nature structural & molecular biology

IDN1 and IDN2 are required for *de novo* DNA methylation in *Arabidopsis thaliana*

Israel Ausin[1], Todd C Mockler[2], Joanne Chory[3,4] & Steven E Jacobsen[1,5]

DNA methylation is an epigenetic mark affecting genes and transposons. Screening for mutants that fail to establish DNA methylation yielded two we termed "*involved in de novo*" (*idn*) 1 and 2. *IDN1* encodes DMS3, an SMC-related protein, and *IDN2* encodes a previously unknown double-stranded RNA–binding protein with homology to SGS3. *IDN1* and *IDN2* control *de novo* methylation and small interfering RNA (siRNA)-mediated maintenance methylation and are components of the RNA-directed DNA methylation pathway.

© 2009 Nature America, Inc. All rights reserved.

npg

All known *de novo* DNA methylation in *Arabidopsis thaliana* is carried out by domains rearranged methylase 2 (DRM2)[1,2], a homolog of DNA methyltransferase 3. DRM2 is guided by siRNAs in the RNA-directed DNA methylation (RdDM) pathway[3]. Current models suggest that the RNA polymerase IV (Pol IV) complex is recruited to target sequences by an unknown mechanism, resulting in the production of single-stranded RNA, which is converted into double-stranded RNA by RNA-dependent RNA polymerase 2 (RDR2) and subsequently processed by dicer-like 3 (DCL3) into 24-nucleotide siRNAs. These siRNAs are then loaded into argonaute 4 (AGO4), which interacts with the Pol V machinery, which in turn transcribes RNA at silent loci[4,5]. AGO4–siRNA complexes are then thought to pair with nascent RNA transcripts to guide DRM2 to initial target loci[6]. To better understand RdDM, we performed a screen to identify mutants affecting *de novo* methylation using *FWA* gene silencing as a model.

The endogenous *FWA* gene is heritably silenced by methylation at two tandem repeats in its upstream region. Unmethylated *fwa* epialleles show *FWA* ectopic expression, resulting in a dominant late-flowering phenotype that is easily scored[7]. In wild-type plants, methylation of *FWA* transgenes is efficiently established, whereas RdDM mutants fail to methylate and silence *FWA* and therefore flower late. By screening for mutants that affect the establishment of silencing at *FWA* transgenes but that do not affect pre-existing silencing at the *FWA* endogene, one can screen for *de novo* methylation mutants[1,2].

We screened a collection of 429 T-DNA insertion mutants using *FWA Agrobacterium tumefaciens*–mediated transformation. This collection includes insertions in genes encoding RNA-binding proteins, Agenet domain–containing proteins, Jumonji domain–containing proteins and other chromatin-related proteins (**Supplementary Table 1**). This screen identified two mutants that were late flowering after transformation but showed normal flowering before transformation (**Fig. 1a**). We termed these mutants "*involved in de novo*" 1 and 2 (*idn1-1* and *idn2-1*). The late-flowering phenotype of these mutants was correlated with reduced methylation of the *FWA* transgenes as revealed by genomic bisulfite sequencing (**Supplementary Table 2, Supplementary Methods**). Methylation was severely reduced in all DNA sequence contexts (CG, CHG and CHH, where H = A, T or C); however, pre-established CG methylation remained unaffected at the *FWA* endogenous gene (**Fig. 1b**).

Another gene that can be used to test for *de novo* methylation is suppressor of *drm1 drm2 cmt3* (*SDC*). *SDC* possesses seven tandem repeats in its promoter region. Hypomethylation at these repeats leads to overexpression of *SDC* and a characteristic morphological phenotype including curled leaves and short stature. All RdDM mutants tested so far fail to efficiently establish methylation at either *SDC* or *FWA* transgenes[8]. We found that the *idn* mutants also showed a reduced ability to establish silencing of incoming *SDC* transgenes. This was especially evident in the case of *idn1-1*: T_1 *idn1-1* plants transformed with *SDC* showed a marked increase in *SDC* expression, whereas *idn2-1* T_1 transformants had nearly wild-type *SDC* expression levels (**Fig. 1c**). *SDC* overexpression was correlated with moderate hypomethylation at *SDC* tandem repeats as assayed by bisulfite sequencing (**Fig. 1c**). These data reinforce the idea that the proteins IDN1 and IDN2 control *de novo* methylation.

Thus far, all mutations known to block *de novo* methylation also affect maintenance of methylation at several loci[1,2,9]. We analyzed the methylation state of representative endogenous targets, including the *MEA-ISR*, *5S*, *SDC* and *FWA* repeats, using methylation-sensitive enzymes coupled with Southern blots and/or bisulfite sequencing techniques. The *idn1-1* and *idn2-1* mutations both caused a reduction of non-CG methylation at *MEA-ISR* and *FWA* repeats (**Figs. 1c** and **2a**,**b**). Furthermore, the mutations also caused reduced CG methylation at the *5S* loci and a minor reduction in CHH methylation at the *SDC* repeats (**Figs. 2b**,**c**). These methylation phenotypes are consistent with those seen in *drm2* and other RdDM mutants[3], suggesting that *IDN1* and *IDN2* are important in the RdDM pathway.

Because *SDC* silencing is redundantly controlled by both the RdDM and chromomethylase 3 (CMT3) pathways[8], we crossed the *idn* mutants with the null CMT3 allele *cmt3-11*. As previously reported for *drm1 drm2 cmt3* triple mutants[8], we observed strong *SDC* overexpression and the accompanying *SDC* morphological

[1]Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, Los Angeles, California, USA. [2]Department of Botany and Plant Pathology and Center for Genome Research and Biocomputing, Oregon State University, Corvallis, Oregon, USA. [3]Plant Biology Laboratory, Salk Institute for Biological Studies, La Jolla, California, USA. [4]Howard Hughes Medical Institute, La Jolla, California, USA. [5]Howard Hughes Medical Institute, University of California, Los Angeles, Los Angeles, California, USA. Correspondence should be addressed to S.E.J. (jacobsen@ucla.edu).

Received 30 March; accepted 11 September; published online 15 November 2009; doi:10.1038/nsmb.1690

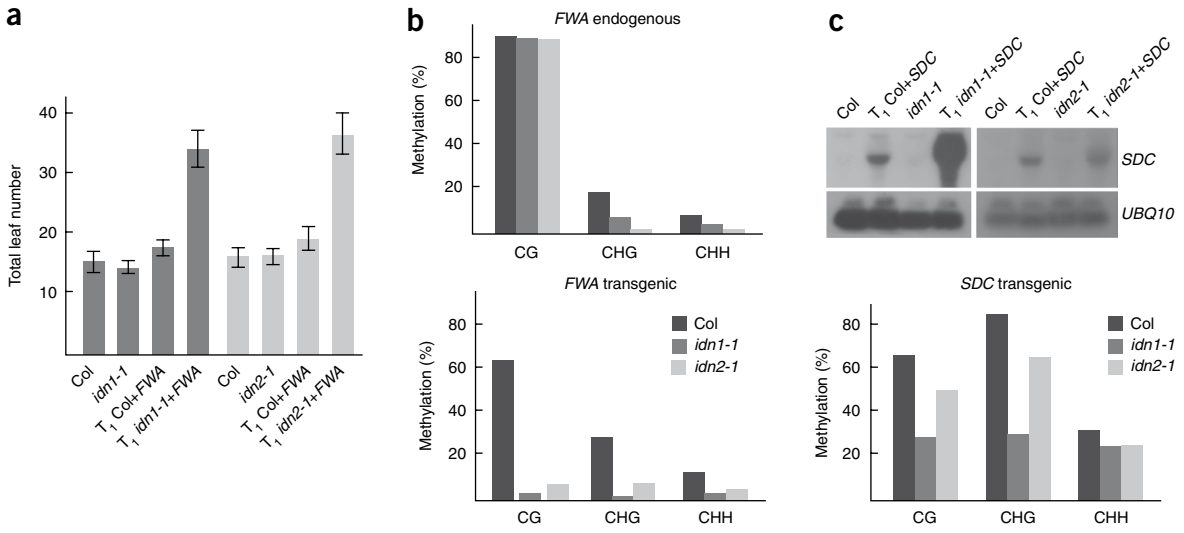

Figure 1 *De novo* methylation phenotype of *idn* mutants. (**a**) Flowering time measured as total number of leaves produced by wild-type (Col), *idn* mutants and *FWA*-transformed T_1 plants under long-day conditions. Error bars, s.d. (**b**) Methylation status of wild-type and *idn* mutants at endogenous and transgenic *FWA*. (**c**) *SDC* expression levels in *idn* mutants before and after *SDC* transformation. Hybridization with a polyubiquitin (*UBQ10*) probe is shown as loading control. Panel below shows methylation status of wild type and *idn* mutants at transgenic *SDC*.

© 2009 Nature America, Inc. All rights reserved.

phenotype in the *idn1-1 cmt3-11* double mutant. However, the *idn2-1 cmt3-11* double mutant did not overexpress *SDC*. The *idn1-1 cmt3-11* double mutant consistently showed reduced CG and CHG methylation at *SDC*, whereas *idn2-1 cmt3-11* showed only a partial decrease in CHH methylation (**Figs. 2b**,**c**). These results suggest that *IDN1* has a stronger or more general role in RdDM than *IDN2*.

To determine where the IDN proteins act within the RdDM pathway, we examined siRNA abundance at several loci. Comparison between the *idn* mutants and other RdDM pathway mutants showed that both *idn* mutants had normal (*SDC* and *siR02*) or slightly reduced (*FWA*, *AtSN1*, *MEA-ISR*, *5S* and *siR1003*) siRNAs levels, but they did not completely lack siRNAs as observed in *rdr2-1* or *nrpd1a* mutants (**Supplementary Fig. 1**). This suggests that IDN1 and IDN2 are likely to act downstream of initial siRNA biogenesis.

To gain further insight into the molecular mechanism of *de novo* methylation, we cloned both *IDN1* and *IDN2*. Because neither of the T-DNA insertions co-segregated with the respective mutant phenotypes, we isolated both *IDN1* and *IDN2* genes using a map-based approach. The *idn* phenotype at *MEA-ISR* segregated in a mendelian fashion in F_2 Landsberg *erecta* × *idn* populations, allowing us to follow the mutant phenotypes. Fine-mapping strategies localized each mutation to a small genomic interval on chromosome III (**Supplementary Fig. 2a**). We identified *IDN1* and *IDN2* initially by DNA sequence mutations in these regions. The *idn1-1* allele carries a point mutation, whereas the *idn2-1* allele carries both a point mutation and a 24-base-pair deletion. To confirm gene identification, we complemented *idn1-1* and *idn2-1* with genomic fragments containing wild-type *IDN1* or *IDN2*, respectively (**Supplementary Figs. 2b**,**c**). In addition, we characterized additional

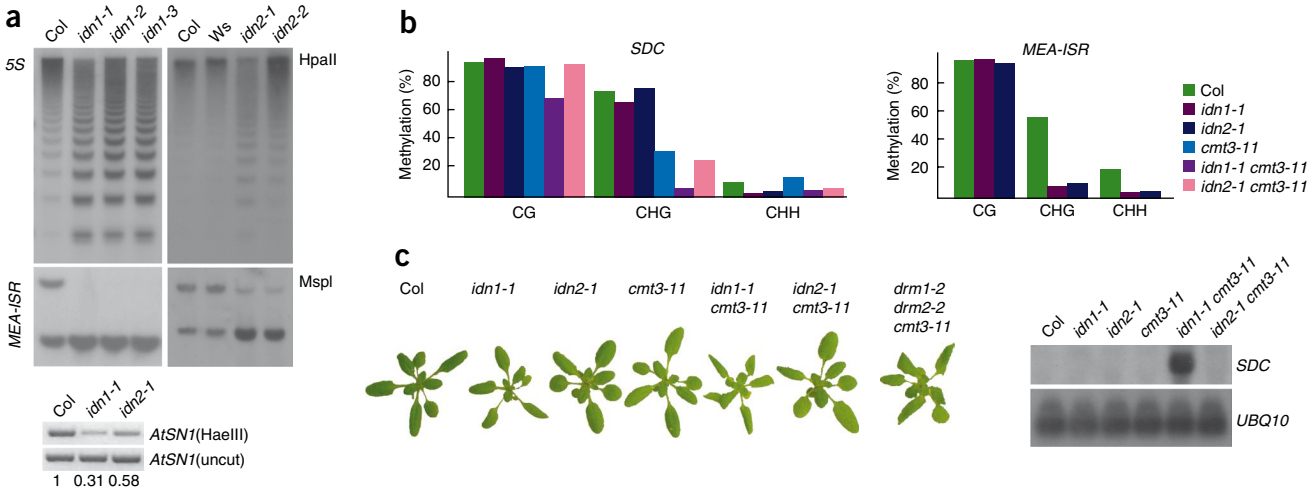

Figure 2 Maintenance methylation phenotype of *idn* mutants. (**a**) Methylation-sensitive enzyme Southern hybridization assay at *5S* and *MEA-ISR* loci, and HaeIII cutting assay at *AtSN1*. HaeIII is blocked by C methylation in GGCC context, HpaII is blocked by C methylation in CCGG context and MspI is blocked by methylation of the external C in CCGG context. (**b**) Methylation status of wild type (Col), *idn* and *idn cmt3-11* double mutants at endogenous *SDC* and methylation status of wild type and *idn* mutants at *MEA-ISR*. (**c**) Morphological and molecular phenotype of *idn cmt3-11* double mutants.

alleles of *IDN1* and *IDN2* obtained from available T-DNA collections, which we termed *idn1-2*, *idn1-3*, and *idn2-2* (**Supplementary Fig. 2b**), and found similar methylation defects at the *5S* and *MEA-ISR* loci (**Fig. 2a**).

IDN1 encodes defective in meristem silencing 3 (DMS3, or At3g49250), a previously described protein that shares homology with structural maintenance of chromosome (SMC) proteins[10]. The *dms3* alleles were shown to be required for RdDM and for the production of secondary siRNAs[10] as well as of Pol V noncoding transcripts[6], consistent with a role of IDN1 (DMS3) in a downstream part of the RNA-directed *de novo* methylation pathway.

IDN2 (*At3g48670*) encodes a predicted protein of 648 amino acids containing a zinc finger domain, and XS and XH domains separated by a coiled-coil region. An XS/XH protein family has been recently defined based on homology with the rice gene *X* and the protein suppressor of gene silencing 3 (SGS3), which is involved in post-transcriptional gene silencing[11,12]. In addition to *IDN2*, there are related homologous genes in the *Arabidopsis* genome containing XS/XH domains, including *At1g15910*, *At4g00380*, *At4g01780*, *At1g13790*, *At3g12550*, *At1g80790*, *At5g59390* and *At4g01180*, and one or more of these might encode proteins that act redundantly with IDN2, which would explain why the *idn2* loss-of-function phenotype is not as strong as that of *idn1*.

The XS domain has been predicted to be an RNA-binding domain, and regions around residues Asn16 and Phe64 are conserved between XS and ribonucleoprotein domain 1 and 2 functional motifs in RNA–binding proteins[13]. Recently, it has been shown that SGS3 XS domain can bind 5′ overhanging double-stranded RNA (dsRNA) species[14]. To test the RNA-binding ability of the IDN2 protein, we performed gel mobility shift assays using a truncated version of IDN2 lacking the zinc finger and XH domain IDN2ΔZnFΔXH and several RNA species previously described[14]. Notably, IDN2ΔZnFΔXH showed specificity for dsRNA species possessing 5′ overhangs (**Supplementary Fig. 3**). Mutations at Asn16 and Phe64 abolished this binding, indicating that XS is the domain responsible for the RNA binding. Though it is unclear which double-stranded 5′ overhang substrate might be involved in RdDM, it is tempting to speculate that IDN2 could recognize an siRNA that is bound to its target noncoding RNA, facilitating the downstream targeting of chromatin factors such as DRM2 to methylation targets. Although future studies will be required to determine the precise function of both IDN1 (DMS3) and IDN2, it is clear that they are critical factors in the RdDM pathway.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

We thank J. Goodrich and D. Schubert for kindly providing T-DNA homozygous lines. We thank members of the Jacobsen laboratory for supportive discussions. Jacobsen lab research was supported by US National Institutes of Health grant GM60398. I.A. was supported by a postdoctoral fellowship from the Ministerio de Educacion y Ciencia. S.E.J. and J.C. are investigators of the Howard Hughes Medical Institute.

AUTHOR CONTRIBUTIONS

T.C.M. and J.C. provided the majority of the insertion mutagenized population. I.A. and S.E.J. designed the experiments. I.A. performed the experiments and wrote the manuscript.

Published online at http://www.nature.com/nsmb/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Chan, S.W. *et al. Science* **303**, 1336 (2004).
2. Cao, X. & Jacobsen, S.E. *Curr. Biol.* **12**, 1138–1144 (2002).
3. Henderson, I.R. & Jacobsen, S.E. *Nature* **447**, 418–424 (2007).
4. Li, C.F. *et al. Cell* **126**, 93–106 (2006).
5. Pontes, O. *et al. Cell* **126**, 79–92 (2006).
6. Wierzbicki, A.T., Haag, J.R. & Pikaard, C.S. *Cell* **135**, 635–648 (2008).
7. Soppe, W.J. *et al. Mol. Cell* **6**, 791–802 (2000).
8. Henderson, I.R. & Jacobsen, S.E. *Genes Dev.* **22**, 1597–1606 (2008).
9. Cao, X. & Jacobsen, S.E. *Proc. Natl. Acad. Sci. USA* **99** Suppl 4, 16491–16498 (2002).
10. Kanno, T. *et al. Nat. Genet.* **40**, 670–675 (2008).
11. Bateman, A. *BMC Bioinformatics* **3**, 21 (2002).
12. Mourrain, P. *et al. Cell* **101**, 533–542 (2000).
13. Zhang, D. & Trudeau, V.L. *Cell Cycle* **7**, 2268–2270 (2008).
14. Fukunaga, R. & Doudna, J.A. *EMBO J.* **28**, 545–555 (2009).

© 2009 Nature America, Inc. All rights reserved.