REVIEW

The top-down, middle-down, and bottom-up mass spectrometry approaches for characterization of histone variants and their post-translational modifications

Annie Moradian[1], Anastasia Kalli[2], Michael J. Sweredoski[1] and Sonja Hess[1]

[1] Proteome Exploration Laboratory, Beckman Institute, California Institute of Technology, Pasadena, CA, USA
[2] Department of Pathology and Laboratory Medicine, Children's Hospital Los Angeles, Los Angeles, CA, USA

Epigenetic regulation of gene expression is, at least in part, mediated by histone modifications. PTMs of histones change chromatin structure and regulate gene transcription, DNA damage repair, and DNA replication. Thus, studying histone variants and their modifications not only elucidates their functional mechanisms in chromatin regulation, but also provides insights into phenotypes and diseases. A challenge in this field is to determine the best approach(es) to identify histone variants and their PTMs using a robust high-throughput analysis. The large number of histone variants and the enormous diversity that can be generated through combinatorial modifications, also known as histone code, makes identification of histone PTMs a laborious task. MS has been proven to be a powerful tool in this regard. Here, we focus on bottom-up, middle-down, and top-down MS approaches, including CID and electron-capture dissociation/electron-transfer dissociation based techniques for characterization of histones and their PTMs. In addition, we discuss advances in chromatographic separation that take advantage of the chemical properties of the specific histone modifications. This review is also unique in its discussion of current bioinformatic strategies for comprehensive histone code analysis.

Received: June 28, 2013
Revised: July 30, 2013
Accepted: August 15, 2013

Keywords:
Histone / Middle-down proteomics / MS / PTM / Sytems biology / Top-down proteomics

1 Introduction

Epigenetic regulation of gene expression is, at least in part, mediated by PTMs of histones. As part of the nucleosomes, histones are generally tightly packed proteins that form octamers (H2A/H2B dimers and H3/H4 tetramer) with unstructured N-terminal tails that are highly modified in vivo. DNA is tightly packed around these structures. Depending on its transcriptional activity, lightly packed, active euchromatin and tightly packed, inactive heterochromatin can be distinguished. The transition from inactive to active state and vice versa is mostly accomplished via PTMs. Adding, for instance, enzymatically catalyzed acetylation to the ε-amino group of a lysine leads to its neutralization, thereby reducing its interactions with the negatively charged DNA in the promoter region of a gene. As a result of this decondensation, transcription factors have better access to chromatin to induce gene expression [1]. Therefore, the transcriptionally active state, euchromatin, has been associated with acetylation, especially acetylation of H4 [2]. In addition, H3K4, H3K36, and H3K79 trimethylation as well as H2B ubiquitination has been associated with euchromatin [2]. In contrast, the restricted inactive state, heterochromatin, has been mostly associated with H3K27Me3, H3K9Me3, and H2A ubiquitination [2] (Fig. 1). In addition to acetylation, propionylation, butyrylations, and crotonylations have been reported [3,4]. Lysine can be mono-, di-, or trimethylated while arginine can be mono- or dimethylated. Arginine dimethylation can be either symmetric

Correspondence: Dr. Sonja Hess, Proteome Exploration Laboratory, Beckman Institute, California Institute of Technology, BI 211, MC139-74, Pasadena, CA 91125, USA
E-mail: shess@caltech.edu
Fax: 001-626-449-4159

Abbreviations: ECD, electron-capture dissociation; **ETD**, electron-transfer dissociation; **MILP**, mixed integer linear programming; **WCX-HILIC**, weak cation exchange hydrophilic interaction LC

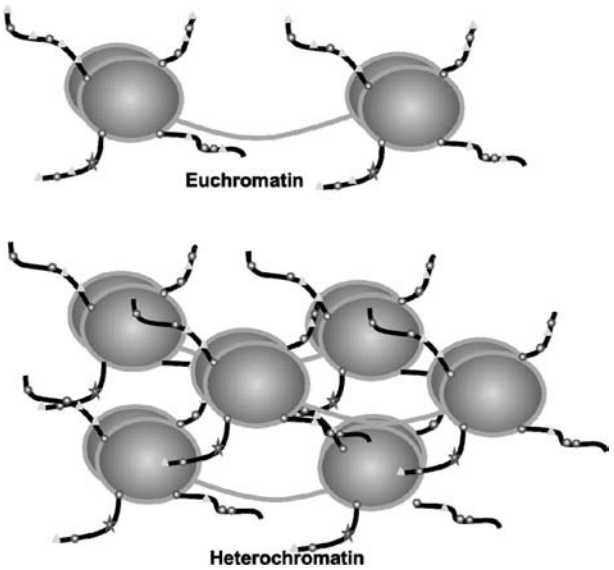

Figure 1. Euchromatin and heterochromatin. The active form euchromatin (top) is generally high in acetylation (shown as triangles). Acetylation of the N-terminal tails of histones leads to deprotonation of the lysine side chains and thus a loss of positive charge. This ultimately results in chromatin decondensation, which enables better access for transcription. In contrast, the tightly packed heterochromatin (bottom) is inactive. Positively charged lysines interact with the negatively charged DNA.

or asymmetric [5]. Additional PTMs include citrullination (deamidation), phosphorylation, O-GlcNAcylation (where O-GlcNAc is O-acetyl-D-glucosamine), sumoylation, biotinylation, and adenosine diphosphate (ADP)-ribosylation [6]. A list of currently known PTMs in histones is shown in Table 1, together with their elemental composition and residual monoisotopic and average masses. A major focus in the area is currently directed toward acetylations and methylations at the expense of other PTMs. This development is due, in part, to the abundance of the PTMs, with higher abundance PTMs being detected more easily than lower abundance PTMs. In addition, in an effort to reduce the search space to a manageable size, most database search engines limit the number of possible modifications that can be studied at once. Naturally, many researchers focus on the most abundant PTMs. We only begin to understand the impact of all these PTMs and may find in the future that some of the less-studied PTMs (e.g., ubiquitination) have greater impact than currently anticipated. Nevertheless, it is clear that studying histone variants and their modifications not only elucidates their functional mechanisms through transcription, DNA replication, and DNA repair [7], but also provides insights into transcriptionally controlled processes such as cell proliferation and differentiation. It, therefore, increases our fundamental understanding in functional processes during development [8], aging [9], and diseases such as cancer [10] and cardiovascular [11] and neurological [12] diseases. Recently, it was discovered that regulator of methylation A of *Legionella pneumophila* trimethylates H3K14 of the host and thus inactivates gene expression of its proteins [13]. In addition, it activates gene expression of bacterial proteins [13]. This gives an interesting glimpse at the possibilities of cross-species regulation through histone PTMs.

Despite tremendous progress, one major challenge is the enormous complexity of the histone combinatorial modifications. In addition, the high number of closely related histone variants and their high sequence homology complicates their analysis even more. For instance, it can be difficult to distinguish an isobaric methylated Ala variant from an unmethylated Val variant, when insufficient fragment ions are present. For histone H3.1 alone, 40 trillion theoretical proteoforms [14] have been calculated [15]. Taken together, the large number of different histone variants, possible modifications, and the enormous diversity that can be generated through combinatorial modifications, also known as histone code, makes identification of histone PTMs a challenging task. In

Table 1. Common PTMs in histones

Modification	Δ Monoisotopic mass	Δ Average mass	Elemental composition
Citrullination (deamidation)	0.984016	0.9848	H(-1) N(-1) O(1)
Monomethylation	14.01565	14.0266	H(2) C(1)
Formylation	27.994915	28.0101	C(1) O(1)
Dimethylation	28.0313	28.0532	H(4) C(2)
Acetylation	42.010565	42.0367	H(2) C(2) O(1)
Trimethylation	42.04695	42.0797	H(6) C(3)
Propionylation	56.026215	56.0633	H(4) C(3) O(1)
Crotonylation	68.026215	68.074	H(4) C(4) O(1)
Butyrylation	70.041865	70.0898	H(6) C(4) O(1)
Phosphorylation	79.966331	79.9799	H(1) O(3) P(1)
Ubiquitinylation (glygly)	114.042927	114.1026	H(6) C(4) N(2) O(2)
Biotinylation	226.077598	226.2954	H(14) C(10) N(2) O(2) S(1)
O-GlcNAcylation	203.079373	203.1925	H(13) C(8) N(1) O(5)
ADP-ribosylation	541.06111	541.3005	H(21) C(15) N(5) O(13) P(2)

ADP, adenosine diphosphate; O-GlcNAc, O-acetyl-D-glucosamine.

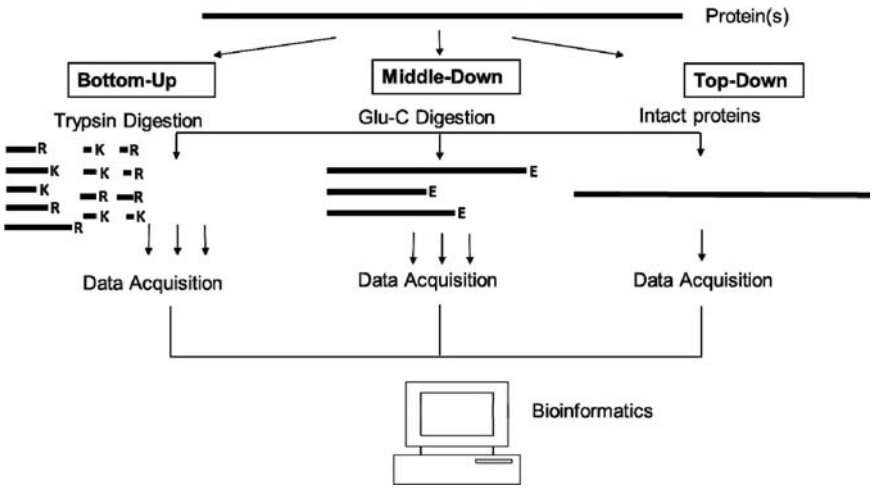

Figure 2. Bottom-up, middle-down, and top-down strategies. In the bottom-up approach (left), proteins are generally digested with trypsin leading to short (and in the case of histones ultrashort) peptides. The middle-down approach (middle) uses a limited digest (e.g., Glu-C or Asp-N), while the top-down approach (right) uses intact proteins without any digestions. All peptides and proteins are acquired under approach-specific conditions and analyzed by bioinformatics tools. See text for discussion of advantages and disadvantages of each approach.

the past, many studies relied on Western blots against modified histone peptides. Recent studies showed that antibodies against histone modifications showed cross-reactivity [16] and failed specificity tests [17], questioning at least some of the results obtained with these antibodies. More recently, MS has been proven to be a powerful method for characterization, including identification and quantitation of biomolecules and their PTMs. Thus, MS has become the method of choice for the comprehensive analysis of combinatorial histone proteoforms.

2 Analytical challenges and solutions

A challenge in the field of histone code analysis is to find the best mass spectrometric approach(es) to identify the histone variants and their PTMs using a robust high-throughput analysis. To this end, top-down, middle-down, and bottom-up approaches have been used to study histones and their PTMs (Fig. 2).

2.1 Top-down approach

Theoretically, all PTMs are conserved in the top-down approach [18]. In fact, the field was largely pioneered using histones, because they are particularly amenable to top-down analysis due to their small size (11–21 kDa) and their relative high abundance [19–21]. There is no digestion step required in this technique. Furthermore, this approach should be able to differentiate between different histone variants and proteoforms. It could also provide information about the stoichiometry of all the modifications. It should, therefore, be most suited for the comprehensive characterization of the histone code. However, despite tremendous advances in instrumentation, top-down approaches remain relatively insensitive and require large amounts of sample. While not prohibitive, the necessary expensive high-end instrumentation somewhat limits its widespread use. In addition, the top-down approach is not always successful in fully fragmenting the entire protein, which sometimes leads to ambiguous PTM localization assignment [22]. The effective analysis of histones also requires extensive prefractionation to enable the comprehensive analysis of as many proteoforms as currently possible [23–25]. Because of the many fractions that are needed to accomplish a comprehensive histone analysis, it cannot be currently considered a true high-throughput method that would allow researchers to test many biological samples. When deciphering the histone code, a sensitive, automated high-throughput method is deemed important, rendering the top-down approach not the best choice for comprehensive histone analysis.

2.2 Bottom-up approach

In contrast to the top-down approach, the bottom-up approach is significantly more sensitive and, in fact, has been successful in identifying some novel modifications on histones [26], but it seems not particularly well suited for characterization of concurrently occurring, distant PTMs that would be present in different tryptic peptides. Trypsin digestion of the N-terminal tails of the histones that are rich in Lys and Arg residues leads to ultra-short peptides with an average length of three to four amino acids that are not amenable for LC-MS identification [27, 28]. First, some of the highly hydrophilic ultra-short peptides will elute with the dead volume of the column and may not even be detected. Second, even if detected and fragmented, many six or seven amino acid histone peptides are shared between several histone paralogs, which complicates the inference of assigning modifications to individual variants. Lastly, even if a peptide is uniquely assigned to one variant, the connectivity of the individually identified peptides is lost in tryptic digestions. In addition, it is impossible to obtain information of PTMs occurring in two adjacent tryptic peptides in combination [27, 29, 30].

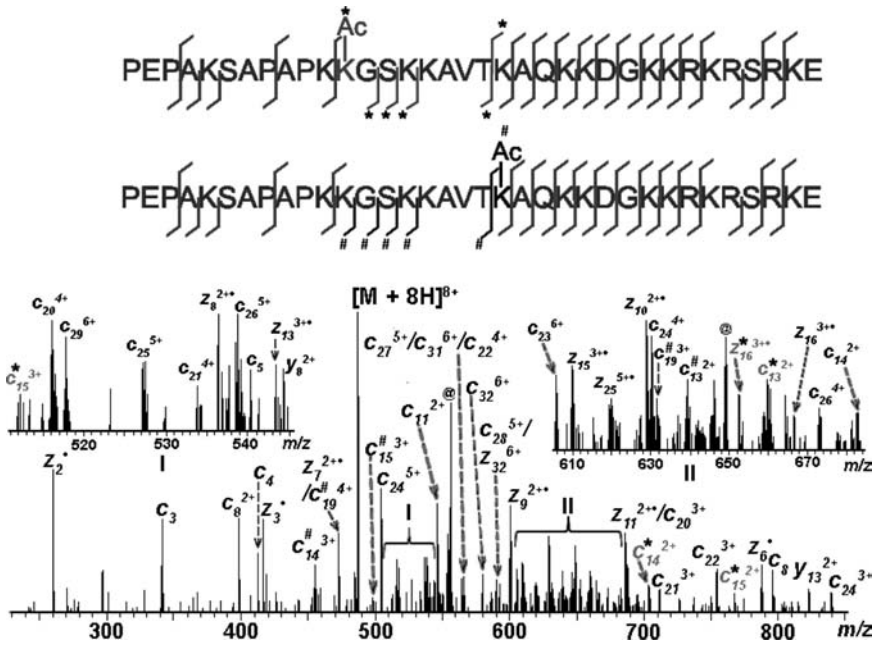

Figure 3. Hybrid spectra of the N-terminal H2b 1-K histone tail. The N-terminal tail 1–35 of H2B 1-K (and/or H2B 2-E, H2B 1-J, H2B 1-C/E/F/G/I) was obtained from a Glu-C digest of Jurkat cells. Product ions labeled with "*" exclusively belong to K12Ac proteoform and product ions labeled with "#" exclusively belong to the K20Ac proteoform. Only the major peaks are labeled in the figure. @, charge-reduced species and neutral or side chain losses.
Reprinted with permission and adaption from Kalli [37]. Copyright 2013 American Chemical Society.

2.3 Middle-down approach

For histone analysis, the middle-down approach, in which proteins are digested into peptides commonly in the 3 to 9 kDa range, is emerging as an attractive alternative [31–33]. Middle-down approaches have recently become more popular, both because of advancements in instrumentation and the realization that middle-down approaches largely preserve the combinatorial modifications of the histone tail, while approaching the sensitivity of the bottom-up approach [22, 27, 32, 34]. Due to the paucity of Asp and Glu in the N-terminal histone tails, larger peptides can be generated using either Asp-N or Glu-C [31]. The comprehensive propionylation of all free lysines, which renders them less susceptible for tryptic cleavage, also produces larger peptides, although they are for the most part technically not yet in the middle-down range [35]. Unless heavy labels for these propionylations are used, it would also be impossible to distinguish between biologically occurring [3] and man-made propionylation events. Initial attempts to characterize the nuclear fraction including histones via middle-down approaches took advantage of the conventional CID-based fragmentation technique [36]. Complementary electron-based fragmentation techniques such as electron-transfer dissociation (ETD) and electron-capture dissociation (ECD) have recently been shown to be better suited for the characterization of posttranslationally modified large histone peptides [30, 37–39]. Both, ETD and ECD were used with online LC, making this a high-throughput methodology [30, 37–39]. Kalli et al. used a C_{18AQ} material of the unfractionated Glu-C- and Asp-N-digested (unfractionated) histones [37] together with ECD fragmentation. ECD of large polypeptides has previously been considered unfavorable if not unsuitable for online separation [40, 41]. Thus, it has only been used in an offline fashion for the ECD analysis of prefractionated Glu-C-digested histone isoforms [42, 43]. However, ECD has previously been shown to produce extensive fragmentation [44–48], particularly for the analysis of highly modified, large polypeptides [49, 50]. We thus reasoned that ECD should produce superior sequence coverage and due to the high resolution of the Fourier transformation ion cyclotron resonance mass analyzer, high specificity in identifying and characterizing histone modifications. The unambiguous characterization of chimeric spectra is shown in Fig. 3, where the N-terminal tail of H2B 1-K (and/or H2B 2-E, H2B 1-J, H2B 1-C/E/F/G/I) features an acetylation at K12 and K20. Distinct fragment ions enable the differentiation and concomitant presence of the two proteoforms. Ambiguity can arise when low mass accuracy ETD in the ion trap is used since it is often impossible to assign charge states of fragment ions. When high mass accuracy ETD data are obtained in an Orbitrap (which would allow greater confidence in the individual fragment-ion assignment) however, sensitivity is much lower in our experience.

In case of the ETD analysis, Phanstiel et al. focused on the Asp-N-digested, prefractionated H4 (aa1-23) [39] using a C_{18} column. In an effort to separate histones based on their acetylation state, an online weak cation exchange hydrophilic interaction chromatography (WCX-HILIC) column was developed [30, 38]. The WCX-HILIC fractionation proves to be superior to reverse-phase (C_{18}) fractionation in terms of its ability to separate highly modified histone tails of H3 into several peaks [30, 38]. Ultimately, this leads to a more in-depth analysis of the present histone tails. The high resolving power of this method is based on separation by charge (acetylation status) due to weak cation exchange and by hydrophilic interactions of histone tail (methylation status). The important

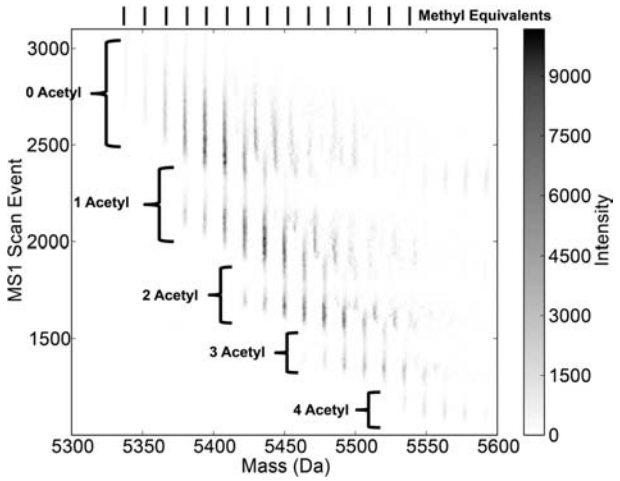

Figure 4. Heat map of the N-terminal H3 histone tail spectra. A heat map of the MS1 scan events versus mass of the N-terminal tail of histone H3 using a WCX-HILIC separation clearly shows the separation of the different acetylated and methylated proteoforms. Note that the number of acetylations decreases over time in the chromatographic analysis.

aspect introduced by Young et al. was the pH gradient used in WCX-HILIC column in order to avoid the salt gradients that are not compatible with high-throughput MS analysis [30, 38].

This methodology combined with ECD fragmentation was applied to study H3 modifications in histone deacetylase inhibitor treated and untreated murine erythroleukemia cells in our laboratory and the resulting mass spectral heat map of the pH gradient WCX-HILIC of histone H3 modified forms is shown in Fig. 4: N-terminal tails of H3 are separated by their acetylation status and methylation status, enabling an in-depth analysis of the modifications. The chromatographic selectivity of this method provides a better chance of triggering relatively separated MS/MS spectra of different H3 proteoforms. The distinct separation based on acetylation and methylation status decreases the complexity of isobaric species, which improves confidence in spectral assignments compared to previous methods. We also reasoned that the high mass accuracy acquisition of the ECD data should be advantageous for correct PTM characterization. When analyzed by WCX-HILIC combined with ECD, we obtained promising results. The top proteoforms identified for the N-terminus of H3.2 alone are shown in Fig. 5, indicating the power of this technique (unpublished data).

3 Postacquisition challenges and solutions

Histone analysis also remains a challenge from the informatics perspective. The first issue to be addressed is whether preprocessing of the raw data should occur or multiply charged fragment ions should be searched as acquired. The problem with the latter approach is that the potential for a false fragment ion assignment increases when several charge states are considered for every possible fragment ion. The options for processing include deisotoping and decharging MS2 spectra and the possibility of MS2 spectral averaging. Each of these operations is not without potential pitfalls either. Tools such as MS-Deconv [51], YADA [52], Thrash [53], and Xtract (Thermo) can be used for generating a list of monoisotopic peaks, but it is not clear how well these algorithms perform with ETD or ECD, where the radical and nonradical fragment ions could hinder interpretation. In fact, just a simple determination of a monoisotopic peak that is readily accomplished in bottom-up proteomics remains challenging when done with larger peptides [54]. Additionally, problems can arise when trying to average MS2 spectra to increase the S/N ratio (as is commonly done in top-down analyses), particularly when online separations are used. Isobaric peptides with slightly different PTM combinations, which were chromatographically resolved, could be accidently merged into a single spectrum, making the interpretation more difficult.

The next challenge in analysis pipeline is the identification of histone peptides or proteins. While there are relatively few sequences that need to be considered, the combinatorial nature of the modifications and the number of potential sites of modification make brute-force searches of middle-down and top-down histone data very challenging for traditional bottom-up search engines. As an example, a brute-force search of the 40 trillion theoretical proteoforms of H3.1 would require searching 1.2 million proteoforms a second for an entire year, when no restrictions are made with respect to the allowable modifications. The number of unmodified tryptic peptides in the entire human proteome (~5 million) is modest in comparison. In practice, one would not expect all 40 trillion theoretical proteoforms to be present at once, but the challenge is to determine which ones are present in any given sample. The current way to deal with this challenge is to limit the searched proteoforms to the most likely anticipated ones. As a consequence, this practical solution can miss unknown modifications. In addition to the combinatorial nature of the modifications, there are also challenges associates with identical or near-identical masses of the modifications. For instance, very few fragment ions would differentiate mono-methylation on neighboring residues or dimethylation on a single residue. Insufficient fragmentation would automatically result in ambiguous assignments. Additionally, formylation and dimethylation as well as acetylation and trimethylation only differ in mass by 0.0364 Da (see Table 1). At the mass accuracy of high-resolution MS/MS typically used during data analysis, these modifications are on the edge of being indistinguishable, yet often one can observe a slight difference in distribution of fragment mass errors when multiple fragment-ion mass errors are considered. One would have to conclude that tools that incorporate a fragment mass error penalty would, therefore, have an advantage over those that only employ a hard fragment ion mass tolerance. This is illustrated in Fig. 6, where a histogram of the observed fragment-ion mass

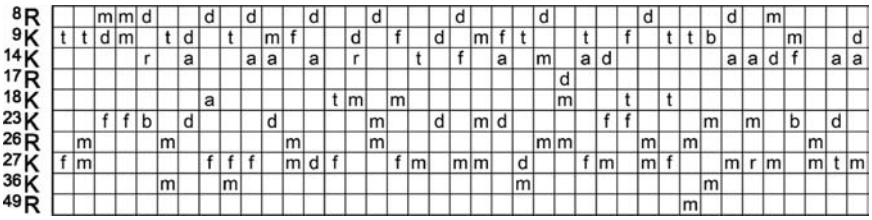

Figure 5. Proteoforms of the N-terminal histone H3 tail. Representative PTM identifications of the N-terminal H3 tail show the combinatorial nature of the modifications. For clarity, only the modified sites are shown. a, acetylation; b, butyrylation; d, dimethylation; f, formylation; m, monomethylation; t, trimethylation; r, propionylation.

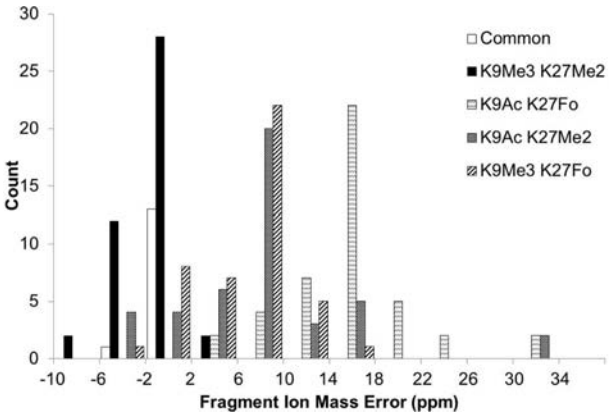

Figure 6. The importance of high-accuracy fragment-ion determination. A 15 or 20 ppm precursor and fragment-ion tolerance is not sufficient to differentiate formylation or dimethylation modifications of the H3.2 N-terminus tail using middle-down MS/MS. This histogram shows the distribution of parts per million errors of the fragment ions, true and false assignments can be distinguished. The correct modification K27Me2 centers around 0 ppm, whereas the incorrect modifications center around 8 and 16 ppm, respectively.

error for the correct K9Me3K27Me2 is plotted against possible K9AcK27Fo, K9AcK27Me2, and K9Me3K27Fo. All of these possibilities are within a 20 ppm mass error. The fragment-ion mass error of the correct assignment, however, centers around 0 ppm, while the incorrect assignments center either at 8 or 16 ppm. This example also shows the importance of high mass accuracy for the correct fragment-ion assignments.

Few tools have been developed for searching middle-down or top-down data. Of particular note, MS-Align+ [55] and ProSightPC [56] have been developed for searching top-down data. Additionally, several algorithms have been proposed specifically for analyzing highly modified histones, including a mixed integer linear programming (MILP) approach [57], an approach that incorporates prior knowledge and no deisotoping of spectra [58] and an approach that uses spectral alignments, MS-Align-E [59]. Lastly, a few database search engines such as Phenyx (GeneBio) [60, 61], ProsightPTM [62], and ROCCIT (roccit.caltech.edu) [37] incorporate known modification sites. By restricting the search space to mostly known locations of specific modifications, more probable peptides can be given priority over less probable (e.g., no evidence exists to suggest H2A K96 is ubiquitinated, while H2A K120 is known to be ubiquitinated).

Yet another hurdle in interpreting histone data is the inevitable chimeric nature of the spectra. While significant advances have been made in chromatographically separating peptides, it is currently still unavoidable that peptides with different numbers of methylations are co-isolated. As an example, the typical H3.1 peptide has a mass of approximately 5300 Da. At charge state 8, the single-, double-, and triple-methylated versions would have mass to charge ratios of 664.25, 666.00, and 667.75, which would be co-isolated and fragmented together using typical ECD isolation widths of 4 Da. Both DiMaggio et al. [57] and Guan and Burlingame [58] attempt to handle this challenge and quantify the relative abundances of the different proteoforms. DiMaggio et al. employ MILP to identify and quantify chimeric spectra [57]. The first MILP identifies all possible proteoforms that would match the observed precursor mass within a given tolerance by formulating the problem as one of feasibility. The second MILP attempts to quantify the relative abundance of the potential proteoforms identified in the first MILP by formulating the problem as one of superposition. Lastly, DiMaggio et al. show how retention time can be used to help differentiate acetylation and trimethylation when certain chromatographic separations are employed [57].

Guan and Burlingame developed a collection of algorithms to identify and quantify chimeric spectra [58]. The first step in their workflow is to perform a standard database search to identify the peptide sequence and a representative set of modifications present. The next step is to enumerate all other possible modification configurations and related nonredundant ions (i.e., fragment ions that can differentiate potential modification configurations). Next, the nonredundant ions are identified and quantified in the observed raw spectrum. Independent modification configurations are then enumerated using a greedy algorithm similar to that used when inferring proteins from identified peptides. Lastly, the relative abundance of the independent modification configurations is calculated using a nonnegative least squares procedure [58].

One potential problem with both approaches is that there is an assumption that only one peptide sequence is present in each spectrum. While this may be the case for extensively fractionated samples, it will be less likely so in any high-throughput workflow.

Considering these challenges, there is the potential to employ tools and computational methods from data-independent acquisition approaches using gas-phase fragmentation techniques [63] including SWATH-MS [64]. The constant monitoring of all precursor and fragments will give

more in-depth data than the current data-dependent analysis, where often the most abundant ions are triggered for MS/MS and identified at the expense of the lower abundant ions. In fact, as stated above, in most data-dependent settings, the isolation width of the precursor ion will be large enough (3 Da or more) to isolate more than one precursor, but most database search engines will only allow the identification of one precursor per MS/MS spectrum, resulting in a lot of PTMs to be missed. Using a data-independent approach with all ion fragmentation, all precursor and fragment ions within certain mass range are monitored. A slight mass shift in a precursor and fragment ion due to a variation in the PTM (e.g., acetylation vs. trimethylation, unmodified vs. citrullinated) of an isoform can easily be visualized in contour plots, enabling the comprehensive detection of the vast majority of the PTMs present in the sample.

4 Conclusions

Bottom-up, middle-down, and top-down approaches have all contributed to our current knowledge in histone modifications, and will likely do so in the future, either alone, or more likely in combination. With the recent development of WCX-HILIC [30, 38] and online ECD applications in the middle-down range [37], we do see a major breakthrough on the analytical side. The current bottleneck is the development of robust bioinformatics pipelines that can automatically interpret the collected data. On the instrument side, one can assume that more advanced techniques (higher energy collision dissociation triggered ETD, electron transfer and higher energy collision dissociation combination, all ion fragmentation) will become available and/or more efficient with, for example, the latest Orbitrap Fusion instrumentation. Together with the prospect of analyzing all coeluting ions at once and a robust bioinformatics pipeline, one can expect the next revolution in the characterization of histone modifications.

This research is funded by the Beckman Institute and the Gordon and Betty Moore Foundation through grant GBMF775.

Potential conflict of interest: S.H. is a consultant to Genentech.

5 References

[1] 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.

[2] Wong, J. J., Hawkins, N. J., Ward, R. L., Colorectal cancer: a model for epigenetic tumorigenesis. *Gut* 2007, *56*, 140–148.

[3] Chen, Y., Sprung, R., Tang, Y., Ball, H. et al., Lysine propionylation and butyrylation are novel post-translational modifications in histones. *Mol. Cell Proteomics* 2007, *6*, 812–819.

[4] Tan, M., Luo, H., Lee, S., Jin, F. et al., Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* 2011, *146*, 1016–1028.

[5] Wang, Y., Wysocka, J., Sayegh, J., Lee, Y. H. et al., Human PAD4 regulates histone arginine methylation levels via demethylimination. *Science* 2004, *306*, 279–283.

[6] Hottiger, M. O., ADP-ribosylation of histones by ARTD1: an additional module of the histone code? *FEBS Lett.* 2011, *585*, 1595–1599.

[7] Turner, B. M., Histone acetylation and an epigenetic code. *Bioessays* 2000, *22*, 836–845.

[8] Avrahami, D., Kaestner, K. H., Epigenetic regulation of pancreas development and function. *Semin. Cell Dev. Biol.* 2012, *23*, 693–700.

[9] Kamminga, L. M., de Haan, G., Cellular memory and hematopoietic stem cell aging. *Stem Cells* 2006, *24*, 1143–1149.

[10] Taby, R., Issa, J. P., Cancer epigenetics. *CA Cancer J. Clin.* 2010, *60*, 376–392.

[11] Webster, A. L., Yan, M. S., Marsden, P. A., Epigenetics and cardiovascular disease. *Can. J. Cardiol.* 2013, *29*, 46–57.

[12] Gray, S. G., Epigenetic treatment of neurological disease. *Epigenomics* 2011, *3*, 431–450.

[13] Rolando, M., Sanulli, S., Rusniok, C., Gomez-Valero, L. et al., Legionella pneumophila effector RomA uniquely modifies host chromatin to repress gene expression and promote intracellular bacterial replication. *Cell Host Microbe* 2013, *13*, 395–405.

[14] Smith, L. M., Kelleher, N. L., Proteoform: a single term describing protein complexity. *Nat. Methods* 2013, *10*, 186–187.

[15] Perkel, J. M., Tearing the top off 'top-down' proteomics. *Biotechniques* 2012, *53*, 75–78.

[16] Peach, S. E., Rudomin, E. L., Udeshi, N. D., Carr, S. A., Jaffe, J. D., Quantitative assessment of chromatin immunoprecipitation grade antibodies directed against histone modifications reveals patterns of co-occurring marks on histone protein molecules. *Mol. Cell Proteomics* 2012, *11*, 128–137.

[17] Egelhofer, T. A., Minoda, A., Klugman, S., Lee, K. et al., An assessment of histone-modification antibody quality. *Nat. Struct. Mol. Biol.* 2011, *18*, 91–93.

[18] Siuti, N., Kelleher, N. L., Decoding protein modifications using top-down mass spectrometry. *Nat. Methods* 2007, *4*, 817–821.

[19] Eliuk, S. M., Maltby, D., Panning, B., Burlingame, A. L., High resolution electron transfer dissociation studies of unfractionated intact histones from murine embryonic stem cells using on-line capillary LC separation: determination of abundant histone isoforms and post-translational modifications. *Mol Cell Proteomics* 2010, *9*, 824–837.

[20] Fraga, M. F., Ballestar, E., Villar-Garea, A., Boix-Chornet, M. et al., Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nat. Genet.* 2005, *37*, 391–400.

[21] 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.

[22] Meyer, B., Papasotiriou, D. G., Karas, M., 100% protein sequence coverage: a modern form of surrealism in proteomics. *Amino Acids* 2011, *41*, 291–310.

[23] Pesavento, J. J., Bullock, C. R., LeDuc, R. D., Mizzen, C. A., Kelleher, N. L., Combinatorial modification of human histone H4 quantitated by two-dimensional liquid chromatography coupled with top down mass spectrometry. *J. Biol. Chem.* 2008, *283*, 14927–14937.

[24] Tian, Z., Tolic, N., Zhao, R., Moore, R. J. et al., Enhanced top-down characterization of histone post-translational modifications. *Genome Biol.* 2012, *13*, R86.

[25] Tian, Z., Zhao, R., Tolic, N., Moore, R. J. et al., Two-dimensional liquid chromatography system for online top-down mass spectrometry. *Proteomics* 2010, *10*, 3610–3620.

[26] Sidoli, S., Cheng, L., Jensen, O. N., Proteomics in chromatin biology and epigenetics: elucidation of post-translational modifications of histone proteins by mass spectrometry. *J. Proteomics* 2012, *75*, 3419–3433.

[27] Arnaudo, A. M., Molden, R. C., Garcia, B. A., Revealing histone variant induced changes via quantitative proteomics. *Crit. Rev. Biochem. Mol.* 2011, *46*, 284–294.

[28] Garcia, B. A., Mollah, S., Ueberheide, B. M., Busby, S. A. et al., Chemical derivatization of histones for facilitated analysis by mass spectrometry. *Na. Protoc.* 2007, *2*, 933–938.

[29] Garcia, B. A., Mass spectrometric analysis of histone variants and post-translational modifications. *Front. Biosci. (Schol. Ed.)* 2009, *1*, 142–153.

[30] Young, N. L., DiMaggio, P. A., Garcia, B. A., The significance, development and progress of high-throughput combinatorial histone code analysis. *Cell. Mol. Life Sci.* 2010, *67*, 3983–4000.

[31] Cannon, J., Lohnes, K., Wynne, C., Wang, Y. et al., High-Throughput Middle-Down Analysis Using an Orbitrap. *J. Proteome Res.* 2010, *9*, 3886–3890.

[32] Garcia, B. A., What does the future hold for top down mass spectrometry? *J. Am. Soc. Mass Spectrom.* 2010, *21*, 193–202.

[33] Wu, S. L., Kim, J., Hancock, W. S., Karger, B., Extended range proteomic analysis (erpa): a new and sensitive LC-MS platform for high sequence coverage of complex proteins with extensive post-translational modifications-comprehensive analysis of beta-casein and epidermal growth factor receptor (EGFR). *J. Proteome Res.* 2005, *4*, 1155–1170.

[34] Zhou, H., Ning, Z. B., Starr, A. E., Abu-Farha, M., Figeys, D., Advancements in top-down proteomics. *Anal. Chem.* 2012, *84*, 720–734.

[35] Fischle, W., Tseng, B. S., Dormann, H. L., Ueberheide, B. M. et al., Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. *Nature* 2005, *438*, 1116–1122.

[36] Boyne, M. T., Garcia, B. A., Li, M. X., Zamdborg, L. et al., Tandem mass spectrometry with ultrahigh mass accuracy clarifies peptide identification by database retrieval. *J. Proteome Res.* 2009, *8*, 374–379.

[37] Kalli, A., Sweredoski, M. J., Hess, S., Data dependent middle-down nano-liquid chromatography-electron capture dissociation tandem mass spectrometry: an application for the analysis of unfractionated histones. *Anal. Chem.* 2013, *85*, 3501–3507.

[38] Young, N. L., DiMaggio, P. A., Plazas-Mayorca, M. D., Baliban, R. C. et al., High throughput characterization of combinatorial histone codes. *Mol. Cell Proteomics* 2009, *8*, 2266–2284.

[39] Phanstiel, D., Brumbaugh, J., Berggren, W. T., Conard, K. et al., Mass spectrometry identifies and quantifies 74 unique histone H4 isoforms in differentiating human embryonic stem cells. *Proc. Natl. Acad. Sci. USA* 2008, *105*, 4093–4098.

[40] Liu, T., Belov, M. E., Jaitly, N., Qian, W. J., Smith, R. D., Accurate mass measurements in proteomics. *Chem. Rev.* 2007, *107*, 3621–3653.

[41] Garcia, B. A., Mass spectrometric analysis of histone variants and post-translational modifications. *Front Biosci. (Schol. Ed.)* 2009, *1*, 142–153.

[42] Garcia, B. A., Siuti, N., Thomas, C. E., Mizzen, C. A., Kelleher, N. L., Characterization of neurohistone variants and post-translational modifications by electron capture dissociation mass spectrometry. *Int. J. Mass Spectrom.* 2007, *259*, 184–196.

[43] Garcia, B. A., Thomas, C. E., Kelleher, N. L., Mizzen, C. A., Tissue-specific expression and post-translational modification of histone H3 variants. *J. Proteome Res.* 2008, *7*, 4225–4236.

[44] Axelsson, J., Palmblad, M., Hakansson, K., Hakansson, P., Electron capture dissociation of substance P using a commercially available Fourier transform ion cyclotron resonance mass spectrometer. *Rapid Commun. Mass Spectrom.* 1999, *13*, 474–477.

[45] Creese, A. J., Cooper, H. J., Liquid chromatography electron capture dissociation tandem mass spectrometry (LC-ECD-MS/MS) versus liquid chromatography collision-induced dissociation tandem mass spectrometry (LC-CID-MS/MS) for the identification of proteins. *J. Am. Soc. Mass Spectrom.* 2007, *18*, 891–897.

[46] Kruger, N. A., Zubarev, R. A., Carpenter, B. K., Kelleher, N. L. et al., Electron capture versus energetic dissociation of protein ions. *Int. J. Mass Spectrom.* 1999, *182*, 1–5.

[47] Kruger, N. A., Zubarev, R. A., Horn, D. M., McLafferty, F. W., Electron capture dissociation of multiply charged peptide cations. *Int. J. Mass Spectrom.* 1999, *185*, 787–793.

[48] Zubarev, R. A., Kelleher, N. L., McLafferty, F. W., Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J. Am. Chem. Soc.* 1998, *120*, 3265–3266.

[49] Kalli, A., Hakansson, K., Electron capture dissociation of highly charged proteolytic peptides from Lys N, Lys C and Glu C digestion. *Mol. Biosyst.* 2010, *6*, 1668–1681.

[50] Mikesh, L. M., Ueberheide, B., Chi, A., Coon, J. J. et al., The utility of ETD mass spectrometry in proteomic analysis. *Biochim. Biophys. Acta* 2006, *1764*, 1811–1822.

[51] Liu, X., Inbar, Y., Dorrestein, P. C., Wynne, C. et al., Deconvolution and database search of complex tandem mass spectra of intact proteins: a combinatorial approach. *Mol. Cell Proteomics* 2010, *9*, 2772–2782.

[52] Carvalho, P. C., Xu, T., Han, X., Cociorva, D. et al., YADA: a tool for taking the most out of high-resolution spectra. *Bioinformatics* 2009, *25*, 2734–2736.

[53] 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.

[54] Hüttenhain, R., Hess, S., A combined top-down and bottom-up MS approach for the characterization of hemoglobin variants in Rhesus monkeys. *Proteomics* 2010, *10*, 3657–3668.

[55] Liu, X., Sirotkin, Y., Shen, Y., Anderson, G. et al., Protein identification using top-down. *Mol. Cell Poteomics* 2012, *11*, M111.008524.

[56] Zamdborg, L., LeDuc, R. D., Glowacz, K. J., Kim, Y. B. et al., ProSight PTM 2.0: improved protein identification and characterization for top down mass spectrometry. *Nucleic Acids Res.* 2007, *35*, W701–W706.

[57] DiMaggio, P. A., Jr., Young, N. L., Baliban, R. C., Garcia, B. A., Floudas, C. A., A mixed integer linear optimization framework for the identification and quantification of targeted post-translational modifications of highly modified proteins using multiplexed electron transfer dissociation tandem mass spectrometry. *Mol. Cell Proteomics* 2009, *8*, 2527–2543.

[58] Guan, S., Burlingame, A. L., Data processing algorithms for analysis of high resolution MSMS spectra of peptides with complex patterns of posttranslational modifications. *Mol. Cell Proteomics* 2010, *9*, 804–810.

[59] Liu, X., Hengel, S., Wu, S., Tolić, N. et al., in: Deng, M., Jiang, R., Sun, F., Zhang, X. (Eds.), *Research in Computational Molecular Biology*, Springer, Heidelberg 2013, pp. 132–144.

[60] Colinge, J., Masselot, A., Cusin, I., Mahe, E. et al., High-performance peptide identification by tandem mass spectrometry allows reliable automatic data processing in proteomics. *Proteomics* 2004, *4*, 1977–1984.

[61] Pitzer, E., Masselot, A., Colinge, J., Assessing peptide de novo sequencing algorithms performance on large and diverse data sets. *Proteomics* 2007, *7*, 3051–3054.

[62] Durbin, K. R., Tran, J. C., Zamdborg, L., Sweet, S. M. et al., Intact mass detection, interpretation, and visualization to automate top-down proteomics on a large scale. *Proteomics* 2010, *10*, 3589–3597.

[63] Venable, J. D., Dong, M. Q., Wohlschlegel, J., Dillin, A., Yates, J. R., Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat. Methods* 2004, *1*, 39–45.

[64] Gillet, L. C., Navarro, P., Tate, S., Rost, H. et al., Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell Proteomics* 2012, *11*, O111.016717.

