



Published in final edited form as:

Epigenomics. 2009 October 1; 1(1): 163–175. doi:10.2217/epi.09.3.

Bioinformatic Identification of Novel Methyltransferases

Tanya Petrossian and Steven Clarke

Department of Chemistry and Biochemistry and the Molecular Biology Institute, UCLA, Los Angeles, CA 90095-1570

Summary

Methylation of DNA, protein, and even RNA species are integral processes in epigenesis. Enzymes that catalyze these reactions using the donor *S*-adenosylmethionine fall into several structurally distinct classes. The members in each class share sequence similarity that can be used to identify additional methyltransferases. Here, we characterize these classes and *in silico* approaches to infer protein function. Computational methods such as hidden Markov model profiling and the Multiple Motif Scanning program can be used to analyze known methyltransferases and relay information into the prediction of new ones. In some cases, the substrate of methylation can be inferred from hidden Markov model sequence similarity networks. Functional identification of these candidate species is much more difficult; we discuss one biochemical approach.

Keywords

methyltransferases; protein methylation; RNA methylation; *S*-adenosylmethionine; SET domain; SPOUT domain; HMM profiling

Methyltransferases and Epigenomics

Recent years have witnessed greatly expanded attention to the enzymes that catalyze the transfer of methyl groups from *S*-adenosylmethionine (AdoMet) to DNA, RNA, proteins, lipids, and small molecules [1]. The central role of methyltransferases in epigenetics was first realized with enzymes modifying DNA [2,3]; subsequent work has demonstrated the importance of a number of enzymes that modify lysine and arginine residues in histones [4-7]. RNA methyltransferases also appear to play a role. For example, microRNA species can be modified by methyltransferases such as HEN1 to affect DNA methylation in paramutation [8-11]. It is likely that additional methyltransferases for protein, RNA, DNA, and perhaps even lipids or small molecules, may be involved in epigenetic phenomena. For the human proteome, only a fraction of the potential methyltransferases has been functionally identified and enzymes of importance to epigenetics may be lurking among the unknown species. Thus, it is of interest to be able to identify the complete “cast” of methyltransferases and their substrates in the proteomes of various organisms.

In this paper, we review recent progress on the identification and characterization of new methyltransferases. We have focused our discussion largely on the situation in the budding yeast *Saccharomyces cerevisiae*, where the “methyltransferasome” has been most fully characterized [12,13] (Table I). The successful identification of yeast methyltransferases will hopefully help pave the way to the identification of new enzymes in higher plants, mammals, and other organisms.

We will first introduce the different methyltransferase families and what bioinformatic methods have helped reveal about them, along with the limitations of these methods. We will then describe one biochemical approach to determining the function of candidate methyltransferases identified using bioinformatics methods.

Structural Classes of Methyltransferases: Approaches to Bioinformatic Identification of New Enzyme Species

The success of the identification of novel methyltransferases using bioinformatics methods ultimately lies on the information known about previously discovered methyltransferases. Each topologically-distinct family of methyltransferases is described along with computational methods for the identification and characterization of new family members. These topographical classes have been identified in references [1], [14], and [15].

The Broad Swath of Classic Seven Beta Strand Methyltransferases

Seven beta strand enzymes (also referred to as “Class I” methyltransferases) appear to make up the majority of methyltransferases in organisms [14,16]. This group includes the mammalian *de novo* and maintenance DNA methyltransferases [3,17,18], the Dot1 histone lysine methyltransferase [19], and the HEN1 microRNA methyltransferase [8,10], all known enzymes that play roles in epigenesis. Remarkably, sequence similarity is shared between methyltransferases ranging from the *Saccharomyces cerevisiae* enzyme active on small molecules (Tmt1), the *Mycoplasma arthritidis* enzyme active on DNA (HhaI), the *Arabidopsis thaliana* enzyme active on lipids (UbiE), the human enzyme active on protein (PCMT1), to even the *Bos taurus* enzyme active on inorganic arsenite (AS3MT). Despite vastly different substrates of methylation, primary sequence similarity was found in small regions of these proteins before any structural information was available [20].

In Fig. 1a, we give the histone lysine methyltransferase Dot1 as an example of this class of enzyme [19]. Enzymes in this class of methyltransferases share a common seven strand twisted beta sheet with a C-terminal beta hairpin, sandwiched between alpha helices [15,16]. Four signature motifs are present (I, Post I, II, and III; [13,21]). Residues of Motif I and Motif Post I contact AdoMet. The conserved aspartate amino acids in these motifs are key in stabilizing charged AdoMet species as well as hydrogen bonding to two different locations of the cofactor for positioning the methyl group to transfer. The last residues of $\beta 4$ and $\beta 5$, which make up portions of Motifs II and III, respectively, form part of the catalytic domain and can bind the methyl-accepting substrate [13]. A few enzymes in this methyltransferase class deviate from this structural core, most notably the protein arginine methyltransferase PRMT1 that lacks $\beta 6$ and $\beta 7$ [15,16] and the circularly permuted motifs in plant DRM enzymes [18].

Some proteins in this superfamily contain conserved sequences between Motifs II and III that are methyl-acceptor substrate specific. For instance, the “DPPY” motif is seen in several *N*-methyltransferases active on sp^2 -hybridized nitrogen atoms in adenosine or glutamine residues [15,22], while the “EE” motif is present in protein arginine methyltransferases [23]. Inserts and deletions to the core structure have also been found to reflect substrate identity [16]. Yeast histone H3K79 methyltransferase Dot1, originally discovered for its role in telomeric silencing [24], has several basic residues in the N-terminal domain which bind nucleosomes [19,25]; the same stretch is seen in the C-terminal domain of human Dot1 [26]. To date, Dot1 is the only non-SET histone lysine methyltransferase (see below) and interestingly the only histone lysine methyltransferase which methylates in the globular domain of histones [27].

Initially, *in silico* searches for novel methyltransferases were performed using known methyltransferase sequences as probes against protein databases with BLAST. The discovery

of Hmt1/Rmt1 protein arginine methyltransferase is an example of the success from this approach [28].

The shift from whole sequence comparisons to motif-based searches has led to the generation of a comprehensive list of putative seven beta strand (Class I) methyltransferases [21,29]. Katz *et al.* used MEME [30] to build position-based amino acid frequency matrices, or profiles, of Motifs I and Post I [31] from multiple alignments of known methyltransferases and utilized these profiles in a comprehensive MAST [32] search of the genome [31]. As a result, the search is based on information from multiple methyltransferases rather than simply amino acids similarity (as in the 20×20 matrix “BLOSUM 62” used in BLAST searching).

Methyltransferase domain identification was further refined by Ansari *et al.*, who aligned sequences through additional secondary structure information [33]. In their database search, the authors used hidden Markov model (HMM) profiles that take into account not only the log-odds amino acids frequency but also the frequency of inserts and deletions to account for gaps in the alignment. HMM profiles can be created from large superfamily reference sets (such as all Class I methyltransferases) to identify a general list of proteins, or alternatively can be generated from a specific subclass of proteins to restrict the search. Ansari *et al.* specifically identified O-, N- and C- methyltransferases from a non-redundant database using HMM profiles from the methyltransferase domain of polyketide synthase (PKS) and nonribosomal peptide synthetase (NRPS) [33]. However, such global search profiles spanning the entire methyltransferases domain assign penalties for mismatches between motifs that may leave true previously unidentified methyltransferases undetected.

More recently, a new approach involving motif-based searches along with HMM profile-profile local alignments were used to solve some of the computational hurdles of the past [13]. Independent matrices describing all of the motifs, including II and III, were developed through better motif identification from either solved methyltransferase structures or from HMM profile primary and secondary structure prediction alignments [13] using the program HHpred [34]. Additionally, a novel program, Multiple Motif Scanning (MMS) [101], was used to rank the yeast database of proteins to the sequence similarity of the methyltransferase motif profiles. Here, the position-based matrices were entered into MMS which includes a parameter for the conserved number of amino acids between the motifs and outputs the overall highest scoring combinations of multiple best-fit motifs [13]. The success of this program relies on the input matrices; using matrices derived from different methyltransferase reference sets output slightly different rankings and putative methyltransferases. MMS is advantageous for proteins containing multiple ungapped motifs such as methyltransferases because it does not allow for inserts within a motif as do HMM profiles. For example, when we use HHpred to search against the yeast proteome with HMM profiles constructed from the identical motif sequences employed in Ref. 13 we did not detect a number of putative yeast methyltransferases (YMR209C, YLR063W, YKL155C and YNL092W) that were identified using MMS (see Table I). On the other hand, as shown in Table I, HHpred did find YKL162C and YLR137W that were overlooked by MMS and were not reported previously [13]. Together, these results highlight the importance of combining these methodologies to create a comprehensive list of putative methyltransferases.

To determine the potential biological function of the candidate methyltransferases, identification of the methyl-accepting substrate would be valuable. Bioinformatic approaches for identifying substrates has so far resulted in a mixed record of success. A widespread approach has been comparing protein sequences across species to reveal homologs. In fact, databases of phylomes such as PhylomeDB contain compilations of phylogenetic trees which can be used to assess the orthological and paralogical relationships of a given protein [35]. More recently, similarity sequence networks have been used as a high-throughput method for substrate predictions [36]. Sequence similarities in or outside of methyltransferase motifs that

reflect substrate recognition allow for enzymes acting on similar substrates to cluster with each other. HMM profile-profile comparisons with a stringent E value as cutoff ($<10^{-20}$) separated yeast methyltransferases into protein arginine methyltransferases, protein glutamine methyltransferases, wybutosine-forming transferases, 2'-O-ribose methyltransferases, cytosine 5-methyltransferases, and small molecule/lipid methyltransferase clusters [13]. Although not all yeast methyltransferases clustered in this protocol, one can gain useful information when a putative methyltransferase was found grouped with enzymes of known catalytic activity.

A Different SET of Protein Lysine Methyltransferases

Sequence similarity between the plant Rubisco protein lysine methyltransferase and three *Drosophila* proteins involved in epigenetics - Suppressor of position-effect variegation 3-9 (Su (var) 3-9), Enhancer of zeste (E(z)), and Trithorax (Trx) - led to the discovery of the family of **SET** methyltransferases [37]. This family includes a number of histone lysine methyltransferases involved in transcriptional control through chromatin structural modification [38]. These proteins contain the SET domain consisting of eight curved beta strands arranged into three sheets and a characteristic pseudoknot structure. An example of this domain is shown for the MLL1 histone lysine methyltransferase in Fig. 1b [39]. The SET proteins share sequence similarity in the N-terminal (N-SET) and C-terminal domains (C-SET) that contain residues responsible for catalysis, cofactor-binding, and substrate interaction. The first two motifs reside in N-SET while the last two motifs lie in C-SET and form the knot-like structure (Fig. 1b) [26,40,41]. AdoMet binds to Motif I, N-terminal residues of motif III, and tyrosine in Motif IV [26]. Interestingly, the "GxG" sequence of Motif I interacts with AdoMet as does the seven beta strand Motif I "GxGxG" sequence despite the lack of any overall structural similarity between SET and seven-beta strand methyltransferases [42]. The catalytic site is located on the opposite side of the enzyme and includes a key catalytic tyrosine residue in Motif II. Interactions with the lysine substrate occur in the hydrophobic pocket formed by the remaining portions of Motifs III and IV [26,40]. It has been hypothesized that variability within this domain defines the substrate and in fact, residues in C-SET have been integral in the determining mono-, di-, or trimethylation of the enzyme. Point mutations in the "Y/F switch" have proved successful in converting SET7/9 to a di-/tri-methyltransferases [43], SET8 to a dimethyltransferase [44], and Dim5 to a mono-methyltransferase [45].

Amino acid sequences between the N-SET and C-SET domains are highly variable among the SET superfamily and have been dubbed the I-SET region. I-SET residues can interact with the substrate [46,47]. In fact, several non-histone methyltransferases have a "SRA" motif in the I-SET domain [48]. I-SET is not always indicative of the binding ligand; two pairs of enzymes - SUV39H1 and SETDB1, and SET7/9 and MLL - have non-homologous I-SET despite sharing identical substrates [46,47]. Many SET proteins also have a Pre-SET and Post-SET domain composed of several conserved cysteine residues that coordinate zinc ions in triangular clusters. The function of these domains is not clear, although Post-SET seems to shape the channel for the lysine substrate. In enzymes that lack the cysteine-rich Post-SET, such as SET7/9 and Rubisco, additional alpha helices are oriented to create this channel [26,40]. Several SET methyltransferases also have additional domains such as PWWP, PHD, and SANT that appear to recruit the chromatin substrate [49].

Yeast candidate SET-domain methyltransferases, including YHL039W and YBR030W (now Rkm3), were identified by Porras-Yakushi *et al.* through reiterative PSI- and PHI-BLAST searches [48]. However, the inherent nature of BLAST searching does not lead to a single list of SET methyltransferases; instead, two self-contained "subfamilies" of proteins were found [48]. When we now search with HHpred using SET protein sequences compiled from the SMART 6 database [50], we find that we can produce a single list of all of these

methyltransferases proteins (Table I). Additionally, profiles obtained from the same reference dataset using only MEME-derived matrices of Motifs I-IV in MMS also identified all of these proteins, confirming their identification through the SET domain (data not shown).

The “subfamilies” described by Porras-Yakushi [48] ultimately differentiated between what has been described as Class I-VI histone and Class VII non-histone methyltransferases based on their substrate specificity (Table II). Initially, four classes of SET proteins were discovered through BLASTP searches and ClustalW clustering analysis of Arabidopsis genome using *Drosophila* genes E(z) (Class I, H3K27), Ash1 (Class II, H3K36), Trx (Class III, H3K4), Su (var)3-9 (Class V, H3K9) [51]. Springer *et al.* expanded this analysis to include other genomes and with an updated SET protein list [52]. These authors identified Class IV of proteins that contain Phd finger but lack Pre-SET and Post-SET, and several proteins whose I-SET domain was extended, dubbed the disrupted S-ET proteins. The S-ET proteins were later divided into two classes: Class VI histone and Class VII nonhistone proteins that includes Rubisco, cytochrome c, and ribosomal proteins [52,53]. This classification of SET proteins based on the families or substrates of methylation may need to be expanded in the future upon the discovery of new methylation sites and SET proteins. In fact, methylation on substrates H1K26 and H4K20 have been recently discovered in mammalian cells [46].

We have now created individual HMM profiles of each SET Class from the reference set of proteins in Ref. ⁵³ and performed an HHpred search against the complete yeast protein database (Table II). Every one of the twelve yeast SET methyltransferases fit into its appropriate class: H3K36-methylating Set2 in Class II, H3K4-methylating Set1 in Class III, Set3 and Set4 (which contain PHD domains) in Class IV, the interrupted domain Set5 and Set6 in Class VI, and lastly ribosomal methyltransferases Rkm1-4, YHL039W and Ctm1 in Class VII (Table II). Interestingly, the substrates of yeast Set3, Set4, Set5, and Set6 proteins are not known. It appears likely that these will be histone lysine methyltransferases but it will be important to confirm this tentative identification experimentally.

SPOUTing Additional RNA Methyltransferases

The SPOUT methyltransferase family was first described based on the primary sequence and predicted secondary structural similarities of bacterial SpoU and TrmD methyltransferases [55]. This new topology of methyltransferases became apparent with the solved structures of RrmA and RlmB [56,57] revealing a characteristic knot distinct from SET methyltransferases. To date, SPOUT methyltransferases have been found to exclusively methylate RNA. Members of this family may thus possibly methylate RNA species involved in epigenesis. The core structure consists of a beta sheet with five parallel beta strands in a 5-3-4-1-2 orientation between two layers of helices. An example of this structure is given for TrmH in Fig. 1c [58]. A partial Rossmann-like fold similar to that in the seven beta strand (Class I) methyltransferases is formed by the first two N-terminal strands; variability can exist with additional alpha/beta units in this region. Unlike the seven beta strand (Class I) enzymes, AdoMet binds to the C-terminal alpha-beta “trefoil” knot that characterizes the SPOUT superfamily [12].

Primary sequence similarity is not very strong among the members of this superfamily that is largely defined by its tertiary structure [12,58,59]. Nonetheless, common motifs have been described. Motif 1 is not widely conserved among all subclasses of SPOUT methyltransferases but contains amino acids integral for tRNA binding, the release of AdoHcy, and catalysis [59]. The latter residues of β 3 bind AdoMet (here termed Motif Post 1; Fig. 1c). Although the topology of SPOUT methyltransferases is unique from the seven beta strand (Class I) and SET enzymes, Motif 2 of the SPOUT domain has several shared residues with both of these classes: the glycine rich coil preceding β 4 binds both the tRNA substrate along with AdoMet and the catalytic glutamyl residue is catalytic much like the asparagine/aspartate in the seven beta strand (Class I) β 4 and the asparagine in the SET Motif III [12,42]. Motif 3, originally described

as the coil preceding $\beta 5$ [59], can be expanded to include an extended helix with a catalytic tyrosine, and is involved in AdoMet-binding and catalysis [12] (Fig 1c). The active site is created upon dimerization, and additional catalytic residues for SPOUT methyltransferases are family specific and lie on the antiparallel or perpendicular mode of dimerization. Like the SET superfamily, several SPOUT methyltransferases have additional domains flanking the SPOUT domain including, not surprisingly, THUMP, OB-fold, L30e, and PUA domains that are associated with nucleic acid binding or modification [12].

Tkaczuk *et al.* have also used similar computational techniques to identify new SPOUT methyltransferases [12]. Crystal structures of known SPOUT methyltransferases were collected and were used to search the PDB with DALI to find proteins with similar structures [12]. PSI-BLAST searches using different members of COG families were performed on a non-redundant database to discover previously unidentified putative SPOUT methyltransferases, which were corroborated by secondary structural predictions [12]. HMM profiles of aligned sequences were created and searched by HHpred to identify as many protein families with even remote similarities to the SPOUT domain, where proteins were further validated by reciprocal searches and fold-recognition methods [12]. These methods identified known yeast methyltransferases Trm10, Mrm1, Trm3 as well as putative methyltransferases Emg1, YGR283C, YMR310C. The crystal structure of Emg1 later confirmed these predictions [60,61]. We have also used these methods to predict YOR021C as an additional yeast putative SPOUT methyltransferase (Table I).

The pairwise PSI-BLAST searches performed by Tkaczuk *et al.* revealed a core “supercluster” of five COG families along with four satellite clusters that are all 2'-O- methyltransferases [12]. Therefore, proteins such as *Escherichia coli* YibK, LasT, and YfiF were predicted to be 2'-O-ribose methyltransferases [12]. Interestingly, we find that yeast Tan1, currently annotated as a putative tRNA acetyltransferase, has high similarity by HHpred to the one of these satellite clusters (COG1818; $e = 1.6^{-20}$, $p = 2.8^{-24}$), indicating that it may be a 2'-O-ribose methyltransferase as well (Table I). Additionally, enzymes responsible for m1G and m3U methylation form independent clusters which were distinct from the other COG groups [12]. These analyses may thus reveal the substrate specificity of a putative methyltransferase.

Hitting Three Methyltransferase Superfamilies with One Stone

Although most methyltransferases are found in the seven beta strand, SET domain, and SPOUT families, there are, however, a number of these enzymes that have other types of structural folds. Interestingly, the crystal structure of a single enzyme, cobalamin-dependent methionine synthase (MetH), has given insight into three additional distinct classes of AdoMet-binding methyltransferases [62]. This enzyme uses the methyl group of N-5-methyl-tetrahydrofolate to produce methionine from homocysteine through a methylcob(III)alamin intermediate. These classes include the MetH-reactivation domain, the homocysteine methyltransferases, and radical SAM methyltransferases .

AdoMet binds to the reactivation domain; its methyl group is then transferred to the oxidized B12 cofactor on a separate domain [62]. The unique arrangement of this AdoMet-binding domain can be best described as a twisted center beta strand surrounded by several shorter antiparallel beta-strands forming two perpendicular sheets (Fig. 1d) [63]. AdoMet binds to the helices and coils in the middle of this C-shaped structure, specifically the acidic residue of $\alpha 2$, RLAEAF in $\alpha 6$, the RPAPG coil following $\alpha 7$, and a C-terminal aromatic residue [63]. Interestingly, we find that the AdoMet-binding domain of MetH does not show homology to any protein in yeast by sequence analysis using HHpred (Table I). It is presently unclear whether this domain architecture is utilized in any other methyltransferase reactions; although we did not find any homologs by fold recognition programs utilizing automated modeling (MODELLER [64]) or threading approaches (PHYRE [65]) (Table I).

The second methyltransferase domain illuminated by methionine synthase is the homocysteine-binding domain. Our HHpred searches using this N-terminal domain of MetH as a probe against the yeast protein database detects the yeast homocysteine methyltransferase family proteins Mht1 and Sam4 with very high sequence similarity (Table 1). Additional searches with the homocysteine COG group against the yeast proteome confirms this observation (Table I). Mht1 and Sam4 catalyze the same homocysteine to methionine reaction as MetH but utilize AdoMet or *S*-methylmethionine as methyl donors [66]. The similarity in sequences of these enzymes suggest a similarity in overall structures as well. The homocysteine-binding domain of MetH is composed of a beta-barrel from eight parallel strands (Fig. 1e) [67]. A zinc ion is also bound to the structure and functions in MetH to draw the cobalamin closer to the catalytic domain as well as activate the thiol for nucleophilic attack. The metal coordinates with tetrahedral geometry with three cysteines following $\beta 6$ (GXNC) and $\beta 8$ (GGCC) with the last binding partner being either substrate homocysteine or a nitrogen/oxygen containing side-chain residue of $\beta 7$ (N in the case of MetH) [67,68]. Interestingly, the latter half of this domain, is homologous to YMR321C. It is unclear whether YMR321C is a putative methyltransferase; the AdoMet-binding domain of these proteins remains to be determined.

Finally, the cobalamin-binding domain of MetH is often present in proteins that also include the “radical SAM” domain [69]. Radical SAM enzymes generally form methionine and the deoxyadenosine radical from AdoMet, where crystal structure determinations have demonstrated a TIM barrel domain (Fig 1f) [70]. These proteins are distinguished by their CxxxCxxC motif, which is used to bind an iron sulfur cluster necessary for radical generation. Although many of these family members catalyze non-methyltransferase reactions (typically involving the deoxyadenosyl radical formed by a one electron transfer to AdoMet), there are at least several members that are known to participate in methylation reactions despite the fact that the mechanisms of these transfers are still unclear [71,72]. These include the florfenicol/chloramphenicol resistance protein (Cfr), the fortimicin methyltransferase (fmrO), and the fosfomycin methyltransferase (Fom3) [72]. Radical SAM methyltransferases are difficult to distinguish from other radical SAM enzymes by sequence analysis. This was highlighted by our HHpred searches against the yeast proteome using multiple alignments of radical SAM methyltransferases found in the RefSeq database [73] (Table I). This search identifies apparent non-methyltransferases including the Bio2 biotin synthase, the C-terminal portion of Elp3 – a histone acetyltransferase thought to also be involved in histone demethylation initiated by 5'-deoxyadenosyl radical, Lip5 involved in biosynthesis of the lipoic acid, and Tyw1 in the wybutosine pathway (Table 1). Further work will be needed to ask if there are additional methyltransferases in the radical SAM family in other organisms.

Another Boost from B12

Cobalamin is a link to yet another structurally distinct methyltransferase, this time not as a partner in methylation but as a necessity in its own biosynthesis. The structure of CbiF, a precorrin-4-C11 methyltransferase, revealed two asymmetric domains of a five beta-strand, four-helix structure – the first domain containing strands in parallel while the second are in antiparallel orientation (Fig.1g) [74]. The GxGxG motif at the end of $\beta 1$ residues on the beta sheet does not bind AdoMet in the absence of precorrin [15]. Unlike the other classes of methyltransferases, AdoMet is distorted to 82° between the two beta sheets; it is thought that this orientation is favorable for the transfer of the methyl group to the bulky precorrin substrate in the active site [74]. However, other less bulky substrates are methylated by this superfamily of enzymes. Dph5, involved in diphthamide biosynthesis, shows sequence similarity to CbiF (Table I), yet interestingly does not share the six amino acids known to bind the substrate precorrin-4 methyltransferase in CbiF [74].

Two Additional Distinct Classes of Methyltransferases – Structures to be Determined

Although their three-dimensional structures are currently unknown, membrane-bound methyltransferases share no sequence homology to structurally solved methyltransferases. Biochemical studies of the isoprenylcysteine carboxyl methyltransferases Ste14 has led to a topology model describing its structure as six membrane spans, with two forming helical hairpin [75]. The conserved region A contains motif RHPxYxG that is trailed by a hydrophobic stretch ending in two conserved adjacent glutamates in region B. This C-terminal domain, where five of six point mutations lead to a loss-of-function, is conserved not only in isoprenylcysteine carboxyl methyltransferases but also phospholipid methyltransferases. Interestingly, searches of Ste14 through BLAST [75] and HHpred yield yeast phospholipids methyltransferases Opi3 and Cho2 as well as several fatty acid/steroid reductases and the C-terminal residues of ergosterol biosynthetic enzymes Erg4p and Erg24p. However, when we searched the database using multiple alignments of the proteins in the PEMT family present in the Pfam database [76], this list only included Opi3, Ste14, and Cho2 (Table 1).

Evidence has been presented for a final class of methyltransferases represented by enzymes that modify the N-6 position of adenosine in mRNA [77]. The yeast Ime4 protein appears to be in this group. Weak sequence similarity is found with the “DPPY” motif between Motifs II and III of some Class I seven beta strand N-methyltransferases. Our HMM analysis using the Ime4 protein sequence as probe against the yeast database indicates that this family includes members Kar4 and YGR001C (Table 1). Further work will be needed to confirm the methyltransferase activity of these proteins.

Biochemical Identification of Putative Methyltransferases

Although computational methods are very powerful, the functional identification of a methyltransferases can only be made with biochemical evidence. In some cases, bioinformatic approaches can suggest one or more specific functions that can be specifically tested. Often, however, this is not the case. Thus, general biochemical approaches that can at least confirm the binding of AdoMet become more important. A useful procedure here is to take advantage of the fact that many methyltransferases can covalently be linked to [³H-methyl]AdoMet after UV treatment and detected as a crosslinked product on SDS gels [78]. An example of this is shown in Fig. 2, where the crosslinked proteins are detected by fluorography. Here AdoMet-binding was confirmed for the yeast YHR209W protein. It is also possible to cut out the Coomassie-stained band, dissolve the gel with hydrogen peroxide, and directly measure the radioactivity associated with the protein [79].

This method could be adapted to include high throughput methods by separating proteins from cell extracts, cross-linking to AdoMet, and separating by two-dimensional gel electrophoresis for identification of radioactive species [80]. Advancements in proteome chip technologies can lead to the identification of methyltransferases by these crosslinking methods if sensitivity and resolution of tritium fluorography can be achieved [81]. Recently, a number of new approaches in this area have been described [82].

Not all enzymes that bind and catalyze reactions with the cofactor AdoMet or its derivatives are methyltransferases. AdoMet or its decarboxylated derivative can also be used in reactions of adenosyltransfer, formation of the deoxyadenosyl radicals, aminotransfer, aminobutryltransfer, and aminopropyltransfer [14,16]. For example, two clear yeast seven beta strand Class I “methyltransferases” are actually aminopropyltransferases - spermidine synthase (Spe3) and spermine synthase (Spe4).

Clearly, the crucial indicator of methyltransferase function is the sure identification of the methyl-accepting substrates and products. However, such identification can be difficult

because most enzymes are very specific and substrates can be unique to each methyltransferase reaction. Some clues can be surmised from mutant phenotypes. However, many knockouts have either no phenotype or ones that are not readily interpreted in terms of specific methylation events. There is a large amount of high-throughput data in yeast, including localization and expression profiles; only in rare cases has this information been useful to date in identifying new methyltransferases. In fact, it is often hard to rationalize the data available for known methyltransferases. In the end, there is probably no substitute for direct biochemical assays of methylation!

Future perspective

It is of course hard to predict how this field will evolve in the next five to ten years. However, at the present rate of discovery, it seems likely that we will know the function of most of the methyltransferases of yeast and a good fraction of the human enzymes in this time frame. From the success of the bioinformatic approaches, the rate-limiting step in fully characterizing the biological function of the candidate methyltransferases is soon likely to be biochemical analysis. One question is whether high-throughput approaches will achieve more success than they have to date. Unfortunately, the information content derived from these approaches, even in species such as yeast that have been intensively studied, has been limited and has generally not permitted assignments of functions to candidate methyltransferases. However, advances here may allow identification of these roles in the next few years; alternatively the tried and true biochemical approaches on individual or small groups of proteins may remain the best approach.

Executive summary

bulleted summary points that illustrate the main topics or conclusions made under each of the main headings of the article

Methyltransferases and epigenomics

- Methyltransferases modify a variety of substrates and fall into several topologically distinct structures
- Even though only enzymes from the Class I seven beta strand methyltransferases and SET methyltransferases families are known to be involved in epigenetics to date by methylation of histone protein, DNA, and even miRNA, methyltransferase species from other families may be involved in epigenetics as well. Interestingly, at least one histone acetyltransferase and putative demethylase Elp5 falls into the radical SAM enzyme family.

Structural Classes of Methyltransferases: Approaches to Bioinformatic Identification of New Enzyme Species

- Each structural class of methyltransferases is discussed and analyzed for the development of a comprehensive list of methyltransferasome
- Analysis of sequence and structural information of known methyltransferases allows one to build a profile of the methyltransferases family to find putative methyltransferases computationally
- Utilizing information derived from a profile of proteins (HMM profiles or MEME) versus a single sequence exponentially enhances the sensitivity to discover new enzymes

- Additionally, advanced computational programs such as HHpred and Multiple Motif Scanning increases the searching power from previous methods of BLAST
- Unique sequence characteristics based on substrate specificity allow for the creation of sequence similarity networks to predict substrates of methylation

Biochemical Identification of Putative Methyltransferases

- Functional identification of candidate methyltransferases can only be completed with biochemical verification
- The list of putative methyltransferases with predicted substrates of methylation from bioinformatic analysis makes the biochemical identification of the methyltransferasome a feasible task
- High-throughput biochemical methods such as UV-crosslinking can be used as an initial screen for methyltransferases

Acknowledgments

This work was supported by National Institutes of Health Grant GM026020. T.C.P. was supported by the UCLA Chemistry-Biology Interface Training Grant GM008496. We are grateful to Professor Christopher Lee for his comments on this work.

References

1. Cheng, X.; Blumenthal, RM., editors. *S-Adenosylmethionine-Dependent Methyltransferases: Structures and Functions..* World Scientific; Singapore: 1999.
2. Berman BP, Weisenberger DJ, Laird PW. Locking in on the human methylome. *Nature* 2009;27(4): 341–342.
3. Jeltsch A. Molecular enzymology of mammalian DNA methyltransferases. *Curr. Top. Microbiol. Immunol* 2006;301:203–225. [PubMed: 16570849]
4. Kim JK, Samaranayake M, Pardhan S. Epigenetics mechanisms in mammals. *Cell. Mol. Life Sci* 2009;66(4):596–612. [PubMed: 18985277]
5. Ng SS, Yue WW, Oppermann U, Klose RJ. Dynamic protein methylation in chromatin biology. *Cell. Mol. Life Sci* 2009;66(3):407–422. [PubMed: 18923809]
6. Zhao Q, Rank G, Tan YT, et al. PRMT5-mediated methylation of histone H4R3 recruits DNMT3A, coupling histone and DNA methylation in gene silencing. *Nat. Struct. Mol. Biol* 2009;16(3):304–311. [PubMed: 19234465]
7. Couture JF, Trievel RC. Histone-modifying enzymes: encrypting an enigmatic epigenetic code. *Curr. Opin. Struct. Biol* 2006;16(6):753–760. [PubMed: 17070031]
8. Horwich MD, Li C, Matranga C, et al. The *Drosophila* RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Current Biol* 2007;17(14):1265–1272.
9. Wassenegger M. The Role of the RNAi Machinery in Heterochromatin Formation. *Cell* 2005;122(1): 13–16. [PubMed: 16009128]
10. Yu B, Yang Z, Li J, et al. Methylation as a Crucial Step in Plant microRNA Biogenesis. *Science* 2005;307(5711):932–935. [PubMed: 15705854]
11. Li J, Yang Z, Yu B, Liu J, Chen X. Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in *Arabidopsis*. *Current Biol* 2005;15(16):1501–1507.
- 12**. Tkaczuk KL, Dunin-Horkawicz S, Purta E, Bujnicki JM. Structural and evolutionary bioinformatics of the SPOUT superfamily of methyltransferases. *BMC Bioinformatics* 2007;8:73. [PubMed: 17338813] [This paper comprehensively characterizes the SPOUT domain through bioinformatic analysis.]
- 13**. Petrossian TC, Clarke SG. Multiple motif scanning to identify methyltransferases from the yeast proteome. *Mol. Cell. Proteomics.* 2009 (in press). [This paper refines the definition of the motifs

of Class I methyltransferases, provides a new program for scoring multiple motifs, and utilizes HMM methods to generate lists of putative methyltransferases.]

- 14** Kozbial PZ, Mushegian AR. Natural history of *S*-adenosylmethionine-binding proteins. *BMC Struct. Biol* 2005;5:19. [PubMed: 16225687] [This review summarizes the structures of AdoMet binding proteins.]
15. Schubert HL, Blumenthal RM, Cheng X. Many paths to methyltransfer: a chronicle of convergence. *Trends Biochem. Sci* 2003;28(6):329–335. [PubMed: 12826405]
- 16*. Martin JL, McMillan FM. SAM (dependent) I AM: the *S*-adenosylmethionine-dependent methyltransferase fold. *Curr. Opin. Struct. Biol* 2002;12(6):783–793. [PubMed: 12504684] [This review of Class I methyltransferase structure and topology highlights substrate specificity due to residues outside the signature motifs.]
17. Okano M, Bell DW, Haber DA, Li E. DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. *Cell* 1999;99(3):247–257. [PubMed: 10555141]
18. Pavlopoulou A, Kossida S. Plant cytosine-5 DNA methyltransferases: structure, function, and molecular evolution. *Genomics* 2007;90(4):530–541. [PubMed: 17689048]
19. Sawada K, Yang Z, Horton JR, Collins RE, Zhang X, Cheng X. Structure of the Conserved Core of the Yeast Dot1p, a Nucleosomal Histone H3 Lysine 79 Methyltransferase. *J. Biol. Chem* 2004;279(41):43296–43306. [PubMed: 15292170]
20. Ingrosso D, Fowler AV, Bleibaum J, Clarke S. Sequence of the D-aspartyl/L-isoaspartyl protein methyltransferase from human erythrocytes. Common sequence motifs for protein, DNA, RNA, and small molecule *S*-adenosylmethionine-dependent methyltransferases. *J. Biol. Chem* 1989;264(33):20131–20139. [PubMed: 2684970]
21. Niewmierzycka A, Clarke S. *S*-Adenosylmethionine-dependent methylation in *Saccharomyces cerevisiae*. Identification of a novel protein arginine methyltransferase. *J. Biol. Chem* 1999;274(2):814–824. [PubMed: 9873020]
22. Kossykh VG, Schlagman SL, Hattman S. Conserved sequence motif DPPY in region IV of the phage T4 Dam DNA-[N6-adenine]-methyltransferase is important for *S*-adenosyl-L-methionine binding. *Nucl. Acids Res* 1993;21(20):4659–4662. [PubMed: 8233814]
23. Zhang X, Zhou L, Cheng X. Crystal structure of the conserved core of protein arginine methyltransferase PRMT3. *EMBO J* 2000;19(14):3509–3519. [PubMed: 10899106]
24. Singer MS, Kahana A, Wolf AJ, et al. Identification of high-copy disruptors of telomeric silencing in *Saccharomyces cerevisiae*. *Genetics* 1998;150(2):613–632. [PubMed: 9755194]
25. Fingerman IM, Li HC, Briggs SD. A charge-based interaction between histone H4 and Dot1 is required for H3K79 methylation and telomere silencing: identification of a new trans-histone pathway. *Genes Develop* 2007;21(16):2018–2029. [PubMed: 17675446]
- 26** Cheng XD, Collins RE, Zhang X. Structural and sequence motifs of protein (histone) methylation enzymes. *Annu. Rev. Biophys. Biomol. Struct* 2005;34:267–294. [PubMed: 15869391] [This paper provides an comprehensive review of the structure and motifs of lysine and arginine protein methyltransferases.]
27. Ng HH, Feng Q, Wang H, et al. Lysine methylation within the globular domain of histone H3 by Dot1 is important for telomeric silencing and Sir protein association. *Genes Develop* 2002;16(12):1518–1527. [PubMed: 12080090]
28. Gary JD, Lin WJ, Yang MC, Herschman HR, Clarke S. The predominant protein-arginine methyltransferase from *Saccharomyces cerevisiae*. *J. Biol. Chem* 1996;271(21):12585–12594. [PubMed: 8647869]
29. Kagan RM, Clarke S. Widespread occurrence of three sequence motifs in diverse *S*-adenosylmethionine-dependent methyltransferases suggests a common structure for these enzymes. *Arch. Biochem. Biophys* 1994;310(2):417–427. [PubMed: 8179327]
30. Timothy, LB.; Charles, E. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*; AAAI Press, Menlo Park, California. 1994; p. 28–36.
31. Katz JE, Dlakić M, Clarke S. Automated Identification of Putative Methyltransferases from Genomic Open Reading Frames. *Mol. Cell. Proteomics* 2003;2(8):525–540. [PubMed: 12872006]

32. Bailey TL, Gribskov M. Combining evidence using p-values: application to sequence homology searches. *Bioinformatics* 1998;14(1):48–54. [PubMed: 9520501]
33. Ansari MZ, Sharma J, Gokhale RS, Mohanty D. In silico analysis of methyltransferase domains involved in biosynthesis of secondary metabolites. *BMC Bioinformatics* 2008;9:454. [PubMed: 18950525]
- 34**. Söding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. *Nucl. Acids Res* 2005;33:W244–248. [PubMed: 15980461] [This paper describes the HHpred program for homology detection and structure prediction by HMM-HMM comparison.]
35. Huerta-Cepas J, Bueno A, Dopazo J, Gabaldón T. PhylomeDB: a database for genome-wide collections of gene phylogenies. *Nucl. Acids Res* 2007;36:D491–496. [PubMed: 17962297]
36. Atkinson HJ, Morris JH, Ferrin TE, Babbitt PC. Using Sequence Similarity Networks for Visualization of Relationships Across Diverse Protein Superfamilies. *PLOS One* 2009;4(2):e4345. [PubMed: 19190775]
37. Rea S, Eisenhaber F, O'Carroll D, et al. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* 2000;406(6796):593–599. [PubMed: 10949293]
38. Wang, Y. Methylation and Demethylation of Histone Arg and Lys Residues in Chromatin Structure and Function. In: Clarke, SG.; Tamanoi, F., editors. *The Enzymes: Protein Methyltransferases*. Elsevier; 2006. p. 123-153.
39. Southall SM, Wong PS, Odho Z, Roe SM, Wilson JR. Structural Basis for the Requirement of Additional Factors for Mll1 Set Domain Activity and Recognition of Epigenetic Marks. *Mol. Cell* 2009;33(2):181–191. [PubMed: 19187761]
- 40*. Dillon SC, Zhang X, Trievel RC, Cheng X. The SET-domain protein superfamily: protein lysine methyltransferases. *Genome Biol* 2005;6(8):227. [PubMed: 16086857] [This paper provides an analysis of SET domain motifs and substrate specific domains for these proteins.]
41. Xiao, B.; Gamblin, SJ.; Wilson, JR. Structure of SET Domain Protein Lysine Methyltransferases. In: Clarke, SG.; Tamanoi, F., editors. *The Enzymes: Protein Methyltransferases*. Elsevier; 2006.
42. Aravind L, Iyer LM. Provenance of SET-Domain Histone Methyltransferases Through Duplication of a Simple Structural Unit. *Cell Cycle* 2003;2(4):369–376. [PubMed: 12851491]
43. Xiao B, Jing C, Wilson JR, et al. Structure and catalytic mechanism of the human histone methyltransferase SET7/9. *Nature* 2003;421(6923):652–656. [PubMed: 12540855]
44. Couture JF, Dirk LM, Brunzelle JS, Houtz RL, Trievel RC. Structural origins for the product specificity of SET domain protein methyltransferases. *Proc. Natl. Acad. Sci. U. S. A* 2008;105(52):20659–20664. [PubMed: 19088188]
45. Zhang X, Yang Z, Khan SI, et al. Structural basis for the product specificity of histone lysine methyltransferases. *Mol. Cell* 2003;12(1):177–185. [PubMed: 12887903]
- 46*. Qian C, Zhou M. SET domain protein lysine methyltransferases: Structure, specificity and catalysis. *Cell. Mol. Life Sci* 2006;63(23):2755–2763. [PubMed: 17013555] [This paper also provides an analysis of SET domain motifs and substrate specific domains for these proteins.]
47. Marmorstein R. Structure of SET domain proteins: a new twist on histone methylation. *Trends Biochem. Sci* 2003;28(2):59–62. [PubMed: 12575990]
48. Porras-Yakushi TR, Whitelegge JP, Clarke S. A novel SET domain methyltransferase in yeast: Rkm2-dependent trimethylation of ribosomal protein L12ab at lysine 10. *J. Biol. Chem* 2006;281(47):35835–35835. [PubMed: 17005568]
49. Bottomley MJ. Structure of protein domains that create or recognize histone modifications. *EMBO Reports* 2004;5(5):464–469. [PubMed: 15184976]
50. Letunic I, Doerks T, Bork P. SMART 6: recent updates and new developments. *Nucleic Acids Res* 2009;37:D229–32. (Database issue). [PubMed: 18978020]
51. Baumbusch LO, Thorstensen T, Krauss V, et al. The Arabidopsis thaliana genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes. *Nucl. Acids Res* 2001;29(21):4319–4333. [PubMed: 11691919]
52. Springer NM, Napoli CA, Selinger DA, et al. Comparative Analysis of SET Domain Proteins in Maize and Arabidopsis Reveals Multiple Duplications Preceding the Divergence of Monocots and Dicots. *Plant Physiol* 2003;132(2):907–925. [PubMed: 12805620]

- 53*. Ng DW-K, Wang T, Mahesh B, Chandrasekharan, Aramayo R, Kertbundit S, Hall TC. Plant SET domain-containing proteins: Structure, function and regulation. *Biochim. Biophys. Acta* 2007;1769(5-6):316–329. [PubMed: 17512990] [A system of classification of SET domain proteins is presented in this paper.]
54. Dirk, LMA.; Trievel, RC.; Houtz, RL. Non-Histone Protein Lysine Methyltransferases: Structure and Catalytic Roles. In: Clarke, SG.; Tamanoi, F., editors. *The Enzymes: Protein Methyltransferases*. Elsevier; 2006. p. 178-228.
55. Anantharaman V, Koonin EV, Aravind L. SPOUT: a class of methyltransferases that includes spoU and trmD RNA methylase superfamilies, and novel superfamilies of predicted prokaryotic RNA methylases. *J. Mol. Microbiol. Biotechnol* 2002;4(1):71–75. [PubMed: 11763972]
56. Das K, Acton T, Chiang Y, Shih L, Arnold E, Montelione GT. Crystal structure of RlmAI: implications for understanding the 23S rRNA G745/G748-methylation at the macrolide antibiotic-binding site. *Proc. Natl. Acad. Sci. U. S. A* 2003;101(12)
57. Michel G, Sauve V, Larocque R, Li Y, Matte A, Cygler M. The structure of the RlmB 23S rRNA methyltransferase reveals a new methyltransferase fold with a unique knot. *Structure* 2002;10(10): 1303–1315. [PubMed: 12377117]
58. Nureki O, Watanabe K, Fukai S, et al. Deep Knot Structure for Construction of Active Site and Cofactor Binding Site of tRNA Modification Enzyme. *Structure* 2004;12(4):593–602. [PubMed: 15062082]
59. Watanabe K, Nureki O, Fukai S, et al. Roles of conserved amino acid sequence motifs in the SpoU (TrmH) RNA methyltransferase family. *J. Biol. Chem* 2005;280(11):10368–10377. [PubMed: 15637073]
60. Taylor AB, Meyer B, Leal BZ, et al. The crystal structure of Nep1 reveals an extended SPOUT-class methyltransferase fold and a pre-organized SAM-binding site. *Nucleic Acids Res* 2008;36(5):1542–1554. [PubMed: 18208838]
61. Leulliot N, Bohnsack MT, Graille M, Tollervey D, Van TH. The yeast ribosome synthesis factor Emg1 is a novel member of the superfamily of alpha/beta knot fold methyltransferases. *Nucleic Acids Res* 2008;36(2):626–639.
62. Matthews RG, Koutmos M, Datta S. Cobalamin-dependent and cobamide-dependent methyltransferases. *Curr. Opin. Struct. Biol* 2009;18(6):658–666. [PubMed: 19059104]
63. Dixon MM, Huang S, Matthews RG, Ludwig M. The structure of the C-terminal domain of methionine synthase: presenting S-adenosylmethionine for reductive methylation of B12. *Structure* 1996;4(11): 1263–1275. [PubMed: 8939751]
64. Sali A, Potterton L, Yuan F, van Vlijmen H, Karplus M. Evaluation of comparative protein modelling by MODELLER. *Proteins* 1995;23(3):318–326. [PubMed: 8710825]
65. Kelley LA, Sternberg MJE. Protein structure prediction on the web: a case study using the Phyre server. *Nat. Protoc* 2009;4(3):363–371. [PubMed: 19247286]
66. Vinci CR, Clarke SG. Recognition of age-damaged (R,S)-adenosyl-L-methionine by two methyltransferases in the yeast *Saccharomyces cerevisiae*. *J. Biol. Chem* 2007;282(12):8604–8612. [PubMed: 17264075]
67. Koutmos M, Pejchal R, Bomer TM, Matthews RG, Smith JL, Ludwig ML. Metal active site elasticity linked to activation of homocysteine in methionine synthases. *Proc. Natl. Acad. Sci. U. S. A* 2008;105(9):3286–3291. [PubMed: 18296644]
68. Ferrer J-L, Ravanel S, Robert M, Dumas R. Crystal Structures of Cobalamin-independent Methionine Synthase Complexed with Zinc, Homocysteine, and Methyltetrahydrofolate. *J. Biol. Chem* 2004;279(43):44235–44238. [PubMed: 15326182]
69. Sofia HJ, Chen G, Hetzler BG, Reyes-Spindola JF, Miller NE. Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms: functional characterization using new analysis and information visualization methods. *Nucl. Acids Res* 2001;29(5):1097–1106. [PubMed: 11222759]
70. Berkovitch F, Nicolet Y, Wan JT, Jarrett JT, Drennan CL. Crystal structure of biotin synthase, an S-adenosylmethionine-dependent radical enzyme. *Science* 2004;303(5654):76–79. [PubMed: 14704425]

71. Wang SC, Frey PA. *S*-adenosylmethionine as an oxidant: the radical SAM superfamily. *Cell* 2007;32(3):101–110.
72. Booker SJ. Anaerobic functionalization of unactivated C-H bonds. *Curr. Opin. Chem. Biol* 2009;13(1):58–73. [PubMed: 19297239]
73. Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* 2007;35:D61–D65. (Database issue). [PubMed: 17130148]
74. Schubert HL, Wilson KS, Raux E, Woodcock SC, Warren MJ. The X-ray structure of a cobalamin biosynthetic enzyme, cobalt-precorrin-4 methyltransferase. *Nature Struct. Biol* 1998;5(7):585–592. [PubMed: 9665173]
75. Romano JD, Michaelis S. Topological and Mutational Analysis of *Saccharomyces cerevisiae* Ste14p, Founding Member of the Isoprenylcysteine Carboxyl Methyltransferase Family. *Mol. Biol. Cell* 2001;12(7):1957–1971. [PubMed: 11451995]
76. Finn RD, Tate J, Mistry J, et al. The Pfam protein families database. *Nucleic Acids Res* 2008;36:D281–D288. (Database Issue). [PubMed: 18039703]
77. Clancy MJ, Shambaugh ME, Timpte CS, Bokar JA. Induction of sporulation in *Saccharomyces cerevisiae* leads to the formation of N6-methyladenosine in mRNA: a potential mechanism for the activity of the IME4 gene. *Nucl. Acids Res* 2002;30(20):4509–4518. [PubMed: 12384598]
78. Subbaramaiah K, Simms SA. Photolabeling of CheR methyltransferase with *S*-adenosyl-L-methionine (AdoMet). Studies on the AdoMet binding site. *J. Biol. Chem* 1992;267(12):8636–8642. [PubMed: 1349019]
79. Zhou S, Bailey MJ, Dunn MJ, Preedy VR, Emery PW. A systematic investigation into the recovery of radioactively labeled proteins from sodium dodecyl sulfate-polyacrylamide gels. *Electrophoresis* 2003;25(1):1–7. [PubMed: 14730561]
80. Zhou S, Mann CJ, Dunn MJ, Preedy VR, Emery PW. Measurement of specific radioactivity in proteins separated by two-dimensional gel electrophoresis. *Electrophoresis* 2006;27(5-6):1147–1153. [PubMed: 16470777]
81. Zhu H, Bilgin M, Bangham R, et al. Global analysis of protein activities using proteome chips. *Science* 2001;293(5537):2101–2105. [PubMed: 11474067]
82. Rathert P, Dhayalan A, Ma H, Jeltsch A. Specificity of protein lysine methyltransferases and methods for detection of lysine methylation of non-histone proteins. *Mol. Biosyst* 2008;4(12):1186–1190. [PubMed: 19396382]
83. Laskowski RA. PDBsum new things. *Nucl. Acids Res* 2009;37:D355–D359. [PubMed: 18996896]
84. Tatusov RL, Fedorova ND, Jackson JD, et al. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 2003;4:41. [PubMed: 12969510]
101. <https://www.chem.ucla.edu/files/MotifSetup.Zip>
102. <http://www.yeastgenome.org/>

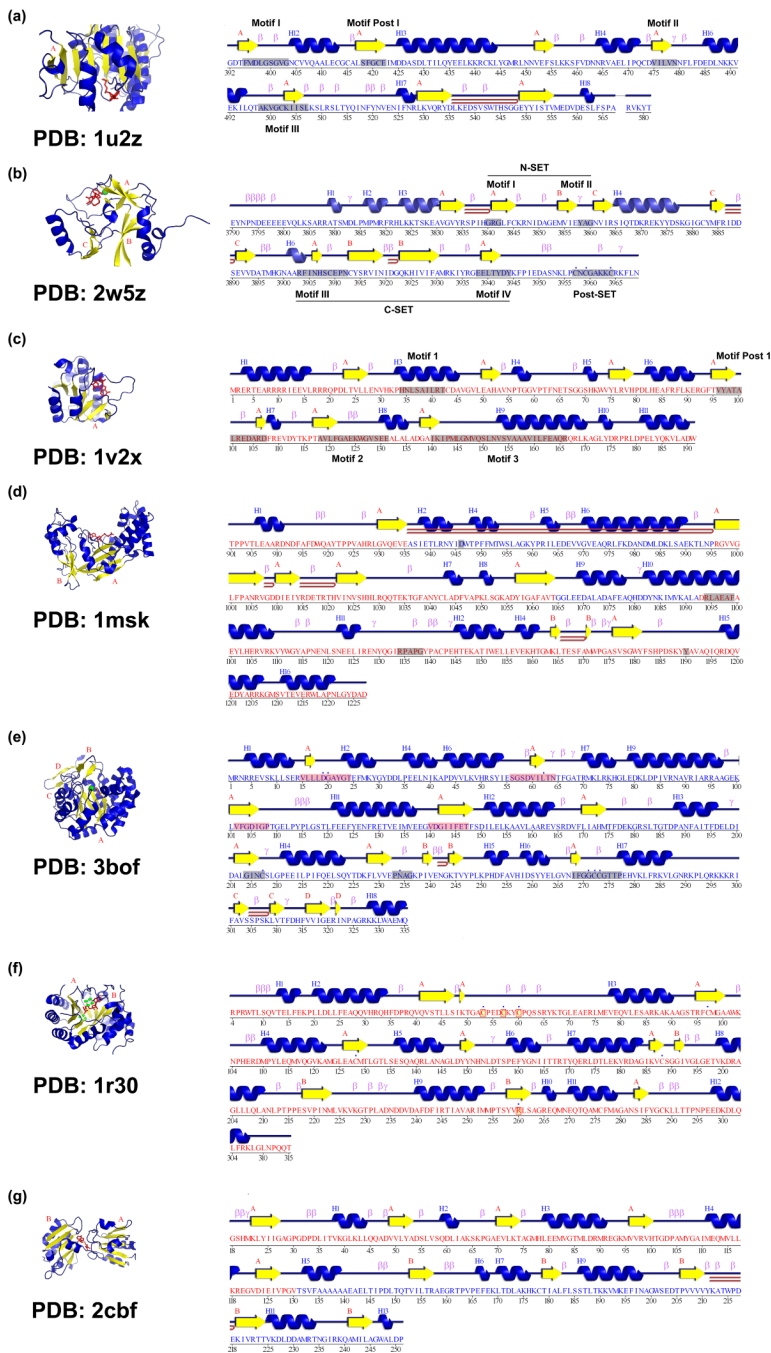


Figure 1. Seven structural strategies for methyltransferases. The methyltransferase domain for a representative member of each topologically distinct family created in PyMOL is shown on the left. Beta strands are indicated by yellow arrows, alpha helices and coils are depicted in blue. Zinc ion and iron sulfur clusters are indicated as green spheres, the cofactors AdoMet/AdoHcy are colored red, and substrate homocysteine is shown in orange. The corresponding amino acid sequence with annotated secondary structure (adapted from PDBsum [83]) is shown on the right. These annotations include: β , beta turn; γ , gamma turn; D , beta hairpin; H1, H2, ..., helices; A, B, ..., individual beta sheets; \blacksquare , metal-binding residue. Additionally, residues in well-defined methyltransferase motifs are shaded in gray, a) Class I seven beta strand

methyltransferase, Dot1 (pdb: 1u2z, [19]), (b) SET methyltransferase, MLL1 (pdb: 2w5z, [39]), (c) SPOUT methyltransferase, TrmH (pdb: 1v2x, [58]), (d) reactivation domain of MetH, (pdb: 1msk, [63]), (e) Hcy-binding domain of MetH, (pdb: 3bof, [67]), (f) Radical SAM, BioB (pdb: 1r30, [70]), (g) Precorrin-4 methyltransferase, CbiF (pdb: 2cbf, [74]). Residues shaded in red are homologous regions between homocysteine methyltransferases Sam4/Hmt1 and MetH. The representative member of the radical SAM family (BioB) is not a methyltransferase; however there are no structures known yet for family members that participate in methylation reactions.

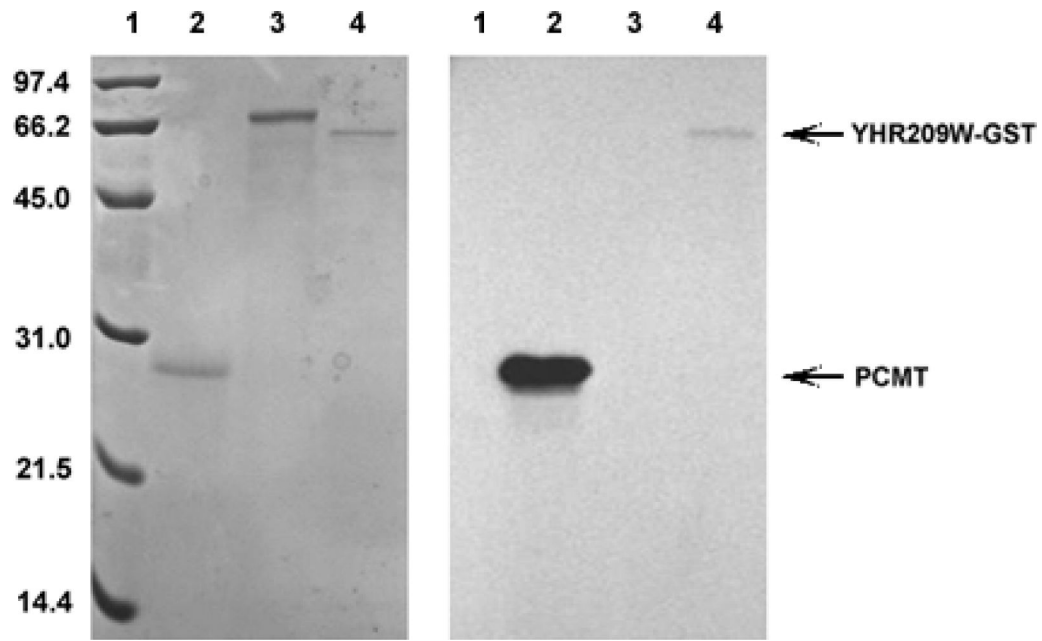


Figure 2.

Biochemical determination of AdoMet binding by UV crosslinking. Proteins (2 μ g) were mixed in a final volume of 200 μ l containing 1.7 μ M [*methyl*- 3 H]AdoMet (80 Ci/mmol) in 10 mM potassium phosphate, 100 mM NaCl, 2 mM EDTA, 1 mM dithiothreitol, 5% glycerol, pH 7.0 [78] in an open plastic 96-well plate. The reaction mixture was exposure to UV irradiation in a Stratalinker 2400 apparatus for 30 min at 4 $^{\circ}$ C, and the products electrophoresed on a SDS-PAGE gel. After staining with Coomassie Blue and treatment with En3Hance, the dried gel was exposed to film for 72 h at -80 $^{\circ}$ C. In this experiment, the human isoaspartyl protein methyltransferase was used as a positive control (lane 2) and bovine serum albumin was used as a negative control (lane 3). The putative yeast methyltransferase YHR209W was purified as a GST-fusion protein (lane 4). Molecular weight standards (Bio-Rad LMW) were loaded in lane 1.

Table 1
The Yeast ‘Methyltransferase’ – Known and Putative Enzymes in Each Structural Family

Proteins in non-italic type are known methyltransferases in *Saccharomyces cerevisiae*. Proteins in italics are candidate methyltransferases; question marks indicate similarity within the class but these species are not likely to be methyltransferases (see text). All gene/protein and open reading frame designations use the standard nomenclature of the *Saccharomyces* Genome Database (SGD), which also includes genetic, biochemical, and functional characterization of these species [102]. Evidence supporting the assignment of each of the candidate methyltransferases is provided below by footnotes.

	Seven-Beta Strand (Class I)^{a,b,c,d}	SET^{e,f}	SPOUTG^{g,h}	MetH (activation domain)ⁱ	Radical SAM^j	Homocysteine^k	Precorrin-like^l	Membrane^m	N6-Adenosineⁿ
Nop1	Trm44	Set1	Trm10	NONE	<i>Lip5?j</i>	Mht1	Dph5	Opi3	<i>Ime4ⁿ</i>
Nop2	Ppm2	Set2	Mfm1		<i>Eip3?j</i>	Sam4	Met1	Stel4	<i>Kar4ⁿ</i>
Rrp8	Gcd14	<i>Set3^{e,f}</i>	Trm3		<i>Tyw1?j</i>	<i>YMR321C^{?k}</i>		Cho2	<i>YGR001Cⁿ</i>
Mrm2	Tgs1	<i>Set4^{e,f}</i>	<i>Emg1^{g,h}</i>		<i>Bit2?j</i>				
Spb1	Dot1	<i>Set5^{e,f}</i>	<i>YGR283C^{g,h}</i>						
Dim1	Hmt1	<i>Set6^{e,f}</i>	<i>YMR310C^{g,h}</i>						
Bud23	Rmt2	Rkm1	<i>YOR021C^h</i>						
Adb1	Hsl7	Rkm2	<i>Tan1^h</i>						
Trm1	Ppm1	Rkm3							
Trm2	Mtq1	Rkm4							
Ncl1	Mtq2	Ctm1							
Trm5	Erg6	<i>YHL039W^{e,f}</i>							
Trm7	Coq3								
Trm8	Coq5								
Trm9	Tmt1								
Trm11	Nnt1								
Tyw3									
Trm13									

^aRef. 21.

^bRef. 31.

^cRef. 13.

- ^dThe program HHpred is used with a multiple alignment generated by ClustalW of the amino acids in the motif sequences obtained from the “yeast reference set” of Ref. ¹³ against the yeast proteome.
- ^eRef. 48.
- ^fThe program HHpred is used with the multiple alignment of the amino acids in SET domain defined obtained from the SMART database (SMART family: SM00317) against the yeast proteome (see text).
- ^gRef. 12.
- ^hThe program HHpred is used with the multiple alignment of each COG group described in Ref. ¹² against the yeast proteome.
- ⁱThe program HHpred with the PSI-BLAST parameter is used with the single sequence of the C-terminal of Meth (UniProtKB: locus METH_ECOLI, accession P13009[897-1227]) against the yeast proteome. Additionally, use of the fold recognition programs MODELLER and PHYRE with the same sequence indicated that no other proteins have significant homology to Meth in the yeast proteome (see text).
- ^jThe program HHpred with the PSI-BLAST parameter is used with the single sequence of N-terminal sequence of Meth (UniProtKB: locus METH_ECOLI, accession P13009[2-325]) against the yeast proteome. Additionally, HHpred is used with the multiple alignment of the corresponding COG group [84] (COG0646; Methionine synthase I (cobalamin-dependent), methyltransferase domain) against the yeast proteome.
- ^kThe program HHpred is used with the multiple alignment generated by ClustalW of proteins identified by the keyword search “Radical SAM methyltransferase” in RefSeq database against the yeast proteome.
- ^lThe program HHpred with the PSI-BLAST parameter is used with the single sequence of CbfF sequence (UniProtKB: locus CBIF_BACME, accession O87696) against the yeast proteome. Additionally, HHpred is used with the multiple alignment of the corresponding COG group (COG1798; Diphthamide biosynthesis methyltransferase DPH5) against the yeast proteome.
- ^mThe program HHpred is used with the multiple alignment of the PEMT family obtained from Pfam (Pfam family: PF04191) against the yeast proteome.
- ⁿThe program HHpred with the PSI-BLAST parameter is used with the single sequence of Ime4 sequence [102] against the yeast proteome. Additionally, HHpred is used with the multiple alignment of the corresponding COG group (COG4725; Predicted N6-adenine RNA methylase) against the yeast proteome.

Table II

Classes of SET Domain Methyltransferases [51-53]

	Substrate ^a	Additional Motifs/Domains ^b	Function ^c	<i>S. cerevisiae</i> proteins ^d	<i>D. melanogaster</i> proteins ^d
Class I	H3K27	Pre-SET, SANT	euchromatic silencing X-inactivation		EZ
Class II	H3K36	Pre-SET, Post-SET, PWWP, PHD, HMG, BAH, eaf3	transcriptional elongation transcriptional silencing	Set2	ASH1 Mes-4 SET2
Class III	H3K4	Post SET, PWWP, PHD, RRM, Chd1, MBT, JMJD2A, WDR5	transcriptional activation transcriptional elongation	Set1	Trx
Class IV	H3K4?	PHD	?	Set3 Set4	Mes-4
Class V	H3K9	Pre-SET, Post SET, HPI, CDY1, JMJD2A, ankyrin repeats, MBD, YDG	euchromatic silencing-, heterochromatic silencing transcriptional repression transcriptional activation		Su(var) 3-9 G9a
Class VI	H3K36?	pre-SET, Post-SET	cell-cycle silencing transcriptional silencing	Set5 Set6	Bzd ^e
Class VII	Non-histones			Rkm1-4 Ctm1 YHL039W	Dmel\CG32732 ^f
EZH2 family	H1K26	PreSET, SANT	transcriptional silencing		HP1
H4K20 family	H4K20	PostSET, Crb2, JMJD2A	transcriptional silencing transcriptional activation cell cycle silencing mitosis, cytokinesis, heterochromic silencing		SET8 SUV4-H20

^aRef. 46,51,53.^bFor domain/motif assignments and designations, see Ref. 40,46,52,53.^cFunction includes information from all SET proteins among a variety of organisms. Adapted from Refs. 40 and 46.^dThe program HHpred is used with a multiple alignment generated by ClustalW of each SET class of proteins obtained from Ref. 53 against the yeast and *Drosophila melanogaster* proteomes.^eOrthologous to mammalian ASH1.^fContains the substrate-binding region of Rubisco large subunit methyltransferase.