RNA-directed DNA methylation in plants

Where to start?

Heng Zhang[1], Xinjian He[2], and Jian-Kang Zhu[1,3,*]

[1]Shanghai Center for Plant Stress Biology; Shanghai Institute of Biological Sciences; Chinese Academy of Sciences; Shanghai, P.R. China

[2]National Institute of Biological Sciences; Beijing, P.R. China

[3]Department of Horticulture and Landscape Architecture; Purdue University; West Lafayette, IN USA

Plants use 24-nucleotide small interfering RNAs (24-nt siRNAs) and long non-coding RNAs (lncRNAs) to direct de novo DNA methylation and transcriptional gene silencing. This process is called RNA-directed DNA methylation (RdDM). An important question in the RdDM model is what explains the target specificity of RNA polymerase IV (Pol IV), the enzyme that initiates siRNA production. Two recent papers addressed this question by characterizing the DTF1/SHH1 protein, which contains a homeodomain in the N terminus and a novel histone-binding domain SAWADEE in the C terminus. Here we review the main results of the two studies and discuss several possible mechanisms that could contribute to Pol IV and Pol V recruitment.

RNA-mediated transcriptional gene silencing (TGS) is a conserved phenomenon that occurs in fungi, plants, and animals.[1] It is required for many important cellular functions, such as transposon silencing, genome stability, cell identity maintenance, and defense against exogenous DNAs. A surprising feature of RNA-mediated TGS is that silencing requires transcriptional activity, i.e., low levels of transcripts can usually be detected at the silenced locus.[2] This seemingly contradictory observation can be explained by the action of small RNAs (sRNAs) and long non-coding RNAs (lncRNAs). Eighteen to 40 nt sRNAs, including small interfering RNAs (siRNA) and piwi-interacting RNAs (piRNA), can direct molecular machinery that catalyzes heterochromatic histone modifications or DNA methylation to loci with sequence homology, usually by base pairing with long non-coding RNAs (lncRNAs) that are associated with the chromatin. Thus, a low level of transcripts needs to be generated to provide positional information for TGS. A recent study that is part of the human ENCODE project reports that primary transcripts cover at least 74.7% of the human genome and that ncRNA expression is more cell type-specific than are protein-coding genes, suggesting that non-coding RNAs (ncRNAs) have a broad range of roles in various cellular functions.[3]

In plants, RNA-mediated TGS is mainly effected by a pathway called RNA-directed DNA methylation (RdDM) (**Fig. 1**), a term coined in 1994 by Sanger and colleagues.[4] Since then, carefully designed genetic screens in the model organism *Arabidopsis thaliana* have identified many components of the RdDM pathway, some of which, like the small RNA-binding protein ARGONAUTE, were subsequently found to exhibit conserved biochemical activities and to have similar roles in RNA-mediated TGS in other organisms.[5] While ncRNAs play important and similar roles in plants and animals, ncRNA generation differs in these two systems. In animals, the DNA-dependent RNA polymerase II (Pol II) is the main polymerase responsible for ncRNA generation, but ncRNA generation in plants involves at least three RNA polymerases: Pol II, Pol IV, and Pol V.

Pol IV and Pol V evolved from Pol II but have distinct functions in the RdDM

pathway. Pol IV is believed to produce single-stranded RNAs that serve as precursors of siRNAs. Pol IV is tightly coupled to RDR2 (RNA-DEPENDENT RNA POLYMERASE 2), which transcribes newly synthesized Pol IV transcripts and generates double-stranded RNAs (dsRNAs).[6] This process is likely aided by the putative ATP-dependent chromatin remodeling factor CLSY1 (CLASSY1).[7] The dsRNAs are diced by DICER-LIKE 3 (DCL3) into 24-nt siRNAs with 3′ overhangs, which are then methylated by HUA ENHANCER 1 (HEN1). Single-stranded 24-nt siRNAs are finally loaded into ARGONAUTE 4 (AGO4) or ARGONAUTE 6 (AGO6) to form the RdDM effector complex (**Fig. 1**).

Pol V and Pol II, in contrast, are involved in producing ncRNA scaffolds with which 24-nt sRNAs form base pairs.[8,9] Transcription by Pol V requires a protein complex DDR, named after its three components: DRD1 (DEFECTIVE IN RNA-DIRECTED DNA METHYLATION), DMS3 (DEFECTIVE IN MERISTEM SILENCING 3), and RDM1 (RNA-DIRECTED DNA METHYLATION).[10] DRD1 is a putative ATP-dependent chromatin remodeling factor, DMS3 is a putative cohesin-like protein, and RDM1 contains single-stranded DNA-binding activity. Researchers have proposed that the DDR complex is required to unwind the DNA and maintain the single-stranded DNA structure required for Pol V transcription.[11] RDM1 is also part of the RdDM effector complex and associates with AGO4 and the de novo methyltransferase DRM2 (DOMAIN REARRANGED METHYLTRANSFERASE 2) that locally catalyzes CG, CHG, and CHH methylation.[12] The scaffolding function of Pol V transcripts is further exemplified by several RNA-binding proteins that are necessary for proper methylation of RdDM targets. SPT5-LIKE/KOW DOMAIN-CONTAINING TRANSCRIPTION FACTOR 1 (SPT5L/KTF1) is recruited to Pol V transcripts independently of AGO4 and is thought to play a role in stabilizing the effector complex.[13,14] The IDN2-IDP protein complex also binds to Pol V transcripts[15] and recruits the SWI/SNF chromatin remodeling complex, whose
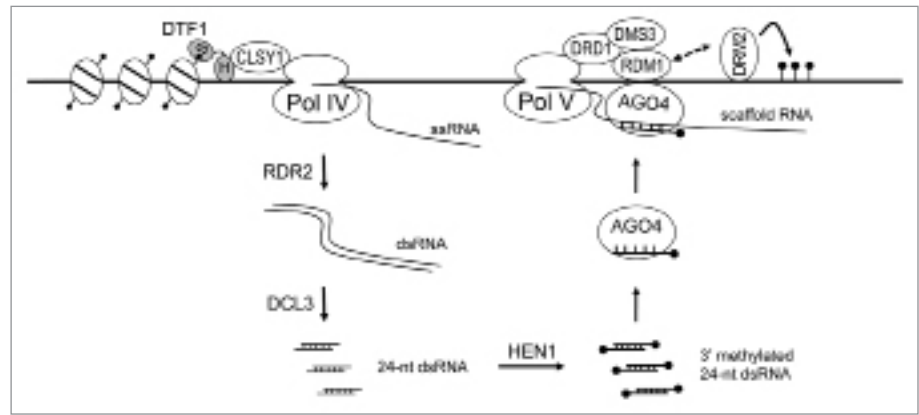
Figure 1. A simplified model of RNA-directed DNA methylation. The two domains (S, SAWADEE; H, homeodomain) of DTF1 protein are indicated by gray filled ovals. Filled black circles indicate methyl groups on histones, 24-nt dsRNA, or DNA; the dashed line with double arrows indicates the two proteins can interact, either directly or indirectly.

activity is required for heterochromatin formation.[16]

An important question is "What explains the target specificity of the RdDM pathway or of Pol IV and Pol V transcription?" Affinity purification followed by tandem mass spectrometry analyses indicates that Pol IV and Pol V are each composed of 12 subunits, half of which are shared by Pol II, Pol IV, and Pol V.[17] Phylogenetic analyses based on subunit 4, 5, and 7 of Pol IV and Pol V also showed that Arabidopsis Pol IV and Pol V evolved from Pol II via a multistep process.[18] Thus, it is possible that Pol IV and/or Pol V could be recruited by sequence-specific transcription factors as occurs with Pol II. Two recent studies used the ChIP-seq technique to characterize Pol V occupancy in the Arabidopsis genome.[10,19] A search for consensus sequences in the Pol V binding peaks, however, did not yield informative results, possibly because it is difficult to define the promoter region of Pol V transcription.

It is also possible that specific epigenetic marks play a role in Pol IV and/or Pol V recruitment. In Arabidopsis, the RdDM pathway mainly acts on repetitive sequences and transposons, most of which are marked with strong DNA methylation and other heterochromatic histone modifications such as monomethylation of lysine 27 of histone H3 (H3K27me1) and mono- and dimethylation of lysine 9 of histone H3 (H3K9me1/2). Pol IV and Pol V may have an intrinsic preference for

chromatin regions with such signatures. Alternatively, proteins that specifically recognize one or more of these marks may help recruit Pol IV and Pol V to the targets. The findings that different types of lncRNAs in mammalian cells are associated with different chromatin signatures suggest that epigenetic marks play a role in directing Pol II transcription of lncRNAs.[20,21]

Researchers have also proposed that specific nucleic acid structures may be involved in recruitment of Pol IV and Pol V.[11] In vitro transcription by Pol IV and Pol V requires a RNA primer hybridized to the DNA template.[6] Thus, the "R loop" structure formed by the RNA primer displacing one strand of the DNA template may be recognized by Pol IV and Pol V and may help recruit them to target loci.

A protein that could possibly contribute to the target specificity of Pol IV is DTF1/SHH1 (DNA TRANSCRIPTION FACTOR 1/SAWADEE HOMEODOMAIN HOMOLOG 1). DTF1 was first identified as a protein that co-purifies with NRPD1, the largest subunit of Pol IV.[22] Other co-purified proteins include the RdDM components RDR2, CLSY1, and RDM4. DTF1 was also discovered independently in a genetic screen for RdDM mutants.[23] DTF1 function is required for proper DNA methylation and for proper 24-nt siRNA levels at the selected RdDM loci examined in the latter two studies. The *DTF1* gene encodes a plant-specific protein with 167 amino acid residues. The

N-terminal part of the protein contains a cryptic homeodomain, and the C-terminal part contains a conserved plant-specific domain called SAWADEE.[24] The homeodomain is a well-characterized DNA-binding domain and usually adopts a 3-helix structure with the third helix involved in DNA recognition.[25] The homeodomain sequence of DTF1 is only remotely similar to other homeodomain sequences, but the sequence corresponding to the third helix is conserved, suggesting that it could be involved in recognition of specific DNA motifs (HZ and JKZ, unpublished data). The SAWADEE domain, on the other hand, was predicted to adopt a chromo barrel structure that may be involved in binding to specific histone modifications[26] (see below). Thus, the predicted functions of its two domains suggest that DTF1 could recognize specific chromatin signatures and may be involved in Pol IV recruitment.

Two recent papers (published in May 2013) further characterized the DTF1/SHH1 protein and reported that it is involved in Pol IV recruitment.[26,27] Both research groups used high-throughput sequencing to characterize the DNA methylome and whole-genome 24-nt siRNA levels in the *dtf1* plants and several other RdDM mutants. Mutation of *DTF1* has a dramatic effect on the 24-siRNA levels.[26] While 24-nt siRNAs decrease to less than 10% in the *nrpd1* mutant, they decrease to 28% in the *dtf1* mutant, which is substantially greater than the 50% decrease observed in the *nrpe1* mutant (a knockout mutant in the largest subunit of Pol V). These numbers are consistent with the number of siRNA clusters that show decreased expression in each mutant. The *dtf1* and *nrpe1* differentially expressed siRNA clusters are subsets of those of *nrpd1*. Each accounts for 83% and 68%, respectively, of the differentially expressed siRNA clusters identified in *nrpd1*, indicating that *DTF1* is involved in the upstream steps of RdDM pathway. This is unlike *NRPE1*, which presumably affects 24-nt siRNA accumulation indirectly through affecting DNA methylation levels. Not surprisingly, mutation of *DTF1* also leads to a decrease of DNA methylation across the whole genome. The number of hypo DMRs (differentially methylated regions)

identified in the *dtf1* mutant accounts for ~40% of those identified in the *nrpd1* or *nrpe1* mutant. Immunoaffinity purification of DTF1 followed by tandem mass spectrometry indicates that DTF1 specifically associates with Pol IV but not Pol V.[26] Thus, DTF1 is an upstream RdDM factor that specifically associates with Pol IV.

To understand DTF1 function, Zhang et al. followed an in silico approach in order to predict its protein structure using the Phyre2 server.[28] The prediction indicated that the SAWADEE domain could adopt a structure similar to that of the chromo barrel domain.[29] Proteins with such structures have been shown to bind specific histone modifications. Thus, purified recombinant SAWADEE domain was applied to a histone peptide array. The results indicate that the SAWADEE domain can specifically recognize histone H3 peptide with methylated lysine 9 (H3K9me1/2/3) but that the binding is blocked by the presence of methylation on lysine 4 of the same peptide.[26] Law et al. obtained similar results using another type of histone peptide array. In addition, they solved the crystal structure of the SAWADEE domain complexed with histone peptide. The structure indicates that the SAWADEE domain contains two binding pockets, one that recognizes unmethylated H3K4 (H3K4me0) and the other that recognizes H3K9me1/2/3.[27] Thus, SAWADEE is a new reader of histone marks.

The results so far suggest that DTF1 functions in recruiting Pol IV to chromatin regions with H3K9 methylation. Law et al., who tested this hypothesis by examining NRPD1 occupancy genome wide in the *dtf1* mutant, observed a decreased NRPD1 association with the chromatin at loci where 24-nt siRNA levels are decreased in the *dtf1* mutant.[27]

It is unknown whether the homeodomain of DTF1 is also involved in recognizing specific DNA sequences and, thus, in providing additional targeting information for Pol IV. That possibility is supported, however, by the observation that DTF1 apparently only affects 24-nt siRNA levels of subsets of Pol IV targets.[26,27] The Arabidopsis genome contains a DTF1 homolog named DTF2/

SHH2 that has the same domain arrangement as DTF1 except that its SAWADEE domain has a longer C-terminal region. The SAWADEE domain of DTF2 also binds to H3K9me1/2/3 in in vitro binding assays.[26] It is possible that the homeodomains of DTF1 and DTF2 preferentially recognize different DNA sequences, allowing DTF1 and DTF2 to target methylated H3K9 regions with different DNA sequence contexts. In support of this inference, the siRNA clusters that depend on DTF1 and Pol IV are enriched in euchromatin, whereas the siRNA clusters that depend only on Pol IV are enriched in pericentromeric heterochromatin.[27]

The exact molecular mechanism by which DTF1 affects Pol IV recruitment is not clear. The possibility that DTF1 could interact with any of the Pol IV-specific subunits has not been tested. The yeast homologs of NRPD4/NRPE4 and NRPB5/NRPD5 interact with transcription factors.[18] Yeast two-hybrid assays indicated that DTF1 directly associates with CLSY1 but not with RDR2.[26] Previous experiments found that the subnuclear localization patterns of RDR2 and NRPD1 were disrupted in the *clsy1* mutant,[7] suggesting that DTF1 could affect RDR2 and Pol IV localization through CLSY1.

The finding that DTF proteins bind to H3K9 methylation provides the first direct link between histone modifications and RNA-directed DNA methylation. This also indicates, however, that we are only beginning to understand how Pol IV and Pol V are recruited. Although DTF1 is required for Pol IV targeting, it is not clear whether it is sufficient. If it is sufficient, then the question remains "What directs H3K9 methylation in general?" That 24-nt siRNAs and scaffolding ncRNAs are each generated by different RNA polymerases in plants means that the two processes can now be studied separately. Factors involved in Pol V recruitment have not been reported. How Pol IV and Pol V transcriptional processes at the same locus are coordinated also remains unknown. Direct evidence showing dynamic changes of the DNA methylome in plants under biotic stress has been reported, indicating that DNA methylation can be dynamically regulated in response to environmental cues.[30] Could

factors like DTF1 be involved in targeting the RdDM machinery to those newly methylated regions? Answers to these and related questions will greatly increase our understanding of the regulated genesis of ncRNAs.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

References

1. Castel SE, Martienssen RA. RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. Nat Rev Genet 2013; 14:100-12; http://dx.doi.org/10.1038/nrg3355; PMID:23329111

2. Grewal SI, Elgin SC. Transcription and RNA interference in the formation of heterochromatin. Nature 2007; 447:399-406; http://dx.doi.org/10.1038/nature05914; PMID:17522672

3. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, et al. Landscape of transcription in human cells. Nature 2012; 489:101-8; http://dx.doi.org/10.1038/nature11233; PMID:22955620

4. Wassenegger M, Heimes S, Riedel L, Sänger HL. RNA-directed de novo methylation of genomic sequences in plants. Cell 1994; 76:567-76; http://dx.doi.org/10.1016/0092-8674(94)90119-8; PMID:8313476

5. Vaucheret H. Plant ARGONAUTES. Trends Plant Sci 2008; 13:350-8; http://dx.doi.org/10.1016/j.tplants.2008.04.007; PMID:18508405

6. Haag JR, Ream TS, Marasco M, Nicora CD, Norbeck AD, Pasa-Tolic L, Pikaard CS. In vitro transcription activities of Pol IV, Pol V, and RDR2 reveal coupling of Pol IV and RDR2 for dsRNA synthesis in plant RNA silencing. Mol Cell 2012; 48:811-8; http://dx.doi.org/10.1016/j.molcel.2012.09.027; PMID:23142082

7. Smith LM, Pontes O, Searle I, Yelina N, Yousafzai FK, Herr AJ, Pikaard CS, Baulcombe DC. An SNF2 protein associated with nuclear RNA silencing and the spread of a silencing signal between cells in Arabidopsis. Plant Cell 2007; 19:1507-21; http://dx.doi.org/10.1105/tpc.107.051540; PMID:17526749

8. Wierzbicki AT, Haag JR, Pikaard CS. Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. Cell 2008; 135:635-48; http://dx.doi.org/10.1016/j.cell.2008.09.035; PMID:19013275

9. Zheng B, Wang Z, Li S, Yu B, Liu JY, Chen X. Intergenic transcription by RNA polymerase II coordinates Pol IV and Pol V in siRNA-directed transcriptional gene silencing in Arabidopsis. Genes Dev 2009; 23:2850-60; http://dx.doi.org/10.1101/gad.1868009; PMID:19948763

10. Zhong X, Hale CJ, Law JA, Johnson LM, Feng S, Tu A, Jacobsen SE. DDR complex facilitates global association of RNA polymerase V to promoters and evolutionarily young transposons. Nat Struct Mol Biol 2012; 19:870-5; http://dx.doi.org/10.1038/nsmb.2354; PMID:22864289

11. Pikaard CS, Haag JR, Pontes OM, Blevins T, Cocklin R. A Transcription Fork Model for Pol IV and Pol V-dependent RNA-Directed DNA Methylation. Cold Spring Harb Symp Quant Biol 2013.

12. Gao Z, Liu HL, Daxinger L, Pontes O, He X, Qian W, Lin H, Xie M, Lorkovic ZJ, Zhang S, et al. An RNA polymerase II- and AGO4-associated protein acts in RNA-directed DNA methylation. Nature 2010; 465:106-9; http://dx.doi.org/10.1038/nature09025; PMID:20410883

13. He XJ, Hsu YF, Zhu S, Wierzbicki AT, Pontes O, Pikaard CS, Liu HL, Wang CS, Jin H, Zhu JK. An effector of RNA-directed DNA methylation in arabidopsis is an ARGONAUTE 4- and RNA-binding protein. Cell 2009; 137:498-508; http://dx.doi.org/10.1016/j.cell.2009.04.028; PMID:19410546

14. Rowley MJ, Avrutsky MI, Sifuentes CJ, Pereira L, Wierzbicki AT. Independent chromatin binding of ARGONAUTE4 and SPT5L/KTF1 mediates transcriptional gene silencing. PLoS Genet 2011; 7:e1002120; http://dx.doi.org/10.1371/journal.pgen.1002120; PMID:21738482

15. Zhang CJ, Ning YQ, Zhang SW, Chen Q, Shao CR, Guo YW, Zhou JX, Li L, Chen S, He XJ. IDN2 and its paralogs form a complex required for RNA-directed DNA methylation. PLoS Genet 2012; 8:e1002693; http://dx.doi.org/10.1371/journal.pgen.1002693; PMID:22570638

16. Zhu Y, Rowley MJ, Böhmdorfer G, Wierzbicki ATA. A SWI/SNF chromatin-remodeling complex acts in noncoding RNA-mediated transcriptional silencing. Mol Cell 2013; 49:298-309; http://dx.doi.org/10.1016/j.molcel.2012.11.011; PMID:23246435

17. Ream TS, Haag JR, Wierzbicki AT, Nicora CD, Norbeck AD, Zhu JK, Hagen G, Guilfoyle TJ, Pasa-Tolić L, Pikaard CS. Subunit compositions of the RNA-silencing enzymes Pol IV and Pol V reveal their origins as specialized forms of RNA polymerase II. Mol Cell 2009; 33:192-203; http://dx.doi.org/10.1016/j.molcel.2008.12.015; PMID:19110459

18. Tucker SL, Reece J, Ream TS, Pikaard CS. Evolutionary history of plant multisubunit RNA polymerases IV and V: subunit origins via genome-wide and segmental gene duplications, retrotransposition, and lineage-specific subfunctionalization. Cold Spring Harb Symp Quant Biol 2010; 75:285-97; http://dx.doi.org/10.1101/sqb.2010.75.037; PMID:21447813

19. Wierzbicki AT, Cocklin R, Mayampurath A, Lister R, Rowley MJ, Gregory BD, Ecker JR, Tang H, Pikaard CS. Spatial and functional relationships among Pol V-associated loci, Pol IV-dependent siRNAs, and cytosine methylation in the Arabidopsis epigenome. Genes Dev 2012; 26:1825-36; http://dx.doi.org/10.1101/gad.197772.112; PMID:22855789

20. Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature 2009; 458:223-7; http://dx.doi.org/10.1038/nature07672; PMID:19182780

21. De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, Muller H, Ragoussis J, Wei CL, Natoli G. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. PLoS Biol 2010; 8:e1000384; http://dx.doi.org/10.1371/journal.pbio.1000384; PMID:20485488

22. Law JA, Vashisht AA, Wohlschlegel JA, Jacobsen SE. SHH1, a homeodomain protein required for DNA methylation, as well as RDR2, RDM4, and chromatin remodeling factors, associate with RNA polymerase IV. PLoS Genet 2011; 7:e1002195; http://dx.doi.org/10.1371/journal.pgen.1002195; PMID:21811420

23. Liu J, Bai G, Zhang C, Chen W, Zhou J, Zhang S, Chen Q, Deng X, He XJ, Zhu JK. An atypical component of RNA-directed DNA methylation machinery has both DNA methylation-dependent and -independent roles in locus-specific transcriptional gene silencing. Cell Res 2011; 21:1691-700; http://dx.doi.org/10.1038/cr.2011.173; PMID:22064704

24. Mukherjee K, Brocchieri L, Bürglin TR. A comprehensive classification and evolutionary analysis of plant homeobox genes. Mol Biol Evol 2009; 26:2775-94; http://dx.doi.org/10.1093/molbev/msp201; PMID:19734295

25. Billeter M. Homeodomain-type DNA recognition. Prog Biophys Mol Biol 1996; 66:211-25; http://dx.doi.org/10.1016/S0079-6107(97)00006-0; PMID:9284451

26. Zhang H, Ma ZY, Zeng L, Tanaka K, Zhang CJ, Ma J, Bai G, Wang P, Zhang SW, Liu ZW, et al. DTF1 is a core component of RNA-directed DNA methylation and may assist in the recruitment of Pol IV. Proc Natl Acad Sci U S A 2013; 110:8290-5; http://dx.doi.org/10.1073/pnas.1300585110; PMID:23637343

27. Law JA, Du J, Hale CJ, Feng S, Krajewski K, Palanca AM, Strahl BD, Patel DJ, Jacobsen SE. Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. Nature 2013; 498:385-9; http://dx.doi.org/10.1038/nature12178; PMID:23636332

28. Kelley LA, Sternberg MJ. Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc 2009; 4:363-71; http://dx.doi.org/10.1038/nprot.2009.2; PMID:19247286

29. Nielsen PR, Nietlispach D, Buscaino A, Warner RJ, Akhtar A, Murzin AG, Murzina NV, Laue ED. Structure of the chromo barrel domain from the MOF acetyltransferase. J Biol Chem 2005; 280:32326-31; http://dx.doi.org/10.1074/jbc.M501347200; PMID:15964847

30. Dowen RH, Pelizzola M, Schmitz RJ, Lister R, Dowen JM, Nery JR, Dixon JE, Ecker JR. Widespread dynamic DNA methylation in response to biotic stress. Proc Natl Acad Sci U S A 2012; 109:E2183-91; http://dx.doi.org/10.1073/pnas.1209329109; PMID:22733782