Molecular Mechanism of Action of Plant DRM *De Novo* DNA Methyltransferases

Xuehua Zhong[1, 7, 8], Jiamu Du[2, 7], Christopher J. Hale[1], Javier Gallego-Bartolome[1, 3], Suhua Feng[1, 4], Ajay A. Vashisht[5], Joanne Chory[3, 6], James A. Wohlschlegel[5], Dinshaw J. Patel[2, *] and Steven E. Jacobsen[1, 4, *]

[1]Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, Los Angeles, CA 90095, USA.

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

[3]Plant Biology Laboratory, The Salk Institute for Biological Studies, La Jolla, CA 92037.

[4]Howard Hughes Medical Institute and Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, University of California, Los Angeles, Los Angeles, CA 90095, USA.

[5]Department of Biological Chemistry, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California, 90095, USA.

[6]Howard Hughes Medical Institute, The Salk Institute for Biological Studies, La Jolla, CA 92037

[7]X.Z. and J.D. contribute equally to this work

[8]Present address: Wisconsin Institute for Discovery, Laboratory of Genetics, University of Wisconsin, Madison, WI 53706, USA

[*]Correspondence: pateld@mskcc.org (D.J.P.), jacobsen@ucla.edu (S.E.J.)

SUMMARY

DNA methylation is a conserved epigenetic gene regulation mechanism. DOMAINS REARRANGED METHYLTRANSFERASE (DRM) is a key *de novo* methyltransferase in plants, but how DRM acts mechanistically is poorly understood. Here, we report the crystal structure of the methyltransferase domain of tobacco DRM (NtDRM) and reveal a molecular basis for its rearranged structure. NtDRM forms a functional homo-dimer critical for catalytic activity. We also show that Arabidopsis DRM2 exists in complex with the siRNA effector ARGONAUTE4 (AGO4) and preferentially methylates one DNA strand, likely the strand acting as the template for RNA polymerase V mediated non-coding RNA transcripts. This strand-biased DNA methylation is also positively correlated with strand-biased siRNA accumulation. These data suggest a model in which DRM2 is guided to target loci by AGO4-siRNA and involves base-pairing of associated siRNAs with nascent RNA transcripts.

INTRODUCTION

DNA methylation is a conserved epigenetic gene regulation mechanism that is utilized by cells to regulate gene expression and suppress transposon activity. Unlike in mammals where DNA methylation predominantly occurs in CG context (Lister et al., 2009), plant DNA is frequently methylated in three different sequence contexts: CG, CHG and CHH (H=A, T, or C) (Law and Jacobsen, 2010). In *Arabidopsis thaliana*, while the maintenance of CG and CHG methylation is primarily controlled by METHYLTRANSFERASE1 (MET1, an ortholog of mammalian Dnmt1) and CHROMOMETHYLASE3 (CMT3, a plant specific methyltransferase), respectively (Du et al., 2012; Finnegan and Dennis, 1993; Finnegan and Kovac, 2000; Lindroth et al., 2001; Stroud et al., 2013), the maintenance of CHH methylation is controlled by DOMAINS REARRANGED METHYLTRANSFERASE2 (DRM2, an ortholog of mammalian Dnmt3) (Law and Jacobsen, 2010) and CMT2 (Stroud et al., 2014; Zemach et al., 2013).

De novo DNA methylation in all sequence contexts is mediated by DRM2 and is dependent on RNA interference (RNAi) like machinery via a process termed RNA-directed DNA methylation (RdDM) (Law and Jacobsen, 2010). This pathway involves two main phases: an upstream small interference RNA (siRNA) biogenesis phase and a downstream methylation targeting phase. Biogenesis of siRNAs is initiated by a plant specific RNA polymerase IV (Pol IV), which generates single-stranded RNA transcripts that are copied into double-stranded RNA by an RNA dependent RNA polymerase (RDR2). The resulting transcripts are cleaved into 24nt siRNAs by a Dicer endonuclease (DCL3) and further loaded into ARGONAUTE 4 (AGO4) forming AGO4-siRNA complexes. The targeting phase involves another plant specific RNA polymerase V (Pol V), which produces noncoding RNA transcripts that are proposed to act as a scaffold to recruit AGO4 through base-pairing of associated siRNAs (Law and Jacobsen, 2010;

Wierzbicki et al., 2009). While genome-wide occupancy of Pol V is dependent on the DDR complex consisting of DEFECTIVE IN MERISTEM SILENCING 3 (DMS3), DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1 (DRD1), and RNA-DIRECTED DNA METHYLATION 1 (RDM1) (Zhong et al., 2012), global chromatin association of Pol IV is dependent on a H3K9 methyl binding domain protein SHH1/DTF1 (Law et al., 2013; Zhang et al., 2013). A recent study suggests that DNA methylation is also required for Pol V association to chromatin, demonstrating the nature of the RdDM pathway as a self reinforcing loop mechanism (Johnson et al., 2014). The co-occurrence of Pol IV dependent siRNAs and Pol V dependent non-coding transcripts is thought to determine the sites of DRM2 action. However, despite the identification of a large number of proteins required for the RdDM pathway, the specific mechanism of DRM2 action including its biochemical activities, interacting partners, and how DRM2 is recruited to specific loci remain largely unknown.

To further understand the molecular mechanism of DRM2 action, we carried out structural and functional studies. We solved the crystal structure of the methyltransferase domain of a DRM2 homologue from tobacco, NtDRM. The structure reveals that although DRM proteins have a rearrangement of their methyltransferase sequence motifs, the overall structure retains a classic class-I methyltransferase fold (Schubert et al., 2003). In the crystal, NtDRM forms a homo-dimer with the dimer interface mimicking the Dnmt3a-Dnmt3L hetero-dimer interface. Mutations disrupting this dimerization significantly reduce its *in vitro* methyltransferase activity, which is similar to the behavior of Dnmt3a-Dnmt3L. These results suggest that dimerization may be a commonly used mechanism to initiate DNA methylation. To further understand the mechanism of DRM2 action, we performed affinity purification followed by mass spectrometry and found that Arabidopsis AGO4 co-purified with DRM2. Given that

AGO4 binds siRNAs, and that siRNAs have the potential to base pair either with the complementary DNA strand or nascent RNA transcripts, we examined the relationship between the strandedness of DNA methylation and siRNAs. We found that strand-biased DNA methylation is positively correlated with strand-biased siRNAs, suggesting that DRM2 preferentially methylates the template DNA strand for Pol V transcription. Collectively, our data suggest a model wherein AGO4-siRNAs guide a DRM2 dimer to methylate a template DNA strand for Pol V transcription and this process is mediated by base-pairing of associated siRNAs with Pol V transcripts.

RESULTS AND DISCUSSION

Overall Structure of the NtDRM Catalytic Domain

To begin to reveal the mechanism of DRM action, we sought to determine the crystal structure of DRM2. Despite extensive efforts to crystallize Arabidopsis DRM2, we failed to obtain diffraction quality crystals. Instead, we successfully crystallized the DRM methyltransferase domain from a related plant *Nicotiana tabacum* (NtDRM MTase, residues 255-608). NtDRM shares a similar domain architecture and function with DRM2 (Figure 1A) (Wada et al., 2003). The structure of NtDRM MTase in complex with sinefungin, an analog of the cofactor substrate S-adenosyl-L-methionine (SAM), was solved by the SAD method and refined to 2.8 Å resolution yielding an *R* factor of 20.1% and a free *R* factor of 22.1%. In the asymmetric unit, there is an NtDRM MTase dimer with each molecule bound to a cofactor analog sinefungin in the active site (Figure 1B, Table S1). Overall, the protein exhibits well-defined electron density except that the catalytic loop regions (residues 567-584 in monomer A and residues 569-584 in monomer B) were not well defined and we were unable to build these segments into the final model. The NtDRM MTase dimer exhibits a butterfly like arrangement with the two monomers related by a 2-fold non-crystallographic symmetry axis. The two catalytic domains dimerize in the middle and the two target recognition domains (TRDs) extend on two sides as the wings (Figure 1B).

The extreme N-terminal 30 residues of NtDRM MTase (residues 259-288) form a long loop wrapped on the surface of the core methyltransferase domain, which is composed of the remaining residues (residues 289-608). Although the primary sequence of NtDRM MTase (and all other DRM2 proteins) is rearranged as compared to that of class-I methyltransferases, its overall structure adopts a typical class-I methyltransferase fold with a catalytic domain and a TRD domain (Figures 2A and 2B). The catalytic domain features a central seven-stranded β-sheet flanked by one layer of three α-helices on one side and another layer of four α-helices on

the other side (Figures 2A and 2B) resembling other class-I DNA methyltransferases including M.HhaI, Dnmt3a, Dnmt1, and ZMET2 (Cheng et al., 1993; Du et al., 2012; Jia et al., 2007; Song et al., 2011) (Figure 2C). The catalytic loop of NtDRM is disordered probably due to the absence of the substrate DNA. In the region near the catalytic domain, the TRD domain of NtDRM is composed of a two-stranded anti-parallel β-sheet similar to that of Dnmt3a. In the region away from the catalytic domain, the TRD domain has two anti-parallel α-helices connected by a loop which defines a novel arrangement of the TRD domain (Figures 2A and 2C), indicating a novel DNA substrate binding mode different from other known DNA methyltransferases.

The Rearranged Domain Structure of DRM

The first residue of the core methyltransferase domain, Pro288, is adjacent to the C-terminal end of the protein in three-dimensional space (Figures 2A and 2C). Similarly, the N and C termini of Dnmt3a are also adjacent to each other (Figure 2C). If the N and C termini of NtDRM MTase were fused together as a closed loop and then broken around Gly480 (black arrow in Figure 2B), its sequence folding topology would be identical to Dnmt3a. Thus, while the DRM MTase domains are rearranged in the linear sequence, it retains the overall fold of a classic class-I methyltransferase. The domain rearrangement mechanism confirms previous speculation that DRM folds similarly to other typical class-I methyltransferases despite the motif rearrangement (Cao et al., 2000). The point of rearrangement is identical in many plant species, at the bottom side of the catalytic domain opposite against and far away from the catalytic center or the cofactor binding site of the catalytic and TRD domains (Figures 2C and S1), suggesting that the rearrangement occurred during an early stage of plant evolution. Based on structures, it seems likely that DRM proteins have a similar catalytic mechanism as other class-I methyltransferases.

NtDRM MTase Forms a Functional Homo-Dimer Critical for Catalytic Activity

It was reported that mammalian Dnmt3a and Dnmt3L form a Dnmt3L-Dnmt3a-Dnmt3a-Dnmt3L tetramer and that this oligomeric status is essential for its DNA methylation activity (Jia et al., 2007). A Dnmt3a F728A mutant, disrupting the Dnmt3a-Dnmt3L hetero-dimer interface, abolishes the methyltransferase activity (Jia et al., 2007). Interestingly, when we superpose Dnmt3a onto one monomer of the NtDRM MTase dimer, we found that the other monomer of the NtDRM MTase dimer can be well superposed with the Dnmt3a dimerized Dnmt3L molecule (Jia et al., 2007) (Figure 3A). The NtDRM MTase homo-dimer interface mimics the Dnmt3a-Dnmt3L hetero-dimer interface with the former stabilized by a hydrophobic core composed of aromatic amino acids Phe310 and Tyr590 from each monomer, and a hydrophilic periphery involved in salt bridges and hydrogen bond interactions between positively charged Arg309 and Arg605 and negatively charged Asp591 and Glu283 (Figure 3B). However, no interface of NtDRM MTase mimics the Dnmt3a-Dnmt3a interface in the crystal, indicating that unlike the Dnmt3a-Dnmt3L that forms a hetero-tetramer, DRM likely utilizes a homo-dimer as a functional unit. By analyzing plants containing two different DRM2 constructs with different epitope tags, we also confirmed that DRM forms multimers *in vivo* (Figure 3C), consistent with the structural data.

To determine the importance of NtDRM dimerization, we mutated all the residues involved in the dimerization to serine (E283S/R309S/F310S/Y590S/D591S, designated as NtDRM-M5) and solved the crystal structure of the mutant protein (Table S1). The overall structure of NtDRM-M5 monomer is almost identical to the wild type NtDRM MTase with an RMSD of only 0.77 Å for 326 aligned Cα atoms by aligning their monomer structures (Figure

S2A); however, the dimer interface of NtDRM MTase is completely disrupted. In addition, enzymatic activity assays show that NtDRM-M5 has lost virtually all DNA methyltransferase activity compared with the wild type protein **(**Figure 3D). This result indicates that like the Dnmt3a-Dnmt3L interface, the DRM dimer interface is essential for catalysis. One plausible explanation is that dimerization might help stabilize the conformation of the catalytic loop because the C-terminal portion of the active site loop is involved in dimer interface formation (Jia et al., 2007). To further examine the functional significance of DRM2 dimerization *in vivo*, we generated a transgenic version of DRM2 in which the five key residues involved in dimerization were mutated to serine (E301S, R327S, H328S, F610S, E611S), designated as DRM2-M5. The wild type DRM2 (FLAG-DRM2) and mutant DRM2 (FLAG-DRM2-M5) transgenes were transformed into *drm1 drm2* and the effects of loss of DRM2 dimerization on DNA methylation were assessed by whole genome bisulfite sequencing approach. As shown in Figure 3E, DNA methylation was significantly reduced in the representative FLAG-DRM2-M5 mutant line compared to that of wild type FLAG-DRM2 lines even though they show similar expression levels of the DRM2 transgene (Figure S2B). This result suggests that dimerization is also critical for *in vivo* DRM2 activity.

Besides the Dnmt3a-Dnmt3L hetero-dimer interface, the Dnmt3a-Dnmt3a homo-dimer interface was also reported to be essential for the catalytic activity of Dnmt3a (Jia et al., 2007). Dnmt3a has one of the smallest TRD domains in comparison to other DNA methyltransferases. However, the dimerization of two Dnmt3a molecules doubles the DNA binding surface and enables the DNA substrate to be more accessible to the enzyme (Jia et al., 2007). In our NtDRM MTase structure, the TRD domain is larger than that of Dnmt3a (Figure 2C). In addition, the TRD domain and the catalytic site of NtDRM MTase form a large continuous positively charged

surface suitable for DNA substrate binding (Figure S3A). Despite extensive efforts, we were not able to crystallize NtDRM with DNA oligomer duplexes of varying length and overhangs. We instead modeled the NtDRM MTase with a DNA substrate based on the structure of the productive covalently linked Dnmt1-DNA complex (Song et al., 2012). The model reveals that the substrate DNA duplex can be positioned within the substrate cleft between the catalytic domain and TRD, with the looped out to-be-methylated cytosine base positioned within the active site near the cofactor analog sinefungin (Figure S3B). The two α-helices of the TRD approach the major groove of the putative substrate DNA duplex, most likely participating in binding and sequence-specific DNA recognition (Figure S3B). Given that this model predicts that NtDRM is most likely sufficient to capture the substrate DNA duplex, it appears unnecessary to form a Dnmt3a-Dnmt3a like dimer to enlarge the DNA binding surface. This may explain why only the Dnmt3a-Dnmt3L surface is conserved in NtDRM, while the Dnmt3a-Dnmt3a surface is not present in NtDRM.

Collectively, our results reveal a possible conserved dimerization mechanism for plant and animal *de novo* DNA methyltransferases, suggesting that dimerization may be a commonly used mechanism to initiate DNA methylation.

UBA Domains are Important for DRM Function *In Vivo*

Besides the methyltransferase domain, DRM proteins also contain ubiquitin-associated (UBA) domains (Figure 1A) of unknown function (Cao et al., 2000). Previously, DRM2 UBA domains were shown to be required for the maintenance of DNA methylation at the *MEA-ISR* locus (Henderson et al., 2010). However, it remains unclear to what extent UBA domains are required for DNA methylation in the genome. To address these questions, we first examined whether the

UBA domains are required for global DNA methylation *in vivo*. We performed a whole genome bisulfite sequencing on previously published DRM2uba mutant lines where conserved residues within UBA domains were mutated and the DRM2uba mutant transgene was transformed into a *drm2* null mutant (Henderson et al., 2010). As shown in Figure 4A, DRM2uba showed a strong global loss of DNA methylation that was only slightly weaker than a catalytically inactive DRM2cat mutant (negative control), indicating that the UBA domains are required for genome-wide DRM2 activity *in vivo*. We further showed that loss of DNA methylation in DRM2uba is unlikely due to reduced expression of DRM2 as DRM2uba has similar protein level as that of wild type DRM2 (Figure S4D).

It is possible that the failure in DNA methylation restoration by DRM2uba is due to the loss of DRM2 catalytic activity. Despite extensive testing, we have, been unable to find *in vitro* conditions that allow for robust Arabidopsis DRM2 activity. Thus, we compared the activity of the full length NtDRM with the truncated NtDRM containing only the catalytic domain used for crystallization. As shown in Figure 4B, the NtDRM MTase domain alone exhibited activity very similar to that of the full length NtDRM, suggesting that UBA domains are not necessary for DRM catalytic activity. It is therefore possible that the UBA domains are involved in other aspects of DRM function, such as targeting DRM to specific loci. Consistent with this possibility, we noted a bimodal distribution of methylation change in the DRM2uba line as compared to that of DRM2cat (Figure S4A), suggesting that some DRM2 target sites are more sensitive to the loss of the UBA domains than others. Furthermore, we note that the sites most strongly hypomethylated in the DRM2uba line tend to have less broadly distributed heterochromatic marks than those sites weakly affected in a DRM2uba mutant (Figures S4B-S4C), suggesting that the UBA domains may help reinforce DRM2 activity at euchromatic

regions of the genome that contain smaller patches of heterochromatin.

AGO4 Co-Purifies With DRM2 *In Vivo*

To further explore the biochemical activity of DRM2, we performed immunoprecipitation and mass spectrometry (MS) to identify DRM2-interacting proteins. We generated an epitope-tagged 9xMYC-Biotin Ligase Recognition Peptide (BLRP)-DRM2 transgenic line where the expression of DRM2 is under the control of its own promoter. After affinity purification, copurifying proteins were identified through MS analysis. Peptides corresponding to AGO4 (At2g27040) were the most abundant in two independent purifications (Figure 5A). Less abundant peptides from a few other proteins were also found in both replicas (Figure 5A), although the biological significance of these interactions has not been tested. We validated the interaction between DRM2 and AGO4, by performing MYC pull-down assays in which tagged DRM2 was isolated using immobilized MYC beads and the presence of AGO4 in the purified DRM2 fraction was examined with an AGO4 endogenous antibody (Figure 5B). Taken together, the MS analyses together with affinity purification data indicate that DRM2 is associated with AGO4 *in vivo*.

DRM2 Mediates Strand-Biased DNA Methylation That is Positively Correlated With Strand-Biased SiRNAs

Given that AGO4 binds siRNAs (Qi et al., 2006) and interacts with DRM2 (Figure 5), we sought to examine the relationship between DRM2-dependent methylation and siRNA populations. Genomic studies have established a strong correlation between endogenous siRNAs and DRM2-mediated DNA methylation throughout the genome (Cokus et al., 2008; Law et al., 2013; Lee et al., 2012; Lister et al., 2008). However, the mechanism by which siRNAs guide DRM2

methylation is poorly understood. Previous observations of strand-biased DNA methylation that tended to associate with siRNAs in Arabidopsis (Lister et al., 2008; Luo and Preuss, 2003) prompted us to investigate DRM2 sites and to specifically test if there is a relationship between siRNA strandedness and the respective strand of targeted cytosines. To this end, we used a set of previously identified DRM2-dependent total siRNA clusters (Law et al., 2013) and defined a subset of these clusters that showed a strand-biased distribution of siRNAs, as well as clusters that showed little to no strand bias. Strand-biased clusters were defined as having a significant excess of siRNA reads mapping to either the positive or negative strand relative to the complementary strand (see Materials and Methods). We then used previously published whole-genome bisulfite sequencing datasets (Law et al., 2013) to calculate strand bias of both the methylcytosine and cytosine content at these clusters. As shown in Figure 6A, the strand-biased siRNA clusters were correlated with a strand bias for both cytosine content and methylcytosine content. Moreover, the direction of the bias was the same between siRNAs and cytosines or methylcytosines (Figures 6B-S5A), consistent with the general trend previously noted in whole genome bisulfite sequencing data (Lister et al., 2008). Thus, positive-strand siRNA clusters correlated with regions with a positive strand bias for methylcytosine and total cytosine content and vice versa. These results suggest that DRM2 preferentially methylates the same DNA strand as the siRNA, rather than the complementary strand to which the siRNA could base pair.

As a confirmation of these results, we used whole-genome bisulfite sequencing data from a *drm2* mutant line to define individual cytosines whose methylation was most strongly lost upon loss of the DRM2 protein (Figure 6C). We then plotted 24 nt siRNA abundance around these cytosines. Similar to what we observed at strand-biased siRNA clusters, we noted that siRNA abundance at these differentially methylated cytosines strongly correlated with the strandedness

of the methylcytosine assayed; differentially methylated cytosines on a given strand are more likely to be covered by siRNAs of the same strand as opposed to siRNAs of the complement strand (Figure 6D). We also noticed a pattern in the abundance of siRNA 5' ends distributed around differentially methylated cytosines, with the highest abundance of 5' ends 23 nt upstream of the cytosine in question (Figure S5B). In other words, the strongest 24 nt siRNA signal at differentially methylated cytosines correlates to a strand-matched siRNA positioned such that the 3' nt of the siRNA corresponds to the cytosine methylated by DRM2. One possible trivial explanation for this result is that, since 24 nt siRNAs have an over-representation of cytosines at their 3' end (Figure S5B), by centering our analysis on a cytosine we may be identifying patterns that are simply a consequence of the underlying sequence composition of the *Arabidopsis* 24 nt siRNA population. Alternatively it is possible that AGO4 and the associated 24 nt siRNAs are physically positioning the DRM2 active site to the targeted cytosine, which could also explain the over-representation of cytosine at 24 nt 3' ends. In support of this latter hypothesis, we observed that the pattern of siRNA strandedness is much greater for *drm2* DMCs as compared to DMCs defined in other methyltransferase mutants that are presumed to operate largely independent from siRNA pathways (Figures S5C-S5E).

DRM2 Preferentially Methylates DsDNA Templates

The observation that preferentially DRM2-targeted cytosines correlate with the same strand as RdDM associated siRNAs suggests that the targeting of DRM2 enzymatic activity by these RNAs is likely through a mechanism other than direct base-pairing between the siRNA and its complementary DNA sequence. Furthermore, previous work suggests that AGO4 and its associated siRNAs interact with nascent Pol V transcripts (Wierzbicki et al., 2009). These

observations suggest models in which DRM2 might directly methylate single-stranded DNA, or perhaps DNA-RNA hybrids produced by annealing of the Pol V transcript with the complementary DNA strand (Figures 7A-7B). To shed additional light on these potential models, we performed *in vitro* methylation reactions with NtDRM MTase using a variety of templates. We observed robust methylation on a dsDNA template, but not on single-stranded DNA or DNA-RNA hybrids (Figure 6E). Therefore, it seems likely that despite being targeted by siRNAs and nascent non-coding Pol V RNAs, DRM2 is likely methylating duplex DNA. In order to reconcile this observation with the strand-biased nature of siRNA-guided DNA methylation, we hypothesize that the tethering of the AGO4-siRNA complex to a Pol V transcript positions DRM2 to methylate the Pol V template strand of DNA near the Pol V exit channel where perhaps the structure of the Pol V complex or associated proteins allows for the transfer of strand information for DRM2 target selection (Figure 7C). This hypothesis is consistent with previous observations that AGO4 is physically associated with the CTD of Pol V (El-Shami et al., 2007; Li et al., 2006). In this model, AGO4 would acts as a bridge between the siRNA-Pol V transcript and DRM2.

CONCLUSION

The results reported here provide molecular details on the functioning of plant *de novo* DNA methyltransferases. Our structural data reveal that, despite its rearranged structure, DRM shares a classic class-I methyltransferase fold with other known class-I methyltransferases. We also uncovered that DRM forms a homo-dimer and that dimerization is essential for catalytic activity, and *in vivo* function. These results suggest a conserved mechanism for eukaryotic *de novo* DNA methyltransferases, in which dimerization is commonly used to initiate DNA methylation. A key

finding from our *in vivo* analysis is that DRM2 interacts with AGO4, and that DRM2-mediated strand-biased DNA methylation is correlated with strand-biased siRNAs. These results are consistent with a model in which DRM2 is acting on DNA immediately after Pol V transcription such that one of the two DNA strands is a preferential target.

EXPERIMENTAL PROCEDURES

Plant Materials

The *drm1/2* mutant plants were previously described (Cao and Jacobsen, 2002). Myc-DRM2, Myc-DRM2cat and Myc-DRM2uba transgenic lines were previously described (Henderson et al., 2010).

Construction of Vectors and Generation of Transgenic Plants

DNA fragments containing NtDRM MTase and NtDRM MTase-M5 were amplified by PCR and were cloned into pENTRD_TOPO vector (Invitrogen) to create pENTRD_NtDRM and pENTRD_NtDRM-M5. These constructs were recombined into the binary vectors pEarleyGate202 and pEarleyGate201 (Earley et al., 2006) to create FLAG and HA fusions, respectively. Each construct was then introduced into *Agrobacterium* AGL1 cells, which were used subsequently to infiltrate leaves of *Nicotiana benthamiana.* A pENTRD vector containing a genomic fragment of DRM2 with N-terminal fusion of 3xFLAG-9xMYC was mutated to generate DRM2-M5 mutant lines using the multiquick change kit (Stratagene). These constructs were recombined into a modified pEarlyGate302 binary vector as previously described (Du et al., 2012). These constructs were transformed into *drm1 drm2* mutant. The detail information for oligos can be found in Table S2.

Protein Preparation

A construct encoding the *Nicotiana tabacum* DRM MTase domain (255-608) was inserted into a self-modified vector, which fuses an N-terminal hexa-histidine plus a yeast sumo tag to the target gene. The plasmid was transformed into *E. coli* strain BL21 (DE3) RIL (Stratagene). The cells

were cultured at 37 °C till OD_{600} reached 1.0, then the media was cooled to 17 °C and 0.2 mM IPTG was added to induce protein expression overnight. The hexa-histidine-sumo tagged protein was initially purified using a HisTrap FF column (GE Healthcare). Then, the tag was cleaved by Ulp1 protease, which was subsequently removed by a second step HisTrap FF column purification. The pooled target protein was further purified by a Heparin FF column (GE Healthcare) and a Hiload Superdex G200 16/60 column (GE Healthcare) with buffer 300 mM NaCl, 20 mM Tris pH 8.0, and 5 mM DTT. The Se-methionine substituted protein was expressed in Se-methionine (Sigma) containing M9 medium and purified using the same protocol as the wild-type protein. The NtDRM-M5 (E283S/R309S/F310S/Y590S/D591S) mutant was generated using a Phusion Site-Directed Mutagenesis Kit (New England Biolabs) and was expressed and purified with the same protocol as the wild-type protein. For enzymatic assays, full length NtDRM (1-608) was cloned into the same vector and expressed and purified with the same protocol as the MTase domain.

Crystallization

Before crystallization, the purified proteins were concentrated to 8 mg/ml and mixed with sinefungin at a molar ratio of 1:3 at 4 °C for 30 minutes. Crystallization was conducted at 20 °C using the hanging drop vapor diffusion method. The wild-type NtDRM MTase was crystallized under 0.19 M $CaCl_2$, 5% glycerol, 26.6% PEG400, and 0.095 M HEPES, pH 7.5 condition. The Se-methionine substituted NtDRM MTase was crystallized under 4.5 M NaCl, and 0.1 M HEPES, pH 7.5 condition. The $NtDRM^{M5}$ was crystallized under 0.2 M sodium nitrate, 20% PEG3350, and 0.1 M BisTris propane, pH 7.5 condition. All the crystals were soaked into the reservoir solution supplement with 15% glycerol for 1 minute. Then the crystals were mounted

on a nylon loop and flash-cooled into liquid nitrogen. The diffraction data for the Se-methionine substituted NtDRM MTase were collected at beamline BL17U, Shanghai Synchrotron Radiation Facility (SSRF), Shanghai, China. The diffraction data for the wild-type and mutant NtDRM MTase were collected at beamline 24IDE, Advanced Photon Source (APS) at the Argonne National Laboratory, Chicago. All the data were processed with the program HKL2000 (Otwinowski and Minor, 1997). The statistics of the diffraction data are summarized in Table S1.

Structure Determination and Refinement

The structure of NtDRM MTase in the presence of sinefungin was solved using single-wavelength anomalous dispersion method as implemented in the program Phenix (Adams et al., 2010). The model building was carried out using the program Coot (Emsley et al., 2010). Because the Se anomalous data had strong anisotropy and significant twin fraction, a rough model was build based on the anomalous data and the model was subsequently used as the search model to perform molecular replacement for the native data. The molecular replacement and structural refinement were carried out using the program Phenix (Adams et al., 2010). Throughout the refinement, a free *R* factor was calculated using 5% random chosen reflections. The stereochemistry of the structural models was analyzed using the program Procheck (Laskowski et al., 1993). The structure of NtDRM-M5 was solved using molecular replacement method with the program Phenix and refined with the same protocol as the wild-type protein (Adams et al., 2010). The statistics of the refinement and structure models are shown in Table S1. All the molecular graphics were generated with the program Pymol (DeLano Scientific LLC).

Affinity Purification and Mass Spectrometry

Approximately 10 g flowers from 9xMyc-BLRP-DRM2 or WT (negative control) were ground and resuspended in 50 mL of Lysis Buffer (LB: 50 mM Tris pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 10% glycerol, 0.1% NP-40, 0.5 mM DTT, 1 μg/μl pepstatin, 1 mM PMSF and 1 protease inhibitor cocktail tablet (Roche, 14696200)). The resulting crude cell extracts were incubated with 200 μl of monoclonal c-Myc 9E10 agarose beads (Covance; AFC-150P) at 4ºC for 2-3 hours. The bead-bound complex was then washed two times with 40 mL of LB and four additional times with 1 mL of LB by mixing at 4ºC for 5 mins each wash. Bound proteins were released by two times 10-min incubation with 400 μl of 8 M UREA at room temperature. The eluted protein complexes were precipitated by trichloroacetic acid and subjected to mass spectrometric analyses as previously described (Du et al., 2012). The interaction between DRM2-AGO4 was performed by using 1.5 g of flowers from Myc-DRM2 transgenic plants and WT plants. The powders were resuspended in 3 ml of Low Salt Lysis Buffer (50 mM Tris pH 7.5, 50 mM NaCl, 5 mM $MgCl_2$, 10% glycerol, 0.1% NP-40, 0.5 mM DTT, 1 μg/μl pepstatin, 1 mM PMSF and 1 protease inhibitor cocktail tablet). The presence of AGO4 was determined by anti-AGO4 antibody (a gift from Dr. Craig Pikaard, Indiana University) at a dilution of 1:1000.

Co-Immunoprecipitation Analyses

The *Nicothiana Benthamiana* leaves (1.5 g) co-expressing FLAG-tagged and HA-tagged NtDRM MTase were grinded in liquid nitrogen and resuspended in 10 ml of LB buffer. Lysates were cleared by filtration through miracloth followed by centrifugation at 13,200 rpm for 10 minutes at 4°C. The supernatants were incubated with 50 ul M2 FLAG magnetic beads (50% slurry, Sigma M8823) for 40 minutes at 4°C with rotation. The beads were then washed 5 times

with 1 ml of LB buffer with incubation of 5 minutes between each wash. The copurification of HA-DRM was deteced by using ANTI-HA-Peroxydase High Affinity 3F10 antibody (Roche 13800200). All western blots were developed using ECL Plus Western Blotting Detection System (GE healthcare RPN2132).

DNA Methyltransferase Activity Assays

The methyltransferase assay was modified from previous studies (Du et al., 2012; Wada et al., 2003). Briefly, the activity assay was carried out at room temperature for one hour in a total volume of 25 μl containing 2.5 μl *of S*-adenosyl-l- [methyl-^{3}H] methionine (SAM) (15 Ci/mmol, GE Healthcare), 125 ng substrate DNA and 100 ng NtDRM protein in assay buffer (20 mM MOPS, pH 7.0, 1 mM DTT, 5 mM EDTA, 200 μg/ml BSA and 25% glycerol) and stopped by placing tubes into dry ice/ethanol bath and subsequently adding 2 μl *of* Proteinase K. 10 μl from each reaction was applied onto DE81 paper (Whatman) and washed two times with 200 mM ammonium bicarbonate, two times with water and two times with ethanol. The paper was dried and placed into liquid scintillation cocktail (Ecolite, MP) and the activity was measured by Beckman scinallation counter, model LS1701 (U.K.). The DNA oligos JP3010 and JP3011 were annealed and purified from as previously described (Du et al., 2012).

Whole-Genome Bisulfite Sequencing

Libraries were prepared as previously described (Stroud et al., 2013) and sequenced on an Illumina HiSeq instrument. Alignment of resulting reads and methods for calculating percent methylation shown in Figure 4A are also as previously described (Stroud et al., 2013). Percent methylation for Figure 3E (complementation of the DRM2-M5 mutant) was calculated similarly

although the reads were aligned using the BSmap program (Xi and Li, 2009). For consistency, the Col WT and DRM2cat data shown in Figure 4A were also remapped with BSmap. The *drm2* DMRs were defined as previously described (Greenberg et al., 2013) and the *drm2* mutant methylome was previously described (Stroud et al., 2013).

Strand Specificity Analysis

To define siRNA clusters with a strand bias, we used a previously defined set of total DRM2-depended siRNA clusters as well as previously published small RNA sequencing datasets (Law et al., 2013) (GSE45368). Small RNA coverage of both the positive and negative strands at these clusters was calculated using unique and non-redundant reads. Strand bias value was calculated as number of aligning reads on the positive strand divided by the number of aligning reads on the negative strand. To avoid artifacts of low coverage we did not consider clusters that were in the bottom 25th percentile of coverage by small RNA reads. To classify clusters as biased, we chose clusters in the top 10 percentiles of bias values (positive-strand bias) and those clusters in the bottom 10 percentiles of bias values (negative-strand bias). Neutral clusters (those without strand bias) were defined as clusters with bias values in 40th to 60th percentiles of bias values. Cytosine bias over the resulting groups of small RNA clusters was calculated by simply tallying the number of cytosines on either strand. To calculate methylcytosine bias over these regions, we defined methylcytosines from a wildtype bisulfite library (GSE49090) using a methodology similar to that previously described (Lister et al., 2009) with the exception that an FDR<0.001 was used and the chloroplast genome was used to control for bisulfite conversion efficiency.

To define the hypomethylated cytosines in Figure 6C, we compared a *drm2* methylome (GSE39901) to a wildtype methylome (GSE49090) and called individual hypomethylated

cytosines of as those significantly hypomethylated ($p<0.001$, Fisher's exact test). The other methyltransferase methylomes were published previously (GSE39901). To avoid over-sampling small RNA profiles from clusters of hypomethylated cytosines groups of DMCs within 24 nt of each other were sampled as to only have one DMC. Small RNA reads from three wildtype libraries (Law et al., 2013; Stroud et al., 2014) (GSE45368, GSE49090, GSE52041) were plotted about these identified hypomethylated cytosines.

Accession Codes

Coordinates and structure factors for NtDRM MTase domain and NtDRM-M5, both in the presence of sinefungin, have been deposited in the RCSB Protein Data Bank with the accession code: 4ONJ and 4ONQ, respectively. Sequencing data were deposited into GEO with the accession number GSE54944.

SUPPLEMENTAL INFORMATION

Supplemental data include 5 figures and 2 tables.

AUTHOR CONTRIBUTIONS

X.Z., J.D., D.J.P., and S.E.J. designed the project; X.Z., J.D., J.G.B, S.F., A.A.V. and J.A.W. performed experiments; X.Z., J.D., C.J.H. D.J.P., J.C. and S.E.J. analyzed the data. X.Z., J.D., C.J.H., and S.E.J. wrote the manuscript.

ACKNOWLEDGENTS

We thank the staff members at Shanghai Synchrotron Radiation Facility (SSRF) and Advanced

Photon Source (APS) for their support in diffraction data collection and at UCLA BSCRC BioSequencing core for high-throughput sequencing. We are grateful to Craig Pikaard for discussions on strand-biased DNA methylation, to Dr. Jianping Ding for access to data collection at the SSRF, and to Dr. Eerappa Rajakumara for assistance with cloning. X. Z. is a research fellow of Ruth L. Kirschstein National Research Service Award (F32GM096483-01). C.J.H. is a HHMI Fellow of the Damon Runyon Cancer Research Foundation. S.F. is a Special Fellow of the Leukemia & Lymphoma Society. J.G-B is a Human Frontiers Science Program Fellow (LT000425/2012-L). This work was supported by Abby Rockefeller Mauze Trust and the Maloris and STARR foundations to D.J.P. NIH grant GM089778 and the UCLA Jonsson Cancer Center to J.A.W. and NIH grant GM60398 to S.E.J. S.E.J. is an Investigator of the Howard Hughes Medical Institute.

FIGURE LEGENDS

Figure 1. Overall Structure of NtDRM

(A) Color-coded domain architecture of AtDRM2, NtDRM, and NtDRM MTase domain used to grow crystal. UBA stands for ubiquitin-associated domain.

(B) Ribbon representation of the overall structure of NtDRM MTase domain dimer with bound sinefungin. One monomer (Mol A) is colored in green and the other one (Mol B) in magenta. The sinefungin cofactors bound to each monomer are shown in space filling model.

See also Figures S1 and Table S1.

Figure 2. Structural Basis of the Domain Rearrangement Mechanism

(A) The structure of NtDRM MTase in two orientations rotated by 90°. The catalytic domain is colored in magenta and the TRD in blue.

(B) The schematic representation of the secondary structural assembly of NtDRM. The catalytic domain and TRD are as indicated, respectively. The disordered catalytic loop (CL) is shown by a dashed line. The break point corresponding to the N and C termini of Dnmt3a is indicated by an arrow.

(C) Superposition of NtDRM monomer with Dnmt3a. NtDRM is colored the same as in Figure 2A and Dnmt3a is in silver. The N and C termini of the two proteins are indicated, respectively. The initiation site of NtDRM MTase domain, Pro288, is highlighted to be near the C terminus of the protein.

Figure 3. NtDRM MTase Forms a Homo-Dimer and Dimerization is Required for Catalytic Activity

(A) Upon superposition of Dnmt3a with one monomer of NtDRM MTase, the other NtDRM MTase monomer can be well superposed with the Dnmt3a dimerized Dnmt3L. NtDRM MTase is colored as in Fig. 1B and Dnmt3a-Dnmt3L dimer is in silver.

(B) Detailed interaction of the NtDRM MTase homo-dimer interface. The interacting residues are shown in stick representation and hydrogen bonds are shown by dashed red lines.

(C) Coimmunoprecipitation assays confirming that DRM2 forms multimers in *Nicotiana Benthamiana*.

(D) *In vitro* methyltransferase activity assays on NtDRM MTase and dimerization-disrupting mutant NtDRM-M5. Error bars represent standard deviation for three replicates.

(E) Boxplot of CHH methylation at *drm2* CHH hypomethylated DMRs in wild type (WT), a *drm2* mutant transformed with a catalytic mutant DRM2 transgene (DRM2cat), a wild type transgene (DRM2) and a dimerization disruptive mutant (DRM2-M5).

See also Figures S2-S3 and Table S2.

Figure 4. UBA Domains are Required for Global DNA Methylation

(A) Boxplots showing the DNA methylation at CG, CHG and CHH contexts for wild type (WT), *drm2* mutant, or wild-type DRM2, catalytic mutant (DRM2cat) or UBA mutant (DRM2uba) transformed back into *drm2,* respectively.

(B) *In vitro* methyltransferase activity assays on full length NtDRM and truncated NtDRM with catalytic domain (NtDRM MTase). Error bars represent standard deviation for three replicates.

See also Figure S4.

Figure 5. DRM2 is Associated With AGO4 *In Vivo*

(A) Summary of proteins associated with DRM2 identified by MS. Only proteins represented in both replicas are shown. NSAF, normalized spectral abundance factor.

(B) Affinity purification confirming DRM2-AGO4 interaction.

Figure 6. SiRNA Strand Biases are Correlated With DNA Methylation Strand Biases

(A) Separation of DRM2-dependent siRNA clusters into strand-biased siRNA clusters and clusters with no strand bias, and assessment of methylcytosine and cytosine strand bias over these clusters.

(B) The direction of methylcytosine strand bias correlates with the direction of siRNA strand bias at strand-biased siRNA clusters.

(C) Number and context of identified hypomethylated differentially methylated cytosines (DMCs) in a *drm2* mutant.

(D) Plot of the relative number of 24 nt siRNA 5' ends around *drm2* DMCs for siRNAs homologous to the same strand as the DMCs or the opposite strand.

(E) NtDRM exhibits robust methyltransferase activity on duplex DNA templates, but not on single-stranded DNA or RNA-DNA hybrids. ssDNA: single-stranded DNA; DNA-RNA: DNA and RNA hybrid; dsDNA: double-stranded DNA; dsDNA-Me: double-stranded control premethylated DNA. Error bars represent standard deviation for three replicates.

See also Figure S5.

Figure 7. Models for the Strand-Specific Nature of DRM2 Methylation

(A) DRM2 activity on single-stranded DNA exposed by Pol V transcription. The activity would be positioned by DRM2's interaction with AGO4, as well as the AGO4-siRNA interaction with

Pol V and basepairing to the nascent Pol V transcript (orange).

(B) DRM2 activity on a RNA-DNA hybrid formed by interaction between the Pol V transcript and template DNA strand. DRM2 activity is positioned by the interaction with AGO4 and the AGO4-siRNA interaction with Pol V and base pairing to the coding DNA strand displaced by Pol V transcription.

(C) DRM2 activity on a double-stranded DNA template formed immediately after passage of the transcription bubble wherein DRM2 activity and strand selection is mediated by an interaction with AGO4. The AGO4-siRNA complex would be mediated by an interaction with Pol V and base-pairing with the nascent Pol V transcript. The solid and open lollipops in A-C represent methylated and unmethylated cytosines, respectively.

REFERENCES

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W.*, et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta crystallographica Section D, Biological crystallography *66*, 213-221.

Cao, X., and Jacobsen, S.E. (2002). Role of the arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing. Current biology : CB *12*, 1138-1144.

Cao, X., Springer, N.M., Muszynski, M.G., Phillips, R.L., Kaeppler, S., and Jacobsen, S.E. (2000). Conserved plant genes with similarity to mammalian de novo DNA methyltransferases. Proc Natl Acad Sci U S A *97*, 4979-4984.

Cheng, X., Kumar, S., Posfai, J., Pflugrath, J.W., and Roberts, R.J. (1993). Crystal structure of the HhaI DNA methyltransferase complexed with S- adenosyl-L-methionine. Cell *74*, 299-307.

Cokus, S.J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C.D., Pradhan, S., Nelson, S.F., Pellegrini, M., and Jacobsen, S.E. (2008). Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature *452*, 215-219.

Du, J., Zhong, X., Bernatavichute, Y.V., Stroud, H., Feng, S., Caro, E., Vashisht, A.A., Terragni, J., Chin, H.G., Tu, A.*, et al.* (2012). Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. Cell *151*, 167-180.

Earley, K.W., Haag, J.R., Pontes, O., Opper, K., Juehne, T., Song, K., and Pikaard, C.S. (2006). Gateway-compatible vectors for plant functional genomics and proteomics. The Plant journal : for cell and molecular biology *45*, 616-629.

El-Shami, M., Pontier, D., Lahmy, S., Braun, L., Picart, C., Vega, D., Hakimi, M.A., Jacobsen, S.E., Cooke, R., and Lagrange, T. (2007). Reiterated WG/GW motifs form functionally and

evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. Genes & development *21*, 2539-2544.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta crystallographica Section D, Biological crystallography *66*, 486-501.

Finnegan, E.J., and Dennis, E.S. (1993). Isolation and identification by sequence homology of a putative cytosine methyltransferase from Arabidopsis thaliana. Nucleic Acids Res *21*, 2383-2388.

Finnegan, E.J., and Kovac, K.A. (2000). Plant DNA methyltransferases. Plant Mol Biol *43*, 189-201.

Greenberg, M.V., Deleris, A., Hale, C.J., Liu, A., Feng, S., and Jacobsen, S.E. (2013). Interplay between active chromatin marks and RNA-directed DNA methylation in Arabidopsis thaliana. PLoS genetics *9*, e1003946.

Henderson, I.R., Deleris, A., Wong, W., Zhong, X., Chin, H.G., Horwitz, G.A., Kelly, K.A., Pradhan, S., and Jacobsen, S.E. (2010). The de novo cytosine methyltransferase DRM2 requires intact UBA domains and a catalytically mutated paralog DRM3 during RNA-directed DNA methylation in Arabidopsis thaliana. PLoS genetics *6*, e1001182.

Jia, D., Jurkowska, R.Z., Zhang, X., Jeltsch, A., and Cheng, X. (2007). Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature *449*, 248-251.

Johnson, L.M., Du, J., Hale, C.J., Bischof, S., Feng, S., Chodavarapu, R.K., Zhong, X., Marson, G., Pellegrini, M., Segal, D.J*., et al.* (2014). SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. Nature. doi: 10.1038/nature12931.

Laskowski, R.A., Macarthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst *26*, 283-291.

Law, J.A., Du, J., Hale, C.J., Feng, S., Krajewski, K., Palanca, A.M., Strahl, B.D., Patel, D.J., and Jacobsen, S.E. (2013). Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. Nature *498*, 385-389.

Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nature reviews Genetics *11*, 204-220.

Lee, T.F., Gurazada, S.G., Zhai, J., Li, S., Simon, S.A., Matzke, M.A., Chen, X., and Meyers, B.C. (2012). RNA polymerase V-dependent small RNAs in Arabidopsis originate from small, intergenic loci including most SINE repeats. Epigenetics : official journal of the DNA Methylation Society *7*, 781-795.

Li, C.F., Pontes, O., El-Shami, M., Henderson, I.R., Bernatavichute, Y.V., Chan, S.W., Lagrange, T., Pikaard, C.S., and Jacobsen, S.E. (2006). An ARGONAUTE4-containing nuclear processing center colocalized with Cajal bodies in Arabidopsis thaliana. Cell *126*, 93-106.

Lindroth, A.M., Cao, X., Jackson, J.P., Zilberman, D., McCallum, C.M., Henikoff, S., and Jacobsen, S.E. (2001). Requirement of *CHROMOMETHYLASE3* for maintenance of CpXpG methylation. Science *292*, 2077-2080.

Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., and Ecker, J.R. (2008). Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell *133*, 523-536.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M.*, et al.* (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature *462*, 315-322.

Luo, S., and Preuss, D. (2003). Strand-biased DNA methylation associated with centromeric regions in Arabidopsis. Proc Natl Acad Sci U S A *100*, 11133-11138.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode. Methods in Enzymology *276*, 307-326.

Qi, Y., He, X., Wang, X.J., Kohany, O., Jurka, J., and Hannon, G.J. (2006). Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. Nature *443*, 1008-1012.

Schubert, H.L., Blumenthal, R.M., and Cheng, X. (2003). Many paths to methyltransfer: a chronicle of convergence. Trends in biochemical sciences *28*, 329-335.

Song, J., Rechkoblit, O., Bestor, T.H., and Patel, D.J. (2011). Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. Science *331*, 1036-1040.

Song, J., Teplova, M., Ishibe-Murakami, S., and Patel, D.J. (2012). Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. Science *335*, 709-712.

Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Johnson, L., Patel, D.J., and Jacobsen, S.E. (2014). Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. Nature structural & molecular biology *21*, 64-72.

Stroud, H., Greenberg, M.V., Feng, S., Bernatavichute, Y.V., and Jacobsen, S.E. (2013). Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. Cell *152*, 352-364.

Wada, Y., Ohya, H., Yamaguchi, Y., Koizumi, N., and Sano, H. (2003). Preferential de novo methylation of cytosine residues in non-CpG sequences by a domains rearranged DNA methyltransferase from tobacco plants. J Biol Chem *278*, 42386-42393.

Wierzbicki, A.T., Ream, T.S., Haag, J.R., and Pikaard, C.S. (2009). RNA polymerase V transcription guides ARGONAUTE4 to chromatin. Nature genetics *41*, 630-634.

Xi, Y., and Li, W. (2009). BSMAP: whole genome bisulfite sequence MAPping program. BMC bioinformatics *10*, 232.

Zemach, A., Kim, M.Y., Hsieh, P.H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer, S.L., and Zilberman, D. (2013). The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell *153*, 193-205.

Zhang, H., Ma, Z.Y., Zeng, L., Tanaka, K., Zhang, C.J., Ma, J., Bai, G., Wang, P., Zhang, S.W., Liu, Z.W.*, et al.* (2013). DTF1 is a core component of RNA-directed DNA methylation and may assist in the recruitment of Pol IV. Proc Natl Acad Sci U S A *110*, 8290-8295.

Zhong, X., Hale, C.J., Law, J.A., Johnson, L.M., Feng, S., Tu, A., and Jacobsen, S.E. (2012). DDR complex facilitates global association of RNA polymerase V to promoters and evolutionarily young transposons. Nature structural & molecular biology *19*, 870-875.

SUPPLEMENTAL FIGURE LEGENDS

Figure S1. Structure based sequence alignment of NtDRM and AtDRM2 catalytic domains with the secondary structure of NtDRM catalytic domain labelled on the top. The conserved residues forming the NtDRM dimer interface are marked with black dots at the bottom of the alignment. Related to Figure 1

Figure S2. (A) The superposition of structures of wild type NtDRM MTase and the multiple mutant of NtDRM, which disrupts the dimer interface shows the mutant protein shares the same monomer structure as wild type counterpart. (B) Western Blot of DRM2 protein levels of each representative line of FLAG-DRM2 and FLAG-DRM2-M5. Related to Figure 3

Figure S3. A Model Positioning DNA Within the Active Site of NtDRM, Related to Figure 1 and 2

(A) A electrostatics surface representation of NtDRM in two orientations rotated by 90°. The TRD and active site form a continuous positively charged surface cleft, which is suitable for the DNA substrate binding.

(B) A model positioning DNA onto the NtDRM structure based on the DNMT1-DNA active complex. The DNA can be well positioned into the cleft between the TRD and catalytic domains. The flipped out cytosine base can insert into the active site which is near the cofactor sinefungin. The two a-helices of the TRD approach the major groove of the substrate DNA and might play a role in the DNA recognition.

Figure S4. Loss of the DRM2 UBA Domain Affects Some *drm2* CHH DMRs More Strongly Than Others, Related to Figure 4

(A) Kernel density plots of mean weighted change in CHH methylation ((mutant - WT)/ mean

(mutant, WT)) for drm2 mutant and transgenic lines over drm2 DMRs. The orange dotted line represents the median change and the gray dotted lines represent the 25th and 75th percentiles of change.

(B) Metaplots of DNA methylation in a wildtype genome at drm2 DMRs strongly (>75th percentile of mean-weighted change) and weakly (<25th percentile of mean-weighted change) hypomethylated in the CHH context in the DRM2uba mutant.

(C) Metaplots of two broadly heterochromatin chromatin marks at drm2 DMRs strongly and weakly affected in DRM2uba.

(D) Western blot of DRM2 protein levels of DRM2, DRM2cat and DRM2uba lines.

Figure S5. Relationship Between DNA Methylation and SiRNA Strand and Sequence Composition, Related to Figure 6

(A) The direction of siRNA strand bias over strand-biased drm2-dependent siRNA clusters correlates with the direction of total cytosine strand bias on these clusters.

(B) The sequence composition of non-redundant, uniquely mapping 24 nt siRNAs showing an A-bias at the 5' position and a C/T bias at the 3' position.

(C-E) Number of called differentially methylated cytosines (DMCs) for different DNA methyltransfersase mutants and the distribution of siRNA 5' ends around each class of DMC.

Table S1. Summary of X-ray Diffraction Data and Structure Refinement Statistics, Related Figure 1

Summary of diffraction data			
Crystal	Se-NtDRM + sinefungin	NtDRM + sinefungin	NtDRM-M5 + sinefungin
Beamline	SSRF-BL17U	APS-24IDE	APS-24IDE
Wavelength (Å)	0.9795	0.9792	0.9792
Space group	$P4_3$	$P3_121$	$P6_5$
Cell parameters			
a, b (Å)	83.8	153.3	131.5
c (Å)	154.9	149.0	88.2
Resolution (Å)	50.0-3.3 (3.42-3.30) [a]	50.0-2.8 (2.90-2.80)	50.0-2.5 (2.59-2.50)
R_{merge} (%)	14.7 (68.4)	12.2 (72.3)	7.1 (69.4)
Observed reflections	97,226	348,674	335,243
Unique reflections	15,822	49,582	30,076
Redundancy	6.1 (6.4)	7.0 (7.3)	11.1 (11.2)
Average I/σ(I)	21.4 (3.6)	24.2 (2.6)	45.6 (3.7)
Completeness (%)	99.5 (100.0)	99.9 (100.0)	100.0 (100.0)
Refinement and structure model			
R factor / Free *R* factor (%□)		20.1 / 22.1	18.5 / 23.0
Number of non-H atoms		5,386	5301
Protein		5,332	5134
Sinefungin		54	54
Water		-	113
Average B factor (Å^2)		81.5	62.6
Protien		81.1	62.6
Sinefungin		115.4	85.9
Water		-	53.3
RMS deviations			
Bond lengths (Å)		0.007	0.015
Bond angles (°)		1.288	1.410

[a] Values in parentheses are for highest-resolution shell.

Table S2. List of Primers Used for Cloning NtDRM, NtDRM-M5, AtDRM2 and AtDRM2-M5. Related Figures 3 and 5

Name	Primer sequences from 5' to 3'
NtDRM point mutation R309S F310S	TCTTTGTACGATATTGAGCCAGAGTTTG AGAAGAAATTGTGTCCCACACAC
NtDRM point mutation Y590S D591S	TCTTATGTTCGGATATTGGACTTGGTC AGAAAATAGGGAAGACTCTTTGCC
NtDRM point mutation E283S	CAAGCTGTTGGACCCCCGTTTTTC AGAGGGAAGAGTTCTTCGAACCATTG
NtDRM 255-608 cDNA cloning into pENTR/D	CACCGAGACAATTCGTTTGCCCAAACC CTAATGTTTATGTCTGGACATTATGG
Full length genomic AtDRM2 cloning into pENTR/D	CACCGTAATGGAGATAGCTTCTCAGGATTATC AACCAGATTGGGGCAATATACATATAGAAGAGCC
AtDRM2 point mutation primer 1	CACATAGATCGCTTCCATCGTTAGCCCGAGGGCCAC
AtDRM2 point mutation primer 2	GTTTGGGAGACTATTTCCAGctcCTTGTTCGAGATCCCACCTG
AtDRM2 point mutation primer 3	CTTGAAGGTGATCAATCTTCGTTGTTCTcTtcGTATTGCCGTATTC TAG
Quikchange primers to destroy EcoR I site of AtDRM2	CTAAAAATTCTCTGGGTATTCGCCCTTACT CTAGTAAGGGCGAATACCCAGAGAATTTTTAG

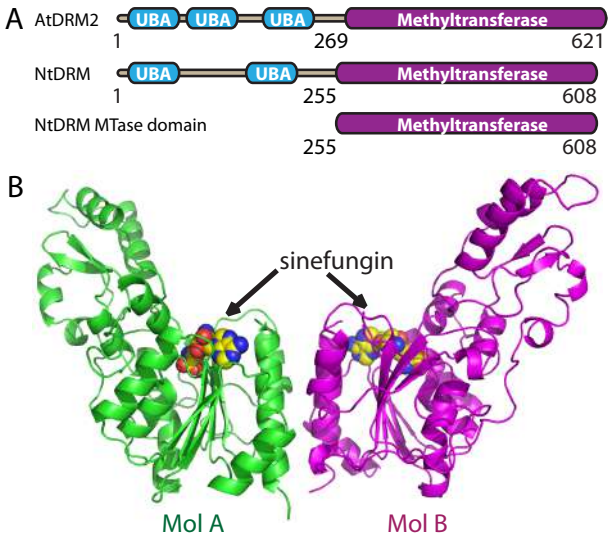

Figure 1. Overall structure of NtDRM. (A) Color-coded domain architecture of AtDRM2, NtDRM, and NtDRM MTase domain used to grow crystal. (B) Ribbon representation of the overall structure of NtDRM MTase domain dimer with bound sinefungin. One monomer (Mol A) is colored in green and the other one (Mol B) in magenta. The sinefungin cofactors bound to each monomer are shown in space filling model.

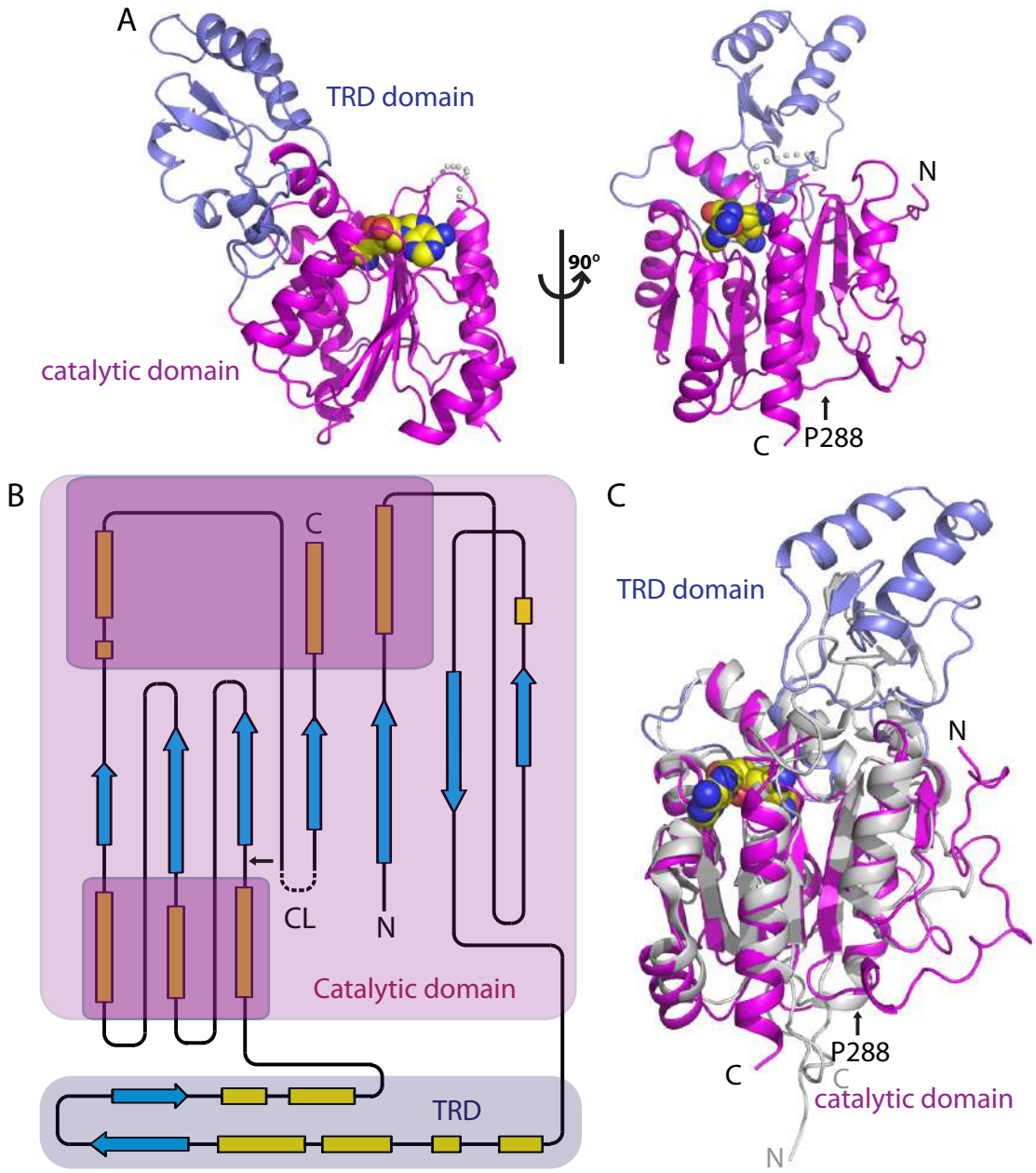

Figure 2. Structural basis of the domain rearrangement mechanism. (A) The structure of NtDRM MTase in two orientations rotated by 90 °. The catalytic domain is colored in magenta and the TRD in blue. (B) The schematic representation of the secondary structural assembly of NtDRM. The catalytic domain and TRD are as indicated, respectively. The disordered catalytic loop (CL) is shown by a dashed line. The break point corresponding to the N and C termini of Dnmt3a is indicated by an arrow. (C) Superposition of NtDRM monomer with Dnmt3a. NtDRM is colored the same as in Fig. 2A and Dnmt3a is in silver. The N and C termini of the two proteins are indicated, respectively. The initiation site of NtDRM MTase domain, Pro288, is highlighted to be near the C terminus of the protein.

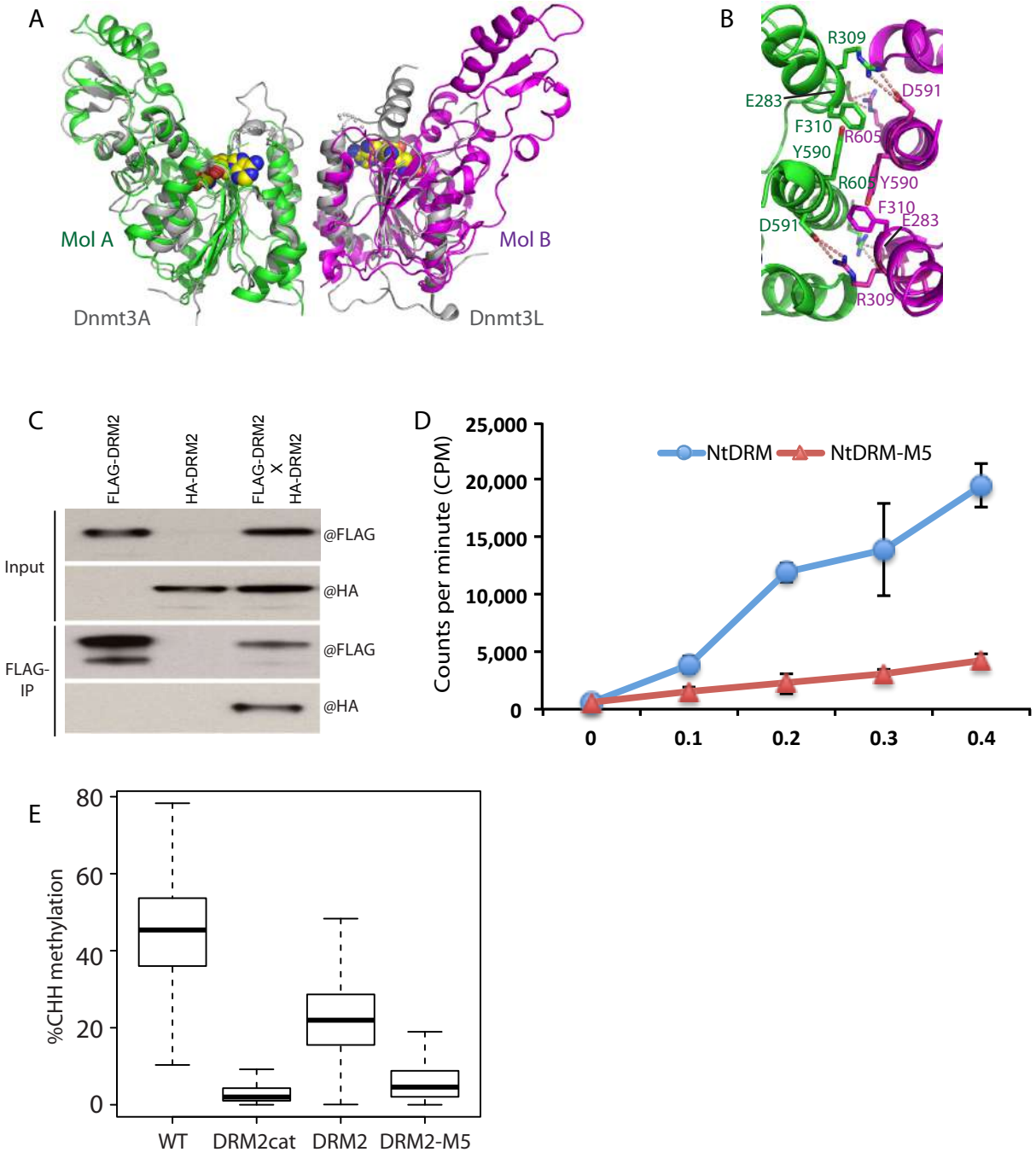

Figure 3. DRM MTase forms a homo-dimer and dimerization is required for catalytic activity. (A) Upon superposition of Dnmt3a with one monomer of NtDRM MTase, the other NtDRM MTase monomer can be well superposed to the Dnmt3a dimerized Dnmt3L. NtDRM MTase is colored as in Fig. 1B and Dnmt3a-Dnmt3L dimer is in silver. (B) The detailed interaction of the NtDRM MTase homo-dimer interface. The interacting residues are shown in stick representation and hydrogen bonds are shown by dashed red lines. (C) Coimmunoprecipitation assays confirming DRM2 forms dimer in Nicotiana Benthamiana. (D) The in vitro methyltransferase activity assay on NtDRM MTase and dimerization-disrupting mutant NtDRM-M5. Error bars represent standard deviation for three replicates. (E) Boxplot of CHH methylation at drm2 CHH hypomethylated DMRs in WT (Col), a drm2 mutant transformed with a catalytic mutant DRM2 transgene (DRM2cat), a wildtype transgene (DRM2) and a dimerization disruptive mutant (DRM2-M5).

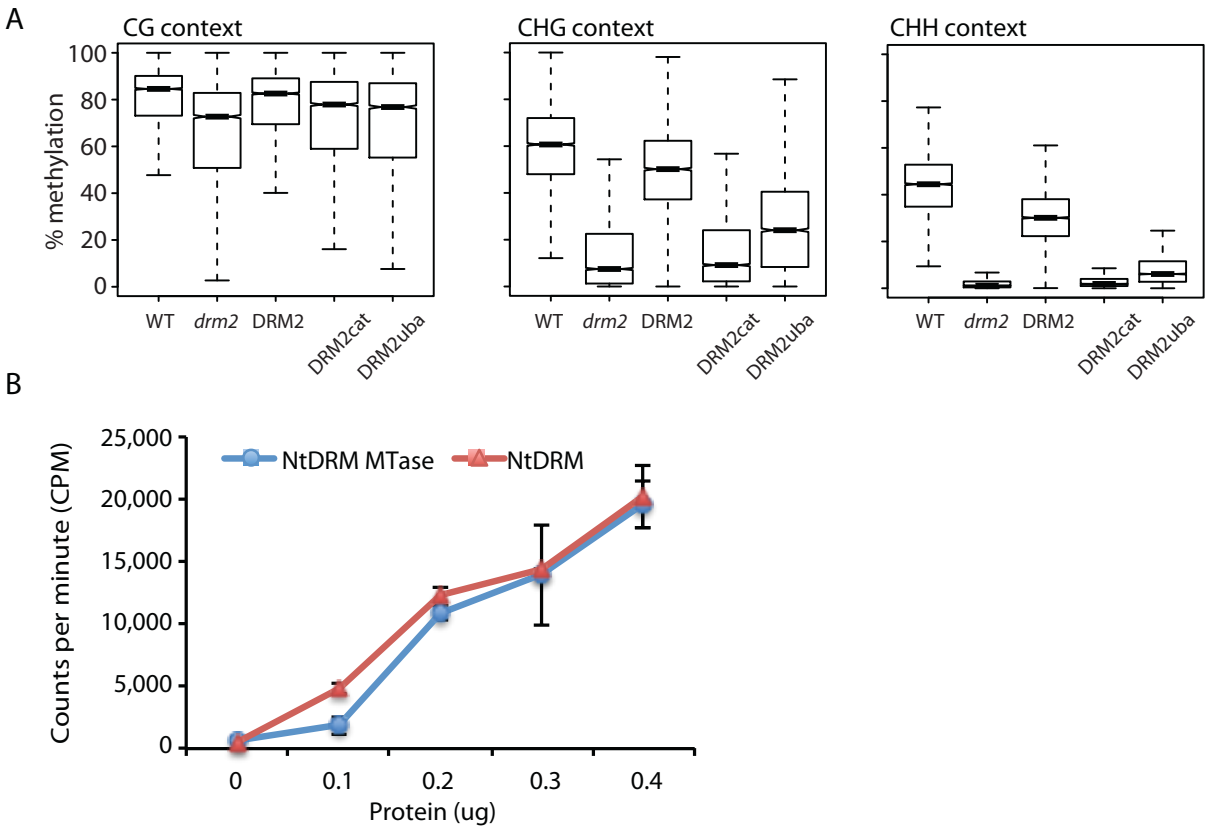

Figure 4. UBA domains are required for global DNA methylation. (A) Boxplots showing the DNA methylation at CG, CHG and CHH contexts for wild type (WT), drm2 mutant, or wild-type DRM2, catalytic mutant (DRM2cat) or UBA mutant (DRM2uba) transformed back into drm2, respectively. (B) In vitro methyltransferase activity assays on full length NtDRM and truncated NtDRM with catalytic domain (NtDRM MTase). Error bars represent standard deviation for three replicates.

A

	Experiment I				Experiment II			
	Spectra	Unique Peptides	% Coverage	NSAF	Spectra	Unique Peptides	% Coverage	NSAF
DRM2	310	37	43.3	1025	331	32	45.8	2500
At2g27040 (AGO4)	35	20	22.1	78	15	8	10.4	77
At5g03740 (HD2C)	11	6	25.8	79	5	3	11.8	82
At5g21150 (AGO9)	11	10	12.9	25	4	2	2.5	21
At3g16830	11	7	5.4	20	5	5	4.6	21
At2g19520 (MSI4)	25	8	21.3	75	3	2	5.9	28
At3g45980 (HTB9)	10	5	35.3	138	12	5	35.3	379
At1g75950	4	2	17.5	51	2	2	13.8	59

MS analyses of DRM2 purifications (Only proteins present in both experiments are shown)

B

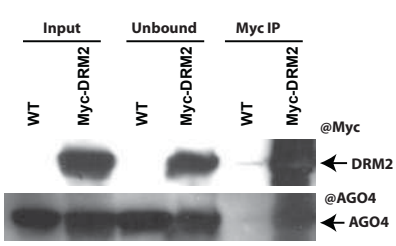

Figure 5. DRM2 is associated with AGO4 in vivo. (A) Summary of proteins associated with DRM2 identified by MS. Only proteins represented in both replicas are shown. NSAF, normalized spectral abundance factor. (B) Affinity purification confirming DRM2-AGO4 interaction.

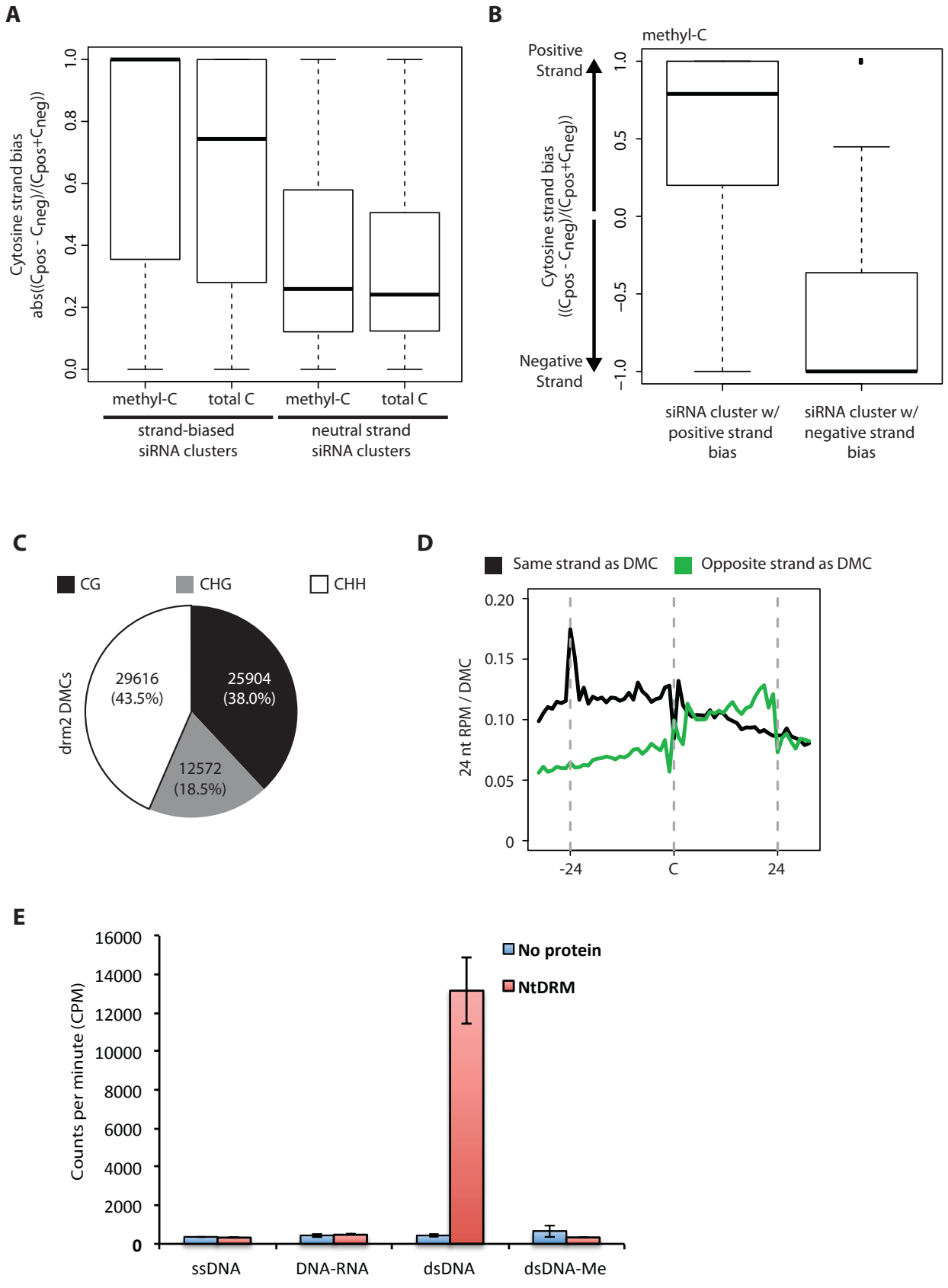

Figure 6. siRNA strand biases are correlated with DNA methylation strand biases. (A) Separation of DRM2-dependent siRNA clusters into strand-biased siRNA clusters and clusters with no strand bias, and assessment of methylcytosine and cytosine strand bias over these clusters. (B) The direction of methylcytosine strand bias correlates with the direction of siRNA strand bias at strand-biased siRNA clusters. (C) Number and context of identified hypomethylated differentially methylated cytosines (DMCs) in a drm2 mutant. (D) Plotting of relative number of 24 nt siRNA 5′ ends around drm2 DMCs for siRNAs homologous to the same strand as the DMCs or the opposite strand. (E) NtDRM exhibits robust methyltransferase activity on duplex DNA templates, but not on single-stranded DNA or RNA-DNA hybrids. ssDNA: single-stranded DNA; DNA-RNA: DNA and RNA hybrid; dsDNA: double-stranded DNA; dsDNA-Me: double-stranded control premethylated DNA. Error bars represent standard deviation for three replicates.

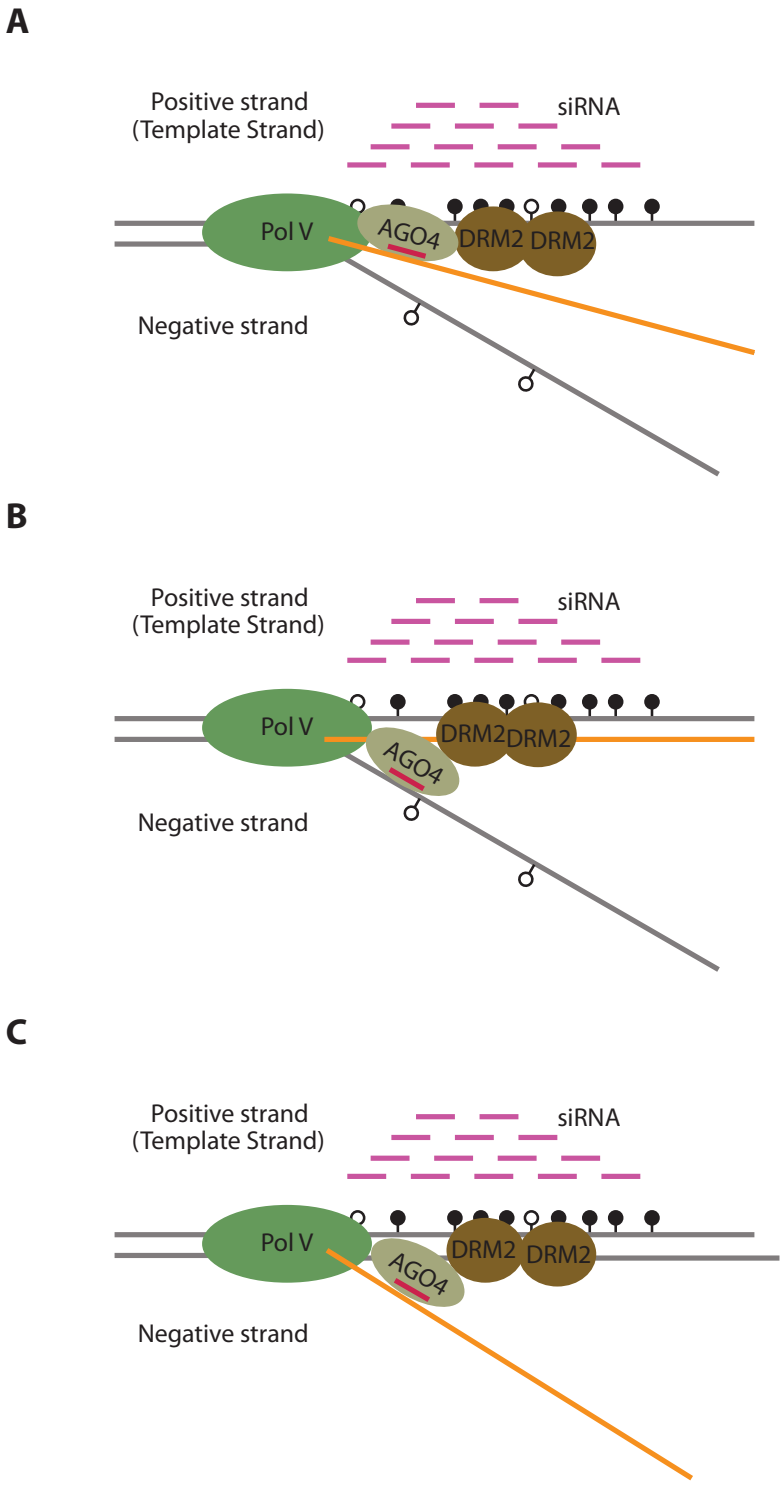

Figure 7. Models for the strand-specific nature of DRM2 methylation. (A) DRM2 activity on single-stranded DNA exposed by Pol V transcription. The activity would be positioned by DRM2's interaction with AGO4, as well as the AGO4-siRNA interaction with Pol V and basepairing to the nascent Pol V transcript (orange). (B) DRM2 activity on a RNA-DNA hybrid formed by interaction between the Pol V transcript and template DNA strand. DRM2 activity is positioned by the interaction with AGO4 and the AGO4-siRNA interaction with Pol V and basepairing to the coding DNA strand displaced by Pol V transcription. (C) DRM2 activity on a double-stranded DNA template formed immediately after passage of the transcription bubble wherein DRM2 activity and strand selection is mediated by an interaction with AGO4. The AGO4-siRNA complex would be mediated by an interaction with Pol V and base-pairing with the nascent Pol V transcript.

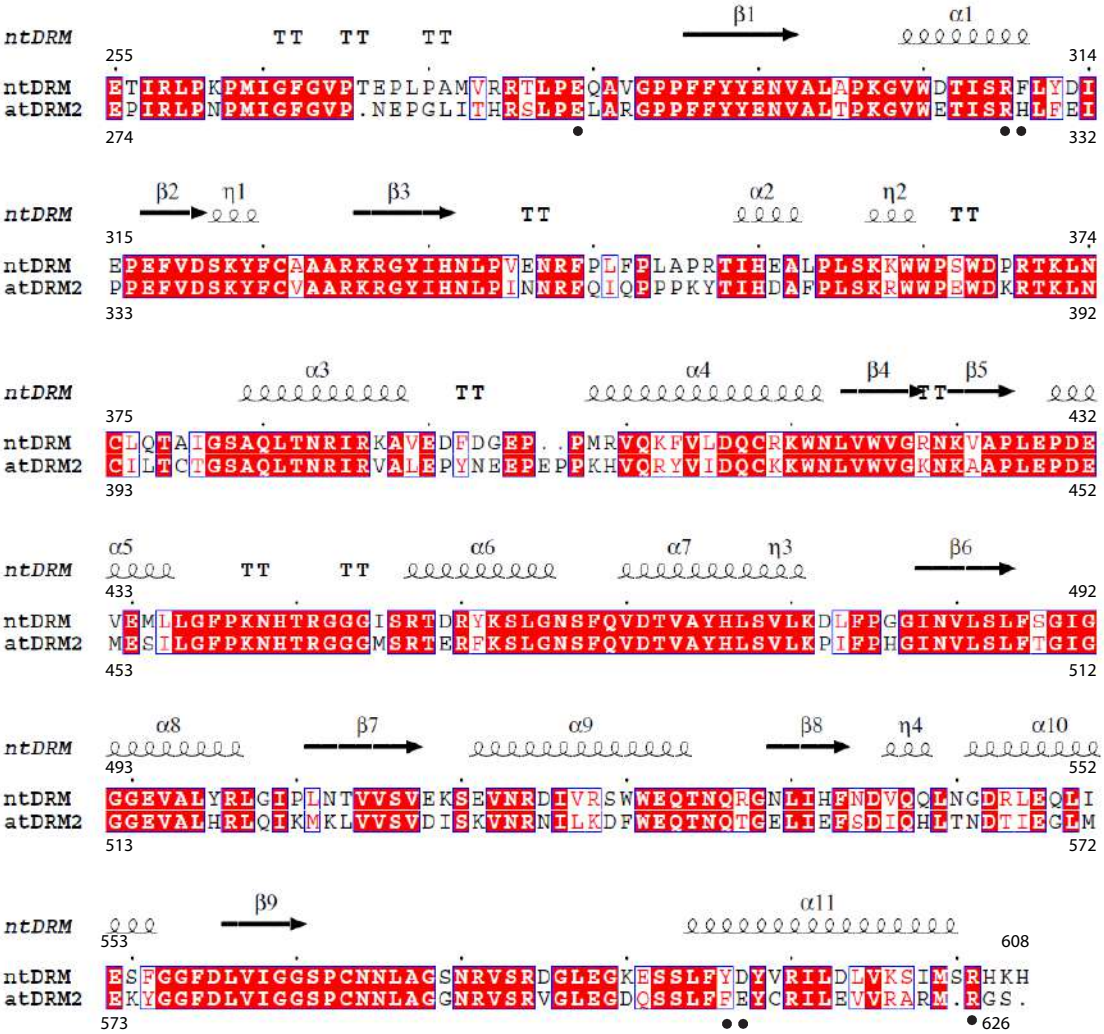

Figure S1. Structure based sequence alignment of NtDRM and AtDRM2 catalytic domains with the secondary structure of NtDRM catalytic domain labelled on the top. The conserved residues forming the NtDRM dimer interface are marked with black dots at the bottom of the alignment. Related to Figure 1.

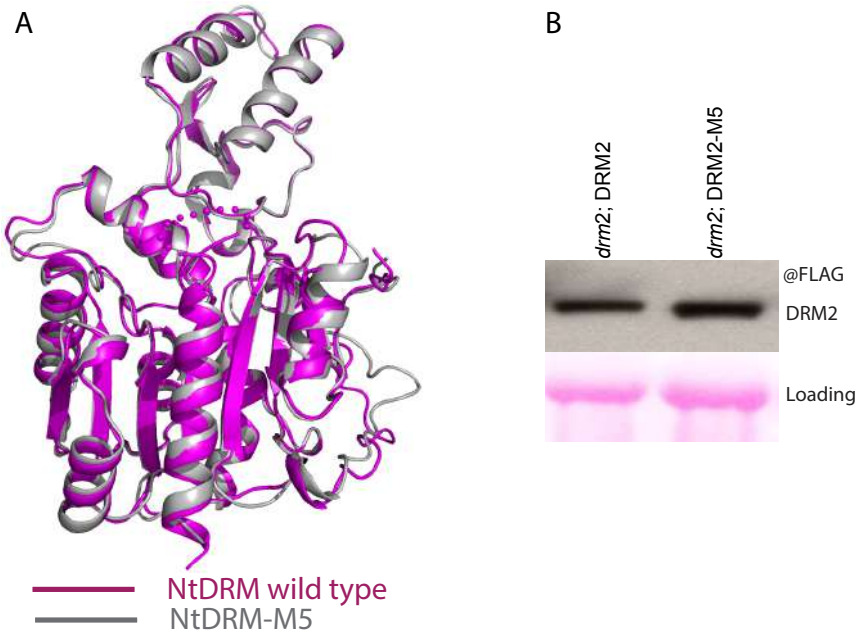

Figure S2. (A) The superposition of structures of wild type NtDRM MTase and the multiple mutant of NtDRM which disrupts the dimer interface shows the mutant protein shares the same monomer structure as wild type counterpart. (B) Western blot of DRM2 protein levels of each representative line of FLAG-DRM2 and FLAG-DRM2-M5. Related to Figure 3.

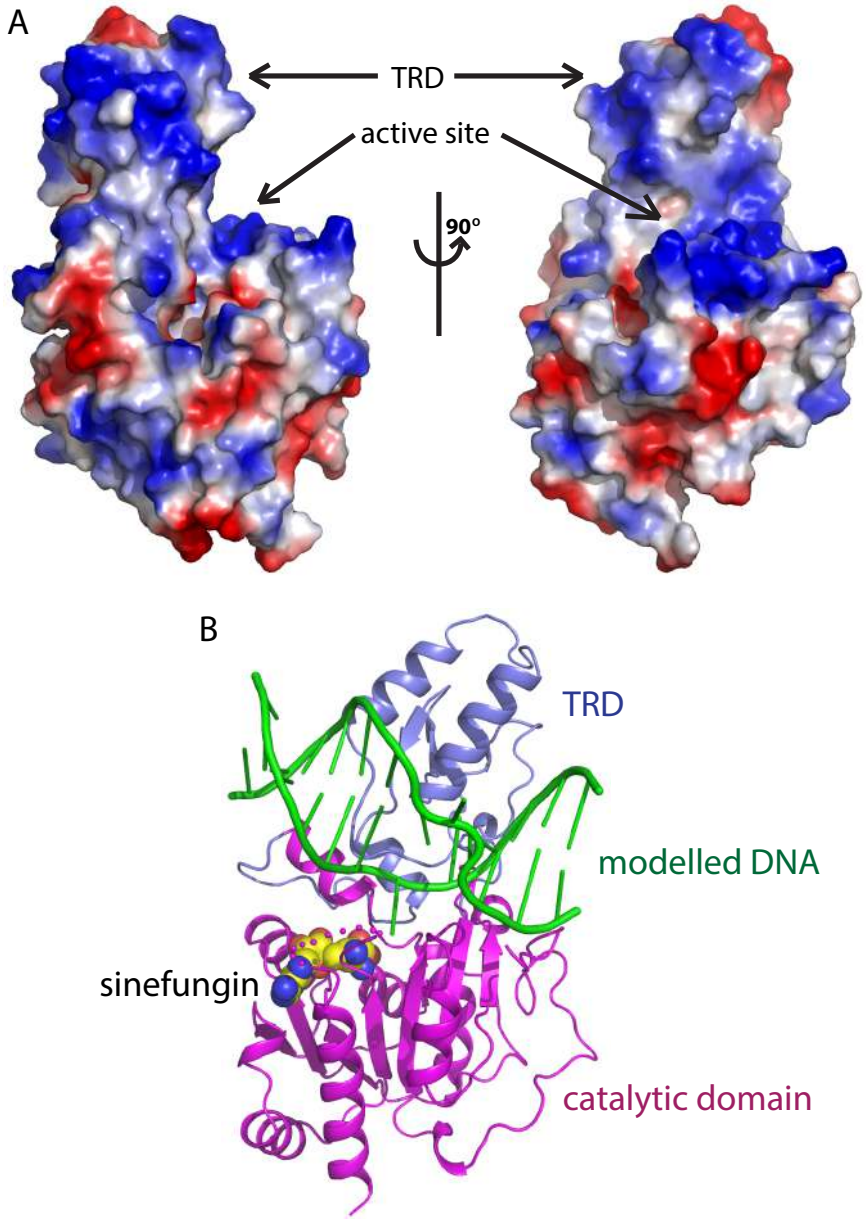

Figure S3. A model positioning DNA within the active site of NtDRM. (*A*) A electrostatics surface representation of NtDRM in two orientations rotated by 90 °. The TRD and active site form a continuous positively charged surface cleft which is suitable for the DNA substrate binding. (*B*) A model positioning DNA onto the NtDRM structure based on the DNMT1-DNA active complex. The DNA can be well positioned into the cleft between the TRD and catalytic domains. The flipped out cytosine base can insert into the active site which is near the cofactor sinefungin. The two α-helices of the TRD approach the major groove of the substrate DNA and might play a role in the DNA recognition. Related to Figures 1 and 2.

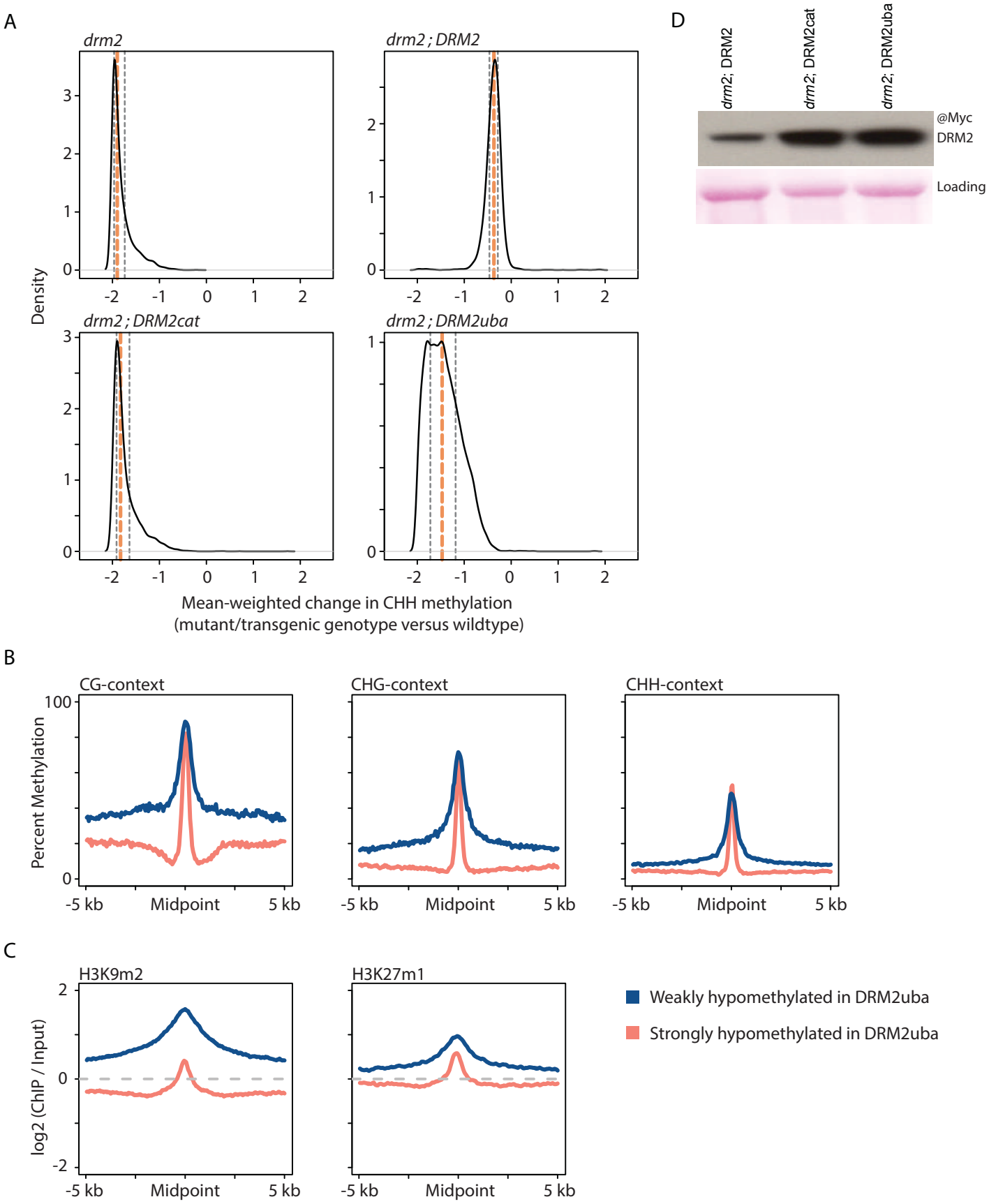

Figure S4. Loss of the DRM2 UBA domain affects some *drm2* CHH DMRs more strongly than others. (A) Kernel density plots of mean weighted change in CHH methylation ((mutant - WT)/ mean (mutant, WT)) for *drm2* mutant and transgenic lines over *drm2* DMRs. The orange dotted line represents the median change and the gray dotted lines represent the 25th and 75th percentiles of change. (B) Metaplots of DNA methylation in a wildtype genome at *drm2* DMRs strongly (>75th percentile of mean-weighted change) and weakly (<25th percentile of mean-weighted change) hypomethylated in the CHH context in the DRM2uba mutant. (C) Metaplots of two broadly heterochromatin chromatin marks at *drm2* DMRs strongly and weakly affected in DRM2uba. (D) Western blot of DRM2 protein levels of DRM2, DRM2cat and DRM2uba lines. Related to Figure 4.

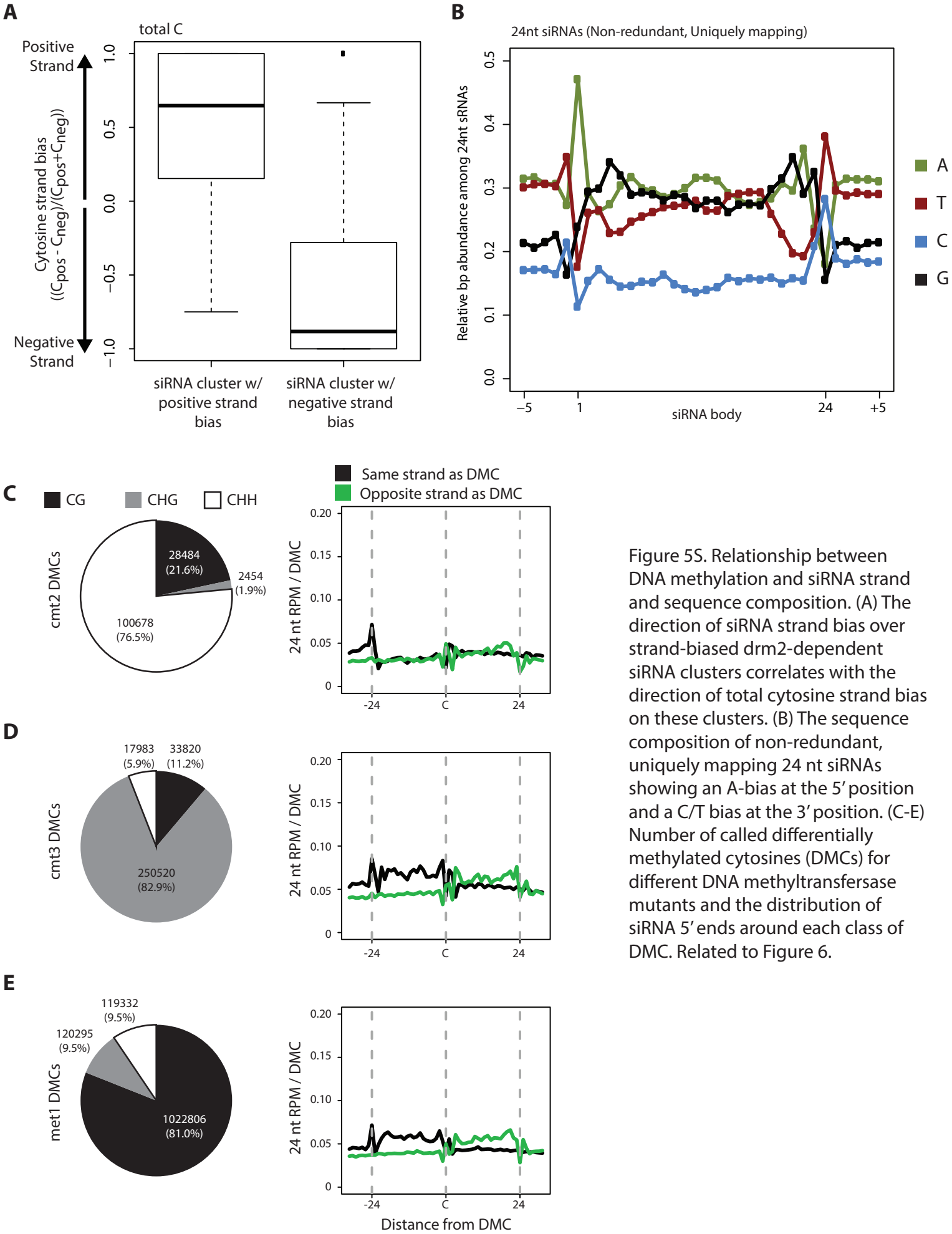

Figure 5S. Relationship between DNA methylation and siRNA strand and sequence composition. (A) The direction of siRNA strand bias over strand-biased drm2-dependent siRNA clusters correlates with the direction of total cytosine strand bias on these clusters. (B) The sequence composition of non-redundant, uniquely mapping 24 nt siRNAs showing an A-bias at the 5′ position and a C/T bias at the 3′ position. (C-E) Number of called differentially methylated cytosines (DMCs) for different DNA methyltransfersase mutants and the distribution of siRNA 5′ ends around each class of DMC. Related to Figure 6.