UCLA
UCLA Previously Published Works

Title
DNA methylation pathways and their crosstalk with histone methylation

Permalink
https://escholarship.org/uc/item/22t6647z

Journal
Nature Reviews Molecular Cell Biology, 16(9)

ISSN
1471-0072

Authors
Du, Jiamu
Johnson, Lianna M
Jacobsen, Steven E
et al.

Publication Date
2015-09-01

DOI
10.1038/nrm4043

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Nat Rev Mol Cell Biol. Author manuscript; available in PMC 2015 December 08.

Published in final edited form as:

Nat Rev Mol Cell Biol. 2015 September ; 16(9): 519–532. doi:10.1038/nrm4043.

DNA methylation pathways and their crosstalk with histone methylation

Jiamu Du[1,*], **Lianna M. Johnson**[2,*], **Steven E. Jacobsen**[2], and **Dinshaw J. Patel**[3]

[1]Shanghai Center for Plant Stress Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 201602, China

[2]Howard Hughes Medical Institute and Department of Molecular, Cell and Developmental Biology, University of California at Los Angeles, Los Angeles, California 90095, USA

[3]Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA

Abstract

Methylation of DNA and of histone 3 at Lys 9 (H3K9) are highly correlated with gene silencing in eukaryotes from fungi to humans. Both of these epigenetic marks need to be established at specific regions of the genome and then maintained at these sites through cell division. Protein structural domains that specifically recognize methylated DNA and methylated histones are key for targeting enzymes that catalyse these marks to appropriate genome sites. Genetic, genomic, structural and biochemical data reveal connections between these two epigenetic marks, and these domains mediate much of the crosstalk.

Epigenetic marks regulate gene expression and suppress transposon activity. Methylation of histone 3 at lysine 4 (H3K4) and H3K9 are among the most highly conserved epigenetic marks that correlate well with gene activation and gene silencing, respectively, in plants, animals and fungi[1]. One, two or three methyl groups may be added to lysine, and these different methylation states often have different functions.

DNA methylation, although not found in all organisms, is also highly correlated with gene silencing[2,3]. DNA methylation is established by specialized *de novo* DNA methyltransferase enzymes and is present in three different DNA sequence contexts: CG and CHG (where H corresponds to A, T, or C), which are symmetrical sequences, and CHH, which is an asymmetrical sequence[3]. After establishment, DNA methylation is perpetuated through both mitotic divisions and meiotic divisions by maintenance DNA methyltransferases. The mechanisms that maintain these different types of methylation vary widely between different eukaryotes.

Research in model organisms has shown that there are extensive links and crosstalk between histone modifications and DNA methylation. Key to these links are the readers of histone

Correspondence to S.E.J. and D.J.P. jacobsen@ucla.edu; pateld@mskcc.org.

*These authors contributed equally to this work.

Competing interests statement: The authors declare no competing interests.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

methylation including plant homeodomains (PHDs), chromodomains and bromo adjacent homology domains (BAH domains), and readers of DNA methylation such as the SRA (SET- and RING-associated), CXXC domain and methyl-CpG-binding domain (MBD).

In this article, we review the connections and crosstalk between histone and DNA methylation marks and the structural domains that facilitate these connections in fungi, plants and mammals. Although additional histone marks such as H3K36 and H3K27 methylation may affect DNA methylation[4–7], the focus of this Review will be on the more conserved H3K4 and H3K9 methylation marks.

A unidirectional link in Fungi

Neurospora crassa, a red bread mould of the phylum Ascomycota, provided the first direct evidence for a link between histone H3K9 methylation and DNA methylation. This link was shown to be unidirectional from histones to DNA.

DNA methylation during N. crassa life cycle

The *N. crassa* life cycle includes a vegetative (asexual) and a sexual cycle. At the beginning of the sexual cycle, just after two haploid spores fuse but before nuclear fusion (see FIG. 1a), a genome defence system is activated to protect the genome from repeated sequences, such as transposable elements (TEs). This genome defence system (known as repeat-induced point mutation (RIP)) mutates repeated invasive DNA sequences by changing numerous Cs to Ts, rendering the sequences AT-rich[8–11]. Most cytosine methylation in *N. crassa* is restricted to the remaining Cs in these mutated regions, though traces are found in other sequences including bona fide genes[12–14]. Cs in both symmetrical and asymmetrical sequences are methylated[14,15]. Therefore, maintenance of this methylation pattern in vegetative cells is potentially more complex than that for symmetrically methylated DNA, which is ‘remembered’ via hemimethylated sites that result from replication. It was demonstrated that although ‘maintenance methylation’ (some form of a copying mechanism) can occur in *N. crassa*[16], it is not completely sequence-independent and the vast majority of methylation in the wild-type genome results from reiterative *de novo* methylation in vegetative cells[17].

A unidirectional pathway from histone to DNA methylation

A key advance in understanding how cytosine methylation is established in vegetative cells of *N. crassa* came with the identification of a mutation that abolished DNA methylation and mapped to a gene, *dim-5*, which was predicted to encode a H3K9 methyltransferase[18]. Several experiments established that DIM-5 is responsible for trimethylation (not dimethylation) of H3K9 and that this modification is required for DNA methylation in this organism[17–19]. DIM-5 recognizes the AT rich-sequences resulting from RIP as part of a complex known as DCDC, which comprises DIM-5, DIM-7, DIM-9, CUL4 and DDB1 (DNA damage-binding protein 1)[20]. DIM-7 is specifically required to recruit DIM-5 to chromatin[21] (FIG. 1b), but the exact mechanism by which the signal created by the RIP pathway is recognized by DCDC is not yet understood.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

H3K9 methylation by DIM-5 is regulated by several factors, such as protein complexes including histone deacetylases, chromodomain proteins and a putative histone demethylase[22,23]. After the H3K9me3 mark has been incorporated into the histones associated with RIP affected sequences, this mark must be read and relayed to the DNA cytosine methyltransferase, DIM-2. Heterochromatin protein 1 (HP1), which forms a complex with DIM-2, plays a central role as it functions as an adaptor (FIG. 1b). The chromodomain of HP1 recognizes H3K9me3 and its chromo-shadow domain interacts with the two PXVXL-related domains of DIM-2 (REF. 24). Knock out of the *hpo* gene, which encodes HP1, leads to complete loss of DNA methylation, indicating its essential role in DNA methylation[25]. Moreover, this pathway is unidirectional, from histone methylation to DNA methylation via HP1, as knock out of either *hpo* or *dim-2* has little effect on histone methylation[17].

Methylation links in *Arabidopsis thaliana*

In *A. thaliana*, cytosines are also methylated in all sequence contexts[26–28]; however, two distinct methylation patterns have been observed. Heavy cytosine methylation in all sequence contexts is observed in transposable elements, which are found primarily in the pericentromeric heterochromatin, but also in small patches in euchromatin. Methylation at CG residues only is observed in the exons of approximately one third of transcribed genes and is referred to as gene body methylation[26,27]. Transposable element methylation results in transcriptional silencing, whereas gene body methylation is correlated with moderately high transcription.

DNA methylation is established *de novo* by the RNA-directed DNA methylation (RdDM) pathway and maintained by three pathways.

Enzymes for DNA and histone methylation

In *A. thaliana* there are seven DNA methyltransferase encoding genes: domains rearranged DNA methylase 1 (*DRM1*) and *DRM2*; chromomethylase 1 (*CMT1*), *CMT2* and *CMT3*; methyltransferase 1 (*MET1*) and *MET2* and fifteen putative H3K9 methyltransferase encoding genes (*SUVH1–10* and *SUVR1–5*)[3,29]. Four DNA methyltransferase genes have been genetically shown to be active (*DRM2*, *CMT2*, *CMT3*, *MET1*)[30–33] and three histone methyltransferases are responsible for the majority of H3K9 methylation (the SU(var)3–9 homologues KRYPTONITE (KYP; also known as SUVH4), SUVH5 and SUVH6)[34,35]. In addition, two catalytically inactive homologues (SUVH2 and SUVH9) have a key role in RdDM[36–39].

Pathways for maintenance of DNA methylation

Three different maintenance pathways exist in *Arabidopsis* spp.: RdDM, the CMT2–CMT3 pathway and the MET1 pathway. RdDM maintains methylation of small euchromatic sites via a reiterative *de novo* mechanism, similar to *N. crassa* but involving siRNAs. DNA methylation by CMT2 and CMT3 is dependent on histone methylation. CMT2 prefers unmethylated DNA as a substrate and can catalyse the methylation of both CHH and CHG, making this a *de novo* DNA methyltransferase that is recruited by histone methylation[40].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

However, this pathway is not required for *de novo* methylation of unmethylated DNA introduced by *Agrobacterium tumefaciens*-mediated transformation. CMT3 prefers hemimethylated CHG sites, consistent with it being a maintenance methyltransferase[41]. MET1 functions strictly as a maintenance pathway for CG methylation at all sites in the genome and is not dependent on histone methylation. We first discuss the CMT2–CMT3 pathway as it has the most direct link to histone methylation.

Direct links between CMT2–CMT3 and KYP histone methylation

Historically, the first indication that histone and DNA methylation were linked in *Arabidopsis* spp. was provided by the discovery of the chromodomain-containing DNA methyltransferase CMT3, which is responsible for CHG methylation[31,32]. Subsequently, it was found that mutation in the *KYP* gene also reduced CHG methylation[42–44], placing histone methylation upstream of CHG methylation. However, this pathway turned out to be more complex than in *N. crassa* as knockout of *CMT3* also resulted in reduction in histone methylation[40,45–47]. These and subsequent studies have provided support for a self-reinforcing loop between histone and DNA methylation[34,48,49] (FIG. 2a).

The primary sequence of CMT3 contains a unique arrangement of three domains: an amino-terminal BAH domain, a carboxy-terminal DNA methyltransferase domain and a chromodomain embedded within the DNA methyltransferase domain[50,51] (FIG. 2b). The structure of maize ZMET2 (REF. 52), which is an orthologue of *A. thaliana* CMT3, revealed a unique triangular-shaped scaffold[41] (FIG. 2c). The BAH and the chromodomains align along two edges of the centrally positioned DNA methyltransferase domain, despite the chromodomain being embedded within the DNA methyltransferase domain in the primary sequence[41].

Isothermal titration calorimetry (ITC) binding studies of either ZMET2 or CMT3 with H3K9me2-containing H3 peptides indicated that both proteins contain a pair of H3K9me2 binding sites[40,41]. The structure of ZMET2 with H3K9me2-containing peptides of different lengths established that both the BAH and chromodomains contain H3K9me2 binding sites (FIG. 2c). Both domains recognize the H3K9me2 peptide involving a classical binding model whereby the K9me2 side chain specifically inserts into and is anchored within aromatic cages (FIG. 2d,e), while the main chain of the bound peptide forms extensive hydrogen bonding interaction with the protein[41,53].

Importantly, the H3K9me2-containing peptides are bound to both BAH and chromodomains with directionality such that the C terminus of the bound peptides are directed towards the catalytic centre of ZMET2, while the N terminus of the peptides extend out towards the solvent[41] (FIG. 2c). The parallel orientation of the two peptides raises the possibility that the BAH and chromodomains can simultaneously recognize a pair of H3K9me2 containing tails associated with either the same or adjacent nucleosomes. Indeed, disruption of the aromatic cage of either the BAH or the chromodomain results in a substantial loss of *in vivo* CHG methylation, while the *in vitro* DNA methyltransferase activity remains unchanged[41]. This observation indicates that both the BAH and chromodomains are essential for the *in vivo* function of CMT3, showing that the enzyme requires both H3K9me2 binding domains for *in vivo* targeting of the protein.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

In plants, the KYP protein possesses an N-terminal SRA domain capable of recognizing methylated DNA and C-terminal pre-SET (Su(var)3–9, Enhancer-of-zeste and Trithorax), SET and post-SET domains, which adopt a typical histone methyltransferase fold[42,43,48] (FIG. 2f). In a recently reported crystal structure of KYP in complex with methylated non-CG DNA, cofactor SAH and unmodified H3 peptide substrate (FIG. 2g), the SRA domain of KYP forms a positively charged surface cleft to accommodate the methylated DNA: the methylated cytosine is flipped out from the DNA duplex and inserted into the pocket within the SRA domain[54] (FIG. 2h). The histone peptide and cofactor SAH are positioned in between the SET and post-SET domains[54] (FIG. 2i). The position of the SRA domain relative to the SET domain is the same in both the KYP complex and the free-form structure of KYP homologue SUVH9, suggesting there is no significant conformational change in KYP upon binding of DNA and histonesubstrate[37,54]. The methylated DNA thus may serve as a platform for recruitment of KYP to nucleosomes, with subsequent methylation of the H3 tail. This simple system allows for a reinforcing loop that requires no adaptors. KYP and CMT3 do not interact directly with each other, but rather bind to epigenetic marks that are installed by the other partner in the feedback loop[41,52,54] (FIG. 2a).

More recently, a second chromomethylase gene, *CMT2*, was discovered to be active and responsible for the majority of methylation at CHH sites in pericentromeric heterochromatin[40,55]. Knocking out the three major H3K9 methyltransferases, *kyp*, *suvh5* and *suvh6*, eliminates CMT2- and CMT3-dependent DNA methylation genome-wide, indicating that these enzymes are completely dependent on H3K9 methylation for binding to their target sites[56] (FIG. 2a). Biochemical studies have revealed that CMT2 can specifically recognize H3K9me2 peptides with a 1:2 molar ratio, indicative of a dual recognition mode similar to that of CMT3 (REF. 40). However, *in vitro* binding studies revealed that CMT2 prefers H3K9me2 over H3K9me1, whereas CMT3 has no preference. These observations suggest that although the BAH and chromodomains of CMT2 and CMT3 are very similar, they have different binding specificities. ChIP–seq was used to analyse histone methylation genome wide, and it was found that CMT2 sites are enriched in H3K9me2, suggesting that the number of methyl groups on H3K9 may determine which DNA methyltransferase is recruited[40].

Further analysis of mutants revealed that elimination of all non-CG methylation (such as organisms with mutations in *drm1*, *drm2*, *cmt2* and *cmt3* (ddcc mutants)) also resulted in total loss of all H3K9 methylation[40]. This suggests that KYP, SUVH5 and SUVH6 bind specifically to non-CG methylation sites through their SRA domains, as observed *in vitro*[48,54]. The observations that eliminating non-CG methylation results in loss of all H3K9 methylation, and that eliminating all H3K9 methylation results in loss of all non-CG methylation, confirms the reinforcing loop model (FIG. 2a).

Indirect DNA-histone methylation links during RdDM

Many of the small methylation patches in euchromatin are maintained by the reiteration of the *de novo* methylation RdDM pathway[57] (FIG. 3a). There are two main steps in this pathway that lead to the recruitment of DRM2, and that involve pre-existing methylated histones and DNA.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

In the first step, 24-nucleotide siRNAs are synthesized by the concerted actions of RNA POLYMERASE IV (Pol IV, also known as NRPD), RNA-DIRECTED RNA POLYMERASE 2 (RDR2) and DICER-LIKE 3 (DCL3)[58]. This key step in RdDM is dependent on the function of SAWADEE HOMEODOMAIN HOMOLOGUE 1 (SHH1)[59–61], which can bind to methylated H3K9 tails. The SHH1 protein was identified as a Pol IV-interacting factor that is required for the generation of siRNAs at a subset of sites[59] as well as in a forward genetic screen[62]. SHH1 contains an N-terminal homeodomain and a C-terminal SAWADEE domain[59] (FIG. 3b). Using a peptide chip, the SAWADEE domain of SHH1 was shown to specifically bind to the H3K9me2 mark and to unmodified H3K4 (REF. 60). The subsequent crystal structure determination of the SAWADEE domain of SHH1 revealed a tandem tudor-like fold, with a unique zinc-binding motif within the second tudor domain[60]. Further, the structure of SHH1 SAWADEE domain with bound H3K9me2 peptide revealed the molecular mechanism underlying specific recognition of this histone mark (FIG. 3c). The SAWADEE domain uses a classic three-aromatic-residue-lined cage to accommodate the H3K9me2 side chain (FIG. 3d) and a negatively charged pocket to specifically recognize unmodified H3K4 (REF. 60) (FIG. 3e). Mutations that disrupt either the K4 or the K9me2 binding pockets reduced DNA methylation at RdDM sites[60]. Thus, SHH1 can directly target Pol IV to unmethylated H3K4 (H3K4me0)-containing and H3K9me2-containing chromatin regions, leading to the production of siRNAs from these sites.

The second step in RdDM involves the production of scaffold transcripts by RNA POLYMERASE V (Pol V; also known as NRPE) with the help of the DDR complex (which is composed of DRD1, a SWI and SNF2 chromatin remodeller; DMS3, a chromosomal architectural protein; and RDM1, with unknown function)[63]. ARGONAUTE 4 (AGO4), loaded with 24-nucleotide siRNAs, is thought to bind the Pol V transcripts[3,64,65] and recruit DRM2 to chromatin[66]. Although this second step seems to be independent of histone methylation, it requires the catalytically inactive histone methyltransferases SUVH2 and SUVH9, which bind methylated DNA via their SRA domains[36,38], creating a self-reinforcing loop between RdDM and existing DNA methylation[37,39].

SUVH2 and SUVH9 have the same domain architecture (FIG. 3f) and redundant functions in the RdDM pathway[36]. They have an N-terminal SRA domain, which can recognize methylated DNA (modelled in FIG. 3g) and C-terminal pre-SET and SET domains, which were assumed to have histone methyltransferase activity[36,67]. Studies on the structure of SUVH9 confirmed that its SRA domain adopts a fold capable of binding methylated DNA, and that the histone methyltransferase domain contains a pre-SET and SET domains, but it lacks the post-SET domain and thus forms an incomplete substrate-binding pocket, resulting in a catalytically inactive enzyme[37]. Importantly, targeting SUVH2 with a sequence-specific zinc finger led to DNA methylation of an unmethylated epiallele, showing that the tethering of SUVH2 results in the recruitment of Pol V and subsequent DNA methylation[37]. These results suggest that DNA methylation, through SUVH2 binding, has a role in targeting Pol V.

The final step of RdDM is catalysed by the *de novo* methyltransferase DRM2. DRM2, which is found in all higher plants, possesses the signature motifs of other type I

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

methyltransferases, but their arrangement is different. DRM2 was identified in *Arabidopsis* spp. by genetic approaches and subsequent structural and functional studies have shed light on its molecular features and DNA targeting mechanism. The structure of its close homologue in *Nicotiana tabacum* DRM (which lacks the three UBA sequences towards its N-terminus; see FIG. 3h) was solved and found to have a dimer interface mimicking the mammalian DNA methyltransferase 3A (DNMT3A)–DNMT3L heterodimer interface[66] (FIG. 3i). The DNA methyltransferase domain of *N. tabacum* DRM retains the type I methyltransferase fold (FIG. 3j) despite rearrangement of methyltransferase sequence motifs[66] (FIG. 3j). Biochemical studies established that *Arabidopsis* spp. DRM2 exists in complex with AGO4 and preferentially methylates one DNA strand. These results support a model in which DRM2 is guided to target loci through base pairing of AGO4-associated siRNAs with nascent transcripts[66].

The removal of active marks such as H3K4me3 and histone acetylation may function as indirect regulators of DNA methylation[68–71]. Several groups have independently found that the histone H3K4me3 demethylase JMJ14 is a component of the RdDM pathway[68,69,71] (FIG. 3a). Mutations in JMJ14 do not affect *de novo* methylation by RdDM but instead affects maintenance methylation by RdDM[68]. More recently, the histone demethylases LDL1 and LDL2 were also shown to function in the RdDM pathway (FIG. 3a). The evidence suggests that they remove H3K4me2 and H3K4me3 to allow for SHH1 binding and the synthesis of Pol IV-dependent siRNAs[71]. Similarly, the removal of histone acetylation can be linked to RdDM through HDA6 (REFS 72–75). HDA6 was recently shown to function upstream of Pol IV and siRNA generation, again suggesting that the active mark must be erased before the silencing signal is generated[76].

The dependence of RdDM on both histone methylation (via SHH1) and DNA methylation (via SUVH2 and SUVH9) is readily observed in loss-of-function mutant plants. All RdDM sites lose both DNA methylation and H3K9 methylation in mutants lacking the H3K9 methyltransferases (in plants with mutations in *kyp*, *suvh5* and *suvh6*) or lacking the non-CG methyltransferases (in ddcc mutants). The dependence of SHH1 on H3K9 methylation, presumably, explains the drastic reduction in 24-nucleotide siRNAs in both the ddcc mutant and the *kyp suvh5 suvh6* triple-mutant plants[40].

MET1 methylation is independent of histone methylation

CG methylation in *Arabidopsis* spp. is maintained by the enzyme MET1 (REF. 77) and is dependent on three redundant variant in methylation (VIM) proteins[78,79]. The VIM proteins are ubiquitin E3 ligases that also contain an SRA domain (that specifically binds hemimethylated DNA) and a PHD domain of unknown function[48,80]. Homologues in mammals (such as UHRF1) have been investigated in more detail and are discussed below.

The observation that gene body methylation is not associated with H3K9me was the first indication that CG methylation is independent of H3K9me in plants[35]. However, CG methylation may attract some H3K9 methylation that is subsequently removed by the active histone demethylase IBM1 (REF. 81). Genome-wide bisulphite sequencing in *kyp suvh5 suvh6* mutant plants was found to have little effect on CG methylation, indicating H3K9 methylation is not required for targeting CG methylation[56]. Loss of MET1 activity in a *met1*

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

knockout is more complicated to interpret as it causes a reduction in non-CG methylation, and subsequently in H3K9me[40,45,47,56]. The exact mechanisms coupling of CG and non-CG methylation are not understood.

DNA and histone methylation in mammals

DNA methylation in mammals occurs primarily at CG residues; non-CG methylation is observed only in stem cells in the body of actively transcribed genes[82–85]. Genome wide, 60–80% of the CG residues are methylated. However, in CpG islands and active regulatory regions only 10% of the CGs are methylated[84,86]. These active promoters are protected from methylation, whereas other promoters are repressed by methylation during differentiation (see below). Methylation of repetitive DNA, which is found near centromeres and dispersed throughout the genome, is extremely important in maintaining genome integrity. In mice, two types of repetitive DNA are found near centromeres: major satellites, in the pericentromeric region, and minor satellites, in the centromeric region[87]. The main classes of dispersed repetitive sequences include LINEs and SINEs (long and short interspersed nuclear elements) and long terminal repeat-containing endogenous retroviruses (ERVs).

There are three active DNA methyltransferases in mammals: DNMT1, DNMT3A and DNMT3B[88]. DNMT1 is primarily a maintenance methyltransferase and DNMT3A and DNMT3B are primarily *de novo* methyltransferases[89]; however, it is clear that these distinctions are not absolute[90]. Depending on the type of repetitive element, DNMT1 can exhibit *de novo* activity and DNMT3A or DNMT3B may also be required for maintenance[90].

In mammals, DNA and H3K9 methylation are strongly associated[85,91]. H3K9 methylation is catalysed by one of five members of the SET-containing SUV39 protein family: SUV39H1, SUV39H2, G9A, GLP, and SETDB1 (REFS 92–94). G9A and GLP catalyse mono- and dimethylation of H3K9 primarily found associated with silent genes in euchromatin[95], while SUV39H1 and SUV39H2 are trimethyltransferases responsible primarily for centromeric and pericentromeric heterochromatin[95–97]. SETDB1 (also known as ESET) is an H3K9 trimethyltransferase responsible for methylating ERVs and the inactive X chromosome[98,99]. Some of the histone methyltransferases contain domains that are also important for their targeting. SUV39H1 and SUV39H2 each contain a chromodomain, which binds H3K9me3 (REFS 92,100), and G9a and GLP proteins contain ankyrin repeats, which bind H3K9me1 or H3K9me2 (REF. 101). These enzymes therefore bind to the mark that they create on chromatin, facilitating a feedforward loop. Although the targets of these enzymes seem distinct, there is some evidence that at times they may act together in a single complex[102].

Mutant analysis in mouse embryonic stem (ES) cells revealed that *Suv39H1*$^{-/-}$*Suv39H2*$^{-/-}$ mice had reduced DNA methylation in major satellites but not minor satellites or C-type retroviruses[90,103], knockout of *G9a* in mouse ES cells resulted in DNA hypomethylation at specific loci throughout the genome[104,105], and knockout of *Setdb1* resulted in minor loss of DNA methylation at a subset of loci including imprinted genes[99,106]. No effect on H3K9 methylation was observed in *Dnmt1*$^{-/-}$*Dnmt3a*$^{-/-}$*Dnmt3b*$^{-/-}$ mouse ES cells (the three DNA methyltransferase genes)[107,108]. However H3K9 methylation was found to be dependent on

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

DNA methylation in human cancer cells[109–111]. These differences may reflect the fact that mouse ES cells are undifferentiated and utilize different mechanisms for maintaining H3K9 methylation. As described below, there are numerous proteins involved in linking these two marks to the same genetic targets.

Establishment of DNA Methylation

Two waves of global demethylation occur in mammals: one in early embryogenesis and the other during primordial germ cell (PGC) specification[3]. Global *de novo* DNA methylation takes place around the time of embryo implantation and is accomplished through the activities of DNMT3A and DNMT3B. In PGCs, *de novo* DNA methylation is crucial for establishment of imprints and requires the catalytically inactive homologue DNMT3L[112]. Multiple mechanisms are involved in the initial establishment of methylation — some are independent, and others dependent, of H3K9 methylation[113].

At active promoters, CpG islands are protected from methylation by binding of transcription factors and recruitment of H3K4 methyltransferases[113]. The DNMT3 enzymes each contain an ADD domain (ATRX–DNMT3–DNMT3L) that recognizes unmodified H3 and is inhibited by H3K4 methylation[114–118] (FIG. 4a,b). Genetic evidence for the inhibitory effect of H3K4 methylation is also observed at imprinted genes, which fail to become methylated in cells reduced in a H3K4 demethylase[119].

During differentiation of mouse ES cells, some gene promoters undergo DNA methylation that is dependent on G9a or GLP. 126 genes in this category have recently been identified[120]. G9A and GLP-dependent H3K9me2 appears before DNA methylation, and DNA methylation is lost in cells with mutations in *G9a* or *Glp*. G9A and GLP can recruit DNMT3A and DNMT3B directly[120] or indirectly through the chromodomain protein MPP8, resulting in *de novo* DNA methylation[121,122]. These interactions nicely explain the observation that DNA methylation is dependent on G9A, even in the absence of its catalytic activity[105,120].

The minor satellites found in the centromeres are associated with centromeric proteins (CENP)[87]. CENP-B (centromeric protein B) binds a specific sequence known as the CENP-B box[123]. Integration of naked DNA containing 32 copies of the human CENP-B box in mouse embryonic fibroblast (MEF) cells was shown to recruit SUV39H1 or SUV39H2 histone methyltransferases and lead to H3K9 trimethylation[123]. CENP-B is in a complex with CENP-A and CENP-C[124] and subsequent studies revealed that DNMT3B interacts directly with CENP-C. This targeting of DNMT3B by CENP-C results in DNA methylation independent of histone H3K9 methylation[125]. This is consistent with the observation that knocking out *Suv39h1* and *Suv39h2* in mouse ES cells does not affect DNA methylation in minor satellite repeats[90,103].

In the major satellite repeats, SUV39H1 or SUV39H2 recruit DNMT3A directly, and this DNA methylation is lost in *Suv39h1*$^{-/-}$*Suv39h2*$^{-/-}$ mice[90,103,126]. Major satellites are also enriched in HP1, which has been shown to recruit DNMTs as well, which provides an additional method by which DNA methylation can be targeted to regions enriched in H3K9 methylation[126,127]. HP1 not only binds H3K9me, but has been shown to interact with G9a

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

and SUV39H1 or SUV39H2 (REFS 126,128). However, using a mutant *Suv39h1* in which the HP1-interacting region was deleted, it was found that both H3K9me3 and DNA methylation could be restored to *Suv39h1*$^{-/-}$*Suv39h2*$^{-/-}$ ES cells without restoration of HP1 binding[128]. Moreover, SUV39H1 and SUV39H2 are stable components of heterochromatin, whereas HP1 has a rapid on–off rate[129]. These results suggest that HP1 may act downstream of both H3K9me3 and DNA methylation unlike in *N. crassa*.

DNA methylation is established at a large subset of retroelements and retroviruses using a very different pathway. In early embryos, KAP (KRAB-associated protein A; also known as TRIM28) is targeted to specific sequences through zinc-finger proteins (such as KRAB–ZFP809) and recruits SETDB1, the H3K9 trimethyltransferase[108,130–132]. Enrichment of H3K9me3 is found even in *Dnmt1*$^{-/-}$*Dnmt3a*$^{-/-}$*Dnmt3b*$^{-/-}$ mutant ES cells, indicating DNA methylation is not required for recruitment of SETDB1 (REF. 108). Once established, though, DNMT1 and DNMT3B take over maintenance methylation[133]. At this point, KAP and SETDB1 are no longer required for silencing and H3K9me3 is lost[91,108,134,135].

DNA methylation in gene bodies is associated with high levels of DNMT3B enrichment and association with RNA polymerase II[136]. This targeting may also involve the PWWP domain binding to H3K36me3, which is tied to transcription[4,137,138].

De novo methylation is tightly linked to unmethylated H3K4

The ADD domain found in DNMT3A, DNMT3B and DNMT3L specifically recognizes unmethylated H3K4 (REFS 114, 115, 117, 118) (FIG. 4a,b). DNMT3A and DNMT3L form a DNMT3L–DNMT3A–DNMT3A–DNMT3L tetramer (FIG. 4c) that, when modelled on nucleosomal DNA, positions the two DNMT3A active sites on adjacent DNA major grooves[139]. Such an alignment could facilitate DNMT3A-mediated methylation of a pair of CpG sites separated by one helical turn[139], consistent with the observed 10 bp methylation periodicity[140].

The multiple ADD domains within the tetramer facilitate scanning and capturing the H3K4me0 state, coupling the reading of H3K4me0 and DNA methylation establishment. In addition to the ADD domain, DNMT3A and DNMT3B possess the PWWP domain at their N terminus[4] (FIG. 4b). The DNMT3A PWWP domain recognizes the H3K36me3 mark[5,85]. However, in mouse ES cells, using chromatin immuno-precipitation of tagged DNMT3A and DNMT3B, or in *Saccharomyces cerevisiae* that expresses a heterologous DNMT3B, it was found that DNMT3B, not DNMT3A, is specifically enriched at H3K36me3 sites and that recruitment is based on binding affinity[137,138].

It has been reported that the histone tail can stimulate the enzymatic activity of mammalian *de novo* DNA methyltransferase DNMT3A[70,141]. Recent structural studies on this system revealed that in the absence of the H3 tail, the ADD domain of DNMT3A specifically interacts with the catalytic domain, with the binding interface positioned on one face of the active site[142] (FIG. 4d). In such an alignment, the ADD domain blocks access of the substrate DNA to the active site of the catalytic domain, reflecting an auto-inhibitory mode of the enzyme[142] (FIG. 4e). When the ADD domain interacts with an H3K4me0 (but no H3K4me3) peptide, DNMT3A undergoes a substantial conformational change, which

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

exposes two acidic residues that are important for mediating the interaction between the ADD domain and the catalytic domain in the auto-inhibitory conformation. As a result, the H3 peptide-bound ADD domain interacts with another face on the catalytic domain, resulting in the release of auto-inhibition (FIG. 4f). The H3K4me0 peptide-bound ADD domain allows access of the DNA to the catalytic site, thereby facilitating the methylation reaction, and revealing the allosteric regulatory role of histone H3 in DNA methylation[142].

Maintenance DNA Methylation by DNMT1

Maintenance DNA methylation takes place at replication foci shortly after the DNA is replicated. DNMT1 is primarily responsible for this methylation and is recruited to replication foci by PCNA and other factors[143]. Direct interactions between DNMT1 and SUV39H1, SUV39H2 or G9a may play a role in targeting both types of histone methyltransferases to the appropriate sites during replication[143]. In addition to these direct interactions, adaptor proteins have been found to be essential for maintaining DNA methylatio n in the appropriate regions.

Structural studies of DNMT1 have established how a combination of active and auto-inhibitory mechanisms ensures the high fidelity of DNMT1-mediated maintenance DNA methylation[144,145]. DNMT1 has a replication foci domain (RFD), a CXXC domain and two BAH domains in addition to the DNA methyltransferase domain (FIG. 5a). Structural studies on DNMT1 using a construct lacking the RFD revealed that the CXXC domain of DNMT1 can specifically recognize unmethylated CpG DNA (FIG. 5b), thereby positioning a loop that connects the CXXC and BAH1 domains between the active site of the DNA methyltransferase domain and the DNA. Such an alignment constitutes an auto-inhibitory conformation preventing potential *de novo* methylation activity[144]. In a construct lacking both the RFD and CXXC domains, the DNA targets hemimethylated DNA, with the cytosine in the target strand looped out and anchored in the catalytic pocket[145] (FIG. 5c). In the structure of this complex, side chains from catalytic and recognition loops insert from both grooves to fill an intercalation site cavity associated with a dual base flip-out on partner strands (FIG. 5d). The DNA is positioned outside the binding pocket in the auto-inhibitory complex[144], while it fits snugly within the binding channel in the productive complex[145] (FIG. 5e). By contrast, in the absence of DNA substrate, the N-terminal RFD domain of DNMT1, which targets the protein to the replication fork, blocks the DNA substrate binding site of the DNA methyltransferase domain (FIG. 5f), achieving an auto-inhibitory effect in the free state[146].

The C-terminal catalytic cassette of DNMT1 is composed of two tandem BAH domains of as yet unknown function and the DNA methyltransferase domain (FIG. 5a) which can convert hemimethylated CG DNA to full methylated DNA using the methylated parental strand as a guide to target the daughter strand[145]. To date, there is no direct evidence supporting interaction between histones and DNMT1, although the BAH1 domain of DNMT1 contains an aromatic cage[144], and BAH domains have been identified as histone binding modules for methylated lysine residues[147,148].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

UHRF1 is an adaptor between histone methylation and DNMT1

The VIM homologue UHRF1 (ubiquitin-like, containing PHD and RING finger domains 1; also known as ICBP90 in humans and NP95 in mice) is a key adaptor protein. Early studies showed that knockout of *Uhrf1* resulted in reduction in DNA methylation similar to knockout of *Dnmt1* (REFS 149–151), indicating that UHRF1 is required for DNMT1 function (FIG. 5g). More recently, bisulphite sequencing in mouse ES cells revealed that knockout of *Uhrf1* was more effective than knockout of *Dnmt1* at reducing methylation, suggesting it may be functioning with other methyltransferases in addition to DNMT1 (REF. 90). This multi-modular protein contains five recognizable domains: PHD, tandem tudor domain, SRA, RING (real interesting new gene), and UBL (ubiquitin-like) (FIG. 5h). The tandem tudor and PHD finger domains act together to function as a histone-reader cassette. The tandem tudor domains specifically recognize the H3K9me3 mark using a classical aromatic cage recognition mode (FIG. 5i), while the PHD finger both assists the tandem tudor recognition of H3K9me3 and recognizes unmodified H3R2 (REFS 152–158) (FIG. 5i). A recent report biochemically identified a lipid molecule (phosphatidylinositol 5-phosphate) bound to UHRF1, and showed that it could regulate the cooperative binding by the tandem tudor and PHD finger domains[159]. In the absence of phosphatidyl-inositol 5-phosphate, UHRF1 recognizes the unmodified H3 tail predominately through the PHD finger[159]. Phosphatidylinositol 5-phosphate allosterically regulates UHRF1 so that the tandem tudor domain preferentially binds to the H3K9me3 mark[159], revealing a dynamic regulation of UHRF1 binding specificity.

In addition to binding to the methylated H3K9 tail (FIG. 5i), UHRF1 binds hemimethylated CG residues generated at replication foci via its SRA domain[149,150,160–163] (FIG. 5j). A number of groups were able to detect direct interactions between UHRF1 and DNMT1 suggesting a direct recruitment and activation model[164–167]. However, an alternative model was proposed when it was discovered that the UHRF1 RING domain functions to ubiquinate H3K23 and is required for the recruitment of DNMT1 to chromatin[151,168,169]. The second model proposes that UHRF1 recognizes hemimethylated DNA that is bound by H3K9me3-containing nucleosomes and ubiquinates H3K23 (REF. 169). DNMT1 then binds ubiquinated H3K23 through its RFD domain[170], which induces a conformational change in DNMT1 that promotes its activation[169].

Conclusions

Histone and DNA methylation have important and connected roles in the epigenetic control of gene expression in all three kingdoms of eukaryotic organisms. In some cases the relationships between these two epigenetic marks are linear. For example, histone methylation in *N. crassa*. is clearly upstream of DNA methylation. In other cases the relationships are more interdependent. In plants, for example, histone and DNA methylation are linked in a codependent feedforward loop, and RNA-directed DNA methylation both promotes, and is dependent on, histone and DNA methylation through self-reinforcing loop mechanisms. In mammals the situation is more complex. DNA methylation in some genomic sites is dependent on histone methylation, whereas at other sites histone and DNA methylation occur independently; at yet other sites there is evidence of self-reinforcing

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

loops. Thus, whereas links between histone and DNA methylation are present in fungi, plants, and animals, the relationships vary widely.

Although much has been learned about the mechanisms by which histone and DNA methylation control gene expression, there are many aspects yet to be uncovered. For example, although we have structural information about the interaction of specific chromatin domains with particular epigenetic modifications, there are still very few studies of the interaction of chromatin proteins with their native substrate, the nucleosome. In addition, for DNA methyltransferase enzymes that are stably localized to chromatin, such as CMT3 in plants and DNMT3s in mammals, it is not clear if these proteins play roles in chromatin compaction separate from their roles in modifying DNA. More generally, while tremendous progress has been made in understanding the enzymes controlling epigenetic marks, much less is known about the processes downstream of these marks that ultimately control the activation or repression of genes.

Acknowledgments

The authors thank W. Pastor and E. Selker for helpful comments on the manuscript. This work was supported by the Chinese Academy of Sciences (to J.D.), NIH grant GM60398 (to S.E.J.), and LLSCOR Program Project and STARR Foundation grants (to D.J.P.). S.E.J. is an investigator of the Howard Hughes Medical Institute.

References

1. Grewal SI, Jia S. Heterochromatin revisited. Nat Rev Genet. 2007; 8:35–46. [PubMed: 17173056]
2. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet. 2012; 13:484–492. [PubMed: 22641018]
3. Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet. 2010; 11:204–220. [PubMed: 20142834]
4. Dhayalan A, et al. The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. J Biol Chem. 2010; 285:26114–26120. [PubMed: 20547484]
5. Hodges E, et al. High definition profiling of mammalian DNA methylation by array capture and single molecule bisulfite sequencing. Genome Res. 2009; 19:1593–1605. [PubMed: 19581485]
6. Boulard M, Edwards JR, Bestor TH. FBXL10 protects Polycomb-bound genes from hypermethylation. Nat Genet. 2015; 47:479–485. [PubMed: 25848754]
7. Saksouk N, et al. Redundant mechanisms to form silent chromatin at pericentromeric regions rely on BEND3 and DNA methylation. Mol Cell. 2014; 56:580–594. [PubMed: 25457167]
8. Selker EU, Cambareri EB, Jensen BC, Haack KR. Rearrangement of duplicated DNA in specialized cells of *Neurospora*. Cell. 1987; 51:741–752. [PubMed: 2960455]
9. Selker EU, Garrett PW. DNA sequence duplications trigger gene inactivation in *Neurospora crassa*. Proc Natl Acad Sci USA. 1988; 85:6870–6874. [PubMed: 2842795]
10. Selker EU. Premeiotic instability of repeated sequences in *Neurospora crassa*. Annu Rev Genet. 1990; 24:579–613. [PubMed: 2150906]
11. Cambareri EB, Jensen BC, Schabtach E, Selker EU. Repeat-induced G-C to A-T mutations in *Neurospora*. Science. 1989; 244:1571–1575. [PubMed: 2544994]
12. Belden WJ, Lewis ZA, Selker EU, Loros JJ, Dunlap JC. CHD1 remodels chromatin and influences transient DNA methylation at the clock gene frequency. PLoS Genet. 2011; 7:e1002166. [PubMed: 21811413]
13. Dang Y, Li L, Guo W, Xue Z, Liu Y. Convergent transcription induces dynamic DNA methylation at disiRNA loci. PLoS Genet. 2013; 9:e1003761. [PubMed: 24039604]
14. Selker EU, Fritz DY, Singer MJ. Dense nonsymmetrical DNA methylation resulting from repeat-induced point mutation in *Neurospora*. Science. 1993; 262:1724–1728. [PubMed: 8259516]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

15. Selker EU, Stevens JN. DNA methylation at asymmetric sites is associated with numerous transition mutations. Proc Natl Acad Sci USA. 1985; 82:8114–8118. [PubMed: 2415981]
16. Selker EU, et al. Induction and maintenance of nonsymmetrical DNA methylation in *Neurospora*. Proc Natl Acad Sci USA. 2002; 99(Suppl. 4):16485–16490. [PubMed: 12189210]
17. Lewis ZA, et al. Relics of repeat-induced point mutation direct heterochromatin formation in *Neurospora crassa*. Genome Res. 2009; 19:427–437. [PubMed: 19092133]
18. Tamaru H, Selker EU. A histone H3 methyltransferase controls DNA methylation in *Neurospora crassa*. Nature. 2001; 414:277–283. [PubMed: 11713521]
19. Tamaru H, et al. Trimethylated lysine 9 of histone H3 is a mark for DNA methylation in *Neurospora crassa*. Nat Genet. 2003; 34:75–79. [PubMed: 12679815]
20. Lewis ZA, et al. DNA methylation and normal chromosome behavior in *Neurospora* depend on five components of a histone methyltransferase complex, DCDC. PLoS Genet. 2010; 6:e1001196. [PubMed: 21079689]
21. Lewis ZA, Adhvaryu KK, Honda S, Shiver AL, Selker EU. Identification of DIM-7, a protein required to target the DIM-5 H3 methyltransferase to chromatin. Proc Natl Acad Sci USA. 2010; 107:8310–8315. [PubMed: 20404183]
22. Honda S, et al. The DMM complex prevents spreading of DNA methylation from transposons to nearby genes in *Neurospora crassa*. Genes Dev. 2010; 24:443–454. [PubMed: 20139222]
23. Honda S, et al. Heterochromatin protein 1 forms distinct complexes to direct histone deacetylation and DNA methylation. Nat Struct Mol Biol. 2012; 19:471–477. [PubMed: 22504884]
24. Honda S, Selker EU. Direct interaction between DNA methyltransferase DIM-2 and HP1 is required for DNA methylation in *Neurospora crassa*. Mol Cell Biol. 2008; 28:6044–6055. [PubMed: 18678653]
25. Freitag M, Hickey PC, Khlafallah TK, Read ND, Selker EU. HP1 is essential for DNA methylation in *Neurospora*. Mol Cell. 2004; 13:427–434. [PubMed: 14967149]
26. Cokus SJ, et al. Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. Nature. 2008; 452:215–219. [PubMed: 18278030]
27. Lister R, et al. Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. Cell. 2008; 133:523–536. [PubMed: 18423832]
28. Zhang X, et al. Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. Cell. 2006; 126:1189–1201. [PubMed: 16949657]
29. Zhao Z, Shen WH. Plants contain a high number of proteins showing sequence similarity to the animal SUV39H family of histone methyltransferases. Ann NY Acad Sci. 2004; 1030:661–669. [PubMed: 15659850]
30. Cao X, Jacobsen SE. Role of the arabidopsis DRM methyltransferases in *de novo* DNA methylation and gene silencing. Curr Biol. 2002; 12:1138–1144. [PubMed: 12121623]
31. Bartee L, Malagnac F, Bender J. *Arabidopsis cmt3* chromomethylase mutations block non-CG methylation and silencing of an endogenous gene. Genes Dev. 2001; 15:1753–1758. [PubMed: 11459824]
32. Lindroth AM, et al. Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. Science. 2001; 292:2077–2080. [PubMed: 11349138]
33. Finnegan EJ, Peacock WJ, Dennis ES. Reduced DNA methylation in *Arabidopsis thaliana* results in abnormal plant development. Proc Natl Acad Sci USA. 1996; 93:8449–8454. [PubMed: 8710891]
34. Ebbs ML, Bender J. Locus-specific control of DNA methylation by the *Arabidopsis* SUVH5 histone methyltransferase. Plant Cell. 2006; 18:1166–1176. [PubMed: 16582009]
35. Bernatavichute YV, Zhang X, Cokus S, Pellegrini M, Jacobsen SE. Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in *Arabidopsis thaliana*. PloS ONE. 2008; 3:e3156. [PubMed: 18776934]
36. Johnson LM, Law JA, Khattar A, Henderson IR, Jacobsen SE. SRA-domain proteins required for DRM2-mediated *de novo* DNA methylation. PLoS Genet. 2008; 4:e1000280. [PubMed: 19043555]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

37. Johnson LM, et al. SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. Nature. 2014; 507:124–128. [PubMed: 24463519]
38. Kuhlmann M, Mette MF. Developmentally non-redundant SET domain proteins SUVH2 and SUVH9 are required for transcriptional gene silencing in *Arabidopsis thaliana*. Plant Mol Biol. 2012; 79:623–633. [PubMed: 22669745]
39. Liu ZW, et al. The SET domain proteins SUVH2 and SUVH9 are required for Pol V occupancy at RNA-directed DNA methylation loci. PLoS Genet. 2014; 10:e1003948. [PubMed: 24465213]
40. Stroud H, et al. Non-CG methylation patterns shape the epigenetic landscape in *Arabidopsis*. Nat Struct Mol Biol. 2014; 21:64–72. This paper describes the tight links between non-CG methylation and both H3K9 methylation and siRNAs. [PubMed: 24336224]
41. Du J, et al. Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. Cell. 2012; 151:167–180. This paper describes structural and functional insights underlying a dual-binding mode of histone-regulating DNA methyltransferase. [PubMed: 23021223]
42. Jackson JP, Lindroth AM, Cao X, Jacobsen SE. Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature. 2002; 416:556–560. [PubMed: 11898023]
43. Malagnac F, Bartee L, Bender J. An *Arabidopsis* SET domain protein required for maintenance but not establishment of DNA methylation. EMBO J. 2002; 21:6842–6852. [PubMed: 12486005]
44. Tran RK, et al. Chromatin and siRNA pathways cooperate to maintain DNA methylation of small transposable elements in *Arabidopsis*. Genome Biol. 2005; 6:R90. [PubMed: 16277745]
45. Soppe WJ, et al. DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in *Arabidopsis*. EMBO J. 2002; 21:6549–6559. [PubMed: 12456661]
46. Mathieu O, Probst AV, Paszkowski J. Distinct regulation of histone H3 methylation at lysines 27 and 9 by CpG methylation in *Arabidopsis*. EMBO J. 2005; 24:2783–2791. [PubMed: 16001083]
47. Tariq M, et al. Erasure of CpG methylation in *Arabidopsis* alters patterns of histone H3 methylation in heterochromatin. Proc Natl Acad Sci USA. 2003; 100:8823–8827. [PubMed: 12853574]
48. Johnson LM, et al. The SRA methyl-cytosine-binding domain links DNA and histone methylation. Curr Biol. 2007; 17:379–384. [PubMed: 17239600]
49. Inagaki S, et al. Autocatalytic differentiation of epigenetic modifications within the *Arabidopsis* genome. EMBO J. 2010; 29:3496–3506. [PubMed: 20834229]
50. Henikoff S, Comai LA. DNA methyltransferase homolog with a chromodomain exists in multiple polymorphic forms in *Arabidopsis*. Genetics. 1998; 149:307–318. [PubMed: 9584105]
51. Finnegan EJ, Kovac KA. Plant DNA methyltransferases. Plant Mol Biol. 2000; 43:189–201. [PubMed: 10999404]
52. Papa CM, Springer NM, Muszynski MG, Meeley R, Kaeppler SM. Maize chromomethylase *Zea methyltransferase2* is required for CpNpG methylation. Plant Cell. 2001; 13:1919–1928. [PubMed: 11487702]
53. Patel DJ, Wang Z. Readout of epigenetic modifications. Annu Rev Biochem. 2013; 82:81–118. [PubMed: 23642229]
54. Du J, et al. Mechanism of DNA methylation-directed histone methylation by KRYPTONITE. Mol Cell. 2014; 55:495–504. This paper outlines the structural basis underlying methylated DNA-mediated regulation of lysine methylation by a histone lysine methyltransferase. [PubMed: 25018018]
55. Zemach A, et al. The *Arabidopsis* nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell. 2013; 153:193–205. [PubMed: 23540698]
56. Stroud H, Greenberg MV, Feng S, Bernatavichute YV, Jacobsen SE. Comprehensive analysis of silencing mutants reveals complex regulation of the *Arabidopsis* methylome. Cell. 2013; 152:352–364. [PubMed: 23313553]
57. Matzke MA, Mosher RA. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nat Rev Genet. 2014; 15:394–408. [PubMed: 24805120]
58. Pontier D, et al. Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in *Arabidopsis*. Genes Dev. 2005; 19:2030–2040. [PubMed: 16140984]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

59. Law JA, Vashisht AA, Wohlschlegel JA, Jacobsen SE. SHH1, a homeodomain protein required for DNA methylation, as well as RDR2, RDM4, and chromatin remodeling factors, associate with RNA polymerase IV. PLoS Genet. 2011; 7:e1002195. [PubMed: 21811420]
60. Law JA, et al. Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. Nature. 2013; 498:385–389. This paper elucidated the structure-function links between RdDM and the H3K9me mark. [PubMed: 23636332]
61. Zhang H, et al. DTF1 is a core component of RNA-directed DNA methylation and may assist in the recruitment of Pol IV. Proc Natl Acad Sci USA. 2013; 110:8290–8295. [PubMed: 23637343]
62. Liu J, et al. An atypical component of RNA-directed DNA methylation machinery has both DNA methylation-dependent and -independent roles in locus-specific transcriptional gene silencing. Cell Res. 2011; 21:1691–1700. [PubMed: 22064704]
63. Law JA, et al. A protein complex required for polymerase V transcripts and RNA- directed DNA methylation in *Arabidopsis*. Curr Biol. 2010; 20:951–956. [PubMed: 20409711]
64. Pikaard CS, Haag JR, Ream T, Wierzbicki AT. Roles of RNA polymerase IV in gene silencing. Trends Plant Sci. 2008; 13:390–397. [PubMed: 18514566]
65. Wierzbicki AT, Ream TS, Haag JR, Pikaard CS. RNA polymerase V transcription guides ARGONAUTE4 to chromatin. Nat Genet. 2009; 41:630–634. [PubMed: 19377477]
66. Zhong X, et al. Molecular mechanism of action of plant DRM *de novo* DNA methyltransferases. Cell. 2014; 157:1050–1060. [PubMed: 24855943]
67. Naumann K, et al. Pivotal role of AtSUVH2 in heterochromatic histone methylation and gene silencing in *Arabidopsis*. EMBO J. 2005; 24:1418–1429. [PubMed: 15775980]
68. Deleris A, et al. Involvement of a Jumonji-C domain-containing histone demethylase in DRM2-mediated maintenance of DNA methylation. EMBO Rep. 2010; 11:950–955. [PubMed: 21052090]
69. Searle IR, Pontes O, Melnyk CW, Smith LM, Baulcombe DC. JMJ14, a JmjC domain protein, is required for RNA silencing and cell-to-cell movement of an RNA silencing signal in *Arabidopsis*. Genes Dev. 2010; 24:986–991. [PubMed: 20478993]
70. Zhang Y, et al. Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. Nucleic Acids Res. 2010; 38:4246–4253. [PubMed: 20223770]
71. Greenberg MV, et al. Interplay between active chromatin marks and RNA-directed DNA methylation in *Arabidopsis thaliana*. PLoS Genet. 2013; 9:e1003946. [PubMed: 24244201]
72. Aufsatz W, Mette MF, van der Winden J, Matzke M, Matzke AJ. HDA6, a putative histone deacetylase needed to enhance DNA methylation induced by double-stranded RNA. EMBO J. 2002; 21:6832–6841. [PubMed: 12486004]
73. Earley K, et al. Erasure of histone acetylation by *Arabidopsis* HDA6 mediates large-scale gene silencing in nucleolar dominance. Genes Dev. 2006; 20:1283–1293. [PubMed: 16648464]
74. Probst AV, et al. *Arabidopsis* histone deacetylase HDA6 is required for maintenance of transcriptional gene silencing and determines nuclear organization of rDNA repeats. Plant Cell. 2004; 16:1021–1034. [PubMed: 15037732]
75. Aufsatz W, Stoiber T, Rakic B, Naumann K. *Arabidopsis* histone deacetylase 6: a green link to RNA silencing. Oncogene. 2007; 26:5477–5488. [PubMed: 17694088]
76. Blevins T, et al. A two-step process for epigenetic inheritance in *Arabidopsis*. Mol Cell. 2014; 54:30–42. [PubMed: 24657166]
77. Kankel MW, et al. *Arabidopsis* MET1 cytosine methyltransferase mutants. Genetics. 2003; 163:1109–1122. [PubMed: 12663548]
78. Woo HR, Pontes O, Pikaard CS, Richards EJ. VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization. Genes Dev. 2007; 21:267–277. [PubMed: 17242155]
79. Woo HR, Dittmer TA, Richards EJ. Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in *Arabidopsis*. PLoS Genet. 2008; 4:e1000156. [PubMed: 18704160]
80. Kraft E, Bostick M, Jacobsen SE, Callis J. ORTH/VIM proteins that regulate DNA methylation are functional ubiquitin E3 ligases. Plant J. 2008; 56:704–715. [PubMed: 18643997]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

81. Rigal M, Kevei Z, Pelissier T, Mathieu O. DNA methylation in an intron of the IBM1 histone demethylase gene stabilizes chromatin modification patterns. EMBO J. 2012; 31:2981–2993. [PubMed: 22580822]
82. Lister R, et al. Global epigenomic reconfiguration during mammalian brain development. Science. 2013; 341:1237905. [PubMed: 23828890]
83. Varley KE, et al. Dynamic DNA methylation across diverse human cell lines and tissues. Genome Res. 2013; 23:555–567. [PubMed: 23325432]
84. Lister R, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature. 2009; 462:315–322. [PubMed: 19829295]
85. Meissner A, et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature. 2008; 454:766–770. [PubMed: 18600261]
86. Stadler MB, et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature. 2011; 480:490–495. [PubMed: 22170606]
87. Guenatri M, Bailly D, Maison C, Almouzni G. Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. J Cell Biol. 2004; 166:493–505. [PubMed: 15302854]
88. Robertson KD, Wolffe AP. DNA methylation in health and disease. Nat Rev Genet. 2000; 1:11–19. [PubMed: 11262868]
89. Cheng X. Structural and functional coordination of DNA and histone methylation. Cold Spring Harb Perspect Biol. 2014; 6:a018747. [PubMed: 25085914]
90. Arand J, et al. *In vivo* control of CpG and non-CpG DNA methylation by DNA methyltransferases. PLoS Genet. 2012; 8:e1002750. [PubMed: 22761581]
91. Mikkelsen TS, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature. 2007; 448:553–560. [PubMed: 17603471]
92. Rea S, et al. Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature. 2000; 406:593–599. [PubMed: 10949293]
93. Tachibana M, Sugimoto K, Fukushima T, Shinkai Y. Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. J Biol Chem. 2001; 276:25309–25317. [PubMed: 11316813]
94. Schultz DC, Ayyanathan K, Negorev D, Maul GG, Rauscher FJ 3rd. SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. Genes Dev. 2002; 16:919–932. [PubMed: 11959841]
95. Shinkai Y, Tachibana M. H3K9 methyltransferase G9a and the related molecule GLP. Genes Dev. 2011; 25:781–788. [PubMed: 21498567]
96. Peters AH, et al. Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. Mol Cell. 2003; 12:1577–1589. [PubMed: 14690609]
97. Martin C, Zhang Y. The diverse functions of histone lysine methylation. Nat Rev Mol Cell Biol. 2005; 6:838–849. [PubMed: 16261189]
98. Minkovsky A, et al. The *Mbd1–Atf7ip–Setdb1* pathway contributes to the maintenance of X chromosome inactivation. Epigenetics Chromatin. 2014; 7:12. [PubMed: 25028596]
99. Karimi MM, et al. DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. Cell Stem Cell. 2011; 8:676–687. [PubMed: 21624812]
100. Firestein R, Cui X, Huie P, Cleary ML. Set domain-dependent regulation of transcriptional silencing and growth control by SUV39H1, a mammalian ortholog of *Drosophila* Su(var)3-9. Mol Cell Biol. 2000; 20:4900–4909. [PubMed: 10848615]
101. Collins RE, et al. The ankyrin repeats of G9a and GLP histone methyltransferases are mono- and dimethyllysine binding modules. Nat Struct Mol Biol. 2008; 15:245–250. [PubMed: 18264113]
102. Fritsch L, et al. A subset of the histone H3 lysine 9 methyltransferases Suv39h1, G9a, GLP, and SETDB1 participate in a multimeric complex. Mol Cell. 2010; 37:46–56. [PubMed: 20129054]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

103. Lehnertz B, et al. Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Curr Biol. 2003; 13:1192–1200. [PubMed: 12867029]
104. Ikegami K, et al. Genome-wide and locus-specific DNA hypomethylation in G9a deficient mouse embryonic stem cells. Genes Cells. 2007; 12:1–11. [PubMed: 17212651]
105. Dong KB, et al. DNA methylation in ES cells requires the lysine methyltransferase G9a but not its catalytic activity. EMBO J. 2008; 27:2691–2701. [PubMed: 18818693]
106. Leung D, et al. Regulation of DNA methylation turnover at LTR retrotransposons and imprinted loci by the histone methyltransferase Setdb1. Proc Natl Acad Sci USA. 2014; 111:6690–6695. [PubMed: 24757056]
107. Tsumura A, et al. Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. Genes Cells. 2006; 11:805–814. [PubMed: 16824199]
108. Matsui T, et al. Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. Nature. 2010; 464:927–931. [PubMed: 20164836]
109. Bachman KE, et al. Histone modifications and silencing prior to DNA methylation of a tumor suppressor gene. Cancer Cell. 2003; 3:89–95. [PubMed: 12559178]
110. Espada J, et al. Human DNA methyltransferase 1 is required for maintenance of the histone H3 modification pattern. J Biol Chem. 2004; 279:37175–37184. [PubMed: 15220328]
111. Nguyen CT, et al. Histone H3-lysine 9 methylation is associated with aberrant gene silencing in cancer cells and is rapidly reversed by 5-aza-2′-deoxycytidine. Cancer Res. 2002; 62:6456–6461. [PubMed: 12438235]
112. Rose NR, Klose RJ. Understanding the relationship between DNA methylation and histone lysine methylation. Biochim Biophys Acta. 2014; 1839:1362–1372. [PubMed: 24560929]
113. Smith ZD, Meissner A. DNA methylation: roles in mammalian development. Nat Rev Genet. 2013; 14:204–220. [PubMed: 23400093]
114. Otani J, et al. Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. EMBO Rep. 2009; 10:1235–1241. [PubMed: 19834512]
115. Ooi SK, et al. DNMT3L connects unmethylated lysine 4 of histone H3 to *de novo* methylation of DNA. Nature. 2007; 448:714–717. This paper reports on structural evidence that a histone modification regulated DNA methyltransferase activity. [PubMed: 17687327]
116. Edwards JR, et al. Chromatin and sequence features that define the fine and gross structure of genomic methylation patterns. Genome Res. 2010; 20:972–980. [PubMed: 20488932]
117. Iwase S, et al. ATRX links atypical histone methylation recognition mechanisms to human mental retardation syndrome. Nat Struct Mol Biol. 2011; 18:769–776. [PubMed: 21666679]
118. Eustermann S, et al. Combinatorial readout of histone H3 modifications specifies localization of ATRX to heterochromatin. Nat Struct Mol Biol. 2011; 18:777–782. [PubMed: 21666677]
119. Ciccone DN, et al. KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. Nature. 2009; 461:415–418. [PubMed: 19727073]
120. Epsztejn-Litman S, et al. *De novo* DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. Nat Struct Mol Biol. 2008; 15:1176–1183. [PubMed: 18953337]
121. Kokura K, Sun L, Bedford MT, Fang J. Methyl-H3K9-binding protein MPP8 mediates E-cadherin gene silencing and promotes tumour cell motility and invasion. EMBO J. 2010; 29:3673–3687. [PubMed: 20871592]
122. Chang Y, et al. MPP8 mediates the interactions between DNA methyltransferase Dnmt3a and H3K9 methyltransferase GLP/G9a. Nat Commun. 2011; 2:533. [PubMed: 22086334]
123. Okada T, et al. CENP-B controls centromere formation depending on the chromatin context. Cell. 2007; 131:1287–1300. [PubMed: 18160038]
124. Ando S, Yang H, Nozaki N, Okazaki T, Yoda K. CENP-A, -B, and -C chromatin complex that contains the I-type α-satellite array constitutes the prekinetochore in HeLa cells. Mol Cell Biol. 2002; 22:2229–2241. [PubMed: 11884609]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

125. Gopalakrishnan S, Sullivan BA, Trazzi S, Della Valle G, Robertson KD. DNMT3B interacts with constitutive centromere protein CENP-C to modulate DNA methylation and the histone code at centromeric regions. Hum Mol Genet. 2009; 18:3178–3193. [PubMed: 19482874]

126. Fuks F, Hurd PJ, Deplus R, Kouzarides T. The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. Nucleic Acids Res. 2003; 31:2305–2312. [PubMed: 12711675]

127. Smallwood A, Esteve PO, Pradhan S, Carey M. Functional cooperation between HP1 and DNMT1 mediates gene silencing. Genes Dev. 2007; 21:1169–1178. [PubMed: 17470536]

128. Muramatsu D, Singh PB, Kimura H, Tachibana M, Shinkai Y. Pericentric heterochromatin generated by HP1 protein interaction-defective histone methyltransferase Suv39h1. J Biol Chem. 2013; 288:25285–25296. This paper shows that the SUV39H HP1 interaction is dispensable for pericentric H3K9me3 and DNA methylation, but is necessary for HP1 recruitment and completion of heterochromatin formation. [PubMed: 23836914]

129. Krouwels IM, et al. A glue for heterochromatin maintenance: stable SUV39H1 binding to heterochromatin is reinforced by the SET domain. J Cell Biol. 2005; 170:537–549. [PubMed: 16103223]

130. Rowe HM, et al. *De novo* DNA methylation of endogenous retroviruses is shaped by KRAB-ZFPs/KAP1 and ESET. Development. 2013; 140:519–529. [PubMed: 23293284]

131. Wolf D, Goff SP. Embryonic stem cells use ZFP809 to silence retroviral DNAs. Nature. 2009; 458:1201–1204. [PubMed: 19270682]

132. Jacobs FM, et al. An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. Nature. 2014; 516:242–245. [PubMed: 25274305]

133. Leung DC, et al. Lysine methyltransferase G9a is required for *de novo* DNA methylation and the establishment, but not the maintenance, of proviral silencing. Proc Natl Acad Sci USA. 2011; 108:5718–5723. [PubMed: 21427230]

134. Rowe HM, et al. KAP1 controls endogenous retroviruses in embryonic stem cells. Nature. 2010; 463:237–240. [PubMed: 20075919]

135. Leung DC, Lorincz MC. Silencing of endogenous retroviruses: when and why do histone marks predominate? Trends Biochem Sci. 2012; 37:127–133. [PubMed: 22178137]

136. Jin B, et al. Linking DNA methyltransferases to epigenetic marks and nucleosome structure genome-wide in human tumor cells. Cell Rep. 2012; 2:1411–1424. [PubMed: 23177624]

137. Morselli M, et al. *In vivo* targeting of *de novo* DNA methylation by histone modifications in yeast and mouse. eLife. 2015; 4:e06205. [PubMed: 25848745]

138. Bell O, et al. Localized H3K36 methylation states define histone H4K16 acetylation during transcriptional elongation in *Drosophila*. EMBO J. 2007; 26:4974–4984. [PubMed: 18007591]

139. Jia D, Jurkowska RZ, Zhang X, Jeltsch A, Cheng X. Structure of Dnmt3a bound to Dnmt3L suggests a model for *de novo* DNA methylation. Nature. 2007; 449:248–251. [PubMed: 17713477]

140. Zhang Y, et al. DNA methylation analysis of chromosome 21 gene promoters at single base pair and single allele resolution. PLoS Genet. 2009; 5:e1000438. [PubMed: 19325872]

141. Li BZ, et al. Histone tails regulate DNA methylation by allosterically activating *de novo* methyltransferase. Cell Res. 2011; 21:1172–1181. [PubMed: 21606950]

142. Guo X, et al. Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. Nature. 2015; 517:640–644. This paper shows a structure-based mechanism underlying histone-based allosteric regulation of*de novo* DNA methyltransferase activity. [PubMed: 25383530]

143. Esteve PO, et al. Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. Genes Dev. 2006; 20:3089–3103. [PubMed: 17085482]

144. Song J, Rechkoblit O, Bestor TH, Patel DJ. Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. Science. 2011; 331:1036–1040. [PubMed: 21163962]

145. Song J, Teplova M, Ishibe-Murakami S, Patel DJ. Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. Science. 2012; 335:709–712. This paper,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

with reference 144, defines the structural basis for autoinhibitory and active mechanisms for regulation of DNMT1 activity. [PubMed: 22323818]

146. Takeshita K, et al. Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1). Proc Natl Acad Sci USA. 2011; 108:9055–9059. [PubMed: 21518897]
147. Yang N, Xu RM. Structure and function of the BAH domain in chromatin biology. Crit Rev Biochem Mol Biol. 2013; 48:211–221. [PubMed: 23181513]
148. Kuo AJ, et al. The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier–Gorlin syndrome. Nature. 2012; 484:115–119. [PubMed: 22398447]
149. Bostick M, et al. UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science. 2007; 317:1760–1764. [PubMed: 17673620]
150. Sharif J, et al. The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature. 2007; 450:908–912. [PubMed: 17994007]
151. Karagianni P, Amazit L, Qin J, Wong J. ICBP90, a novel methyl K9 H3 binding protein linking protein ubiquitination with heterochromatin formation. Mol Cell Biol. 2008; 28:705–717. [PubMed: 17967883]
152. Nady N, et al. Recognition of multivalent histonestates associated with heterochromatin by UHRF1 protein. J Biol Chem. 2011; 286:24300–24311. [PubMed: 21489993]
153. Rajakumara E, et al. PHD finger recognition of unmodified histone H3R2 links UHRF1 to regulation of euchromatic gene expression. Mol Cell. 2011; 43:275–284. [PubMed: 21777816]
154. Hu L, Li Z, Wang P, Lin Y, Xu Y. Crystal structure of PHD domain of UHRF1 and insights into recognition of unmodified histone H3 arginine residue 2. Cell Res. 2011; 21:1374–1378. [PubMed: 21808300]
155. Wang C, et al. Structural basis for site-specific reading of unmodified R2 of histone H3 tail by UHRF1 PHD finger. Cell Res. 2011; 21:1379–1382. [PubMed: 21808299]
156. Cheng J, et al. Structural insight into coordinated recognition of trimethylated histone H3 lysine 9 (H3K9me3) by the plant homeodomain (PHD) and tandem tudor domain (TTD) of UHRF1 (ubiquitin-like, containing PHD and RING finger domains, 1) protein. J Biol Chem. 2013; 288:1329–1339. [PubMed: 23161542]
157. Arita K, et al. Recognition of modification status on a histone H3 tail by linked histone reader modules of the epigenetic regulator UHRF1. Proc Natl Acad Sci USA. 2012; 109:12950–12955. References 157 and 156 report on the mechanism of histone recognition by UHRF1, thereby linking H3K9me3 mark to CG methylation maintanance. [PubMed: 22837395]
158. Rothbart SB, et al. Multivalent histone engagement by the linked tandem Tudor and PHD domains of UHRF1 is required for the epigenetic inheritance of DNA methylation. Genes Dev. 2013; 27:1288–1298. [PubMed: 23752590]
159. Gelato KA, et al. Accessibility of different histone H3-binding domains of UHRF1 is allosterically regulated by phosphatidylinositol 5-phosphate. Mol Cell. 2014; 54:905–919. [PubMed: 24813945]
160. Rottach A, et al. The multi-domain protein Np95 connects DNA methylation and histone modification. Nucleic Acids Res. 2010; 38:1796–1804. [PubMed: 20026581]
161. Arita K, Ariyoshi M, Tochio H, Nakamura Y, Shirakawa M. Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. Nature. 2008; 455:818–821. [PubMed: 18772891]
162. Avvakumov GV, et al. Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. Nature. 2008; 455:822–825. [PubMed: 18772889]
163. Hashimoto H, et al. The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. Nature. 2008; 455:826–829. The above three papers describe the structural basis underlying recognition of hemimethylated DNA by UHRF1, thereby linking UHRF1 to CG methylation maintenance. [PubMed: 18772888]
164. Felle M, et al. The USP7/Dnmt1 complex stimulates the DNA methylation activity of Dnmt1 and regulates the stability of UHRF1. Nucleic Acids Res. 2011; 39:8355–8365. [PubMed: 21745816]
165. Berkyurek AC, et al. The DNA methyltransferase Dnmt1 directly interacts with the SET and RING finger-associated (SRA) domain of the multifunctional protein Uhrf1 to facilitate

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

accession of the catalytic center to hemi-methylated DNA. J Biol Chem. 2014; 289:379–386. [PubMed: 24253042]

166. Bashtrykov P, Jankevicius G, Jurkowska RZ, Ragozin S, Jeltsch A. The UHRF1 protein stimulates the activity and specificity of the maintenance DNA methyltransferase DNMT1 by an allosteric mechanism. J Biol Chem. 2014; 289:4106–4115. [PubMed: 24368767]
167. Achour M, et al. The interaction of the SRA domain of ICBP90 with a novel domain of DNMT1 is involved in the regulation of *VEGF* gene expression. Oncogene. 2008; 27:2187–2197. [PubMed: 17934516]
168. Citterio E, et al. Np95 is a histone-binding protein endowed with ubiquitin ligase activity. Mol Cell Biol. 2004; 24:2526–2535. [PubMed: 14993289]
169. Nishiyama A, et al. Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication. Nature. 2013; 502:249–253. [PubMed: 24013172]
170. Leonhardt H, Page AW, Weier HU, Bestor TH. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. Cell. 1992; 71:865–873. [PubMed: 1423634]

Abbrevations

Symmetrical sequences	DNA fragments that display the same sequence on both DNA strands
Asymmetrical sequence	A sequence that is only present on one strand of the DNA
Readers	Proteins and domains capable of binding to a specific epigenetic mark to recruit certain proteins to the target epigenetic mark
Chromodomains	A type of reader module that targets histone lysine methylation marks
Bromo adjacent homology domains	(BAH domains). Another type of reader module that targets histone lysine methylation marks
Hemimethylated sites	DNA sequences that are methylated on only one of the two complementary DNA strands
***Agrobacterium tumefaciens*-mediated transformation**	The process of experimentally inserting foreign DNA in the genome of a plant via infiltration with *Agrobacterium tumefaciens*
Isothermal Titration Calorimetry	(ITC). A biophysical technique that quantifies the solution interaction features by measuring the reaction thermodynamic changes
Epiallele	An allele that differs in its epigenetic marks, not in its DNA sequence
Major satellites	Refers to the 234 bp repeat sequence found in the pericentromeric region in mice
Minor satellites	Refers to the 120 bp repeat sequences found in the centromeric region in mice

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Retroviruses	RNA viruses that use reverse transcriptase to convert their genome into DNA, which is then integrated into the host genome
Repetitive element	A sequence that is found in multiple copies in the genome. Examples are telomeric repeats, transposable elements, and centromeric repeats
Inactive X chromosome	One of the two X chromosomes in females is inactivated to prevent overexpression of X gene products in females compared to males. This silencing is done through formation of heterochromatin
Primordial germ cell	A Cell that gives rise to both spermatozoa and oocytes
Retroelements	Transposable elements that move via the transcription of an RNA intermediate

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

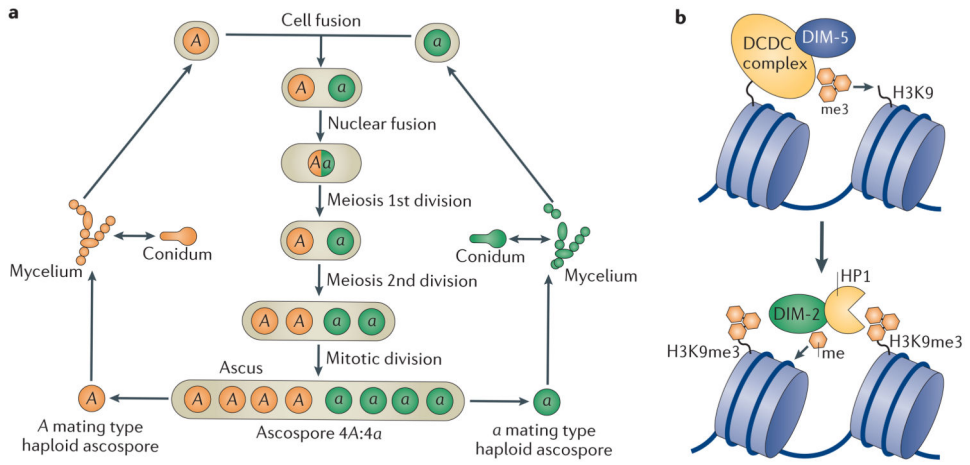

Figure 1. A unidirectional pathway in *Neurospora crassa*

a | A cartoon model of the life cycle of *Neurospora crassa*. *Neurospora* grows from a single haploid spore into a multinucleated branched thread called mycelium. This vegetative stage continues as the mycelium expands and can originate new mycelia as colonies bud off and disperse (conidium stage). This cycle is asexual and continues until two colonies of opposite mating types (*A* and *a*) interact and give rise to a fruiting body. The fusion of the two haploid spores gives rise to a dikaryon that then proliferates within the fruiting body. After premeiotic DNA synthesis, nuclei of the opposite mating type fuse, and meiosis is initiated. Each of the meiotic spores then undergoes mitosis to give rise to an octad of haploid spores. This series of events constitutes the sexual cycle. **b** | A schematic representation of the unidirectional pathway from H3K9me3 to DNA methylation in *N. crassa*. The DCDC complex associates with H3K9 methyltransferase DIM-5 and targets it to certain chromatin loci to create the H3K9me3 mark. Once the H3K9me3 is established, the heterochromatin protein 1 (HP1) can specifically recognize the H3K9me3 mark to facilitate targeting of the associated DNA methyltransferase DIM-2 to methylate DNA at the same sites.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

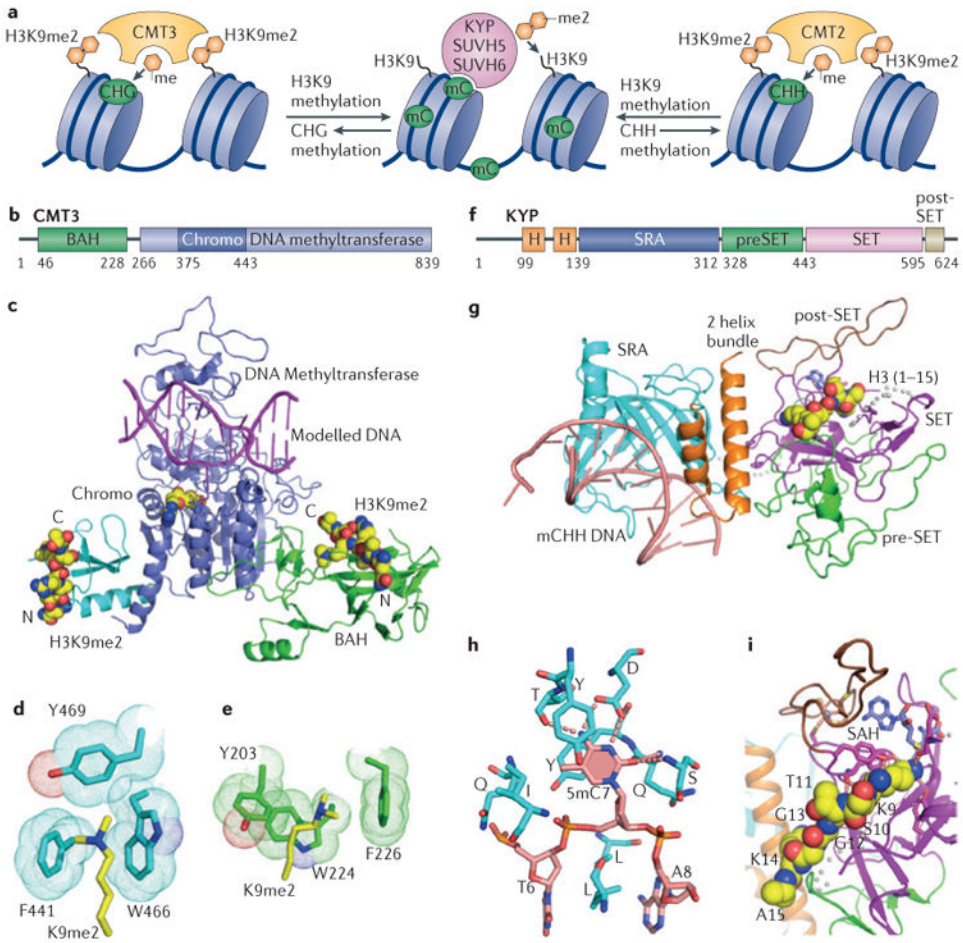

Figure 2. Structural basis underlying the cross-regulation of CHG DNA methylation and histone H3K9me2 in *Arabidopsis thaliana* by the self-reinforcing loop between CMT3 and KRYPTONITE

a | A cartoon representation of the self-reinforcing model between Chromomethylase 3 (CMT3) and CMT2 and KRYPTONITE (KYP) in CHG and CHH methylation, respectively, and H3K9me2 methylation. CMT3 and CMT2 are targeted by H3K9me2, which catalyses the transfer of a methyl group to CHG and CHH sites on the DNA in the corresponding region. Similarly, the CHG and CHH methylation mark can be captured by KYP, which then catalyses the transfer of a methyl group to H3K9 of nearby nucleosomes, creating the binding sites for CMT3 and CMT2 to establish the reinforcing loop. **b** | Domain architecture of CMT3 in a colour-coded representation. **c** | A superposed composite structural model of ZMET2 (the maize homologue of CMT3) in complex with H3K9me2 peptide and modelled DNA (based on the RCSB protein databank (PDB) codes: 4FSX, 4FT2, 4FT4 and 4DA4). The bromo adjacent homology (BAH), DNA methyltransferase, and chromodomains are coloured in green, purple, and cyan, respectively. The bound peptide and cofactor SAH are shown in space-filling representations, and modelled DNA is magenta ribbon representation. **d** | The aromatic residues Phe441, Trp466, and Tyr469 of the ZMET2 chromodomain form an aromatic cage for recognition of the H3K9me2 in a methyllysine-dependent manner. **e** | The aromatic residues Tyr203, Trp224, and Phe226 of the ZMET2 BAH domain form an aromatic cage for recognition of H3K9me2 in a methyllysine-dependent manner. **f** | Domain

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

architecture of *Arabidopsis thaliana* KYP in a colour-coded representation. **g** | A ribbon representation of the crystal structure of KYP in complex with methylated non-CG DNA, cofactor SAH, and unmodified H3(1–15) substrate peptide (PDB codes: 4QEN, 4QEO, and 4QEP). The amino-terminal 2-helix bundle, SRA (SET- and RING-associated), pre-SET, SET, and post-SET domains of KYP are coloured in orange, cyan, green, magenta and brown, respectively. The bound methylated DNA, cofactor SAH, and peptide substrate are shown in salmon ribbon, lilac stick, and space-filling representation, respectively. **h** | Structural basis underlying specific recognition by the SRA domain of the flipped out 5mC base of the bound methylated DNA. The base is stacked between two tyrosine residues from the top and bottom directions. The Watson-Crick edge of the 5mC forms several hydrogen bonding interactions with the surrounding residues as indicated by dashed red lines. The methyl group of 5mC is accommodated within a small hydrophobic pocket. **i** | The peptide substrate and cofactor SAH are embedded in between the post-SET and SET domains. The to-be-methylated lysine forms several hydrogen-bonding interactions with important tyrosine residues as shown by dashed pink lines.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

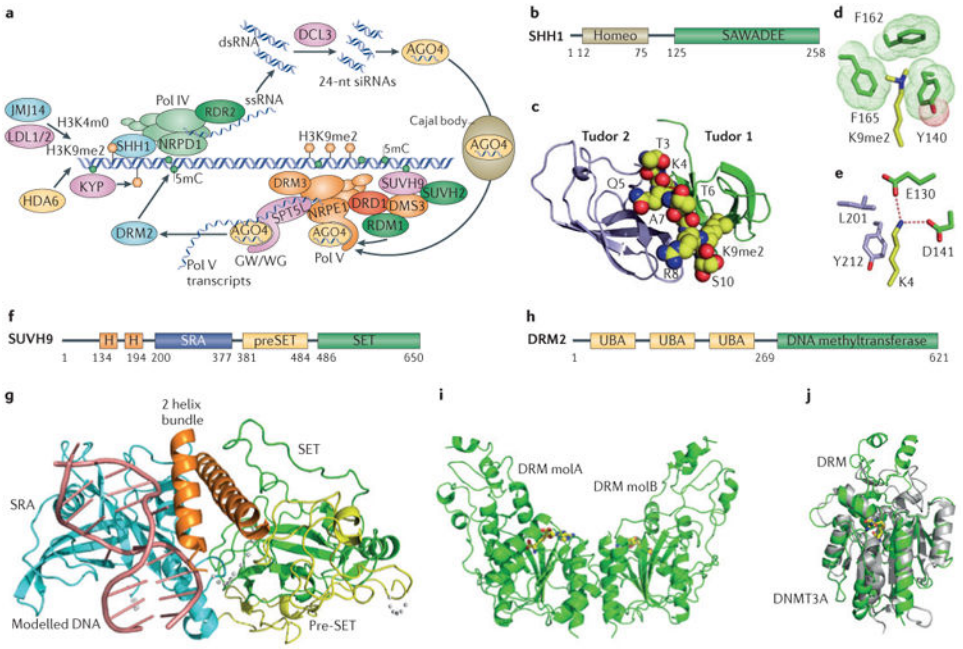

Figure 3. Structures of regulators of *Arabidopsis thaliana* RdDM pathway

a | An updated scheme of the RNA-directed DNA methylation (RdDM) pathway. The RdDM is initiated when RNA POLYMERASE IV (Pol IV) is targeted by SAWADEE HOMEODOMAIN HOMOLOGUE 1(SHH1)-bound H3K4me0K9me2 (orange circles) to produce single RNA transcripts. The H3K4me0K9me2 state is regulated by the histone modification enzymes JMJ14, LDL1 and LDL2, KRYPTONITE (KYP) and HDA6. Pol IV produced RNA is replicated by Pol IV-associated RNA-DIRECTED RNA POLYMERASE 2 (RDR2) to generate double-stranded RNA, which is further processed by DICER-LIKE 3 (DCL3) and ARGONAUTE 4 (AGO4) to produce 24-nucleotide (nt) siRNAs which are loaded onto AGO4. Meanwhile, the DRD1, DMS3, RDM1 (DDR) complex is directed to the methylated DNA (green circle) region by its associated SUVH2 and SUVH9 and targets Pol V to produce scaffold non-coding RNA. siRNA-bound AGO4 can interact with Pol V either directly or indirectly through SPT5L and through base-pairing between siRNA and non-coding RNA to further target DNA methylase 2 (DRM2) to methylate target DNA. **b** | Domain architecture of *Arabidopsis thaliana* SHH1. **c** | Ribbon representation of the crystal structure of the SAWADEE domain of SHH1 in complex with H3K9me2 peptide (PDB code: 4IUT). The first tudor domain, second tudor domain and the bound peptide are coloured in green, purple and yellow, respectively. The peptide is shown in a space-filling representation. **d** | The structural basis underlying specific recognition of H3K9me2. Three aromatic residues of SHH1, Tyr140, Phe162, and Phe165, form an aromatic cage to accommodate the methyllysine side chain, involving stabilization by cation-π interactions. **e** | The structural basis underlying specific recognition of unmodified H3K4. Two acidic residues, Glu130 and Asp141, form salt bridges and hydrogen bonding interactions with the amide protons of unmodified H3K4, with hydrogen bonding alignments shown as dashed pink lines. **f** | Domain architecture of *Arabidopsis thaliana* SUVH9, which possesses SET- and RING-associated (SRA), pre-SET and SET domains but lacks the post-SET domain. **g** | Crystal structure of SUVH9 (PDB code: 4NJ5) with a modelled DNA in ribbon

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

representation. The amino-terminal 2-helix bundle, SRA domain, pre-SET domain, SET domain, and the modelled DNA are coloured in orange, cyan, yellow, green and salmon, respectively. **h** | Domain architecture of *Arabidopsis thaliana de novo* DNA methyltransferase DRM2. **i** | Ribbon representation of the structure of the symmetric dimer (coloured in green) formed by the catalytic domain of *Nicotiana tabacum* DRM (PDB code: 4ONJ). The cofactor analogue sinefungin is shown in a stick representation. **j** | A superposition of *N. tabacum* DRM monomer (in green) with DNA methyltransferase 3A (DNMT3A) catalytic domain (in silver, PDB code: 2QRV) reveals NtDRM possesses classic type I methyltransferase fold like DNMT3A. Mol, molecule. Part **a** is from REF. 3, Nature Publishing Group.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

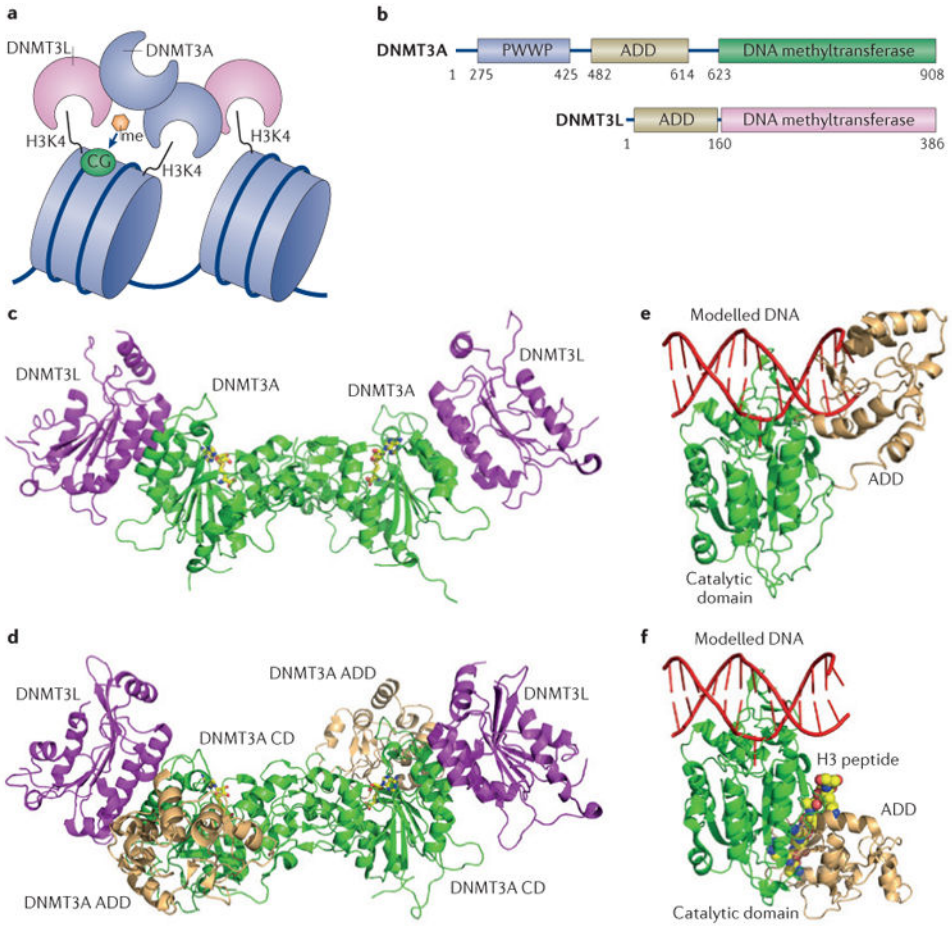

Figure 4. Structure of mammalian *de novo* DNA methyltransferases DNMT3A and DNMT3L

a | A cartoon model showing the DNA methyltransferase 3A (DNMT3A)–DNMT3L tetramer binds to unmodified H3K4 and then catalyses CG DNA methylation. **b** | Domain architecture of mammalian *de novo* DNA methyltransferase DNMT3A and its catalytically inactive cofactor DNMT3L in a colour-coded representation. **c** | Ribbon representation of the structure of the DNMT3L–DNMT3A–DNMT3A–DNMT3L functional tetramer with the catalytic domain of DNMT3L coloured in magenta and the catalytic domain of DNMT3a in green (RCSB protein databank (PDB) code: 2QRV). The cofactor SAH is shown in yellow in a stick representation. **d** | Structure of the DNMT3L–DNMT3A–DNMT3A–DNMT3L tetramer catalytic domains including the ADD (ATRX–DNMT3–DNMT3L) domain of DNMT3A (in light brown) (PDB code: 4U7P). **e** | Structure of the autoinhibitory conformation with a superposed model of bound DNA (PDB code: 4U7P, the DNA is modelled from Methyltransferase HhaI (M.HhaI)–DNA complex with a PDB code: 1MHT). The ADD domain, catalytic domain, and the modelled DNA are coloured in light brown, green, and red, respectively. The ADD domain interacts with the catalytic domain and blocks access to modelled DNA along one face. **f** | The structure of DNMT3a with captured H3K4me0 peptide together with a superposed model of bound DNA (PDB code: 4U7T). The bound peptide is shown in space-filling representation. Upon H3 peptide binding, the ADD domain interacts with the catalytic domain along another face that is positioned further

away from the catalytic site, thereby releasing autoinhibition and providing access to modelled DNA. Panels e and f are aligned in the same orientation.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

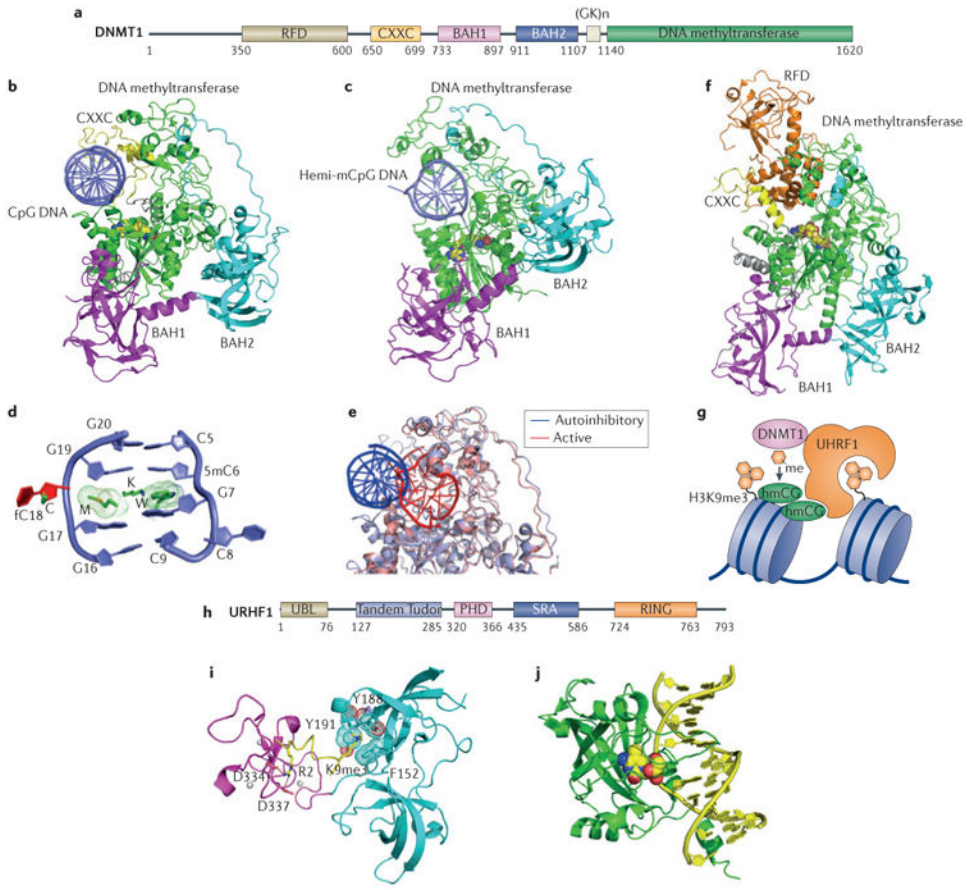

Figure 5. Structures of DNMT1 and its epigenetic regulator UHRF1 in maintenance DNA methylation

a | Domain architecture of DNA methyltransferase 1 (DNMT1). **b** | Structure of DNMT1 in complex with unmethylated CpG DNA in an autoinhibitory conformation (RCSB protein databank (PDB) code: 3PT6). The CXXC, bromo adjacent homology 1 (BAH1), BAH2 and DNA methyltransferase domains are coloured in yellow, magenta, cyan, and green, respectively. The unmethylated CpG DNA is shown in a purple ribbon representation, with the DNA interacting with the CXXC domain. The linker between the CXXC and BAH domain is positioned between the DNA and the catalytic pocket, thereby blocking access to the catalytic site. **c** | Structure of DNMT1 in complex with a hemimethylated CpG DNA in a productive conformation (PDB code: 4DA4). The to-be-methylated cytosine is flipped out from the DNA duplex and inserts into the active site of the methyltransferase domain. **d** | The base-flipping mechanism of DNMT1. The to-be-methylated fC (this cytosine analogue was used to covalently trap a productive complex) is highlighted in red and forms a covalent bond with a cysteine residue of the active site. Lysine and methionine residues insert into the space vacated by the flipped-out 5fC, with the alignment buttressed by a tryptophan residue. **e** | A superposition of the unmethylated DNA bound in an autoinhibitory conformation of DNMT1 (protein in light blue and DNA in dark blue) and hemimethylated DNA bound productive conformation of DNMT1 (protein in light red and DNA in dark red). **f** | The structure of a replication foci domain (RFD)-containing DNMT1 (free state, with RFD domain in orange) in an autoinhibitory conformation (PDB code: 3AV5). **g** | A model

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

proposing that UHRF1 could target DNMT1 to hemimethylated CG (hmCG) DNA by recognizing and potentially binding to both H3K9me3 and hmCG DNA. **h** | Domain architecture of UHRF1. **i** | Structure of the tandem tudor-plant homeodomain (PHD) cassette of UHRF1 in complex with H3K9me3 peptide (PDB code: 4GY5). The tandem tudor and PHD domains are coloured in cyan and magenta, respectively. The unmodified R2 is specifically recognized by the acidic residues Asp334 and Asp337 of the PHD finger through salt bridges and hydrogen-bonding interactions, which are highlighted by dashed red lines. The trimethylated H3K9 is accommodated within an aromatic cage formed by Phe152, Tyr188 and Tyr191 of the tandem tudor domain. **j** | The SET- and RING-associated (SRA) domain of UHRF1 can specifically recognize hemimethylated CpG DNA (PDB code: 3CLZ). The SRA domain and DNA are coloured in green and yellow, respectively. The flipped out 5mC is highlighted in a space-filling representation. UBL, ubiquitin-like domain