Methyl-CpG-binding domain protein MBD7 is required for active DNA demethylation in Arabidopsis

Chunlei Wang1, Xiaomei Dong2, Dan Jin1, Yusheng Zhao1, Shaojun Xie3, Xiaojie Li1, Xinjian He4, Zhaobo Lang3, Jinsheng Lai2, Jian-Kang Zhu3, 5, Zhizhong Gong1

1State Key Laboratory of Plant Physiology and Biochemistry, College of Biological Sciences, China Agricultural University, Beijing 100193, China.
2State Key Laboratory of Agrobiotechnology, China Agricultural University, Beijing 100193, China National Maize Improvement Center, Department of Plant Genetics and Breeding, China Agricultural University, Beijing 100193, China.
3Department of Horticulture and Landscape Architecture, Purdue University, West Lafayette, IN 47906, USA.
4National Institute of Biological Sciences, Beijing, China
5Shanghai Center for Plant Stress Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China;

Corresponding author:
Zhizhong Gong
State Key Laboratory of Plant Physiology and Biochemistry, College of Biological Sciences, China Agricultural University, Beijing 100193, China.
Email: gongzz@cau.edu.cn; Tel: 86-10-62733733

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Zhizhong Gong (gongzz@cau.edu.cn).
Abstract

Although researchers have established that DNA methylation and active
demethylation are dynamically regulated in plant cells, the molecular mechanism for
regulation of active DNA demethylation is not well understood. By using an
Arabidopsis line expressing the ProRD29A:LUC and Pro35S:NPTII transgenes, we
isolated an mbd7 (Methyl-CpG-binding domain protein 7) mutant. The mbd7
mutation causes an inactivation of the Pro35S:NPTII transgene but does not affect the
expression of the ProRD29A:LUC transgene. The silencing of the Pro35S:NPTII
reporter gene is associated with DNA hypermethylation of the reporter gene. MBD7
physically interacts with ROS5/IDM2, a protein in the small heat shock protein family.
MBD7 prefers to target the genomic loci with high densities of DNA methylation
around chromocenters. The gypsy type LTR retrotransposons mainly distributed
around chromocenters are most affected by mbd7 in all transposons. Our results
suggest that MBD7 is required for active DNA demethylation and antisilencing of the
genomic loci with high densities of DNA methylation in Arabidopsis.

Key words: DNA methylation, active DNA demethylation, Arabidopsis,
Methyl-CpG-binding domain protein, transposons

Running title: Active DNA demethylation by MBD7

One sentence summary: Active DNA demethylation is dynamically regulated in cells
and a methyl-CpG-binding domain protein MBD7 is required for this process.
Introduction

DNA methylation is an important epigenetic marker for genome stability and regulation of gene expression in both plants and animals (Law and Jacobsen, 2010; He et al., 2011). In plants, the molecular mechanisms for DNA methylation have been well characterized by the use of powerful genetic screening systems (Bartee et al., 2001; Lindroth et al., 2001; Matzke et al., 2004; He et al., 2009). A transgene or an endogenous gene may be silenced because of DNA hypermethylation in the promoter region. Screenings for mutants with release of the silenced marker genes have identified many components that are involved in RNA-directed DNA methylation (RdDM) and in maintaining DNA methylation (Matzke and Birchler, 2005; Law and Jacobsen, 2009; He et al., 2011; Bender, 2012). DNA methylation is catalyzed by DNA methyltransferases including MET1 and CMT3, which maintain symmetric CG and CHG methylation, respectively, during DNA replication, and DRM2 and CMT2, which are required for establishing CHG and asymmetric CHH methylation during each cell cycle. DRM2 also catalyzes CG methylation (Law and Jacobsen, 2010; Haag and Pikaard, 2011; He et al., 2011; Zemach et al., 2013; Stroud et al., 2014).

24-nt small RNAs produced through the RdDM pathway target genomic regions to guide the establishment of DNA methylation by DRM2 (Cao et al., 2003).

DNA methylation can be actively removed by a subfamily of bifunctional DNA glycosylases/lyases including Repressor of Silencing (ROS)1 (Gong et al., 2002) and its paralogs DEMETER and DEMETER-LIKE2/3 (Gehring et al., 2006; Ortega-Galisteo et al., 2008). DNA methylation can also be passively lost during DNA replication when DNA methylation cannot be maintained (Zhu, 2009).

ProRD29A:LUC in the ProRD29A:LUC /Pro35S:NPTII transgenic Arabidopsis line has been used as a marker to identify ros1 and ros3 mutants in which both ProRD29A:LUC and Pro35S:NPTII are silenced (Gong et al., 2002; Zheng et al., 2008). ROS3 is an RNA-binding protein that facilitates the function of ROS1 in active DNA demethylation at certain genomic loci. Using Pro35S:NPTII as a selection marker for kanamycin-sensitive mutants and the 35S-SUC2 transgene or a Chop PCR
marker for assaying DNA methylation at the 3’ region of At1g26400 from T-DNA insertion mutants, researchers recently identified two genes involved in active DNA demethylation: ROS4/Increased DNA methylation (IDM)1 and ROS5/IDM2 (Li et al., 2012; Qian et al., 2012; Qian et al., 2014; Zhao et al., 2014). ROS4/IDM1 is a PHD-finger domain-containing histone acetyltransferase that catalyzes H3K18 and H3K23 acetylation (Li et al., 2012; Qian et al., 2012). ROS5/IDM2 is a member of the small heat shock protein family that physically interacts with IDM1/ROS4 for regulation of active DNA demethylation. Genetic analysis indicates that ROS1, ROS4/IDM1, and ROS5/IDM2 are in the same genetic pathway and ROS4/IDM1 and ROS5/IDM2 may form a protein complex for regulation of active DNA demethylation (Qian et al., 2014; Zhao et al., 2014).

During the genetic screening for kanamycin-sensitive mutants using the ProRD29A:LUC/Pro35S:NPTII transgenic line in the current study, we identified another mutant, mbd7, where the Pro35S:NPTII transgene is specifically silenced. MBD7 is a methyl-CpG-binding domain (MBD) protein containing three MBD motifs that bind in vitro to methylated CpG sites. MBD7 localizes to all highly CpG-methylated chromocenters in vivo (Zemach and Grafi, 2003; Zemach et al., 2008). Recruitment of MBD7 to chromocenters is disrupted in ddm1 and met1, two mutants with great reductions in DNA methylation, suggesting that DNA methylation is required for proper MBD7 localization (Zemach et al., 2005). In this study, we found that MBD7 physically interacts with ROS5/IDM2 and is required for active DNA demethylation of certain genomic loci, especially for the gypsy type LTR retrotransposons with high densities of DNA methylation around chromocenters in Arabidopsis.

Results

MBD7 is required for active DNA demethylation at the 3’-region of Pro35S:NPTII

The C24 transgenic Arabidopsis line carries ProRD29A:LUC and Pro35S:NPTII
transgenes, which are actively expressed (Gong et al., 2002). We used an ethyl
methanesulfonate–mutagenized population of this line (as the wild type in this study)
to screen for kanamycin-sensitive mutants (Li et al., 2012; Zhao et al., 2014). The
kanamycin-sensitive mutant, mbd7, was isolated during the screening (Figure 1A,
left). Like NTPII expression in ros1-1, NTPII expression in mbd7 was only weakly
detected but was clearly detected in the C24 wild type (Figure 1B). mbd7 was
backcrossed with the wild type at least three times to remove background mutations
before further analysis. F1 seedlings of mbd7 crossed with the wild type exhibited the
wild-type phenotype for kanamycin resistance, and F2 seedlings were about 1:3
kanamycin-sensitive (52 seedlings) to -resistant (178 seedlings), indicating that mbd7
is caused by a recessive, single nuclear gene mutation. Like ros4 and ros5-1 mutants
(Li et al., 2012; Zhao et al., 2014), the mbd7 mutant did not affect the expression of
ProRD29A:LUC (Figure 1C, right, and 1D), but ros1-1 silenced both ProRD29A:LUC
and Pro35S:NPTII (Figure 1A-D).

We used map-based cloning to isolate the MBD7 gene. The mbd7 mutant in the
C24 accession was crossed with the wild-type Col-0 accession. The 582 F2 plants that
contained NPTII gene (checked by PCR), but were kanamycin sensitive were isolated
and used for mapping. MBD7 was narrowed to a region between BAC clone MNC17
and F15L12 on chromosome 5 (Figure 1E). Sequencing the candidate genes on the
BAC clone MTH12 revealed a G-A point mutation in AT5G59800, which changes an
Arg42 to a putative stop codon. mbd7 should be a null mutant. To confirm that the
point mutation in AT5G59800 leads to Pro35S:NPTII transgene silencing in mbd7, we
transformed a wild-type genomic fragment containing the full-length AT5G59800
including the 2643-bp promoter and the 278-bp 3’- region into mbd7 mutant plants.
As shown in Figure 1G and 1F, three randomly selected, independent transgenic lines
(#9, #14, and #15) complemented the kanamycin-sensitive phenotype with expression
of NPTII protein. Analysis of GUS staining in the transgenic plants expressing
ProMBD7:GUS indicated that MBD7 is highly expressed at the early seedling stage;
the timing of this expression overlaps with that of ROS4/IDM1 and ROS5/IDM2
In previously isolated ros mutants, ros1-1 increased the DNA methylation in both the RD29A promoter and the 3’-NOS terminator region, and ros4 and ros5-1 increased DNA methylation only in the 3’-NOS terminator (Zhao et al., 2014). Whole-genome bisulfite sequencing indicated that, like ros4 and ros5-1, mbd7 increased DNA methylation only in the 3’-NOS terminator region and did not change DNA methylation in the RD29A promoter (Figure 2A); this is consistent with the lower expression level of NPTII but the similar expression level of LUC relative to the wild type (Figure 1B and 1D). Bisulfite sequencing confirmed the DNA hypermethylation in the 3’-NOS terminator region in mbd7 (Figure 2B). The DNA methylation level in 3’-NOS was not greater in the ros5-1 mbd7 double mutant than in the ros5-1 or mbd7 single mutants. We further checked four loci (AT2G37840, AT3G60961, AT1G42980, and AT4G11430) in which DNA methylation was found to be increased by ros5-1 mutation in a previous study (Zhao et al., 2014). As was the case for 3’-NOS, the methylation level at the four loci was not greater in the ros5-1 mbd7 double mutant than in the ros5-1 or mbd7 single mutant. These results suggest that ROS5/IDM2 and MBD7 work in the same pathway to mediate DNA methylation. Treatment with 5’-aza-2’-deoxycytidine (5’-Aza) (a cytosine methylation inhibitor) consistently rescued the kanamycin-sensitive phenotype of mbd7 seedlings (Figure S2A) and induced a high level of NPTII expression (Figure S2B). The introduction of a DNA replication mutation, rfc1-1, into mbd7 released the silencing of Pro35S:NPTII (Figure S2C), which is similar to previous reports for ros1-1, ros4, and ros5-1 mutants (Liu et al., 2010; Li et al., 2012; Zhao et al., 2014). Previous study indicates that mutations in RdDM pathway greatly reduce the expression of ROS1. However, ROS1 expression was not affected by mbd7 (Figure S2D).

MBD7 physically interacts with ROS5/IDM2

ROS5/IDM2 is a small, heat shock protein-like protein with a conserved α-crystallin domain that forms a ~16-mer complex (Zhao et al., 2014). ROS5 physically interacts with ROS4/IDM1 and is required for histone H3K18 acetylation by ROS4/IDM1.
(Qian et al., 2014; Zhao et al., 2014). Because ROS1, ROS4, ROS5, and MBD7 are all required for active DNA demethylation in the 3’-NOS terminator region of 35S-NPTII, and because they are all identified by the same genetic screening, we wanted to know whether they operate in the same protein complex. We first used a firefly luciferase complementation imaging assay to determine whether MBD7 interacts with ROS1, ROS4, or ROS5 in tobacco leaves (Chen et al., 2008). We found that MBD7 interacts with ROS5 (Figure 3A) but not with ROS4 or ROS1 (Supplemental Figure S3A). The interaction of FLS2 with PUB13-ARM domain was used as a positive control (Lu et al., 2011). The yeast two-hybrid assay indicated that ROS5 interacts with the MBD7 C-terminal domain but not with the MBD7 N-terminal region, which contains three MBD domains (Figure 3B). MBD7 did not interact with ROS1 or IDM1/ROS4 in the yeast two-hybrid assay (Figure S3B). An in vivo coimmunoprecipitation assay using proteins from Arabidopsis protoplasts transiently expressing different plasmids indicated that MBD7-Flag was co-immunoprecipitated with Myc-ROS5 (Figure 3C). An in vitro protein pull-down assay, which used purified proteins expressed in Escherichia coli, also showed that GST-MBD7 pulled down His-ROS5 (Figure 3D). We also analyzed the subnuclear localization of mCherry-ROS5 and MBD7-GFP in transiently co-expressing tobacco epidermal cells. As shown by the yellow signals in Figure 3E, mCherry-ROS5 was co-localized with MBD7-GFP. Because ROS4/IDM1 interacts with ROS5/IDM2, and because ROS1 is likely in the complex of IDM1/ROS5 (Zhao et al., 2014), these results suggest that ROS1, ROS4/IDM1, ROS5/IDM2, and MBD7 form a protein complex in the nucleus.

**MBD7 prevents DNA methylation mainly in the genomic regions around chromocenters**

To determine whether MBD7 plays a role in active DNA demethylation at the whole-genome level, we compared the genome-wide DNA methylation profiles of mbd7 and the C24 wild type by whole-genome bisulfite sequencing (Zhao et al., 2014). Previous studies reported no significant difference in the whole-genome DNA
methylation level between *ros4* or *ros5-1* and the wild type (Zhao et al., 2014).

Similarly, the whole-genome DNA methylation level in *mbd7* didn’t show too much
difference when compared with that in the wild type (Figure S4A). However, by
analyzing the windows with significant methylation difference (P-value<0.01)
between C24 wild type and *mbd7*, we found that the methylation level of *mbd7* was
higher than that of C24 in these analyzed loci (Figure S4B), which indicates that
MBD7 prevents hypermethylation at these genomic loci. The total methylation levels
in *mbd7* were 20.6% in CG, 6.0% in CHG and 2.2% in CHH, which were a little
higher than 20.2% (CG), 5.4% (CHG) and 2.0% CHH) in the C24 wild type (Figure
S4C). In total, we identified 2664 differentially methylated regions (DMRs) with
increase DNA methylation, including 497 hyper-DMRs in CG methylation, 192
hyper-DMRs in CHG methylation and 2041 hyper-DMRs in CHH methylation in
*mbd7* (Table S1). The number of hyper-DMRs that are localized in gene bodies, TEs,
upstream and downstream of gene bodies and the intergenic regions are 701, 1348,
241, 114 and 260 in *mbd7*, respectively (Table S2). Among 2041 CHH hyper-DMRs
in *mbd7*, 1235 are in TEs, 272 in gene bodies, 231 in upstream of gene bodies and 93
in downstream of gene bodies, and 210 in intergenic regions. Hyper-DMR
distribution analysis on the whole genome indicated that hyper-DMRs in *mbd7* were
intensively localized around the chromocenters in 5 individual chromosomes (Figure
4B). As more than 50% hyper-DMRs are TEs, we analyzed whether these TEs have
any distinguishing features. The length of TEs that overlapped with hyper-DMRs in
*mbd7* was a little shorter than that of total TEs (Figure 4C), suggesting that MBD7 has
a bias to target short TEs. Comparing the ratio of individual TEs to total TEs in the
whole genome with the ratio of MBD7 affected individual TEs to total MBD7
affected TEs, four categories of TEs including DNA/En-Spm, LINE/L1, LTR/Copia
and LTR/Gypsy were more affected than other TEs by MBD7 (Figure 4D). Among
these four TEs, the gypsy type LTR retrotransposons (LTR/Gypsy) are most affected
by MBD7 (Figure 4D). The gypsy type LTR retrotransposons targeted by MBD7
were intensively distributed around the 5 chromocenters (Figure 4B, blue line), which
is consistent with the total Gypsy TE distribution in *Arabidopsis* (Zemach et al., 2013).
These results suggest that MBD7 prefers to target genomic loci around chromocenters, and the gypsy type LTR retrotransposons are the most affected TEs.

In order for comparing with other mutants, we also included the DNA methylation profiles of ros1-1, ros4, and ros5-1 mutants (Zhao et al., 2014). The heat-map analysis of these DMRs indicated a high co-relativity between mbd7 and ros1-1, ros4, or ros5-1 (Figure 5A). The DMRs of mbd7 overlapped with those of each mutant with different ratio (Figure 5B). Among these DMRs in mbd7, about 34% in CG methylation, 29.1% in CHG methylation and 2.6% in CHH methylation are found to be overlapped in all four mutants (mbd7, ros1-1, ros4, and ros5-1) (Figure 5B). The low overlapped ratio in CHH is likely due to the low methylation level which reduces the number after quality control. We also found that 50.6% of the DMRs in mbd7 overlapped with TEs, which is higher than in ros1-1 (30.6%), ros4 (23.7%), or ros5-1 (16.3%). The CG, CHG, and CHH DMRs in ros1-1 were evenly distributed in different genomic regions, while CG DMRs in mbd7, ros4, and ros5-1 were mainly localized in genic regions, and CHH DMRs were mainly in TEs (Figure 5C). CHG DMRs in TEs were more abundant in mbd7 than other mutants. These results suggest that although these proteins may work together to target some common genomic loci, each of them has its own preference to some specific regions (Li et al., 2012; Qian et al., 2012; Qian et al., 2014; Zhao et al., 2014).

**MBD7 preferably binds to the high-density DNA methylation regions**

To determine which regions MBD7 binds to, we performed a chromatin immunoprecipitation (ChIP) assay using transgenic plants expressing Pro35S:MBD7-GFP. We selected several DNA regions that have higher DNA methylation in mbd7 than in the wild type, and we checked them by PCR after ChIP. In the T-DNA region of ProRD29A:LUC/Pro35S:NPTII transgenic plants, the affinity was highest near the 35S promoter region, followed by the 3'-NOS terminator and the linker region between 3'-NOS and RD29A promoter (Figure 6A, also see Figure 2A for examined regions). The affinity was lowest in the LUC coding region. These results indicate that MBD7 binds to most of the Pro35S:NPTII:NOS region with high...
affinity. We selected four hypermethylated loci \((AT4G11430, AT3G60961, AT2G37840\) and \(AT2G22350\)) that are regulated by MBD7, ROS1, ROS4, and ROS5, and we confirmed their DNA methylation levels by bisulfite sequencing in each mutant (Figure 6B). \(AT3G60961\) and \(AT2G22350\) are transposon elements that are targeted by RdDM pathway, and \(AT4G11430\) is near transposons and not targeted by RdDM pathway. As shown in Figure 6A, MBD7 bound to these four regions. MBD7 did not bind to \(AT1G42980\) (a gene without nearby transposons, targeted by RdDM), whose methylation was increased by \(ros1-1, ros4,\) and \(ros5-1\) mutations but not by the \(mbd7\) mutation. The \(AT5G30942\) locus is localized in the middle of a TE-rich region that has a high density of DNA methylation and that is in the chromocenter of chromosome V, and that is targeted by DDM1 pathway; MBD7 bound to this locus (Figure 6A), although DNA methylation in \(AT5G30942\) was not changed by \(mbd7, ros1-1, ros4,\) or \(ros5-1\) mutations (Figure 6B). Perhaps MBD7 targets \(AT5G30942\) but the DNA methylation level in this locus is already very high and thus cannot be enhanced by \(mbd7\). In contrast to \(AT5G30942, AT1G10950\) is localized in a gene region without TEs. \(AT1G10950\) has hypermethylated DNA, but the DNA methylation density is lower than in \(AT5G30942\), and its DNA methylation was not changed by \(mbd7, ros1-1, ros4,\) or \(ros5-1\) mutations (Figure 6B). MBD7 did not bind to \(AT1G10950\) (Figure 6A). As negative controls, MBD7 did not bind to \(AT1G10260\) or \(ACTIN2\) loci that lacked DNA methylation. These results suggest that MBD7 binds to genomic regions with high densities of DNA methylation and might target loci in addition to those detected by the whole-genome bisulfite sequencing.

**Discussion**

Active DNA demethylation is carried out by DNA demethylases in the ROS1 protein family through a base excision-repair pathway (Zhu, 2009; Gong and Zhu, 2011). However, how these DNA demethylases are recruited to the specific target regions for removal of DNA methylation is not known. Previous studies suggest that ROS4/IDM1 interacts with ROS5/IDM2, which regulates H3K18 acetylation, and creates a favorable chromatin environment for recruiting of ROS1 (Li et al., 2012;
Qian et al., 2012; Qian et al., 2014; Zhao et al., 2014). By using the same genetic screen that was used to isolate the *ros4* and *ros5* mutants, we identified MBD7, a methyl-CpG-binding domain protein in the current study. MBD7 physically interacts with ROS5/IDM2, and previous studies indicated that ROS5/IDM2 also interacts with IDM1 (Qian et al., 2014; Zhao et al., 2014) and that ROS1 is in the same complex as ROS5/IDM2 (Zhao et al., 2014). These related proteins evidently target some common genomic sites for active DNA demethylation because any mutation in these four genes increases the DNA methylation in some common sites, and because the *mbd7* *ros5-1* double mutant does not have an additive effect on DNA methylation level relative to *mbd7* or *ros5-1* single mutants. Because ROS1, IDM1/ROS4, IDM2/ROS5, and MBD7 all belong to a small family of proteins, functional redundancy of the paralogous proteins may work at different loci for regulation of active DNA demethylation. For an example, for the *AT1G42980* locus (Figure 6), the DNA methylation in *mbd7* is not changed, and MBD7 does not bind to this locus, but DNA methylation is increased in *ros5-1, ros4* and *ros1-1* mutants. It is speculated that the MBD7 paralogs may function together with IDM1/ROS4, ROS5/IDM2 and ROS1 in this locus. The molecular mechanism for the different genomic regions preferably targeted by different proteins needs further exploring.

Our study indicates that MBD7 prefers to target the genomic regions with high density of DNA methylation around chromocenters, which is consistent with the previous studies to show that MBD7 is intensively localized in the chromocenters (Zemach et al., 2008). The mutations in *met1* and *ddm1* abolish the recruitment of MBD7 to chromocenters (Zemach et al., 2005). MBD7 binds to CG methylation sites (Zemach and Grafi, 2003). These results suggest that the dense DNA methylation is required for MBD7 functions. Our study indicates that the hyper-DMRs in *mbd7* are highly enriched with CHH methylation and MBD7 prefers to target the short TEs, especially the gypsy type LTR retrotransposons. Given that CHH methylation is usually catalyzed by DRM2 in RdDM pathway (Cao and Jacobsen, 2002; Cao et al., 2003), and by CMT2 in DDM1-dependent pathway (Zemach et al., 2013), these
results suggest that MBD7 is involved in active DNA demethylation at genomic sites targeted by both RdDM pathway and DDM1-dependent pathway, which is supported by our DNA methylation analyses on individual loci (Figure 6).

A previous study indicated that the MBD7 C-terminal domain has very strong chromatin-binding ability and that one of its MBD domains binds to chromosomal methylated DNA sites but does not affect MBD7 subnuclear localization (Zemach et al., 2009). In this study, we found that the MBD7 C-terminal domain, but not its MBD domains, interacts with ROS5/IDM2. ROS5/IDM2 is an α-crystallin domain-containing protein that also interacts with ROS4/IDM1, a histone acetyltransferase (Qian et al., 2014; Zhao et al., 2014). We speculate that MBD7 binds to methylated CG through one of its MBD domains in the N-terminus, and then the C-terminal domain recruits ROS5/IDM2, which further recruits IDM1 and other proteins (Figure 7) (Zemach et al., 2005; Scebba et al., 2007). IDM1 will acetylate histone H3K18 and 23 to create a chromatin environment that facilitates the recruitment of ROS1 protein for removal of DNA methyl groups. Because MBD domain proteins are conserved in mammals, the results reported here may also apply to mammalian systems (Fournier et al., 2012).

**Materials and Methods**

**Plant Materials and Mutant Screening.**

All mutants used in this study are in Arabidopsis thaliana C24 background, which carries transgenes of ProRD29A:LUC and Pro35S:NPTII (Which we refer as the wild type in this study). The mutants ros1-1, ros4/idm1-3, ros5-1, and nrpd2, rfc1-1 were previously isolated in our laboratory. The mutant mbd7 was identified from an EMS-mutagenized population of the wild type as described previously (Li et al., 2012). In short, wild-type seedlings are resistant to 50 mg/L kanamycin. Individual mutants that were sensitive to 50 mg/L kanamycin were selected, and their kanamycin phenotype was confirmed in the next generation. The putative mutants were crossed with Col-0 (the wild type without T-DNA insertion), and the F2 plants that were sensitive to a kanamycin-amended medium and that carried the NPTII gene (detected
by PCR) were used for mapping. We obtained 582 F2 plants that contain NPTII, but show kanamycin sensitivity for the map-based cloning.

For the complementation assay, the MBD7 genomic sequence (from -2643 to +343 bp, which contains the 2643-bp promoter, the coding region, and the 278-bp 3’ region) was cloned into pCAMBIA1300. Then the construct in Agrobacterium tumefaciens GV3101 was transformed into the mbd7 mutant, and transgenic plants were selected and analyzed for their kanamycin phenotypes. All primers used in this study are listed in Table S3.

**Real-Time RT-PCR Assay**

Real-time RT-PCR was used to assess the transcription levels of different genes. Total RNAs were extracted from 7-d-old seedlings by use of Trizol reagent (Invitrogen). The reverse-transcribed cDNAs were used for real-time RT-PCR in SYBR Green Master Mix (Takara) on a Step One Plus machine (Applied Biosystems). Primers specific for each gene are listed in Table S3. ACTIN was used as an internal control.

**Histochemical GUS Staining**

The MBD7 promoter region (from -2643 to -1 bp) and the GUS coding region were cloned into the pCAMBIA1391 vector. Then the recombinant plasmid was transformed into C24 wild-type plants with the help of Agrobacterium tumefaciens GV3101. Histochemical GUS staining was performed on the T2 homozygous transgenic plants as previously described (Xia et al., 2006).

**Bisulfite Sequencing**

Genomic DNA was extracted from 7-d-old seedlings, and about 500 ng was treated with the EZ Methylation-Gold Kit (Zymo Research) following the protocol supplied by the manufacturer. About 80 ng of treated DNA was used in the PCR with the specific primers listed in Table S3. The PCR products were cloned into the pMD18-T vector (Takara), and at least 15 independent clones of each sample were sequenced for each region.

**Co-Immunoprecipitation Assay**

The full-length cDNA of MBD7 fused with the Flag coding sequence was cloned into
the Multiple Cloning Site (MCS) of a modified vector pCAMBIA1300. A vector expressing Myc-ROS5 was described previously (Zhao et al., 2014). The two purified plasmids were co-transformed into *Arabidopsis* protoplasts. After the protoplasts were cultured overnight in a light chamber, the proteins were extracted from protoplasts with the immunoprecipitation buffer (50 mM Tris-HCl, pH 7.6, 5 mM MgCl₂, 150 mM NaCl, 10% glycerol, 0.5% Nonidet P-40, 0.5 mM DTT, 1 mM phenylmethylsulfonyl fluoride [PMSF], and protease inhibitor cocktail [one tablet per 50 mL]; Roche). The protein solution was centrifuged, and the supernatant was added to 20 μL of Anti-C-Myc-Agarose Affinity Gel (Sigma-Aldrich). After incubation on a rotary shaker at 4 °C for 2 h, the agarose beads were washed four times with the immunoprecipitation buffer in which the concentration of Nonidet P-40 was adjusted to 0.1%. The immunoprecipitated products were then detected by western blot.

**Luciferase Complementation Imaging Assay**

The full-length cDNA of *MBD7* was fused with NLUC in the pCAMBIA vector. The vectors expressing CLUC-ROS1, CLUC-ROS4/IDM1, and CLUC-ROS5/IDM2 were previously constructed in our laboratory (Zhao et al., 2014). The FLS2-NLUC and CLUC-PUB13ARM constructs were prepared as previously described (Lu et al., 2011). After the plasmid constructs were transformed into *Agrobacterium tumefaciens* GV3101 and then into *Nicotiana benthamiana*, fluorescence image was observed with a CCD camera (1300B; Roper).

**Protein Purification and Pull-down Assay**

To prepare the purified GST-MBD7 protein, the GST in the pGEX4T-1 vector was fused with the full-length cDNA of MBD7; the construct was then transformed into *E. coli* BL21 cells. Constructs of HiS-ROS5 and His-PYR1 in the pET30a vector were previously described (Zhao et al., 2014). A 2-μg quantity of GST-MBD7 and the same quantity of His-ROS5 or His-PYR1 protein were added in a binding buffer (50 mM Tris-HCl, pH 7.5, 250 mM NaCl, 0.1% Nonidet P-40, 1 mM PMSF). After incubating at 4 °C overnight, each binding system was treated with 1 mL of binding buffer and 25 μL of nickel-nitrilotriacetic acid agarose beads, and was then incubated at 4 °C...
for another 2 h. The nickel-nitrilotriacetic acid agarose beads were then washed four times with the binding buffer and finally one time with the Nonidet P-40-free binding buffer. The proteins were eluted twice with 160 μL of elution buffer (20 mM Tris-HCl, pH 7.5, 300 mM NaCl, 5 mM MgCl₂, 250 mM imidazole). The pull-down products were detected by western blot using GST or His antibodies.

Yeast Two-Hybrid Assay

We used the pGADT7 and pGBKT7 vectors for the yeast two-hybrid assay. BD-MBD7 was constructed by cloning full-length cDNA of MBD7 into the pGBKT7 vector. The plasmids of AD-ROS5, AD-ROS4, and AD-ROS1 were previously generated in our laboratory (Zhao et al., 2014). To determine which domain of MBD7 is required for interaction with ROS5, we generated two truncated MBD7 constructs: BD-MBD (1 to 231) and BD-CTD (232-306). About 400 ng of plasmid DNAs of each construct were co-transformed into the yeast strain AH109, which was then cultured on two different synthetic dropout media that either lacked Trp and Leu or that lacked Trp, Leu, His, and ADE. After 3 d at 28 ºC, the growth of the strains on the medium lacking Trp and Leu indicates the transformation efficiency and the positive strains survived on the medium lacking Trp, Leu, His and ADE indicate that two proteins interact each other.

Whole-Genome Bisulfite Sequencing and Data Analysis

Genomic DNA was extracted from 7-d-old seedlings using the DNeasy plant Mini Kit (QIAGEN). We performed MethylC-seq with HiSeq 2000 (Illumina) after bisulfite treatment and Illumina library construction.

The reads obtained from sequencing were cleaned with SolexaQA software (Cox et al., 2010) and were mapped to the Arabidopsis reference sequence (TAIR 10) by use of Bismark (Krueger and Andrews, 2011). DMRs were determined as described previously (Zhao et al., 2014). Bins applied in this study were 100 bp from the reference genome. The DMR identification used the standard that the absolute methylation levels of mutant in CG, CHG and CHH contexts were at least 0.4, 0.2, and 0.1, respectively.
The genome distribution of DMRs was annotated according to GFF files (including gene and TE) downloaded from TAIR10. The Arabidopsis genome was divided into five parts: TE, gene, upstream, downstream, and intergenic regions. The regions upstream and downstream of a gene were 2 kb away from TSS and TSE.

**Chromatin Immunoprecipitation Assay**

Ten-day-old seedlings of wild-type C24 and its transgenic lines generated from the transformation of Pro35S:MBD7-GFP were used for ChIP assays as described previously (Saleh et al., 2008). Anti-GFP antibodies (AB290) were used. The ChIP products were dissolved in 100 μL of water, and 1.5 μL of the solution was used in each qPCR reaction with the specific primers listed in Table S3.

**Colocalization of MBD7 and ROS5**

The full-length cDNA of MBD7 was fused with GFP under control of the 35S promoter, and the construct was cloned into a modified pCAMBIA1300 vector. A vector expressing mCherry-ROS5 was previously described (Zhao et al., 2014). The procedures of transient expression in tobacco and image acquisition were also previously described (Liu et al., 2010).

**Accession Numbers**

The gene accession numbers used in this study are AT5G59800 (MBD7), AT2G36490 (ROS1), AT3G14980 (ROS4/IDM1), AT1G54840 (ROS5), AT5G22010 (RFC1), AT5G46330 (FLS2), AT3G46510 (PUB13), AT4G17870 (PYR1) and AT3G18780 (ACTIN2). We used whole-genome bisulfite sequencing to analyze the methylome of mbd7 mutant plants. The original data set was deposited in the Gene Expression Omnibus database with the accession number SRP051989.

**Supplemental Data**

The following materials are available in the online version of this article.

- **Supplemental Table S1.** Hyper-DMRs in mbd7.
- **Supplemental Table S2.** The genomic locations of hyper-DMRs in mbd7
- **Supplemental Table S3.** Primers used in this study.
Supplemental Figure S1. *GUS* expression driven by *MBD7* promoter in seedlings growing on MS medium for 2.5, 3, 6, and 10 days.

Supplemental Figure S2. A DNA methylation inhibitor and a DNA replication mutation in *rfc1-1* release *Pro35S:NPTII* gene silencing in *mbd7*.

Supplemental Figure S3. MBD7 does not interact with ROS4/IDM1 or ROS1.

Supplemental Figure S4. The whole genome methylation difference between *mbd7* and the wild type (C24).

Acknowledgments

This work was supported by the Natural Science Foundation of China (Grants 31330041 and 31421062).
References


2 ubiquitination of pattern recognition receptor FLS2 attenuates plant innate immunity. Science
3 332: 1439-1442
8 proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks.
9 Plant Mol Biol 67: 671-681
11 Regulation of Active DNA Demethylation by an alpha-Crystallin Domain Protein in Arabidopsis.
12 Mol Cell 55: 361-371
15 demethylation in Arabidopsis. Science 336: 1445-1448
17 protocol for studying histone modifications in Arabidopsis plants. Nat Protoc 3: 3018-3025
19 Arabidopsis MBD7 protein partner with arginine methyltransferase activity. Plant J 52:
20 210-222
22 patterns shape the epigenetic landscape in Arabidopsis. Nat Struct Mol Biol 21: 64-72
24 a putative replication protein A2, functions in epigenetic gene silencing and in regulation of
29 proteins. Plant J 34: 565-572
31 (2013) The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to
32 access H1-containing heterochromatin. Cell 153: 193-205
34 DDM1 binds Arabidopsis methyl-CpG binding domain proteins and affects their subnuclear
35 localization. Plant Cell 17: 1549-1558
37 Arabidopsis AtMBD7 protein confers strong chromatin binding activity. Exp Cell Res 315:
38 3554-3562
40 Member of the Small Heat Shock Protein Family and Is Required for DNA Demethylation in
41 Arabidopsis. Plant Cell 26: 2660-2675
43 an RNA-binding protein required for DNA demethylation in Arabidopsis. Nature 455:
Figure legends

Figure 1. Identification of MBD7 by map-based cloning

A. mbd7 mutants silence Pro35S:NPTII but do not affect ProRD29A:LUC.

Seedlings were grown on MS medium or MS medium supplemented with 50 mg/L kanamycin. C24 accession was used as the wild type. ros1-1, in which both Pro35S:NPTII and ProRD29A:LUC are silenced, was included for comparison.

B. NPTII expression in C24, ros1-1, and mbd7 as determined by real-time PCR.

Three independent experiments were done with similar results, each with three technical repeats. The results of one representative experiment are shown. Values are means ± SE (n=3).

C. mbd7 does not affect the expression of ProRD29A:LUC. Seedlings of C24, mbd7, and ros1-1 were treated with 300 mM NaCl for 3 h before luminescence images were captured.

D. Real-time PCR analyses of LUC transcripts from samples in (C). Three independent experiments were done with similar results, each with three technical repeats. The results of one representative experiment are shown. Values are means ± SE (n=3).

E. Identification of MBD7 by map-based cloning. The location of MBD7 was narrowed to BAC clone MNC17 and F15L12 by use of SSLP markers. A G-to-A mutation, which changes Arg42 to a stop codon, was found in AT5G59800.

F. The kanamycin-sensitive phenotype of mbd7 was complemented by the MBD7 gene as shown in three independent transgenic lines.

G. NPTII protein levels in C24, mbd7, and three transgenic lines as indicated by western blot with NPTII antibodies. ACTIN was used as a loading control.

Figure 2. MBD7 prevents DNA hypermethylation in the 3’-NOS region in
ProRD29A:LUC/Pro35S:NTPII transgenic plants.

A. Diagram of the whole T-DNA region and corresponding DNA methylation patterns in mbd7, ros5-1, ros4, ros1-1, and C24 (the wild type) as determined by whole-genome bisulfite sequencing. The red bars with numbers 1, 2, 3, and 4 indicate the fragments used for the ChIP assay in Figure 5A.

B. Confirmation of DNA methylation in the NOS region in mbd7 and ros1-1 by bisulfite sequencing.

C. DNA methylation levels in ros5-1 and mbd7 mutants are not additive in the 3’-NOS region or in AT2G37840, AT3G60961, AT1G42980, or AT4G11430 in the mbd7 ros5-1 double mutant.

Figure 3. MBD7 interacts with ROS5/IDM2

A. MBD7 interacts with ROS5/IDM2 as indicated by a firefly luciferase complementation imaging assay in tobacco leaves. The interaction of FLS2 with the PUB13 ARM domain was used as a positive control.

B. MBD7 interacts with ROS5/IDM2 in a yeast two-hybrid assay. The full-length, C-terminal domain (CTD) and MDB domain of MBD7 were analyzed by yeast two-hybrid assay for interaction with ROS5.

C. Interaction of MBD7 with ROS5 as indicated by coimmunoprecipitation in a protoplast transient assay. Three independent experiments were done with similar results.

D. Interaction of MBD7 with ROS5 as indicated by protein pull-down assay using proteins expressed from E. coli.

E. Colocalization of MBD7-GFP with mCherry-ROS5 as indicated by a transient assay in tobacco epidermal cells. MBD7 was fused to the N-terminus of GFP, and mCherry was fused to the N-terminus of ROS5. A. tumefaciens carrying MBD7-GFP and mCherry-ROS5 was co-injected into tobacco epidermis cells. Bars = 1 μm.
**Figure 4. Hyper-DMRs in mbd7 are enriched around chromocenters.**

A. The distribution of hyper-DMRs (differentially methylated regions) in mbd7 on 5 chromosomes of Arabidopsis. The red line in diagrams represented the distribution of hype-DMRs in mbd7, and the blue line represented the distribution of Gypsy transposons that overlap with hyper-DMRs in mbd7. The red points on the bar under the diagrams represent the chromocenters of different chromosomes.

B. Comparison of the average length of TEs overlapped with hyper-DMRs in mbd7 with that of total TEs in the whole genome.

C. Comparison between the percentage of individual TEs to total TEs in the whole genome and the percentage of individual TEs to total TEs overlapping with hyper-DMRs in mbd7. Numbers above the bars are the percentage. The classified information of TEs was downloaded from TAIR 10 (ftp://ftp.arabidopsis.org/home/tair/Genes/TAIR10_genome_release/TAIR10_transposable_elements/TAIR10_Transposable_Elements.txt).

**Figure 5. Comparison of hyper-DMRs among different mutants.**

A. Heat-map showing the hyper-DMRs in mbd7 overlapping with those regions in ros5-1, ros4, ros1-1 in CG, CHG and CHH context. The color key is presented at right (light yellow indicates low methylation and black indicates high methylation).

B. The overlapping ratio of the hyper-DMRs (P <0.01) in mbd7 with those regions in ros5-1, ros4, and ros1-1 mutants in CG, CHG, and CHH context.

C. Compositions of the genomic locations of hyper-DMRs in ros5-1, ros4, ros1-1, and mbd7 mutants in CG, CHG, and CHH context. TE, transposon elements. The regions upstream and downstream of a gene were 2 kb away from TSS (the transcription start site) and TES (transcriptional end site).

**Figure 6. MBD7 binds to DNA regions with high methylation density.**
A. ChIP assay indicates that MBD7 binds to different parts of the T-DNA region and to some DMRs. The C24 wild-type plants expressing MBD7-GFP were used for the ChIP assay with anti-GFP antibodies. The fragments used for PCR in the T-DNA region are indicated in Figure 2A with red bars and with the numbers 1 (35S), 2 (NOS), 3 (Linker), and 4 (LUC). Other fragments for PCR in different DMRs are indicated by red bars in Figure 5B.

B. DNA methylation in different DMRs as indicated by whole-genome bisulfite sequencing and individual bisulfite sequencing in \textit{mbd7}, \textit{ros5-1}, \textit{ros4}, \textit{ros1-1}, and the C24 wild type. TE, transposon element.

**Figure 7. The proposed working model for MBD7.** MBD7 binds to genomic region with high density of CG methylation, which recruits ROS5/IDM2, and then IDM1/ROS4. ROS4/IDM1 may acetylate H3K18 and H2K23 to create a chromatin environment for further facilitating the recruitment of ROS1 protein for removal of DNA methyl groups. The proteins that directly recruit ROS1 need to be identified in the future study. \textit{RD29A} is regulated by ROS1, which is independent of MBD7-ROS5/IDM2-ROS4/IDM1 complex. Ac: H3K18 or 23 acetylation; $^m$CG or $^m$C: 5’-cytosine methylation. The light color $^m$C means the dynamically removed $^m$C.
Supplemental tables

**Supplemental Table S1.** Hypermethylated DMRs in *mbd7.*

**Supplemental Table S2.** The genomic locations of hypermethylated DMRs in *mbd7*

**Supplemental Table S3.** Primers used in this study.

**Figure S1.** *GUS* expression driven by *MBD7* promoter in seedlings growing on MS medium for 2.5, 3, 6, and 10 days.

**Figure S2.** A DNA methylation inhibitor and a DNA replication mutation in *rfc1-1* release *Pro35S:NPTII* gene silencing in *mbd7.*

A. 5’-Aza treatment (7 mg/L) rescues the kanamycin-sensitive phenotypes of *mbd7* and *ros1-1* on MS medium containing 50 mg/L kanamycin (Kan).

B. *NPTII* transcripts analyzed by real-time PCR in C24, *ros1-1*, and *mbd7* without or with 5’-Aza treatment (7 mg/L). Three independent experiments, each with three technical repeats, were done with similar results. Values are means ± SE (n=3) from one representative experiment.

C. Relative expression of *NPTII* in the C24 wild type, *mbd7*, *rfc1-1*, and *mbd7 rfc1-1* as determined by real-time PCR. Three independent experiments, each with three technical repeats, were done with similar results. Values are means ± SE (n=3) from one representative experiment.

**Figure S3. MBD7 does not interact with ROS4/IDM1 or ROS1**

A. MBD7 does not interact with ROS4/IDM1 or ROS1 in a firefly luciferase complementation imaging assay in tobacco leaves. The interaction of FLS2 with the PUB13 ARM domain was used as a positive control.

B. MBD7 does not interact with ROS4/IDM1 or ROS1 in a yeast two-hybrid assay. The full-length of *MBD7* was analyzed by yeast two-hybrid assay for interaction with the full-length of *ROS4/IDM1* and *ROS1.*

**Supplemental Figure S4.** The whole genome methylation difference between
mbd7 and the wild type (C24).

A. Frequency distribution histograms of methylation difference between C24 (the wild type) and mbd7 over the whole genome. The histograms were made with 100-bp analyzable windows over the genome-wide scale and the methylation levels of C24 and mbd7 in CG, CHG and CHH context were calculated, separately. The methylation difference between mbd7 and C24 was shown.

B. Frequency distribution histograms of significant methylation difference (P-value<0.01) between C24 (the wild type) and mbd7. The windows with significant methylation difference (P-value<0.01) between the C24 wild type and mbd7 were selected for comparison.

C. The genome wide methylation ratios in mbd7 and C24 in CG, CHG and CHH context.