

Reading and writing	650
Finding the marks	650
The sum of the marks	652
Box 1: Managing mass spectrometry	651

# Mass spectrometry for chromatin biology

Monya Baker

To discover new histone marks and interactions, researchers turn to the sophisticated instruments of proteomics.

Investigators intent on cracking one code may overlook another. So back in the days when the DNA sequence alone was believed to hold all cellular secrets, histones, the protein cores that spool DNA, were dismissed as little more than passive packing material. Now, that packing material is studied as a crucial component of genetic regulation.

A histone complex is made of eight small proteins: two each of H2A, H2B, H3 and H4. One-fifth to one-third of each protein's amino acids dangle out from the core; these contain many positively charged residues that might serve to attract negatively charged DNA. In the 1990s, researchers such as C. David Allis showed that these histone tails are more than indiscriminate DNA lures: they are platforms for sophisticated signals that are communicated via chemical modifications that decorate the

amino acids. These histone marks subtly govern DNA packing and guide specific protein complexes controlling gene expression. Allis and his former postdoctoral student Brian Strahl went on to propose the existence of a 'histone code' of chemical signals for gene transcription and gene repression<sup>1</sup>.

The field is still in its early days, says Strahl, now a biochemist at the University of North Carolina, Chapel Hill. "The more we scratch the surface, the more doors open," he says. "As much as the field has learned over the past two decades, we hardly know anything."

To study histone modifications, researchers use a variety of tools. Perhaps the most widespread technique is using antibodies to pull down particular histone marks and associated DNA for subsequent analysis. This technique, called ChIP-Seq (chromatin immunoprecipitation followed by sequencing), can reveal which marks are associated with particular genetic sequences. A series of genome-wide studies overlaying sequences associated with dozens of marks in different cell types have found combinations representing chromatin states linked to repression, long-term silencing and active transcription. (See <http://www.nature.com/nmeth/journal/v8/n9/full/nmeth.1673.html>.)

But as powerful as antibody-based techniques are, they also have limitations of variability, sensitivity and specificity. An antibody made to recognize particular modifications can confuse one mark for another (such as a monomethylated lysine for a trimethylated lysine) or fail to find its mark if other marks are nearby. More fundamentally, antibodies can only be used to enrich for known modifications. They cannot discover new marks, quantify marks in

an unbiased way or detect whether marks occur together on the same histone or individually on several.

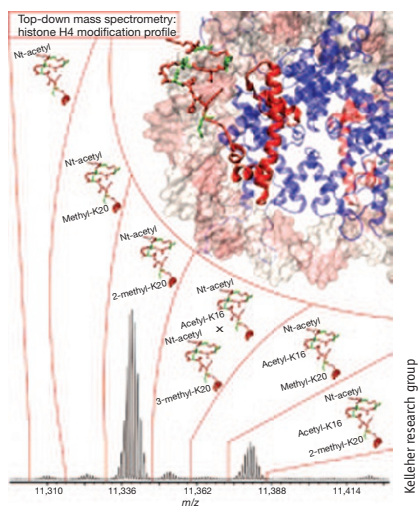
Such investigations require mass spectrometry (MS), says Strahl. "Mass spectrometry really is the fundamental tool to identify these modifications." The technique has become a powerful force in the chromatin world<sup>2</sup>. "If you go to an MS conference, there will be a histone alley," says Neil Kelleher, a scientist at Northwestern University who is known for mapping many histone variants.

The questions MS can address are diverse. Quantitative labeling techniques can find the proteins that bind modifications as well as monitor how modifications change over time. So-called 'bottom-up' techniques are best for detecting the rarest modifications. 'Top-down' or 'middle-down' techniques can detect multiple modifications on the same histone.

## The language of histones

Modifications occur all along the histone tails and even down into the core. Three methyl groups on a lysine that is four amino acids from the exposed tip of H3 (noted in common shorthand as H3K4me3) indicate the promoter of an actively transcribed gene. Three methyl groups on the 9th (H3K9me3) or 27th (H3K27me3) lysine indicates repression. And depending on the particular amino acid, more modifications are possible: besides mono-, di- and trimethylation, there are also acetylation, phosphorylation, sumoylation, ubiquitylation and more.

Hundreds of proteins have been found to interact with these modifications. 'Writers', such as histone acetyltransferases and methyltransferases, place marks on histones. 'Erasers', such as lysine demethylases and



Mass spectrometry can reveal the many chemical modifications that decorate histones. Nt, (N)-terminal or amino terminus.

Kelleher research group

histone deacetylases, remove them. ‘Readers’ bind to marks and signal gene-reading and gene-packing machinery.

Modifications are implicated in the most basic biological processes as well as common diseases. They appear to guide pluripotent cells through differentiation. Drugs that inhibit histone deacetylases are used against cancer and mental diseases. More than a dozen other inhibitors are in clinical trials for indications including cancer, inflammation and neurodegenerative disorders<sup>3</sup>.

### Reading and writing

Chromatin biologist Michiel Vermeulen at the University Medical Center Utrecht was drawn into MS because he realized that it was the most effective way to find histone readers. “To visualize those specific interactions, there was a need for a quantitative filter,” explains Vermeulen. “If you run the protein mixture out on a gel, you won’t see them because they are masked by background. Quantitative MS seemed like a perfect tool.” But sophisticated MS techniques are not in the typical repertoire of protein biochemists. Vermeulen accepted a postdoc with Matthias Mann at the Max Planck Institute of Biochemistry, Martinsried to learn the ropes and gain access to high-end instruments.

Mann is known for a technique called SILAC (stable isotope labeling with amino acids in cell culture), which allows proteins from different cell populations to be compared quantitatively. Typically, one cell culture is fed nutrients containing heavy carbon atoms in the form of arginine and lysine, whereas the other is fed common ‘light’ isotopes. The cell lysates are mixed together, and the ratio of heavy and light proteins reveals relative abundance in each population.

Instead of growing cells under different conditions, Vermeulen and colleagues compare histone modifications in differently labeled cell lysates. “We take light and heavy cells, make extracts, incubate the lysates with different baits and then use qMS to see which peptides react with the modified peptide more than the unmodified peptide,” he explains.

The baits are synthetic versions of histone tails. In one set of experiments, Vermeulen made five baits that each contained a common trimethyl mark: two associated with gene activation (H3K4me3 and H3K36me3) and three with repression (H3K9me3, H3K27me3 and H4K20me3). This revealed a host of bound proteins, including previously



Michiel Vermeulen uses quantitative mass spectrometry to find histone readers.

the activating mark H3K4me3 (ref. 4).

Vermeulen’s group also used three sets of cell extracts (labeled with ‘heavy,’ ‘medium’ and ‘light’ amino acids) to compare unmodified, singly modified and multiply modified baits. In one set of experiments, they tested the marks H3K27me3 and H3K9me3 in the presence of H3S28 or H3S10 phosphorylation, respectively. MS showed that different readers responded differently. HP1, an important regulator of chromatin packing and gene repression, was unaffected by phosphorylation, but ORC bound phosphorylated baits less strongly. Because ORC is known to bind all repressive marks, this observation hints at an unrecognized layer of regulation, says Vermeulen. “It allows a sophisticated fine-tuning of the interactions.”

Nucleosomes, which comprise the histone complex along with DNA, can also be reconstituted and used as baits. These allow investigation of DNA methylation and histone modifications that are far apart from each other. One such study revealed that several readers that normally recognized histone modifications, including ORC, bound less strongly if the DNA around the histone was methylated<sup>5</sup>.

The next step, says Vermeulen, is to find out how readers vary between cell types. So far, studies have been conducted in HeLa cancer cell lines, but Vermeulen is now studying embryonic stem cells and differentiated cells to discover whether the reader complexes vary. Another avenue is to study histone complexes at different points in the cell cycle or cells that have been exposed to DNA-damaging conditions.

Vermeulen uses orbitrap instruments in his work because of their speed, sensitivity and high mass accuracy—they can sequence thousands of peptides in a matter of hours, he says. However, more sophisticated data acquisition could make studies more efficient. Right now much of the data gathered represent the 99% of peptides that are not

known readers and dozens of new candidates; one of these bound proteins was the origin-of-recognition complex (ORC), which was found with all three repressive marks. Another was the SAGA complex, a transcription activator that binds to

enriched for one bait or another; ignoring these “boring proteins” would vastly improve throughput, he says. “That is something that the coming years will significantly improve.”

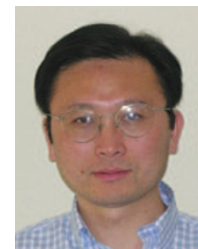
In addition to finding readers, writers and erasers, SILAC can be used to monitor the actions of such proteins. Rather than feeding two sets of cells labeled amino acids and comparing them, researchers such as Ben Garcia at the University of Pennsylvania are studying how heavy amino acids are incorporated over time as cells divide, differentiate or just grow. It is unexplored terrain, he says. “No one really understands how fast histone modifications turn over, how fast they can be induced or how long the half-lives are.” Such information is critical for understanding potential cancer drugs that work by inhibiting these modifications, he adds.

### Finding the marks

Other implementations of MS are used to find new modifications. In a typical workflow for bottom-up proteomics, researchers purify proteins from cell lysates and digest them into peptides. These are fractionated so they can be introduced a few at a time into a mass spectrometer. Once inside the machine, peptides are separated again and fragmented into tinier pieces. After the mass-to-charge ratio is measured for as many pieces as possible, the spectra for observed fragments are matched up with those predicted based on protein sequence and past experiments.

Unlike protein sequences, post-translational modifications cannot be predicted from genetic information alone. However, modifications alter the mass and sometimes the charge of a peptide, and bioinformatics tools can be programmed to look for these telltale changes.

The broader the range of modifications studied, the more difficult detection is. “It is not difficult to detect a mass shift caused by a post-translational modification, but it is not easy to transform from the mass shift to chemical structures because there are multiple structural possibilities that can lead to the same mass shift,” says Yingming Zhao, a proteomics researcher at the University of Chicago.



Yingming Zhao hunts for new histone modifications.

"If you don't know [which marks are on what residues], you need to do unrestricted sequence alignment, meaning that you scan all the possible modifications."

Such sequence alignment requires not only sophisticated data analysis but also getting the very best data possible, which in turn requires highly sensitive analysis once peptides have been digested. Zhao uses a technique called isoelectric focusing to separate molecules by charge and performs tandem MS using a nanoflow HPLC/LTQ orbitrap.

Last year alone, researchers led by Zhao identified three new types of histone modification: crotonylation, succinylation and malonylation. Previously, his team had found propionylation and butyrylation, both on lysines. There are many enzymes classified as lysine deacetylases without known substrates, and butyryl and propionyl are very similar to acetyl groups, leading Zhao to suspect that these modifications existed. "So we tested," he says, "and we found them on the histones."

Another modification was unanticipated. Zhao's lab had programmed software to look for amino acids with the added weight of a butyryl group and was using it in follow-up studies. Unexpectedly, although their analysis identified mass peaks with the heaviest isotopes as butyryl, it was unable to classify accompanying peaks representing common isotopes. Further inspection (and follow-up experiments) revealed that the modifications were actually crotonyl groups, which are 2 Da lighter than butyryl groups. Genome-wide studies using modification-specific antibodies showed that the mark was associated with genomic DNA sequences distinct from those that undergo histone acetylation<sup>6</sup>. In fact, crotonylation is particularly enriched in testis-specific genes, which hints at potential functions of this modification.

But discovering the mark is just the first step to understanding biology, says Zhao. "When we find a new modification, we want to find the enzymes that put it there or the proteins that bind to it." After that, the goal is to find associations with disease or biological processes.

Such biology is starting to unfold for *N*-acetylglucosamine, or *O*-GlcNAc. Gerald Hart, director of biological chemistry at Johns Hopkins School of Medicine, became convinced that this derivative of glucose was important for cell signaling decades ago, when a graduate student in his lab studying how lymphocytes communicate found

## BOX 1 MANAGING MASS SPECTROMETRY

Histones are hard to analyze. In 2007, the *International Journal of Mass Spectrometry* dedicated a special issue to Hunt for working out ways to study histone modifications.

It was a problem Hunt was initially leery to tackle. Allis, who had been working out the function of histone modification and wanted more precise measurements, brought purified histones to Hunt. But the analysis did not produce the characteristic sparse peaks of an analytic spectrum. "All I saw was a blob," Hunt recalls. "I looked at it and decided I didn't want [those proteins] anywhere near my mass spectrometer." Luckily, curiosity and a good protein digest protocol changed his mind.

The reason for the "many, many overlapping signals" Hunt observed was that even highly pure histones are actually a complex mixture of thousands of forms, all with different masses. As a further complication, histones contain high levels of the amino acids lysine and arginine, and so treatment with trypsin, which cleaves peptide bonds at these residues, produces short, overlapping and highly positively charged peptides that are not retained well by chromatography and so don't enter the mass spectrometer. "Some of the most interesting modifications never make it; they get washed off the column," explains Hunt.

New versions of chromatography are helping. Weak cation exchange, hydrophilic interaction liquid chromatography, developed by Garcia and colleagues, reverses the normal order by first eluting the most hydrophobic residues. The positively charged residues typical of histones can be slowly washed off with formic acid, thus efficiently fractionating species that have similar mass but different charge. Another option is to treat histones chemically to protect the sites trypsin would normally cleave. In particular, a propionyl group can be added to unmodified lysines so that trypsin cannot cut there.

Other innovations involve how peptides are fragmented inside the mass spectrometer. Commonly used triple quadrupole instruments rely on collisions with inert gas molecules, but this technique can remove modifications such as phosphorylation. Alternatively, the Fourier-transform ion cyclotron (FT-ICR) and some time-of-flight instruments can use electrons for fragmentation; this tends to break proteins along the backbone, leaving chemical modifications intact. Electron-transfer dissociation, a technique Hunt and colleagues at the University of Virginia developed, pairs readily with liquid chromatography and is used with ion traps, orbitraps and some time-of-flight instruments. Another technique, electron-capture dissociation, generally requires higher-end magnet-based FT-ICR but is also available in some time-of-flight instruments.

Much still depends on separating highly similar proteins: mass spectrometers become more powerful when samples contain fewer species. "The issue with histones is that if you use standard chromatography, all the modified forms loop together," explains Garcia. Researchers specializing in marks such as acetylation, phosphorylation or ubiquitination can enrich for such modifications with specialized chromatography or custom-made antibodies. Though enrichment introduces bias, it also reduces background.

Indeed, says Hunt, some of the most important future innovations will involve what happens to a sample before it enters a mass spectrometer. "The modifications we are missing are probably the ones that are at low levels, and perhaps transient," he says. "Any approach that allows you to isolate or enrich a subset of biologically important material before you get to the mass spectrometer is important."



Don Hunt in front of the first mass spectrometer that performs electron-transfer dissociation, a technique for retaining post-translational modifications.

that most of the sugars were not on the cell surface but in the interior of cells. He began investigating *O*-GlcNAc as a post-translational modification using tactics much like those used to study phosphorylation, manipulating the enzymes that take the group on and off other proteins. Eventually, another graduate student, Kaoru Sakabe, suggested looking for the modification on histones, and found them<sup>7</sup>.

But before such experiments became possible, the researchers needed better techniques. Lysing cells activates enzymes that remove *O*-GlcNAc; standard peptide fragmentation breaks the modification off the peptide, and the spectrometry signal from any still-modified peptides is suppressed by the more readily ionized unmodified peptide.

For Hart, solving these problems took many steps and a collaboration with Donald Hunt at the University of Virginia. The work also illustrates many of the MS challenges in studying post-translational modifications (Box 1). The first step was to modify *O*-GlcNAc so that modified peptides could be enriched. More specifically, the researchers used click chemistry to add an azido moiety to *O*-GlcNAc. This allows the peptides to be biotinylated with a photocleavable tag and captured on a streptavidin-coated column. Ultraviolet light releases the peptides from their biotin tag so that they can be eluted and collected. What's more, the tag that contains the azido group adds an additional positive charge to peptides, which makes them more visible during MS<sup>8</sup>.

Since the initial work showing that histones are modified with *O*-GlcNAc, several more papers from Hart's lab and elsewhere show that the mark cycles on and off, and that it seems to coordinate with activity of other histone marks. "It also affects methylation and acetylation and phosphorylation and ubiquitination. The cross-talk is going to be incredibly important for understanding this," says Hart.

### The sum of the marks

To understand cross-talk, it is crucial to know what modifications occur together on the same protein. Combinations of histone marks are the norm. "A single site by itself is almost never found on a histone; it's always found in combination with other modifications that are nearby," says Garcia. "We've only been looking at the modifications one at a time. Is that good enough? Or do we need to understand all the combinations together?" He compares looking at modifications singly as reading only every third or fifth word of a sentence. "Will we truly understand the meaning of the sentence this way?"

Many combinations are possible; in H3, for example, seven known sites of modification exist just within the first 20 amino acids. The theoretical number of combinations is in the tens of millions, but studies capable of detecting millions of modifications seem to find only a few thousand. Such observations indicate that the combinations occur in carefully coordinated patterns full of biological meaning, says Garcia.

However, it is hard to detect combinations in typical bottom-up experiments that start by shredding proteins into peptides with the enzyme trypsin. "If you just digest, then you have a gnarly mixture; you can't assign the combinations with confidence," says Kelleher, who has pioneered top-down MS, a method that introduces intact proteins into the instrument. Such techniques are less sensitive than bottom-up techniques and require two or three times as much sample, but researchers such as Kelleher are betting that understanding combinations can be as enlightening as finding the rarest modifications.

Some researchers, such as Garcia, are restricting studies to histone tails, thereby simplifying analysis to a few dozen or so amino acids (rather than 100 or more). This middle-down approach boosts sensitivity and throughput, and focuses



Neil Kelleher and former postdoc Ben Garcia pick apart combinations of histone marks.

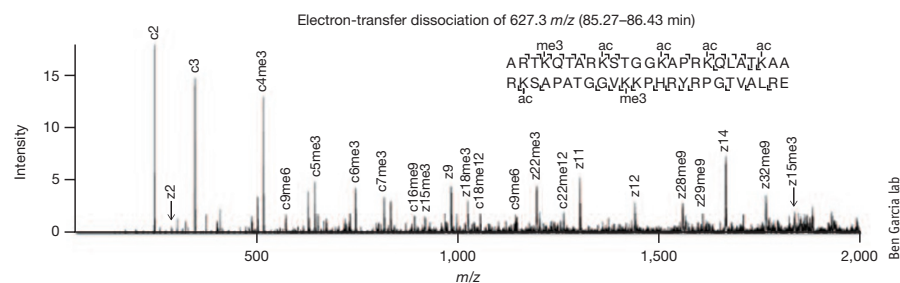
on the most highly decorated parts of the protein. Still, middle-down approaches risk overlooking important marks: intriguing modifications have already been found on histone cores.

Working with intact histones or even histone tails requires specialized fragmentation and informatics strategies. The Kelleher lab has built software to handle all possible combinations of disparate modifications. Versions of the software, called ProSight, are available in various forms from the Kelleher lab and the company Thermo Fisher Scientific.

To learn what these modifications mean, researchers must be able to compare their appearance over time and in different populations of cells. Labs are starting to compare pluripotent stem cells and differentiated cells as well as cancer and normal cells, the better to understand how gene expression is governed in those cell types.

The experiments that can be performed are infinite in number. The human genome is roughly the same in every human cell, but histone modifications are dynamic and can be very different across diverse cell types. To understand the range of histone modifications, more cell types need to be studied under more conditions. But for researchers to even be able to study a reasonable spectrum of these, techniques and machines must become more efficient and sensitive, says Kelleher. "There's a whole language there, and we have trouble even finding the words."

1. Strahl, B.D. & Allis, C.D. *Nature* **403**, 41–45 (2000).
2. Britton, L.M., Gonzales-Cope, M., Zee, B.M. & Garcia, B.A. *Expert Rev. Proteomics* **8**, 631–643 (2011).
3. Arrowsmith, C.H. et al. *Nat. Rev. Drug Discov.* **11**, 384–400 (2012).
4. Vermeulen, M. et al. *Cell* **142**, 967–980 (2010).
5. Bartke, T. et al. *Cell* **143**, 470–484 (2010).
6. Tan, M. et al. *Cell* **146**, 1016–1028 (2011).
7. Sakabe, K., Wang, Z. & Hart, G. *Proc. Natl. Acad. Sci. USA* **107**, 19915–19920 (2010).
8. Wang, Z. et al. *Mol. Cell. Proteomics* **9**, 153–160 (2010).



Mass spectra of a highly modified histone tail.

Monya Baker is technology editor for *Nature* and *Nature Methods* (m.baker@us.nature.com).