Supplementary Material

Role of the *Arabidopsis DRM* Methyltransferases in De Novo DNA Methylation and Gene Silencing

Xiaofeng Cao and Steven E. Jacobsen

Supplementary Results and Discussion

Characterization of the *clk-st* Transgene and Its Silencing Properties

We previously described a line, *clk-st*, which shows a stable (nonreverting) epigenetic *clark kent* phenotype due to the presence of additional transgenic copies of the *SUP* gene [S1]. We further characterized this line by Southern blot analysis of the *SUP* loci with the restriction enzymes EcoRI, EcoRV, PvuII, BamHI, HincII, KpnI, SalI, and XhoI and found that *clk-st* contains a single transgene locus consisting of a 24 kilobase inverted repeat of the entire T-DNA cassette, including the 6.7 kilobase *SUP* gene and the associated Kanamycin resistance gene. The right borders of the T-DNA face each other in the center of the repeat. We used TAIL PCR [S2] and DNA sequencing to determine the *clk-st* T-DNA insertion site. The T-DNA is inserted between positions 66180 and 66186 on BAC clone F26G5 (GeneBank accession AL353814) on chromosome III, creating a five base pair deletion in an intergenic region. PCR and sequencing of both ends of the insertion confirmed the inverted repeat structure of the *clk-st* insert, since left border sequences were found on both sides of the insertion site. Furthermore, 3.7 kilobase PCR products were obtained by using an oligonucleotide primer in the 5′ end of *SUP* gene (JP947, 5′-GTAGAATGAGG TCTTGGGAACTTTGTCAA-3′) in combination with primers in the genomic DNA on either side of the T-DNA insertion (JP1214, TCCCCGATATATTGTGGGTTAAAATTTTG-3′; and JP1215, CAAAT GTGGTGGTGGTGATCAAAAGTGAT-3′) but not when JP947 was used alone. A three-oligonucleotide system was developed for genotyping the *clk-st* inverted repeat insertion, which consisted of the oligonucleotides JP700, 5′-TTTCTCCATATTGACCATCATACT CATTG-3′, JP1214, and JP1215. We further confirmed the genetic location of this T-DNA in an F2 mapping population derived from a cross of the *clk-st* line in the Ler ecotype with the wild-type ecotype Columbia. In this population, the Kanamycin resistance locus showed linkage (1 recombinant out of 38 chromosomes) with a PCR-based marker, CER479314, on BAC clone T28A8 (GenBank accession AL162691).

We found that the inverted repeat transgene *SUP* locus present in *clk-st* induces de novo methylation and gene silencing of a previously unmethylated and active *SUP* endogene. This silencing phenomenon occurs after two or more generations of exposure of the *SUP* endogene to the *SUP* inverted repeat. The silencing properties of the *clk-st* inverted repeat are illustrated by the results of the following three experiments.

In the first experiment, we utilized the *cmt3-7* mutation (a null *CMT3* allele), which eliminates the majority of CpNpG and asymmetric methylation of *SUP* toward the 5′ end of the gene causing reactivation of *SUP* expression [S1]. A line homozygous for *cmt3-7* in the *clk-st* homozygous background was backcrossed to the *clk-st* strain (which contains methylated and silenced *SUP* genes). While the F1 of this cross showed a wild-type *SUP* phenotype, the F2 segregated 3:1 for *clark kent*:wild-type plants, and we observed a perfect correlation between wild-type *SUP* plants and *cmt3-7* homozygosity. Therefore, all plants with a wild-type *CMT3* allele had undergone resilencing of the *SUP* genes.

In a second experiment, a line homozygous for *cmt3-7* in the *clk-st* homozygous background was crossed to the wild-type ecotype Ler two successive times. Ten F1 plants from the second backcross were self-pollinated, and the F2 progeny were tested on kanamycin plates for the presence of the *clk-st* transgene and also sown on soil. None of the six F2 families that had lost the *clk-st* transgene segregated plants with a *clark kent* phenotype. However, all four of the F2 families that retained the *clk-st* transgene segregated *clark kent* plants (10/74 = 14% *clark kent* plants).

In a third experiment, the *clk-st* inverted repeat *SUP* transgene locus was introgressed into either the Columbia or WS ecotypes by four successive backcrosses. After the fourth backcross, the plants were allowed to self-pollinate, and F2 plants were sown. While the F1 plants of each backcross showed a wild-type *SUP* phenotype, 15% of the F2 plants in the Columbia population and 29% of the plants in the WS population developed a *clark kent* phenotype. All of the plants showing a *clark kent* phenotype (total of 50) contained the *clk-st* transgene. Thirty-four of these plants were homozygous for the *clk-st* transgene, while only 16 were hemizygous, indicating that silencing of the endogenous *SUP* locus occurs more frequently in plants containing two copies of the inverted repeat *SUP* locus ($\chi^2 = 27$, $p < 0.005$, null hypothesis 2:1 hemizygous:homozygous). Furthermore, the F2 plants homozygous for the *clk-st* transgene showed a stronger *sup* phenotype (56% ± 5.0% [mean ± standard error] defective carpels) than plants hemizygous for the *clk-st* transgene (21% ± 3.5% defective carpels, t test p value = 6.4×10^{-5}). Thus, two copies of the *SUP* inverted repeat appear to be more effective in silencing the *SUP* endogene than one copy.

Cosegregation of the *SUP* De Novo Methylation Phenotype with the *drm* Loci

Since only one allele of *drm1* and one allele of *drm2* was used in this study, there remains the possibility that a non-*drm* mutation was present in the T-DNA lines used to isolate the *drm* mutations and that this non-*drm* mutation is responsible for the de novo methylation phenotypes observed. Therefore, to ensure that the de novo methylation phenotype cosegregates with the *drm* loci, we used an extensively backcrossed *drm1 drm2* double mutant plant in the segregation experiment shown in Figure 3C. Based on the crossing scheme, one can calculate the probability that an unlinked non-*drm* mutation could explain our results. The calculation follows.

First, we backcrossed the *drm1* and *drm2* mutations to the wild-type WS strain three times in the process of creating the *drm1 drm2* double mutant strain (this is explained in Experimental Procedures). We then crossed the *drm1 drm2* double mutant to the wild-type *Ler* strain an additional five times. In each of these backcrosses, we selected for kanamycin resistance (a kanamycin resistance marker is present in the T-DNA) and then confirmed that the *drm1* and *drm2* mutations were present by PCR genotyping. A *drm1 drm2* double heterozygote from this last backcross was used in the cross diagrammed in Figure 3C. Since there were a total of eight backcrosses, there would be a $(1/2)^8$ or 1/256 chance that an unlinked non-*drm* mutation would still have been present in the double heterozygote used in Figure 3.

Further, for a non-*drm* mutation to explain the de novo methylation results presented in Figure 3, it would have to cosegregate with the *drm1*/*drm2* mutations in all 11 F1 plants. For an unlinked mutation, the chances of this are $(1/2)^{11}$ or 1/2048.

Finally, in order to explain the results of Figure 3 in terms of a non-*drm* mutation, we also need to consider the probability of the mutation being homozygous in line 30. To create line 30, we first crossed the *drm1 drm2* double mutant to *clk-st* two successive times and then self-pollinated to isolate a *drm1 drm2* homozygote (described in Experimental Procedures). This *clk-st drm1 drm2* line was then crossed to *cmt3-7 clk-st*, and the F1s were allowed to self (described in Figure 3B). The chance of an unlinked non-*drm* mutation being homozygous in line 30 is therefore $(1/2)^6$ or 1/64.

Thus, the chance of an unlinked non-*drm* mutation explaining the results of Figure 3C is the product of these separate probabilities, which is $(1/2)^{25}$ or 1/33,554,432.

If the hypothesized non-*drm* mutation were instead 15 centimorgans away from the *drm* loci, then the chances of cosegregation in the Figure 3 experiments would be $(0.85)^{25}$ or 1.7%. Thus, if there is an unrelated de novo methylation mutation, it is likely to be linked to *drm1* and *2*.

Supplementary References

S1. Lindroth, A.M., Cao, X., Jackson, J.P., Zilberman, D., McCallum, C.M., Henikoff, S., and Jacobsen, S.E. (2001). Requirement of *CHROMOMETHYLASE3* for maintenance of CpXpG methylation. Science *292*, 2077–2080.

S2. Liu, Y.G., Mitsukawa, N., Oosumi, T., and Whittier, R.F. (1995). Efficient isolation and mapping of Arabidopsis thaliana T-DNA insert junctions by thermal asymmetric interlaced PCR. Plant J. *8*, 457–463.