis of in Vivo PTMs on Tryptic and Arg-C Peptides of HMGA1a Proteins Isolated from HA8A Cells[a]

postion	sequence
73−83	R.KTTTTPGRKPR.G
1−25	SESSS*KS*QPLAS*KQEK#DGT*EK#RGR.G
30−57	R.KQPPVSPGTALVGSQKEPSEVPTPKRPR.G
58−83	R.GRPKGSKNKGAAKTRKT*T*TTPGRKPR.G
58−80	R.GRPKGSKNK#GAAKTRK#T*TTT*PGR.K
30−57	R.KQPPVSPGT*ALVGSQK%%EPSEVPTPKRPR.G
26−57	R.GRPRKQPPVSPPGTALVGSQK#EPSEVPTPKR%PR.G
73−88	R.K#TT*TTPGRKPRGRPK#K.L
18−44	K.DGTEK#RGRGEPRKQPPVS*PGTALVGSQK.E
31−61	K.QPPVSPGTALVGS*QK#EPS*EVPTPK#RPRGRPK.G
84−106	R.GRPK#KLEKEEEEGISQESSEEEQ
1−25	S*ES*SSKSSQPLASK#QEKDGTEK#RGR.G

HMGA1a Sequence

SESSSKSSQPLASKQEKDGTEKRGRGRPRKQPPVSPGTALVGSQ
KEPSEVPTPKRPRGRPKGSKNKGAAKTRKTTTTPGRKPRGRP
KKLEKEEEEGISQESSEEEQ

[a] Tandem mass spectrometry analysis of HMGA1a tryptic and Arg-C digestions. The HMGA1a sequence is shown in the lower section of the table. Key: * = phosphate group; # = acetyl group; % = methyl group; %% = dimethyl group.

found on these proteins are much more complex and labile than previously thought. The present study not only characterizes several new sites of phosphorylation, acetylation, and monomethylation on the HMGA1 proteins but also identifies two previously unreported types of modification (dimethylation) on specific arginine and lysine residues. Furthermore, the present data suggest that many of the modifications and PTM patterns present on HMGA1 proteins in vivo are labile and change with the physiological or phenotypic state or both of cells. Importantly, however, these results also confirm the existence and sites of many of the previously reported biochemical modifications on HMGA1a proteins found by others in various cell types (*16*−*19*, *22*, *25*−*27*). An apparent exception is that, in contrast to a number of previous reports indicating that serine residues (i.e., Ser_{98}, Ser_{101}, and Ser_{102}) in the acidic carboxyl tail of HMGA1 proteins are constitutively phosphorylated in vivo (*20*, *21*, *23*, *27*, *42*), we did not observe this feature in any of the MCF-7 cell lines studied. In fact, in these particular cell lines, phosphorylation of the acidic tail appeared to be dynamic, varying from zero to two phosphates, on these serine residues (Figure 8). The reason(s) for this difference in phosphorylation patterns in MCF-7 cells compared to other cell types is unknown but highlights the labile and cell-specific nature of in vivo PTMs on HMGA1 proteins.

Earlier studies have demonstrated that dynamic in vivo phosphorylations and acetylations of specific amino acid residues modulate the DNA-binding properties of HMGA1 proteins and, as a consequence, influence their ability to participate in gene regulatory activities and other cellular functions (*4*). For example, phosphorylation of specific threonine residues in the AT-hook motifs by cdc2 kinase during certain stages of the cell cycle has been demonstrated to markedly reduce the DNA-binding affinity of HMGA1 proteins (*17*). Likewise, reversible acetylation of specific lysine residues (Lys_{65} and Lys_{71}) located between AT-hooks II and III, the region of HMGA1 proteins know to be the primary site for physical interaction with other transcription factors (*3*, *4*, *41*), has been shown to reversibly regulate transcription of the IFN-β gene by controlling the formation/destabilization of an enhancesome on its promoter region (*25*, *26*). Interestingly, in our studies, we also identified acetylation of both of these lysine residues (e.g., Lys_{64} and Lys_{70}) in the MCF-7/Tet off and HA7C cell lines but not in the HA8A cell line (Figure 8; note that our amino acid numbering system differs from that used in refs *25* and *26* since in these latter studies the first methionine residue of HMGA1a, which is normally removed in vivo, was designated as amino acid number one). It is therefore reasonable to suspect that the new modifications identified in this study, particularly the mono- and dimethylations, might also influence gene regulation in cells. This possibility seems likely given that the methylations occurs on all three of the DNA-binding AT-hook motifs of the HMGA1a protein (Figure 8), but whether these modifications actually affect DNA binding and the in vivo function of these proteins remains to be determined.

PTMs and Cancer. Developing effective methods for molecularly forecasting the malignant state of a tumor is a major medical goal, and many researchers are exploring gene expression microarray technology as a means to achieve this objective (*47*−*50*). Messenger RNA expression levels, however, do not always correlate with protein levels, and it has been demonstrated that when analysis of genomic information is measured against comparable information for protein expression, protein and mRNA levels often differ (*51*). Furthermore, DNA microarray analyses are not able to assess important posttranslational processes such as protein modifications or protein degradation (*52*−*54*). Identification of relevant protein markers for early diagnosis of cancer is vital. For HMGA1a proteins, the elevated expression levels that have been proposed to be diagnostic markers for both

Table 5: Amino Acid Sites and Types of in Vivo HMGA1A PTMs[a]

cell line	modification	residue(s) modified
MCF-7/Tet off	phosphorylation	S63, T71
(nonmetastatic)	acetylation	S1, K6, K17, K22, K30, K61, K64, K66, K70, K81
	monomethylation	K14
	dimethylation	R80, R83, R85
HA7C	phosphorylation	T20, S63
(moderately metastatic)	acetylation	S1, K14, K17,K22, K30, K45, K54, K61, K64, K70, K73, K81
	monomethylation	**R72**
	dimethylation	R59, **K66**, R80
HA8A	phosphorylation	S63
(highly metastatic)	acetylation	S1, K17, K22, K30, K45, K54, K61, K64, K66, K73, K87
	monomethylation	multiple sites (unassigned)
	dimethylation	**K45**, R80, R83, R85

[a] Summary of modifications from Figure 8. Bold amino acid residues (depicted by the amino acid residue followed by the sequence location) indicate unique modification sites.

neoplastic transformation and metastatic progression in a number of different types of naturally occurring tumors (reviewed in refs *3*, *4*, and *7*) may only be part of the picture relating to the role(s) of these proteins in cancer. Indeed, results from the present study suggest that certain types and patterns of posttranslational modifications on HMGA1 proteins may also be important features of cancer cells. Thus, as summarized in Table 5, HMGA1a proteins present in MCF-7 cells that are malignant and metastatic (e.g., HA7C and HA8A cells) are more highly acetylated and methylated than those in the nonmalignant/nonmetastatic parental cells from which they were experimentally derived (e.g., MCF-7/Tet off cells). Furthermore, HMGA1a proteins in both the HA7C and HA8A metastatic cell lines but not in the nonmetastatic MCF-7/Tet off line are dimethylated at specific amino acid residues. These findings are of potential biological interest because methylation patterns on other nuclear proteins (e.g., histones) have also been associated with neoplastic malignancies (*43*−*46*). Nevertheless, more studies need to be conducted on HMGA1a proteins from both nonmalignant and malignant tissue cell lines, as well as from naturally occurring tumors, before the biological significance of these PTM patterns can be critically assessed and their role, if any, in cancer established.

HMGA1a PTM Patterns and a Potential HMGA1a Biochemical Code. There are distinct similarities and differences between the PTMs of the HMGA1a proteins in all three-cell MCF-7 lines studied (Table 4). As noted above, HMGA1a proteins isolated from the parental, noninvasive MCF-7/Tet off cells have lower levels of acetylation and methylation than proteins isolated from the more metastatic HA7C and HA8A cell lines and, in addition, are also devoid of certain site-specific dimethylations found in these more malignant cell types (Figure 8 and Table 5). Likewise, lysine residue 45 was modified by either mono- or dimethylation in the HA7C and HA8A cells, respectively, but was not modified in the MCF-7/Tet off cells. Similarly, lysine residue 73 was acetylated in the metastatic HA7C and HA8A cells but not in the MCF-7/Tet off cells. These are only a few examples to illustrate the fact that the population of endogenous HMGA1a protein isolated from the different MCF-7 cell lines each exhibits what appear to be unique combinations or patterns of PTMs that could potentially reflect a biochemical "code" relating to the physiology or phenotype of the cells or both.

Identification of Potential Molecular "Switches". Lysines can undergo both methylation and acetylation making these residues potential sites for differential in vivo modification of proteins. For example, it has been shown that histones undergo alternate transitions between methylation and acetylation on specific lysine residues and these changes are correlated with altered biological processes (*55*, *56*). HMGA1a proteins are very lysine-rich, but until now, modifications of lysines alternating between acetylation and methylation have not been reported. This study demonstrates that the HMGA1a lysine 14 residue is methylated in the nonmetastatic MCF-7/Tet off cell line but is acetylated in the more aggressive HA7C cell line, providing evidence for a possible biochemical "switching mechanism" for HMGA1a. In addition, HMGA1a from the highly metastatic HA8A cell line also exhibited alternate modifications between acetylation and dimethylation of lysine 45 further supporting the possible biochemical "switch" theory. These cumulative results provide strong support for the existence of a dynamic HMGA1a protein "modification code" analogous to that proposed earlier for histone proteins (*55*, *56*), and they also persuasively suggest that this "code" may be expressed in different ways in normal and malignant cells.

ACKNOWLEDGMENT

We greatly appreciate and acknowledge Dr. Ying Li for generating the HMGA1 overexpressing MCF-7 cell lines. We thank Dr. Geoff Banks for stimulating our interest in HMGA1 posttranslational modifications and helping solve MALDI-TOF MS problems. Special thanks are likewise extended to Jennifer Adair, Gregory Dement, and Laurie Johnson for their critical reading of the manuscript.

REFERENCES

1. Bustin, M., and Reeves, R. (1996) High-mobility-group chromosomal proteins: architectural components that facilitate chromatin function, *Progress in Nucleic Acid Research and Molecular Biology* (Cohn, W. E., and Moldave, K., Eds.) Vol. 84, pp 35−100, Academic Press, New York.
2. Bustin, M. (1999) Regulation of DNA-dependent activities by the functional motifs of the high-mobility-group chromosomal proteins, *Mol. Cell. Biol. 19*, 5237−5246.
3. Reeves, R. (2001) Molecular biology of HMGA proteins: hubs of nuclear function, *Gene 277*, 63−81.
4. Reeves, R., and Beckerbauer, L. (2001) HMGI/Y proteins: flexible regulators of transcription and chromatin structure, *Biochim. Biophys. Acta 1519*, 13−29.
5. Friedmann, M., Holth, L. T., Zoghbi, H. Y., and Reeves, R. (1993) Organization, inducible-expression and chromosome localization of the human HMG-I(Y) nonhistone protein gene, *Nucleic Acids Res. 21*, 4259−4267.
6. Nagpal, S., Ghosn, C., DiSepio, D., Molina, Y., Sutter, M., Klein, E. S., and Chandraratna, R. A. (1999) Retinoid-dependent recruitment of a histone H1 displacement activity by retinoic acid receptor, *J. Biol. Chem. 274*, 22563−22568.
7. Tallini, G., and Dal Cin, P. (1999) HMGI(Y) and HMGI−C dysregulation: a common occurrence in human tumors, *Adv. Anat. Pathol. 6*, 237−246.
8. Ram, T. G., Reeves, R., and Hosick, H. L. (1993) Elevated high mobility group-I(Y) gene expression is associated with progressive transformation of mouse mammary epithelial cells, *Cancer Res. 53*, 2655−2660.
9. Holth, L. T., Thorlacius, A. E., and Reeves, R. (1997) Effects of epidermal growth factor and estrogen on the regulation of the HMG-I/Y gene in human mammary epithelial cell lines, *DNA Cell Biol. 16*, 1299−1309.
10. Liu, W. M., Guerra-Vladusic, F. K., Kurakata, S., Lupu, R., and Kohwi-Shigematsu, T. (1999) HMG-I(Y) recognizes base-unpairing regions of matrix attachment sequences and its increased expression is directly linked to metastatic breast cancer phenotype, *Cancer Res. 59*, 5695−5703.
11. Wood, L. J., Maher, J. F., Bunton, T. E., and Resar, L. M. (2000) The oncogenic properties of the HMG-I gene family, *Cancer Res. 60*, 4256−4261.
12. Reeves, R., Edberg, D. D., and Li, Y. (2001) Architectural transcription factor HMGI(Y) promotes tumor progression and mesenchymal transition of human epithelial cells, *Mol. Cell. Biol. 21*, 575−594.
13. Dolde, C. E., Mukherjee, M., Cho, C., and Resar, L. M. (2002) HMG-I/Y in human breast cancer cell lines, *Breast Cancer Res. Treat. 21*, 181−191.
14. Treff, N. R., Pouchnik, D., Dement, G. A., Britt, R. L., and Reeves, R. (2004) High mobility group A1a protein regulates Ras/ERK signaling in MCF-7 human breast cancercells, *Oncogene*, in press.
15. Baldassarre, G., Battista, S., Belletti, B., Thakur, S., Pentimalli, F., Trapasso, F., Fedele, M., Pierantoni, G., Croce, C. M., and Fusco, A. (2003) Negative regulation of BRCA1 gene expression

by HMGA1 proteins accounts for the reduced BRCA1 protein levels in sporadic breast carcinoma, *Mol. Cell. Biol. 23*, 2225−2238.

16. Elton, T. S., and Reeves, R. (1986) Purification and postsynthetic modifications of Friend erythroleukemic cell high mobility group protein HMG-I, *Anal. Biochem. 157*, 53−62.
17. Reeves, R., and Nissen, M. S. (1995) Cell cycle regulation and functions of HMG-I(Y), *Prog. Cell Cycle Res. 1*, 339−349.
18. Banks, G. C., Li, Y., and Reeves, R. (2000) Differential in vivo modifications of the HMGI(Y) nonhistone chromatin proteins modulate nucleosome and DNA interactions, *Biochemistry 39*, 8333−8346.
19. Xiao, D. M., Pak, J. H., Wang, X., Sato, T., Huang, F. L., Chen, H. C., and Huang, K. P. (2000) Phosphorylation of HMG-I by protein kinase C attenuates its binding affinity to the promoter regions of protein kinase C gamma and neurogranin/RC3 genes, *J. Neurochem. 74*, 392−399.
20. Palvimo, J., and Linnala-Kankkunen, A. (1989) Identification of sites on chromosomal protein HMG-I phosphorylated by casein kinase II, *FEBS Lett. 257*, 101−104.
21. Ferranti, P., Malorni, A., Marino, G., Pucci, P., Goodwin, G. H., Manfioletti, G., and Giancotti, V. (1992) Mass spectrometric analysis of the HMGY protein from Lewis lung carcinoma. Identification of phosphorylation sites, *J. Biol. Chem. 267*, 22486−22489.
22. Wang, D. Z., Ray, P., and Boothby, M. (1995) Interleukin 4-inducible phosphorylation of HMG-I(Y) is inhibited by rapamycin, *J. Biol. Chem. 270*, 22924−22932.
23. Diana, F., Sgarra, R., Manfioletti, G., Rustighi, A., Poletto, D., Sciortino, M. T., Mastino, A., and Giancotti, V. (2001) A link between apoptosis and degree of phosphorylation of high mobility group A1a protein in leukemic cells, *J. Biol. Chem. 276*, 11354−11361.
24. Fedele, M., Pierantoni, G. M., Berlingieri, M. T., Battista, S., Baldassarre, G., Munshi, N., Dentice, M., Thanos, D., Santoro, M., Viglietto, G., and Fusco, A. (2001) Overexpression of proteins HMGA1 induces cell cycle deregulation and apoptosis in normal rat thyroid cells, *Cancer Res. 61*, 4583−4590.
25. Munshi, N., Merika, M., Yie, J., Senger, K., Chen, G., and Thanos D. (1998) Acetylation of HMG I(Y) by CBP turns off IFN beta expression by disrupting the enhanceosome, *Mol. Cell 2*, 457−467.
26. Munshi, N., Agalioti, T., Lomvardas, S., Merika, M., Chen, G., and Thanos, D. (2001) Coordination of a transcriptional switch by HMGI(Y) acetylation, *Science 293(5532)*, 1133−1136.
27. Sgarra, R., Diana, F., Bellarosa, C., Dekleva, V., Rustighi, A., Toller, M., Manfioletti, G., and Giancotti, V. (2003) During apoptosis of tumor cells HMGA1a protein undergoes methylation: identification of the modification site by mass spectrometry, *Biochemistry 42*, 3575−3585.
28. Reeves, R., and Nissen, M. S. (1999) Purification and assays for high mobility group HMG-I(Y) protein function, *Methods Enzymol. 304*, 155−188.
29. Johnson, K. R., Lehn, D. A, and Reeves R. (1989) Alternative processing of mRNAs encoding mammalian chromosomal high-mobility-group proteins HMG-I and HMG-Y, *Mol. Cell. Biol. 9*, 2114−2123.
30. Chong, B. E., Lubman, D. M., Rosenspire, A., and Miller, F. (1998) Protein profiles and identification of high performance liquid chromatography isolated proteins of cancer cell lines using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, *Rapid Commun. Mass Spectrom. 12*, 1986−1993.
31. Mann, M., Hendrickson, R. C., and Pandey, A. (2001) Analysis of proteins and proteomes by mass spectrometry, *Annu. Rev. Biochem. 70*, 437−473.
32. Eng, J. K., Mccormack, A. L., and Yates, J. R. (1994) An Approach to Correlate Tandem Mass-Spectral Data of Peptides with Amino-Acid-Sequences in a Protein Database. *J. Am. Soc. Mass Spectrom. 5*, 976−989.
33. Yates, J. R., 3rd, Carmack, E., Hays, L., Link, A. J., and Eng, J. K. (1999) Automated protein identification using microcolumn liquid chromatography-tandem mass spectrometry, *Methods Mol. Biol. 112*, 553−569.
34. Yates, J. R., 3rd, McCormack, A. L., and Eng, J. K. (1996) *Analytical Chemistry News and Features*, pp 534−540, American Chemical Society, Washington, DC.
35. Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., Garvik, B. M., and Yates, J. R., 3rd (1999) Direct analysis of protein complexes using mass spectrometry, *Nat. Biotechnol. 17*, 676−682.
36. Washburn, M. P.,Wolters, D., and Yates, J. R., 3rd (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol. 19*, 242−247.
37. Kinter, M., and Sherman, N. E. (2000) *Protein Sequencing and Identification Using Tandem Mass Spectrometry*, pp 1−351, Wiley-Interscience, New York.
38. Liao, P. C., Leykam, J., Andrews, P. C., Gage, D. A., and Allison, J. (1994) An approach to locate phosphorylation sites in a phosphoprotein: mass mapping by combining specific enzymatic degradation with matrix-assisted laser desorption/ionization mass spectrometry, *Anal. Biochem. 219*, 9−20.
39. Rice, R. H., Means, G. E., and Brown, W. D. (1977) Stabilization of bovine trypsin by reductive methylation, *Biochim. Biophys. Acta 492*, 316−321.
40. Fraenkel-Conrat, H., Bean, R. S., and Lineweaver, H. (1949) Essential groups for the interaction of ovomucoid (egg white trypsin inhibitor) and trypsin, and for tryptic activity, *J. Biol. Chem. 177*, 385−402.
41. Reeves, R., and Edberg, D. D. (2004) HMGA proteins: multifaceted players in nuclear function, in *Chromatin Structure and Dynamics: State-of-the-Art*, (Zlatanova, J., and Leuba, S. H., Eds.) pp 155−180, Elsevier Science B. V., New York.
42. Giancotti, V., Bandiera, A., Buratti, E., Fusco, A., Marzari, R., Coles, B., and Goodwin, G. H. (1991) Comparison of multiple forms of the high mobility group I proteins in rodent and human cells. Identification of the human high mobility group I−C protein, *Eur. J. Biochem. 198*, 211−216.
43. Nguyen, C. T., Weisenberger, D. J., Velicescu, M., Gonzales, F. A., Lin, J. C., Liang, G., and Jones, P. A. (2002) Histone H3-lysine 9 methylation is associated with aberrant gene silencing in cancer cells and is rapidly reversed by 5-aza-2′-deoxycytidine, *Cancer Res. 62*, 6456−6461.
44. Rice, J. C., and Allis, C. D. (2001) Code of silence, *Nature 414*, 258−261.
45. Stallcup, M. R. (2001) Role of protein methylation in chromatin remodeling and transcriptional regulation, *Oncogene 20*, 3014−3020.
46. Fischle, W., Wang, Y., and Allis, C. D. (2003) Binary switches and modification cassettes in histone biology and beyond, *Nature 425*, 475−479.
47. van de Vijver, M. J., He, Y. D., van't Veer, L. J., Dai, H., Hart, A. A., Voskuil, D. W., Schreiber, G. J., Petersen, J. L., Roberts, C., Marton, M. J., Parrish, M., Atsma, D., Witteveen, A., Glas, A., Delahaye, L., van der Velde, T., Bartelink, H., Rodenhuis, S., Rutgers, E. T., Friend, S. H., and Bernards, R. (2002) A gene-expression signature as a predictor of survival in breast cancer, *N. Engl. J. Med. 347*, 1999−2009.
48. Van't Veer, L. J., and Weigelt, B. (2003) Road map to metastasis, *Nat. Med. 9*, 999−1000.
49. Guo, Q. M. (2003) DNA microarray and cancer, *Curr. Opin. Oncol. 15*, 36−43.
50. Schmidt, U., and Begley, C. G.(2003) Cancer diagnosis and microarrays, *Int. J. Biochem. Cell Biol. 35*, 119−124.
51. Pandey, A., and Mann, M. (2000) Proteomics to study genes and genomes, *Nature 405*, 837−846.
52. Anderson, L., and Seilhamer, J. (1997) A comparison of selected mRNA and protein abundances in human liver, *Electrophoresis 18*, 533−537
53. Mann, M. (1999) Quantitative proteomics? *Nat. Biotechnol. 17*, 954−955.
54. Srinivas, P. R., Srivastava, S., Hanash, S., and Wright, G. L., Jr. (2001) Proteomics in early detection of cancer, *Clin. Chem. 47* 1901−1911.
55. Strahl, B. D., and Allis, C. D. (2000) The language of covalent histone modifications, *Nature 403*, 41−45.
56. Hake, S. B., Xiao, A., and Allis, C. D., (2004) Linking the epigenetic 'language' of covalent histone modifications to cancer, *Br. J. Cancer. 90* (4), 761−769.

BI049833I