SUMMARY

DNA methylation and histone modification exert epigenetic control over gene expression. CHG methylation by CHROMOMETHYLASE3 (CMT3) depends on histone H3K9 dimethylation (H3K9me2), but the mechanism underlying this relationship is poorly understood. Here, we report multiple lines of evidence that CMT3 interacts with H3K9me2-containing nucleosomes. CMT3 genome locations nearly perfectly correlated with H3K9me2, and CMT3 stably associated with H3K9me2-containing nucleosomes. Crystal structures of maize CMT3 homolog ZMET2, in complex with H3K9me2 peptides, showed that ZMET2 binds H3K9me2 via both bromo adjacent homology (BAH) and chromo domains. The structures reveal an aromatic cage within both BAH and chromo domains as interaction interfaces that capture H3K9me2. Mutations that abolish either interaction disrupt CMT3 binding to nucleosomes and show a complete loss of CMT3 activity in vivo. Our study establishes dual recognition of H3K9me2 marks by BAH and chromo domains and reveals a distinct mechanism of interplay between DNA methylation and histone modification.

INTRODUCTION

DNA methylation and histone modification are two major epigenetic marks regulating gene expression and chromatin state. Eukaryotic DNA is wrapped around a histone octamer that consists of two of each of the four core histones H2A, H2B, H3, and H4 and forms the basic structural unit of chromatin: the nucleosome (Luger et al., 1997). The formation of heterochromatin involves the recruitment of silencing complexes containing histone-binding and -modifying proteins and the generation of specific silent epigenetic modification patterns. In animals and Schizosaccharomyces pombe, methylation of histone H3 at lysine 9 (H3K9) is associated with silenced heterochromatic regions and is bound by the silencing protein Heterochromatin-associated protein 1 (Jenuwein and Allis, 2001). In plants, methylation of H3K9 (H3K9me2) is required for the silencing of transposable elements (TEs) and other repetitive DNA, which are enriched in heterochromatic regions (Bernatavichute et al., 2008; Jackson et al., 2002, 2004; Malagnac et al., 2002).

DNA methylation in mammals is mainly found at CG sites and is initially catalyzed by the de novo methyltransferase DNMT3A/B and subsequently maintained by DNMT1. In contrast, DNA methylation in plants exists in three sequence contexts: CG, CHG (where H is either C, T, or A), and CHH (or asymmetric) (Cokus et al., 2008). While METHYLTRANSFERASE1 (MET1, an ortholog of mammalian DNMT1) maintains CG methylation and DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2, an ortholog of mammalian DNMT3) primarily maintains CHH methylation, CHROMOMETHYLASE3 (CMT3) is a plant-specific methyltransferase responsible for CHG methylation in Arabidopsis thaliana (Bartee and Bender, 2001; Law and Jacobsen, 2010; Lindroth et al., 2001). The CMT3 ortholog in maize, the Zea mays methyltransferase2 (ZMET2), was also shown to be required for CHG DNA methylation (Papa et al., 2001).

DNA methylation and histone modifications have been correlated in multiple organisms. In mammals, de novo methyltransferase DNMT3A and its cofactor DNMT3L form a complex that specifically recognizes unmodified H3K4 through the ADD domains (Ooi et al., 2007; Otani et al., 2009). UHRF1 is a key factor connecting DNMT1 and H3K9 methylation by binding methylated DNA through the SRA domain (Bostick et al., 2007), H3K9me3 through a tandem Tudor (Nady et al., 2011; Xie et al., 2012) and unmodified H3R2 through an atypical PHD
finger domain (Rajakumara et al., 2011b; Xie et al., 2012). In plants, H3K9me2 is mainly associated with heterochromatic TEs and correlated with CHG methylation (Bernatavichute et al., 2008). The SRA domain of the H3K9 methyltransferase KRYPTONITE (KYP; also called SUVH4) directly binds to methylated CHG-containing oligonucleotides (Johnson et al., 2007). Based on these results, a self-reinforcing loop was proposed, in which CMT3 is recruited by H3K9me2 deposited by KYP and its close homolog (SUVH5 and SUVH6), to methylate CHG. In turn, methylated CHG DNA recruits KYP to maintain methylation of H3K9 (Johnson et al., 2007; Law and Jacobsen, 2010).

To investigate the molecular mechanism underlying the relationship between CHG methylation and histone H3K9 methylation, we carried out structural and functional studies on CMT3 and ZMET2. We found that CMT3 is stably associated with heterochromatic nucleosomes. Consistent with its role in DNA methylation maintenance, we found that CMT3 was predominantly expressed in actively replicating cells and was specifically associated with replication-dependent histone H3 variants. Genome-wide mapping of CMT3 binding sites demonstrated that CMT3 is nearly perfectly correlated with H3K9me2 in vivo. Chromatin association of CMT3 is dependent on specific binding of H3K9me2 via the bromo adjacent homology (BAH) and chromo domains as shown by our crystal structures of ZMET2 in complex with H3K9me2 peptides. The structures identify an aromatic cage within both the BAH and the chromo domains as interaction interfaces that capture H3K9me2. Mutations that abolish either interaction cause a failure of CMT3 binding to nucleosomes, and a complete loss of CMT3 activity in vivo. Together, these results suggest that CMT3 associates with H3K9me2-containing nucleosomes through dual binding of its BAH and chromo domains to H3K9me2 in order to target DNA methylation.

**RESULTS**

**CMT3 Is Associated with H3K9me2-Containing Nucleosomes**

To identify CMT3-interacting proteins, we generated epitope-tagged CMT3 lines. As shown in Figure 1A, both pCM3:BLRP-3xFLAG-CMT3 and pCM3:BLRP-9xMYC-CMT3, designated as FLAG-CMT3 and MYC-CMT3, respectively, complemented the cmt3 mutation and restored DNA methylation at the SS rDNA locus, demonstrating that epitope-tagged CMT3 was functional in vivo. We found a major band at the size of CMT3 and a series of smaller bands that corresponded to the size of individual histone proteins (Figure 1B). As shown in Table 1 from mass spectrometry (MS) analysis, peptides corresponding to CMT3 were the most abundant proteins with 63.4% coverage of the CMT3 protein. We also identified all four core histones: H3, H4, H2A, and H2B at roughly equivalent levels and observed several linker H1 histones, as well as Ku70/80 proteins (Table 1; Figure S1 available online) that were likely bound to the sheared DNA ends, suggesting that CMT3 interacts with bona fide nucleosomes.

To test whether CMT3 specifically interacts with histone H3, we treated plant extracts with benzonase, a nonspecific nuclease, to remove DNA prior to affinity purification. We found that benzonase treatment eliminated the low molecular weight histone bands (Figure 1C). MS analysis further confirmed that H2A, H2B, H3, H4, and Ku70/80 were no longer detectable (Table S2). This suggests the integrity of nucleosomes is critical for CMT3 binding, and that CMT3 does not interact with H3 strongly on its own in vivo.

To investigate whether CMT3 is associated with H3K9me2-containing nucleosomes, we tested for the presence of H3K9me2 marks in histones copurified with CMT3 and found that H3K9me2, but not H3K4me2, was precipitated with CMT3 (Figure 1D). This indicates that CMT3 preferentially binds to H3K9me2-containing nucleosomes.

**CMT3 Is Enriched in Heterochromatic Regions and Highly Correlated with H3K9me2**

To further explore the in vivo binding pattern of CMT3, we performed chromatin immunoprecipitation coupled with sequencing (ChIP-seq) (Figure 1E). We found that CMT3 is highly enriched in Arabidopsis pericentromeric regions and is nearly perfectly colocalized with H3K9me2 (Figure 2A). CMT3 is also enriched in heterochromatic patches in the euchromatin arms, where high levels of H3K9me2 were also observed (Figure 2B). CMT3 tends to bind large uninterrupted blocks in pericentromeric regions, while in chromosome arms CMT3 forms smaller and isolated patches (Figure 2C). These results are consistent with the general trend of H3K9me2 distribution in these regions (Bernatavichute et al., 2008). Furthermore, CMT3 was not enriched over protein coding genes (Figure 2D), consistent with previous data showing that the majority of genes are devoid of CHG methylation and H3K9me2 (Cokus et al., 2008). We also found a strong enrichment of CMT3 over TEs (Figure 2E). To complement the ChIP-seq data, we examined the transcriptome profile of cmt3 mutants by whole genome RNA sequencing (RNA-seq) analysis. As expected, CMT3 was strongly enriched over the TEs that were upregulated in cmt3 mutants (Figure 2F), as were H3K9me2 and CHG methylation (Figures 2G and S1A).

To determine the specificity of CMT3 binding to histone marks in vitro, we screened full-length CMT3 on a peptide array and found that CMT3 can bind to mono-, di-, and trimethylated H3K9 peptides, with preference for di- and trimethylated H3K9 peptides (Figure 2H). In plants, H3K9me2 is highly enriched in heterochromatin and TEs, and functionally associated with DNA methylation. In contrast, H3K9me3 is found in low abundance in euchromatic regions in plants and its functional role is still unknown (Roudier et al., 2011). In addition, CMT3 binding was blocked by phosphorylation of Ser10 and Thr11 on H3 (Figure 2H).

Collectively, the nearly perfect correlation between CMT3 binding sites with H3K9me2 sites in the genome and the direct interaction of CMT3 and H3K9me2-containing nucleosomes and histone tail peptides support the role of H3K9me2 in recruiting CMT3 to chromatin in vivo.

**CMT3 Is Primarily a CHG Methyltransferase In Vitro and Is Predominantly Expressed in Actively Replicating Cells**

To determine whether CMT3 is active in vitro, recombinant CMT3 protein was assayed on a set of oligonucleotide
substrates with cytosines either unmethylated or premethylated at CG, CHG, or CHH sites (Table S3). As shown in Figure 3A, CMT3 methylated oligonucleotides that were either unmethylated or hemimethylated at CHG sites, while it lost the majority of its activity when CHG sites were blocked by premethylation on both strands. We also performed bisulfite sequencing of an in vitro methylated plasmid DNA and observed 10% CHG methylation, 1.4% CHH methylation, and 0% CG methylation, with an error rate of 0.4% (Figure 3B). These results provide direct evidence that CMT3 is indeed a methyltransferase that preferentially methylates CHG sites.

During DNA replication, fully methylated DNA becomes hemimethylated, and can become completely unmethylated in the following round of replication if methylation is not properly maintained. If CMT3 efficiently converts hemimethylated DNA to fully methylated DNA in vivo, then one should observe a large number of fully methylated CHG sites and a smaller number of hemimethylated sites in the genome. To test this, we utilized Hairpin Bisulfite Sequencing (Laird et al., 2004) at the Ta3 transposon. We found that the number of observed fully methylated dyads exceeded the number of expected fully methylated dyads occurring at random by a small but significant number (Figure 3C). These results suggest that CMT3 has a significant in vivo preference for conversion of hemimethylated sites to fully methylated sites and in this way acts as a maintenance methyltransferase.

Our MS results showed that the replication-dependent H3 (H3.1), known to be expressed in S phase (Okada et al., 2005), but not the replication-independent H3 (H3.3), copurified with CMT3 (Table S4). In addition, an analysis of publicly available microarray data shows that CMT3 expression is highly
correlated with proteins that are actively expressed during DNA replication (Figure S1B). We also directly examined the localization of CMT3 in replicating root cells. To detect cells undergoing replication, 5-ethynyl-2'-deoxyuridine (EdU), a thymidine analog that is incorporated into dividing cells during S phase (Chehrehasa et al., 2009), was used to label newly synthesized DNA. We observed high accumulations of CMT3 in actively replicating cells as labeled by EdU (Figure 3D). CMT3 also accumulated in cells not undergoing DNA replication, but at a much lower level than that in replicating cells (Figure 3D).

Together, these results suggest that CMT3 is an active DNA methyltransferase that is predominately expressed in the S phase of actively replicating cells and binds to nucleosomes that contain replication dependent histone variants.

**Structure of SAH-Bound ZMET2**

The CMT3 homolog in *Zea mays*, ZMET2 (Figures 4A and S2A), has a similar binding specificity to methylated H3K9 on the peptide array (Figure 2H) and similar methyltransferase activity on CHG sites (Figure S2B). Although we were unable to

### Table 1. Summary of Proteins Associated with CMT3 Identified by Mass Spectrometric Analyses

<table>
<thead>
<tr>
<th>Protein</th>
<th>AGI Code</th>
<th>Spectra</th>
<th>Unique Peptides</th>
<th>% Coverage</th>
<th>NSAFe5</th>
<th>% CMT3</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMT3</td>
<td>At1g69770</td>
<td>756</td>
<td>131</td>
<td>63.4</td>
<td>7,240</td>
<td>100</td>
</tr>
<tr>
<td>H1</td>
<td>At1g06760</td>
<td>19</td>
<td>11</td>
<td>31.8</td>
<td>557</td>
<td>7.7</td>
</tr>
<tr>
<td>H1</td>
<td>At3g18035</td>
<td>29</td>
<td>17</td>
<td>22.3</td>
<td>485</td>
<td>6.7</td>
</tr>
<tr>
<td>H1/H5</td>
<td>At1g48620</td>
<td>28</td>
<td>12</td>
<td>17.3</td>
<td>470</td>
<td>6.5</td>
</tr>
<tr>
<td>H1.2</td>
<td>At2g30620</td>
<td>46</td>
<td>16</td>
<td>47.3</td>
<td>1,353</td>
<td>18.7</td>
</tr>
<tr>
<td>HTA13</td>
<td>At3g20670</td>
<td>50</td>
<td>7</td>
<td>21.2</td>
<td>3,043</td>
<td>42.0</td>
</tr>
<tr>
<td>HTA6</td>
<td>At5g59870</td>
<td>59</td>
<td>18</td>
<td>17.3</td>
<td>964</td>
<td>13.3</td>
</tr>
<tr>
<td>HTA9</td>
<td>At1g52740</td>
<td>15</td>
<td>4</td>
<td>30.6</td>
<td>899</td>
<td>12.4</td>
</tr>
<tr>
<td>HTA7</td>
<td>At5g27670</td>
<td>15</td>
<td>3</td>
<td>8.0</td>
<td>803</td>
<td>11.1</td>
</tr>
<tr>
<td>HTA12</td>
<td>At5g02560</td>
<td>5</td>
<td>2</td>
<td>7.8</td>
<td>262</td>
<td>3.6</td>
</tr>
<tr>
<td>HTA8</td>
<td>At2g38810</td>
<td>15</td>
<td>4</td>
<td>32.4</td>
<td>251</td>
<td>3.5</td>
</tr>
<tr>
<td>H2B</td>
<td>At5g02570</td>
<td>59</td>
<td>13</td>
<td>32.6</td>
<td>3,675</td>
<td>50.8</td>
</tr>
<tr>
<td>H2B</td>
<td>At3g9480</td>
<td>10</td>
<td>8</td>
<td>30.4</td>
<td>1,863</td>
<td>25.7</td>
</tr>
<tr>
<td>H2B</td>
<td>At3g53650</td>
<td>62</td>
<td>15</td>
<td>46.4</td>
<td>3,610</td>
<td>49.9</td>
</tr>
<tr>
<td>H2B</td>
<td>At3g53650</td>
<td>32</td>
<td>8</td>
<td>30.4</td>
<td>1,863</td>
<td>25.7</td>
</tr>
<tr>
<td>H2B</td>
<td>At1g07790</td>
<td>79.6</td>
<td>16.6</td>
<td>42.2</td>
<td>4,327</td>
<td>59.8</td>
</tr>
<tr>
<td>H2B</td>
<td>At1g07790</td>
<td>15</td>
<td>4</td>
<td>32.6</td>
<td>3,675</td>
<td>50.8</td>
</tr>
<tr>
<td>H3a</td>
<td>At5g10400</td>
<td>63</td>
<td>21</td>
<td>52.9</td>
<td>3,674</td>
<td>50.7</td>
</tr>
<tr>
<td>H3a</td>
<td>At1g09200</td>
<td>59</td>
<td>13</td>
<td>42.1</td>
<td>2,028</td>
<td>28.0</td>
</tr>
<tr>
<td>HTR12</td>
<td>At1g01370</td>
<td>2</td>
<td>2</td>
<td>25.3</td>
<td>90.3</td>
<td>1.2</td>
</tr>
<tr>
<td>H4a</td>
<td>At1g07820</td>
<td>26</td>
<td>13</td>
<td>62.1</td>
<td>2,028</td>
<td>28.0</td>
</tr>
<tr>
<td>Ku80</td>
<td>At1g16970</td>
<td>10</td>
<td>8</td>
<td>16.7</td>
<td>129</td>
<td>1.8</td>
</tr>
</tbody>
</table>

The percentage of CMT3 (% CMT3) column indicates the approximate stoichiometry of each copurifying protein as a function of the normalized spectral abundance factor (NSAF) (Law et al., 2010). See also Table S1.

*Unique peptides map to the group of these genes.
Figure 2. CMT3 Is Enriched in Heterochromatic Regions and Highly Correlated with H3K9me2
(A) Distribution of CMT3 (black) and H3K9me2 (red) along the five Arabidopsis chromosomes.
(B) CMT3 ChIP-seq signal is enriched in heterochromatic patches in the arms.
(C) Genome browser views of pericentromeric and euchromatic regions. mCG, CG methylation; mCHG, CHG methylation; mCHH, CHH methylation; TE, transposable element; PCG, protein coding gene.
(D and E) CMT3 ChIP-seq signal is not enriched in the protein coding gene regions (D), but is enriched in TEs (E).
(F and G) CMT3 (F) and H3K9me2 (G) are enriched in TEs that are upregulated in cmt3 null mutants.
(H) CMT3 and ZMET2 specifically bind to mono-, di-, and trimethylation of H3K9. Blue, red, and yellow circles in right panel represent the locations of mono-, di-, and trimethylation at lysine 9-containing peptides on the array, respectively.
See also Figure S1A.
crystallize CMT3, we were successful in obtaining diffraction quality crystals for an N-terminal truncated version of ZMET2 (residues 130–912, Figure 4A), containing all the functional domains, along with the cofactor SAH and the structure was refined to 3.2 Å resolution (Figure 4B; Table S5). The BAH and chromo domains project outward in opposite directions relative to the methyltransferase domain (Figure 4B). The overall topology is that of an equilateral triangle with each edge of about 85–100 Å in length. The BAH domain, the target recognition sub-domain (TRD) of the methyltransferase domain, and the chromo-domain are positioned at the three vertices, while the catalytic subdomain of the methyltransferase domain is positioned within the interior of the triangular architecture (Figure 4B).

The BAH domain of ZMET2, composed of a twisted β barrel with some extended segments, resembles BAH1 domain of mouse DNMT1. Although they only share a sequence identity of 26%, the superposition of the BAH domains of ZMET2 and mouse DNMT1 yields an rmsd of 1.9 Å for 121 aligned Cα atoms, showing a similar folded topology (Figure S2C). The position of the BAH domain relative to its methyltransferase domain in ZMET2 is also similar to that of BAH1 domain of DNMT1 relative to its methyltransferase domain, indicating a plausible similar function for the BAH domains of these two proteins.

The methyltransferase domain of ZMET2 adopts a typical class-I methyltransferase fold composed of catalytic and TRD subdomains. The catalytic subdomain of ZMET2 resembles the

Figure 3. CMT3 Is Primarily a CHG Methyltransferase and Predominantly Expressed in Actively Replicating Cells
(A) CMT3 has activity on both unmethylated and hemimethylated DNA oligos, but not fully methylated DNA oligos containing methylations at all cytosines. Error bars represent the SD of three replicates.
(B) Bisulfite sequencing of in vitro methylated plasmid DNA.
(C) Hemimethylated dyads at the Ta3 locus were significantly lower than expected (chi-square value: 32.84). Probability of obtaining the observed distribution by random chance is less than 0.1%.
(D) Accumulation of CMT3 in cells with active replication. DAPI staining shows the localization of nuclei (blue). Actively replicating cells were labeled by 5-ethynul-2’-deoxyuridine (EdU, Red) and CMT3 was immunostained by MYC antibody (green).
See also Figure S1B and Tables S3 and S4.
corresponding folds adopted by DNMT1 (Song et al., 2011) and DNMT3A (Jia et al., 2007), with a central seven-stranded β sheet flanked by two layers of α-helical segments on either side. Similar to its topology in DNMT1, the central seven-stranded β sheet of the catalytic subdomain of ZMET2 forms a continuous nine-stranded β sheet by pairing with two additional β strands from the BAH domain (Figure 4B), thereby probably stabilizing the relative positioning of the BAH and methyltransferase domains of ZMET2. A SAH molecule is positioned in the active site of the catalytic subdomain. The TRD subdomain of ZMET2, consisting of 206 residues (612–818), has a well-defined continuous peptide backbone, but poor density for some side chains, which were consequently modeled as polyalanine.

Although the primary sequence implies that the chromodomain of ZMET2 is embedded within the methyltransferase domain, it buds out from one side of the methyltransferase domain in the structure (Figure 4B), therefore forming an independent structural module that does not affect the topology of the methyltransferase domain. The chromodomain uses several hydrophobic residues to form a continuously hydrophobic surface that interacts with its hydrophobic counterpart on the surface of the methyltransferase domain (Figure S2D). This interaction positions the chromodomain on one side of the methyltransferase domain, indicative of a relatively rigid alignment between chromo and methyltransferase domains. The ZMET2 chromodomain adopts a typical chromodomain topology, which resembles that of methylated H3K9-binding chromodomains, such as the MPP8 chromodomain (rmsd of 1.1 Å for 57 aligned Cα atoms) and the HP1 chromodomain (rmsd of 1.2 Å for 52 aligned Cα atoms) (Chang et al., 2011; Jacobs and Khorasanizadeh, 2002).

**Structural Basis for H3(1–15)K9me2 Peptide Recognition by ZMET2 Chromodomain**

To investigate the molecular basis for H3K9me2 recognition by ZMET2, we first performed isothermal calorimetric (ITC) binding studies and established that the chromodomain of ZMET2-bound H3K9me2 and H3K9me3 with a Kd of 2.3 (Figure 4C) and 2.0 μM (data not shown), respectively, whereas it bound only weakly to unmodified H3, as well as H3K27me2 and H4K20me2 peptides (Figure 4C). Unexpectedly, ITC binding studies also established that the BAH domain of ZMET2 also exhibits strong binding affinity for H3K9me2 and H3K9me3 with a Kd of 0.5 μM (Figure 4D) and 0.7 μM (data not shown)
and weak binding for unmodified H3 and H4K20me2 peptides, and exhibited very reduced affinity to H3K27me2 peptide ($K_d$ of 17 μM) (Figure 4D).

We then crystallized ZMET2 in complex with H3(1–15)K9me2 peptide in the presence of SAH at 3.2 Å resolution (Figure 5A; Table S5). There are two molecules of ZMET2 and one molecule of H3K9me2 peptide bound to one of the two chromodomains in the asymmetric unit (Figures S3A and S3B). The overall structure of ZMET2 bound to the H3(1–15)K9me2 peptide resembles the structure of ZMET2 in the free state (Figure 4B and Figure 5A). The H3(1–15)K9me2 peptide binds to the chromodomain in a classic chromodomain-binding mode (Figure 5A) and the bound peptide can be traced from Gln5 to Thr11 (Figure S3C). The peptide is clamped within the chromodomain in a β-strand-like conformation. The K9me2 side chain inserts into an aromatic cage, which is formed by Phe441, Trp466, and Tyr469, and stabilized by extensive hydrophobic interactions (Figures 5B and 5C). These three aromatic cage-forming residues are
conserved with CMT3 and the classic H3K9me3-recognizing HP1 chromodomain (Figures S3D and S3E), suggesting a similar methylated-lysine binding properties within this family. The residues adjacent to K9me2 are also involved in sequence specific recognition, with Arg8 of the histone peptide forming two intermolecular hydrogen bonds, as well as electrostatic interactions with the side chain of Glu440 of the protein. The main chain of Arg8 also forms two hydrogen bonds with Asn481 and the backbone and side chain of Ser10 forms hydrogen bonds with the side chain of Glu477 of the protein (Figures 5B and 5C), consistent with our observation that Ser10 phosphorylation inhibits its binding to CMT3 (Figure 2H). In addition, the backbone of Gin5, Thr6, and Ala7 form intermolecular hydrogen bonds with the protein (Figures 5B and 5C).

**Structural Basis for H3(1–32)K9me2 Peptide Recognition by ZMET2 BAH Domain**

We further determined the structure of another crystal form of ZMET2 in complex with a longer H3(1–32)K9me2 peptide at 2.7 Å resolution that revealed a distinct binding mode of the H3K9me2 peptide to the BAH domain of ZMET2 (Figure 5D; Table S5). Due to the different packing arrangement in this crystal, the peptide–binding site within the chromodomain was blocked by the other molecule in the asymmetric unit (Figures S4A and S4B), and hence no peptide was found to bind to the chromodomain.

The structure contains two molecules of ZMET2, with H3K9me2 peptides bound to each of the BAH domains, in the asymmetric unit (Figure S4A). The packing results in the chromodomain of one molecule being clamped by the TRD subdomain and the catalytic loop of the second molecule (Figure S4A). These intermolecular interactions stabilize the conformation of the TRD subdomain and catalytic loop, and hence it became possible to trace all the side chains of the TRD subdomain and the catalytic loop, while they could not be fully built in the H3(1–15)K9me2 complex. The electron density was traceable from Gin5 to Thr11 for the H3(1–32)K9me2 peptide bound to the BAH domain in one complex (Figure S4C), while a somewhat shorter segment was traceable in the other complex. The side chain of K9me2 inserts into an aromatic cage formed by Tyr203, Trp224, and Phe226 of the BAH domain (Figure 5E). These aromatic residues are also conserved in CMT3, suggesting of a conserved function (Figure S4D). By contrast, flanking peptide residues form fewer interactions with the protein in the H3(1–32)K9me2 complex (Figures 5E and 5F) than they do in the H3(1–15)K9me2 complex (Figures 5B and 5C). The methyl group Ala7 is recognized within a small hydrophobic pocket formed by Val194, Tyr203, and Trp224 in the H3(1–32)K9me2 complex (Figures 5E and 5F). In addition, the backbone amide protons of Ala7, Arg8, and Ser10 form intermolecular hydrogen bonds with the BAH domain (Figures 5E and 5F). These studies establish the BAH domain of ZMET2 also as a reader of H3K9me2, with its aromatic cage playing a key role in methyl-lysine recognition.

ZMET2 has a relatively large TRD subdomain of around 206 residues with a unique fold (Figure SSA). In regions proximal to the catalytic center (middle and bottom segments of Figure S5A), the ZMET2 TRD subdomain adopts 7 helices and loops similar to those observed for DNMT1 (Song et al., 2011), indicating the likelihood of a similar DNA substrate recognition mechanism. However, in regions further from the catalytic center (top segment of Figure S5A), the TRD subdomains of DNMT1 and ZMET2 show different structural features. For DNMT1, regions of the TRD subdomain further from the catalytic center are enriched with loops and a coordinated zinc ion. In addition, a long loop extending outward from BAH2 domain of DNMT1 (Figure S5A) interacts at its tip with the TRD subdomain, suggestive of a plausible regulatory role. By contrast, ZMET2 has neither a second BAH domain, nor a bound zinc ion. The space occupied by the BAH2 loop-interacting region on the TRD subdomain in DNMT1 is partially occupied by a two-stranded anti-parallel β sheet in ZMET2 (Figure S5A). These observations imply that though ZMET2 might have a potentially similar DNA recognition mode as DNMT1, its regulatory mechanism could be different.

The catalytic loop region is disordered for structures of ZMET2 in the free state and in the complex with H3(1–15)K9me2 peptide, but is well defined in the complex with H3(1–32)K9me2 peptide. We observe that the catalytic loop (in red, stereo view in Figure S5B) and catalytic cysteine (labeled C in red, Figure S5B) of ZMET2 in its peptide-bound form is directed outward from the catalytic center, similar to that of DNMT1 bound to DNA in its autoinhibitory conformation (labeled C in blue, Figure S5B) (Song et al., 2011). By contrast, the catalytic loop (in green) and catalytic cysteine (labeled C in green, Figure S5B) are directed inward toward the catalytic center in the active form of M.Hhal bound to DNA (Klimasauskas et al., 1994) and the active form of DNMT1 bound to hemimethylated DNA (Song et al., 2012). These results are consistent with ZMET2 being in an inactive conformation in the peptide-bound state in the absence of bound DNA.

**Both BAH and Chromo Domains Are Important for CMT3 Function In Vivo**

To study the biological significance of CMT3 and H3K9me2 interaction, we mutated the three aromatic residues lining the methyl-lysine-binding aromatic coves in the chromo (Phe382, Trp409, and Tyr412) and the BAH (Phe127, Trp148, and Tyr150) domains to generate triple point mutants named CMT3chr3 and CMT3bah3, respectively. We also generated a catalytically inactive C460A mutant (CMT3cat). Notably, CMT3cat, CMT3chr3 and CMT3bah3 failed to restore DNA methylation at the 5S rDNA (Figure 6A) and at the Ta3 (Figure 6B), even though the mutated proteins were expressed at a similar level (Figure 6C) and CMT3chr3 had similar catalytic activity as wild-type CMT3 (Figure 6D). In summary, mutations in either BAH or chromo domains of CMT3 abolish its methylation activity in vivo.

We also analyzed the effect of the chromo and BAH mutations on the binding of CMT3 to nucleosomes. Silver staining of proteins copurifying with CMT3 showed that, unlike the situation in wild-type CMT3, proteins copurifying with the CMT3chr3 and CMT3bah3 mutants lacked the low molecular weight bands corresponding to histone proteins (Figure 1B), which was further confirmed by MS analysis on affinity-purified CMT3chr3 samples (Table S2) and by ChIP experiments showing that mutations of either BAH or chromodomain abolish CMT3 binding to...
Figure 6. Both BAH and Chromo Domains Are Important for CMT3 Function In Vivo

(A) Southern blot of the 5S rDNA locus.

(B) Graphical representation of bisulfite sequencing analysis of Ta3 locus.

(C) Western blots showing the expression of the wild-type and mutant versions of the CMT3 protein. The bottom panel is the Coomassie blue staining and serves as a loading control.

(D) In vitro activity assay. Error bars represent the standard deviation of three biological replicates.

(E) ChIP data showing the failure of CMT3chr3 and CMT3bah3 to bind to chromatin at Ta3 and SDC. Error bars represent the SD of three biological replicates.

(F) A working model of dual-recognition mechanism of CMT3 BAH and chromo domains simultaneously reading the H3K9me2 on a single nucleosome to position the methyltransferase domain over the nucleosomal DNA.

See also Figures S6 and Table S2.
chromatin (Figure 6E). These results indicate that both the chromo and BAH domains are required for stable association of CMT3 with nucleosomes.

While the CMT3chr3 mutation abolished the interaction of CMT3 with nucleosomes, we found that CMT3chr3 retained similar in vitro catalytic activity as that of wild-type CMT3 (Figure 6D). This suggests that H3K9me2 acts as a recruitment factor for CMT3 in vivo.

**DISCUSSION**

Nucleosome occupancy is an important determinant of global DNA methylation patterns in vivo (Chodavarapu et al., 2010). The association of CMT3 with nucleosomes suggests that CMT3 may methylate nucleosome-bound DNA, which may explain our previous observations from whole genome shotgun bisulfite sequencing that CHG methylation was enhanced on nucleosomal DNA relative to linker DNA and that CHG sites facing on the outside of nucleosomes were methylated at a higher level than those facing on the inside (Chodavarapu et al., 2010). This model also explains the 167 nt periodicity pattern of CHG methylation previously observed in bulk analysis of whole genome methylation data (Cokus et al., 2008).

The observation that CMT3 is predominately expressed in cells undergoing active replication implies that methylation by CMT3 takes place during DNA replication, when H3.1 is incorporated onto newly synthesized chromatin and modified by the histone H3K9 methyltransferase, KYP. One plausible model is that CMT3 is displaced from nucleosomes during replication. Once replication is completed, it is recruited back to the newly assembled nucleosomes that have been methylated by KYP (or other SUVHs) on H3K9. Alternatively, CMT3 may be recruited to chromatin during DNA replication by premodified H3K9me2 marks from parental histones. In turn, CHG methylation would recruit KYP to deposit methylation marks on newly synthesized histones.

We attempted to test whether the H3K9-specific histone methyltransferase KYP (SUHV4) or possibly SUHV5 and SUHV6 (two other H3K9 methyltransferases) would form a complex with CMT3. However, the CMT3 IP-MS analysis did not detect any KYP, SUHV5, or SUHV6 peptides. Reciprocal IP experiments with protein extracts from Nicotiana benthamiana leaves coexpressing MYC-tagged CMT3 and FLAG-tagged KYP also failed to detect interaction between CMT3 and KYP (data not shown). These findings suggest either that KYP is not stably associated with nucleosomes, or that KYP and CMT3 cannot simultaneously remain bound to the same nucleosomes.

Our in vitro activity assays show that CMT3 has some activity on unmethylated DNA. We speculate that this “de novo” activity is in fact part of the “maintenance” loop of KYP and CMT3. Unlike the mammalian maintenance DNA methyltransferase DNMT1, which has much greater activity on hemimethylated DNA than on unmethylated DNA (Bestor and Ingram, 1983), CMT3 only has slightly higher activity on hemimethylated substrates than unmethylated substrates. In addition, our data suggested that CMT3 showed only a small tendency to fully methylate CHG sites in vivo rather than leaving them hemimethylated, which is in contrast to the strong tendency of DNMT1 to fully methylate CG sites in vivo. This suggests that CMT3 is not as good as DNMT1 at restoring hemimethylated sites to fully methylated sites and consequently, methylation at many sites will be completely lost during multiple cycles of replication. The “de novo” activity of CMT3, however, could persistently target methylation to regions marked with H3K9 methylation. CMT3’s inefficiency at maintaining methylation is also reflected by the fact that the overall global levels of CHG methylation (6.7%) are much lower than those of CG methylation (24%) (Cokus et al., 2008). In addition to the preference for hemimethylated DNA as substrate, DNMT1 and MET1 work with the UHRF1/VIM cofactor that specifically recognizes hemimethylated DNA through its SRA domain (Bostick et al., 2007). However, the SRA domain of KYP and SUVH5 bind equally well to hemimethylated and fully methylated DNAs (Johnson et al., 2007; Rajakumara et al., 2011a). Therefore, CMT3’s “maintenance” activity is likely to be driven mainly by the feedback loop between KYP and CMT3, rather than by the inherent preference of CMT3 for hemimethylated DNA.

H3K27me1 is enriched in heterochromatin and was proposed to be involved in CMT3 function based on in vitro data showing that the chromodomain of CMT3 bound to histone H3 peptides only when both H3K9 and H3K27 were methylated (Lindroth et al., 2004). However, these doubly methylated peptides were longer than the singly methylated control peptides, which might be an alternative explanation for this result. More critically, a recent study showed that depletion of H3K27me1 in vivo did not reduce CHG methylation (Jacob et al., 2009). In addition, our peptide array and structural analyses showed that H3K9me2 is necessary and sufficient to bind CMT3. Therefore, H3K27me1 is unlikely to play a role in CMT3 targeting.

The BAH domain functions as a mediator of protein-protein interactions. Yeast Sir3 BAH domain in complex with mononucleosomes revealed that the BAH domain makes contacts through unmodified H4K16 and H3K79 (Armache et al., 2011), which are important for BAH-nucleosome binding (Onish et al., 2007). Recent studies of the mouse ORC1 BAH domain bound to H4K20me2 peptide have established that methylated lysine is recognized through positioning within an aromatic cage on the surface of the BAH domain, and the methyl-lysine recognition mode is similar as the ZMET2 BAH domain recognizing H3K9me2 peptide reported here (Figure S4E) (Kuo et al., 2012). In the current study, we also observed that the methylated lysine of H3K9me2 peptide was positioned within an aromatic cage of the BAH domain of ZMET2. The aromatic cage residues in the BAH domain of ZMET2, that are conserved among the BAH domains of other plant chromomethylases, are also observed within the BAH1 domain of human DNMT1. These results imply that the BAHI domain of DNMT1 may also be a reader of methylated-lysine marks using aromatic cage capture. By contrast, no aromatic cage was identified following alignment of the BAH2 domain of DNMT1.

How is CHG methylation precisely controlled and faithfully maintained? We propose a dual-recognition mechanism, in which the CMT3 BAH and chromo domains simultaneously read the H3K9me2 on the two tails emanating from a single nucleosome to ensure a higher fidelity of CHG DNA methylation (Figure 6F). In the crystal, we are restricted to one or the other
complex most likely due to packing interactions, but there should be no such constraints for simultaneous recognition of H3K9me2 marks by both BAH and chromo domains in solution. Indeed, ITC binding studies yield a stoichiometry of 1.8 for binding of H3(1–15)K9me2 peptide to ZMET2 (Figure S6A). Modeling efforts indicate that it is possible that two H3K9me2 modified tails could extend from the single nucleosome and bind to the chromo and BAH domains simultaneously. The nucleosomal DNA between the two H3 tails could be clamped by the TRD subdomain and the catalytic loop and subsequently methylated (Figures S6B and S6C). The H3 tail has sufficient length so as to likely allow the enzyme enough flexibility to catalyze on a range of nucleosomal DNA. The two-domain binding mode could potentially increase CMT3’s binding affinity and specificity toward H3K9me2 marks and thus help to recruit CMT3 to H3K9me2-containing nucleosomes. It is also conceivable that the dual binding modules could help CMT3 walk from nucleosomes to adjacent nucleosomes, characteristic of a spreading mechanism. Interestingly, the mammalian de novo DNA methyltransferase DNMT3A/3L complex also binds stably to nucleosomes (Jeong et al., 2009), and DNMT3A and DNMT3L each have an unmodified H3K4-recognizing ADD domain, suggesting that the DNMT3A/3L complex may also have the potential for dual recognition of two unmodified H3K4-containing tails of a nucleosome. By contrast, DNMT1 only contains one BAH domain that has a potential methyl-lysine binding cage and is not tightly associated with nucleosomes (Jeong et al., 2009). Thus, a dual histone tail recognition mode may be a common feature of DNA methyltransferase that are stably bound to chromatin.

EXPERIMENTAL PROCEDURES

Immunoprecipitation and Mass Spectrometry
Epitope-tagged CMT3 transgenic plants were generated and used for IP-MS analysis as described previously (Law et al., 2010) and detailed information can be found in the Extended Experimental Procedures.

Western Blot
The FLAG and MYC epitope tags were detected using the anti-FLAG M2 monoclonal antibody (Sigma, A8592) and anti-Myc 9E10 monoclonal antibody (Covance, MMS-150B), respectively. Primary antibodies used for histones included: anti-histone H3 antibody (Abcam, ab10799), anti-histone H3K9me2 (Abcam, ab1220-100), and anti-histone H3K4me2 (Abcam, ab32356).

Immunofluorescence
EdU labeling of BLRP-9xMYC-CMT3 transgenic root cells was based on the Click-IT kit (Invitrogen) and immunofluorescence analysis was described in detail in the Extended Experimental Procedures.

Protein Purification and Histone Peptide Array
Recombinant CMT3 and ZMET2 protein expression and purification were based on (Song et al., 2011) and described in detail in the Extended Experimental Procedures. MOdified histone peptide array slide (Active Motif, 13001) was blocked by incubation in TBS buffer (10 mM Tris-HCl [pH 7.5] and 150 mM NaCl) containing 5% milk at room temperature (RT) for 1 hr and washed three times with TTBS buffer (10 mM Tris-HCl [pH 7.5], 150 mM NaCl, and 0.05% Tween-20). The array slide was then incubated with 30 μg purified proteins in 3 ml binding buffer (50 mM HEPES [pH 7.5], 50 mM NaCl, 5% glycerol, 0.4% BSA, and 2 mM DTT) overnight at 4°C, washed three times with TTBS buffer and incubated with anti-His antibody in TBS buffer for 1 hr at RT. The array was washed three times with TTBS and developed by enhanced chemiluminescence. The images were analyzed according to the instruction of array analysis software.

DNA Methyltransferase Assay
Baculovirus-mediated CMT3 expression and purification were performed as previously described (Patnaik et al., 2004). DNA methyltransferase activity assay was described in detail in the Extended Experimental Procedures.

Hairpin Bisulfite Sequencing
Genomic DNA was digested with BglII restriction enzyme and ligated with oligo (5’/5Phos/GATCTGCGATCGDDDDDDDCATCGCA) by T4 DNA ligase (NEB). JP3200 and JP1615 primers were used to amplify Ta3 hairpin bisulfite treated DNA. Methylation on both strands of the DNA was assayed in 21 clones from wild-type plants. Chi-square value was calculated by using standard chi-square formula (the sum of squared observed minus expected over expected). The expected number of dyads was calculated based on the observed cytosine methylation percentage and four possible combinations of methylated and unmethylated cytosines in a dyad. Calculated chi-square value, together with degree of freedom of two, was used to obtain probability based on the chi-square distribution table.

Chromatin Immunoprecipitation
ChiP was based on IP with the following modifications. Micrococcal nuclease was included in crude extracts to shear chromatin before incubation with beads. The eluted protein-DNA complexes were treated by proteinase K followed by phenol:chloroform purification. The enriched DNA was ethanol precipitated and subjected to library generation following Illumina’s manufacturer instructions (see Extended Experimental Procedures).

ChiP-Seq and RNA-Seq Data Analyses
Illumina base-called reads were mapped with Bowtie (Langmead et al., 2009) allowing up to two mismatches. For ChiP-seq, identical reads were collapsed into one read. Both gene and transposon expressions were measured by calculating reads per kilobase per million mapped reads (RPKM). p values were calculated using Fisher’s exact test and Benjamini corrected for multiple testing. Differentially expressed elements in wild-type and mutants were defined by applying log2(mutant/wild-type) > 2 and p < 0.05 cutoffs.

Crystallization, Structure Determination, and Refinement
Crystals were grown using the hanging drop vapor diffusion method and the diffraction data were collected at the NE-CAT beamline 24ID-E, Advanced Photon Source (APS) at the Argonne National Laboratory, Chicago, and processed with the program HKL2000 (Otwinowski and Minor, 1997). The structure of Se-substituted ZMET2 in the presence of SAH and the structures of ZMET2 in complex with H3K9me2 peptides in the presence of SAH were solved using SAD method and molecular replacement method, respectively, as implemented in the program Phenix (Adams et al., 2010). The model building was carried out using the program Coot (Emsley et al., 2010). The statistics of the diffraction data and the refinement are summarized in Table S5. Additional details are provided in the Extended Experimental Procedures.

ACCESSION NUMBERS
Coordinates and structure factors for Se-ZMET2 in the free state (PDB code 4F5X), and in complex with H3(1–15)K9me2 (PDB code 4FT2) and H3(1–32) K9me2 (PDB code 4FT4) peptides, all in the presence of SAH, have been deposited in the RCSB Protein Data Bank. Sequencing data have been deposited in the GEO database under ID codes GSE39097 and GSE38286.

SUPPLEMENTAL INFORMATION
Supplemental Information includes Extended Experimental Procedures, six figures, and five tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2012.07.034.
ACKNOWLEDGMENTS

We thank Anders Lindroth and Debasis Patnak for initiating CMT3 in vitro activity assays, staff members at APS for support in diffraction data collection and at UCLA BSCRC BioSequencing core for high-throughput sequencing, and Nathan Springer and Shawn Kaeppeler for providing zMET2 gene. We are also grateful to Lianna Johnson for help with histone peptide array experiments, Greg Horwitz for help with baculoviral protein expression, and their critical reading of this manuscript. X.Z. is supported by Ruth L. Kirschstein National Research Service Award (F32GM086463-01). S.F. is supported by Leukemia & Lymphoma Society. This work was supported by the 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 GM63038 to S.E.J. S.E.J. is an Investigator at the Howard Hughes Medical Institute.

Received: June 6, 2012
Revised: July 6, 2012
Accepted: July 31, 2012
Published: September 27, 2012

REFERENCES


