siRNA-mediated gene silencing in plants

Introduction

We have seen that small interfering RNAs (siRNAs) act in the biological gene-silencing process called RNA interference (RNAi). The stimulus that triggers RNAi is the presence of a long double-stranded RNA (dsRNS) in the cell, which was introduced artificially into organisms such as plants, worms, or flies using injection or transgenic expression. However, it is clear that RNAi does not exist in cells to be able to silence transgenes or experimental dsRNA. siRNAs were observed during transgene- and virus-induced silencing in plants, which is consistent with their role in the defense of the genome against foreign or invasive nucleic acids, such as viruses or transposons. However, are there also endogenous siRNAs? If so, what is their origin?

Comparing siRNAs and miRNAs

We have already discussed the similarities and differences between siRNAs and miRNAs. Indeed, they are similar in both their biogenesis and activity in the cell; however, there are a few key differences between them. Let's recall the main points they have in common and the main characteristic that distinguishes these molecules.

siRNAs induce the degradation of homologous mRNAs in a similar way miRNAs that are perfectly matched to target transcripts would do, i.e., via "slicing" followed by exonuclease-mediated degradation of the cleaved mRNA fragment. Thus, to some extent, siRNAs and miRNAs share a similar machinery to exert their action. In RNAi, the long dsRNA is converted into a population of siRNAs of about 20-25 nucleotides in length by the enzyme Dicer. The siRNAs are incorporated into AGO proteins as part of a RISC involved in target RNA recognition. siRNA-mediated gene silencing is mainly associated with RNA degradation by slicing; however, recent findings indicate that they may also function in translational inhibition as well as in other sequence-specific gene-silencing mechanisms (which we will discuss later).

Although siRNAs and miRNAs seem to be very similar in both their biogenesis and activity in the cell, the main criterion to differentiate these molecules is the following: miRNAs are excised by Dicer from their short (roughly 70 nucleotides long in mammals) and imperfect stem-loop precursor as a **single small RNA species**, which may accumulate to tremendous levels in the cell (up to 50'000 molecules/cell). siRNAs, on the other hand, are produced by consecutive cuts by Dicer along a long perfectly double-stranded dsRNA precursor. Therefore, siRNAs, unlike miRNAs, **are always part of a population** that is distributed along their long dsRNA precursor.

Experimental evidence for endogenous siRNAs

When researchers went out to define the repertoire of small RNAs in different organisms, some interesting findings were made. To identify small RNAs, they used deep sequencing, which not only allows determining the sequence but also the size of an RNA. Deep sequencing of *Arabidopsis* showed that the majority of small RNAs present in plant cells are actually siRNAs. By far the largest part of the RNAs discovered were 24 nucleotides long (see figure 5-1). These 24-nucleotide-long RNAs could not be miRNAs, because a similar distribution pattern was found in *Arabidopsis* mutant for DCL1 (remember that DCL1 is the Dicer homolog in plants that produces miRNAs).

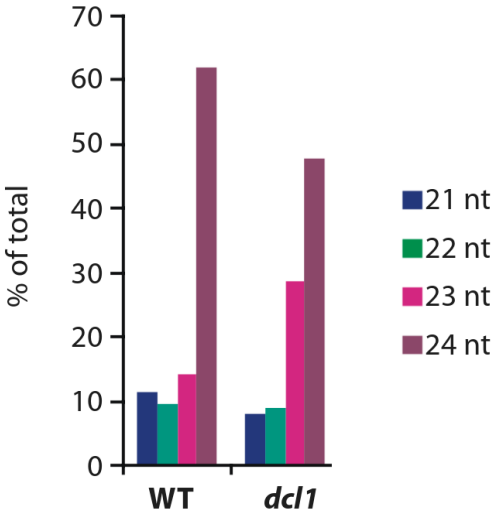

Figure 5-1 Distribution of small RNAs in Arabidopsis as detected in deep RNA sequencing. The small RNAs of 24 nucleotides length constitute the largest class within the total small RNAs detected. These RNAs are not miRNAs, because plants mutant for DCL1 show a similar distribution. (Adapted from K.D. Kasschau *et al.*, *PLOS,* 2007)

What are all these small, 24-nucleotide long RNAs if they are not produced by DCL1 and are therefore not miRNAs? *Arabidopsis* has four Dicer genes, named *dcl1-4*. As mutations exist in each dcl gene, the products made by each of the Dicers could be assigned very precisely. Except for DCL1, which produces 20- to 24-nucleotide-long miRNAs, every DCL produces small RNAs of a very specific size. DCL2 produces short RNAs of 22 nucleotides length, DCL3 produces short RNAs of 24 nucleotides length and DCL4 produces short RNAs of 21 nucleotides length.

Thus, it was clear that the large majority of small RNAs present in plant cells must be 24-nucleotides long siRNAs produced by DCL3. However, the plants used for the RNA analysis were healthy, neither transfected nor infected with viruses. This means that the small RNAs extracted from these plants must represent endogenous, naturally occurring small RNAs that exist in the absence of an exogenous inducer. Thus, endogenous RNAi-silencing activity may exist for siRNAs, similar to the activity of miRNA genes.

Transposable elements as an origin of endogenous siRNAs

If endogenous RNAi exists, where do the endogenous siRNAs come from? An answer to this was found when the small RNAs detected in the deep-sequencing analysis described above were mapped onto the *Arabidopsis* reference genome. Interestingly, the 24-nucleotide-long siRNAs could be mapped to genomic sites containing transposons and retrotransposons (see figure 5-2A). The researchers then determined how the small RNA-generating loci are distributed on the chromosomes. In line with the first observation, for each chromosome, the density of small RNA-generating loci was the highest at the centromere regions and the regions very close to the centromere. These regions are known to contain a high density of repeated elements, including transposons and retrotransposons. Most of the small RNA-generating loci were mapped to these regions (see figure 5-2B).

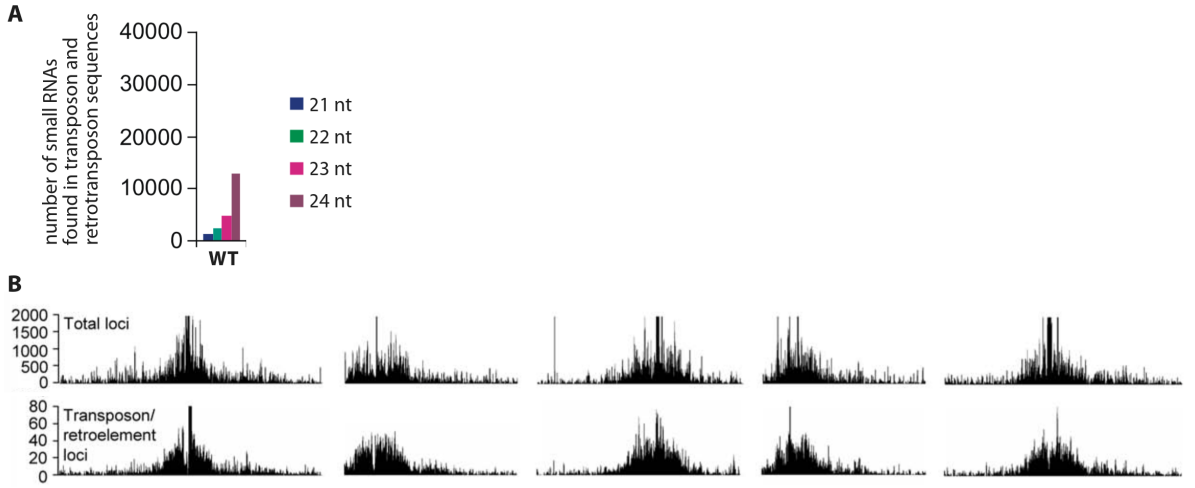

Figure 5-2 The small RNAs of 24 nucleotides length are highly correlated with repetitive elements in the Arabidopsis genome. (A) The small RNAs of 24 nucleotides lengths often map to transposons or retrotransposon sequences within the genome. **(B)** Analysis of small RNA loci on the chromosomes. The upper panel shows all small RNA loci and the lower panel only those that map to transposons and retrotransposons. In both cases, the small RNA loci are located at the centromeric region of each chromosome (centromeres are indicated by the grey ellipse). (Adapted from K.D. Kasschau *et al., PLOS*, 2007)

Role of endogenous siRNAs

Thus, it was clear that most of the DLC3-dependent, 24-nucleotide-long siRNAs originate from regions at or around the centromeres that are very rich in transposable elements (TE or transposons). These regions are enriched in condensed chromatin called heterochromatin and are poorly transcribed by RNA polymerase II. This is in contrast to euchromatic regions on the chromosome arms, which are rich in protein-coding genes and highly transcribed by RNA polymerase II.

There are also 24 nucleotide long siRNAs that are found in euchromatic regions of the chromosome arms. Interestingly, these also map to TEs. They derive from these TEs as populations made from long dsRNA precursor molecules. Remember that many TEs form inverted repeat loci, the transcription of which can lead to self-complementary dsRNA. Other mechanisms exist to generate dsRNA from transposons, which we will discuss in the high-level lecture. Independent of the way it is made, transposon-derived dsRNAs is then converted into siRNAs by DLC3 (and to a lower extent, by DLC2). These 24 nt siRNAs are then loaded mostly into AGO4, which belongs to a specific clade of so called "heterochromatic" AGO proteins (AGO4/6/9). Remarkably, siRNAs loaded onto these AGO proteins anneal to nascent transcripts of the targeted DNA (that is, the DNA of the transposon) and recruit *de novo* methyl-transferase to add methyl groups to the cytosines in the target DNA in an siRNA-complementary, and thus sequence-specific manner, resulting in silencing of the TE. As we have discussed in the reverse genetics part already, the mobilization of TEs can cause gene disruptions and cause chromosomal breakage; thus, cells must have evolved strategies to protect their genomic integrity by guarding their DNA from the activity of mobile elements. One of these strategies is, as we have just seen, to silence TEs by RNA-directed DNA methylation (RdDM). RdDM therefore is another mechanism how siRNAs can silence genes – not by mRNA degradation on a post-transcriptional level, but by epigenetically modifying the DNA to reduce its transcription. Gene silencing mechanisms acting on the level of transcriptional repression are collectively termed transcriptional gene silencing (TGS). We will discuss RdDM and TGS in more detail in the high level lecture.

Role of exogenous siRNAs in plants: antiviral silencing

The results from deep RNA sequencing performed in different plant groups showed that endogenous siRNA products of DCL2 (22 nucleotides) and DCL4 (21 nucleotides) are very underrepresented in healthy plants, raising questions about the cellular functions of these two enzymes. However, in

plants infected with viruses, these siRNA species dramatically increase and are of the most abundant ones. Interestingly, these siRNAs do not map onto the plant genome, but rather onto the genome of the virus used for infection. Furthermore, plants lacking DLC4 and DCL2 function show increased susceptibility to plant viruses; suggesting that the virus-derived siRNA products of DCL4 and DCL2 mediate antiviral defense. Indeed, it was shown that DCL4 and DCL2 act redundantly against all RNA and DNA viruses tested so far.

The mechanism by which virus infections lead to the production of siRNAs can be explained the easiest by the replication cycle of an RNA virus (most plant viruses have an RNA genome). The 21- and 22-nucleotide-long siRNAs derive from a long dsRNA precursor that forms by replicating the genomic positive, single-sense (+) RNA strand of the virus into a complementary negative-sense (-) RNA (see figure 5-3). This dsRNA precursor, called the replication form (RF), is recognized by the host-plant-encoded DCL2 and DCL4 enzymes and cut into siRNAs, which can be loaded onto AGO proteins to mediate gene silencing by degrading the viral mRNA (+ strand). DCL4 is the major enzyme in antiviral silencing, and is seconded by its surrogate, DCL2. The major AGO proteins that mediate antiviral RNAi are AGO1 and AGO2. Note that AGO1 is also the main effector of most plant miRNAs.

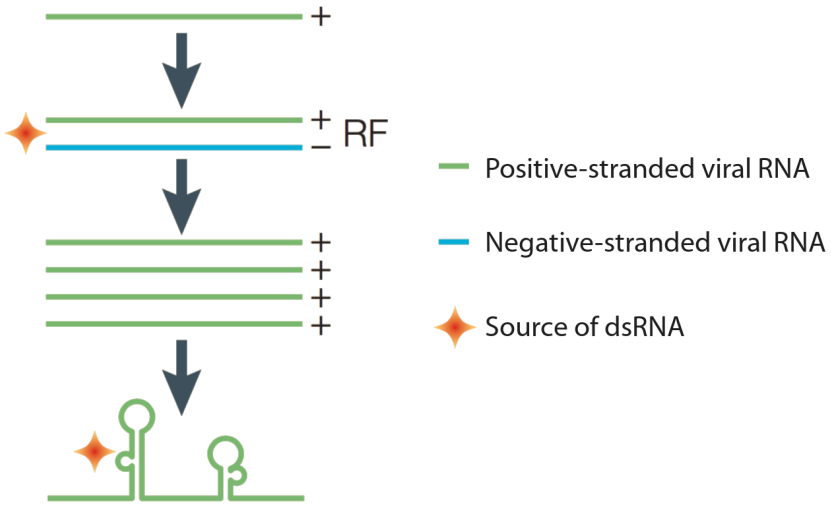

Figure 5-3 Molecules serving for RNA silencing are produced during the replication of viruses. Most plant viruses, and some animal viruses, have genomes of positive, ssRNA that are replicated within the cytoplasm of the host. Viral RNA-dependent RNA polymerases synthesize complementary negative-stranded (-) genomic RNA, from which numerous copies of positive-stranded (+) RNA are reproduced. These (+) strands are used for protein synthesis. Partial or complete annealing of (+) and (-) RNA strands provides one source of dsRNA. A second source is provided by the folding of replicated, single-stranded genomic (+) RNA, which forms secondary double-stranded structures. (Adapted from O. Voinnet, *Nat Rev Genet*, 2005)

The question remains how the RNA-silencing machinery of the host can keep up with the rapid replication pace of the virus. As you can see in figure 5-3, the viral (-) strand is copied to produce multiple (+) strands, leading to a multifold amplification of the (+) strands. These (+) strands serve for protein synthesis. Thus, the plant RNA-silencing machinery faces many RNAs to be targeted for destruction, and to do this, plants encode RNA-dependent RNA polymerases (RDR) (which, interestingly, have also been found in *C. elegans*). The host-encoded RDRs can amplify the antiviral silencing action by generating dsRNA *de novo* from virus single-stranded RNA substrates. Here, the cleaved RNA fragments resulting from AGO1/AGO2-mediated slicing serve as templates to synthesize new dsRNA that again can be processed by DCL4 and DCL2, loaded onto AGO1/AGO2 and mediate

silencing of more viral RNAs (see figure 5-4). The multiplication of antiviral Dicers in plants and other organisms can be seen as the result of an arms race between hosts and their parasites.

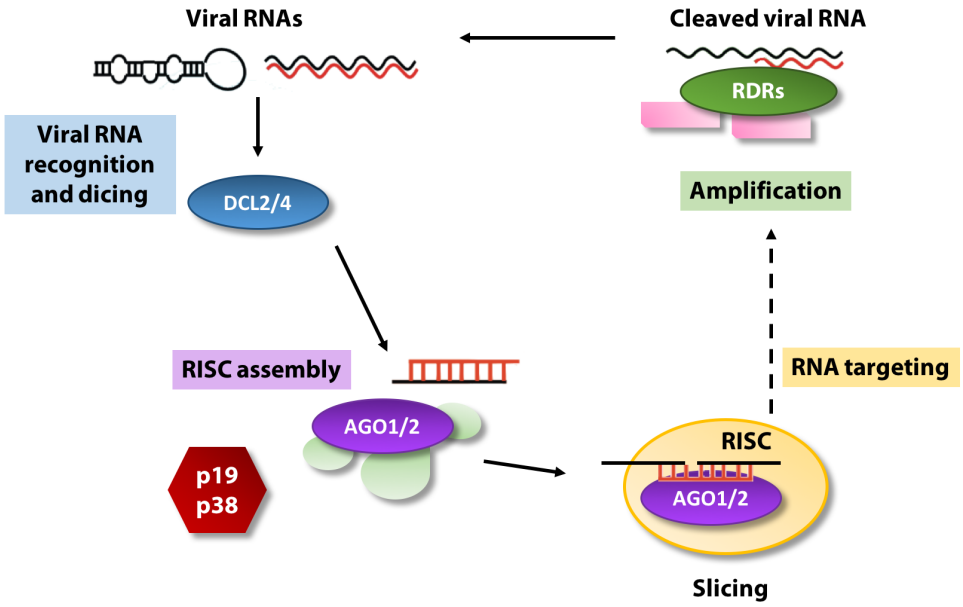

Figure 5-4 Current model of virus-induced gene silencing (VIGS) in plants. RNA silencing is initiated by the recognition of viral dsRNAs or dsRNAs with a partial hairpin, which are processed to siRNAs by DCL2 and DCL4. In the next step, AGO1/2 proteins are loaded with siRNA, thereby forming large RISCs, which probably also incorporate other unidentified proteins (green). Afterwards, the siRNA-loaded RISC targets viral RNAs by slicing or translational inhibition. Secondary siRNAs are produced in an amplification loop through the actions of host-encoded RNA-dependent RNA polymerases (RDRs) and their cofactors (pink). The viral-silencing suppressors (VSR) p19 and p38 interact with the silencing pathways (red hexagon). (Adapted from J. Burgyán and Z. Havelda, *Trends Plant Sci*, 2011)

As you can imagine, viruses have evolved counter-defense strategies to escape virus-induced gene silencing (VIGS). These viral suppressors of RNA silencing (VSR) are proteins encoded by the virus that interfere at different steps with the silencing pathway. We will briefly discuss two strategies that are used by viruses to prevent RISC assembly.

The p19 protein encoded by tombusviruses prevents RNA silencing by binding siRNA with a high affinity, thereby sequestering it. The protein functions as homodimer which is able to bind siRNAs in a size-dependent, sequence-independent manner, acting like a molecular caliper by measuring the length of siRNA duplexes and binding them (see figure 5-5). p19 displays extremely high affinity for 21-bp siRNA duplexes, which are precisely the products of DCL4, the main plant antiviral Dicer. By binding siRNA duplexes, p19 prevents the loading of dsRNAs onto the AGO1/AGO2 protein and thus blocks antiviral silencing.

In a second strategy, the p38 viral protein (which is also part of the viral capsid) binds AGO1/AGO2 and neutralizes their ability to load virus-derived siRNAs.

The antiviral RNAi defense and coordinated production of VSR by viruses are not only found in plants but also in invertebrates including *C. elegans* and *Drosophila*. Antiviral RNAi is a remarkable process, since it is entirely innate: it is not programmed by the host, but by structural (i.e., dsRNA) and nucleotide sequence features of the pathogen's RNA itself. As such, this mechanism can be virtually adapted to any virus.

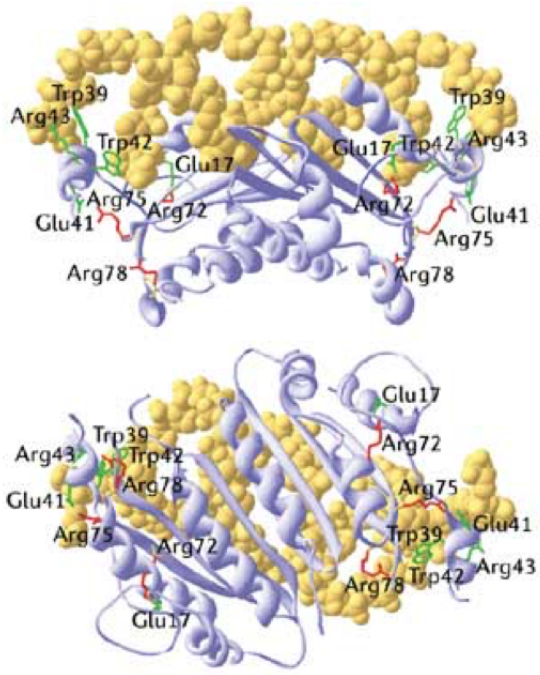

Figure 5-5 Crystal structure of the p19 protein from tombusviruses. The two monomeric p19 units (purple) bind the helical 21-bp siRNA duplex (yellow double helix). The protein-dimer arrangement with the N-terminal portions positioned in opposite directions generate an intramolecular groove to position a dsRNA helix. The protein conformation also results in the surface exposure of several charged or polar residues in the groove that can serve as contact sites for interactions with the dsRNA (residues are indicated). The electrostatic association between p19 and the dsRNA occurs through the sugar-phosphate backbone of the dsRNA. This explains why p19 can bind any 21-bp-long dsRNA, independent of its sequence, and also ensures that dsRNA, but not ssRNA, is bound by p19. The high affinity of p19 for 21-bp siRNAs, rather than longer or shorter molecules, is elegantly orchestrated by tryptophan residues at positions 39 and 42 of each monomer that function as molecular caliper. These residues are outwardly projected and form a hydrogen bond with a 5'-phosphorylated nucleotide on each end of the duplex siRNA. Thereby, the calipers measure the siRNA, resulting in a high affinity for dsRNA molecules of 21 bp. (From H.B. Scholthof, *Nat Rev Microbiol*, 2006)

Systemic transgene RNAi

We have just seen that antiviral RNAi includes an amplification of the RNAi response, which can be seen as a sequence-specific immune system required to keep pace with high viral replication rates, because a few viral RNAs can be converted into many dsRNA molecules that reinforce the silencing system. The idea of the existence of a sequence-specific immune system was reinforced when it was realized that in plants, RNAi induced locally in response to long dsRNA or viruses can move between cells and even over long distances throughout plants. How was this discovery made?

In an experiment, long dsRNA for GFP was locally delivered into only one of the leaves of tobacco plants that constitutively express a GFP transgene. Interestingly, the GFP transgene was not only silenced in this particular part of the leaf, but the treatment also caused the GFP transgene to be silenced over time within the whole plant (see figure 5-6), i.e., systemically. This effect is sequence-specific, because systemic GFP silencing is not initiated by a dsRNA carrying no sequence homology to GFP.

This experiment demonstrates the existence of a systemic silencing signal that gets amplified as it moves away from its sites of initiation. Genetic experiments conducted in *Arabidopsis* have shown that signal amplification occurs through the same RNA-dependent RNA polymerase (RDR6) that also mediates the amplification of antiviral RNAi. Further, these experiments showed that the nucleic-acid component of the signal is made of siRNA duplexes that move independently of AGO proteins between the cells (via plasmodesmata) and through the vasculature (via the phloem) of the plant. Interestingly, these routes are the same that are taken by viruses to spread within plants. This artificial system of non-cell autonomous silencing was found to underpin a systemic immune system that allows plant tissues to become immunized against viruses, ahead of the infection, in a sequence-specific manner. Such a systemic RNAi response has also been reported to exist in *C. elegans*, and may exist in other organisms, too.

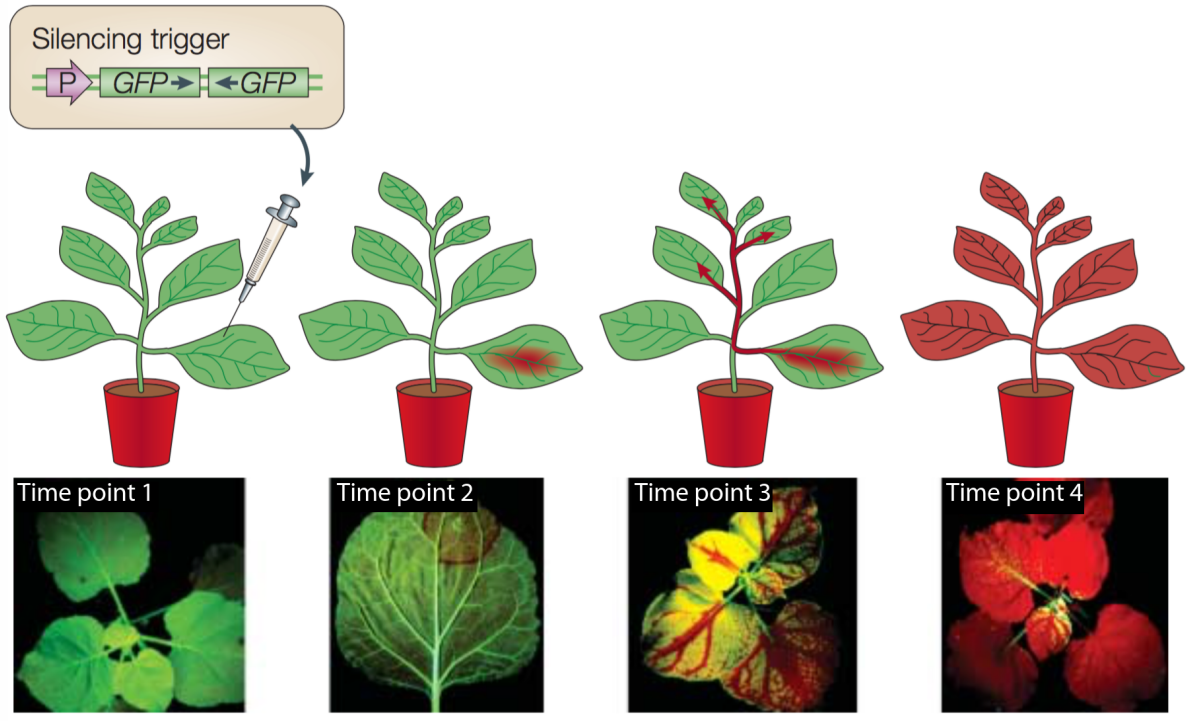

Figure 5-6 Green fluorescence of plants injected with dsRNA against GFP as a silencing trigger. The plants are illuminated with UV light to excite GFP. Just after injection, the plant still fluoresces green due to the presence of GFP protein. Silencing is manifested as the appearance of red tissue under UV light, since chlorophyll fluoresces red when excited with UV light. About three weeks after the local delivery of small amounts of GFP dsRNA into a single leaf, all of the GFP is silenced throughout the plant. (Adapted from O. Voinnet, *Nat Rev Genet*, 2005)

Summary of the role of DCL proteins in gene silencing in plants

In summary, let's look at this table below to remember the most important facts about the role of DCL proteins in gene silencing in plants:

Table 5-1 Functions of the different DCL proteins in small RNA gene silencing pathways in plants.

DCL protein	Localization	Silencing mode	Size of effector sRNAs	Primary function
DCL1	Nucleus	• Processing of miRNA from stem-loop precursors	20-24 nt	• Post-transcriptional silencing (PTGS) of endogenous mRNA by slicing and/or translational inhibition
DCL2	Nucleus and cytoplasm; acts as a surrogate of DCL4 and DCL3	• Processing of siRNAs from viral dsRNA or from inverted-repeat transgenes used for experimental RNAi • Processing of siRNAs from transposon-derived dsRNA	22 nt	• Virus-induced gene silencing (VIGS) • Transposon epigenetic silencing
DCL3	Nucleus	• Processing of siRNAs from transposon-derived dsRNA	24 nt	• Epigenetic modification of transposon DNA and other endogenous loci (TGS)
DCL4	Nucleus and cytoplasm	• Processing of siRNAs from viral dsRNA or from inverted-repeat transgenes used for experimental RNAi	21 nt	• Virus-induced gene silencing (VIGS)

Summary

In this lesson, we have discussed how siRNAs can mediate gene silencing in plants. We have seen that they can mediate silencing of transposable elements (and other repetitive regions within the genome) by a process called transcriptional gene silencing (TGS) via DNA modifications and chromatin remodeling. We will come back to this interesting function in the high-level lecture where we discuss the mechanisms and biological roles of this kind of silencing in more detail. Next, we have discussed how siRNAs function in antiviral silencing (virus-induced gene silencing, VIGS) and how this mechanism is exploited by plants to build an innate immune system based on RNAi amplification by specialized enzymes called RDRs. In turn, we have seen that viruses have evolved VSRs to counteract the antiviral RNAi reaction.

Figure 5-7 summarizes the main functions of small RNAs in plants we have discussed so far.

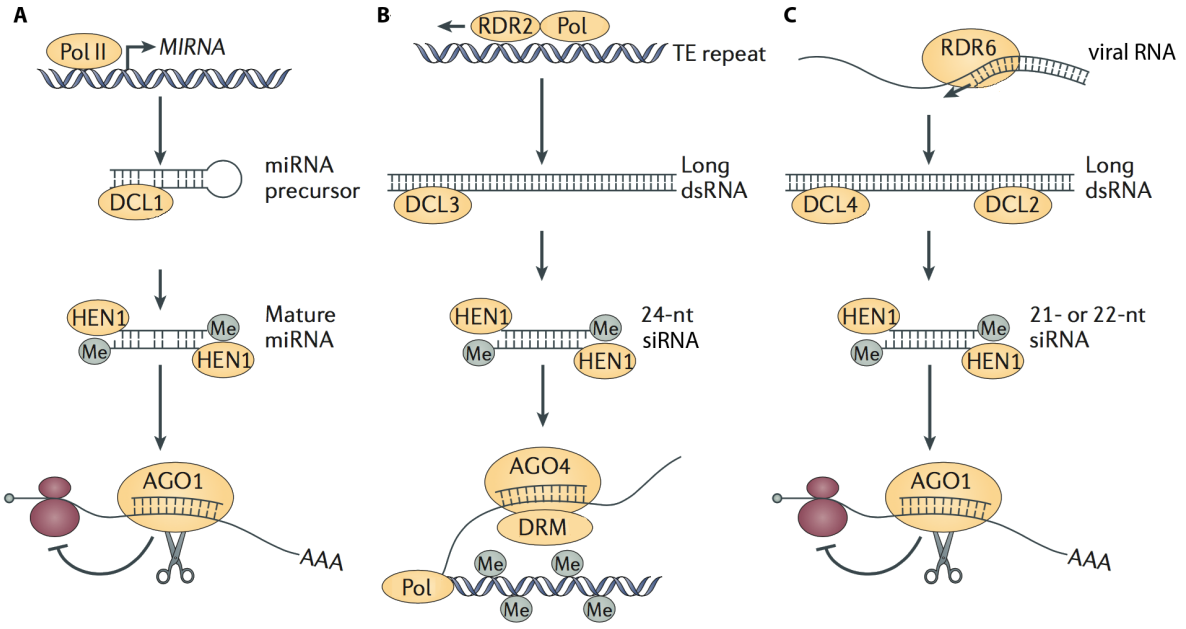

Figure 5-7 Endogenous gene-silencing pathways in plants. RNA silencing is triggered by dsRNA that can arise from various endogenous sources. All dsRNAs are cleaved by DCL proteins to produce mature small RNAs that are methylated (indicated by Me) by HEN1 to be protected from degradation. All processed, methylated small RNAS are incorporated into AGO proteins to form an RNA-induced silencing complex (RISC). RISCs can then target different mRNAs to silence gene transcription or induce transcriptional gene silencing (TGS) by DNA methylation and chromatin modification at target loci. **(A)** The miRNA pathway: Transcription of miRNA genes results in imperfectly matched hairpins that are processed by DCL1 into mature miRNA duplexes. One strand of the duplex is incorporated into AGO1 to form RISCs that target mRNAs for slicing (indicated by the scissors) or inhibit their translation (indicated by blocking ribosomes, red circles). **(B)** The siRNA pathway to silence transposable elements: RNA transcribed from transposable elements can be converted into dsRNA by RNA-dependent RNA polymerases (encoded by the cellular genome). This dsRNA serves as substrate for DCL3 to produce siRNAs that, when incorporated into AGO4, direct RISCs to specific genomic loci to induce transcriptional silencing by DNA and chromatin modifications by recruiting a methyl transferase (DRM). **(C)** The antiviral siRNA pathway: RNA produced by viruses can be converted into dsRNA by host-encoded RNA-dependent RNA polymerases. These dsRNAs are processed by DCL4 and DCL2 to produce siRNAs that are incorporated into AGO1 to form RISCs targeting mRNAs for slicing or translational inhibition. (Adapted from N. Pumplin and O. Voinnet, *Nat Rev Microbiol*, 2013)