

RNA Polymerases IV and V: evolution and function

Carlotta Porcelli, qbp693

October 16, 2017

Abstract

Plants have evolved two RNA Polymerases which are not found in other species. RNA Pol IV and Pol V are involved in the RNA-dependent DNA-methylation pathway. Pol IV initiate short non-coding RNAs (siRNAs) which are bound to proteins from the Argonaute family. Pol V transcribes long-non-coding RNAs (lncRNAs) which, through interactions with the AGO-siRNAs complexes, trigger the recruitment of DNA methyltransferase to guide *de novo* cytosine methylation on complementary DNA transcripts. Word count: 1700 words.

1 Introduction

Most of the known non-coding RNAs (ncRNAs) correspond to intergenic sequences or transcripts with unknown functions. In mammals, two of the increasingly known long ncRNAs are *Xist* and *Tsix* which are involved in the regulation of adjacent genes. In plants, small-interfering RNAs (siRNAs) guide the process of chromatin modifications. These ncRNAs are generated from specialized machinery responsible for synthesis and action of these small RNAs. Two new polymerases have been observed for the production and functional activity of these siRNAs. In flowering plants, RNA Polymerase IV and RNA Polymerase V are the central enzymes for the RNA-dependent DNA methylation pathway (RdDM). The production of siRNAs transcripts is conducted by RNA Pol IV and RNA-dependent RNA polymerase 2 (RDR2) generate short dsRNAs precursors which are then processed by dicer (DCL) enzymes into 21-24 nucleotides long siRNAs. Their association with Argonaute proteins (AGO) guide chromatin modification to homologous DNA sequences [1]. This report explores the key factors of this RdDM path-

way with an insight on the evolution and function of RNA Pol IV and RNA Pol V.

2 The paradox of epigenetic control

Transcriptional silencing happens most of the times in through methylation of CpG islands located in gene promoters. Nevertheless, several tissue-specific genes have CpG islands located downstream of transcription initiation sites. When those regions are methylated elongation is not blocked. Plants use transcription to keep a locus silent. The evolution of specialized RNA Pol IV and V allows transcription of intergenic and non-coding sequences that facilitate heterochromatin formation and silencing of overlapping and adjacent genes. "The existence of this distinct silencing polymerase may explain the paradoxical involvement of an RNA silencing pathway in maintenance of transcriptional silencing" [2].

3 RNA Polymerase IV

Pol IV is thought to initiate the RdDM pathway as it physically colocalizes with loci that undergo RdDM and it is essential for 24-nucleotide siRNA biogenesis from the vast majority of the loci. [3].

3.1 The subunits

Pol IV, as well as Pol II, is composed of 12 subunits. The largest and second largest subunits named NRPD1 and NRPD2 form the catalytic center of the enzyme. The CTD domain lack of the repeating heptad motif found in the CTD of Pol II but it shares with Pol V a single conserved domain Defective Chloroplast and Leaves (DeCL). Six

of the non-catalytic subunits of Pol IV are found in Pol II as well. The remaining four give functional specificity to the enzyme. They are positioned around the DNA entry and RNA exit channels and surround the catalytic subunits. The 4th and 7th subunits form a complex called 'Stalk' positioned near the RNA exit channel. The 5th and 9th subunits interact with the downstream DNA duplex and form the 'Jaw' complex. The subunits schema is presented in Figure 1.

3.2 Transcripts and function

Pol IV is required for the production of 24nt-siRNA in *A. thaliana* [4]. These transcripts are generated from both strands and they lack poly-A tail. On top of that they also lack 5' cap. Downstream of the Pol IV activity a RDR2 (RNA-dependent RNA Polymerase 2) is present, it processes the single stranded siRNA in double stranded siRNAs.

The first step of the biogenesis of the siRNAs is the association of Pol IV with chromatin. This process is mediated by Sawadee Homeodomain Homolog 1 (SHH1) [12] having a specific role in connection with Pol IV. The Sawadee domain recognizes H3K9 (Histone3-lysine9) methylation and adopts a specific folding that enables the interrregation of the methylation status on H3K9 and H3K4 positions. In *A. thaliana* H3K9 maps on DNA regions enriched with methylation. So SHH1 connects Pol IV with the H3K9 methylation establishing a self-reinforcing loop between repressive chromatin modifications. Pol IV transcripts are copied into double-stranded RNAs (dsRNAs) by RNA-DEPENDENT RNA POLYMERASE 2 (RDR2); this biogenesis is assisted by the chromatin remodeller CHR38, helicase domain protein. They are subsequently diced into 24-nt siRNAs by the enzyme Dicer-like 3 (DCL3), stabilized by methylation at their 3' ends by Hua Enhancer (HEN1) and loaded into Argonaute (AGO4) effector protein forming the (RNA-induced silencing complex) RISC. The whole model of action of Pol IV and Pol V is shown in Figure 2.

4 RNA Polymerase V

Nuclear RNA polymerase V (Pol V) is a multi-subunit plant-specific RNA polymerase with a role

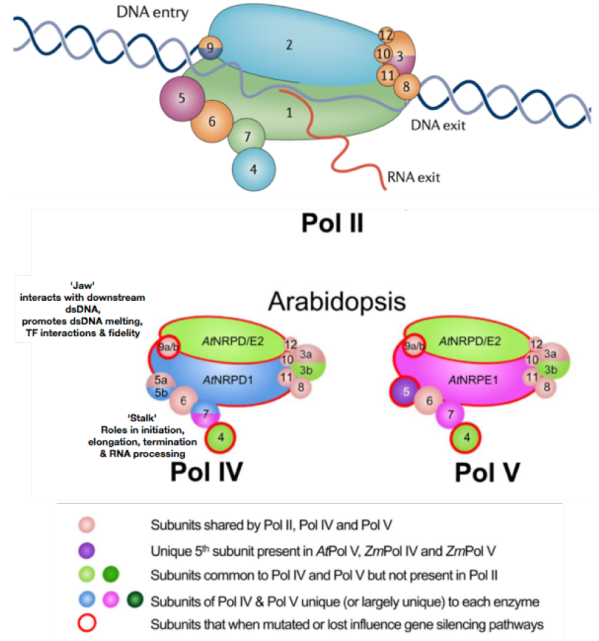


Figure 1: *Arabidopsis thaliana* Pol II, Pol IV and Pol V subunits schema. Taken from [5] and [6]

in the siRNA-directed DNA methylation (RdDM) pathway and transcriptional gene silencing [6].

4.1 The subunits

Pol V is composed of 12 subunits, half of which are encoded by the same genes as Pol II homologous subunits [6]. The two largest subunits, namely NRPE1 and NRPE2, arose from independent duplication events and sub-functionalization from their Pol II counterparts. The subunits presented below have been individuated in the *Arabidopsis thaliana* cauliflower. The NRPE1 is characterized by long a C-terminal domain (CTD) of about 700 amino acids. At the C terminus a subdomain rich in glutamine-serine (QS) motifs has been characterized. This domain is preceded by a Defective Chloroplast and Leaves (DeCL) domain with sequence similarities to chloroplast proteins involved in ribosomal RNA processing [7]. At the N-terminal of this domain ten repeats of 17 amino acids rich in glycine-tryptophan residues has been identified which coordinates AGO4 slicing of Pol V transcripts and subsequent recruitment of DRM2 methyltransferase. NRPE1, together with

Pol V second largest subunit NRPE2, form the catalytic subunits comparable to NRPE1 and NRPE2 of Pol II. The other 10 subunits do not have catalytic activity, and they characterize Pol V functions. Subunits 3, 4, 5 and 7 cluster on the 'leading face' of the polymerase enzyme and surround the catalytic center, the DNA entry and the RNA exit channels. As well as in Pol IV, the 'Stalk' and 'Jaw' complexes are found in Pol V. The subunits schema is presented in Figure 1.

4.2 Transcripts and function

A characterized region of the genome, rich in DNA methylation at intergenic loci, has been analyzed in [8] to prove the role of Pol V in targeting chromatin modifications. The chromomere on the northern arm of *A. thaliana*, on chromosome 4, is rich in heterochromatic repeats and transposons. This area has been investigated for Intergenic non-coding region (IGN). A set of six IGN transcripts has been identified to be reduced or absent in *pol v* mutants. These transcripts can be long up to 200 nt and therefore are defined as lncRNAs (long non-coding RNAs). Their initiation can happen in multiple adjacent sites. They lack poly-A tails and their 5' ends are a mixture of 5' triphosphates and monophosphates. The recruitment of Pol V to chromatin IGN producing loci has been shown to depend on two components: Defecting in RNA-dependent DNA methylation 1 (DRD1) and Defective in Meristem silencing 3 (DMS3) [8]. Moreover a new component RNA-directed DNA METHYLATION 1 (RDM1) has been identified to stably associate with DRD1 and DMS3. These three components form a complex termed DDR that facilitate Pol V association with chromatin via a direct interaction [9]. On top of this, other two methyl-DNA binding proteins (SU(VAR)3-9 homolog 2: SUVH2 and SUVH9) have been found interacting with the DDR complex [10] and required for Pol V chromatin association. These interactions lead to formulate a model in which Pol V recruitment depends on pre-existing DNA methylation bound by SUVH2 and SUVH9 which then interact with the DDR complex that promotes Pol V association. The siRNAs produced by Pol IV, associated with AGO4 proteins, guide cytosine methylation through base-pairing interaction with the nascent lncRNAs transcripts of Pol V. [1] This interaction facilitate the recruitment of

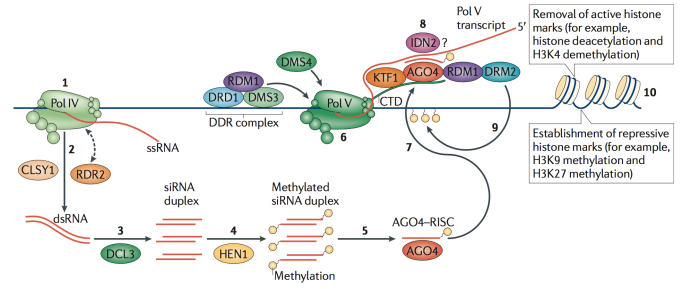


Figure 2: Model for the RNA-directed DNA methylation pathway in *Arabidopsis thaliana*. Taken from [5]

Domains rearranged methyltransferase 2 (DRM2) and Chromatin Remodelling 28 and 28 (CHR27 / CHR28) that regulate the activity of Pol V and facilitate *de novo* DNA methylation. As shown in [11] in the C-terminal domain (CTD) of Pol V largest subunit NRPE1, and its associated factor SPT5L (a Pol V auxiliary protein), is present a conserved domain of glycine-tryptophan/tryptophan-glycine (GW/WG) region. Its function is shown to be essential to hook AGO proteins and facilitate the bond of their siRNA sequences to the DNA chromatin targets. These motifs recruit a massive amount of AGO4 siRNA loaded complexes to their site of action. The increase of local concentration of AGO4 is a consequence of the stabilization of Pol V to the DNA strand. The model proposed sees Pol V proceeding through transcription elongation, AGO4 interacting with the lncRNAs Pol V-transcripts generating an intermediate complex that would serve for the transfer of AGO4 to the DNA template. At this stage AGO4 recruits DRM2 (DNA (cytosine-5)-methyltransferase) to the opposite DNA strand to trigger DNA methylation. The whole model of action of Pol IV and Pol V is shown in Figure 2.

5 Future outlook

It is still unclear what templates the two polymerases use *in vivo* and how the transcription is regulated in order not to produce aberrant non-coding RNAs. Further studies could be led to investigate both the transcription initiation and the role of the RNA primers in that and the termina-

tion steps.

5.1 Pol IV

The recruitment of the Pol IV to targets different from H3K9 methylated sites is still unclear. Moreover the SHH1 factor, needed for Pol IV association with chromatin, recognizes H3K9 methylation and unmethylated H3K4 on the loci where the Pol IV downstream is going to be recruited. [12] It is still not clear how the histone methylation are initially established in the specific loci.

5.2 Pol V

The association of Pol V to the chromatin involves seems to involve more factors than the few one discussed in this work, as shown in [13]. Nevertheless their specific function is still unclear and could be elucidated in further studies. Not much is known on the nature and function of the IGN Pol V transcripts and this could be a case of future studies as well.

References

1. Wierzbicki, A.T., Ream, T., Haag, J.R., and Pikaard, C.S. (2009). RNA Polymerase V transcription guides ARGONAUTE4 to chromatin. *Nature genetics* *41*, 630–634. URL <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2674513/>.
2. Herr, A.J., Jensen, M.B., Dalmay, T., and Baulcombe, D.C. (2005). RNA Polymerase IV Directs Silencing of Endogenous DNA. *Science* *308*, 118–120. URL <http://science.sciencemag.org/content/308/5718/118>.
3. Pontes, O., Li, C.F., Nunes, P.C., Haag, J., Ream, T., Vitins, A., Jacobsen, S.E., and Pikaard, C.S. The *jem* Arabidopsis *em* Chromatin-Modifying Nuclear siRNA Pathway Involves a Nucleolar RNA Processing Center. *Cell* *126*, 79–92. URL <http://dx.doi.org/10.1016/j.cell.2006.05.031>.
4. Zhang, X., Henderson, I.R., Lu, C., Green, P.J., and Jacobsen, S.E. (2007). Role of RNA polymerase IV in plant small RNA metabolism. *Proceedings of the National Academy of Sciences* *104*, 4536–4541. URL <http://www.pnas.org/content/104/11/4536.abstract>.
5. Haag, J.R. and Pikaard, C.S. (2011). Multi-subunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. *Nat Rev Mol Cell Biol* *12*, 483–492. URL <http://dx.doi.org/10.1038/nrm3152>.
6. Zhou, M. and Law, J.A. (2015). RNA Pol IV and V in Gene Silencing: Rebel Polymerases Evolving Away From Pol II's Rules. *Current opinion in plant biology* *27*, 154–164. URL <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4618083/>.
7. Huang, Y., Kendall, T., Forsythe, E.S., Dorantes-Acosta, A., Li, S., Caballero-Pérez, J., Chen, X., Arteaga-Vázquez, M., Beilstein, M.A., and Mosher, R.A. (2015). Ancient Origin and Recent Innovations of RNA Polymerase IV and V. *Molecular Biology and Evolution* *32*, 1788–1799. URL <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4476159/>.
8. Wierzbicki, A.T., Haag, J.R., and Pikaard, C.S. (2008). Noncoding Transcription by RNA Polymerase Pol IVb/Pol V Mediates Transcriptional Silencing of Overlapping and Adjacent Genes. *Cell* *135*, 635 – 648. URL <http://www.sciencedirect.com/science/article/pii/S0092867408011926>.
9. Law, J.A., Ausin, I., Johnson, L.M., Vashisht, A.A., Zhu, J.K., Wohlschlegel, J.A., and Jacobsen, S.E. (2010). A protein complex required for polymerase V transcripts and RNA-directed DNA methylation in plants. *Current biology : CB* *20*, 951–956. URL <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2972704/>.
10. Johnson, L.M., Du, J., Hale, C.J., Bischof, S., Feng, S., Chodavarapu, R.K., Zhong, X., Marson, G., Pellegrini, M., Segal, D.J., et al. (2014). SRA/SET domain-containing proteins link RNA polymerase V occupancy to DNA methylation. *Nature* *507*, 124–128. URL <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3963826/>.
11. Lahmy, S., Pontier, D., Bies-Etheve, N., Laudie, M., Feng, S., Jobet, E., Hale, C.J.,

- Cooke, R., Hakimi, M.A., Angelov, D., et al. (2016). Evidence for ARGONAUTE4–DNA interactions in RNA-directed DNA methylation in plants. *Genes & Development* *30*, 2565–2570. URL <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5204349/>.
12. Law, J.A., Du, J., Hale, C.J., Feng, S., Krajewski, K., Palanca, A.M.S., Strahl, B.D., Patel, D.J., and Jacobsen, S.E. (2013). Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. *Nature* *498*, 385–389. URL <http://dx.doi.org/10.1038/nature12178>.
 13. Stroud, H., Greenberg, M.V.C., Feng, S., Bernatavichute, Y.V., and Jacobsen, S.E. Comprehensive Analysis of Silencing Mutants Reveals Complex Regulation of the *Arabidopsis* Methylome. *Cell* *152*, 352–364. URL <http://dx.doi.org/10.1016/j.cell.2012.10.054>.