Published in final edited form as:

J Mol Biol. 2014 October 9; 426(20): 3350–3362. doi:10.1016/j.jmb.2014.04.024.

Emerging technologies to map the protein methylome

Scott M. Carlson^{a,*} and Or Gozani^a

^aDepartment of Biology, Stanford University, 371 Serra St, Stanford, CA, 94305, USA

Abstract

Protein methylation plays an integral role in cellular signaling, most notably by modulating proteins bound at chromatin, and increasingly through regulation of non-histone proteins. One central challenge in understanding how methylation acts in signaling is identifying and measuring protein methylation. This includes locus-specific modification of histones, on individual non-histone proteins, and globally across the proteome. Protein methylation has traditionally been studied using candidate approaches such as methylation-specific antibodies, mapping of post-translational modifications by mass spectrometry, and radioactive labeling to characterize methylation on target proteins. Recent developments have provided new approaches to identify methylated proteins, measure methylation levels, identify substrates of methyltransferase enzymes, and match methylated proteins to methyl-specific reader domains. Methyl-binding protein domains and improved antibodies with broad specificity for methylated proteins are being used to characterize the “protein methylome”. They also have the potential to be used in high-throughput assays for inhibitor screens and drug development. These tools are often coupled to improvements in mass spectrometry to quickly identify methylated residues, and to protein microarrays, where they can be used to screen for methylated proteins. Finally, new chemical biology strategies are being used to probe the function of methyltransferases, demethylases and methyl-binding “reader” domains. These tools are creating a “system-level” understanding of protein methylation, and integrating protein methylation into broader signaling processes.

Keywords

Lysine methylation; arginine methylation; methyltransferase; signaling; proteomics

© 2014 Elsevier Ltd. All rights reserved.

*Corresponding author To whom correspondence should be addressed: scottmc@stanford.edu, Phone: (617) 301-0242, Fax: (650) 725-8309.

Competing Interests

O.G. is a cofounder of EpiCypher, Inc. Stanford University has submitted a provisional US patent describing the use of the 3xMBT domain as an affinity reagent.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Introduction

Post-translational methylation of arginine and lysine on histone proteins has well-established functions in epigenetics and chromatin regulation, and methylation of non-histone proteins has recently emerged as an important actor in broader biological regulation (Figure 1, discussed in several recent reviews^{1; 2; 3}). Arginine methylation has many critical functions such as in signal transduction⁴, regulation of mRNA splicing and translation⁵, and physiological processes such as glucose homeostasis⁶ (reviewed in ^{7; 8; 9}). Non-histone lysine methylation regulates pathways including tumor suppression by p53 and signaling by NFκB^{10; 11} (reviewed in ¹²). Until recently most sites of protein methylation had been identified through candidate approaches, arising through detailed investigation of individual proteins.

Development of high-throughput approaches to identify and characterize methylated proteins has been ongoing for the past decade. In the last two years many of these techniques have reached fruition, revealing thousands of sites of arginine methylation and hundreds of sites of lysine methylation across the human proteome. These discoveries raise important questions about the function and regulation of protein methylation. For example, which of the at least nine known protein arginine methyltransferases (PRMTs) in the human proteome are responsible for all these methylation events, or do additional enzymes remain to be discovered³, and which lysine methyltransferases (KMTs) act to regulate non-histone methylation^{13; 14}? The SPOUT family of RNA methyltransferases has recently been shown to methylate arginine in yeast, perhaps this or other families are also acting in mammalian systems¹⁵. How is methylation regulated during dynamic biological processes, and how does it interact with other modes of regulation such as kinase signaling, transcription factor activity and mRNA splicing? Addressing these questions will require a combination of traditional biological approaches along with new systems-level tools to investigate methylation across the entire proteome.

In this review we will focus on emerging “system-level” tools to investigate protein methylation. We discuss approaches for proteome-wide identification of methylated proteins using target enrichment and tandem mass spectrometry (LC-MS/MS), affinity and high-throughput screening to identify PRMT and KMT substrates, and chemical biology approaches to investigate methylation signaling and discover PRMT and KMT substrates. These technologies have already enabled the first steps toward integrating methylation with broader signaling processes. We expect that future developments will make it possible to track methylation signaling dynamics directly, and to examine how methylation contributes to signaling that regulates development and disease.

Introduction to methods for proteomic analysis of arginine and lysine methylation

Most strategies for proteomic analysis of post-translational modifications begin by enriching modified proteins from cell lysate, or by digesting the lysate into short peptides and then enriching modified peptides. Such techniques are well-established for phosphorylation, where selective resins are used to enrich phosphorylated peptides from complex mixtures¹⁶.

Other modifications, notably phosphorylated tyrosine, lysine acetylation, and ubiquitinylation have been targeted effectively with “pan-specific” antibodies that bind modified peptides with little or no dependence on the local amino acid sequence^{17; 18; 19}. Once modified peptides are enriched they can be identified by LC-MS/MS²⁰ (Figure 2). It has become routine to incorporate isotopic labels or other quantitative approaches to compare modifications across multiple biological conditions²¹.

Similar techniques were first used to study arginine methylation more than a decade ago but only since 2012 have they improved to the point that analysis of the arginine methylation proteome could begin to become routine^{22; 23}. Arginine methylation is especially abundant, occurring on close to 1% of arginine residues in some types of mammalian cells²⁴. Antibodies recognizing methylated arginine in the context of an Arg-Gly motif have been very effective for enriching modified proteins and peptides^{22; 25; 26}, and recent developments have extended this approach to thousands of arginine methylation sites²³. In contrast, it has proven difficult to develop antibodies capable of enriching lysine methylation from crude protein extracts or digested peptides²⁶. Only in the past year have several groups reported successful large-scale identification of lysine methylation using antibodies, while other groups have developed alternative approaches based on promiscuous methyl-lysine binding protein domains^{23; 27; 28; 29; 30} (Table 1 summarizes the proteomic studies of protein methylation discussed here). These techniques will make it possible to address system-level questions about the regulation and function of arginine and lysine methylation.

Global approaches for proteome-wide analysis of arginine methylation

Foundations of arginine methylation proteomics using broad-specificity antibodies

The group of Stéphane Richard conducted early work to identify protein complexes containing symmetric or asymmetric di-methylated arginine (SDMA and ADMA) using broadly-specific antibodies in 2003²². They used antibodies recognizing ADMA and SDMA embedded in Arg-Gly sequences frequently methylated by PRMT1 and PRMT5 (which generate ADMA and SDMA respectively). This work identified over 200 proteins enriched by immunoprecipitation using antibodies against ADMA or SDMA but it did not directly identify any of their methylated residues. Unlike approaches that enrich modified peptides from digested protein, enriching modified proteins requires that modified residues be identified in a separate experiment or by high-coverage MS/MS of the precipitated proteins. This work suggested that arginine methylation is a widespread modification and formed a groundwork for proteomic identification of specific modified residues. Similar approaches have identified additional methylated proteins bound by the same antibodies, such as the PRMT1 substrate MRE11³¹.

Building on this idea, Ong and Mann conducted protein immunoprecipitation (IP) with an antibody recognizing methyl-arginine to enrich methylated proteins from lysate of HeLa S3 cells²⁶. They used LC-MS/MS to identify and 57 methylated arginine residues containing various methylation states of arginine across 31 proteins. Importantly, this work pioneered the use of *in vivo* heavy isotopic labeling to generate a distinctive mass shift for methylated residues. The authors prepared cells in media containing methionine labeled with deuterium and carbon-13 at the side-chain methyl group (¹³CD₃). This methionine is converted by cells

into S-adenosyl methionine (SAM) with isotopic label at the sulfonium methyl group, which is then transferred to proteins during the methylation reaction (Figure 3). Changing the mass shift for methylation from 14 Da to 18 Da increases confidence in identification of methylated residues because it differentiates legitimate methylation sites from artifacts arising by chemical methylation, such as conversion of acidic residues to methyl esters due to methanol used during processing³². Ong and Mann also attempted to enrich methylated lysine, but found identified peptides containing histone H3 lysine 27 and histone H4 lysine 20 from methyl-lysine IP. They noted that only a small fraction of proteins recovered by IP for methyl-lysine seemed to be were methylated and that they actually seemed to recover some amount of methylated arginine, indicating that available antibodies were limited in their selectivity.

Subsequent work focused on biochemical techniques to separate methylated proteins or peptides, optimizing mass spectrometry methods, and improving antibodies and techniques for IP. Li and colleagues used 2-dimensional gel electrophoresis (2D-PAGE) and Western blotting for methyl-arginine to select spots for identification by LC-MS/MS³³. This approach identified a small number of methylated proteins but the authors noted that many spots were identified as abundant proteins that are likely to co-migrate with proteins producing methyl-arginine signals in the Western blot. A similar strategy was used by Lin and colleagues to identify changes in protein methylation in LA29 rat fibroblasts expressing a temperature-sensitive form of the v-Src oncogene³⁴. They used *in vitro* methylation of cell lysates from cells grown at different temperatures to identify differential methylation, and 2D-PAGE with LC-MS/MS to identify likely methylated targets. As with earlier work using 2D-PAGE, this work identified a small number of likely methylated proteins. These were all very abundant proteins with roles in translation, metabolism or cellular structure. Further improvements in antibody specificity and mass spectrometry techniques would be needed to achieve truly proteome-wide analysis of arginine methylation.

Global proteomic analysis of arginine methylation using antibodies and high-throughput LC-MS/MS

Several studies published since 2013 have improved antibodies and mass spectrometry techniques to extend proteomic analysis of arginine methylation to thousands of methylation sites. Read and colleagues used protein IP with several commercial antibodies to identify 167 arginine methylation sites in mitochondria of the parasite *Trypanosoma brucei*³⁵. They separated material from each IP into 25 fractions using strong cation exchange (SCX). They increased the number of peptides identified by employing multiple strategies to fragment each peptide in the MS/MS analysis. In addition to the standard collision induced dissociation (CID) they fragmented peptides using electron transfer dissociation (ETD) because of its utility in identifying post-translational modifications (reviewed by Afjehi-Sadat *et al.*²⁰). They also increased coverage by performing additional analyses using the protease Glu-C instead of trypsin. Glu-C cleaves C-terminal to glutamic and aspartic acids depending on the buffer conditions, producing peptides that provide information complementary to those produced by trypsin. In subsequent work the same group analyzed arginine methylation across the entire *T. brucei* proteome³⁶, reporting 1,332 sites of arginine methylation. This work shows that the techniques for enrichment and analysis of arginine

methylation have reached the point where deep coverage can be achieved. It will be exciting to see how directly these data apply to the human arginine methylome.

Comb and colleagues produced rabbit monoclonal antibodies using peptide libraries containing mono-methyl arginine (MMA) at the central position, MMA in the context of an RGG motif, or asymmetric di-methyl arginine at four positions²³. These antibodies were used for peptide IP to identify sites of MMA and ADMA in HCT116 lung adenocarcinoma cells, mouse brain tissue and mouse embryos. In total they identified over 1,000 sites of MMA and ADMA. They also used label-free quantitation to compare the abundance of arginine methylation sites between mouse brain and embryo samples. In contrast to earlier studies this analysis did not require additional fractionation or multi-dimensional chromatography. Nielsen and coworkers used the same antibodies for peptide IP of MMA from HEK 293T and U2OS cell³⁷. By combining peptide IP with fractionation using SCX they identified 1,027 sites of MMA. They also used stable isotopic labeling by amino acids in cell culture (SILAC)³⁸ to examine changes in arginine methylation following transcriptional arrest by actinomycin D, identifying dynamic arginine methylation on many proteins involved in transcriptional regulation. To our knowledge, these are the only examples to-date of global quantitative analysis of arginine methylation. We expect that such experiments will make it possible to examine substrates of arginine methyltransferases, track dynamic methylation during signaling transduction, and measure how methylation changes through biological processes such as cellular differentiation.

Biochemical techniques for analyzing the methyl proteome

In addition to analyzing protein methylation using antibodies, at least two recent studies have used biochemical techniques to enrich methylated peptides for identification by LC-MS/MS. Acuto and colleagues identified 41 arginine methylation sites using peptide IP, and compared these sites with several strategies to enrich arginine methylation based on its chemical properties²⁵. They used SCX to enrich tryptic peptides containing arginine methylation based on the fact that cleavage of methyl-arginine by trypsin is inefficient. This causes methylated peptides to contain both an internal arginine and a C-terminal arginine or lysine, and thus to carry a charge of at least +3 at pH less than 3. In comparison, about 70% of non-methyl tryptic peptides have +2 charge at the same pH. Using SCX they identified 39 methylated residues, most of which were not previously reported. The authors next observed that most methylated peptides observed so far are basic, with isoelectric point (pI) of close to 11. They used isoelectric focusing to enrich and identify 66 methylated residues, again most of which were novel. They finally observed that most methylated peptides were highly hydrophilic and reasoned that hydrophilic interaction liquid chromatography (HILIC) would be able to separate methylated from non-methyl peptides. After fractionating cell lysate using HILIC the authors identified 215 methylation sites, most of which had not been identified using the other techniques.

A similar study was conducted by Bonaldi and colleagues³⁹. They used protein IP with five antibodies recognizing various methylation states of lysine and six recognizing methylation states of arginine. Protein IP was followed by separation into fractions using SDS-PAGE and analysis by LC-MS/MS. They compared methylation sites identified using this approach

to sites identified by direct separation using SDS-PAGE and LC-MS/MS, and by isoelectric focusing of peptides from a tryptic digest. Importantly, they incorporated isotopic labeling using the method of Ong and Mann discussed above. They authors combined light and heavy cell lysates in a 1:1 mixture so that methylated peptides could be identified by the distinctive mass shift between their two isotopic forms. This work identified in total 74 sites of lysine methylation and 323 sites of arginine methylation. In contrast to earlier work by Acuto and colleagues, this study found that protein IP and fractionation by SDS-PAGE was the most effective strategy to identify methylated residues (they used it to identify 54 and 254 methylation sites for lysine and arginine respectively). This difference is likely due to the use of 11 distinct antibodies and deep protein sequencing by analyzing many molecular weight fractions separated by SDS-PAGE.

A key observation of these studies is that many arginine methylation sites are not contained within the Arg-Gly motif enriched in earlier work using peptide IP. This result suggested that better pan-specific antibodies could greatly increase coverage of the methyl-arginine proteome. It is also notable that there is little overlap among the methylation sites identified by the different enrichment strategies. It may be necessary to use multiple physical and chemical approaches to achieve thorough coverage of arginine methylation. Both of these studies separated their samples into many fractions for analysis using independent runs of the mass spectrometer. This strategy can greatly increase peptide coverage but also requires a large input of time and resources. Even with improvements in automated sample handling, high-throughput LC-MS/MS and better bioinformatics tools, it is unclear whether such approaches can be used routinely to characterize the methyl proteome across biological contexts, or to examine how dynamic methylation contributes to cellular behavior.

Proteomic strategies for non-histone lysine methylation

Over the past decade many non-histone proteins have been found to be modified by mono-, di- or tri-methylation of lysine, and these modifications have been shown to participate in a wide range of signaling pathways and regulatory processes^{1; 12}. Unlike arginine methylation, early efforts to develop broad-specificity antibodies recognizing the methylation states of lysine were largely unsuccessful²⁶. This has made it difficult to establish a system-level understanding of lysine methylation; instead requiring methylated proteins to be identify and investigated individually. It has only been since 2013 that approaches have become available to probe lysine methylation across the proteome using antibodies and other methyl-lysine binding proteins^{23; 27; 28; 29}. We next discuss early efforts toward the methyl-lysine proteome, and then a series of recent breakthroughs in antibodies and protein affinity tools that have made it possible to examine lysine methylation at a proteomic scale.

Lysine methylation proteomics using antibodies

An early effort to enrich lysine methylation for proteomic analysis was described by Ong and Mann along with their work discussed earlier on methylated arginine³. Using methyl-lysine antibodies available at the time they were only able to identify one methylation site each on the histones H3 and H4. Further work was conducted over the next eight years to develop antibodies against methylation states of lysine but they all suffered from some

combination of low affinity, poor selectivity for methylation, and specificity for particular peptide sequences (for example, Iwabata *et al.*⁴ describe using an methyl-lysine antibody with 2D-PAGE and Western blotting and Levy *et al.*⁵ used peptide arrays to evaluate commercially available methyl-lysine antibodies).

In 2013 Garcia and colleagues developed a series of polyclonal antibodies against mono-, di- and tri-methylated lysine²⁷. They used these antibodies for peptide IP, followed by separation into up to 11 fractions using SCX to increase coverage by LC-MS/MS. In total they reported 323 sites of mono-methylation, 127 sites of di-methylation and 102 sites of tri-methylation across a total of 413 proteins. To our knowledge, this is the most extensive single list of lysine methylation produced so far. Similar to much of the work on arginine methylation, the authors found it necessary to separate material from the peptide IP into many fractions for analysis by LC-MS/MS, suggesting that methylated peptides make up only a small part of the total. It is likely that the extensive fractionation was required because of limited selectivity of the antibodies for methylated relative to non-methylated peptides.

Independently, Comb and colleagues immunized rabbits with peptide libraries containing mono-, di- or tri-methylated lysine to generate pan-specific methyl-lysine antibodies²³. Using these antibodies for peptide IP identified 132 sites of mono-methylation, 35 sites of di-methylation and 31 sites of tri-methylation from HCT116 lung adenocarcinoma cells (representing 111, 19 and 23 distinct proteins respectively). Unlike the Garcia lab, the authors did not perform additional fractionation following peptide IP. This difference may account for the larger number of peptides identified in the earlier experiment. The authors did not find any specific sequence motif enriched by their methyl-lysine antibodies, suggesting that sequence specificity was not a limiting factor. Both this work and the study by Garcia and colleagues identified many new methylated residues but also did not identify many methylated lysine residues that have been discovered by other approaches. It may be that methylated species vary greatly between biological contexts, or that these antibodies have limited ability to enrich low-abundance peptides from tryptic digest of the entire proteome.

These two studies have shown that it is possible to produce broad-specificity antibodies against methylated lysine, and that these antibodies can be used to identify hundreds of distinct methylation sites. There is little overlap in the methylation sites identified by these two studies, perhaps indicating that the antibodies have only partly overlapping specificity. It may also be that lysine methylation changes dramatically between the cell types used in these studies. Future work will help determine how thoroughly these experiments probe the methyl-lysine proteome. It will also be important to establish whether the depth of coverage, reproducibility and ease-of-use will allow these techniques to be used for system-level investigation into cellular signaling through lysine methylation, and how dynamic lysine methylation participates in biological regulation.

Proteomics using methyl-binding domains

As an alternative to antibodies, two groups have recently used natural methyl-lysine binding domains to isolate and identify proteins modified by lysine methylation. This approach is

based on the observation that some methyl-lysine binding domains, such as the triple MBT domains (3xMBT) of the protein L3MBTL1, recognize methylated lysine through a hydrophobic binding pocket that forms limited contacts with surrounding residues^{40; 41}. Taken out of their natural context, such domains can be used in much the same way as pan-specific antibodies²⁸.

We established that 3xMBT will bind to almost any protein or peptide containing mono- or di-methylated lysine (Figure 4a). We then introduced a structure-based negative control by using a point mutant in the 3xMBT hydrophobic pocket to disrupt binding to methylated lysine⁴⁰, and used SILAC labeling to quantitatively compare proteins captured by native 3xMBT and the inactive point mutant (Figure 4b). This work identified several hundred proteins that are consistently enriched by the MBT domain from nuclear extract of HEK 293T cells, but only directly identified methylated lysine on 26 of these. Importantly, only five of these methylation sites were previously reported, strongly suggesting that the majority of methylated residues remain to be identified. Many of the methylated proteins identified by peptide IP are strongly enriched in this experiment but there is little overlap among the methylation sites identified directly. Advantages of this approach are that 3xMBT can be easily produced in *e.coli*, that it has well-established specificity for a broad range of peptide sequences, that protein methylation can be measured even when methylated residues are embedded in sequences difficult to identify by LC-MS/MS, and that the point-mutant provides a rigorous negative control. The major disadvantage is that the domain only efficiently enriches methylated proteins, so additional experiments are often required to identify specific methylated residues.

In the same work, we used SILAC labeling to examine proteome-wide changes in lysine methylation after inhibiting the lysine methyltransferases G9a and GLP (also called EHMT2 and EHMT1)⁴². In addition to being major methyltransferases producing di-methylation of histone H3 lysine 9^{43; 44}, these enzymes methylate numerous proteins *in vitro* and at least a fraction of these in their biological context^{45; 46}. We observed a strong decrease in methylation of reported G9a substrates WIZ and ACIN1, as well as several potential novel targets, but we did not observe a decrease in bulk methylation of H3, most likely because the effect was buffered by methylation at other residues such as lysine 4, 27 and 36. The ability to monitor lysine methylation across the proteome in a quantitative manner opens the possibility of more sophisticated investigation into methyltransferase and demethylase substrates, as well as providing a route to detailed understanding of how methylation participates in dynamic signaling processes. However, these applications will greatly benefit from residue-specific information instead of total protein methylation.

Independently, Li and colleagues used a proteomic strategy to investigate methyl-specific binding targets of the chromodomain from the heterochromatin protein HP1 β ³⁰. They used immobilized HP1 β chromodomain as bait to capture methylated proteins from cell lysate, identifying a total *in vitro* interactome of 109 proteins in HEK 293T cells. They used peptide arrays to characterize the sequence specificity of HP1 β chromodomain, and then applied a bioinformatics approach to predict likely methylated lysine residues from the overall interactome. They tested specific predicted methylation sites using a targeted approach and confirmed methylation on 40 residues across 29 proteins, representing a 40% hit rate for

predicted sites of methylation. It is notable that many proteins enriched by the 3xMBT domain are also enriched by HP1 β chromodomain, but there is no overlap among the methylation sites directly identified. This may reflect the use of trypsin by our group to digest bound proteins, while the Li group used proteases Arg-C, Glu-C, chymotrypsin and elastase. It is likely that multiple digestion and mass spectrometry strategies will continue to expand coverage of the lysine methylome.

An integrative perspective on the methyl proteome

Emerging proteomic strategies have identified thousands of methylated proteins, revealing methylation as a common post-translational modification. Understanding the biological functions of protein methylation will require better tools to analyze global methylation and methylation dynamics, as well as deeper investigation using traditional approaches of molecular biology. From a reductionist perspective there are two key questions that need to be systematically addressed:

1. Which methyltransferase and demethylase enzymes regulate each methylation event?
2. How do specific methylation events regulate protein function and/or interactions with proteins, nucleic acids or other bio-molecules?

The rest of this review focuses on emerging proteome-wide and high-throughput strategies to address the first of these questions. We will examine strategies to identify non-histone targets of methyltransferase and demethylase enzymes. It will also be necessary to place these questions in a broader context. It will be important to understand how protein methylation is linked to upstream signaling pathways and regulatory processes, and to discover how protein methylation acts to regulate downstream cellular decisions and biological processes. Addressing these questions will call for further development of system-level approaches, as well as experiments probing protein methylation along with other types of signal transduction such as phosphorylation.

Proteome-wide approaches for identifying methyltransferase and demethylase substrates

Identifying proteins directly modified by any signaling enzyme is a challenging task. Fundamentally, it is difficult to know whether all the important substrates of an enzyme are expressed in any particular biological context, and whether an enzyme/substrate interaction might depend on other factors such as interactions with scaffold proteins or prior post-translational modifications. In practice it is necessary to show that a candidate substrate protein is directly modified *in vitro*, and that the same modification site is regulated by the putative enzyme in cells⁴⁷.

Proteomic tools to identify PRMT substrates have had a limited impact because there are only nine known human PRMTs³; it is easier and more reliable to test candidate substrate by screening against the entire set of PRMTs. In contrast, there are at least 55 human proteins containing the Su(var)3-9, enhancer-of-zeste, trithorax (SET) domain, which is responsible for most known lysine methylation¹³. Many of these predicted KMTs have no known

substrates, raising the possibilities that they lack enzymatic activity or that their substrates simply have not been discovered. In addition, some members of the seven- β strand family of methyltransferases are able to methylate lysine⁴⁸, and more such enzymes are still being discovered^{49; 50}. Here we review proteome-wide strategies that have been used to identify substrates of the arginine and lysine protein methyltransferases.

Microarrays

Protein microarrays allow enzymes to be screened against thousands of potential substrate proteins independent of their expression in a biological system^{51; 52; 53; 54}. Modified proteins are detected by radiolabeling⁵⁵, broad-specificity antibodies, or by binding to methyl-protein recognition domains²⁸. The major weaknesses of microarray approaches are that there is no guarantee that every protein is correctly folded, proteins are presented outside their natural biological context of protein interactions and other biomolecules, and the process of immobilizing them on a surface may block recognition of their substrate site(s)⁵⁶. It is necessary to identify methylated residues individually using targeted follow-up experiments to ensure that *in vitro* substrates are bona fide substrates in their biological context⁴⁷.

In the context of protein methylation, high-density microarrays were first used by Bedford and colleagues to identify substrates of the arginine methyltransferases PRMT1 and CARM1 (also called PRMT4)⁵⁵. This work identified five substrates for PRMT1 and two for CARM1. They confirmed the most prominent substrates by verifying that the original bacterial stocks from the array manufacturer expressed the correct protein, and that each substrate was methylated in an independent *in vitro* labeling reaction. They then screened protein fragments and short peptides to identify the precise sites of methylation by CARM1 on Poly(A)-binding protein 1 (PABP1). It is now routine to verify specific modified residues using mass spectrometry²⁰ but site-directed mutagenesis provides strong evidence that all methylation sites on a substrate protein have been identified. They finally used labeled cells in culture with ³H-methionine in the presence of inhibitors of protein synthesis to show that PABP1 is methylated *in vivo*⁵⁷ and they used co-immunoprecipitation (co-IP) to show that it interacts with CARM1.

Protein microarrays have recently emerged as a promising tool for identifying methyltransferase substrates. Levy *et al.* used the ProtoArray system (Life Technologies), consisting of ~9,500 human proteins expressed as a fusion with glutathione-S-transferase (GST) and immobilized on glass slides⁵⁶. They treated arrays with SET domains from the enzymes SETD6 and SETD7, or with GST as a negative control. Methylated proteins were detected by radiolabeling with tritium, or by binding to fluorescent broad-specificity antibodies against methylated lysine. The authors identified 26 high-confidence *in vitro* substrates of SETD6 by both detection methods, validated six out of six tested using independent *in vitro* assays, and tested two, PLK1 and PAK4, in human cell culture. It remains to be determined how many of the lower confidence hits, or hits identified by a single detection method, are real *in vitro* substrates, and how many are relevant in a physiological context.

The use of carefully validated pan-specific antibodies was critical in this work. Several of the commercial antibodies tested exhibited poor selectivity on a panel of methylated peptides, non-methylated peptides, and peptides carrying other modification. Recent work has shown that methyl-binding domains such as the L3MBTL1 3xMBT domain can be used in place of antibodies to probe lysine methylation on protein microarrays²⁸.

Interaction screens for candidate PRMT and KMT substrates

Many signaling molecules form interactions with their substrates that are stable enough to capture by co-IP^{58; 59}. Co-IP is a well-established approach for identifying kinase substrates, and has recently been applied to both arginine and lysine methyltransferases^{9; 50}. A major advantage of this approach is that it captures enzyme/substrate interactions that may require scaffold proteins or interactions through other biomolecules such as DNA or RNA. While this can be a useful approach to generate a relatively small pool of candidate substrates, its utility is limited because many enzyme/substrate interactions are transient and difficult to capture, and because many enzymes interact with large protein complexes that are not direct substrates. Overall, proteomic identification of interacting proteins is often an effective way to begin identifying substrates for a methyltransferase.

Some examples of PRMTs that have been co-purified with their substrates include ribosomal protein S2 methylated by PRMT3⁶⁰, DNA polymerase β by PRMT6⁶¹, of estrogen receptor α by PRMT1^{62; 63}, of ribosomal protein S10 by PRMT5⁶⁴. The same approach has been used to identify the chaperone HSP90 as a direct substrate of SMYD2⁶⁵. In another recent example, Falnes and colleagues identified several novel KMTs from the seven- β strand methyltransferase family, and used tandem affinity purification to identify candidate substrates for these new enzymes. In particular they identified the molecular chaperones VCP and HSP70 as substrates of the KMTs METTL21A and METTL21D^{49; 50}. It remains to be determined whether these KMTs have additional substrates that do not form stable complexes.

Using chemical genetics to identify methyltransferase substrates

A common challenge facing proteomic approaches is differentiating methylation by an enzyme of interest from background methylation by endogenous enzymes. It is therefore desirable to develop strategies that specifically label substrates of an individual enzyme. Chemical genetics, often called “bump-hole”, has been widely used to identify substrates for kinase phosphorylation. In this approach, a mutation is introduced into the enzyme of interest that expands its pocket for binding adenosine triphosphate (ATP); this allows the enzyme to use modified co-factors carrying bulky unnatural substituents^{66; 67; 68; 69}. Replacing the terminal phosphate of ATP with a radiolabel or thiophosphate allows substrates of the enzyme to be uniquely tagged for detection or capture. This strategy was pioneered by the group of Kevan Shokat more than fifteen years ago, and it has recently been applied for proteome-wide identification of many kinase substrates^{59; 70; 71; 72; 73}.

In 2001 Gray and colleagues demonstrated the first example of this approach applied to protein methylation. They introduced a mutation in the yeast methyltransferase RMT1 so that it would be selectively inhibited using bulky analogs of S-adenosylhomocysteine (SAH,

Figure 5)⁷⁴, and would be able to utilize S-adenosyl methionine (SAM) substituted with a benzyl group at the adenosine N⁶ position. In Binda *et al.*, the KMT SETDB1 was shown to utilize an analog of SAM with propargyl-substituted sulfonium (Figure 5)⁷⁵. This causes the enzyme to transfer the propargyl group in place of normal methylation. Click chemistry could then be used to conjugate alkyne-labeled substrates to a detection moiety such as a Flag epitope tag or a fluorophore^{76; 77}. This SAM analog was unstable and would not work with many other KMTs.

This issue was circumvented by Luo and colleagues, who found that mutations to the SAM binding pocket of arginine or lysine methyltransferases could allow them to accept SAM analogs with larger substituents at the sulfonium position (Figure 5)^{78; 79; 80}. This allows enzyme substrates to be labeled in cell lysate using *in vitro* reactions with only minimal background from endogenous protein methyltransferases. Using this approach they engineered the KMTs G9a and GLP to accept bulky SAM analogs (Figure 5)⁸¹. They expressed these enzymes in HEK 293T cells and used lysates from these cells for *in vitro* labeling with an azide-containing SAM analog. Labeled substrates were conjugated to a strained alkyne linked to biotin that allowed substrates to be enriched using immobilized streptavidin and identified by LC-MS/MS. This work identified 82 candidate substrates of G9a and 64 of GLP. It will be exciting to see how many of these substrates have important biological functions. The authors note that their substrates include only a subset of previously reported G9a and GLP targets, perhaps because some targets are present at low level in HEK 293T cells or because *in vitro* labeling does not recapitulate all the molecular interactions present in the cellular context. The Luo group has used the same strategy to identify novel substrates of the arginine methyltransferase PRMT3⁸². This is a promising strategy that can be applied broadly to both PRMT and KMT enzymes. As with all proteomic methods, these studies will need to be paired with direct investigation to elucidate the physiologic importance of individual substrates.

In an exciting extension of this approach Luo and colleagues have engineered the human methionine adenosyltransferase (MAT, the enzyme responsible for synthesizing SAM from methionine) so that it can accept bulky methionine analogs introduced during cell culture. These analogs are processed into the corresponding analog of SAM in living cells. They showed that when appropriately engineered G9a or GLP are expressed in cells along with engineered MAT the bulky alkyl group is transferred to histone H3. To track genome-wide activity of G9a and GLP at chromatin the authors conjugated modified H3 to a cleavage biotin probe and chromatin capture followed by high-throughput sequencing. This allowed them to compare and contrast G9a and GLP activity across the genome in a way that was previously impossible. Such an approach may be a powerful tool to label and identify non-histone methyltransferases in their natural biological context.

Conclusions

Emerging proteomic tools have shown that post-translational methylation of arginine and lysine is common in species ranging from yeast to human. Non-histone methylation is likely involved in a vast range of signaling processes. Only in the last year have system-level tools to probe and measure protein methylation entered widespread use. Pan-specific antibodies

have enabled high-throughput proteomic identification of arginine methylation, while tools to study lysine methylation are rapidly improving, such as broad-specificity antibodies and pan-specific methyl-lysine binding domains. Approaches adapted from work on other signaling processes, especially phosphorylation, has provided tools to map methylation signaling networks using proteomics, protein microarrays and chemical biology. All these tools have allowed us to establish a new perspective on the broad role of protein methylation in biological regulation. They will be critical as we seek to develop an integrated understanding of how methylation interacts with other signaling processes, contributes to physiologic homeostasis, and acts in development of disease.

Acknowledgments

S.M.C. was supported in part by postdoctoral fellowship 123711-PF-13-093-01-TBE from the American Cancer Society Illinois Division. This work was supported in part by NIH grant R01 GM079641 to O.G.

References

1. Zhang X, Wen H, Shi X. Lysine methylation: beyond histones. *Acta Biochim Biophys Sin (Shanghai)*. 2012; 44:14–27. [PubMed: 22194010]
2. Di Lorenzo A, Bedford MT. Histone arginine methylation. *FEBS Lett*. 2011; 585:2024–31. [PubMed: 21074527]
3. Yang Y, Bedford MT. Protein arginine methyltransferases and cancer. *Nat Rev Cancer*. 2013; 13:37–50. [PubMed: 23235912]
4. Mowen KA, Tang J, Zhu W, Schurter BT, Shuai K, Herschman HR, David M. Arginine methylation of STAT1 modulates IFN α /beta-induced transcription. *Cell*. 2001; 104:731–41. [PubMed: 11257227]
5. Boisvert FM, Cote J, Boulanger MC, Cleroux P, Bachand F, Autexier C, Richard S. Symmetrical dimethylarginine methylation is required for the localization of SMN in Cajal bodies and pre-mRNA splicing. *J Cell Biol*. 2002; 159:957–69. [PubMed: 12486110]
6. Iwasaki H, Yada T. Protein arginine methylation regulates insulin signaling in L6 skeletal muscle cells. *Biochem Biophys Res Commun*. 2007; 364:1015–21. [PubMed: 17971302]
7. Bedford MT, Clarke SG. Protein arginine methylation in mammals: who, what, and why. *Mol Cell*. 2009; 33:1–13. [PubMed: 19150423]
8. Bedford MT, Richard S. Arginine methylation an emerging regulator of protein function. *Mol Cell*. 2005; 18:263–72. [PubMed: 15866169]
9. Han HS, Jung CY, Yoon YS, Choi S, Choi D, Kang G, Park KG, Kim ST, Koo SH. Arginine Methylation of CRTC2 Is Critical in the Transcriptional Control of Hepatic Glucose Metabolism. *Sci Signal*. 2014; 7:ra19. [PubMed: 24570487]
10. Shi X, Kachirskaja I, Yamaguchi H, West LE, Wen H, Wang EW, Dutta S, Appella E, Gozani O. Modulation of p53 function by SET8-mediated methylation at lysine 382. *Mol Cell*. 2007; 27:636–46. [PubMed: 17707234]
11. Levy D, Kuo AJ, Chang Y, Schaefer U, Kitson C, Cheung P, Espejo A, Zee BM, Liu CL, Tangsombatvisit S, Tennen RI, Kuo AY, Tanjing S, Cheung R, Chua KF, Utz PJ, Shi X, Prinjha RK, Lee K, Garcia BA, Bedford MT, Tarakhovsky A, Cheng X, Gozani O. Lysine methylation of the NF- κ B subunit RelA by SETD6 couples activity of the histone methyltransferase GLP at chromatin to tonic repression of NF- κ B signaling. *Nat Immunol*. 2011; 12:29–36. [PubMed: 21131967]
12. Huang J, Berger SL. The emerging field of dynamic lysine methylation of non-histone proteins. *Curr Opin Genet Dev*. 2008; 18:152–8. [PubMed: 18339539]
13. Petrossian TC, Clarke SG. Uncovering the human methyltransferasome. *Mol Cell Proteomics*. 2011; 10 M110.000976.

14. Petrossian TC, Clarke SG. Multiple Motif Scanning to identify methyltransferases from the yeast proteome. *Mol Cell Proteomics*. 2009; 8:1516–26. [PubMed: 19351663]
15. Young BD, Weiss DI, Zurita-Lopez CI, Webb KJ, Clarke SG, McBride AE. Identification of methylated proteins in the yeast small ribosomal subunit: a role for SPOUT methyltransferases in protein arginine methylation. *Biochemistry*. 2012; 51:5091–104. [PubMed: 22650761]
16. Ficarro SB, McClelland ML, Stukenberg PT, Burke DJ, Ross MM, Shabanowitz J, Hunt DF, White FM. Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat Biotechnol*. 2002; 20:301–5. [PubMed: 11875433]
17. Choudhary C, Kumar C, Gnad F, Nielsen M, Rehman M, Walther T, Olsen J, Mann M. Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science*. 2009; 325:834–40. [PubMed: 19608861]
18. Zhang Y, Wolf-Yadlin A, Ross P, Pappin D, Rush J, Lauffenburger D, White F. Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. *Mol Cell Proteomics*. 2005; 4:1240–50. [PubMed: 15951569]
19. Xu G, Paige J, Jaffrey S. Global analysis of lysine ubiquitination by ubiquitin remnant immunoaffinity profiling. *Nat Biotechnol*. 2010; 28:868–73. [PubMed: 20639865]
20. Afjehi-Sadat L, Garcia BA. Comprehending dynamic protein methylation with mass spectrometry. *Curr Opin Chem Biol*. 2013; 17:12–9. [PubMed: 23333572]
21. Altelaar AF, Munoz J, Heck AJ. Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat Rev Genet*. 2013; 14:35–48. [PubMed: 23207911]
22. Boisvert FM, Côté J, Boulanger MC, Richard S. A proteomic analysis of arginine-methylated protein complexes. *Mol Cell Proteomics*. 2003; 2:1319–30. [PubMed: 14534352]
23. Guo A, Gu H, Zhou J, Mulhern D, Wang Y, Lee KA, Yang V, Aguiar M, Kornhauser J, Jia X, Ren J, Beausoleil SA, Silva JC, Vemulapalli V, Bedford MT, Comb MJ. Immunoaffinity enrichment and mass spectrometry analysis of protein methylation. *Mol Cell Proteomics*. 2014; 13:372–87. [PubMed: 24129315]
24. Bulau P, Zakrzewicz D, Kitowska K, Wardega B, Kreuder J, Eickelberg O. Quantitative assessment of arginine methylation in free versus protein-incorporated amino acids in vitro and in vivo using protein hydrolysis and high-performance liquid chromatography. *Biotechniques*. 2006; 40:305–10. [PubMed: 16568819]
25. Uhlmann T, Geoghegan VL, Thomas B, Ridlova G, Trudgian DC, Acuto O. A method for large-scale identification of protein arginine methylation. *Mol Cell Proteomics*. 2012; 11:1489–99. [PubMed: 22865923]
26. Ong SE, Mittler G, Mann M. Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. *Nat Methods*. 2004; 1:119–26. [PubMed: 15782174]
27. Cao XJ, Arnaudo AM, Garcia BA. Large-scale global identification of protein lysine methylation in vivo. *Epigenetics*. 2013; 8
28. Moore KE, Carlson SM, Camp ND, Cheung P, James RG, Chua KF, Wolf-Yadlin A, Gozani O. A General Molecular Affinity Strategy for Global Detection and Proteomic Analysis of Lysine Methylation. *Molecular Cell*. 2013; 50:444–456. [PubMed: 23583077]
29. Carlson SM, Moore KE, Green EM, Martin GM, Gozani O. Proteome-wide enrichment of proteins modified by lysine methylation. *Nature Protocols*. 2014; 9:37–50.
30. Liu H, Galka M, Mori E, Liu X, Lin YF, Wei R, Pittock P, Voss C, Dhami G, Li X, Miyaji M, Lajoie G, Chen B, Li SS. A Method for Systematic Mapping of Protein Lysine Methylation Identifies Functions for HP1 β in DNA Damage Response. *Mol Cell*. 2013; 50:723–35. [PubMed: 23707759]
31. Boisvert FM, Déry U, Masson JY, Richard S. Arginine methylation of MRE11 by PRMT1 is required for DNA damage checkpoint control. *Genes Dev*. 2005; 19:671–6. [PubMed: 15741314]
32. Jung SY, Li Y, Wang Y, Chen Y, Zhao Y, Qin J. Complications in the assignment of 14 and 28 Da mass shift detected by mass spectrometry as in vivo methylation from endogenous proteins. *Anal Chem*. 2008; 80:1721–9. [PubMed: 18247584]

33. Hung CJ, Lee YJ, Chen DH, Li C. Proteomic analysis of methylarginine-containing proteins in HeLa cells by two-dimensional gel electrophoresis and immunoblotting with a methylarginine-specific antibody. *Protein J.* 2009; 28:139–47. [PubMed: 19365714]
34. Chiou YY, Fu SL, Lin WJ, Lin CH. Proteomics analysis of in vitro protein methylation during Src-induced transformation. *Electrophoresis.* 2012; 33:451–61. [PubMed: 22228245]
35. Fisk JC, Li J, Wang H, Aletta JM, Qu J, Read LK. Proteomic analysis reveals diverse classes of arginine methylproteins in mitochondria of trypanosomes. *Mol Cell Proteomics.* 2013; 12:302–11. [PubMed: 23152538]
36. Lott K, Li J, Fisk JC, Wang H, Aletta JM, Qu J, Read LK. Global proteomic analysis in trypanosomes reveals unique proteins and conserved cellular processes impacted by arginine methylation. *J Proteomics.* 2013; 91:210–25. [PubMed: 23872088]
37. Sylvestersen KB, Horn H, Jungmichel S, Jensen LJ, Nielsen ML. Proteomic analysis of arginine methylation sites in human cells reveals dynamic regulation during transcriptional arrest. *Mol Cell Proteomics.* 2014
38. Ong S, Blagoev B, Kratchmarova I, Kristensen D, Steen H, Pandey A, Mann M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics.* 2002; 1:376–86. [PubMed: 12118079]
39. Bremang M, Cuomo A, Agresta AM, Stugiewicz M, Spadotto V, Bonaldi T. Mass spectrometry-based identification and characterisation of lysine and arginine methylation in the human proteome. *Mol Biosyst.* 2013
40. Min J, Allali-Hassani A, Nady N, Qi C, Ouyang H, Liu Y, MacKenzie F, Vedadi M, Arrowsmith CH. L3MBTL1 recognition of mono- and dimethylated histones. *Nat Struct Mol Biol.* 2007; 14:1229–30. [PubMed: 18026117]
41. Nady N, Krichevsky L, Zhong N, Duan S, Tempel W, Amaya MF, Ravichandran M, Arrowsmith CH. Histone recognition by human malignant brain tumor domains. *J Mol Biol.* 2012; 423:702–18. [PubMed: 22954662]
42. Vedadi M, Barsyte-Lovejoy D, Liu F, Rival-Gervier S, Allali-Hassani A, Labrie V, Wigle TJ, Dimaggio PA, Wasney GA, Siarheyeva A, Dong A, Tempel W, Wang SC, Chen X, Chau I, Mangano TJ, Huang XP, Simpson CD, Pattenden SG, Norris JL, Kireev DB, Tripathy A, Edwards A, Roth BL, Janzen WP, Garcia BA, Petronis A, Ellis J, Brown PJ, Frye SV, Arrowsmith CH, Jin J. A chemical probe selectively inhibits G9a and GLP methyltransferase activity in cells. *Nat Chem Biol.* 2011; 7:566–74. [PubMed: 21743462]
43. Tachibana M, Sugimoto K, Fukushima T, Shinkai Y. Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. *J Biol Chem.* 2001; 276(2):5309–17.
44. Shinkai Y, Tachibana M. H3K9 methyltransferase G9a and the related molecule GLP. *Genes Dev.* 2011; 25:781–8. [PubMed: 21498567]
45. Rathert P, Dhayalan A, Murakami M, Zhang X, Tamas R, Jurkowska R, Komatsu Y, Shinkai Y, Cheng X, Jeltsch A. Protein lysine methyltransferase G9a acts on non-histone targets. *Nat Chem Biol.* 2008; 4:344–6. [PubMed: 18438403]
46. Sampath SC, Marazzi I, Yap KL, Krutchinsky AN, Mecklenbräuer I, Viale A, Rudensky E, Zhou MM, Chait BT, Tarakhovsky A. Methylation of a histone mimic within the histone methyltransferase G9a regulates protein complex assembly. *Mol Cell.* 2007; 27:596–608. [PubMed: 17707231]
47. Berwick D, Tavaré J. Identifying protein kinase substrates: hunting for the organ-grinder's monkeys. *Trends Biochem Sci.* 2004; 29:227–32. [PubMed: 15130558]
48. Ng HH, Feng Q, Wang H, Erdjument-Bromage H, Tempst P, Zhang Y, Struhl K. Lysine methylation within the globular domain of histone H3 by Dot1 is important for telomeric silencing and Sir protein association. *Genes Dev.* 2002; 16:1518–27. [PubMed: 12080090]
49. Kernstock S, Davydova E, Jakobsson M, Moen A, Pettersen S, Mælandsmo GM, Egge-Jacobsen W, Falnes P. Lysine methylation of VCP by a member of a novel human protein methyltransferase family. *Nat Commun.* 2012; 3:1038. [PubMed: 22948820]

50. Jakobsson ME, Moen A, Bousset L, Egge-Jacobsen W, Kernstock S, Melki R, Falnes P. Identification and characterization of a novel human methyltransferase modulating Hsp70 protein function through lysine methylation. *J Biol Chem*. 2013; 288:27752–63. [PubMed: 23921388]
51. Zhu H, Klemic JF, Chang S, Bertone P, Casamayor A, Klemic KG, Smith D, Gerstein M, Reed MA, Snyder M. Analysis of yeast protein kinases using protein chips. *Nat Genet*. 2000; 26:283–9. [PubMed: 11062466]
52. Zhu H, Bilgin M, Bangham R, Hall D, Casamayor A, Bertone P, Lan N, Jansen R, Bidlingmaier S, Houfek T, Mitchell T, Miller P, Dean RA, Gerstein M, Snyder M. Global analysis of protein activities using proteome chips. *Science*. 2001; 293:2101–5. [PubMed: 11474067]
53. Meng L, Michaud GA, Merkel JS, Zhou F, Huang J, Mattoon DR, Schweitzer B. Protein kinase substrate identification on functional protein arrays. *BMC Biotechnol*. 2008; 8(22)
54. Del Rincón SV, Rogers J, Widschwendter M, Sun D, Sieburg HB, Spruck C. Development and validation of a method for profiling post-translational modification activities using protein microarrays. *PLoS One*. 2010; 5:e11332. [PubMed: 20596523]
55. Lee J, Bedford MT. PABP1 identified as an arginine methyltransferase substrate using high-density protein arrays. *EMBO Rep*. 2002; 3:268–73. [PubMed: 11850402]
56. Levy D, Liu CL, Yang Z, Newman AM, Alizadeh AA, Utz PJ, Gozani O. A proteomic approach for the identification of novel lysine methyltransferase substrates. *Epigenetics Chromatin*. 2011; 4(19)
57. Liu Q, Dreyfuss G. In vivo and in vitro arginine methylation of RNA-binding proteins. *Mol Cell Biol*. 1995; 15:2800–8. [PubMed: 7739561]
58. Sharrocks AD, Yang SH, Galanis A. Docking domains and substrate-specificity determination for MAP kinases. *Trends Biochem Sci*. 2000; 25:448–53. [PubMed: 10973059]
59. Al-Ayoubi AM, Zheng H, Liu Y, Bai T, Eblen ST. Mitogen-activated protein kinase phosphorylation of splicing factor 45 (SPF45) regulates SPF45 alternative splicing site utilization, proliferation, and cell adhesion. *Mol Cell Biol*. 2012; 32:2880–93. [PubMed: 22615491]
60. Swiercz R, Person MD, Bedford MT. Ribosomal protein S2 is a substrate for mammalian PRMT3 (protein arginine methyltransferase 3). *Biochem J*. 2005; 386:85–91. [PubMed: 15473865]
61. El-Andaloussi N, Valovka T, Toueille M, Steinacher R, Focke F, Gehrig P, Covic M, Hassa PO, Schär P, Hübscher U, Hottiger MO. Arginine methylation regulates DNA polymerase beta. *Mol Cell*. 2006; 22:51–62. [PubMed: 16600869]
62. Koh SS, Li H, Lee YH, Widelitz RB, Chuong CM, Stallcup MR. Synergistic coactivator function by coactivator-associated arginine methyltransferase (CARM) 1 and beta-catenin with two different classes of DNA-binding transcriptional activators. *J Biol Chem*. 2002; 277:26031–5. [PubMed: 11983685]
63. Le Romancer M, Treilleux I, Leconte N, Robin-Lespinnas Y, Sentis S, Boucheikoua-Bouzaghrou K, Goddard S, Gobert-Gosse S, Corbo L. Regulation of estrogen rapid signaling through arginine methylation by PRMT1. *Mol Cell*. 2008; 31:212–21. [PubMed: 18657504]
64. Ren J, Wang Y, Liang Y, Zhang Y, Bao S, Xu Z. Methylation of ribosomal protein S10 by protein-arginine methyltransferase 5 regulates ribosome biogenesis. *J Biol Chem*. 2010; 285:12695–705. [PubMed: 20159986]
65. Abu-Farha M, Lanouette S, Elisma F, Tremblay V, Butson J, Figeys D, Couture JF. Proteomic analyses of the SMYD family interactomes identify HSP90 as a novel target for SMYD2. *J Mol Cell Biol*. 2011; 3:301–8. [PubMed: 22028380]
66. Shah K, Liu Y, Deirmengian C, Shokat K. Engineering unnatural nucleotide specificity for Rous sarcoma virus tyrosine kinase to uniquely label its direct substrates. *Proc Natl Acad Sci U S A*. 1997; 94:3565–70. [PubMed: 9108016]
67. Bishop A, Shah K, Liu Y, Witucki L, Kung C, Shokat K. Design of allele-specific inhibitors to probe protein kinase signaling. *Curr Biol*. 1998; 8:257–66. [PubMed: 9501066]
68. Liu Y, Shah K, Yang F, Witucki L, Shokat K. Engineering Src family protein kinases with unnatural nucleotide specificity. *Chem Biol*. 1998; 5:91–101. [PubMed: 9495830]
69. Allen J, Lazerwith S, Shokat K. Bio-orthogonal affinity purification of direct kinase substrates. *J Am Chem Soc*. 2005; 127:5288–9. [PubMed: 15826144]

70. Blethrow J, Glavy J, Morgan D, Shokat K. Covalent capture of kinase-specific phosphopeptides reveals Cdk1-cyclin B substrates. *Proc Natl Acad Sci U S A*. 2008; 105:1442–7. [PubMed: 18234856]
71. Carlson SM, Chouinard CR, Labadorf A, Lam CJ, Schmelzle K, Fraenkel E, White FM. Large-Scale Discovery of ERK2 Substrates Identifies ERK-Mediated Transcriptional Regulation by ETV3. *Science Signaling*. 2011; 4
72. Banko MR, Allen JJ, Schaffer BE, Wilker EW, Tsou P, White JL, Villén J, Wang B, Kim SR, Sakamoto K, Gygi SP, Cantley LC, Yaffe MB, Shokat KM, Brunet A. Chemical genetic screen for AMPK α 2 substrates uncovers a network of proteins involved in mitosis. *Mol Cell*. 2011; 44:878–92. [PubMed: 22137581]
73. Berchowitz LE, Gajadhar AS, van Werven FJ, De Rosa AA, Samoylova ML, Brar GA, Xu Y, Xiao C, Futcher B, Weissman JS, White FM, Amon A. A developmentally regulated translational control pathway establishes the meiotic chromosome segregation pattern. *Genes Dev*. 2013; 27:2147–63. [PubMed: 24115771]
74. Lin Q, Jiang F, Schultz PG, Gray NS. Design of allele-specific protein methyltransferase inhibitors. *J Am Chem Soc*. 2001; 123:11608–13. [PubMed: 11716715]
75. Binda O, Boyce M, Rush JS, Palaniappan KK, Bertozzi CR, Gozani O. A chemical method for labeling lysine methyltransferase substrates. *Chembiochem*. 2011; 12:330–4. [PubMed: 21243721]
76. Kolb HC, Finn MG, Sharpless KB. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew Chem Int Ed Engl*. 2001; 40:2004–2021. [PubMed: 11433435]
77. Best MD. Click chemistry and bioorthogonal reactions: unprecedented selectivity in the labeling of biological molecules. *Biochemistry*. 2009; 48:6571–84. [PubMed: 19485420]
78. Islam K, Zheng W, Yu H, Deng H, Luo M. Expanding cofactor repertoire of protein lysine methyltransferase for substrate labeling. *ACS Chem Biol*. 2011; 6:679–84. [PubMed: 21495674]
79. Wang R, Ibáñez G, Islam K, Zheng W, Blum G, Sengelaub C, Luo M. Formulating a fluorogenic assay to evaluate S-adenosyl-L-methionine analogues as protein methyltransferase cofactors. *Mol Biosyst*. 2011
80. Wang R, Zheng W, Yu H, Deng H, Luo M. Labeling substrates of protein arginine methyltransferase with engineered enzymes and matched S-adenosyl-L-methionine analogues. *J Am Chem Soc*. 2011; 133:7648–51. [PubMed: 21539310]
81. Islam K, Bothwell I, Chen Y, Sengelaub CA, Wang R, Deng H, Luo M. Bioorthogonal Profiling of Protein Methylation (BPPM) Using Azido Derivative of S-adenosyl-L-methionine. *J Am Chem Soc*. 2012
82. Guo H, Wang R, Zheng W, Chen Y, Blum G, Deng H, Luo M. Profiling Substrates of Protein Arginine N-Methyltransferase 3 with S-Adenosyl-L-methionine Analogues. *ACS Chem Biol*. 2014; 9:476–84. [PubMed: 24320160]

Abbreviations

ADMA	asymmetric di-methyl arginine
GST	glutathione-S-transferase
HILIC	hydrophilic interaction liquid chromatography
IP	immunoprecipitation
KMT	lysine methyltransferase
LC-MS/MS	liquid chromatography and tandem mass spectrometry
MMA	mono-methyl arginine
PAGE	polyacrylamide gel electrophoresis
PRMT	protein arginine methyltransferase

SCX	strong cation exchange
SDMA	symmetric di-methyl arginine
SET	<u>Su</u> (var)3-9, <u>e</u> nhancer-of-zeste, <u>t</u> riThorax
SILAC	stable isotopic labeling by amino acids in cell culture

Highlights

- Methylation of histone and non-histone proteins is important in cellular signaling
- Methyl-binding domains are emerging tools for detecting protein methylation
- Quantitative proteomics is revealing signaling through dynamic protein methylation
- Chemical biology is providing new strategies to map methylation signaling networks

Figure 1. Methylation states of arginine and lysine

a. arginine residues are methylated by protein arginine methyltransferases at an N^G position to form mono-methyl arginine (MMA), and then methylated at the same or the opposing nitrogen to form symmetric or asymmetric di-methyl arginine (SDMA and ADMA). **b.** lysine residues are methylated by lysine methyltransferases (KMTs) at the ϵ -nitrogen to form mono-, di-, or tri-methyl lysine.

Figure 2. Enriching and identify methylated peptides by peptide immunoprecipitation (IP) and LC-MS/MS

A protein sample is digested with enzymes such as trypsin, Arg-C or Glu-C to generate a predictable mixture of peptides. Peptides containing modified residues are isolated by IP with a broad-specificity antibody and then identified using LC-MS/MS.

Figure 3. *In vivo* labeling of methylated residues with heavy isotopes

Cells are grown in media containing methionine with a specific isotopic label. Labeled methionine is taken up from the media and converted to S-adenosyl methionine (SAM) by the enzyme methionine adenosyltransferase. The labeled sulfonium methyl group is finally used as a methyl donor by PRMTs and KMTs to produce methylated residues with a distinctive mass shift.

Figure 4. Overview of enrichment and quantitative analysis of protein methylation using 3xMBT

a. Methylated proteins are enriched from cell extract using immobilized 3xMBT^{28; 29}. The complex of domain with methyl-lysine is released from the beads, the domain is depleted to recover methylated proteins, and proteins are identified by LC-MS/MS. Only a subset of methylated residues are directly identified using this approach. **b.** Cells are prepared in media containing isotopically labeled arginine and lysine³⁸. The two extracts are incubated with immobilized 3xMBT or with the negative control 3xMBTD_{355N}. The two pull-downs are then combined and analyzed by LC-MS/MS to identify proteins bound in a methyl-dependent manner. The same approach can be used to compare methylation levels between different biological conditions.

Figure 5. Examples of SAM and S-adenosyl homocysteine analogs used for chemical biology N⁶-benzyl S-adenosyl homocysteine can be used to selectively inhibit methyltransferase enzymes engineered with an expanded ATP binding pocket⁷⁴. Propargyl SAM can be utilized directly by some KMTs to generate alkynlated protein substrates⁷⁵. Larger alkyne or azide substituents at the sulfonium position can only be used efficiently by engineered methyltransferase enzymes, and have improved stability in solution relative to propargyl SAM^{78; 81}. In each case Click chemistry can be used to conjugate proteins labeled with alkyne or azide to moieties for enrichment (such as biotin) or a fluorescent label.

Table 1

Summary of proteome-wide studies identifying protein methylation.

Study ^a	Year	Technique(s)	# Proteins ^b	# Sites ^b
Boisvert et al. ²² (Stéphane Richard)	2003	Protein IP for dimethyl-R	>200 enriched (R)	NA
Ong et al. ²⁶ (Mattias Mann)	2004	Protein IP for methyl-R and K Isotopic labeling with heavy methionine in cell culture	31 (R) 2(K)	57 (R) 2(K)
Uhlmann et al. ²⁵ (Oreste Acuto)	2012	Peptide IP for methyl-R, SCX, isoelectric focusing, HILIC	131 (R)	249 (R)
Bremang et al. ³⁹ (Tiziana Bonaldi)	2013	Protein IP, heavy methionine labeling, fractionation by SDS-PAGE, isoelectric focusing	139	323 (R) 74 (K)
Fisk et al. ³⁵ (Laurie Read)	2013	Protein IP for dimethyl-R followed by SCX and reverse phase chromatography (<i>T. brucei</i> mitochondria)	167 (R)	253 (R)
Lott et al. ³⁶ (Laurie Read)	2013	Protein IP for dimethyl-R followed by SCX and reverse phase chromatography (<i>T. brucei</i>)	676 (R)	1,332 (R)
Moore et al. ²⁸ (Or Gozani)	2013	Methyl-K binding domain pull-down Heavy methionine labeling	313 enriched (K) 24 direct (K)	26 (K)
Liu et al. ³⁰ (Shawn Shun-Cheng)	2013	Methyl-K binding domain pull-down Computational site prediction	109 enriched (K) 29 direct (K)	40 (K)
Cao et al. ²⁷ (Benjamin Garcia)	2013	Peptide IP and SCX	413 (K)	540 (K)
Guo et al. ²³ (Michael Comb)	2014	Peptide IP for methyl-R and K	>800 (R) 130 (K)	>1,000 (R) 165 (K)
Sylvestersen et al. ³⁷ (Michael Nielsen)	2014	Peptide IP for monomethyl-R and SCX	494 (R)	1,027 (R)

^a Studies are listed by first author with corresponding author shown in parentheses^b The number of modified proteins or sites is the number directly shown to be methylated, except when indicated otherwise (Boisvert et al., Moore et al. and Liu et al.). Methylation sites are indicated as arginine (R) or lysine (K) unless the original publication does not distinguish. Studies used mammalian cells or tissues unless indicated otherwise.