Collective Mass Spectrometry Approaches Reveal Broad and Combinatorial Modification of High Mobility Group Protein A1a

Nicolas L. Young,[a,*] Mariana D. Plazas-Mayorca,[b,*] Peter A. DiMaggio,[c] Ian Z. Flaniken,[a] Andrea J. Beltran,[a] Neeli Mishra,[a] Gary LeRoy,[a] Christodoulos A. Floudas,[c] and Benjamin A. Garcia[a,b]

[a] Department of Molecular Biology, Princeton University, Princeton, New Jersey, USA
[b] Department of Chemistry, Princeton University, Princeton, New Jersey, USA
[c] Department of Chemical Engineering, Princeton University, Princeton, New Jersey, USA

Transcriptional states are formed and maintained by the interaction and post-translational modification (PTM) of several chromatin proteins, such as histones and high mobility group (HMG) proteins. Among these, HMGA1a, a small heterochromatin-associated nuclear protein has been shown to be post-translationally modified, and some of these PTMs have been linked to apoptosis and cancer. In cancerous cells, HMGA1a PTMs differ between metastatic and nonmetastatic cells, suggesting the existence of an HMGA1a PTM code analogous to the "histone code." In this study, we expand on current knowledge by comprehensively characterizing PTMs on HMGA1a purified from human cells using both nanoflow liquid chromatography collision activated dissociation mediated Bottom Up and electron-transfer dissociation facilitated middle and Top Down mass spectrometry (MS). We find HMGA1a to be pervasively modified with many types of modifications such as methylation, acetylation, and phosphorylation, including finding novel sites. While Bottom Up MS identified lower level modification sites, Top and Middle Down MS were utilized to identify the most commonly occurring combinatorially modified forms. Remarkably, although we identify several individual modification sites through our Bottom Up and Middle Down MS analyses, we find relatively few combinatorially modified forms dominate the population through Top Down proteomics. The main combinatorial PTMs we find through the Top Down approach are N-terminal acetylation, Arg25 methylation along with phosphorylation of the three most C-terminal serine residues in primarily a diphosphorylated form. This report presents one of the most detailed analyses of HMGA1a to date and illustrates the strength of using a combined MS effort.   (J Am Soc Mass Spectrom 2010, 21, 960–970) © 2010 Published by Elsevier Inc. on behalf of American Society for Mass Spectrometry

Although the human genome has been sequenced for nearly a decade, the understanding of the epigenome, the set of factors superimposed over the DNA sequence that determine gene expression across cell types and developmental stages, remains incomplete [1]. A major locus for epigenetic regulation is the nucleosome, the basic unit of chromatin which consists of 146bp of DNA wrapped around an octamer of histone proteins [2]. Histones are subject to various post-translational modifications (PTMs), including methylation, acetylation, phosphorylation, and ubiquitination, which serve as markers for transcriptionally active or silent chromatin among other roles [3]. Proteins that recognize specific histone modifications bind histones to effect changes in chromatin structure. Heterochromatin Protein 1 (HP1), for example, binds selectively to trimethylated H3 lysine 9 and results in the formation of transcriptionally inert heterochromatin [4, 5].

For many chromatin-associated proteins, however, the mechanisms and consequences of interactions with nucleosomes are less clear. The high mobility group (HMG) proteins are well-characterized nuclear proteins termed "architectural transcription factors" because they associate with nucleosomes to modify chromatin structure and modulate specific gene expression [6]. HMG proteins are divided into three families based on their binding domains: HMGA, which contain AT-hooks and bind preferentially to the minor groove of DNA

in AT-rich regions; HMGB, which contain HMG box domains that bind to the minor groove of DNA with no known specificity; and HMGN, which bind directly to nucleosomes. Mammalian HMGA proteins include HMGA1a, HMGA1b, and HMGA2 [7]. HMGA1a and HMGA1b are simply transcript variants, where HMGA1a is the full length isoform. Among these, the role of HMGA1a in chromatin remodeling is especially intriguing because this protein is an oncoprotein that is only expressed at high levels in cancer cells [8–10] and during embryonic development [11, 12]. HMGA1a is a highly mobile protein that is found throughout the nucleus, however, it has been shown to be enriched in heterochromatin [13, 14]. The three AT-hook binding motifs of HMGA1a enable it to bind to the minor groove of DNA and alter chromatin structure [15, 16]. HMGA1a modifies chromatin structure by altering the rotational setting of DNA bound to nucleosomes [15]. Unlike most transcription factors, HMGA1a possesses the remarkable ability to bind DNA recognition sites that are wrapped around nucleosomes and uncover promoters for transcription factor binding [17]. Like histones, HMGA1a is subject to extensive PTMs, including phosphorylation, acetylation, and methylation [7]. Phosphorylation of serine and threonine residues are the most prevalent PTMs observed [18, 19]. Research suggests that PTMs on HMGA1a have functional implications on this protein's ability to remodel chromatin. Phosphorylation reduces HMGA1a DNA binding affinity, [20] and specific phosphorylation patterns correlate to higher residence in heterochromatin [21]. HMGA1a has also been found to be modified in human cancer cell lines [22, 23].

Recently, mass spectrometry (MS) based proteomic investigations have advanced our knowledge of HMGA PTMs. Sgarra et al. used MS coupled to reverse phase HPLC and found that increased arginine methylation of HMGA1a is associated with apoptosis and possibly with heterochromatin formation [24]. Studies using matrix-assisted laser desorption ionization (MALDI) with time-of-flight (TOF) MS alone or in combination with electrospray ionization (ESI) tandem mass spectrometry (MS/MS) identified differences in HMGA1a PTMs patterns in metastatic and non-metastatic cancer cell lines, suggesting the existence of a PTM code analogous to the histone code [23, 25]. Zou et al. utilized MS/MS to characterize sites of phosphorylation on HMGA1a from human prostate cancer cells; in addition to three distinct phosphorylation sites, Arg25 was found to be mono- and di-methylated [26]. Subsequent LC-MS investigations by this same group identified Arg 57 and Arg 59 as in vivo methylation sites as well [27]. MALDI-MS/MS together with LC-ESI-MS/MS was used to examine HMGA1a PTMs in PC-3 human prostate cancer cells that were either treated or not treated with a histone deacetylase inhibitor [28]. This study found Ser 35, Thr52, Thr77, Ser98, Ser101, and Ser102 to be phosphorylated while Arg25 was found to be methylated. Lys14 was found to be acetylated in the cells that

were not treated with butyrate [28]. The same instrumentation was used in a different investigation to find novel HMGA1a PTMs including Lys30me1, Lys54me1, Ser43phos, and Ser48phos, in addition to previously reported marks [29]. LC-MS/MS was also used to quantitate HMGA1a lysine acetylation [30]. Zou et al. found arginine methylation sites by the protein arginine methyltransferase family through MALDI-MS [31]. Lastly, LC-MS/MS was used to investigate HMGA protein expression and macroscopic PTM differences across different cell lines [32].

Nevertheless, while a significant number of studies have detailed a wide array of HMGA1a PTMs, a comprehensive investigation into potential combinatorial PTM patterns on HMGA has not been reported. In this study, we expand on current knowledge by comprehensively characterizing PTMs on HMGA1a purified from human HeLa S3 cells using Top and Middle Down as well as Bottom Up proteomic approaches. The Bottom Up approach enzymatically digests HMGA1a into short peptides before MS analysis and provided a sensitive platform to detect modifications of sub-stoichiometric abundance [33], while the Top and Middle Down approaches involve the analysis of large polypeptides and intact proteins, and are ideal for probing long-range PTM combinations [34, 35]. Using Bottom Up MS, we find HMGA1a to be extensively modified with all degrees of lysine methylation, lysine acetylation, arginine methylation and phosphorylation of threonine and serine residues at varying abundance levels including detecting some novel HMGA1a modifications. Middle and Top Down analyses revealed the identity of the most abundant combinations of modifications present on HMGA1a. The most abundant isoforms present in HeLa cells were determined to be N-terminally acetylated and di-phosphorylated at Ser98, Ser101, or Ser102. Only the Ser101phSer102ph form was detected by Bottom Up mass spectrometry, suggesting this may be the prevalent form. Additionally the overwhelmingly most abundant site of methylation was found to be Arg25. Un-, mono- and tri- phosphorylated forms were also observed always in combination with N-terminal acetylation and variably with methylation. The amount of methylation observed on each of these forms appeared independent of the other modifications present. Our multifaceted approach proved to produce very complementary data on this highly modified chromatin proteins and underscores the need for multiple approaches in characterization of protein modifications.

Experimental

Cell Culture and Purification of HMGA1a

HeLa S3 cells were grown and harvested as described by Thomas et al. [36], while nuclei were isolated as described by Garcia et al. [37], and HMG proteins were acid extracted from HeLa cells as described by Reeves [38]. Briefly, HMG proteins were acid extracted from

nuclei with 5% perchloric acid and precipitated with trichloroacetic acid (TCA) followed by washes with acetone containing 0.1% HCl and then pure acetone. The resulting pellets were redissolved in 1 mL deionized water and fractionated on a C18 column (4.6 mm i.d. × 250 mm; Vydac, Hesperia, CA, USA) using an Beckman Coulter System Gold HPLC (Fullerton, CA, USA) with a linear gradient of 1% B per min (A = 5% MeCN in 0.2% TFA, B = 90% acetonitrile in 0.188% TFA). Fractions containing HMGA1a proteins were pooled and dried to completion in a SpeedVac concentrator.

Sample Preparation for Bottom Up and Middle Down Mass Spectrometry

HMGA1a was digested into short peptides suitable for Bottom Up mass spectrometry by digestion with trypsin. Briefly, purified HMGA1 was resuspended in 100 mM ammonium bicarbonate (pH = 8.0) and digested for 6 h at 37 °C using trypsin (Promega, Madison, WI, USA) at a protein:enzyme ratio of 20:1. For Middle Down MS, a limited trypsin digestion was used with the reaction being quenched after 2 h. Alternatively, HMGA1a was derivatized by treatment with propionyl anhydride reagent essentially as described before [39, 40]. Briefly, this reagent was created by mixing 75 μL of MeOH with 25 μL of propionic anhydride (Sigma Aldrich, St. Louis, MO, USA). Equal volumes of reagent and HMGA1a protein were mixed and allowed to react at 37 °C for 15 min. Propionylated HMGA1a was then digested with trypsin after dilution of the sample with 100 mM ammonium bicarbonate buffer solution. The reaction was quenched by the addition of concentrated acetic acid and freezing (−80 °C). Following removal of reaction remnants and by-products through vacuum centrifugation, a second round of propionylation was then performed to propionylate the newly created peptide N-termini.

NanoLC-MS/MS for Bottom Up and Middle Down Mass Spectrometry

A small aliquot of the resulting peptides from both of the peptide digests were desalted using homemade stop and go extraction (STAGE) tips as previously described [41]. Briefly, peptides were twice eluted from STAGE Tips using 20 μL of 75% acetonitrile, 5% acetic acid. The resulting volume was reduced to ~5 μL in the SpeedVac. Peptides were acidified with 1 μL of 0.1% acetic acid. All mass spectrometry experiments were performed in the following manner. Peptide samples (2 μL) were loaded by an Eksigent AS-2 autosampler (Eksigent Technologies Inc., Dublin, CA, USA) onto a fused silica micropapillary (75 μm) column constructed with an integrated ESI tip and packed in-house with 5 um C_{18} YMC ODS-A resin. Peptides were HPLC separated with an Agilent 1200 series binary pump with an in-line flow splitter across a 110 min linear gradient ranging from 2 to 35% Buffer B (Buffer A = 0.1 M acetic acid, Buffer B = 70% acetonitrile in 0.1 M acetic acid) with a constant flow of ~100–200 nL/min. Alternatively, a 150 min linear gradient from 2% to 35% buffer B was used if higher chromatographic resolution was desired. The HPLC system was coupled to an LTQ-Orbitrap mass spectrometer (ThermoFisher Scientific, San Jose, CA, USA) operated with a spray voltage set to 2.7 kV. Survey full scan MS spectra (m/z 290–1650) were acquired in the Orbitrap with a resolution of 30,000 at m/z 400 after accumulation of ~500,000 ions. For collisionally activated dissociation (CAD), the 10 most intense ions from the preview scan from the Orbitrap were sequenced (normalized collision energy 35%) in the LTQ after accumulation of ~10,000 ions concurrently to full scan acquisition in the Orbitrap. Maximum filling time was 500 ms for the full scans. Precursor ion charge state screening was enabled and all unassigned charge states as well as singly charged species were rejected. The dynamic exclusion list was restricted to a maximum of 500 entries with a maximum retention period of 120 s and a relative mass window of <1 Da. Data were collected in centroid mode.

Top Down Mass Spectrometry of HMGA1a

HPLC purified HMGA1A was infused at ~0.5 μL/min via nanospray ionization into an LTQ-Orbitrap mass spectrometer equipped with electron-transfer dissociation (ETD). The analyte automatic gain control (AGC) was set to 7E5 for full MS and 2E5 for tandem MS analysis with an isolation width of 3 m/z. The ETD reagent anion AGC was set to 1E6 and a reaction time of 15 ms was used. A microscan count of 10 was used for both full MS and tandem MS and ~10 min of data were acquired and averaged for each spectrum. All data were acquired in the Orbitrap analyzer set at 100,000 resolution.

Data Analysis

For HMGA1a PTM detection, CAD mass spectra were searched using the SEQUEST algorithm within the Bioworks Browser (version 3.3.1 SP1; Thermo Fisher Scientific Inc., San Jose, Ca, USA) against a human protein database derived from sequences obtained from the National Center for Biotechnology Information (NCBI) database (August 2008). Trypsin and keratin contaminant proteins were also included in this database. Enzyme specificity was set to trypsin, fully enzymatic, allowing for up to five missed cleavage sites (since endogenous and chemical modification of lysine residues hinders enzymatic digestion and a 2 h limited trypsin digest was performed). Propionylation (+56.026Da) on the N-terminus of the proteins and peptides was set as a fixed modification. For HMGA1a PTM searches, Lys propionylation, Lys acetylation (+42.010Da), monomethyl on Lys residues (+70.042Da), and Arg residues (14.010 Da), dim-

ethylation on Arg and Lys (+28.031Da), Lys trimethylation (+42.046Da), and Ser and Thr phosphorylation (+79.966Da) were selected as variable modifications. HMGA1a monomethylation was searched as the sum of the masses for propionylation and methylation because monomethylated residues can still be propionylated. Mass tolerances were set to 0.05 m/z for parent ions and fragment ion tolerance was set to 0.5 m/z. Resulting peptide lists were filtered using standard criteria (Protein Probability < 1E-5, Xcorr = 2.0, 2.5, and 3.0 were used for peptide charges of 2, 3, and 4 respectively) as

previously used [42]. The false positive rate was estimated to be the 1% level for protein ID by searching a reverse database. All spectra from modified peptides were also manually inspected.

For analysis of Top Down mass spectrometry data, both computational and manual data interpretation was used to interpret the ETD tandem MS data. Ion peaks in the tandem MS were first identified and converted into their monoisotopic values using a THRASH algorithm and isotopic fitting [43]. Using a set of post-translational modifications that have been reported or discovered by
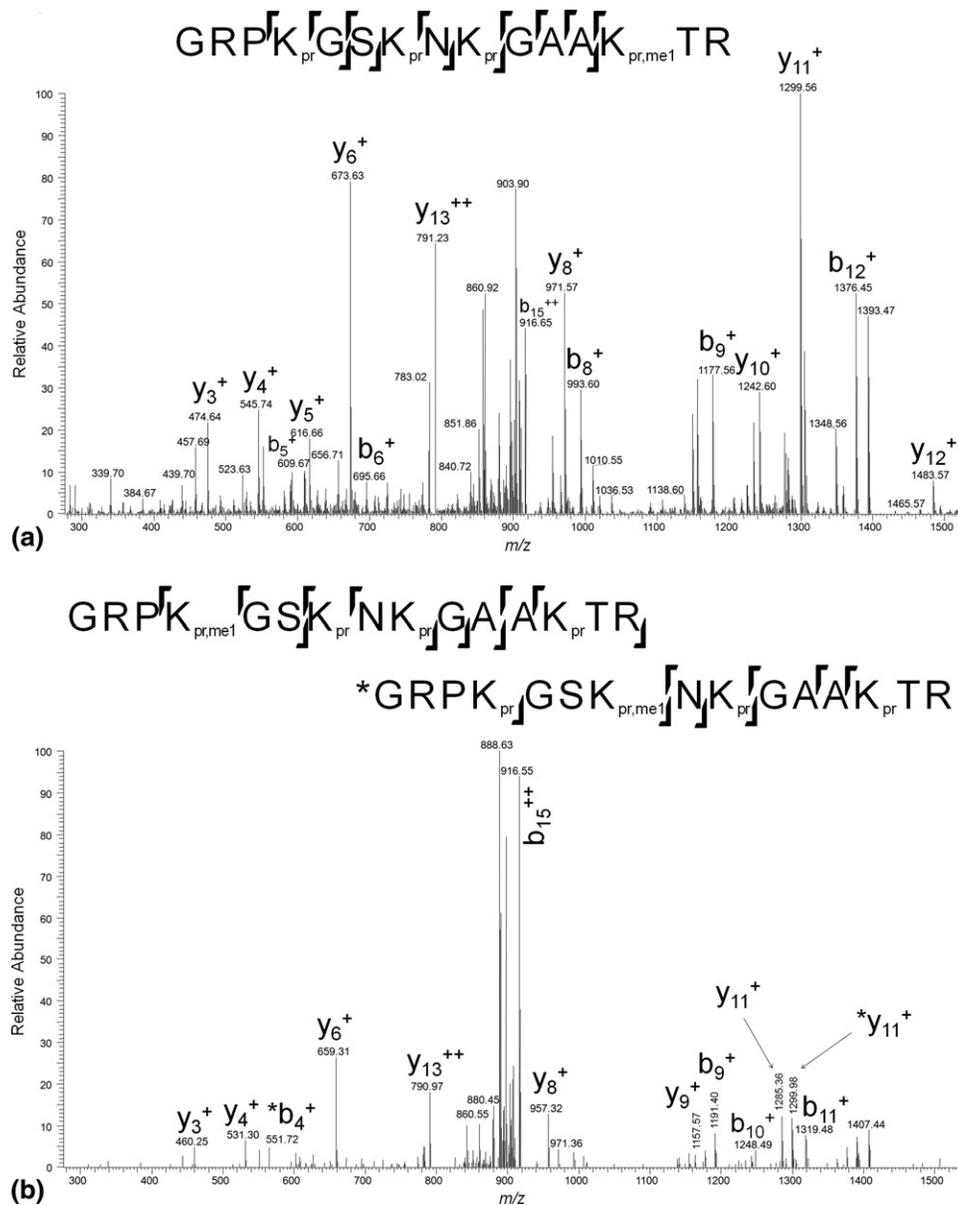
Figure 1. Differentiating post-translational modification isomers by tandem mass spectrometry. (**a**) CAD tandem spectrum of an [M + 2H]2+ ion at 925.535 m/z at from the propionylated HMGA1a trypsin digest showing the peptide with the sequence $_{pr}$-GRPK$_{pr}$GSK$_{pr}$NK$_{pr}$GAAK$_{me1}$TR (Lys70me1). (**b**) CAD MS/MS spectrum and fragment map for a doubly charged ion at the same 925.535 m/z as the species in (**a**) Nevertheless, the b and y type fragments allow for the identification of two different distinct species having the sequences of the HMGA1a peptide $_{pr}$-GRPK $_{prme1}$GSK $_{pr}$NK $_{pr}$GAAK $_{pr}$TR (Lys61me1) and $_{pr}$-GRPK $_{pr}$GSK $_{prme1}$NK $_{pr}$GAAK $_{pr}$TR (Lys64me1); pr = propionyl amide, ac = acetyl.

Bottom Up mass spectrometry, an in silico list of potential fragment masses was generated by independently starting from the N-terminus for the c-ion series and the C-terminus for the z⁻-ion series and then matching these theoretical fragments to the monoisotopic peak list. A modified version of the previously described integer linear optimization model [44] was utilized to discover modified forms that are (1) consistent with the parent mass within 28 Da and (2) maximize the supporting evidence found in the in silico list. The supporting ions corresponding to the modified forms reported for each spectrum were manually validated before concluding their presence or absence. Manual spectral interpretation was also used to thoroughly reveal additional supporting ions as needed.

Results and Discussion

In this article, we describe the intensive characterization of HMGA1a PTMs using a combination of Bottom Up together with Middle and Top Down proteomic approaches. HMGA1a was purified from asynchronously grown HeLa cells using standard perchloric acid extraction followed by reverse phase HPLC separation essentially as previously described [38]. Our overall goal was to identify potential combinatorial modification patterns on HMGA1a. However, before we attempted to resolve any combinatorial PTM patterns by MS, we decided to create a comprehensive PTM inventory of HMGA1a. To accomplish this, we turned to Bottom Up proteomic analyses performed via nanoflow

LC-MS/MS on a high-resolution Orbitrap mass spectrometer; in our laboratory, we find this approach provides the most sensitive measurement of protein PTMs. First, HMGA1a was derivatized by treatment with propionic anhydride reagent followed by trypsin digestion, a procedure we have previously optimized for histone protein PTM analysis [39, 40]. Propionylation blocks endogenously unmodified and monomethylated lysine residues, and thus this protocol generates predictable Arg-C like peptides that are of near optimum length for interrogation by CAD based tandem mass spectrometry. Through this approach, we detected many well known HMGA1a modifications. Supplemental Figure 1, (which can be found in the electronic version of this article) displays an MS/MS spectrum obtained of the $[M + 3H]^{3+}$ ion at 916.118 m/z from the 1-23 peptide of propionylated HMGA1a, and analysis of the fragment ions indicated that both the N-terminus and Lys14 were acetylated. This lysine residue has been previously shown to be acetylated by acetyltransferases PCAF (p300/CBP-associated factor) and p300 [28, 30]. It is interesting to note that in our experiments, we never detected non-N-terminally acetylated HMGA1a protein.

These experiments also revealed the presence of several novel modification sites as well. For instance, we observed monomethylation of HMGA1a on the peptide GRPKGSKNKGAAKTR, and inspection of different MS/MS spectra showed evidence for monomethylation of three distinct internal lysines on that peptide. Figure 1a displays an MS/MS spectrum of the $[M + 2H]^{2+}$ ion confirming the sequence $_{pr}$-

Table 1. List of selected modified HMGA1a peptides detected through nanoLC-MS/MS approaches (bottom up and middle down MS)

Peptide	Calc. mass	Obs. mass	Peptide charge	Relative abundance
$_{ac}$-SESSSK$_{pr}$SSQPLASK$_{pr}$QEK$_{pr}$DGTEK$_{pr}$R	920.789	920.793	3+	93.5%
$_{ac}$-SESSSK$_{pr}$SSQPLASK$_{ac}$QEK$_{pr}$DGTEK$_{pr}$R	916.117	916.115	3+	1.3%
$_{ac}$-SESSSK$_{prme1}$SSQPLASK$_{pr}$QEK$_{pr}$DGTEK$_{pr}$R/ $_{ac}$-SESSSK$_{pr}$SSQPLASK$_{pr}$QEK$_{pr}$DGTEK$_{prme1}$R	925.461	925.469	3+	5.2%
$_{pr}$-GRPK$_{pr}$GSK$_{pr}$NK$_{pr}$GAAK$_{pr}$TR	918.526	918.525	2+	99.5%
$_{pr}$-GRPK$_{pr}$GSK$_{pr}$NK$_{pr}$GAAK$_{prme1}$TR/ $_{pr}$-GRPK$_{pr}$GSK$_{prme1}$NK$_{pr}$GAAK$_{pr}$TR/ $_{pr}$-GRPK$_{prme1}$GSK$_{pr}$NK$_{pr}$GAAK$_{pr}$TR	925.534	925.531	2+	0.3%
$_{pr}$-GRPK$_{pr}$GSK$_{ac}$NK$_{pr}$GAAK$_{pr}$TR/ $_{pr}$-GRPK$_{pr}$GSK$_{pr}$NK$_{ac}$GAAK$_{pr}$TR	911.519	911.515	2+	0.2%
$_{pr}$-K$_{pr}$TTTTPGRK$_{pr}$PR	705.907	705.904	2+	97.8%
$_{pr}$-K$_{prme1}$TTTTPGRK$_{pr}$PR	712.915	712.912	2+	0.4%
$_{pr}$-K$_{ac}$TTTTPGRK$_{pr}$PR	698.899	698.897	2+	1.5%
$_{pr}$-K$_{pr}$TTTT$_{phos}$PGRK$_{pr}$PR	745.890	745.887	2+	0.3%
KLEKEEEGISQESSEEEQ	746.336	746.338	3+	3.7%
KLEKEEEGISQES$_{phos}$S$_{phos}$EEEQ	799.647	799.647	3+	96.3%
KQPPVSPGTALVGSQKEPSEVPTPK	640.352	640.354	4+	51.9%
KQPPVSPGTALVGS$_{phos}$QKEPSEVPTPK/ KQPPVS$_{phos}$PGTALVGSQKEPSEVPTPK/ KQPPVSPGTALVGSQKEPSEVPT$_{phos}$PK/ KQPPVSPGTALVGSQKEPS$_{phos}$EVPTPK/ KQPPVSPGT$_{phos}$ALVGSQKEPSEVPTPK	660.343	660.344	4+	38.0%
KQPPVS$_{phos}$PGTALVGSQKEPSEVPT$_{phos}$PK	680.334	680.335	4+	10.0%
KQPPVSPGTALVGSQKEPSEVPTPK (3 phos)*	700.326	700.326	4+	0.1%

ac = acetyl; me1 = monomethyl; phos = phosphoryl; pr = propionyl amide (chemical derivatization).
*Could not localize modification sites.
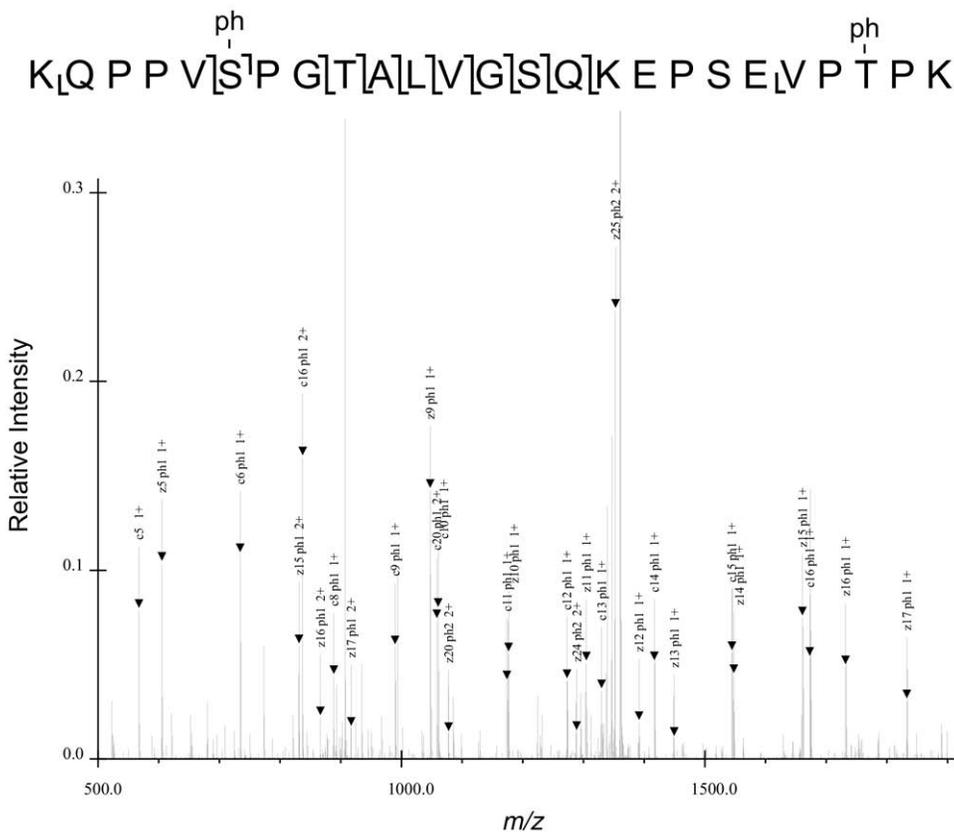
Figure 2. Middle down MS analysis of HMGA1a. ETD MS/MS spectrum of a triply charged ion at 680.335 *m/z* from the limited trypsin digest of HMGA1a. This peptide was determined to primarily be KQPPVSphosPGTALVGSQKEPSEVPTphosPK, displaying a combinatorial phosphorylation pattern.

GRPK$_{pr}$GSK$_{pr}$NK$_{pr}$GAAK$_{prme1}$TR (Lys70me1). Additionally, we observed other tandem mass spectra of the GRPKGSKNKGAAKTR peptide that also contained monomethyl modification, albeit at different lysine residues. For example, manual inspection of an MS/MS spectrum (Figure 1b) of the doubly charged ion at 925.535 *m/z* determined that this spectrum was composed of two positional isomeric species containing monomethyl either at Lys61 or Lys64. This is clearly evident as fragment ions separated by 14 Da can be seen throughout the MS/MS spectrum, such as at 1285 and 1299 *m/z* (y_{11} ions), and at 551 and 565 *m/z* (b_4 ions), providing support for monomethylation at both sites. Lys64 and Lys70 on HMGA1a have only been previously reported to be acetylated; [7] thus we find monomethylation of Lys61, Lys64, and Lys70 to be novel HMGA1a modifications. Interestingly, we did not observe higher degrees of methylation at those specific lysine residues, indicating that if existing, these higher degrees of methylation are present below our current detection limit. Modified peptides detected through this chemical derivatization Bottom Up MS strategy are listed in Table 1.

While this Bottom Up propionylation procedure allowed for extensive sequence coverage of the N- and C-termini of HMGA1a, we obtained little sequence cover-

age of the middle portion of the protein (~residues 25–60) and, hence, potential PTMs in those regions remained elusive. We felt that the peptides generated in this region through the propionylation procedure were possibly too low a charge state to be ionized efficiently for the large size of the peptide (propionylation removes charge from lysine residues producing mostly 2+ ions), and detection of those larger peptides could be actually hindered by this method. Aiming to access peptides covering this section, we then performed a partial tryptic digest (2 h) on HMGA1a with no propionylation derivatization performed either before or after the digest. Such digestion rendered peptides from the mid-region that were essen-

Figure 3. Sequence coverage and post-translational modifications found on HMGA1a through nanoLC-MS/MS approaches (Bottom Up and Middle Down). Sequence coverage is indicated by underlined regions.
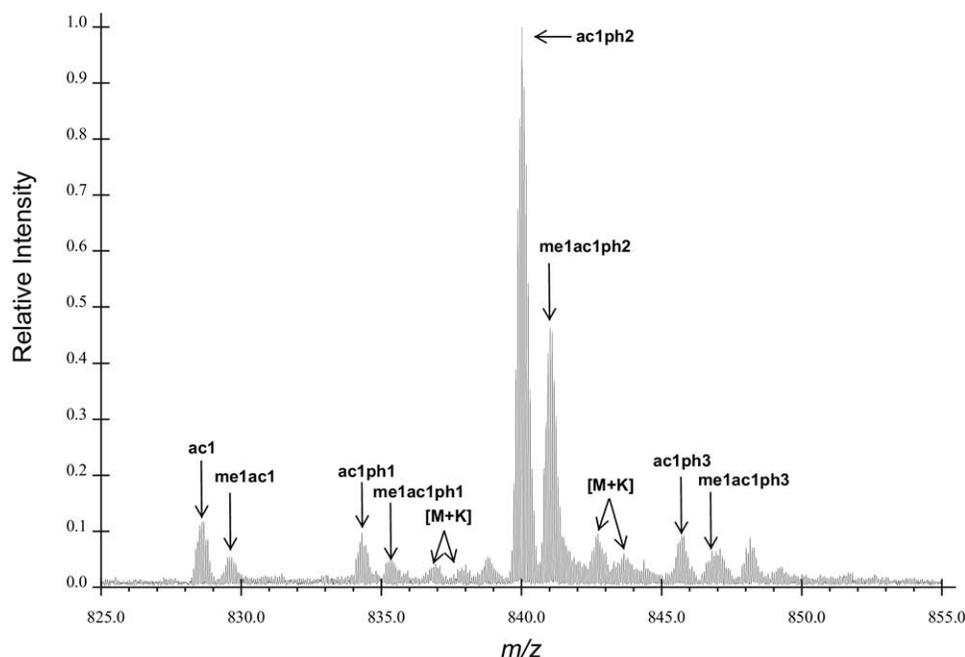
Figure 4. Full mass spectrum of the 14th charge state of HMGA1A derived from HeLa S3 cells grown in suspension. The modification state of the peaks are labeled based on accurate mass analysis (see Table 2) and Top Down ETD tandem MS sequencing (see Figure 5). Note that the spectrum is dominated by the diphosphorylated species. Note also that the relative ratio within a phosphorylation state of the methylated to unmethylated forms is consistent across different degrees of phosphorylation.

tially of the same size as those produced before, but with higher charge states (>3+), thus facilitating a "Middle Down" like characterization of the mid-portion of the HMGA1a protein. As Middle Down MS methods analyze modifications on large polypeptides; the connectivity between different modifications can be partially preserved through this approach as shown on histone proteins [37, 45]. Through this approach, we find that some HMGA1a post-translational modifications can occur in combination simultaneously. For example, an ETD tandem mass spectrum taken of the $[M + 3H]^{3+}$ ion at 680.335 m/z was found to have been generated from the peptide KQPPVSphosPGTALVGSQKEPSEVPTphosPK, phosphorylated on Ser35 and Thr52 (Figure 2), and evidence for other minor di-phosphorylated forms were also present (data not shown). We detected many distinct phosphorylation sites on the peptide

KQPPVSPGTALVGSQKEPSEVPTPK, but unlike the peptide mentioned above, most of them were only detected singly and not in combination (Table 1). This might indicate that the combinational phosphorylation pattern of HMGA1a at these sites is severely limited and/or low level, or may only occur in vivo under some specific cellular conditions.

After Bottom Up and Middle Down characterization, we find HMGA1a to be pervasively modified with many types of PTMs. A summary of all modifications we identify by Bottom Up and Middle Down MS approaches is presented in Figure 3 and Table 1. In all, we detect lysine methylation and acetylation as well as serine and threonine phosphorylation. Formylation has been recently discovered as an abundant modification on several chromatin associated proteins including Lys30 on HMGA1a [46]; unfortunately, we did not find

Table 2. A semiquantitative list of the most abundant modification states of HMGA1A. Both m/z and masses are for the most abundant isotopic peaks. In all cases the ppm error (part per million) in within 2 ppm as expected. In the final column the approximate percent of that each species represents relative to the sum. The two most abundant species that are both diphosphorylated make up about 75% of the mixture and the remaining forms are relatively low level

Modification state	Exp. m/z	Exp. Da	Theo. Da	ppm Error	Approx. %
ac1ph2	840.0117	11746.0619	11746.0650	−0.3	51.6
me1ac1ph2	841.0122	11760.0689	11760.0806	−1.0	23.8
ac1	828.5878	11586.1273	11586.1323	−0.4	6.0
ac1ph1	834.2993	11666.0883	11666.0987	−0.9	5.0
ac1ph3	845.7232	11826.0233	11826.0313	−0.7	4.8
me1ac1ph3	846.7242	11840.0367	11840.0470	−0.9	3.5
me1ac1	829.5900	11600.1581	11600.1480	0.9	2.8
me1ac1ph1	835.2995	11680.0911	11680.1143	−2.0	2.5

evidence of formylation in our study. We found novel modification sites including lysine methylation on Lys6, Lys22, Lys61, Lys64, Lys70, Lys73, as well as phosphorylation on Thr38 (Figure 3 and Table 1). Nevertheless, although we detect a large variety and number of HMGA1a post-translational modifications, we find little evidence for simultaneously occurring combinatorial modifications. This observation prompted us to attempt to estimate the relative modification level of each modified peptide by measuring the abundance of the modified peptide expressed as the ratio of the modified peptide over the sum of the modified and unmodified forms of that peptide, as we have previously performed for histone proteins [40]. The relative abundances of all the modified peptides are also listed in Table 1. Most modified peptides are of very low abundance (<5%); this suggests that most modified forms of HMGA1a are in very few specialized locations in the human genome. However, not all modifications are of low abundance; in particular, phosphorylated forms of the C-terminal serine residues on HMGA1a were found at relatively high levels compared with most modifications, prompting speculation that these sites are constitutively phosphorylated (Table 1). Previous LC-MS profiling experiments have found various phosphorylation states on intact HMGA1a; [32] accordingly, our assessment of the high abundance of C-terminal phosphorylation is in agreement with this prior report.

Despite the many modification sites observed, we could not detect a large variety of combinatorial PTM complexity through Bottom Up and Middle Down approaches. Thus, we decided to take this kind of experiment one step further and explore the possibility of combinatorial PTM codes on HMGA1a through a Top Down MS proteomic approach. We reason that possibly the combinatorial modifications may span much longer regions of the protein than those we produced and examined using Bottom Up and Middle Down MS; therefore, Top Down MS may be the only method suitable to study HMGA1a combinatorial PTMs. Additionally, we found that by using both a Bottom Up and Middle Down approach, we did not detect many peptides with Arg modifications, although those types of modifications are some of the most well known and studied HMG PTMs [24, 26]. A Top Down MS experiment essentially provides 100% sequence coverage (although PTM localization is more challenging), and would enable the identification of PTMs missed in our previous MS analysis of HMGA1a digested samples as well.

HPLC purified HMGA1a extracted from asynchronously grown HeLa cells was dissolved in 50% MeCN/ 0.1% formic acid and nanoelectrosprayed into the Orbitrap mass spectrometer. In Figure 4, the full MS of the purified HMGA1a is zoomed in on the 14+ charge state spanning the 825–865 m/z range. Multiple peaks can be seen corresponding to various combinations of PTMs over the expected baseline molecular mass of HMGA1a. Extrapolated masses of the eight more easily detectable species are presented in Table 2; the high mass accuracy

of the instrument allowed for sub-2 ppm determinations of the masses. The mass of the major peak, 840.0117 m/z is consistent with a doubly phosphorylated and monoacetylated species as shown in Table 2. This species makes up ~50% of the HMGA1A recovered from normally grown HeLa cells. The second most abundant species (~25% of the total mixture) is the diphosphorylated, monoacetylated and monomethylated peak at 841.0122 m/z. Thus the diphosphorylated forms constitute over 75% of the HMGA1a in normally grown HeLa S3 cells. Most other peaks are primarily analogs of these peaks (acetylated and acetylated + monomethylated) with 0, 1, and 3 phosphorylations. Interestingly, the ratio between the acetyl and the acetylated + monomethyl forms remain essentially the same regardless of the phosphorylation state. This spectrum shows relatively limited adduct formation, such as the potassium adducts labeled, which are typical for intact protein mass spectra. Likely other forms are present, accounting for the other modifications observed in the Bottom Up analysis, but are at fairly low levels and obfuscated by interference with adducts. Nevertheless, this data is consistent with our Bottom Up and Middle Down analyses demonstrating that the combinatorial modifications of HMGA1a are fairly limited. A few relatively simple combinatorial forms dominate the population and additional low level modifications likely combine with these apparently constitutive modifications but likely not in extensive combination with each other.

Top Down MS analysis of the major ions observed in the full MS at 840 and 841 m/z indicate that these two major forms are both acetylated and diphosphorylated. From our Bottom Up experiments, we concluded that HMGA1a was always detected as being N-terminally acetylated (Supplemental Figure 1). Therefore, we expected that all these protein forms which contain an acetyl group were merely N-terminally acetylated. Indeed, inspection of the c and z type fragment ions from the ETD-MS/MS spectrum (Figure 5) of the most abundant species at 841 m/z + 1.5 m/z with the 14+ charge state showed that the acetyl group was on the N-terminus. This is also shown in ion map form in Figure 6. The c-ion series locates the acetylation site on the N-terminus as it starts relatively low at c_5ac1^{1+}. This small protein fragment (SESSS) is only capable of acetylation at the N-terminus. Further examination of the fragment ions in the spectrum also pointed to the likely sources of diphosphorylation was through the combination of two residues at Ser98, Ser101 or Ser102 (z_{26} ion at 1600 m/z, Supplemental Figure 2). Unfortunately, our z ion series does not allow for the precise identification of the two residues near the carboxy-terminus that are phosphorylated. HMGA1a ETD spectra usually exhibit extensive sequence coverage except for the last 24 amino acids near the C-terminus. This section of the protein (RGRPKKLEKEEEEGISQESSEEEQ) is very rich in glutamic acid residues making it challenging to maintain sufficient positive charge for MS detection especially

ac-S⌊E⌊S S S⌉K⌉S⌉S⌉Q⌉P⌉L⌉A⌉S⌉K⌉Q⌉E K⌊D⌊G⌊T⌊E⌉K⌉R⌉
    me

⌊G⌊R⌊G⌊R P⌉R⌊K⌊Q P P⌊V⌊S P⌊G⌊T⌊A⌊L⌊V⌊G⌊S⌊Q⌊K⌊E P

S⌊E⌊V P T P K⌊R P⌊R⌊G⌊R P⌊K⌊G⌊S⌊K⌊N⌊K⌊G⌊A⌊A⌊K⌊T⌉

⌊R⌊K⌊T⌊T⌊T⌊T P G⌊R⌊K P⌊R G R P K K L E K E E E E
       2ph
‾G‾I‾S‾Q‾E‾S‾S‾E‾E‾E Q
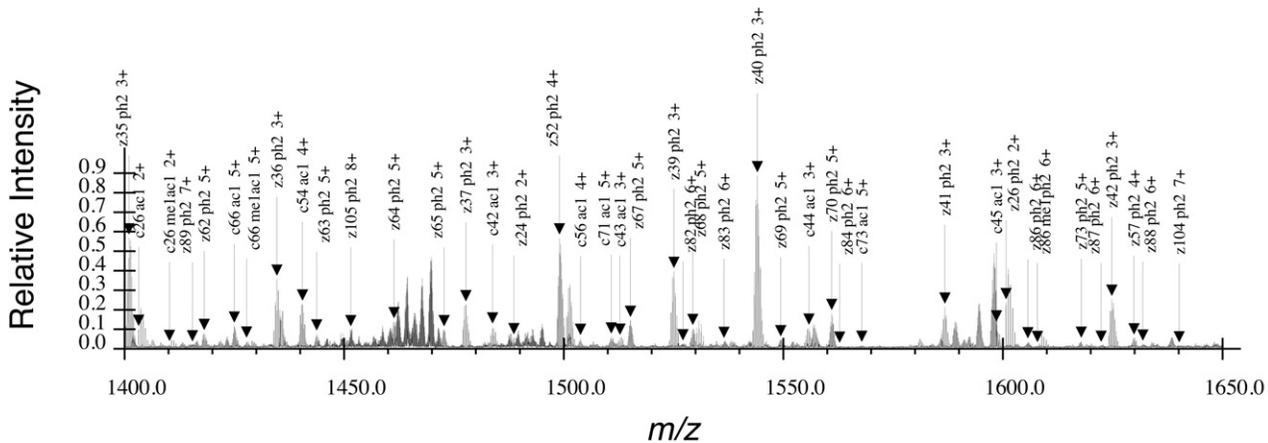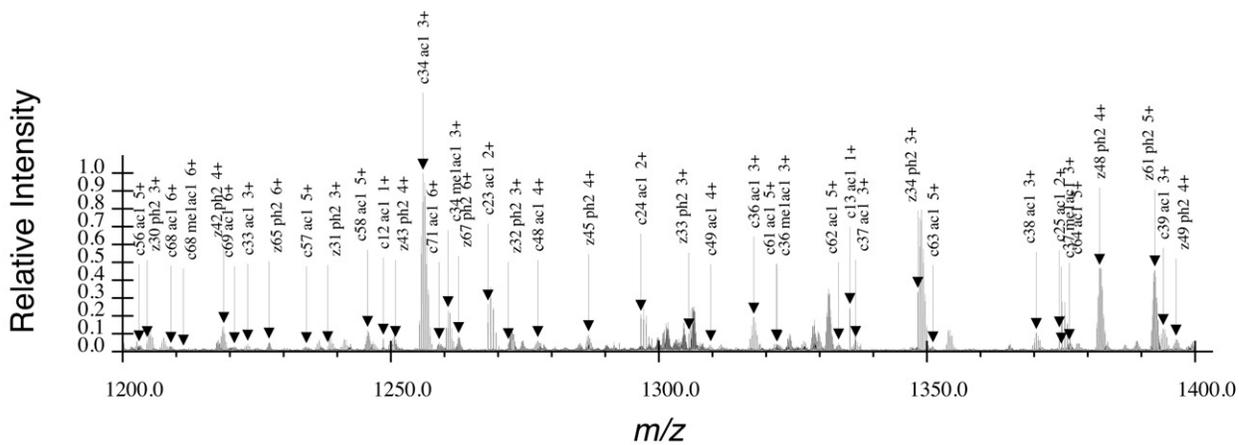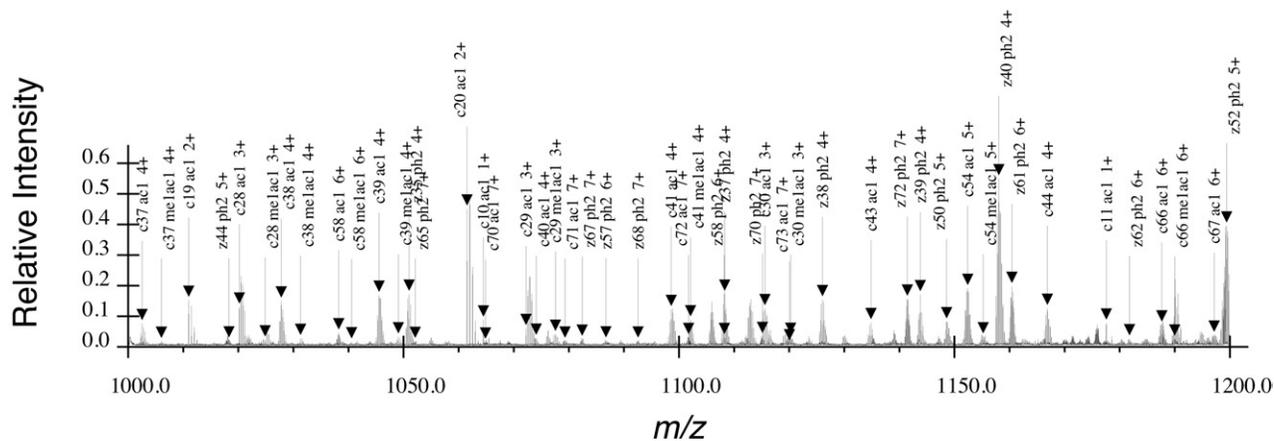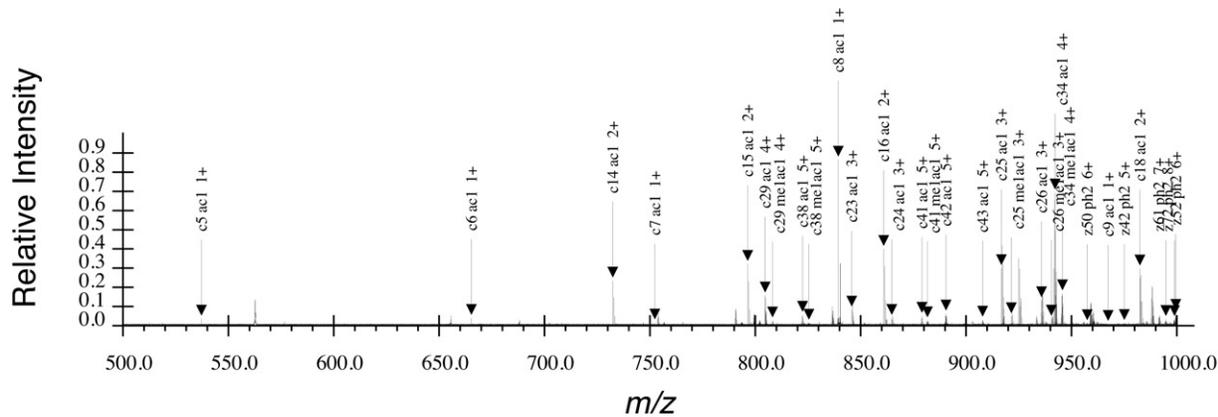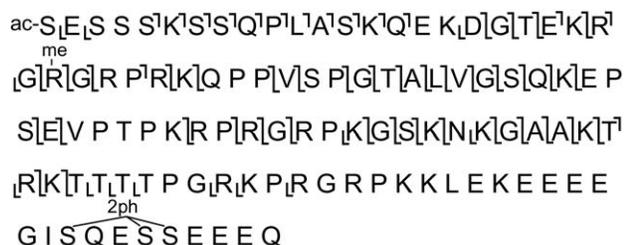
Figure 6. ETD based fragment map of the Top Down MS analysis of HMGA1a obtained from fragmentation produce in Figure 5. The methylation on Arg25 is variable (there is also an unmethylated form in the spectrum). The C-terminal phosphorylations are at two out of three of the serines indicated. The Bottom Up data only supports the two phosphorylations being at the two more C-terminal sites (Ser101phSer102ph).

for the low z-ions. Additionally, the c-ion series, which in theory could maintain charge, did not reach these depths, stopping at c_{73}. Our Bottom Up data is in perfect agreement with the Top Down data indicating that the major sites of phosphorylation are at the C-terminus. The Bottom Up MS experiments only detected positive evidence for the Ser101phSer102ph diphosphorylated form out of the three possible combinations (Ser98phSer101ph, Ser98phSer102ph and Ser101phSer102ph). Although not quantitative, this appears to indicate that Ser101phSer102ph is the dominant form. The c-ion series provides more information to identify the exact site of methylation on the secondary form. Our ETD-MS/MS analysis confirmed that the secondary modified form is monomethylated at Arg25 with supporting ions ($3+$ c_{25} ion at about 920 m/z and beyond) shown in Figure 5. The many ions before c_{25} show no signs of methylation and more than 20 ions after and including c_{25} exhibit clear ac1me1 peaks (Supplemental Figure 2). This is critical information on the primary site of methylation; this site is conspicuously absent from the Bottom Up analysis due to the density of arginine residues in the region. The z-ion series also extends deep enough to also help confirm the location of the methylation of the secondary form (e.g., $z86me1ph2^{6+}$, $6+$ ion at about 1608 m/z) and even exhibit apparent $z104$ and $z105$ ions that constitute the entire protein less the first one or two residues on the N-terminus. Overall, more than 180 peaks were used for identification with an rms error of 1.3ppm, and a fragment map of the ETD driven Top Down analysis of the two major forms of HMGA1a en masse is shown in Figure 6. Multiple alternative hypotheses derived from our computational methods were also tested by manual validation, but generally were found to be caused by interferences from the validated forms, lacking the consistent and voluminous evidence for the top fitting forms. The support for these hypothetical forms was generally relatively weak; we believe such support is due exclusively to the density of peaks in the spectrum and the wide range of possible modifications considered in combination. The algorithm did however correctly identify the top two forms detailed above in the correct order further corroborating the correct identifications.

Conclusions

In this study, we extend the current information concerning HMGA1a PTMs from human cells using various methods including Top, and Middle Down, as well as Bottom Up MS approaches. Through Bottom Up and Middle Down MS proteomics, we find HMGA1a to be highly modified at numerous sites with many different types of PTMs (some novel), and with the exception of the C-terminal phosphorylation, the majority of these modifications are present at low abundance levels (<5%). Middle Down MS analysis allowed us to determine that some HMGA1a modifications can occur simultaneously, suggesting the potential of an HMGA1a combinatorial PTM code analogous to the Histone Code; however, this code was found to be much less complex or much lower in abundance than expected. Results obtained through a Top Down MS analysis were also consistent with those results obtained through the other MS approaches, indicating that combinatorial modifications of HMGA1a are not extensive in asynchronously grown HeLa cells, but rather dominated by a few combinations of the most abundant modifications (very limited specific codes). These data represent the first Top Down MS analysis of HMGA1a, and provide a unique look at HMGA1a modifications. Overall, we found that the combined use of the different MS techniques is highly complementary and allowed for a thorough characterization of HMGA1a modifications that could not be accomplished by any one method alone. It will be interesting to further probe both the single and combinatorial modifications of HMGA1a using this multifaceted MS approach during more specific biological conditions such as during cancer progression or other disease states.

Acknowledgments

Figure 5. A Top Down ETD spectrum of the diphosphorylated forms of HMGA1A isolated at 841.5 m/z ± 1.5 m/z. The major forms are identified are both N-terminally acetylated and diphosphorylated near the C-terminus and the less abundant form is monomethylated at Arg25. The spectrum is broken into four unequal sections for clarity. Due to space limitations, not all identified peaks are labeled.

Appendix A
Supplementary Material

Supplementary material associated with this article may be found in the online version at doi:10.1016/j.jasms.2010.01.020.

References

1. Bernstein, B. E.; Meissner, A.; Lander, E. S. The Mammalian Epigenome. *Cell* **2007,** *128,* 669–681.
2. Cheung, P.; Allis, C. D.; Sassone-Corsi, P. Signaling to Chromatin through Histone Modifications. *Cell* **2000,** *103,* 263–271.
3. Kouzarides, T. Chromatin Modifications and Their Function. *Cell* **2007,** *128,* 693–705.
4. Lachner, M.; O'Carroll, D.; Rea, S.; Mechtler, K.; Jenuwein, T. Methyl-ation of Histone H3 Lysine 9 Creates a Binding Site for Hp1 Proteins. *Nature* **2001,** *410,* 116–120.
5. Bannister, A. J.; Zegerman, P.; Partridge, J. F.; Miska, E. A.; Thomas, J. O.; Allshire, R. C.; Kouzarides, T. Selective Recognition of Methylated Lysine 9 on Histone H3 by the Hp1 Chromo Domain. *Nature* **2001,** *410,* 120–124.
6. Grosschedl, R.; Giese, K.; Pagel, J. HMG Domain Proteins: Architectural Elements in the Assembly of Nucleoprotein Structures. *Trends Genet.* **1994,** *10,* 94–100.
7. Zhang, Q.; Wang, Y. High Mobility Group Proteins and Their Post-Translational Modifications. *Biochim. Biophys. Acta* **2008,** *1784*, 1159–1166.
8. Chiappetta, G.; Bandiera, A.; Berlingieri, M. T.; Visconti, R.; Manfioletti, G.; Battista, S.; Martinez-Tello, F. J.; Santoro, M.; Giancotti, V.; Fusco, A. The Expression of the High Mobility Group Hmgi (Y) Proteins Corre-lates with the Malignant Phenotype of Human Thyroid Neoplasias. *Oncogene* **1995,** *10,* 1307–1314.
9. Giancotti, V.; Berlingieri, M. T.; DiFiore, P. P.; Fusco, A.; Vecchio, G.; Crane-Robinson, C. Changes in Nuclear Proteins on Transformation of Rat Epithelial Thyroid Cells by a Murine Sarcoma Retrovirus. *Cancer Res.* **1985,** *45,* 6051–6057.
10. Giancotti, V.; Buratti, E.; Perissin, L.; Zorzet, S.; Balmain, A.; Portella, G.; Fusco, A.; Goodwin, G. H. Analysis of the Hmg-I Nuclear Proteins in Mouse Neoplastic Cells Induced by Different Procedures. *Exp. Cell. Res.* **1989,** *184,* 538–545.
11. Chiappetta, G.; Avantaggiato, V.; Visconti, R.; Fedele, M.; Battista, S.; Trapasso, F.; Merciai, B. M.; Fidanza, V.; Giancotti, V.; Santoro, M.; Simeone, A.; Fusco, A. High Level Expression of the Hmgi (Y) Gene During Embryonic Development. *Oncogene* **1996,** *13,* 2439–2446.
12. Zhou, X.; Benson, K. F.; Ashar, H. R.; Chada, K. Mutation Responsible for the Mouse Pygmy Phenotype in the Developmentally Regulated Factor Hmgi-C. *Nature* **1995,** *376,* 771–774.
13. Amirand, C.; Viari, A.; Ballini, J. P.; Rezaei, H.; Beaujean, N.; Jullien, D.; Kas, E.; Debey, P. Three Distinct Sub-Nuclear Populations of Hmg-I Protein of Different Properties Revealed by Co-Localization Image Analysis. *J. Cell Sci.* **1998,** *111*(Pt. 23), 3551–3561.
14. Martelli, A. M.; Riccio, M.; Bareggi, R.; Manfioletti, G.; Tabellini, G.; Baldini, G.; Narducci, P.; Giancotti, V. Intranuclear Distribution of Hmgi/Y Proteins. An Immunocytochemical Study. *J. Histochem. Cyto-chem.* **1998,** *46,* 863–864.
15. Reeves, R.; Wolffe, A. P. Substrate Structure Influences Binding of the Non-Histone Protein Hmg-I(Y) to Free Nucleosomal DNA. *Biochemistry* **1996,** *35,* 5063–5074.
16. Reeves, R.; Nissen, M. S. Interaction of High Mobility Group-I (Y) Nonhistone Proteins with Nucleosome Core Particles. *J. Biol. Chem.* **1993,** *268,* 21137–21146.
17. Reeves, R.; Leonard, W. J.; Nissen, M. S. Binding of Hmg-I(Y) Imparts Architectural Specificity to a Positioned Nucleosome on the Promoter of the Human Interleukin-2 Receptor α Gene. *Mol. Cell. Biol.* **2000,** *20,* 4666–4679.
18. Elton, T. S.; Reeves, R.; Purification and Postsynthetic Modifications of Friend Erythroleukemic Cell High Mobility Group Protein Hmg-I. *Anal. Biochem.* **1986,** *157,* 53–62.
19. Lund, T.; Holtlund, J.; Laland, S. G. On the Phosphorylation of Low Molecular Mass Hmg (High Mobility Group) Proteins in Ehrlich Ascites Cells. *FEBS Lett.* **1985,** *180,* 275–279.
20. Nissen, M. S.; Langan, T. A.; Reeves, R. Phosphorylation by Cdc2 Kinase Modulates DNA Binding Activity of High Mobility Group I Nonhistone Chromatin Protein. *J. Biol. Chem.* **1991,** *266,* 19945–19952.
21. Harrer, M.; Luhrs, H.; Bustin, M.; Scheer, U.; Hock, R. Dynamic Interaction of Hmga1a Proteins with Chromatin. *J. Cell. Sci.* **2004,** *117,* 3459–3471.
22. Banks, G. C.; Li, Y.; Reeves, R. Differential In Vivo Modifications of the Hmgi(Y) Nonhistone Chromatin Proteins Modulate Nucleosome and DNA Interactions. *Biochemistry* **2000,** *39,* 8333–8346.
23. Edberg, D. D.; Bruce, J. E.; Siems, W. F.; Reeves, R. In Vivo Posttrans-lational Modifications of the High Mobility Group A1a Proteins in Breast Cancer Cells of Differing Metastatic Potential. *Biochemistry* **2004,** *43,* 11500–11515.
24. Sgarra, R.; Diana, F.; Bellarosa, C.; Dekleva, V.; Rustighi, A.; Toller, M.; Manfioletti, G.; Giancotti, V. During Apoptosis of Tumor Cells Hmga1a Protein Undergoes Methylation: Identification of the Modification Site by Mass Spectrometry. *Biochemistry* **2003,** *42,* 3575–3585.
25. Edberg, D. D.; Adkins, J. N.; Springer, D. L.; Reeves, R. Dynamic and Differential In Vivo Modifications of the Isoform Hmga1a and Hmga1b Chromatin Proteins. *J. Biol. Chem.* **2005,** *280,* 8961–8973.
26. Zou, Y.; Wang, Y. Tandem Mass Spectrometry for the Examination of the Post-Translational Modifications of High-Mobility Group A1 Pro-teins: Symmetric and Asymmetric Dimethylation of Arg25 in Hmga1a Protein. *Biochemistry* **2005,** *44,* 6293–6301.
27. Sgarra, R.; Lee, J.; Tessari, M. A.; Altamura, S.; Spolaore, B.; Giancotti, V.; Bedford, M. T.; Manfioletti, G. The At-Hook of the Chromatin Architectural Transcription Factor High Mobility Group A1a Is Arginine-Methylated by Protein Arginine Methyltransferase 6. *J. Biol. Chem.* **2006,** *281,* 3764–3772.
28. Jiang, X.; Wang, Y. Acetylation and Phosphorylation of High-Mobility Group A1 Proteins in Pc-3 Human Tumor Cells. *Biochemistry* **2006,** *45,* 7194–7201.
29. Zou, Y.; Wang, Y. Mass Spectrometric Analysis of High-Mobility Group Proteins and Their Post-Translational Modifications in Normal and Cancerous Human Breast Tissues. *J. Proteome Res.* **2007,** *6,* 2304–2314.
30. Zhang, Q.; Zhang, K.; Zou, Y.; Perna, A.; Wang, Y. A Quantitative Study on the In Vitro and In Vivo Acetylation of High Mobility Group A1 Proteins. *J. Am. Soc. Mass Spectrom.* **2007,** *18,* 1569–1578.
31. Zou, Y.; Webb, K.; Perna, A. D.; Zhang, Q.; Clarke, S.; Wang, Y. A Mass Spectrometric Study on the in Vitro Methylation of Hmga1a and Hmga1b Proteins by PRMT s: Methylation Specificity, the Effect of Binding to At-Rich Duplex DNA, and the Effect of C-Terminal Phos-phorylation. *Biochemistry* **2007,** *46,* 7896–7906.
32. Sgarra, R.; Maurizio, E.; Zammitti, S.; Lo Sardo, A.; Giancotti, V.; Manfioletti, G. Macroscopic Differences in Hmga Oncoproteins Post-Translational Modifications: C-Terminal Phosphorylation of Hmga2 Affects Its DNA Binding Properties. *J. Proteome Res.* **2009,** *8,* 2978–2989.
33. Garcia, B. A.; Shabanowitz, J.; Hunt, D. F. Characterization of Histones and Their Post-Translational Modifications by Mass Spectrometry. *Curr. Opin. Chem. Biol.* **2007,** *11,* 66–73.
34. Pesavento, J. J.; Mizzen, C. A.; Kelleher, N. L. Quantitative Analysis of Modified Proteins and Their Positional Isomers by Tandem Mass Spectrometry: Human Histone H4. *Anal. Chem.* **2006,** *78,* 4271–4280.
35. Boyne, M. T.; 2nd; Pesavento, J. J.; Mizzen, C. A.; Kelleher, N. L. Precise Characterization of Human Histones in the H2a Gene Family by Top Down Mass Spectrometry. *J. Proteome Res.* **2006,** *5,* 248–253.
36. Thomas, C. E.; Kelleher, N. L.; Mizzen, C. A. Mass Spectrometric Characterization of Human Histone H3: A Bird's Eye View. *J. Proteome Res.* **2006,** *5,* 240–247.
37. Garcia, B. A.; Pesavento, J. J.; Mizzen, C. A.; Kelleher, N. L. Pervasive Combinatorial Modification of Histone H3 in Human Cells. *Nat. Meth-ods* **2007,** *4,* 487–489.
38. Reeves, R. Hmga Proteins: Isolation, Biochemical Modifications, and Nucleosome Interactions. *Methods Enzymol.* **2004,** *375,* 297–322.
39. Garcia, B. A.; Mollah, S.; Ueberheide, B. M.; Busby, S. A.; Muratore, T. L.; Shabanowitz, J.; Hunt, D. F. Chemical Derivatization of Histones for Facilitated Analysis by Mass Spectrometry. *Nat. Protoc.* **2007,** *2,* 933–938.
40. Plazas-Mayorca, M. D.; Zee, B. M.; Young, N. L.; Fingerman, I. M.; Leroy, G.; Briggs, S. D.; Garcia, B. A. One-Pot Shotgun Quantitative Mass Spectrometry Characterization of Histones. *J. Proteome Res.* **2009,** *8*(11), 5367–5374.
41. Rappsilber, J.; Ishihama, Y.; Mann, M. Stop and Go Extraction Tips for Matrix-Assisted Laser Desorption/Ionization, Nanoelectrospray, and LC/Ms Sample Pretreatment in Proteomics. *Anal. Chem.* **2003,** *75,* 663–670.
42. MacCoss, M. J.; Wu, C. C.; Yates, J. R. III. Probability-Based Validation of Protein Identifications Using a Modified SEQUEST Algorithm. *Anal. Chem.* **2002,** *74,* 5593–5599.
43. Horn, D. M.; Zubarev, R. A.; McLafferty, F. W. Automated Reduction and Interpretation of High Resolution Electrospray Mass Spectra of Large Molecules. *J. Am. Soc. Mass Spectrom.* **2000,** *11,* 320–332.
44. Dimaggio, P. A. Jr.; Young, N. L.; Baliban, R. C.; Garcia, B. A.; Floudas, C. A. A Mixed-Integer Linear Optimization Framework for the Identi-fication and Quantification of Targeted Post-Translational Modifica-tions of Highly Modified Proteins Using Multiplexed Electron Transfer Dissociation Tandem Mass Spectrometry. *Mol. Cell. Proteom.* **2009,** *8*(11), 2527–2543.
45. Phanstiel, D.; Brumbaugh, J.; Berggren, W. T.; Conard, K.; Feng, X.; Levenstein, M. E.; McAlister, G. C.; Thomson, J. A.; Coon, J. J. Mass Spectrometry Identifies and Quantifies 74 Unique Histone H4 Isoforms in Differentiating Human Embryonic Stem Cells. *Proc. Natl. Acad. Sci. U.S.A.* **2008,** *105,* 4093–4098.
46. Wisniewski, J. R.; Zougman, A.; Mann, M. Nepsilon-Formylation of Lysine Is a Widespread Post-Translational Modification of Nuclear Proteins Occurring at Residues Involved in Regulation of Chromatin Function. *Nucleic Acids Res.* **2008,** *36,* 570–577.