Breaking the Histone Code with Two-Dimensional Partial Covariance Mass Spectrometry

Taran Driver[†], Rüdiger Pipkorn[‡], Vitali Averbukh[†], Leszek J. Frasinski[†], Jon P. Marangos[†], Marina Edelson-Averbukh[*,†]

[†]Department of Physics, Imperial College London, SW7 2AZ London, UK
[‡]German Cancer Research Centre, Department of Translational Immunology, INF 580, 69120 Heidelberg, Germany

ABSTRACT: A large body of research points to the biological importance of combinatorial post-translational modifications in proteins, such as the active role played by histone modification patterns in the development of cancers, neurodevelopmental disorders, neurodegenerative and other diseases. Nonetheless, our understanding of the precise biological function of different modification patterns is limited by the difficulty of identifying and quantifying different combinatorial isomers in their mixtures as they naturally occur. Tandem mass spectrometry, which infers primary structure from the mass-to-charge ratios of biomolecular fragments, is the preferred method of analysis for proteins and their post-translational modifications. However, the information contained in the mass-to-charge ratios of the individual fragments is frequently insufficient to identify the correct set of modification patterns present in a mixture of combinatorial isomers. This is because no possible single fragment of a combinatorially modified sequence is unique to that sequence in its mass-to-charge ratio. Here we show that the combinatorial post-translational modification problem can be solved by the recently introduced technique of two-dimensional partial covariance mass spectrometry, which provides information about fragment connectivity in a biomolecule by quantifying correlations between the random intensity fluctuations of its fragments, across repeated measurements. Unique fragment-fragment correlations provide the missing link between the non-unique individual fragments to produce unambiguous fingerprints of co-occurring combinatorial isomers, enabling the discovery of biomolecular combinatorial modification patterns by mass spectrometry.

Combinatorial isomers are biomolecules with the same chemical sequence, which are multiply modified by the same number of identical covalent modifications distributed differently across a series of possible modification sites. It is now understood that different combinatorial isomers encode different biological functions. One example is patterns of DNA methylation, which are believed to serve as important epigenetic actors governing the gene expression [1,2]. The best studied systems of combinatorial isomers are the histone proteins, which can be heavily modified on their N-terminal tail [3]. Histones act as spools around which the DNA winds itself, and so function to compact and package the DNA into strands of a higher-order structure named chromatin. Over two decades ago it was demonstrated that enzymes which modify histones play a direct role in the regulation of gene expression [4,5], indicating the biological importance of histone

modification [6]. The initial proposition that modifications on histone tails can act sequentially or in concert to code for different biological functions [7,8] was based on a number of observations suggesting different downstream functions for different combinations of histone modifications [9,10,11,12]. This hypothesis is generally known as the 'histone code'.

The histone code hypothesis has driven a large-scale research effort towards the characterization and understanding of combinatorial modification patterns in order to map biological function to the underlying chemical structure. A critical mass of experimental evidence has been found to support the histone code hypothesis. For example, it was found that certain interaction sites of large proteins called bromodomains, which can act to remodel chromatin and control transcription, specifically interact with particular motifs of histone tail modifications [13]. Multiple further examples of chromatin-modifying proteins talking with particular patterns of histone tail modifications have been reported [14, 15, 16, 17, 18, 19].

Tandem mass spectrometry (MS/MS) is the method of choice for analysis of the primary structure of proteins and in particular their post-translational modifications [20]. This is because a modification typically changes the mass of a residue in a manner which is specific to the modifying group [21]. In a standard tandem mass spectrometry experiment, components of a biological sample separated by high performance liquid chromatography (LC) [22], are first put into the gas phase using a soft ionization method, e.g. electrospray ionization (ESI) [23]. Individual ionized molecules of the co-eluted components are further separated in a mass spectrometer on the basis of their mass-to-charge (m/z) ratio and subsequently fragmented by one of a number of available activation methods, e.g. collision induced dissociation (CID) [24] or electron transfer dissociation (ETD) [25]. Finally, the m/z ratios of the generated fragments are measured by a mass analyzer [20]. The result is a 1D spectrum of relative abundance *vs.* fragment mass-to-charge ratios. The intact mass and fragment ion spectral data of a biomolecule provide the information used to infer its original primary structure.

Mixtures of combinatorial isomers present a particular challenge to the LC-MS/MS approach. Firstly, the efficiency of chromatographic separation techniques depends on the differences in the physical properties of the eluted compounds, which are negligible for combinatorial isomers, frequently leading to their co-elution [26]. Next, all unresolved combinatorial isomers in a mixture are doomed to be isolated together within the mass spectrometer, because they all have an identical mass. Finally, the

m/z degeneracy of the fragment ions produced by a series of combinatorial isomers may not allow for a full set of unique "marker ions", i.e. fragment ions specific to each single isomer within the combinatorial isomer mixture. Instead, each possible fragment ion may be produced by more than one single combinatorial isomer, precluding the exact identification of all the isomers present in the mixture of a combinatorially modified sequence.

There have been a large number of research efforts focused on 'breaking' the histone code by mass spectrometry [3, 26-37]. However, the most significant remaining problem in the mass spectrometric analysis of combinatorial modifications is the resolution and quantification of individual combinatorial isomers

between fragments born in the same or consecutive decomposition processes of an ionized biomolecule. Because of inherent randomness of the dissociation process and experimental parameter instabilities, the abundances of different fragments in an MS/MS measurement fluctuate from scan to scan. By mapping how the fluctuations in the yield of each fragment are synchronized across repeated scans, we identify which pairs of fragments are produced in the same/consecutive decompositions and therefore belong to the same parent ion. Whilst an individual fragment ion could have been produced by a number of different combinatorial isomers, there are such pairs of co-produced fragments which could only have originated from one of them. As we show here, these fragment-fragment correlation signals
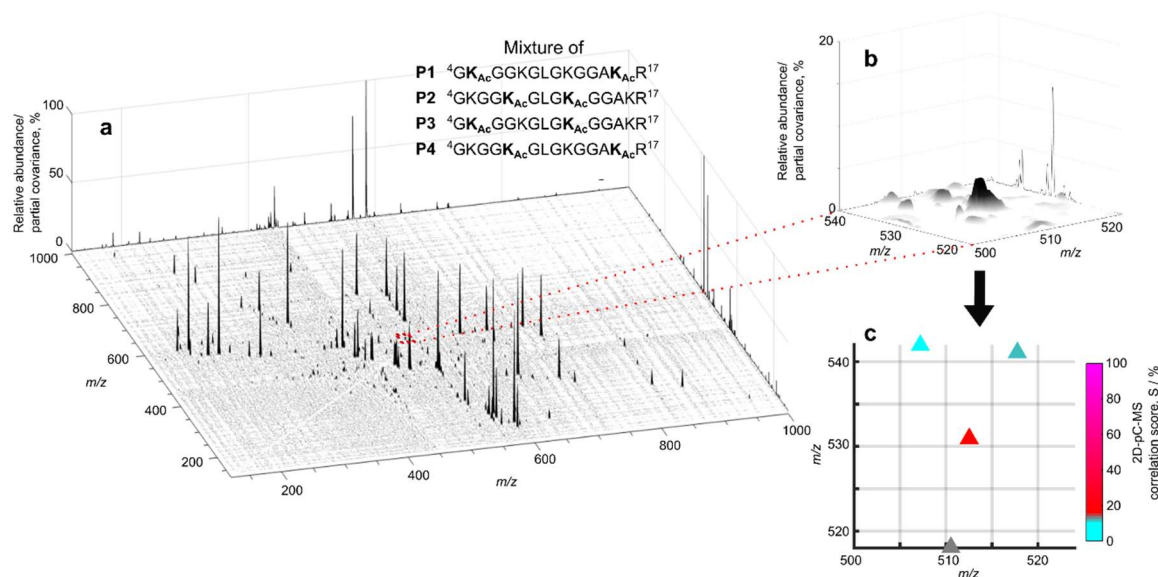

Figure 1. (a) 2D-pC-MS map of the CID measurement of the mixture of four diacetylated isomers, **P1**-**P4**. Peaks at the intersection of (m_1/z_1, m_2/z_2) indicate that the two fragment ions are born in the same or consecutive decompositions or of the same parent ion. Autocorrelation peaks $m_1/z_1 = m_2/z_2$ are removed for clarity. The corresponding 1D CID spectrum is plotted against the back walls of the map. (b) The (m/z 500–524) × (m/z 518–542) region of the map in (a), exhibiting the marker ion correlation of **P4** (large central feature). (c) correlation scores \cite{2D-pC-MS} of the main peaks in the region shown in (b).

[26, 29, 30]. Abshiru et al. [29] compared the overlapping MS/MS spectra of different, co-fragmented isomers to the individually obtained spectra of every possible synthetic isomer in its pure form, and attempted to deconvolve the mixed spectra by modelling the deconvolution as a network flow problem. Although this was successful for some mixtures, the deconvolution and quantification failed in more complex cases, such as the mixtures of diacetylated isomers of the H4 histone tail that we consider in this work. Additionally, the requirement for synthesis of the pure form of each and every individual isomer is highly restrictive. Using MS/MS, Phanstiel et al. [30] solved systems of coupled linear equations to deconvolve the mixed spectra of co-eluting histone structural isomers, but this method also failed in more complex cases such as mixtures of diacetylated H4 histone tails, the deconvolution and quantification of which was conceded to be "mathematically impossible" [30].

Here, we present a new way to resolve mixtures of combinatorial isomers that can not only tackle the more complex cases but does so without requiring synthesis of any model sequences. The new approach is based on the recently developed two-dimensional partial covariance mass spectrometry (2D-pC-MS). The physical principle of 2D-pC-MS is described in detail in Ref. [38]. Briefly, the technique builds a 2D map of correlations

("marker correlations") can be used to straightforwardly resolve complex mixtures of different combinatorial isomers.

The 2D-pC-MS measurement of a mixture of four positional isomers of diacetylated histone H4 fragment 4–17, $^4GK_5GGK_8GLGK_{12}GGAK_{16}R^{17}$ is given in Fig. 1a. The two horizontal axes are the same quantity, which is the *m/z* ratio of the fragment
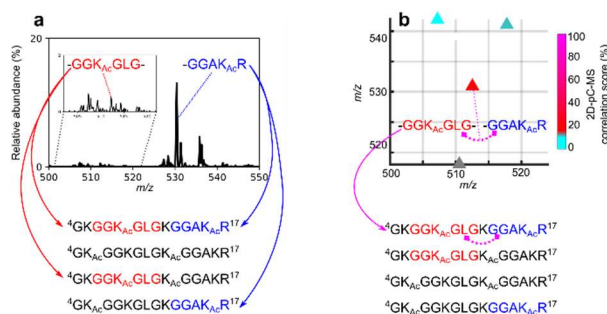

Figure 2. (a) Signals in the 1D mass spectrum could have been produced by more than one combinatorial isomer, making it challenging and often in principle impossible to identify components of a mixture from the 1D mass spectrum. (b) The 2D signal which correlates two *m/z* fragments can only have been produced by one isomer, providing a marker correlation for that isomer.

ion spectrum. The value of the surface is the partial covariance between the two m/z values on the corresponding x- and y-axes [38]. The peaks on the map (Fig. 1a) correspond to pairs of fragments produced in the same or consecutive decomposition processes. The volume of the partial covariance peak is directly proportional to the probability of the decomposition process which produced the two correlated fragments [38]. Obtaining the 2D-pC-MS map requires no hardware modifications and so is immediately realizable on a standard benchtop linear ion trap mass spectrometer.

Here we demonstrate the power of 2D-pC-MS to resolve mixtures of combinatorial isomers by solving the long-standing problem of diacetylated positional isomers of histone H4 [39]. The lysines at positions 5, 8, 12 and 16 are known to be acetylated [40]. Whilst 5 & 8 and 12 & 16 isomers of the diacetylated H4 peptide are straightforwardly identifiable using standard MS/MS [39], resolution of the other 4 positional isomers, shown in Fig. 1a, has been branded as 'mathematically impossible' [30]. As illustrated in Fig. 2a, in the analysis of a fragment ion spectrum of a mixture of the four isomers, detection of the blue C-terminal fragment GGAK$_{Ac}$R provides evidence for the presence of either the K$_{5,Ac}$K$_{16,Ac}$ or K$_{8,Ac}$K$_{16,Ac}$ isomer but fails to differentiate between them. Equally, detection of the internal fragment GGK$_{Ac}$GLG indicates the presence of either isomer K$_{8,Ac}$K$_{16,Ac}$ or isomer
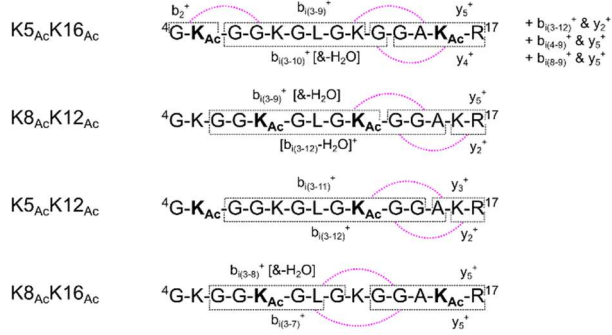

Figure 3. Marker correlations measured for the four diacetylated combinatorial isomers, **P1-P4** (see Fig. 1) considered in this work. Correlating fragments are designated and their correlations are shown by magenta arcs.

K$_{8,Ac}$K$_{12,Ac}$, but is unable to determine which of the two are present. In contrast, if 2D-pC-MS detects the correlation between these two fragments (see Fig. 2b), this means that both fragments are originated from the same parent molecule. Since these two fragments can only be produced together by the single isomer K$_{8,Ac}$K$_{16,Ac}$, this correlation unambiguously confirms the presence of that combinatorial isomer in the mixture. Thus, within 2D-pC-MS, two non-unique individual fragments are connected together to produce a unique fragment-fragment correlation, or a "marker correlation", of a particular combinatorial isomer.

Figure 3 shows the unique marker correlations of the four combinatorial isomers revealed by 2D-pC-MS in a standard linear ion trap CID experiment. Each of the presented correlations between the two fragment ions can only have been produced by one of the isomers. In Fig. 4, mixtures of **a**) two, **b**) three and **c**) four of the isomers have been analyzed by 2D-pC-MS. It is sufficient to interrogate one of the marker correlations (Fig. 3) per isomer to identify components of each of the analyzed mixtures. Probing these marker correlations directly reveals the presence or absence of each of the combinatorial isomers across the different mixtures, as shown in Fig. 4.

In addition to providing a method to qualitatively discriminate between different combinatorial isomers, the partial covariance between two fragment ion signals provides a quantitative measurement of the abundance of the peptide under analysis. This follows from the theory of partial covariance mapping using
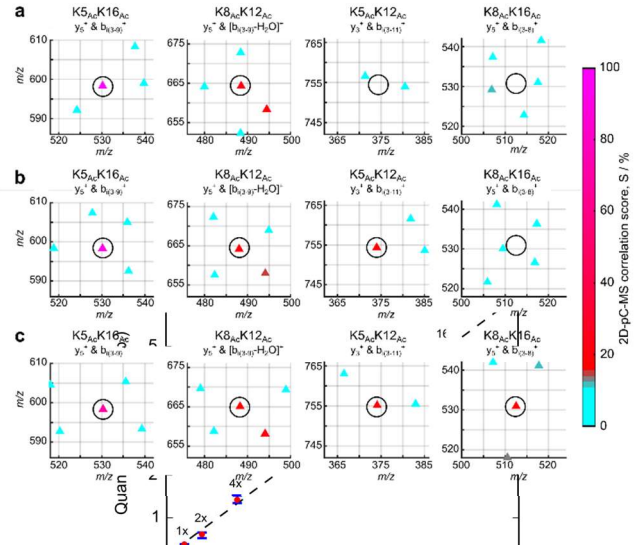

Figure 4. Resolution of different mixtures of combinatorial isomers. For each isomer, one of the marker correlations given in Fig. 3 is shown. These directly reveal the presence or absence of the isomer in a mixture.

the total ion count as a single correction parameter, as presented in [38]. According to the theory, the measured value of any marker correlation provides a quantitative measure of the concentration of the corresponding parent ion. In principle, this requires monitoring of only one marker correlation for the quantification of its specific isomer. Figure 5 shows the experimental verification of this theoretical prediction. It displays the measured value of partial covariance between fragments y$_{5,Ac}^+$ & b$_{i(3-9)}^+$, which is a marker correlation of the isomer K$_{5,Ac}$K$_{16,Ac}$. The relative concentration of this isomer ('Quantity added') has been increased in a mixture of four combinatorial isomers and the measured value ('Quantity measured') of partial covariance rises linearly with the increased concentration. The relative concentration of a particular isomer is obtained by dividing the volume of its marker correlation peak by the volume of any reference partial covariance peak, stemming from another isomer, which cannot have been produced by the parent ion in question. The error bars show the dependence of measured value on the reference peak chosen. This demonstration paves the way
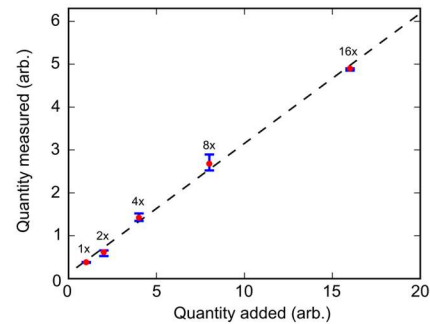

Figure 5. Linear mapping of the measured value of partial covariance between fragments y$_{5,Ac}^+$ & b$_{i(3-9)}^+$ with concentration of isomer K$_{5,Ac}$K$_{16,Ac}$, in a mixture of all four combinatorial isomers.

for the quantification of the combinatorial isomers in their mixtures by 2D-pC-MS.

In conclusion, we have presented a new approach for resolving complex mixtures of combinatorial isomers using tandem mass spectrometry. The technique presented here relies on measuring fragment-fragment correlations that are unique to each possible isomer in a mixture. While demonstrated for diacetylated positional isomers of histone H4 fragment 4–17, previously deemed to pose an intractable problem, the approach is general it can be straightforwardly extended to the analysis of any positional isomers. For example, as 2D-pC-MS has been demonstrated for oligonucleotides [38], the presented approach can be used to study combinatorial patterns of DNA methylation [1,2]. The method presented here should therefore provide key insights into the important biological function of different patterns of protein and nucleotide modification.

AUTHOR INFORMATION

Corresponding Author

*Email: m.edelson-averbukh@imperial.ac.uk

ACKNOWLEDGEMENTS

REFERENCES

1. A. Bird, "DNA methylation patterns and epigenetic memory", Genes Dev 16, 6 (2002).
2. C. Papin, A. Ibrahim, S. Le Gras, A. Velt, I. Stoll, et al., "Combinatorial DNA methylation codes at repetitive elements", Genome Research 27, 934 (2017).
3. L.-M. P. Britton, M. Gonzales-Cope, B. M. Zee, and B. A. Garcia, "Breaking the histone code with quantitative mass spectrometry", Expert Review of Proteomics 8, 631 (2011).
4. J. Taunton, C. A. Hassig, and S. L. Schreiber, "A Mammalian Histone Deacetylase Related to the Yeast Transcriptional Regulator Rpd3p", Science 272, 408 (1996).
5. J. E. Brownell, J. Zhou, T. Ranalli, R. Kobayashi, D. G. Edmondson, et al., "Tetrahymena histone acetyltransferase A: A homolog to yeast Gcn5p linking histone acetylation to gene activation", Cell 84, 843 (1996).
6. C. Sawan and Z. Herceg, "Histone Modifications and Cancer", Adv. Gen. 70, 57 (2010)
7. B. D. Strahl and C. D. Allis, "The language of covalent histone modifications", Nature 403, 41 (2000).
8. T. Jenuwein and C. D. Allis, "Translating the histone code", Science 293, 1074 (2001).
9. Y. Wei, L. Yu, J. Bowen, M. A. Gorovsky, and C. D. Allis, "Phosphorylation of histone H3 is required for proper chromosome condensation and segregation.", Cell 97, 99 (1999).
10. M. J. Hendzel, Y. Wei, M. A. Mancini, A. Van Hooser, T. Ranalli, et al., "Mitosis-specific phosphorylation of histone H3 initiates primarily within pericentromeric heterochromatin during G2 and spreads in an ordered fashion coincident with mitotic chromosome condensation", Chromosoma 106, 348 (1997).
11. M. J. Barratt, C. A. Hazzalin, E. Cano, and L. C. Mahadevan, "Mitogen-stimulated phosphorylation of histone H3 is targeted to a small hyperacetylation-sensitive fraction.", Proceedings of the National Academy of Sciences of the United States of America 91, 4781 (1994).
12. S. Thomson, L. C. Mahadevan, and A. L. Clayton, "MAP kinase-mediated signalling to nucleosomes and immediate-early gene induction", Seminars in Cell and Developmental Biology 10, 205 (1999).
13. P. Filippakopoulos, S. Picaud, M. Mangos, T. Keates, J. P. Lambert, et al., "Histone recognition and large-scale structural analysis of the human bromodomain family", Cell 149, 214 (2012).
14. S. L. Berger, "The complex language of chromatin regulation during transcription", Nature 447, 407 (2007).
15. A. Zippo, R. Serafini, M. Rocchigiani, S. Pennacchini, A. Krepelova, et al., "Histone Crosstalk between H3S10ph and H4K16ac Generates a Histone Code that Mediates Transcription Elongation", Cell 138, 1122 (2009).
16. J. Eswaran, D. Patnaik, P. Filippakopoulos, F. Wang, R. L. Stein, et al., "Structure and functional characterization of the atypical human kinase haspin", Proceedings of the National Academy of Sciences of the United States of America 106, 20198 (2009).
17. Y. Li, L. Sun, Y. Zhang, D. Wang, F. Wang, et al., "The histone modifications governing TFF1 transcription mediated by estrogen receptor", Journal of Biological Chemistry 286, 13925 (2011).
18. W.-S. Lo, R. C. Trievel, J. R. Rojas, L. Duggan, J. Y. Hsu, et al., "Phosphorylation of serine 10 in histone H3 is functionally linked in vitro and in vivo to Gcn5-mediated acetylation at lysine 14", Molecular Cell 5, 917 (2000).
19. W.-S. Lo, L. Duggan, N. C. T. Emre, and S. L. Berger, "Snf1 — a Histone Kinase That Works in Concert with the Histone Acetyltransferase Gcn5 to Regulate Transcription", Science 293, 1142 (2001).
20. D. F. Hunt, J. R. Yates 3rd, J. Shabanowitz, S. Winston and C. R. Hauer, "Protein sequencing by tandem mass spectrometry", Proc. Natl. Acad. Sci. USA 83, 6233 (1986).
21. J. Y. Kim, K. W. Kim, H. J. Kwon, D. W. Lee and J. S. Yoo, "Probing Lysine Acetylation with a Modification-Specific Marker Ion Using High-Performance Liquid Chromatography/Electrospray-Mass Spectrometry with Collision-Induced Dissociation", Anal. Chem. 74, 5443 (2002).
22. M. Holčapek, R. Jirásko and M. Lísa, "Recent developments in liquid chromatography–mass spectrometry and related techniques", J. Chromatography A 1259, 3 (2012).
23. J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, and C. M. Whitehouse, "Electrospray ionization for mass spectrometry of large biomolecules.", Science 246, 64 (1989).
24. A. K. Shukla and J. H. Futrell, "Tandem mass spectrometry: dissociation of ions by collisional activation", J. Mass Spectrom. 35, 1069 (2000).
25. J. E. Syka, J. J. Coon, M. J. Schroeder, J. Shabanowitz and D. F. Hunt, "Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry", Proc. Natl. Acad. Sci. USA 101, 9528 (2004).
26. N. L. Young, P. A. DiMaggio, M. D. Plazas-Mayorca, R. C. Baliban, C. A. Floudas, et al., "High Throughput Characterization of Combinatorial Histone Codes", Molecular & Cellular Proteomics 8, 2266 (2009).
27. R. Y. Tweedie-Cullen, A. M. Brunner, J. Grossmann, S. Mohanna, D. Sichau, et al., "Identification of combinatorial patterns of post-translational modifications on individual histones in the mouse brain", PLoS ONE 7 (2012).
28. A. Moradian, A. Kalli, M. J. Sweredoski and S. Hess, "The top-down, middle-down, and bottom-up mass spectrometry approaches for characterization of histone variants and their post-translational modifications", Proteomics 14, 489 (2014).
29. N. Abshiru, O. Caron-Lizotte, R. E. Rajan, A. Jamai, C. Pomies, et al., "Discovery of protein acetylation patterns by deconvolution of peptide isomer mass spectra", Nature Communications 6, 1 (2015).
30. D. Phanstiel, J. Brumbaugh, W. T. Berggren, K. Conard, X. Feng et al., "Mass spectrometry identifies and quantifies 74 unique

histone H4 isoforms in differentiating human embryonic stem cells", Proc. Natl. Acad. Sci., USA 105, 4093 (2008).

31. N. Siuti and N. L. Kelleher, "Decoding protein modifications using top-down mass spectrometry", Nature Methods 4, 817 (2007).

32. X. Dang, J. Scotcher, S. Wu, R. K. Chu, N. Tolić, et al., "The first pilot project of the consortium for top-down proteomics: A status report", Proteomics 14, 1130 (2014).

33. C. E. Thomas, N. L. Kelleher, and C. A. Mizzen, "Mass spectrometric characterization of human histone H3: a bird's eye view", J.Proteome Res. 5, 240 (2006).

34. A. P. Lothrop, M. P. Torres, and S. M. Fuchs, "Deciphering post-translational modification codes", FEBS Letters 587, 1247 (2013).

35. B. A. Garcia, J. J. Pesavento, C. A. Mizzen, and N. L. Kelleher, "Pervasive combinatorial modification of histone H3 in human cells", Nature Methods 4, 487 (2007).

36. C. Feller, I. Forné, A. Imhof and P. B. Becker, "Global and Specific Responses of the Histone Acetylome to Systematic Perturbation", Mol. Cell 57, 559 (2015).

37. J. J. Pesavento, C. R. Bullock, R. D. LeDuc, C. A. Mizzen and N. L. Kelleher, "Combinatorial modification of human histone H4 quantitated by two-dimensional liquid chromatography coupled with top down mass spectrometry", J. Biol. Chem. 283, 14927 (2008).

38. T. Driver, B. Cooper, R. Ayers, R. Pipkorn, S. Patchkovskii et al., "Two-dimensional mass spectrometry of large molecules based on fragment correlations", Phys. Rev. X, *in press*.

39. Y. Zheng, X. Huang and N. L. Kelleher, "Epiproteomics: quantitative analysis of histone marks and codes by mass spectrometry", Current Opinion in Chemical Biology 33, 142 (2016).

40. C. M. Smith, P. R. Gafken, Z. Zhang, D. E. Gottschling, J. B. Smith and D. L. Smith, "Mass spectrometric quantification of acetylation at specific lysines within the amino-terminal tail of histone H4", Anal. Biochem. 316, 23 (2003).