Reconstitution of siRNA Biogenesis In Vitro: Novel Reaction Mechanisms and RNA Channeling in the RNA-Directed DNA Methylation Pathway

Jasleen Singh[1] and Craig S. Pikaard[1,2]
[1]*Department of Molecular and Cellular Biochemistry and Department of Biology,*
[2]*Howard Hughes Medical Institute, Indiana University, Bloomington, Indiana 47405, USA*
Correspondence: cpikaard@indiana.edu

Eukaryotes deploy RNA-mediated gene silencing pathways to guard their genomes against selfish genetic elements, such as transposable elements and invading viruses. In plants, RNA-directed DNA methylation (RdDM) is used to silence selfish elements at the level of transcription. This process involves 24-nt short interfering RNAs (siRNAs) and longer noncoding RNAs to which the siRNAs base-pair. Recently, we showed that 24-nt siRNA biogenesis could be recapitulated in the test tube using purified enzymes, yielding biochemical answers to numerous questions left unresolved by prior genetic and genomic studies. Interestingly, each enzyme has activities that program what happens in the next step, thus channeling the RNAs within the RdDM pathway and restricting their diversion into alternative pathways. However, a similar mechanistic understanding is lacking for other important steps of the RdDM pathway. We discuss some of the steps most in need of biochemical investigation and important questions still in need of answers.

Small noncoding RNAs (RNAs that do not encode proteins) play important roles in restricting the proliferation of selfish genetic elements, either by base-pairing with messenger RNAs to interfere with their stability or function, thereby inhibiting protein synthesis, or by interfering with transcription to prevent RNA synthesis (Borges and Martienssen 2015; Holoch and Moazed 2015; Martienssen and Moazed 2015; Maillard et al. 2019). In the latter strategy, the small RNAs guide chemical modifications at matching chromosomal DNA sequences. These modifications include addition or removal of chemical groups on the histone proteins that wrap the DNA and, in many eukaryotes, including humans and plants, methylation of the DNA (Law and Jacobsen 2010; Martienssen and Moazed 2015). Collectively, these chemical modifications contribute to chromatin environments that are refractory to promoter-dependent transcription by DNA-dependent RNA polymerases I, II, or III.

In plants, the major transcriptional gene silencing pathway is RNA-directed DNA methylation (RdDM) (Zhou and Law 2015; Wendte and Pikaard 2017). The process has been elucidated, primarily in *Arabidopsis thaliana*, using tools of genetics, genomics, cell biology, and molecular biology. Collectively, these studies have provided a working understanding of what the pathway accomplishes, but how the enzymes and RNAs of the pathway function at a biochemical level remains unclear. Progress in illuminating these biochemical details is the subject of this perspective.

KEY ENZYMES AND RNAs OF THE RdDM PATHWAY

The central aspects of RdDM can be grasped by focusing on the RNA synthesizing and processing enzymes that are key to the pathway. Two are nuclear multisubunit RNA polymerases, abbreviated as Pol IV and Pol V (Haag and Pikaard 2011), that evolved in plants as specialized forms of DNA-dependent RNA polymerase II (Pol II) (Ream et al. 2009). The third is RNA-DEPENDENT RNA POLYMERASE 2 (RDR2) (Xie et al. 2004). Figure 1 provides a simplified view of the RdDM pathway (for a more complete view, see Wendte and Pikaard 2017), with Pol IV and RDR2 collaborating to produce double-stranded RNAs (dsRNAs) that are then cut (diced) by DICER-LIKE 3 (DCL3) (Xie et al. 2004) into short interfering RNA (siRNA) duplexes, single strands of which somehow become associated with ARGONAUTE4 (AGO4) (Zilberman et al. 2003) or related Argonaute proteins. AGO4–siRNA complexes are then recruited to sites of Pol V transcription, apparently via base-pairing between the siRNAs and nascent Pol V transcripts (Wierzbicki et al. 2008, 2009). Protein–protein interactions can also occur between AGO4 and the largest subunit of Pol V (NRPE1) (El-Shami et al. 2007) and between AGO4 and the Pol V–associated protein, SUPPRESSOR of TY INSERTION 5-LIKE (SPT5L) (Lahmy et al. 2016). AGO–siRNA–Pol V complexes then facilitate recruitment of chromatin-modifying enzymes, including the de novo DNA methyltrans-

© 2019 Singh and Pikaard. This article is distributed under the terms of the Creative Commons Attribution-NonCommercial License, which permits reuse and redistribution, except for commercial purposes, provided that the original author and source are credited.

Published by Cold Spring Harbor Laboratory Press; doi:10.1101/sqb.2019.84.039842
Cold Spring Harbor Symposia on Quantitative Biology, Volume LXXXIV

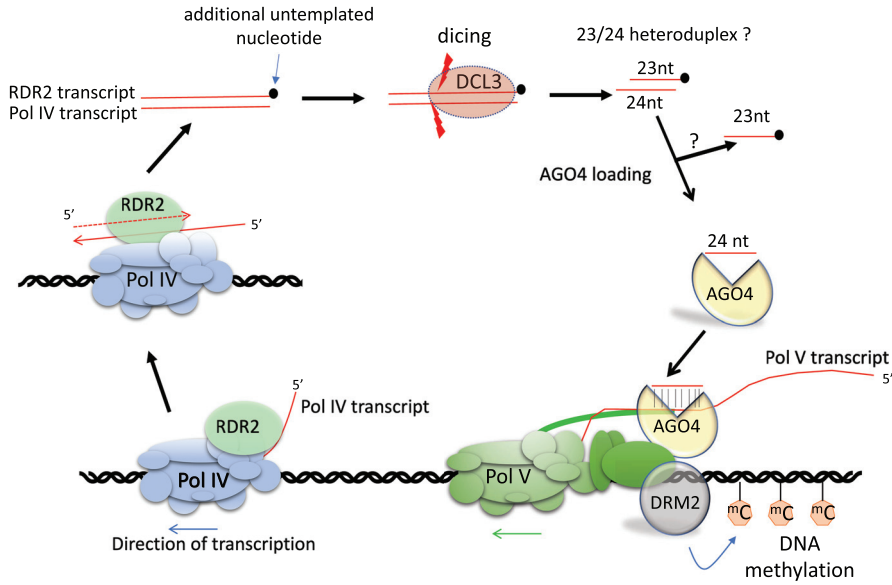

Figure 1. Key events of the RNA-directed DNA methylation pathway. Pol IV and RDR2 physically associate and their reactions are coupled to produce double-stranded RNAs (dsRNAs). These dsRNAs include a nontemplated nucleotide at the 3′ end of the RDR2 strands attributable to RDR2's terminal transferase activity. Upon dicing by DICER-LIKE 3 (DCL3), 24- and 23-nt siRNAs are generated, possibly as a heteroduplex (hence the question mark), with 23-nt RNAs enriched for the untemplated 3′ nucleotides. We hypothesize that the 23-nt siRNAs serve as the passenger strands for the associated 24-nt guide RNAs that are loaded into ARGONAUTE 4 (AGO4). Resulting siRNA–AGO4 complexes find their target sites by base-pairing to Pol V transcripts and interacting with the carboxy-terminal domain (CTD) of the Pol V largest subunit. The cytosine methyltransferase DRM2 is ultimately recruited and carries out regional de novo methylation of cytosines in any sequence context.

ferase, DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) (Cao and Jacobsen 2002), the plant ortholog of the mammalian enzymes, DNMT3a and DNMT3b (Law and Jacobsen 2010). Extensive methylation of local deoxycytosines ensues, in all sequence contexts (CG, CHG, and CHH, in which H represents any nucleotide other than G). Repressive histone modifications also occur, with genetic evidence implicating histone deacetylation, catalyzed by HISTONE DEACETYLASE 6 (HDA6) (Aufsatz et al. 2002; He et al. 2009), histone H3 lysine 9 dimethylation (H3K9me2) catalyzed by suppressor of variegation homologs (SUVH), SUVH4, SUVH5, and SUVH6 (Jackson et al. 2004; Ebbs and Bender 2006; Johnson et al. 2007; Blevins et al. 2014), and histone H3 lysine 4 (H3K4) demethylation catalyzed by the Jumonji C (JmjC) domain–containing protein JMJ14 (Deleris et al. 2010; Searle et al. 2010) and nucleosome repositioning by the SWI/SNF complex (Zhu et al. 2013). Collectively, the DNA and histone modifications contribute to a chromatin environment refractory to promoter-dependent transcription of transposon or viral genes.

IN VIVO STUDIES AND UNANSWERED QUESTIONS

Genetic studies showed that mutants defective for Pol IV, RDR2, or DCL3 lack the 24-nt class of siRNAs associated with DNA methylation and gene silencing (Xie et al. 2004; Herr et al. 2005; Onodera et al. 2005; Henderson et al. 2006; Kasschau et al. 2007). Based on the presumptive functions of these enzymes, deduced from similarities to related proteins, Pol IV and RDR2 were hypothesized to generate RNA precursors that DCL3 then diced into 24-nt siRNAs. It seemed logical that Pol IV, as a presumptive DNA-dependent RNA polymerase, might act first, with RDR2 then using the Pol IV transcripts as templates to generate dsRNAs. Cell biological studies provided indirect support for this hypothesis by showing that RDR2 becomes mislocalized in nuclei of cells lacking Pol IV (Pontes et al. 2006). In contrast, Pol IV immunolocalization was largely unchanged in *rdr2* mutants. However, Pol IV was mislocalized in nuclei treated with ribonuclease A, suggesting that Pol IV might transcribe an RNA template (Pontes et al. 2006); thus, the possibility that RDR2 might act prior to Pol IV could not be ruled out.

Pol IV and RDR2 transcripts proved difficult to identify, remaining elusive until 2015, when several laboratories independently examined RNAs that accumulate in Dicer mutants, as expected for siRNA precursors (Blevins et al. 2015; Li et al. 2015; Zhai et al. 2015). The RNAs were shown to be dependent on both Pol IV and RDR2 for their synthesis, to be sensitive to double-strand specific ribonucleases, to accumulate in *dcl3* mutants but be diced into 24-nt siRNAs by DCL3 in vitro (Blevins et al. 2015), and to map to 24-nt siRNA producing loci. Surprisingly, the siRNA precursors were mostly 25–35 nt in length, so short that they could only be diced once (Blevins et al. 2015; Zhai et al. 2015).

A priori, one might predict Pol IV transcripts to accumulate in a *rdr2* mutant if Pol IV acts first, and RDR2 then converts Pol IV transcripts into dsRNAs. Conversely, if

RDR2 acts first, RDR2 transcripts would be expected to accumulate in a *pol iv* mutant. Surprisingly, in single mutants defective for either Pol IV or RDR2, no precursor transcripts were detected (Blevins et al. 2015; Zhai et al. 2015). Could it be that single-stranded transcripts are simply turned over rapidly, thus evading detection? Or might the physical association of Pol IV and RDR2, shown by previous coimmunoprecipitation experiments (Law et al. 2011; Haag et al. 2012), be needed for each enzyme to be active? Initial in vitro transcription assays of Haag et al. (2012) suggested that this might be true for RDR2, but not Pol IV. Or could it be that Pol IV and RDR2 are both needed to synthesize each strand of a dsRNA precursor, perhaps by RDR2 carrying out a priming step and Pol IV a subsequent elongation step, reminiscent of Okazaki fragment synthesis during DNA replication (Bergsch et al. 2019)?

Another puzzling characteristic of Pol IV and RDR2-dependent precursor RNAs, which we refer to as P4R2 RNAs (Blevins et al. 2015), is that they frequently have a 3′-terminal nucleotide that does not match the corresponding genomic DNA (Blevins et al. 2015; Zhai et al. 2015). Whether these mismatches occur at the ends of Pol IV or RDR2 transcripts or both was unclear because there was no definitive way to distinguish Pol IV transcripts from RDR2 transcripts in vivo. One hypothesis suggested that mismatched nucleotides result from misincorporation errors when Pol IV encounters methylated cytosines in the DNA template (Zhai et al. 2015). It was further hypothesized that if the mistake could not be repaired, transcription termination might result. The latter idea, combined with evidence that Pol IV is recruited to regions displaying CG maintenance methylation (Blevins et al. 2014), provided a possible explanation for why Pol IV transcripts are so short.

The Pol IV-misincorporation hypothesis was tested in subsequent biochemical experiments in vitro. Pol IV transcription is relatively error-prone compared to Pol II and Pol V, but there was no difference in nucleotide misincorporation at methylcytosines versus cytosines (Marasco et al. 2017). Cytosine methylation also does not induce Pol IV termination in vitro (Marasco et al. 2017; Singh et al. 2019).

An alternative explanation for template-mismatched nucleotides in P4R2 RNAs stems from the fact that RDR2 displays terminal nucleotidyl transferase activity that can add one or more untemplated nucleotides to the 3′ end of an RNA in vitro (Blevins et al. 2015). But can RDR2 add an untemplated nucleotide to the 3′ ends of Pol IV transcripts, its own transcripts, or both? Once again, because there was no way to distinguish Pol IV from RDR2 transcripts among P4R2 RNAs in vivo, an answer to this question awaited assays able to distinguish their transcripts in vitro.

NEW BIOCHEMICAL ASSAYS PROVIDE IN VITRO ANSWERS TO THE IN VIVO QUESTIONS

As introduced above, numerous questions concerning Pol IV and RDR2 transcription remained unanswered from in vivo studies. Among these were the following: (1) Which enzyme acts first in the biogenesis of siRNA precursors? (2) Does Pol IV make one strand of each dsRNA precursor and RDR2 the other strand, or are both enzymes needed to synthesize each strand? (3) Is there a way to distinguish Pol IV transcripts from RDR2 transcripts? (4) Are 3′-untemplated nucleotides present at the ends of both Pol IV and RDR2 transcripts? (5) Is there any functional significance to these untemplated nucleotides? (6) Why are P4R2 transcripts so short?

Answers to the first four questions came from a biochemical assay we devised using single-stranded bacteriophage M13 genomic DNA as a source of template for dsRNA synthesis by Pol IV and RDR2 (Singh et al. 2019). Pol IV and RDR2 associate and thus copurify from wild-type plants (Law et al. 2011; Haag et al. 2012). However, epitope-tagged Pol IV can be isolated from a *rdr2* null mutant, and epitope-tagged RDR2 can be purified from a *nrpd1* null mutant, lacking the Pol IV largest subunit, so as to be isolated individually (Haag et al. 2012). We found that dsRNAs were synthesized by Pol IV–RDR2 complexes but not by either enzyme individually (Singh et al. 2019). RNA sequencing showed that Pol IV transcripts are exclusively reverse complements of the DNA. Only if RDR2 is also present are RNA transcripts generated in the same 5′→3′ polarity as the template DNA; these RNAs can only be generated by using Pol IV transcripts as templates. The results showed definitively that each enzyme is wholly responsible for synthesizing individual strands of the dsRNA (Singh et al. 2019). Moreover, the phenomenon of template-mismatched nucleotides present at the 3′ ends of P4R2 RNAs in vivo was recapitulated (Blevins et al. 2015; Singh et al. 2019). Importantly, these untemplated nucleotides occur at the 3′ termini of RDR2 transcripts, not Pol IV transcripts. This indicates that RDR2 uses its terminal transferase activity to add an extra nucleotide to the end of its own transcripts but does not act on Pol IV transcripts.

Terminal transferase activity has been shown for a number of RNA-dependent RNA polymerases, including *Arabidopsis* RDR6, *Neurospora crassa* QDE1, and viral polymerases (Ranjith-Kumar et al. 2001; Curaba and Chen 2008; Poranen et al. 2008; Aalto et al. 2010). But what is the significance of this activity, if any? Our findings using the M13 system suggest that RDR2's terminal transferase activity serves a purpose—in specifying the fates of the two RNA strands following dicing. Reactions that included DCL3, in addition to Pol IV and RDR2, generated both 24- and 23-nt siRNAs in vitro (Singh et al. 2019), as is also the case in vivo (Kasschau et al. 2007). The significance of the 23-nt class of siRNAs has never been clear. Interestingly, we found that the template-mismatched 3′ nucleotides attributable to RDR2 terminal transferase activity are primarily found at the ends of 23-nt siRNAs (Singh et al. 2019). This in vitro result fits with analyses that showed that template-mismatched 3′-terminal nucleotides are also enriched among 23-nt siRNAs in vivo (Wang et al. 2016). Based on these observations, we hypothesize that DCL3 interacts with dsRNAs, measures 24 nt from the 5′ end of the Pol IV strand, and makes a staggered cut that leaves

the 3′ end of the Pol IV strand overhanging the RDR2 strand by 2 nt, consistent with cutting by other Dicers (Park et al. 2011). Because the RDR2 strand has a 1-nt extension at its 3′ end as a result of RDR2's terminal transferase activity, DCL3 cutting generates a 23-nt RNA from the RDR2 strand, with the template-mismatched nucleotide at its 3′ end (Fig. 1). Because 23-nt siRNAs are not abundant among the RNAs that copurify with immunoprecipitated AGO4 (Qi et al. 2006; Havecker et al. 2010; Wang et al. 2016), we hypothesize that 23-nt RNAs may serve as so-called "passenger strands" for the 24-nt guide strands that become stably associated with AGO4 (Singh et al. 2019). If so, 23-nt siRNAs have a function, and so does the terminal transferase activity of RDR2 that contributes to their biogenesis.

An answer to the question of why P4R2 RNAs in vivo are so short came from a different set of in vitro experiments using synthetic oligo- and polynucleotide DNAs. In these experiments, we found that Pol IV engaged in transcription of a single-stranded DNA template DNA strand terminates ∼12–16 nt after encountering the base-paired nontemplate DNA strand (Singh et al. 2019), in a sequence-independent manner. This suggests a model in which Pol IV initiates on one DNA strand following the melting of duplex DNA to form a transcription bubble, like other DNA-dependent RNA polymerases (Holstege et al. 1997; Bae et al. 2015; Barnes et al. 2015). After transcribing single-stranded DNA within the bubble, Pol IV encounters double-stranded DNA at the edge of the bubble and is only able to transcribe an additional 12–16 nt before terminating. Exactly why Pol IV cannot transcribe further than 12–16 nt into duplex DNA is unclear. However, amino acid changes in several key regions of Pol IV's catalytic center offer some clues (Haag et al. 2009, 2012). In prokaryotic and eukaryotic DNA-dependent RNA polymerases, structural features in their largest subunits known as the "trigger," "rudder," and "zipper" loops affect catalytic activity, melting of the DNA strands ahead of the advancing polymerase, and reannealing of the DNA strands in the wake of the passing polymerase, respectively (Toulokhonov and Landick 2006). The largest subunit of Pol IV (NRPD1) contains amino acid deletions in each of these elements (Haag et al. 2009, 2012; Singh et al. 2019). Collectively, these deletions may hamper Pol IV's ability to plow ahead into duplex DNA and propagate a transcription bubble that translocates along with the polymerase, allowing for processivity.

Does RDR2 Interact with Pol IV Transcripts Prior to Pol IV Termination?

Arabidopsis has numerous small RNA-mediated pathways, involving six RNA-dependent RNA polymerases, four Dicers, and 10 Argonaute proteins (Volpe and Martienssen 2011; Borges and Martienssen 2015). With so many alternative enzymes in play, how can RNAs stay channeled within a specific pathway? Our findings suggest that protein–protein interactions and RNA signals, imparted by the enzymes that synthesize or process the RNAs, account for RNA channeling. Pol IV physically associates with RDR2 (Law et al. 2011; Haag et al. 2012) and the interaction is insensitive to RNase A treatment, suggesting that RNA does not link them (Pontes et al. 2006). Therefore, protein–protein interactions are likely, involving one or more Pol IV subunits that interact directly with RDR2 or indirectly via one or more bridging proteins. In *Arabidopsis*, Pol IV's largest subunit, NRPD1, and seventh-largest subunit, NRPD7, are the only subunits not shared by Pols I, II, III, or V (Ream et al. 2009), making them potential candidates for RDR2 interactors. In maize, however, in which Pol IV and RDR2 (MOP1 in maize) also copurify, the only subunit unique to Pol IV is NRPD1 (Haag et al. 2014). Thus, NRPD1 is the prime suspect for mediating the interaction with RDR2 interaction.

RDR2 initiation is coupled to Pol IV termination and results in RDR2 transcripts that mirror the length of Pol IV transcripts in the M13 system (Singh et al. 2019), suggesting that RDR2 transcribes the Pol IV transcripts end-to-end. However, it remains unclear when RDR2 first engages the Pol IV transcript. One possibility is that RDR2 interacts with the nascent Pol IV transcript as soon as the 5′ end of the transcript emerges from Pol IV's RNA exit channel such that the RNA is already engaged by RDR2 prior to Pol IV termination and transcript release. Alternatively, RDR2 may only capture Pol IV transcripts after their release. Additional biochemical tests, such as protein cross-linking experiments and/or structural studies of Pol IV–RDR2 complexes, will be needed to determine how Pol IV and RDR2 interact and how RNAs are channeled from Pol IV to RDR2.

How Is Pol IV Transcription Initiated?

Genetic evidence indicates that sites of Pol IV transcription are not specified by conventional promoters to which transcription factors bind. Instead, the evidence suggests that chromatin modifications are the signals that facilitate Pol IV recruitment (Law et al. 2013; Blevins et al. 2014). These include cytosine methylation marks that can be maintained following every round of DNA replication by DNA METHYLTRANSFERASE 1 (MET1), the plant ortholog of mammalian DNMT1. MET1 maintenance methylation at CG motifs somehow requires the activity of HDA6 (To et al. 2011; Liu et al. 2012; Blevins et al. 2014), as does maintenance methylation at CHG motifs by CHROMOMETHYLTRANSFERASE 3 (CMT3) (Earley et al. 2010). Nucleosomes assembled using DNA methylated by these enzymes include histone H3 proteins that are dimethylated on lysine 9 (H3K9me2). SAWADEE HOMEODOMAIN HOMOLOG 1 (SHH1; also known as DTF1), a protein with TUDOR-like domains that can associate with *Arabidopsis* Pol IV, can bind these H3K9me2 marks (Law et al. 2013; Zhang et al. 2013). Interestingly, in maize, the paralogous proteins, SHH2a and SHH2b, associate with both Pol IV and Pol V (Haag et al. 2014). Like Pol IV, Pol V depends on MET1 to specify its sites of action (Johnson et al. 2014), somehow facilitated by methylcytosine binding by SUVH2 and SUVH9 (Liu et al. 2014; Jing et al. 2016); thus, both

polymerases may have similar epigenetically inherited recruitment signals. But how recruitment leads to Pol IV and Pol V being able to engage and transcribe one strand of DNA wrapped around a histone octamer is unknown. Pol IV–dependent synthesis of 24-nt siRNAs requires one or more members of a family of putative ATP-dependent DNA translocases, CLSY1–CLSY4 (Smith et al. 2007). CLSY1 has been shown to physically interact with SHH1/DTF1 (Zhang et al. 2013) and has been proposed to act as a chromatin remodeler that might alter DNA–histone contacts. However, no biochemical assays of CLSY activity have been reported. Importantly, ATP-dependent DNA translocases can have diverse functions, including helicase activity. This raises the intriguing possibility that CLSY proteins could potentially be involved in the transcription bubble opening to allow Pol IV transcription initiation. Likewise, the putative ATP-dependent DNA translocase, DRD1 (Kanno et al. 2005), is required for Pol V transcription (Wierzbicki et al. 2008; Law et al. 2010; Zhong et al. 2012) and may play a similar role (Pikaard et al. 2012). Clearly, biochemical assays need to be developed to understand how Pol IV and Pol V can initiate transcription in the context of chromatin.

How Does Nontemplate DNA Induce Pol IV Termination?

Pol IV terminates shortly after encountering double-stranded DNA and termination in this manner is somehow required to enable RDR2 to synthesize the second RNA strand (Singh et al. 2019). The biochemical basis for these molecular events is unclear. As discussed previously, amino acid deletions in the Pol IV "trigger," "rudder," and "zipper" loops in the vicinity of the catalytic center may contribute to transcription stalling and termination (for sequence comparisons, see Toulokhonov and Landick 2006; Landick 2009; and Singh et al. 2019). But how termination results in a nucleic acid conformation that enables RDR2 to engage the 3′ end of the released Pol IV transcript and use it as a template is difficult to envision. Structural studies may be necessary to come to a mechanistic understanding of this critical step.

Do Patterns of DCL3 Dicing Program AGO4–siRNA Complex Formation?

AGO4–siRNA complexes immunoprecipitated from *Arabidopsis* contain 24-nt siRNAs, almost exclusively, despite the fact that DCL3 generates both 23- and 24-nt siRNAs (Qi et al. 2006; Kasschau et al. 2007; Havecker et al. 2010; Wang et al. 2016; Singh et al. 2019). As discussed previously, we hypothesize that 23-nt RNAs may serve as passenger strands that help specify the stable association of AGO4 with the 24-nt RNAs base-paired with the 23-nt RNAs (Singh et al. 2019). However, it is not yet clear if 24/23-nt heteroduplexes exist in cells. Likewise, we know little about how AGO4 becomes loaded with siRNAs. In vitro assays that could recapitulate AGO4 loading and siRNA strand choice would provide a means to address these questions.

Are There DCL-Independent Routes of siRNA Biogenesis?

Recent studies suggested that there are DCL-independent routes of siRNA biogenesis, explaining residual levels of DNA methylation in Dicer mutants (Yang et al. 2016; Ye et al. 2016). Briefly, the idea is that Pol IV nascent transcripts might be loaded into AGO4 as single strands that are then trimmed down to 24 nt by the putative 3′ to 5′ exonuclease activity of RRP6L1 or an uncharacterized "trimmer" enzyme. However, using the M13 system in vitro, a significant fraction of Pol IV and RDR2 transcripts happen to be 24 nt in size (Singh et al. 2019). Might these transcripts be capable of being loaded into AGO4? If so, this would be a Dicer- and Trimmer-independent way of producing 24-nt RNAs. Moreover, it is noteworthy that the proposed role for RRP6L1 as an siRNA Trimmer is hypothetical, stemming from genetic evidence that it is involved for RdDM. But there is an alternative explanation for RRP6L1 involvement in RdDM. RRP6L1 interacts with the carboxy-terminal domain of Pol V in yeast two-hybrid assays, and in *rrp6l1* mutants, Pol V transcripts are longer than in wild type (Wendte et al. 2017). Pol V transcripts are similarly longer in *ago4* mutants, suggesting that what RRP6L1 trims are Pol V transcripts sliced by AGO4 (Wendte et al. 2017).

CONCLUSION

Two decades of in vivo studies have identified the players in the RdDM pathway and what happens when they are mutated. However, we are still in the early stages of understanding the biochemical functions of many of these proteins. Having reconstituted the siRNA biogenesis steps of the RdDM pathway in vitro, we can envision a future in which we can load AGO4 with siRNAs of our choosing and have these AGO4–siRNA complexes associate with Pol V engaged in transcription. By considering RdDM pathway as a series of reactions that can be understood in biochemical detail, we are optimistic that RdDM can ultimately be reconstituted in vitro. If so, we may finally understand how the siRNAs, long noncoding RNAs, DNA strands, and proteins are configured to enable the chemical modification of genomic target sites, an understanding not yet in hand for RNA-directed chromatin modification pathways of any organism.

ACKNOWLEDGMENTS

Our work is supported by the National Institutes of Health grant GM077590 and funds to C.S.P. as an Investigator of the Howard Hughes Medical Institute. J.S. was supported, in part, by a Carlos O. Miller fellowship (Indiana University).

REFERENCES

Aalto AP, Poranen MM, Grimes JM, Stuart DI, Bamford DH. 2010. *In vitro* activities of the multifunctional RNA silencing polymerase QDE-1 of *Neurospora crassa*. *J Biol Chem* **285:** 29367–29374. doi:10.1074/jbc.M110.139121

Aufsatz W, Mette MF, Van Der Winden J, Matzke M, Matzke AJ. 2002. HDA6, a putative histone deacetylase needed to enhance DNA methylation induced by double-stranded RNA. *EMBO J* **21:** 6832–6841. doi:10.1093/emboj/cdf663

Bae B, Feklistov A, Lass-Napiorkowska A, Landick R, Darst SA. 2015. Structure of a bacterial RNA polymerase holoenzyme open promoter complex. *eLife* **4:** e08504. doi:10.7554/eLife.08504

Barnes CO, Calero M, Malik I, Graham BW, Spahr H, Lin G, Cohen AE, Brown IS, Zhang Q, Pullara F, et al. 2015. Crystal structure of a transcribing RNA polymerase II complex reveals a complete transcription bubble. *Mol Cell* **59:** 258–269. doi:10.1016/j.molcel.2015.06.034

Bergsch J, Allain FH-T, Lipps G. 2019. Recent advances in understanding bacterial and archaeoeukaryotic primases. *Curr Opin Struct Biol* **59:** 159–167. doi:10.1016/j.sbi.2019.08.004

Blevins T, Pontvianne F, Cocklin R, Podicheti R, Chandrasekhara C, Yerneni S, Braun C, Lee B, Rusch D, Mockaitis K, et al. 2014. A two-step process for epigenetic inheritance in *Arabidopsis*. *Mol Cell* **54:** 30–42. doi:10.1016/j.molcel.2014.02.019

Blevins T, Podicheti R, Mishra V, Marasco M, Tang H, Pikaard CS. 2015. Identification of Pol IV and RDR2-dependent precursors of 24 nt siRNAs guiding de novo DNA methylation in *Arabidopsis*. *eLife* **4:** e09591. doi:10.7554/eLife.09591

Borges F, Martienssen RA. 2015. The expanding world of small RNAs in plants. *Nat Rev Mol Cell Biol* **16:** 727–741. doi:10.1038/nrm4085

Cao X, Jacobsen SE. 2002. Role of the *Arabidopsis* DRM methyltransferases in de novo DNA methylation and gene silencing. *Curr Biol* **12:** 1138–1144. doi:10.1016/S0960-9822(02)00925-9

Curaba J, Chen X. 2008. Biochemical activities of *Arabidopsis* RNA-dependent RNA polymerase 6. *J Biol Chem* **283:** 3059–3066. doi:10.1074/jbc.M708983200

Deleris A, Greenberg MVC, Ausin I, Law RWY, Moissiard G, Schubert D, Jacobsen SE. 2010. Involvement of a Jumonji-C domain–containing histone demethylase in DRM2-mediated maintenance of DNA methylation. *EMBO Rep* **11:** 950–955. doi:10.1038/embor.2010.158

Earley KW, Pontvianne F, Wierzbicki AT, Blevins T, Tucker S, Costa-Nunes P, Pontes O, Pikaard CS. 2010. Mechanisms of HDA6-mediated rRNA gene silencing: suppression of intergenic Pol II transcription and differential effects on maintenance versus siRNA-directed cytosine methylation. *Genes Dev* **24:** 1119–1132. doi:10.1101/gad.1914110

Ebbs ML, Bender J. 2006. Locus-specific control of DNA methylation by the *Arabidopsis* SUVH5 histone methyltransferase. *Plant Cell* **18:** 1166–1176. doi:10.1105/tpc.106.041400

El-Shami M, Pontier D, Lahmy S, Braun L, Picart C, Vega D, Hakimi M-A, Jacobsen SE, Cooke R, Lagrange T. 2007. Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. *Genes Dev* **21:** 2539–2544. doi:10.1101/gad.451207

Haag JR, Pikaard CS. 2011. Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. *Nat Rev Mol Cell Biol* **12:** 483–492. doi:10.1038/nrm3152

Haag JR, Pontes O, Pikaard CS. 2009. Metal A and metal B sites of nuclear RNA polymerases Pol IV and Pol V are required for siRNA-dependent DNA methylation and gene silencing. *PLoS One* **4:** e4110. doi:10.1371/journal.pone.0004110

Haag JR, Ream TS, Marasco M, Nicora CD, Norbeck AD, Pasa-Tolic L, Pikaard CS. 2012. 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* **48:** 811–818. doi:10.1016/j.molcel.2012.09.027

Haag JR, Brower-Toland B, Krieger EK, Sidorenko L, Nicora CD, Norbeck AD, Irsigler A, LaRue H, Brzeski J, McGinnis K, et al. 2014. Functional diversification of maize RNA polymerase IV and V subtypes via alternative catalytic subunits. *Cell Rep* **9:** 378–390. doi:10.1016/j.celrep.2014.08.067

Havecker ER, Wallbridge LM, Hardcastle TJ, Bush MS, Kelly KA, Dunn RM, Schwach F, Doonan JH, Baulcombe DC. 2010. The *Arabidopsis* RNA-directed DNA methylation Argonautes functionally diverge based on their expression and interaction with target loci. *Plant Cell* **22:** 321–334. doi:10.1105/tpc.109.072199

He X-J, Hsu Y-F, Pontes O, Zhu J, Lu J, Bressan RA, Pikaard C, Wang C-S, Zhu J-K. 2009. NRPD4, a protein related to the RPB4 subunit of RNA polymerase II, is a component of RNA polymerases IV and V and is required for RNA-directed DNA methylation. *Genes Dev* **23:** 318–330. doi:10.1101/gad.1765209

Henderson IR, Zhang X, Lu C, Johnson L, Meyers BC, Green PJ, Jacobsen SE. 2006. Dissecting *Arabidopsis thaliana* DICER function in small RNA processing, gene silencing and DNA methylation patterning. *Nat Genet* **38:** 721–725. doi:10.1038/ng1804

Herr AJ, Jensen MB, Dalmay T, Baulcombe DC. 2005. RNA polymerase IV directs silencing of endogenous DNA. *Science* **308:** 118–120. doi:10.1126/science.1106910

Holoch D, Moazed D. 2015. RNA-mediated epigenetic regulation of gene expression. *Nat Rev Genet* **16:** 71–84. doi:10.1038/nrg3863

Holstege FCP, Fiedler U, Timmers HT. 1997. Three transitions in the RNA polymerase II transcription complex during initiation. *EMBO J* **16:** 7468–7480. doi:10.1093/emboj/16.24.7468

Jackson JP, Johnson L, Jasencakova Z, Zhang X, PerezBurgos L, Singh PB, Cheng X, Schubert I, Jenuwein T, Jacobsen SE. 2004. Dimethylation of histone H3 lysine 9 is a critical mark for DNA methylation and gene silencing in *Arabidopsis thaliana*. *Chromosoma* **112:** 308–315. doi:10.1007/s00412-004-0275-7

Jing Y, Sun H, Yuan W, Wang Y, Li Q, Liu Y, Li Y, Qian W. 2016. SUVH2 and SUVH9 couple two essential steps for transcriptional gene silencing in *Arabidopsis*. *Mol Plant* **9:** 1156–1167. doi:10.1016/j.molp.2016.05.006

Johnson LM, Bostick M, Zhang X, Kraft E, Henderson I, Callis J, Jacobsen SE. 2007. The SRA methyl-cytosine-binding domain links DNA and histone methylation. *Curr Biol* **17:** 379–384. doi:10.1016/j.cub.2007.01.009

Johnson LM, Du J, Hale CJ, Bischof S, Feng S, Chodavarapu RK, Zhong X, Marson G, Pellegrini M, Segal DJ, et al. 2014. SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. *Nature* **507:** 124–128. doi:10.1038/nature12931

Kanno T, Aufsatz W, Jaligot E, Mette MF, Matzke M, Matzke AJ. 2005. A SNF2-like protein facilitates dynamic control of DNA methylation. *EMBO Rep* **6:** 649–655. doi:10.1038/sj.embor.7400446

Kasschau KD, Fahlgren N, Chapman EJ, Sullivan CM, Cumbie JS, Givan SA, Carrington JC. 2007. Genome-wide profiling and analysis of *Arabidopsis* siRNAs. *PLoS Biol* **5:** e57. doi:10.1371/journal.pbio.0050057

Lahmy S, Pontier D, Bies-Etheve N, Laudie M, Feng S, Jobet E, Hale CJ, Cooke R, Hakimi MA, Angelov D, et al. 2016. Evidence for ARGONAUTE4–DNA interactions in RNA-directed DNA methylation in plants. *Genes Dev* **30:** 2565–2570. doi:10.1101/gad.289553.116

Landick R. 2009. Functional divergence in the growing family of RNA polymerases. *Structure* **17:** 323–325. doi:10.1016/j.str.2009.02.006

Law JA, Jacobsen SE. 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* **11:** 204–220. doi:10.1038/nrg2719

Law JA, Ausin I, Johnson LM, Vashisht AA, Zhu J-K, Wohlschlegel JA, Jacobsen SE. 2010. A protein complex required for polymerase V transcripts and RNA- directed DNA methylation in *Arabidopsis*. *Curr Biol* **20:** 951–956. doi:10.1016/j.cub.2010.03.062

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

Law JA, Du J, Hale CJ, Feng S, Krajewski K, Palanca AM, Strahl BD, Patel DJ, Jacobsen SE. 2013. Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. *Nature* **498:** 385–389. doi:10.1038/nature12178

Li S, Vandivier LE, Tu B, Gao L, Won SY, Li S, Zheng B, Gregory BD, Chen X. 2015. Detection of Pol IV/RDR2-dependent transcripts at the genomic scale in *Arabidopsis* reveals features and regulation of siRNA biogenesis. *Genome Res* **25:** 235–245. doi:10.1101/gr.182238.114

Liu X, Yu C-W, Duan J, Luo M, Wang K, Tian G, Cui Y, Wu K. 2012. HDA6 directly interacts with DNA methyltransferase MET1 and maintains transposable element silencing in *Arabidopsis*. *Plant Physiol* **158:** 119–129. doi:10.1104/pp.111.184275

Liu Z-W, Shao C-R, Zhang C-J, Zhou J-X, Zhang S-W, Li L, Chen S, Huang H-W, Cai T, He X-J. 2014. The SET domain proteins SUVH2 and SUVH9 are required for Pol V occupancy at RNA-directed DNA methylation loci. *PLoS Genet* **10:** e1003948. doi:10.1371/journal.pgen.1003948

Maillard PV, van der Veen AG, Poirier EZ, Reis ESC. 2019. Slicing and dicing viruses: antiviral RNA interference in mammals. *EMBO J* **38**, e100941. doi:10.15252/embj.2018100941

Marasco M, Li W, Lynch M, Pikaard CS. 2017. Catalytic properties of RNA polymerases IV and V: accuracy, nucleotide incorporation and rNTP/dNTP discrimination. *Nucleic Acids Res* **45:** 11315–11326. doi:10.1093/nar/gkx794

Martienssen R, Moazed D. 2015. RNAi and heterochromatin assembly. *Cold Spring Harb Perspect Biol* **7:** a019323. doi:10.1101/cshperspect.a019323

Onodera Y, Haag JR, Ream T, Costa Nunes P, Pontes O, Pikaard CS. 2005. Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* **120:** 613–622. doi:10.1016/j.cell.2005.02.007

Park J-E, Heo I, Tian Y, Simanshu DK, Chang H, Jee D, Patel DJ, Kim VN. 2011. Dicer recognizes the 5′ end of RNA for efficient and accurate processing. *Nature* **475:** 201–205. doi:10.1038/nature10198

Pikaard CS, Haag JR, Pontes OM, Blevins T, Cocklin R. 2012. A transcription fork model for Pol IV and Pol V–dependent RNA-directed DNA methylation. *Cold Spring Harb Symp Quant Biol* **77:** 205–212. doi:10.1101/sqb.2013.77.014803

Pontes O, Li CF, Costa Nunes P, Haag J, Ream T, Vitins A, Jacobsen SE, Pikaard CS. 2006. The *Arabidopsis* chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center. *Cell* **126:** 79–92. doi:10.1016/j.cell.2006.05.031

Poranen MM, Koivunen MR, Bamford DH. 2008. Nontemplated terminal nucleotidyltransferase activity of double-stranded RNA bacteriophage φ6 RNA-dependent RNA polymerase. *J Virol* **82:** 9254–9264. doi:10.1128/JVI.01044-08

Qi Y, He X, Wang X-J, Kohany O, Jurka J, Hannon GJ. 2006. Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature* **443:** 1008–1012. doi:10.1038/nature05198

Ranjith-Kumar CT, Gajewski J, Gutshall L, Maley D, Sarisky RT, Kao CC. 2001. Terminal nucleotidyl transferase activity of recombinant Flaviviridae RNA-dependent RNA polymerases: implication for viral RNA synthesis. *J Virol* **75:** 8615–8623. doi:10.1128/JVI.75.18.8615-8623.2001

Ream TS, Haag JR, Wierzbicki AT, Nicora CD, Norbeck AD, Zhu J-K, Hagen G, Guilfoyle TJ, Paša-Tolić L, Pikaard CS. 2009. Subunit compositions of the RNA-silencing enzymes Pol IV and Pol V reveal their origins as specialized forms of RNA polymerase II. *Mol Cell* **33:** 192–203. doi:10.1016/j.molcel.2008.12.015

Searle IR, Pontes O, Melnyk CW, Smith LM, Baulcombe DC. 2010. JMJ14, a JmjC domain protein, is required for RNA silencing and cell-to-cell movement of an RNA silencing signal in *Arabidopsis*. *Genes Dev* **24:** 986–991. doi:10.1101/gad.579910

Singh J, Mishra V, Wang F, Huang HY, Pikaard CS. 2019. Reaction mechanisms of Pol IV, RDR2, and DCL3 drive RNA channeling in the siRNA-directed DNA methylation pathway. *Mol Cell* **75:** 576–589.e5. doi:10.1016/j.molcel.2019.07.008

Smith LM, Pontes O, Searle I, Yelina N, Yousafzai FK, Herr AJ, Pikaard CS, Baulcombe DC. 2007. An SNF2 protein associated with nuclear RNA silencing and the spread of a silencing signal between cells in *Arabidopsis*. *Plant Cell* **19:** 1507–1521. doi:10.1105/tpc.107.051540

To TK, Kim J-M, Matsui A, Kurihara Y, Morosawa T, Ishida J, Tanaka M, Endo T, Kakutani T, Toyoda T, et al. 2011. *Arabidopsis* HDA6 regulates locus-directed heterochromatin silencing in cooperation with MET1. *PLoS Genet* **7:** e1002055. doi:10.1371/journal.pgen.1002055

Toulokhonov I, Landick R. 2006. The role of the lid element in transcription by *E. coli* RNA polymerase. *J Mol Biol* **361:** 644–658. doi:10.1016/j.jmb.2006.06.071

Volpe T, Martienssen RA. 2011. RNA interference and heterochromatin assembly. *Cold Spring Harb Perspect Biol* **3:** a003731. doi:10.1101/cshperspect.a003731

Wang F, Johnson NR, Coruh C, Axtell MJ. 2016. Genome-wide analysis of single non-templated nucleotides in plant endogenous siRNAs and miRNAs. *Nucleic Acids Res* **44:** 7395–7405. doi:10.1093/nar/gkw457

Wendte JM, Pikaard CS. 2017. The RNAs of RNA-directed DNA methylation. *Biochim Biophys Acta* **1860:** 140–148. doi:10.1016/j.bbagrm.2016.08.004

Wendte JM, Haag JR, Singh J, McKinlay A, Pontes OM, Pikaard CS. 2017. Functional dissection of the Pol V largest subunit CTD in RNA-directed DNA methylation. *Cell Rep* **19:** 2796–2808. doi:10.1016/j.celrep.2017.05.091

Wierzbicki AT, Haag JR, Pikaard CS. 2008. Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell* **135:** 635–648. doi:10.1016/j.cell.2008.09.035

Wierzbicki AT, Ream TS, Haag JR, Pikaard CS. 2009. RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nat Genet* **41:** 630–634. doi:10.1038/ng.365

Xie Z, Johansen LK, Gustafson AM, Kasschau KD, Lellis AD, Zilberman D, Jacobsen SE, Carrington JC. 2004. Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol* **2:** e104. doi:10.1371/journal.pbio.0020104

Yang D-L, Zhang G, Tang K, Li J, Yang L, Huang H, Zhang H, Zhu JK. 2016. Dicer-independent RNA-directed DNA methylation in *Arabidopsis*. *Cell Res* **26:** 66–82. doi:10.1038/cr.2015.145

Ye R, Chen Z, Lian B, Rowley MJ, Xia N, Chai J, Li Y, He X-J, Wierzbicki AT, Qi Y. 2016. A Dicer-independent route for biogenesis of siRNAs that direct DNA methylation in *Arabidopsis*. *Mol Cell* **61:** 222–235. doi:10.1016/j.molcel.2015.11.015

Zhai J, Bischof S, Wang H, Feng S, Lee T-F, Teng C, Chen X, Park SY, Liu L, Gallego-Bartolome J, et al. 2015. A one precursor one siRNA model for Pol IV-dependent siRNA biogenesis. *Cell* **163:** 445–455. doi:10.1016/j.cell.2015.09.032

Zhang H, Ma Z-Y, Zeng L, Tanaka K, Zhang C-J, Ma J, Bai G, Wang P, Zhang S-W, Liu Z-W, et al. 2013. DTF1 is a core component of RNA-directed DNA methylation and may assist in the recruitment of Pol IV. *Proc Natl Acad Sci* **110:** 8290–8295. doi:10.1073/pnas.1300585110

Zhong X, Hale CJ, Law JA, Johnson LM, Feng S, Tu A, Jacobsen SE. 2012. DDR complex facilitates global association of RNA polymerase V to promoters and evolutionarily young transposons. *Nat Struct Mol Biol* **19:** 870–875. doi:10.1038/nsmb.2354

Zhou M, Law JA. 2015. RNA Pol IV and V in gene silencing: rebel polymerases evolving away from Pol II's rules. *Curr Opin Plant Biol* **27:** 154–164. doi:10.1016/j.pbi.2015.07.005

Zhu Y, Rowley MJ, Bohmdorfer G, Wierzbicki AT. 2013. A SWI/SNF chromatin-remodeling complex acts in noncoding RNA-mediated transcriptional silencing. *Mol Cell* **49:** 298–309. doi:10.1016/j.molcel.2012.11.011

Zilberman D, Cao X, Jacobsen SE. 2003. ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science* **299:** 716–719. doi:10.1126/science.1079695