Review

Protein methylation at the surface and buried deep: thinking outside the histone box

Steven G. Clarke

Department of Chemistry and Biochemistry and the Molecular Biology Institute, UCLA, Los Angeles, CA 90095, USA

Methylated lysine and arginine residues in histones represent a crucial part of the histone code, and recognition of these methylated residues by protein interaction domains modulates transcription. Although some methylating enzymes appear to be histone specific, many can modify histone and non-histone substrates and an increasing number are specific for non-histone substrates. Some of the non-histone substrates can also be involved in transcription, but a distinct subset of protein methylation reactions occurs at residues buried deeply in ribosomal proteins that may function in protein–RNA interactions rather than protein–protein interactions. Additionally, recent work has identified enzymes that catalyze protein methylation reactions at new sites in ribosomal and other proteins. These reactions include modifications of histidine and cysteine residues as well as the N terminus.

Protein methyltransferases: a brief overview

Approximately 1–2% of genes from a variety of prokaryotic and eukaryotic organisms encode methyltransferases, a large fraction of which are specific for protein substrate modification [1–4]. Most of these enzymes are members of the seven-beta-strand [5], SET-domain [6], or SPOUT (SpoU and TrmD) [7] structural protein families. Although seven-beta-strand methyltransferases catalyze a wide variety of methylation reactions at many different types of residues, the SET-domain enzymes characterized to date are all protein lysine methyltransferases. With the recent exception of one protein arginine methyltransferase [8], enzymes from the SPOUT family only appear to catalyze RNA methylation reactions [7]. In humans the largest group of methyltransferases encompasses some 56 SET-domain species [4,9]. Although it seems that the majority of these human SET-domain proteins are histone methyltransferases, it is becoming evident that more and more of them can modify other types of proteins. Protein methylation is perhaps most common at lysine and arginine residues, at least in eukaryotic cells. However, there are many other sites for such modification in proteins including histidine, glutamate, glutamine, asparagine, D-aspartate/L-isoaspartate, cysteine, and N-terminal and C-terminal residues [10,11]. Recent studies have now identified the first enzymes specific for catalyzing the methylation of three of these residues: the N terminus [12,13], histidines [14], and cysteines [15]. This review focuses on current advances in protein methyltransferase enzymology in yeast and mammalian systems, with particular emphasis on reactions in ribosomal systems.

Protein lysine methylation: not just for histones

Although most protein lysine methyltransferases are SET-domain family members [4,9], there is an increasing number of seven-beta-strand enzymes being reported that catalyze similar reactions [16–19]. These enzymes result in the formation of monomethyl-, dimethyl-, and trimethyllysine residues. Some enzymes are specific for one or two of these modifications and some result in the formation of all three derivatives [20]. In histone tails these modified residues are found at the surface of the nucleosome and are recognized by protein interaction domain species that lead to transcriptional activation or repression [17,21–24]. However, in ribosomal proteins these modifications can occur at buried residues that interact directly with rRNA species in the ribosomal interior [25].

The first methyltransferase identified, with what would later be designated the SET domain, was the enzyme responsible for the trimethylation of lysine 14 in the large subunit of ribulose-1,5-bisphosphate carboxylase oxygenase (RuBisCO), the plant enzyme essential for fixing much of the carbon dioxide in the biosphere [26]. At the time, the amino acid sequence for this enzyme showed no similarity with other methyltransferases. Independently, the SET domain had been identified from the encoded amino acid sequences of three *Drosophila* genes associated with development: Su(var)3-9, enhancer of zeste, and trithorax. The realization that the sequences of these domains were similar to those of RuBisCO showed that the SET domain represents a catalytic core of protein lysine methyltransferases and that each of these *Drosophila* proteins catalyzes histone lysine methylation (reviewed in [21]). Interestingly, whereas the function of RuBisCO methylation remains unclear [27], a tremendous amount is now known about the enzymes that methylate histones and their biological role in maintaining and altering the histone code [22–24,28]. Indeed, for many scientists, SET-domain proteins and histone methyltransferases are almost interchangeable terms.

Corresponding author: Clarke, S.G. (clarke@chem.ucla.edu)
Keywords: histone methylation; ribosomal protein methylation; PRMTs; SET-domain methyltransferases

0968-0004/$ – see front matter
 http://dx.doi.org/10.1016/j.tibs.2013.02.004

Analyses of the SET-domain proteins in the yeast *Saccharomyces cerevisiae*, however, revealed that at least half of these methyltransferases recognize non-histone substrates. The yeast genome encodes 12 SET-domain proteins that can be divided into two sequence-related subfamilies containing six members each by position-specific iterated (PSI)-BLAST and pattern hit initiated (PHI)-BLAST searches [29] (Table 1). The first subfamily encodes three proteins that catalyze the methylation of histone proteins: Set1, which modifies lysine 4 of histone H3 [30]; Set2, which modifies lysine 36 of histone H3 [31]; and Set5, which modifies lysines 5, 8, and 12 of histone H4 [32]. No methyltransferase function has yet been assigned for the remaining three members of the group: Set3, Set4, and Set6. Surprisingly, none of the members of the second subfamily of SET-domain proteins in yeast modifies histones. This family consists of four enzymes that modify ribosomal large subunit proteins, one enzyme that modifies elongation factor 1A, which brings aminoacyl-tRNAs into the ribosome, and one enzyme that modifies cytochrome c (Table 1). These results suggest that the biological role of SET-domain methyltransferases includes important translational, as well as transcriptional, components.

Analyses of these yeast SET-domain enzymes have revealed that in most cases the enzymes appear to be specific for modifying a single protein substrate, at a single site in the protein sequence, and to a single degree of methylation (Table 1). However, there are exceptions. The Set5 enzyme monomethylates three nearby lysine residues in histone H4 [32] and the Rkm1 enzyme dimethylates two nearby lysine residues in the ribosomal protein Rpl23ab [33]. Perhaps more interestingly, the Set1 methyltransferase, in its role as the catalytic unit of the complex proteins associated with Set1 (COMPASS) complex, has been shown to form a trimethyllysine residue on histone H3 and dimethyllysine or trimethyllysine residues on the kinetochore Dam1 protein [30]. In this case, it has been suggested that these distinct modifications are linked in a regulatory crosstalk that relays changes in chromatin to the apparatus for chromosome segregation [30]. In any case, it is hard to rule out the possibility that other yeast SET-domain enzymes may also modify additional methyl-accepting substrates in the cell.

In humans the SET-domain family includes some 56 members [4]. Initial analyses indicated that most of these enzymes might be specific for histones, although the evidence was based largely on the comparison of sequence similarities. However, current work, discussed below, has now established that many of these enzymes also catalyze protein lysine methylation of non-histone substrates and that some may not recognize histone substrates at all [34]. These results suggest that this class of enzymes has a significant impact on mammalian cellular physiology.

Additionally, other recent studies have demonstrated that an increasing number of protein lysine methyltransferases are non-SET-domain enzymes. In yeast, four seven-beta-strand enzymes have been identified, including the disruptor of telomeric silencing (Dot)1 histone methyltransferase as well as enzymes that modify ribosomal proteins and translational elongation factors (Table 2). In fact, from the 23 protein methyltransferases with defined functions in yeast, 16 modify proteins of the translational apparatus (e.g., ribosomal proteins, elongation and release factors), highlighting the broad importance of protein methylation in translation [35,36]. In mammalian cells, additional seven-beta-strand protein lysine methyltransferases modify calmodulin [16] and valosin-containing protein (VCP), an ATP-dependent chaperone [19]. Sequence analysis has identified eight additional human

Table 1. SET-domain protein lysine methyltransferases in the yeast *Saccharomyces cerevisiae*[a]

Protein	Substrate(s) (position given from the mature N-terminal residue unless otherwise indicated)	Product(s)	Methylated residue (surface exposed or buried)	Refs
Subfamily one				
Set1 (as the catalytic component of the COMPASS complex)	Histone H3 Lys4 Dam1 Lys233 (from initiator methionine)	Trimethyl Dimethyl (possible trimethyl)	Surface Unknown	[30]
Set2	Histone H3 Lys36	Trimethyl	Surface	[31]
Set3	No known methyltransferase activity			
Set4	No known methyltransferase activity			
Set5	Histone H4 Lys5, Lys8, Lys12 Histone H2A Lys4, Lys7	Monomethyl at all three sites Unknown	Surface	[32]
Set6	No known methyltransferase activity			
Subfamily two				
Rkm1	Ribosomal protein Rpl23ab Lys105, Lys109	Dimethyl at both sites	Surface (interface between small and large subunits)	[33]
Rkm2	Ribosomal protein Rpl12ab Lys3	Trimethyl	Surface	[25]
Rkm3	Ribosomal protein Rpl42ab Lys39	Monomethyl	Buried	[25]
Rkm4	Ribosomal protein Rpl42ab Lys54	Monomethyl	Buried (close contacts to 25S rRNA O2 of cytosine-2764 and OP2 of cytosine-93)	[25]
Efm1	Elongation factor eEF1A Lys30	Monomethyl	Surface	[35,63]
Ctm1	Iso-1-cytochrome c Lys72	Trimethyl	Surface	[64]

[a]Protein designations given from the *Saccharomyces* Genome Database (www.yeastgenome.org/).

Table 2. Non-SET-domain protein methyltransferases and their established substrates in the yeast *Saccharomyces cerevisiae*[a]

Protein (unless otherwise indicated, all are seven-beta-strand methyltransferases)	Substrate and *position* (given from the mature N-terminal residue, unless otherwise designated)	Product(s)	Methylated residue (surface, exposed, or buried)	Refs
Protein lysine methyltransferases				
Dot1	Histone H3 Lys79	Mono-, di-, and tri-methyl	Surface	[17]
See1	Elongation factor eEF1A (Tef1/Tef2) Lys316	Dimethyl	Surface	[63]
Efm2	Elongation factor EF3A (Yef3) Lys186	Trimethyl	Unknown	[35]
Rkm5	Ribosomal protein Rpl1 Lys46	Monomethyl	Surface	[18]
Protein arginine methyltransferases				
Rmt1 (Hmt1)	Many	ω-Monomethyl; ω-asymmetric dimethyl		[65]
Rmt2	Ribosomal protein Rpl12ab Arg66	δ-Monomethyl	Protein surface	[52]
Hsl7	Unknown	ω-Monomethyl; ω-symmetric dimethyl		[66]
Sfm1 (SPOUT family methyltransferase)	Ribosomal protein Rps3 Arg145	ω-Monomethyl	Buried (close contacts with 18S rRNA adenine-1427)	[8]
N-terminal protein methyltransferase				
Ntm1	Ribosomal protein Rpl12ab Pro1 Ribosomal protein Rps25ab Pro1 Several others?	*N*-dimethyl *N*-dimethyl	Surface Surface	[12]
Protein histidine methyltransferase				
Hpm1	Ribosomal protein Rpl3 His242	3-(Tau)methyl	Buried (contacts with 25S rRNA O6 guranine-878 and OP1 adenine-876)	[14]
Protein glutamine methyltransferases				
Mtq1	Mitochondrial translational release factor Mrf1 Gln287 (from initiator Met)	*N*-5-Monomethyl amide	Surface	[67]
Mtq2	Cytoplasmic translational release factor Sup45 Gln182	*N*-5-Monomethyl amide	Surface	[67]
C-terminal protein methyltransferase				
Ppm1	Protein phosphatase 2A catalytic subunit C-terminal Leu	Methyl ester	Surface	[68]
C-terminal protein isoprenylcysteine methyltransferase				
Ste14 (membrane-bound methyltransferase)	Many	C-Terminal isoprenylated cysteine methyl ester	Surface or membrane	[69]

[a]Protein designations given from the *Saccharomyces* Genome Database (www.yeastgenome.org/).

gene products as potential protein lysine methyltransferases of this type [19]. It will be interesting to see whether these additional enzymes also have substrates associated with the translational apparatus. The diversity of SET-domain enzymes and seven-beta-strand methyltransferases specific for protein lysine residues suggests a wide range of physiological roles for the modification reactions that they catalyze. Recent work, described below, has provided evidence that protein lysine methylation may be particularly important not only in the function of proteins involved in translation but also in that of non-histone proteins associated with transcriptional processes.

Methylation of ribosomal proteins, transcription factors, and other non-histone proteins at lysine residues

In *S. cerevisiae* five non-histone proteins are modified at lysine residues by six SET-domain methyltransferases (Table 1) and three non-histone proteins are modified at lysine residues by three seven-beta-strand enzymes (Table 2). Interestingly, with the exception of cytochrome c, all of these proteins are involved in translation, either as ribosomal proteins or elongation factors. The availability of an atomic-resolution structure of the yeast ribosome at 3 Å [37] enables us to map the position of most of the modified residues (Figure 1). Although methyl groups were not modeled into this structure, it is clear that some of the methylated sites are exposed to the surface and some of the sites are buried deeply within the ribosome. For example, in this structure Rpl1 density was not found, presumably because it was easily detached from the ribosomal surface, and the N-terminal tail of Rpl12ab containing the trimethylated lysine 3 appeared to be disordered at the ribosomal surface. Thus, the methylated lysine residues on Rpl1 and Rpl12ab would be expected to be exposed. By contrast, it is apparent that the two methylated residues on Rpl42ab are localized deeply within the ribosomal structure (Figure 2a). Here, the monomethylated lysine 54

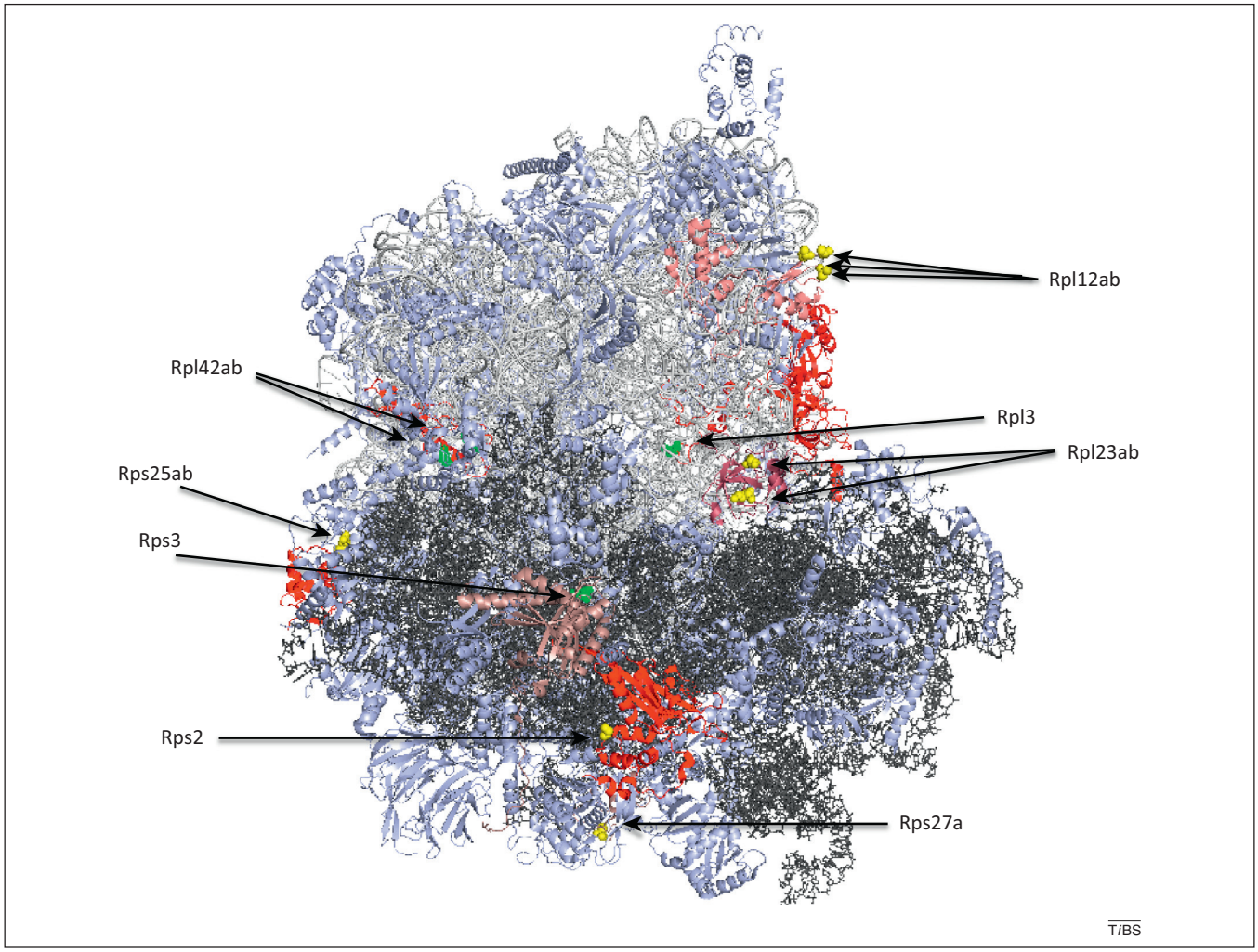

Figure 1. Surface and buried sites of methylation on cytoplasmic ribosomal proteins in the yeast *Saccharomyces cerevisiae*. The 25S ribosomal RNA of the large subunit is shown in light gray; the 18S ribosomal RNA of the small subunit is shown in dark gray. Non-methylated proteins are shown in light blue; methylated proteins (Tables 1 and 2) are shown in pink (Rpl12ab, Rpl23ab, Rps27a, Rps3) and red (Rpl3, Rps2, Rps25ab, Rpl42ab). The approximate positions of surface-exposed methyl groups are shown as yellow spheres; buried methyl groups are represented as green spheres. The illustration was made using PyMOL from the Protein Data Bank (PDB) structures 3U5F, 3U5G, 3U5H, and 3U5I [37].

residue is modeled showing van der Waals contacts with two cytidine residues on the 25S rRNA. For the two dimethyllysine residues on Rpl23ab, it appears that these residues form part of the interface with the small ribosomal subunit, perhaps in a position that interacts with the incoming mRNA (Figure 3). These interactions between protein lysine methyl groups and RNA appear to be novel; all other interactions described to date involve proteins [38]. Modification of Rpl42ab presumably occurs before final ribosomal assembly to allow the methyltransferase access to the lysine side chain, although it is possible that dynamic flexibility in ribosomes may enable the residue to flip out of the ribosomal structure for the methylation reaction.

It can be speculated that the methyl groups may guide productive interactions between the ribosomal RNA and ribosomal proteins. This may occur by methyl groups blocking unfavorable interactions by steric exclusion or disruption of hydrogen bonding patterns. Additionally, it is now clear that hydrogen atoms on methyl groups that are attached to nitrogen atoms can themselves be hydrogen bond donors [39], and may thus contribute to new hydrogen bond networks. For example, such interactions are central to the recognition of trimethyllysine residues by a novel histone H3 binding module [40].

The 56 members of the SET-domain family encoded by the human genome can be divided into ten classes by their amino acid sequences [4]. Nine of these classes, with 53 members in total, were initially surmised to catalyze histone methylation based on sequence similarity to known protein lysine methyltransferases. Only one class (VII), consisting of three members, was initially associated with non-histone substrates because of sequence similarity to the RuBisCO methyltransferase [4]. Recent work has shown that the situation is more complex than previously thought. In many cases, enzymes have been found to have activity against histone and non-histone substrates (Table 3). This development is an important one because so many of the protein methyltransferases that were characterized initially displayed the 'tripartite' specificity seen in most of the yeast enzymes presented in Tables 1 and 2: one protein substrate; one site within that protein; and one

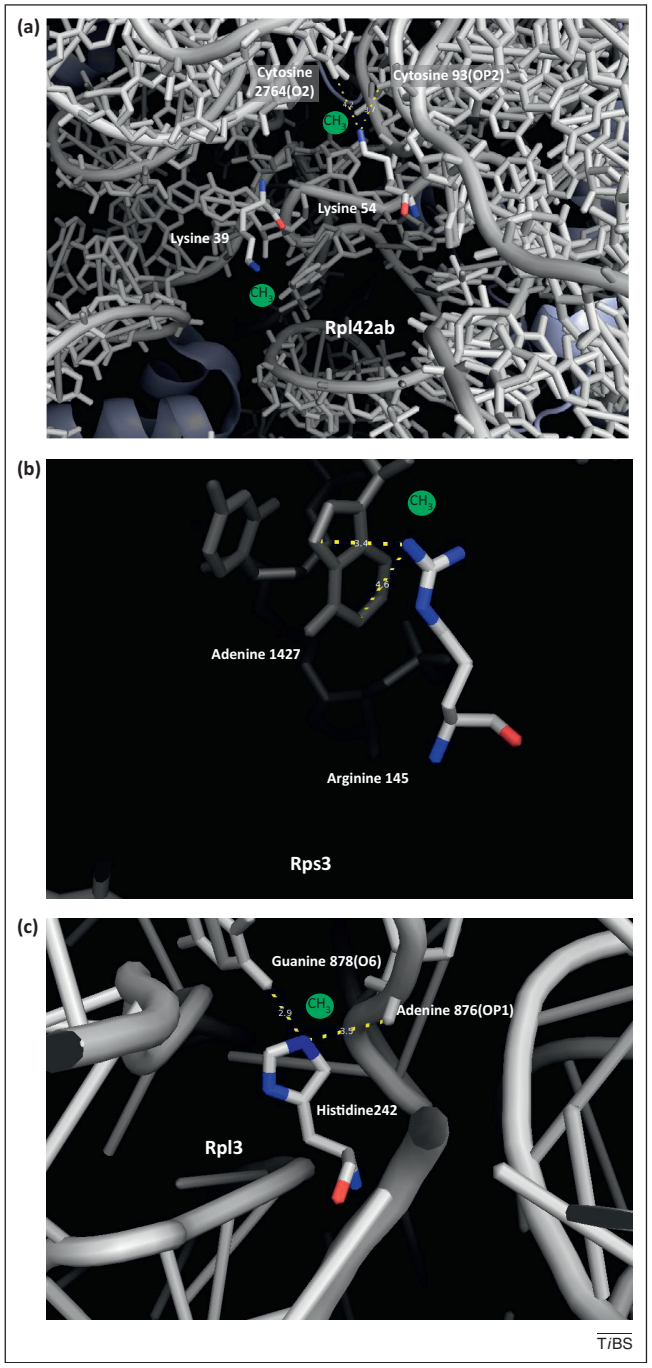

Figure 2. Zoomed in view of methylated lysine, arginine, and histidine residues with ribosomal RNA in yeast cytoplasmic ribosomes. The monomethylated residues of Rpl42ab [(**a**) Lys39; Lys54], Rps3 [(**b**) Arg145], and Rpl3 [(**c**) His242] are shown with nitrogen atoms in blue and oxygen atoms in red [8,14,25]. Ribosomal 25S (a,c) and 18S (b) RNA are shown in gray. Methyl groups were not modeled into these structures; possible locations are suggested by the green spheres. Examples of close contact distances (i.e., less than 5 Å) are shown between the methylated atom in the protein and the RNA to emphasize the apposition of the methylated residue and the RNA, although these interactions may differ in a refined structure that includes the methyl groups. The illustration was made using PyMOL and the Protein Data Bank (PDB) structures 3U5F, 3U5G, 3U5H, and 3U5I [37].

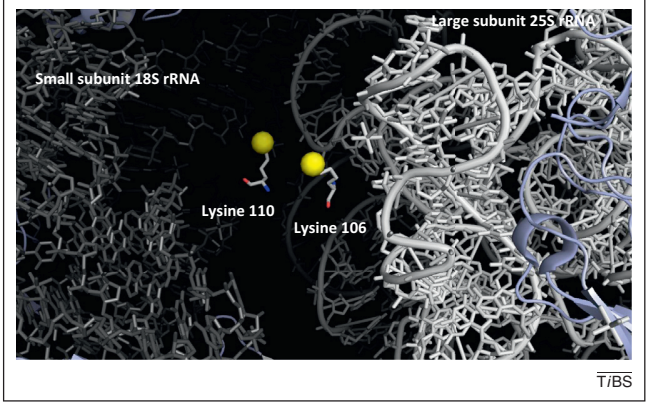

Figure 3. Intrasubunit localization of the methylated lysine residues of Rpl23ab in yeast cytoplasmic ribosomes. Dimethyllysine residues 105 and 109 are shown with the ε-amino group as a yellow sphere. These residues are positioned at the interface of the small and large ribosomal subunits; 18S rRNA is shown in gray on the left and 25S rRNA is shown in white on the right. The illustration was made using PyMOL from the Protein Data Bank (PDB) structures 3U5F, 3U5G, 3U5H, and 3U5I [37].

level of modification of that site for each enzyme. This is clearly not the case for the nine mammalian SET-domain methyltransferases described in Table 3. Eight of these enzymes clearly methylate histone and non-histone substrates. An additional enzyme, the SET and MYND domain-containing protein 2 (SMYD2), modifies multiple non-histone substrates. The rapid pace of discovery of novel non-histone protein substrates for these enzymes suggests that we may only be starting to understand the true range of their methyl-accepting substrates. The recognition of multiple substrates for protein lysine methyltransferases has certainly opened Pandora's Box. How many additional substrates have been missed while characterizing these protein lysine methyltransferases? May other SET-domain enzymes, thought to only modify a single substrate and site, in fact be capable of modifying additional proteins, or additional sites on the protein, in the cell?

What can we conclude at this point about the nature of the identified non-histone substrates compiled in Table 3? It appears that the apple does not fall far from the tree here; many of these non-histone substrates are also involved in transcriptional control. Significantly, 13 of these proteins with lysine methyl-acceptors are transcription factors or closely associated with transcription factors, such as nuclear receptors or the DNA methyltransferase DNMT1 (Table 3). Of the remaining six non-histone substrates, one is involved in apoptosis, two are heat shock proteins, one is a receptor tyrosine protein kinase, and two are serine/threonine protein kinases involved in cell cycle control (Table 3). Certainly, more substrates remain to be identified and it will be of interest to see if the majority of them are also associated with transcriptional regulation. Interestingly, to date no mammalian ribosomal proteins have been found to be modified by SET-domain methyltransferases, suggesting that different organisms may use protein lysine methyltransferases to different ends. However, less is known about mammalian ribosomal methylation and it will be instructive to identify the enzymes responsible for the known sites of methylation on lysine residues, including lysine 4 of RL29 and lysine 22 of RL40 in rat ribosomes [41].

How substrate specificity is determined in SET-domain enzymes has been explored in several recent papers. A study with peptide substrate arrays for the human SETD8

Table 3. Mammalian SET-domain protein methyltransferases that have been shown to modify non-histone substrates

Protein[a]	SET-domain family [4]	Substrate and methylation position (from the mature N terminus, unless otherwise indicated)	Refs
Enzyme(s) that methylate histones and non-histone substrates			
EZH2 (KMT6)	Class I – EZ	Histone H3 Lys27	[70,71]
		Retinoic-acid-related orphan nuclear receptor α (RORα) Lys38	[70]
		Transcription factor GATA4 Lys299	[71]
SETD1A (KMT2F)	Class II – SET2	Histone H3 Lys4	[72]
		Heat-shock protein 70 (HSP70; HSPA1A) Lys560 (dimethyl)	[73]
EHMT1 (KMT1D)(GLP)/ EHMT2 (KMT1C)(G9A)	Class V – Suvar-3-9 – histone H3 Lys9	Histone H1	[74]
		Histone H3 Lys9 (mono and dimethyl)	[74]
		Histone H3 Lys27	[74]
		Chromodomain Y-like protein (CDYL1) repressor of transcription Lys135 (mono, di-, and tri-methyl)	[75]
		Widely interspaced zinc-finger-containing (WIZ) transcription factor Lys1162 (di- and tri-methyl)	[75]
		Apoptotic chromatin condensation inducer in the nucleus (ACINUS) Lys654 (di- and tri-methyl)	[75]
		p53 Transcription factor Lys373 (dimethyl)	[76]
		MyoD transcription factor Lys104	[77]
SETDB1 (KMT1E)	Class V – Suvar-3-9 – histone H3 Lys9	Histone H3 Lys9	[78]
		Human immunodeficiency virus-1 Tat transcriptional activator Lys50, Lys51	
SMYD3	Class VI – SMYD	Histone H4 Lys5	[79]
		Vascular endothelial growth factor receptor 1 tyrosine kinase Lys831	[80]
SETD6	Class VII – SET6	Histone H2AZ Lys7	[81]
		RelA subunit of NF-κB transcription factor Lys310 (monomethyl)	[45]
		Serine/threonine protein kinase PLK1	[45]
		Serine/threonine protein kinase PAK4	[45]
SETD8 (SET8) (KMT5A)(PrR/SET7)	PR/SET	Histone H4 Lys20 (monomethyl)	[42]
		p53 Transcription factor Lys382	[42,82]
SETD7 (KMT7) (SET7/9)(SET7)(SET9)	SET7	A wide variety of substrates including:	[34,43,83]
		histone H2A	
		histone H2B	
		DNMT1 DNA methyltransferase Lys142	
		estrogen receptor α	
		p53 transcription factor Lys372	
		Transcription initiation factor TFIID TAF10 subunit Lys189	
		Forkhead box protein (FOXO)3 transcription factor Lys270	
Enzymes that methylate only non-histone substrates			
SMYD2 (KMT3C)	Class VI – SMYD	Monomethylation at:	
		p53 transcription factor Lys 370	[84]
		Heat-shock protein (Hsp)90 Lys209, Lys615	[85]
		Retinoblastoma-associated RB1 transcription factor Lys850	[86]

[a]Abbreviations: CDYL1, chromodomain Y-like protein; EHMT, euchromatic histone-lysine N-methyltransferase; EZH2, enhancer of zeste homolog 2; SETDB1, SET domain bifurcated 1; SETD1A, SET domain-containing protein 1A; SMYD2, SET and MYND domain-containing protein 2.

protein that methylates histones and p53 (Table 3) revealed a seven-residue consensus sequence R-H-R/K/Y-K-V/I/L/F/Y-L/F/Y-R [42]. Searching human sequences for this consensus sequence identified a number of new candidate substrate proteins, but interestingly none of these appeared to be recognized by SETD8 [42]. A similar approach for the SETD7 methyltransferase was more successful, leading to the identification of nine new non-histone substrates [43]. An alternative approach to identifying new substrates involves the utilization of protein arrays. This approach was used with success to identify new substrates for SETD7 and SETD6 [44] in arrays containing over 9500 human proteins. Finally, it is possible to test banks of SET-domain enzymes against specific substrates to identify which enzyme is responsible for a specific modification [45].

This review has concentrated on enzymes involved in lysine methylation in yeast and mammalian cells. However, one interesting prokaryotic enzyme is worthy of mention. The most promiscuous protein lysine methyltransferase may be an enzyme found in the hyperthermophilic archaeal species *Sulfolubus islandicus* [46]. This enzyme is a seven-beta-strand enzyme that appears to modify numerous proteins at multiple sites within each protein species. The large-scale conversion of lysine residues to dimethyllysine appears to be associated with the resistance of a protein to heat denaturation [46]. Whether such stabilization occurs for the eukaryotic proteins described here, including the proteins of the translational apparatus, remains to be seen.

At this point, our knowledge about the importance of protein lysine methylation is best established for the regulation of chromatin function, although we are beginning to understand better the role of this modification in translation. However, protein methylation occurs at many other residues and may have similar or distinct roles in a wide variety of systems. For all of the modifications described below recent evidence has pointed to roles in ribosomal structure and function, including additional examples suggesting direct interactions between methylated residues and RNA.

Non-histone protein methylation at arginine residues

Protein arginine methylation has been well studied in yeast and mammalian systems (for recent reviews, see [47–49]). In histones and in non-histone proteins, ω-monomethyl, ω-asymmetric, and ω-symmetric dimethylated residues are recognized by tudor protein interaction domains [49], largely in the same way as methylated lysine residues are recognized. In mammalian cells, these methylation reactions are catalyzed by a sequence-related family of nine seven-beta-strand methyltransferases designated protein arginine methyltransferase 1 (PRMT1) through PRMT9. Six of these enzymes (PRMT1, 2, 3, 4, 6, and 8) have been shown to catalyze asymmetric dimethylation, whereas one enzyme (PRMT5) has been shown to catalyze symmetric dimethylation [47,49]. PRMT7 appears unique in that it may only catalyze ω-monomethylation [50], whereas the specificity of PRMT9(4q31) [51], which is erroneously designated as PRMT10 in the UniProt database, has not been established. The specificity of these enzymes for protein substrates is generally much broader than that of the protein lysine methyltransferases. All of the enzymes for which activity has been shown can modify multiple substrates, often at multiple sites within a given protein [47]. In yeast, a smaller family of three seven-beta-strand enzymes includes Rmt1 (the homolog of PRMT1/2/3/4/6/8), Hsl7 (the homolog of PRMT5), and a distinct enzyme (Rmt2) that catalyzes the specific modification of the bridge, or δ-guanidino nitrogen atom, in an arginine residue of the large subunit ribosomal protein Rpl12ab (Table 2) [48,52]. Interestingly, Sfm1, an enzyme of the SPOUT family, members of which generally modify RNA species [7], has been recently shown to catalyze the ω-monomethylation of an arginine residue in the ribosomal small subunit protein Rps3 [8]. Although the methylated arginine residue in Rps3 is on the protein surface, it does not contact the surface of the ribosome or other ribosomal proteins. Rather, the methylated site is buried within the ribosomal RNA and makes close contact with the nitrogen atoms on adenine 1427 [8] (Figure 2b). Genes encoding orthologs of the Rmt2 and Sfm1 enzymes do not appear to be found in animal species. However, the discovery of these proteins does suggest that the family of protein arginine methyltransferases may be broader than previously imagined.

Unlike protein lysine methylation, for which much of the interest and work has centered on histone substrates, protein arginine methylation has been studied extensively not only with histone substrates and transcriptional control but also with substrates involved in signal transduction, DNA repair, and RNA splicing [47–49].

Present challenges in protein arginine methylation include better defining the substrate specificity of mammalian PRMT7 and PRMT9(4q31) enzymes and determining whether additional enzymes are encoded by mammalian genomes. As described above, mammals lack genes encoding proteins with amino acid sequences similar to the *S. cerevisiae* Rtm2 and Sfm1 protein arginine methyltransferases [8,52]. It was previously suggested that the mammalian FBXO10 and FBXO11 proteins had PRMT activity [53], but these claims have not been supported by further work [47]. Finally, it is clear that PRMT enzymes function in the cytoplasm and in the nucleus [54]. However, strong evidence for methylation of rat luminal Golgi proteins [55] suggests that one or more of the existing mammalian PRMTs can localize to the Golgi or that at least one novel enzyme is present there.

Histidine methylation

The modification of protein histidine residues by methylation of the *N*-1(π) or *N*-3(τ) atoms of the imidazole ring has been established for a small group of proteins in prokaryotic and eukaryotic cells. These proteins include mammalian actin, myosin heavy chains, myosin light chain kinase, a migration inhibitor factor related protein, and the alpha chain of methyl-coenzyme M reductase (for a review, see [14]). Much of the attention has been focused on the role of the widely conserved *N*-3 methylation of histidine-73 of actin [56]. The recent discovery that histidine 242 in the 'tryptophan finger' of the cytoplasmic yeast ribosomal protein Rpl3 is methylated on the *N*-3 position enabled the

identification of the first enzyme catalyzing this process [14]. Significantly, this residue is buried deeply within the ribosomal 25S RNA; the methylated *N*-3 atom has close contacts with the O6 atom of guanine 878 and the OP1 atom of adenine 876 (Figure 2c). This site is near the ribosomal A-site and the peptidyltransferase center and may be important in the 'rocker switch' coordinating the binding and dissociation of the incoming aminoacyl-tRNA-elongation factor 1A complex [57]. The human ortholog of this enzyme is C1orf156; it is presently unknown whether or not this enzyme is responsible for the modification of any of the known mammalian proteins methylated at the *N*-1 or *N*-3 positions of histidine residues.

Cysteine methylation

Methylation of cysteine residues can occur in enzymes that transfer methyl groups from alkylated DNA in a repair reaction [58], in intermediate steps of catalysis [59], and in automethylation reactions [60]. However, the first example of a protein cysteine methyltransferase with a separate methyl-accepting substrate was the NleE protein of a pathogenic strain of *Escherichia coli* [15]. This enzyme is secreted into the mammalian host and modifies a cysteine residue in a four-cysteine–zinc cluster in the TAB2 and TAB3 adapter proteins involved in NF-κB signaling. Methylation destabilizes this cysteine–zinc cluster and disrupts the signaling pathway that would normally lead to the inflammatory response against the bacterium.

Another example of methylation of cysteine residues in a cysteine–zinc cluster was recently found in the Rps27a protein of the small subunit of the yeast cytoplasmic ribosome [8]. To date, no evidence has been found to suggest the presence of methyltransferase activity that might be responsible for this modification. Indeed, based on the similarity of the methylated cysteine–zinc cluster in Rps27a to one in the N-terminal domain of the *E. coli* Ada protein involved in DNA repair after phosphotriester damage, it has been speculated that the unmodified form of Rps27a may also be involved in DNA repair by serving as an acceptor site for an unwanted DNA methyl group in a non-enzymatic scavenging reaction [8].

Ribosomal protein methylation has now been shown to occur at lysine, arginine, histidine, and cysteine residues. Recent work has demonstrated enzymes catalyzing one additional site of methylation – at the N terminus. This appears to be more widely dispersed in nature.

Protein modification at the N terminus in eukaryotes

N-terminal methylation has long been established for a small group of prokaryotic and eukaryotic proteins, and it was predicted over 25 years ago that the eukaryotic methyltransferase would recognize Xxx-Pro-Lys sequences [61]. This alpha N-terminal protein methyltransferase, designated Ntm1 in yeast and NTMT1 in humans, was finally identified in two recent studies: an analysis of large subunit ribosomal proteins in yeast [12], and analysis of regulator of chromatin condensation 1 (RCC1) in humans [13]. Both studies confirmed the Xxx-Pro-Lys specificity and pointed to the large range of possible new substrates. A recent study has shown even more relaxed substrate specificity for this methyltransferase, but confirmed that the initially identified motif is the preferred substrate [62]. The specificity of this enzyme suggests that a large group of proteins may be methylated by it [12,13,62]. It has been hypothesized that the quaternization of the N-terminal alpha nitrogen atom by methylation results in a fixed positive charge that would enable the N terminus to maintain its charge even in hydrophobic environments. The methylated N terminus may be recognized by specific protein-binding domains. Alternatively, such methylation might interfere with the recognition of other modified residues near the N terminus by protein-binding domains.

Concluding remarks and future directions

Athough it appears that we may have identified many, perhaps most, of the genes encoding protein methyltransferases, we are still only just beginning to establish the range of their physiological targets. Newer approaches using peptide arrays and protein arrays appear to be powerful, especially when combined with the validation of sites *in vivo*. It has also become clear that crosstalk between modification pathways could be a general phenomenon, therefore it might be essential to understand not only the pathways leading to methylation of specific sites on a 'naked' protein but also those leading to methylation of substrates that have been previously modified by acetylation, phosphorylation, or other post-translational modifications. It has also been of interest to see the examples described here where methylated sites on lysine, arginine, and histidine residues of yeast ribosomal proteins are poised to interact with rRNA. Although there are many cases where protein methylation interactions have been shown to facilitate or block protein–protein interactions [38], the knowledge that methylated residues interact with RNA opens new avenues for understanding the functional range of protein methylation.

Acknowledgments

This work was supported by US Public Health Service grant GM026020. I thank Qais Al-Hadid for his help in the preparation of the figures, and my colleagues for their expert advice, including Albert Courey, Alexander Patananan, Jonathan Lowenson, and Lauren Budenholzer. I am especially indebted to Brian Young, Anne McBride, David Weiss, Cecilia Zurita-Lopez, Kristofor Webb, and Tanya Petrossian for their contributions to our laboratory studies reported here.

References

1 Katz, J.E. *et al.* (2003) Automated identification of putative methyltransferases from genomic open reading frames. *Mol. Cell. Proteomics* 2, 525–540
2 Petrossian, T.C. and Clarke, S. (2009) Bioinformatic identification of novel methyltransferases. *Epigenomics* 1, 163–175
3 Wlodarski, T. *et al.* (2011) Comprehensive structural and substrate specificity classification of the *Saccharomyces cerevisiae* methyltransferome. *PLoS ONE* 6, e23168
4 Petrossian, T.C. and Clarke, S.G. (2011) Uncovering the human methyltransferasome. *Mol. Cell. Proteomics* 10, M110.000976
5 Schubert, H.L. *et al.* (2003) Many paths to methyltransfer: a chronicle of convergence. *Trends Biochem. Sci.* 28, 329–335
6 Del Rizzo, P.A. and Trievel, R.C. (2011) Substrate and product specificities of SET domain methyltransferases. *Epigenetics* 6, 1059–1067
7 Tkaczuk, K.L. *et al.* (2007) Structural and evolutionary bioinformatics of the SPOUT superfamily of methyltransferases. *BMC Bioinformatics* 8, 73

8 Young, B.D. *et al.* (2012) Identification of methylated proteins in the yeast small ribosomal subunit: A role for SPOUT methyltransferases in protein arginine methylation. *Biochemistry* 51, 5091–5104
9 Richon, V.M. *et al.* (2011) Chemogenetic analysis of human protein methyltransferases. *Chem. Biol. Drug Des.* 78, 199–210
10 Clarke, S.G. and Tamanoi, F., eds (2006) *The Enzymes Volume 24: Protein Methyltransferases*, Academic Press, (Amsterdam), pp. 3–570
11 Walsh, C.T. (2006) Protein methylation, In *Posttranslational Modifications of Proteins: Expanding Nature's Inventory*, Roberts and Company, (Englewood Colorado), pp. 121–149
12 Webb, K.J. *et al.* (2010) Identification of protein N-terminal methyltransferases in yeast and humans. *Biochemistry* 49, 5225–5235
13 Tooley, C.E. *et al.* (2010) NRMT is an alpha-N-methyltransferase that methylates RCC1 and retinoblastoma protein. *Nature* 466, 1125–1128
14 Webb, K.J. *et al.* (2010) A novel 3-methylhistidine modification of yeast ribosomal protein Rpl3 is dependent upon the YIL110W methyltransferase. *J. Biol. Chem.* 285, 37598–37606
15 Zhang, L. *et al.* (2012) Cysteine methylation disrupts ubiquitin-chain sensing in NF-κB activation. *Nature* 481, 204–210
16 Magnani, R. *et al.* (2010) Calmodulin methyltransferase is an evolutionarily conserved enzyme that trimethylates Lys-115 in calmodulin. *Nat. Commun.* 1, 43
17 Nguyen, A.T. and Zhang, Y. (2011) The diverse functions of Dot1 and H3K79 methylation. *Genes Dev.* 25, 1345–1358
18 Webb, K.J. *et al.* (2011) The ribosomal L1 protuberance in yeast is methylated on a lysine residue catalyzed by a seven-β-strand methyltransferase. *J. Biol. Chem.* 286, 18405–18413
19 Kernstock, S. *et al.* (2012) Lysine methylation of VCP by a member of a novel protein methyltransferase family. *Nat. Commun.* 3, 1038
20 Couture, J.F. *et al.* (2008) Structural origins for the product specificity of SET domain protein methyltransferases. *Proc. Natl. Acad. Sci. U.S.A.* 105, 20659–20664
21 Jenuwein, T. (2006) The epigenetic magic of histone lysine methylation. *FEBS J.* 273, 3121–3135
22 Bannister, A.J. and Kouzarides, T. (2011) Regulation of chromatin by histone modifications. *Cell Res.* 21, 381–395
23 Greer, E.L. and Shi, Y. (2012) Histone methylation: a dynamic mark in health, disease and inheritance. *Nat. Rev. Genet.* 13, 343–357
24 Black, J.C. *et al.* (2012) Histone lysine methylation dynamics: establishment, regulation, and biological impact. *Mol. Cell* 48, 491–507
25 Webb, K.J. *et al.* (2008) Identification of two SET domain proteins required for methylation of lysine residues in yeast ribosomal protein Rpl42ab. *J. Biol. Chem.* 283, 35561–35568
26 Klein, R.R. and Houtz, R.L. (1995) Cloning and developmental expression of pea ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit N-methyltransferase. *Plant Mol. Biol.* 27, 249–261
27 Houtz, R.L. *et al.* (2008) Co- and post-translational modifications in Rubisco: unanswered questions. *J. Exp. Bot.* 59, 1635–1645
28 Allis, C.D. *et al.* (2007) New nomenclature for chromatin-modifying enzymes. *Cell* 131, 633–636
29 Porras-Yakushi, T.R. *et al.* (2006) A novel SET domain methyltransferase in yeast: Rkm2-dependent trimethylation of ribosomal protein L12ab at the N-terminus. *J. Biol. Chem.* 281, 35835–35845
30 Lantham, J.A. *et al.* (2011) Chromatin signaling to kinetochores: Transregulation of Dam1 methylation by histone H2B ubiquitination. *Cell* 146, 709–719
31 Fuchs, S.M. *et al.* (2012) RNA polymerase II carboxyl-terminal domain phosphorylation regulates protein stability of the Set2 methyltransferase and histone H3 di- and trimethylation at lysine 36. *J. Biol. Chem.* 287, 3249–3256
32 Green, E.M. *et al.* (2012) Methylation of H4 lysines 5, 8, and 12 by yeast Set5 calibrates chromatin stress responses. *Nat. Struct. Mol. Biol.* 19, 361–363
33 Porras-Yakushi, T.R. *et al.* (2007) Yeast ribosomal/cytochrome c SET domain methyltransferase subfamily: Identification of Rpl23ab methylation sites and recognition motifs. *J. Biol. Chem.* 283, 12368–12376
34 Huang, J. and Berger, S.L. (2008) The emerging field of dynamic lysine methylation of non-histone proteins. *Curr. Opin. Genet. Dev.* 18, 152–158
35 Couttas, T.A. *et al.* (2012) Methylation of translation-associated proteins in *Saccharomyces cerevisiae*: Identification of methylated lysines and their methyltransferases. *Proteomics* 12, 960–972
36 Polevoda, B. and Sherman, F. (2007) Methylation of proteins involved in translation. *Mol. Microbiol.* 65, 590–606
37 Ben-Shem, A. *et al.* (2011) The structure of the eucaryotic ribosome at 3.0 Å resolution. *Science* 334, 1524–1529
38 Erce, M.A. *et al.* (2012) The methylproteome and the intracellular methylation network. *Proteomics* 12, 564–586
39 Horowitz, S. and Trievel, R.C. (2012) Carbon-oxygen hydrogen bonding in biological structure and function. *J. Biol. Chem.* 287, 41576–41582
40 Iwase, S. *et al.* (2011) ATRX ADD domain links an atypical histone methylation recognition mechanism to human mental-retardation syndrome. *Nat. Struct. Mol. Biol.* 18, 769–777
41 Williamson, N.A. *et al.* (1997) Post-translational processing of rat ribosomal proteins: Ubiquitious methylation of Lys22 within the zinc-finger motif of RL40 (carboxyl-terminal extension protein 52) and tissue-specific methylation of Lys4 in RL29. *Eur. J. Biochem.* 246, 786–793
42 Kudithipudi, S. *et al.* (2012) The SET8 H4K20 protein lysine methyltransferase has a long recognition sequence covering seven amino acid residues. *Biochimie* 94, 2212–2218
43 Dhayalan, A. *et al.* (2011) Specificity analysis-based identification of new methylation targets of the SET7/9 protein lysine methyltransferase. *Chem. Biol.* 18, 111–120
44 Levy, D. *et al.* (2011) A proteomic approach for the identification of novel lysine methyltransferase substrates. *Epigenet. Chromatin* 4, 19
45 Levy, D. *et al.* (2011) Lysine methylation of the NF-kappaB subunit RelA by SETD6 couples activity of the histone methyltransferase GLP at chromatin to tonic repression of NF-kappaB signaling. *Nat. Immunol.* 12, 29–36
46 Chu, Y. *et al.* (2012) Identification and characterization of a highly conserved crenarchael protein lysine methyltransferase with broad substrate specificity. *J. Bacteriol.* 194, 6917–6926
47 Bedford, M.T. and Clarke, S.G. (2009) Protein arginine methylation in mammals: Who, what, and why. *Mol. Cell* 33, 1–13
48 Low, J.K.K. and Wilkins, M.R. (2012) Protein arginine methylation in *Saccharomyces cerevisiae*. *FEBS J.* 279, 4423–4443
49 Yang, Y. and Bedford, M.T. (2012) Protein arginine methyltransferases and cancer. *Nat. Rev. Cancer* 13, 37–50
50 Zurita-Lopez, C.I. *et al.* (2012) Human protein arginine methyltransferase 7 (PRMT7) is a type III enzyme forming ω-N^G-monomethylated arginine residues. *J. Biol. Chem.* 287, 7859–7870
51 Lee, J. *et al.* (2005) PRMT8, a new membrane-bound tissue-specific member of the protein arginine methyltransferase family. *J. Biol. Chem.* 280, 32890–32896
52 Chern, M.K. *et al.* (2002) Yeast ribosomal protein L12 is a substrate of protein-arginine methyltransferase 2. *J. Biol. Chem.* 277, 15343–15353
53 Krause, C.D. *et al.* (2007) Protein arginine methyltransferases: Evolution and assessment of their pharmacological and therapeutic potential. *Pharmacol. Ther.* 113, 50–87
54 Herrmann, F. *et al.* (2009) Human protein arginine methyltransferases in vivo – distinct properties of eight canonical members of the PRMT family. *J. Cell Sci.* 122, 667–677
55 Wu, C.C. *et al.* (2004) Organellar proteomics reveals Golgi arginine dimethylation. *Mol. Biol. Cell* 15, 2907–2919 Correction *Mol. Biol. Cell* 15 (11)
56 Nyman, T. *et al.* (2002) The role of MeH73 in actin polymerization and ATP hydrolysis. *J. Mol. Biol.* 317, 577–589
57 Meskauskas, A. and Dinman, J.D. (2010) A molecular clamp ensures allosteric coordination of peptidyltransfer and ligand binding to the ribosomal A-site. *Nucleic Acids Res.* 38, 7800–7813
58 Sedgwick, B. *et al.* (2007) Repair of alkylated DNA: recent advances. *DNA Repair (Amst.)* 6, 429–442
59 Boal, A.K. *et al.* (2011) Structural basis for methyl transfer by a radical SAM enzyme. *Science* 332, 1089–1092
60 Siddique, A.N. *et al.* (2011) Auto-methylation of the mouse DNA-(cytosine C5)-methyltransferase Dnmt3a at its active site cysteine residue. *FEBS J.* 278, 2055–2063
61 Stock, A. *et al.* (1987) N-terminal methylation of proteins: structure, function, and specificity. *FEBS Lett.* 220, 8–14

62 Petkowski, J.J. *et al.* (2012) Substrate specificity of mammalian N-terminal alpha-amino methyltransferase. *Biochemistry* 51, 5942–5950
63 Lipson, R.S. *et al.* (2010) Two novel methyltransferases acting upon eukaryotic elongation factor 1A in *Saccharomyces cerevisiae*. *Arch. Biochem. Biophys.* 500, 137–143
64 Polevoda, B. *et al.* (2000) Cytochrome c methyltransferase, Ctm1p, of yeast. *J. Biol. Chem.* 275, 20508–20513
65 Jackson, C.A. *et al.* (2012) Proteomic analysis of interactors for yeast protein arginine methyltransferase Hmt1 reveals novel substrate and insights into additional biological roles. *Proteomics* 12, 3304–3314
66 Sayegh, J. and Clarke, S.G. (2008) Hsl7 is a substrate-specific type II protein arginine methyltransferase in yeast. *Biochem. Biophys. Res. Commun.* 372, 811–815
67 Polevoda, B. *et al.* (2006) The yeast translation release factors Mrf1p and Sup45p (eRF1) are methylated, respectively, by the methyltransferases Mtq1p and Mtq2p. *J. Biol. Chem.* 281, 2562–2571
68 Leulliott, N. *et al.* (2004) Structure of protein phosphatase methyltransferase 1 (PPM1), a leucine carboxyl methyltransferase involved in the regulation of protein phosphatase 2A activity. *J. Biol. Chem.* 279, 8351–8358
69 Hahne, K. *et al.* (2012) Evaluation of substrate and inhibitor binding to yeast and human isoprenylcysteine carboxyl methyltransferases (Icmts) using biotinylated benzophenone-containing photoaffinity probes. *Biochem. Biophys. Res. Commun.* 423, 98–103
70 Lee, J.M. *et al.* (2012) EZH2 generates a methyl degron that is recognized by the DCAF1/DDB1/CUL4 E3 ubiquitin ligase complex. *Mol. Cell* 48, 572–586
71 He, A. *et al.* (2012) PRC2 directly methylates GATA4 and represses its transcriptional activity. *Genes Dev.* 26, 37–42
72 Tate, C.M. *et al.* (2010) CXXC finger protein 1 restricts the Setd1a histone H3K4 methyltransferase complex to euchromatin. *FEBS J.* 277, 210–223
73 Cho, H.S. *et al.* (2012) Enhanced HSP70 lysine methylation promotes proliferation of cancer cells through activation of Aurora kinase B. *Nat. Commun.* 3, 1072
74 Shinkai, Y. and Tachibana, M. (2011) H3K9 methyltransferase G9a and the related molecule GLP. *Genes Dev.* 25, 781–788
75 Rathert, P. *et al.* (2008) Protein lysine methyltransferase G9a acts on non-histone targets. *Nat. Chem. Biol.* 4, 344–346
76 Huang, J. *et al.* (2010) G9a and Glp methylate lysine 373 in the tumor suppressor p53. *J. Biol. Chem.* 285, 9636–9641
77 Ling, B.M.T. *et al.* (2012) Lysine methyltransferase G9a methylates the transcription factor MyoD and regulates skeletal muscle differentiation. *Proc. Natl. Acad. Sci. U.S.A.* 109, 841–846
78 Pagans, S. *et al.* (2010) Characterization of HIV Tat modifications using novel methyl-lysine-specific antibodies. *Methods* 53, 91–96
79 Van Aller, G.S. (2012) Smyd3 regulates cancer cell phenotypes and catalyzes histone H4 lysine 5 methylation. *Epigenetics* 7, 340–343
80 Kunizaki, M. *et al.* (2007) The lysine 831 of vascular endothelial growth factor receptor 1 is a novel target of methylation by SMYD3. *Cancer Res.* 67, 10759–10765
81 Binda, A. *et al.* (2013) SETD6 monomethylates H2AZ on lysine 7 and is required for the maintenance of embryonic stem cell self-renewal. *Epigenetics* 8, 1–7
82 Shi, X. *et al.* (2007) Modulation of p53 function by SET8-mediated p53 methylation at lysine 382. *Mol. Cell* 27, 636–646
83 Calnan, D.R. *et al.* (2012) Methylation by Set9 modulates FoxO3 stability and transcriptional activity. *Aging* 4, 462–479
84 Ferguson, A.D. *et al.* (2011) Structural basis of substrate methylation and inhibition of SMYD2. *Structure* 19, 1262–1273
85 Abu-Farna *et al.* (2011) Proteomic analyses of the SMYD family interactomes identify HSP90 as a novel target for SMYD2. *J. Mol. Cell Biol.* 3, 301–308
86 Cho, H.S. *et al.* (2012) RB1 methylation by SMYD2 enhances cell cycle progression through an increase of RB1 phosphorylation. *Neoplasia* 14, 476–486